# Plasma protein-based organ-specific aging and mortality models unveil diseases as accelerated aging of organismal systems

Ludger J.E. Goeminne [1,4], Anastasiya Vladimirova [1,4], Alec Eames [1,4], Alexander Tyshkovskiy [1,4], M. Austin Argentieri [2,3,4], Kejun Ying [1,4], Mahdi Moqri [1,4], Vadim N. Gladyshev [1,4,*]

[1]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

[2]Department of Medicine, Massachusetts General Hospital, Boston, MA, USA

[3]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA

[4]Broad Institute of MIT and Harvard, Cambridge, MA, USA

[*]Lead contact

**This is the accepted manuscript; when referring to this article, please always cite the original source at https://doi.org/10.1016/j.cmet.2024.10.005.**

**© 2024. This manuscript version is made available under the CC-BY-NC-ND 4.0 license https://creativecommons.org/licenses/by-nc-nd/4.0/**

## In brief

This study develops organ-specific aging models using blood proteomics data from 53,000 UK Biobank participants. These models predict organ-specific diseases and risk of death and reveal that chronic diseases reflect faster aging in specific organs. Different lifestyles affect organ aging differently.

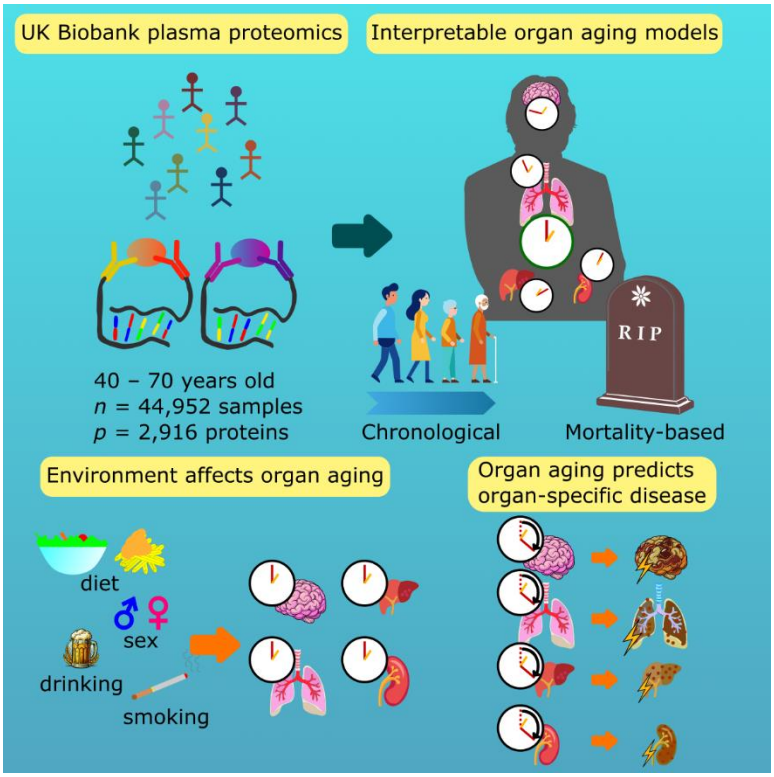
## Highlights

- Organ-specific aging models predict mortality and organ-specific diseases
- Interpretable models reveal proteins that drive links between organ aging and disease
- Lifestyle (diet, occupation, and medication) and sex are linked with organ-specific aging
- Organ-specific aging models suggest that chronic diseases represent (sub)system aging

## Summary

Aging is a complex process manifesting at molecular, cellular, organ, and organismal levels. It leads to functional decline, disease, and ultimately death, but the relationship between these fundamental biomedical features remains elusive. By applying elastic net regularization to plasma proteome data of over 50,000 human subjects in the UK Biobank and other cohorts, we report interpretable organ-specific and conventional aging models trained on chronological age, mortality, and longitudinal proteome data. These models predict organ/system-specific disease and indicate that men age faster than women in most organs. Accelerated organ aging leads to diseases in these organs, and specific diets, lifestyles, professions, and medications influence organ aging rates. We then identify proteins driving these associations with organ-specific aging. Our analyses reveal that age-related chronic diseases epitomize accelerated organ- and system-specific aging, modifiable through environmental factors, advocating for both universal whole-organism and personalized organ/system-specific anti-aging interventions.

## Graphical abstract




## Keywords

organ-specific aging, proteomic clocks, aging models, mortality, disease, blood plasma, longevity interventions, elastic net, diet, lifestyle

## Introduction

Aging is a highly complex process that manifests across multiple scales, from intracellular molecular changes to functional decline at the tissue, organ, and organismal levels, resulting in disease and increased mortality.[1] However, how aging mechanisms across various scales interconnect and contribute to the development of age-related diseases and other outcomes remains poorly understood.

Molecular aging clocks, initially developed based on DNA methylation patterns,[2] have extended to other omics, like the transcriptome[3,4] and proteome.[5,6,7] These tools estimate biological age, "an individual's age defined by the level of age-dependent biological changes, such as molecular and cellular damage accumulation." It may practically be "summarized as a number (in units of time) matching the chronological age where the average person in a reference population shares the individual's level of age-dependent biological changes."[8] Biological age can deviate from chronological age, and this age deviation can serve as a biomarker for age-related diseases and mortality risk.[8] Clocks trained on different tissues and organs can produce strikingly divergent outcomes, implying different organs age at distinct rates.[9]

Recently, Oh et al. demonstrated that individual organs age at different rates across human populations using plasma proteomes in cohorts encompassing 5,676 individuals.[10] However, larger and more extensively phenotyped population-based cohorts, like the UK Biobank, provide an opportunity to even better characterize omics-based organ-specific signatures of aging and investigate the effects of lifestyle, diet, and therapeutic interventions on health outcomes.[11] They also enable the development of robust proteome-based biomarkers to assess biological age.[6,7] However, there remains a significant gap in understanding organ-specific aging processes and their implications for disease prediction and intervention. The Olink Explore 3072 platform, a high-throughput affinity-based proteomics platform that measures protein levels of ~3,000 proteins simultaneously, has recently been applied to the plasma of >53,000 UK Biobank participants.[12]

In this study, we leveraged UK Biobank data to develop organ-specific aging models. We identified plasma proteomic signatures that reflect differential organ aging rates and trained biomarkers for chronological age and mortality. The interpretability of our models allows us

to assess the similarities and differences between the effects of chronological aging and various forms of accelerated organ-specific aging. Their application to various phenotypes offers numerous insights into human organ-level aging and suggests that chronic diseases generally manifest as accelerated organ- and system-specific aging, modifiable through lifestyle interventions, such as diets and medications.

# Results

## Plasma proteome supports robust chronological aging models

We used 5-fold cross-validation, randomly splitting the UK Biobank dataset (*n* = 44,952 participants) into five almost equal-sized, non-overlapping test sets, and trained elastic net models on 2,916 plasma proteins for the ~80% remaining participants to predict chronological age. These conventional 1$^{st}$-generation aging models were consistent, with non-zero coefficients for 2,288–2,315 proteins across folds (Tables S1A and S1B). Predicted age strongly correlates with chronological age in the training (Figure 1A; correlation coefficient (r), 0.94; mean absolute error (MAE) of the residuals, 1.99 years) and test datasets (Figure 1B; r, 0.94; MAE, 2.10 years) of the first fold and in out-of-fold predictions (r, 0.94; MAE, 2.10 years; Figures S1A and S1B; Table S2A), performing on par with gold standard epigenetic[13] and proteomic[6] aging models.

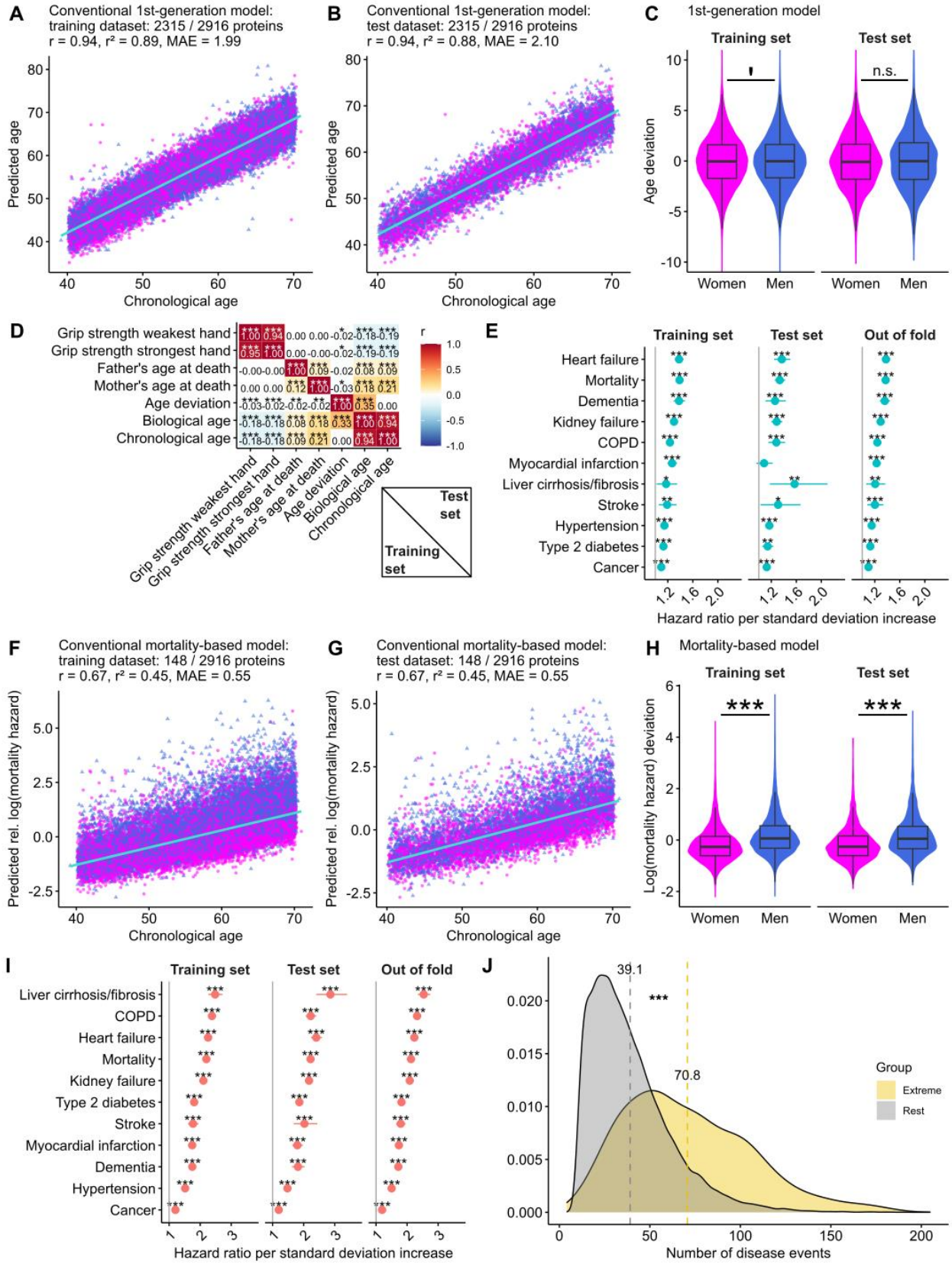



**Figure 1.** Plasma proteome-based biological aging models

**A and B.** Biological age predicted by our 1st-generation proteome aging model correlates strongly with chronological age in training (A) and test datasets (B).

**C.** Violin plot showing no significant difference in age deviation between men and women in training ($n_{men}$ = 16,502 versus $n_{women}$ = 19,460) and test datasets ($n_{men}$ = 4,128 versus $n_{women}$ = 4,862). '$p < 0.1$; n.s., not significant, t test.

**D.** Heatmap of correlation coefficients (r) between biological age predicted by our aging model, age deviation, and some traits in training (bottom triangle) and test (top triangle) datasets.

**E.** Estimated hazard ratios per unit of standard deviation in the chronological aging model are significantly larger than 1 for common age-related diseases and mortality in training and test datasets and out-of-fold predictions, after correcting for chronological age, sex, and their interaction.

**F and G.** Relative log(mortality hazard) correlates with chronological age in training (F) and test datasets (G).

**H.** Violin plot showing significantly lower log(mortality hazard) deviations in women versus men in training ($n_{men}$ = 16,502 versus $n_{women}$ = 19,460) and test datasets ($n_{men}$ = 4,128 versus $n_{women}$ = 4,862). $^{***}p < 1*10^{-16}$, t test.

**I.** Estimated hazard ratios per unit of standard deviation in the mortality-based aging model are significantly larger than 1 for common age-related diseases and mortality in training and test datasets and out-of-fold predictions, after correcting for chronological age, sex, and their interaction.

**J.** Based on the mortality-based aging model, extreme agers ($n$ = 321) have significantly more first occurrences of International Classification of Diseases, 10th edition (ICD-10) disease annotations as compared with other UK Biobank participants ($n$ = 44,631). $p < 1*10^{-16}$, t test. Dotted lines show the mean number of ICD-10 disease codes per group.

**C and H.** Thick black lines represent median values, the boxes span 50% of the data, and the black whiskers span 1.5 times the interquartile range.

**D, E, and I.** 'Benjamini-Hochberg false discovery rate (BH-FDR) < 0.1; *BH-FDR < 0.05; **BH-FDR < 0.01; ***BH-FDR < 0.001.

**E and I.** Whiskers denote 95% confidence intervals.

**A, B, F, and G.** Non-zero protein coefficients are shown as a fraction of total proteins on which the models are trained. Magenta, women; blue, men, with robust regression lines.

$n_{training}$ = 35,962, $n_{test}$ = 8,990, $n_{out\ of\ fold}$ = 44,952.

Our 1st-generation aging model indicates no significant sex difference in age deviation (Figure 1C). Predicted age correlates negatively with grip strength in both hands (Figure 1D), reproducing a well-known age-related decline. Age deviation also correlates negatively with grip strength, demonstrating that our model also captures weak but significant effects

beyond chronological age. Parental age of death correlates positively with both chronological and predicted age (Figure 1D). This is expected because the human longevity heritability is estimated around 15%–35%.[14,15,16] Therefore, older individuals are more likely to have longer-living parents due to survivorship bias. However, age deviation correlates negatively with parental age of death, indicating that biologically younger individuals tend to have longer-living parents. This finding shows that our aging model also captures very weak but significant heritable determinants of longevity. Although trained on chronological age, increased biological age as predicted by our model significantly increases the hazard for a plethora of age-related diseases, such as heart failure, dementia, and stroke, as well as for mortality, even after correcting for chronological age and sex (Figure 1E). These findings demonstrate that our 1st-generation model accurately predicts known age-related traits and diseases, including heritable longevity determinants.

## Mortality-based models of aging provide insights into age-related diseases and capture differences among sexes

We then trained conventional aging models on the same dataset using Cox proportional hazard elastic net regression with mortality as the outcome (STAR Methods; Tables S1C and S1D). Despite chronological age not being included as the training outcome, our mortality-based model retains a reasonably strong correlation with chronological age (Figures 1F and 1G; r, 0.67 in training and test datasets and out-of-fold predictions (r, 0.66; Figure S1C; Table S2A)) and low MAE values (Figure S1D; Table S2A). Our Cox elastic net model performs similar to accelerated failure time models (Figure S1E). We validated both the 1$^{st}$-generation model and the mortality-based conventional models in the external Multi-Ethnic Study of Atherosclerosis (MESA) study described in Bild et al.,[17] in which the plasma proteomes of 921 participants of diverse ancestries (including White, African-American, Hispanic, and Asian) were collected at their first and fifth visits and quantified with the SomaScan platform (Figures S1A–S1D).[18] To improve model performance, we rescaled the SomaScan data with the standard deviations of the Olink normalized protein expression (NPX) values in UK Biobank (Table S3).

Our mortality-based model finds a significantly larger difference in biological age between men and women than the chronological aging model, reflecting men's overall higher mortality risk (Figure 1H), consistent with prior studies.[19] Trained on mortality, this model not only better predicts mortality hazard but also the hazards of various age-related diseases (Figure 1I). Note that with a mortality hazard ratio of 1.37 per standard deviation increase in out-of-fold predictions, our chronological aging model performs on par with established biomarkers of aging such as PhenoAge and extrinsic epigenetic age acceleration (EEAA).[13,20,21] With a mortality hazard ratio equaling 2.13 per standard deviation increase in

out-of-fold predictions, our mortality-based model outperforms even the strongest biomarkers of mortality, including GrimAge,[22] with its hazard ratio of ~1.3–1.9 per unit increase in standard deviation.[23]

We further assessed how the mortality-based model could be informative at the individual level. We identified participants with significantly increased age deviation ("extreme agers"; STAR Methods). Interestingly, these participants have significantly more disease diagnoses compared with the rest of the population (Figure 1J). This demonstrates that the mortality-based model effectively identifies individuals at higher risk for age-related diseases.

## Plasma proteomics reveals markers of aging and mortality

We assessed the associations of the 2,923 measured proteins with chronological age, adjusting for sex. We found 1,664 proteins with significant positive associations with chronological age and 665 proteins with significant negative associations (Figure 2A; Table S4A). Notable hits include growth differentiation factor 15 (GDF15), a mitokine that protects against aging-mediated inflammation[24] and a robust predictor of mortality and various diseases in the UK Biobank,[25] and neurofilament light chain (NEFL), a protein that increases with age and a known mortality marker.[26] Both are also dementia markers.[27] Gene set enrichment analysis (GSEA) revealed significant increases in proteins involved in biological adhesion and locomotion with chronological age (Figure 2B; Table S4B). Interestingly, these proteins also strongly contribute to both our chronological age-based and mortality-based models.

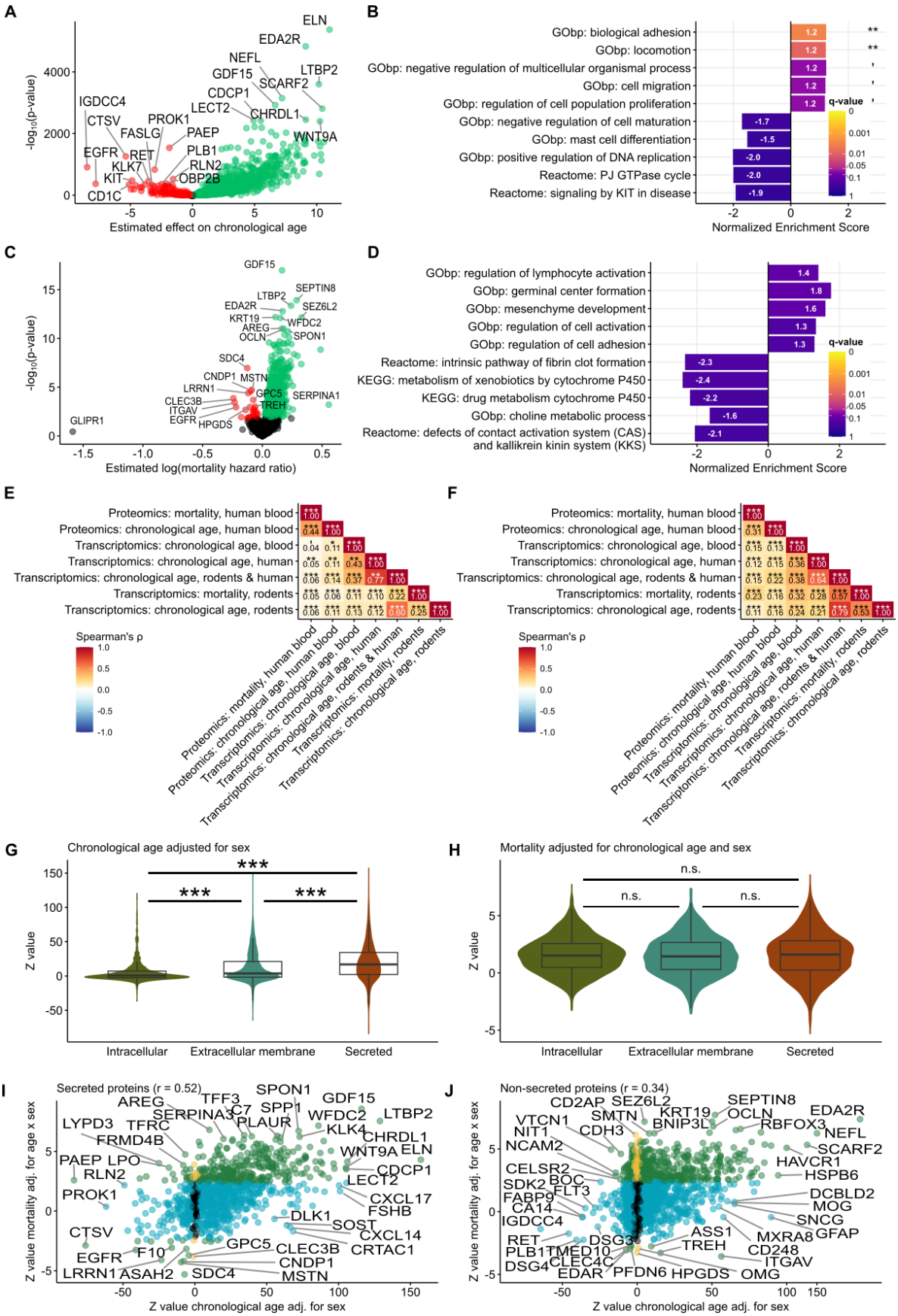

A
-log10(p-value)
Estimated effect on chronological age
ELN
EDA2R
NEFL
GDF15
LTBP2
SCARF2
CDCP1
IGDCC4
LECT2
CHRDL1
CTSV
PROK1
FASLG
PAEP
WNT9A
EGFR
RET
PLB1
KLK7
RLN2
KIT
OBP2B
CD1C
B
GObp: biological adhesion
GObp: locomotion
GObp: negative regulation of multicellular organismal process
GObp: cell migration
GObp: regulation of cell population proliferation
GObp: negative regulation of cell maturation
GObp: mast cell differentiation
GObp: positive regulation of DNA replication
Reactome: PJ GTPase cycle
Reactome: signaling by KIT in disease
q-value
Normalized Enrichment Score
C
-log10(p-value)
Estimated log(mortality hazard ratio)
GDF15
SEPTIN8
LTBP2
EDA2R
SEZ6L2
KRT19
WFDC2
AREG
SPON1
OCLN
SDC4
CNDP1
MSTN
LRRN1
GPC5
SERPINA1
CLEC3B
TREH
ITGAV
EGFR
HPGDS
GLIPR1
D
GObp: regulation of lymphocyte activation
GObp: germinal center formation
GObp: mesenchyme development
GObp: regulation of cell activation
GObp: regulation of cell adhesion
Reactome: intrinsic pathway of fibrin clot formation
KEGG: metabolism of xenobiotics by cytochrome P450
KEGG: drug metabolism cytochrome P450
GObp: choline metabolic process
Reactome: defects of contact activation system (CAS) and kallikrein kinin system (KKS)
q-value
Normalized Enrichment Score
E
Proteomics: mortality, human blood
Proteomics: chronological age, human blood
Transcriptomics: chronological age, blood
Transcriptomics: chronological age, human
Transcriptomics: chronological age, rodents & human
Transcriptomics: mortality, rodents
Transcriptomics: chronological age, rodents
Spearman's ρ
F
Proteomics: mortality, human blood
Proteomics: chronological age, human blood
Transcriptomics: chronological age, blood
Transcriptomics: chronological age, human
Transcriptomics: chronological age, rodents & human
Transcriptomics: mortality, rodents
Transcriptomics: chronological age, rodents
Spearman's ρ
G
Chronological age adjusted for sex
Z value
Intracellular
Extracellular membrane
Secreted
H
Mortality adjusted for chronological age and sex
n.s.
Z value
Intracellular
Extracellular membrane
Secreted
I
Secreted proteins (r = 0.52)
Z value mortality adj. for age x sex
Z value chronological age adj. for sex
SPON1
AREG
TFF3
GDF15
SERPINA3
C7
SPP1
WFDC2
LTBP2
LYPD3
TFRC
PLAUR
KLK4
FRMD4B
CHRDL1
WNT9A
ELN
PAEP
LPO
CDCP1
RLN2
LECT2
PROK1
CXCL17
FSHB
DLK1
CTSV
SOST
CXCL14
GPC5
CLEC3B
CRTAC1
EGFR
F10
CNDP1
LRRN1
ASAH2
SDC4
MSTN
J
Non-secreted proteins (r = 0.34)
Z value mortality adj. for age x sex
Z value chronological age adj. for sex
CD2AP
SEZ6L2
SEPTIN8
VTCN1
SMTN
KRT19
OCLN
EDA2R
NIT1
CDH3
BNIP3L
RBFOX3
NEFL
NCAM2
SCARF2
CELSR2
HAVCR1
SDK2
BOC
HSPB6
FLT3
FABP9
DCBLD2
CA14
MOG
IGDCC4
SNCG
ASS1
MXRA8
GFAP
RET
DSG3
TREH
CD248
PLB1
TMED10
ITGAV
DSG4
CLEC4C
EDAR
PFDN6
HPGDS
OMG

**Figure 2.** Proteomic signatures/biomarkers of chronological age and mortality

**A.** Volcano plot showing the effects on chronological age after correction for sex for 2,923 proteins in 53,014 UK Biobank participants. Increased abundance of 1,664 proteins (green) is significantly linked to higher chronological age, and 665 proteins (red) are significantly linked to lower chronological age.

**B.** Five most significantly up- and downregulated gene sets after GSEA on the effect on chronological age. Proteins in the gene sets "biological adhesion" and "locomotion" are significantly increased with increased chronological age (BH-FDR-adjusted q = 0.003 and 0.009, respectively).

**C.** Volcano plot showing the log(mortality hazard ratio) after correction for chronological age, sex, and their interaction for 2,923 proteins in 53,014 UK Biobank participants. Increased abundance of 892 proteins (green) is significantly linked to increased mortality risk, and 27 proteins (red) are significantly linked to decreased mortality risk.

**D.** Five most significantly up- and downregulated gene sets after GSEA on the log(mortality hazard ratio). No gene sets are significant.

**E and F.** Heatmaps showing the correlation coefficients (r) at gene/protein (E) and pathway (F) levels between blood proteomic signatures of chronological age and mortality with transcriptomic signatures derived from Tyshkovskiy et al.[4,28]

**G.** Violin plots showing the Z values for the effects on chronological age after correction for sex for intracellular proteins, proteins located at the extracellular membrane, and secreted proteins. Plasma concentrations of secreted proteins are significantly increased with age compared with intracellular and plasma membrane proteins, while extracellular membrane proteins are also significantly increased with age compared with intracellular proteins (***single-step adjusted $p < 1.5*10^{-10}$).

**H.** Violin plots showing the Z values for the mortality hazard after correction for chronological age, sex, and their interaction for intracellular proteins, proteins located at the extracellular membrane, and secreted proteins. Z values are not significantly different between the three groups of proteins (n.s., not significant; single-step adjusted $p > 0.1$).

**G and H.** Thick black lines represent median values, the boxes span 50% of the data, and the black whiskers span 1.5 times the interquartile range.

**I and J.** Correlations between Z values for the effects on chronological age after correction for sex and Z values for the mortality hazard ratio after correction for chronological age, sex, and their interaction; secreted proteins (I); and non-secreted (J) proteins. Cyan, significant effects on chronological age; green, significant effects on mortality hazard; yellow, significant effects on both.

**A–F, I, and J.** Significance thresholds were set at 5% BH-FDR. 'BH-FDR < 0.1; *BH-FDR < 0.05; **BH-FDR < 0.01; ***BH-FDR < 0.001.

We also assessed the associations of each plasma protein with mortality using Cox proportional hazards models, adjusting for chronological age, sex, and their interaction. 892 proteins were significantly associated with increased mortality risk, while 27 proteins were significantly associated with decreased mortality risk (Figure 2C; Table S4C). Enrichment analysis showed no significant changes in gene sets with mortality (Figure 2D; Table S4D).

Proteomic and transcriptomic[4,28] signatures of chronological aging and mortality demonstrated statistically significant positive correlations (Figure 2E). The mortality proteomic signature also showed positive correlation with almost all transcriptomic signatures, albeit weaker. Remarkably, at the level of enriched pathways, proteomic and transcriptomic signatures demonstrated even stronger, significant correlations (Figure 2F). Positive associations were mainly driven by upregulation of inflammation-related genes, interleukin (IL) signaling, interferon signaling, p53 signaling, and apoptosis, suggesting that aging is characterized by similar underlying biological processes across different molecular strata.

Extracellular membrane proteins increase on average more with age than intracellular proteins, and secreted proteins increase more than extracellular membrane proteins and intracellular proteins (Figure 2G). Conversely, associations with mortality do not differ significantly between these three protein groups (Figure 2H). This indicates that the chronological aging signature is enriched for secreted or extracellular membrane proteins, reflecting age-associated changes in these proteins, whereas the mortality signature elevated the proportion of intracellular proteins, likely reflecting increased cell death that releases these proteins into the plasma.

Secreted proteins show a much stronger correlation between the associations with chronological aging and mortality than non-secreted proteins (r, 0.52, Figure 2I versus 0.34, Figure 2J). This is likely due to a significant number of non-secreted proteins that do not change with healthy aging but are still indicative of mortality when detected in aberrant amounts in the blood. Conversely, proteins whose secretion changes with age tend to show similar changes with mortality.

## Organ-specific models capture diseases in these organs

We trained organ-specific models to predict chronological age (using elastic net) and mortality (using Cox elastic net) as outcomes (Table S1). Following Oh et al.,[10] a protein was considered organ-specific if its Genotype-Tissue Expression (GTEx) expression[29] is at least four times higher in one organ compared with all others (Figure 3A; STAR Methods). We then assessed their associations with chronological age (Figures 3B and 3C) and retained only

organ-specific models that sufficiently captured features of aging, excluding models without a consistent correlation coefficient (r) with chronological age above 30%.

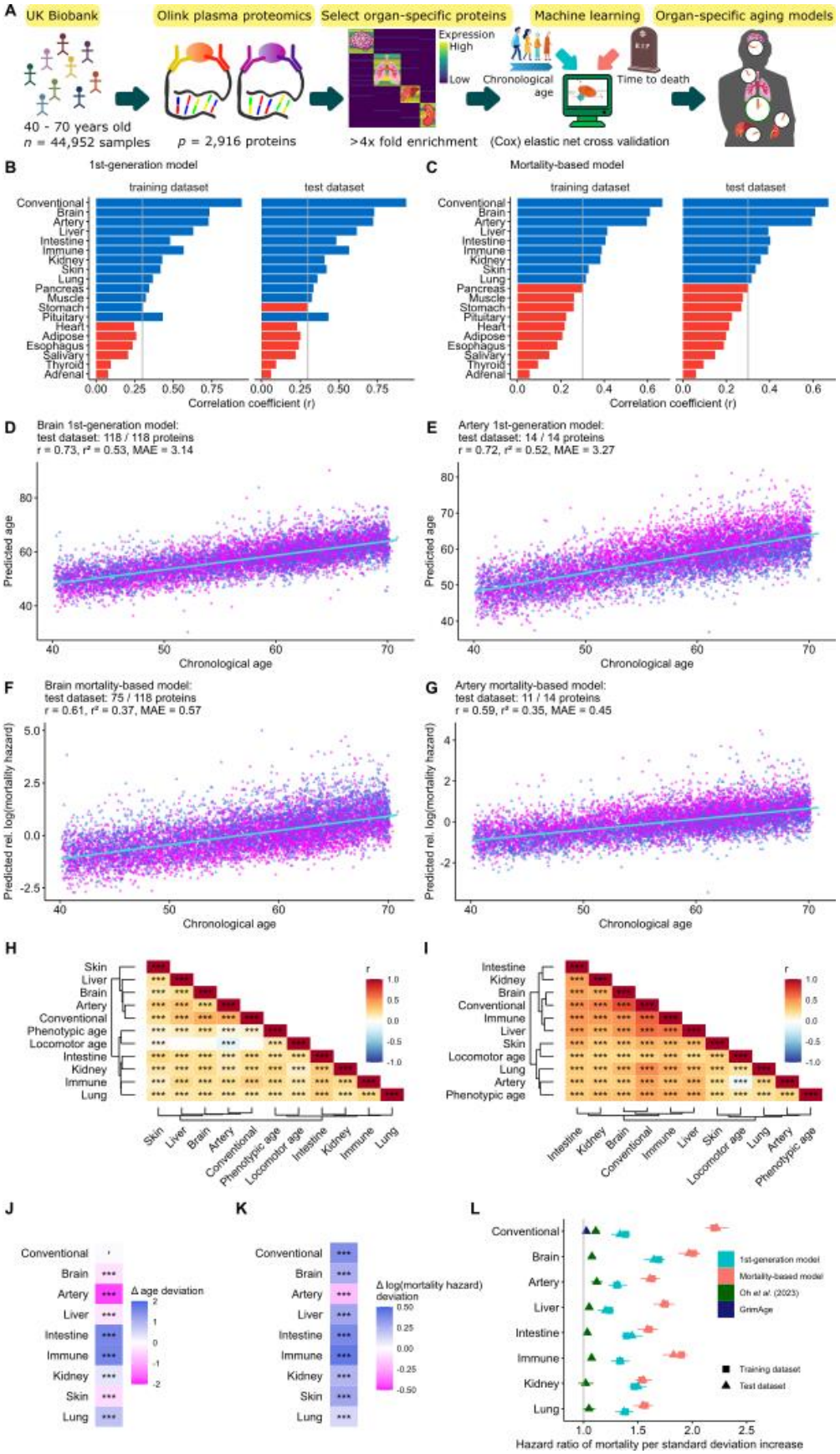

A
UK Biobank
Olink plasma proteomics
Select organ-specific proteins
Machine learning
Organ-specific aging models
40 - 70 years old
n = 44,952 samples
p = 2,916 proteins
>4x fold enrichment
(Cox) elastic net cross validation
Expression
High
Low
Chronological age
Time to death
B
1st-generation model
training dataset
test dataset
Correlation coefficient (r)
C
Mortality-based model
D
Brain 1st-generation model:
test dataset: 118 / 118 proteins
r = 0.73, r² = 0.53, MAE = 3.14
Predicted age
Chronological age
E
Artery 1st-generation model:
test dataset: 14 / 14 proteins
r = 0.72, r² = 0.52, MAE = 3.27
F
Brain mortality-based model:
test dataset: 75 / 118 proteins
r = 0.61, r² = 0.37, MAE = 0.57
Predicted rel. log(mortality hazard)
G
Artery mortality-based model:
test dataset: 11 / 14 proteins
r = 0.59, r² = 0.35, MAE = 0.45
H
I
J
Δ age deviation
K
Δ log(mortality hazard) deviation
L
1st-generation model
Mortality-based model
Oh et al. (2023)
GrimAge
Training dataset
Test dataset
Hazard ratio of mortality per standard deviation increase

**Figure 3.** Organ-specific aging models predict chronological age and mortality

**A.** Overview of organ-specific aging models. Using *n* = 44,952 UK Biobank samples and *p-value* = 2,916 proteins, we defined organ-specific proteins as those with average GTEx expression at least four times higher in one organ versus all other organs. Models were trained with elastic net to predict chronological age (1st-generation) or with Cox elastic net to predict time-to-death (mortality-based).

**B and C.** Bar plots show the correlation coefficients (r) with chronological age for the conventional and organ-specific aging models in training (*n* = 35,962) and test (*n* = 8,990) datasets for the 1st-generation (B) and mortality-based models (C). r > 0.3 are colored in blue, r < 0.3 in red.

**D–G.** Biological age (D and E) and relative log(mortality hazard) (F and G) predicted by our brain (D and F) and artery (E and G) 1st-generation (D and E) and mortality-based (F and G) aging models correlate positively with chronological age in the test dataset (*n* = 8,990). Non-zero protein coefficients are shown as a fraction of total proteins on which the models are trained. Magenta, women; blue, men, with robust regression lines.

**H and I.** Heatmaps showing the correlation coefficients (r) for the age deviations (H) and log(mortality hazard) deviations (I) based on the 1st-generation (H) and mortality-based (I) organ-specific and conventional models on out-of-fold predictions (*n* = 44,952), as well as phenotypic (*n* = 38,025) and locomotor age deviation (*n* = 8,660). ***BH-FDR < 0.001.

**J and K.** Differences in age deviation (J) and log(mortality hazard) deviation (K) between men and women, based on conventional and organ-specific 1st-generation (J) and mortality-based (K) models on out-of-fold predictions (*n* = 44,952). 'Hommel-adjusted $p < 0.1$, ***Hommel-adjusted $p < 0.001$. Blue, men are older; magenta, women are older.

**L.** Mortality hazard ratios per unit of standard deviation in training (*n* = 35,962) and test (*n* = 8,990) datasets for the 1st-generation (cyan) and mortality-based (red) models, compared with hazard ratios reported by Oh et al.[10] (LonGenity cohort, *n* = 962, green) and DNAm-based GrimAge[22] (Framingham Heart Study, *n* = 625, purple). Oh et al.[10] did not report a mortality hazard ratio for their skin model. Whiskers denote 95% confidence intervals.

Organ-specific models, trained on a subset of proteins, show weaker correlations with chronological age than the conventional model with, for example, r ranging from 0.59 to 0.73 in the test dataset for the brain and artery 1st-generation and mortality-based models (Figures 3D–3G). Correlation and accuracy metrics for the other organs can be found in Figures S1A–S1D and Table S2A. The conventional 1st-generation model and the 1st-generation brain-, artery-, and kidney-specific models predict chronological age slightly better than the models of Oh et al.,[10] while other models perform slightly worse (Figure S1F). Biological ages predicted by organ-specific models show significant positive correlations with each other and with chronological age (Figures S1G and S1H). Age deviations also correlate strongly for organ-specific models, reflecting coordinated aging across organs and systems (Figures 3H and 3I), as well as with phenotypic age,[13] and biological age calculated

based on locomotor activity,[30] with the exception of 1st-generation liver, brain, and conventional models, and a significant negative correlation with the artery model. Thus, organ-specific models also align with physiological clocks, consistent with systemic organismal aging. Interestingly, the 1st-generation models predict higher average biological ages for women in brain, artery, liver, and skin and for men in intestine, immune system, and lungs (Figure 3J). Conversely, the mortality-based models indicate men to be older in all organs except arteries (Figure 3K).

Overall, our 1st-generation organ-specific models perform in a range similar to those of Oh et al.[10] in predicting mortality, with our brain-, intestine-, kidney-, and lung-specific models performing better and the others worse (Figure 3L). However, our mortality-based organ-specific models strongly outperform the 1st-generation organ-specific models from Oh et al.[10] and GrimAge, a DNA methylation clock trained on 88 plasma proteins and smoking pack-years (Figure 3L).[22]

We examined if organ-specific models could predict organ-specific diseases. Strikingly, organ-specific aging was indeed associated with increased risk of organ-specific diseases. For example, the mortality-based liver aging model shows the highest hazard ratio for liver cirrhosis/fibrosis, followed by kidney and heart failure, surpassing the hazard ratio for mortality on which the model was trained (Figure 4A). The kidney-specific model excels at predicting liver cirrhosis/fibrosis and kidney failure, followed by type 2 diabetes and heart failure, and only then mortality (Figure 4B), while the lung-specific model best predicts chronic obstructive pulmonary disease (COPD) and only then mortality (Figure 4C).

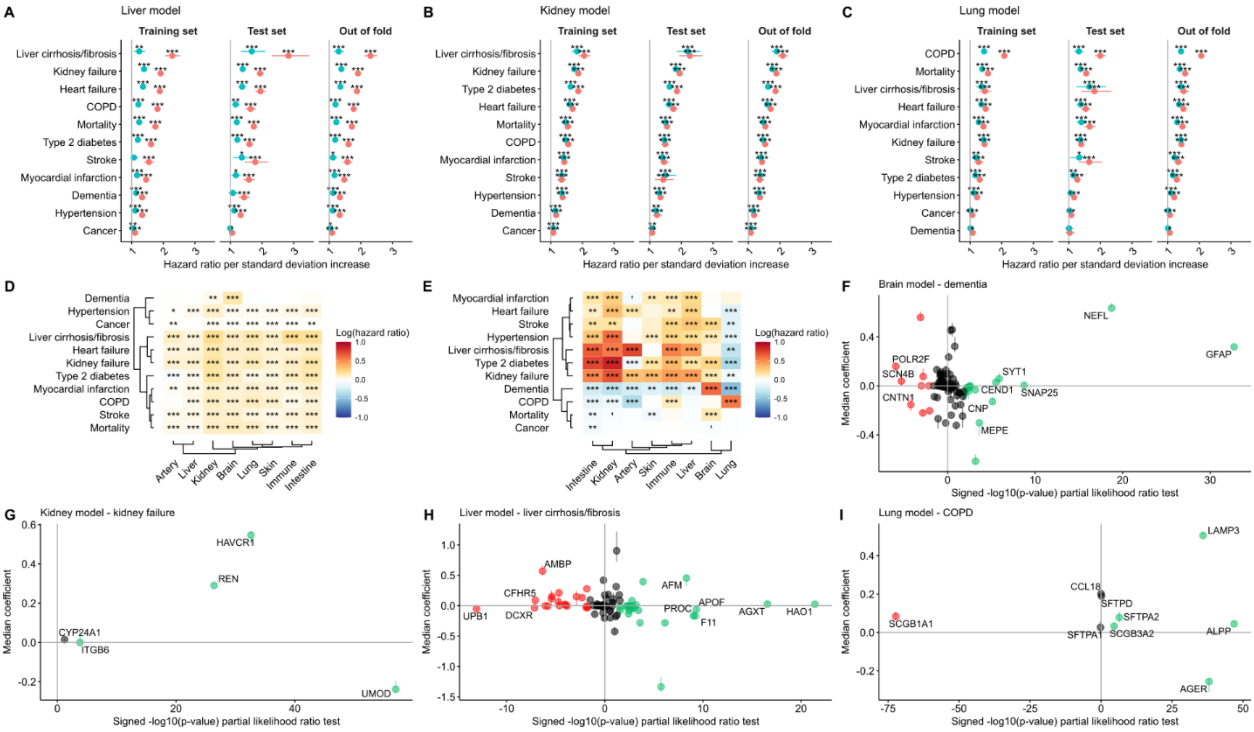



**Figure 4.** Organ-specific aging models predict organ-specific diseases

**A–C.** Hazard ratios per unit of standard deviation in the 1st-generation (cyan) and mortality-based (red) liver (A), kidney (B), and lung (C) aging models for common age-related diseases and mortality in training (*n* = 35,962) and test (*n* = 8,990) datasets and out-of-fold predictions (*n* = 44,952). Whiskers denote 95% confidence intervals.

**D and E.** Heatmaps showing log-hazard ratios per unit of predicted biological age (D) and predicted log(mortality hazard) (E) for the 1st-generation (D) and mortality-based (E) organ-specific models for common age-related diseases and mortality for the out-of-fold predictions (*n* = 44,952), after correcting for conventional 1st-generation (D) or mortality-based (E) biological age, chronological age, sex, and the interaction between chronological age and sex.

**F–I.** The horizontal axis shows the signed $-\log_{10}(p \text{ value})$ from partial likelihood ratio tests comparing the explanatory power of the full model predictions versus predictions where one protein was removed. Outcomes are dementia (F), kidney failure (G), liver cirrhosis/fibrosis (H), or COPD (I). Predictors are chronological age, sex and their interaction, and the log(mortality hazards) predicted by brain- (F), kidney- (G), liver- (H), or lung- (I) specific mortality-based models on full versus incomplete data. Green proteins have significantly better full-data predictions, excluding red proteins significantly improves predictions. The vertical axis shows each protein's median coefficient in the full model predictions. Vertical whiskers denote the coefficient's range across the 5-fold.

'BH-FDR < 0.1; *BH-FDR < 0.05; **BH-FDR < 0.01; ***BH-FDR < 0.001. Significance thresholds were set at 5% BH-FDR.

After correcting for the conventional model's biological age, as well as for chronological age, sex, and their interaction, most 1st-generation organ-specific ages showed positive associations with most diseases (Figure 4D). Conversely, for the mortality-based models, we uncovered strong relations between organ-specific aging and organ-specific diseases. For example, increased aging as predicted by the mortality-based lung model predicts COPD, independent of conventional mortality-based biological age (Figure 4E). Similarly, the brain model uniquely predicts dementia (Figure 4E), whereas the kidney model is the strongest predictor for type 2 diabetes and kidney failure. These findings demonstrate that organ-specific models capture additional features of organ-specific diseases not captured by conventional models, reflecting diseases as accelerated aging of the organs (or systems) involved.

When assessing how individual proteins affect disease associations, we found that the dementia association for the brain-specific model is mainly driven by glial fibrillary acidic protein (GFAP) and NEFL (Figure 4F). Like NEFL,[26] blood GFAP levels increase with dementia.[26] The kidney model's association with kidney failure is most strongly driven by uromodulin (UMOD), followed by hepatitis A virus cellular receptor 1 (HAVCR1) and renin (REN). The liver model's association with liver cirrhosis/fibrosis is most strongly driven by hydroxyacid oxidase 1 (HAO1), alanine-glyoxylate aminotransferase (AGTX), and apolipoprotein F (APOF). Alkaline phosphatase (ALPP), advanced glycosylation end-product specific receptor (AGER), and lysosome-associated membrane glycoprotein 3 (LAMP3) are most strongly driving the lung-specific model's association with COPD. Conversely, uteroglobin (SCGB1A1)'s inclusion significantly weakens the lung model's association with COPD. Interestingly, SCGB1A1 is a club cell marker that is significantly positively associated with chronological age and mortality (*p-value* = 0.01 and $<1*10^{-16}$; Tables S4A and S4C). However, plasma SCGB1A1 levels are significantly reduced in patients with COPD ($p < 1*10^{-16}$, corrected for age, sex, and their interaction). This highlights our models' ability to assess different aspects of aging in diverse contexts of accelerated aging, like smoking.

## Organ-specific aging is associated with environmental factors

We assessed how field of work associates with biological age predicted by our models. Teaching professionals have significantly lower log(mortality hazard) deviations in the immune system, kidney, liver, and lungs (Figure 5A). Conversely, laundry, renovation, and security workers have significantly higher age deviations. This aligns with findings that high autonomy and social status at work are associated with greater health, while repetitive, low-paid jobs are associated with poorer health.[31]

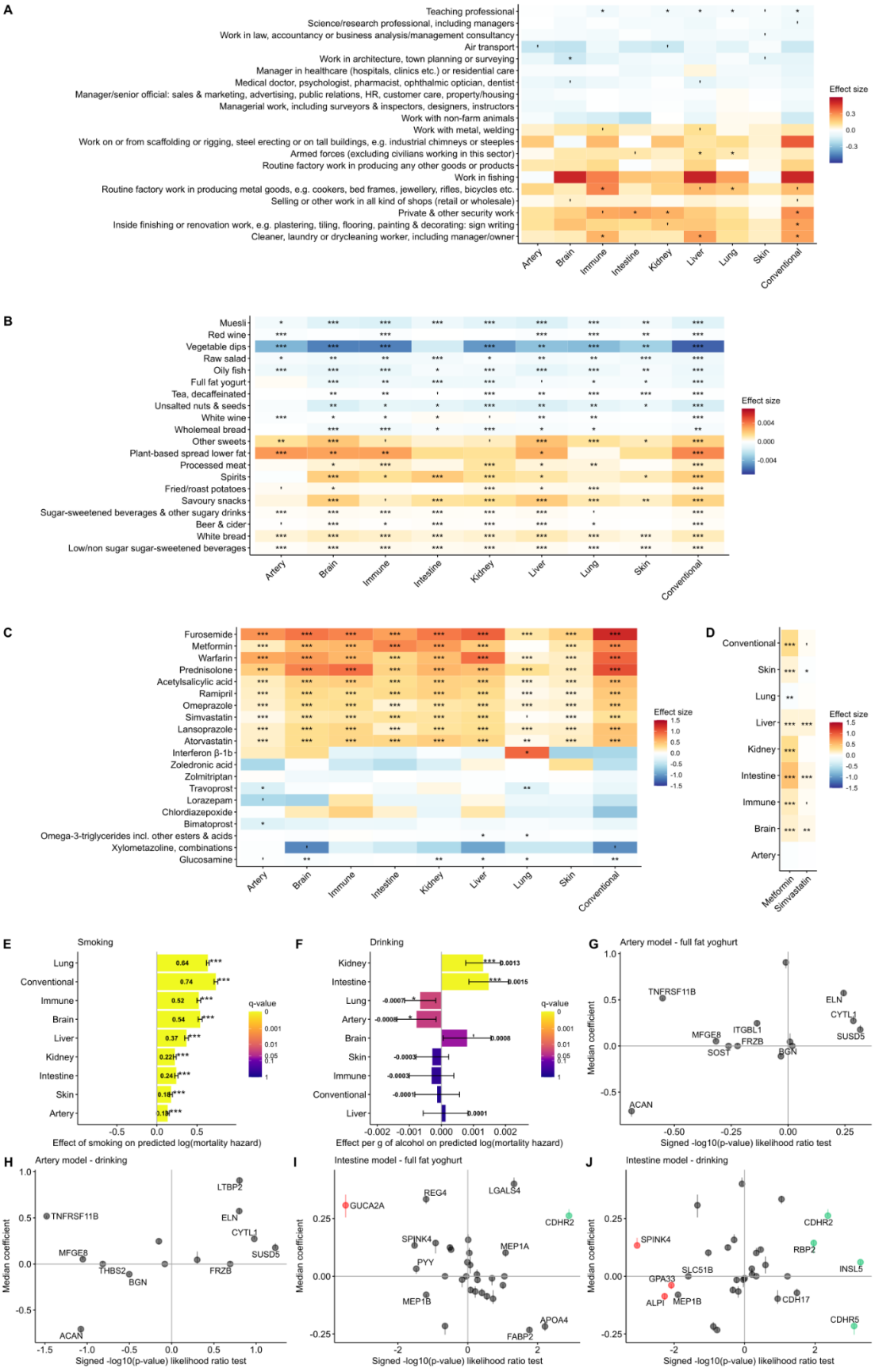

A
Teaching professional
Science/research professional, including managers
Work in law, accountancy or business analysis/management consultancy
Air transport
Work in architecture, town planning or surveying
Manager in healthcare (hospitals, clinics etc.) or residential care
Medical doctor, psychologist, pharmacist, ophthalmic optician, dentist
Manager/senior official: sales & marketing, advertising, public relations, HR, customer care, property/housing
Managerial work, including surveyors & inspectors, designers, instructors
Work with non-farm animals
Work with metal, welding
Work on or from scaffolding or rigging, steel erecting or on tall buildings, e.g. industrial chimneys or steeples
Armed forces (excluding civilians working in this sector)
Routine factory work in producing any other goods or products
Work in fishing
Routine factory work in producing metal goods, e.g. cookers, bed frames, jewellery, rifles, bicycles etc.
Selling or other work in all kind of shops (retail or wholesale)
Private & other security work
Inside finishing or renovation work, e.g. plastering, tiling, flooring, painting & decorating: sign writing
Cleaner, laundry or drycleaning worker, including manager/owner
Artery
Brain
Immune
Intestine
Kidney
Liver
Lung
Skin
Conventional
Effect size
0.3
0.0
-0.3
B
Muesli
Red wine
Vegetable dips
Raw salad
Oily fish
Full fat yogurt
Tea, decaffeinated
Unsalted nuts & seeds
White wine
Wholemeal bread
Other sweets
Plant-based spread lower fat
Processed meat
Spirits
Fried/roast potatoes
Savoury snacks
Sugar-sweetened beverages & other sugary drinks
Beer & cider
White bread
Low/non sugar sugar-sweetened beverages
Artery
Brain
Immune
Intestine
Kidney
Liver
Lung
Skin
Conventional
Effect size
0.004
0.000
-0.004
C
Furosemide
Metformin
Warfarin
Prednisolone
Acetylsalicylic acid
Ramipril
Omeprazole
Simvastatin
Lansoprazole
Atorvastatin
Interferon β-1b
Zoledronic acid
Zolmitriptan
Travoprost
Lorazepam
Chlordiazepoxide
Bimatoprost
Omega-3-triglycerides incl. other esters & acids
Xylometazoline, combinations
Glucosamine
Artery
Brain
Immune
Intestine
Kidney
Liver
Lung
Skin
Conventional
Effect size
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
D
Conventional
Skin
Lung
Liver
Kidney
Intestine
Immune
Brain
Artery
Metformin
Simvastatin
Effect size
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
E
Smoking
Lung
Conventional
Immune
Brain
Liver
Kidney
Intestine
Skin
Artery
0.64
0.74
0.52
0.54
0.37
0.22
0.24
0.18
0.18
q-value
0
0.001
0.01
0.05
0.1
1
-0.5
0.0
0.5
Effect of smoking on predicted log(mortality hazard)
F
Drinking
Kidney
Intestine
Lung
Artery
Brain
Skin
Immune
Conventional
Liver
0.0013
0.0015
-0.0007
-0.0008
0.0008
-0.0003
-0.0003
-0.0001
0.0001
q-value
0
0.001
0.01
0.05
0.1
1
-0.002
-0.001
0.000
0.001
0.002
Effect per g of alcohol on predicted log(mortality hazard)
G
Artery model - full fat yoghurt
TNFRSF11B
ELN
CYTL1
SUSD5
MFGE8
ITGBL1
FRZB
SOST
BGN
ACAN
Median coefficient
1.0
0.5
0.0
-0.5
-0.50
-0.25
0.00
0.25
Signed -log10(p-value) likelihood ratio test
H
Artery model - drinking
LTBP2
TNFRSF11B
ELN
CYTL1
MFGE8
SUSD5
THBS2
FRZB
BGN
ACAN
Median coefficient
1.0
0.5
0.0
-0.5
-1.5
-1.0
-0.5
0.0
0.5
1.0
Signed -log10(p-value) likelihood ratio test
I
Intestine model - full fat yoghurt
REG4
LGALS4
GUCA2A
CDHR2
SPINK4
MEP1A
PYY
MEP1B
APOA4
FABP2
Median coefficient
0.25
0.00
-0.25
-2
0
2
Signed -log10(p-value) likelihood ratio test
J
Intestine model - drinking
CDHR2
SPINK4
RBP2
SLC51B
INSL5
GPA33
ALPI
MEP1B
CDH17
CDHR5
Median coefficient
0.25
0.00
-0.25
-2
0
2
Signed -log10(p-value) likelihood ratio test

**Figure 5.** Organ-specific aging and environmental factors

**A–C.** Heatmaps showing the effects on log(mortality hazard) of the 10 most significant positively and negatively associated professions (A), food groups (B), and medications (C) with conventional log(mortality hazard), after correcting for chronological age, sex, their interaction, Townsend deprivation index, UK Biobank assessment center, IPAQ activity level, and smoking status (out-of-fold predictions, $n$ = 44,952).

**D.** Heatmap showing the effects on log(mortality hazard) of metformin and simvastatin after correcting for the confounders in (A)–(C), plus type 2 diabetes status for metformin and intake of lipid-modifying agents for simvastatin (out-of-fold predictions, $n$ = 44,952).

**E.** Bar plots showing differences in log(mortality hazard) between current smokers and never-smokers after correcting for chronological age, sex, their interaction, Townsend deprivation index, UK Biobank assessment center, and IPAQ activity level. Whiskers denote 95% confidence intervals (out-of-fold predictions, $n$ = 44,952).

**F.** Bar plots showing the effects of one additional gram of alcohol intake on log(mortality hazard) corrected for the confounders in (E), plus smoking status. Whiskers denote 95% confidence intervals (out-of-fold predictions, $n$ = 44,952).

**G–J.** Same as in Figures (4F–4I). Outcomes are full-fat yogurt intake (G and I) or alcohol (H and J), and predictors the log(mortality hazards) predicted by the artery- (G and H) or intestine- (I and J) specific mortality-based models, with chronological age, sex and their interaction, Townsend deprivation index, UK Biobank assessment center, IPAQ activity level, and smoking status as covariates. 'BH-FDR < 0.1; *BH-FDR < 0.05; **BH-FDR < 0.01; ***BH-FDR < 0.001.

We also assessed the associations of diet with biological age (Figure 5B). Using the 24-h recall dietary information as a proxy for dietary habits, we assessed the associations of 93 different food groups on organ-specific biological age. Healthy foods like vegetable dips and raw salads and high-socioeconomic status foods, like red wine, are linked to lower biological age. Unhealthy options like sweetened beverages, processed meat, spirits, beer, and cider are associated with higher biological ages.

We also assessed how 582 medications relate to organ-specific log(mortality hazards) (Figure 5C). Furosemide, used for treating edema in acute life-threatening conditions like heart failure, liver cirrhosis, kidney failure, and nephrotic syndrome, is most strongly associated with high biological age using the conventional model. Drugs such as simvastatin, used against high cholesterol and high triglyceride levels, and metformin, used to treat type 2 diabetes, are also strongly associated with increased biological age, likely reflecting disease status. Indeed, after correcting for diabetes status (for metformin) and regular intake of lipid-modifying agents as a proxy for a high cholesterol diagnosis (for simvastatin), these associations weaken (Figure 5D). Notably, after this correction,

metformin and simvastatin show significant negative associations with lung and skin age, respectively.

Conversely, glucosamine, a food supplement, is significantly associated with lower biological age. However, despite correcting for various confounders, residual effects related to higher socioeconomic status and healthier lifestyle of supplement users may persist.

We furthermore investigated the impact of smoking and alcohol consumption on organ-specific aging. Although the conventional model shows the largest effect size, the difference in predicted log(hazard of mortality) between current and non-smokers is the most significant for the lung model ($-\log_{10}(p$ value): 1,096), followed by the conventional ($-\log_{10}(p$ value): 756) and the immune model ($-\log_{10}(p$ value): 390) (Figure 5E). Alcohol consumption shows the largest increase in predicted log(mortality hazard) in kidney and intestine models, followed by brain (Figure 5F). Interestingly, alcohol has a significant negative effect on the lung- and artery-specific log(mortality hazards) (Figure 5F).

In the artery model, no proteins significantly drive the associations with full-fat yogurt or alcohol intake (Figures 5G and 5H). However, in the intestine model, cadherin-related family member 2 (CDHR2) significantly drives the association between full-fat yogurt intake and reduced biological age (Figure 5I). The increased intestinal age with increased alcohol intake is most strongly driven by insulin-like peptide 5 (INSL5), cadherin-related family member 5 (CDHR5), and CDHR2 (Figure 5J). The contributions of the CDHR5 and CDHR2 cadherins may reflect alcohol-induced epithelial damage.[32]

## Longitudinal models of aging predict future health outcomes

We then fit mortality-based aging models on the 1,451 longitudinally available proteins ("feature-reduced models"; Table S5) and estimated the rate of aging for 985 participants (STAR Methods). The feature-reduced models correlate well with chronological age and have low MAE values (Figures S2 and S3; Table S2B). As expected, our conventional model shows that 983 out of 985 participants (99.8%) increase in biological age over time (Figure 6A), and this is also the case for organ-specific longitudinal models (Figures S4A–S4I).

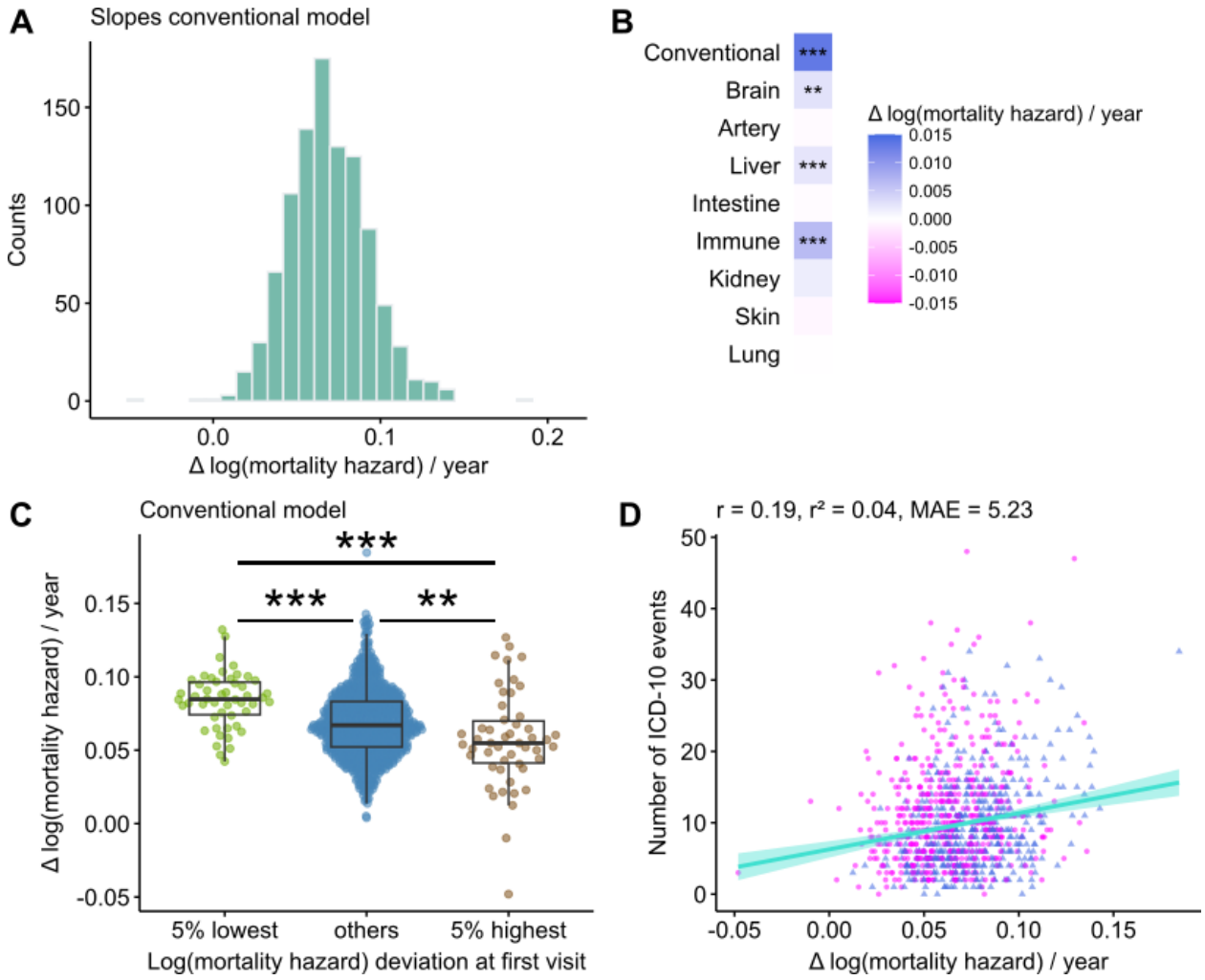



**Figure 6.** Longitudinal mortality-based aging models

**A.** Histogram showing the slopes of the log(mortality hazards) as predicted by the conventional feature-reduced mortality-based model. 983 out of 985 individuals increase in predicted log(mortality hazard).

**B.** The conventional, brain-specific, liver-specific, and immune-specific reduced-features mortality-based models indicate that men age significantly faster than women (Hommel-adjusted *p-value* = $1*10^{-17}$, 0.004, $1*10^{-4}$, and $2*10^{-15}$, respectively, t test, $n_{men}$ = 454, $n_{women}$ = 531). Blue, men are older; magenta, women are older.

**C.** Participants with higher log(mortality hazard deviations) at first visit age slower. *p* values are adjusted with the single-step method. Thick black lines represent median values, the boxes span 50% of the data, and the black whiskers span 1.5 times the interquartile range or the range of the data, whichever is smaller.

**D.** 0.1 units increase in yearly log(mortality hazard) change, estimated by our conventional longitudinal aging model, corresponds to an increase of on average 5.7 extra ICD-10 disease

annotations (ordinary linear regression, $p$ value = $9*10^{-10}$). Magenta, women; blue, men. Robust regression lines with 95% confidence bands are added. $n$ = 985, 'adjusted $p < 0.1$; *adjusted $p < 0.05$; **adjusted $p < 0.01$; ***adjusted $p < 0.001$.

Interestingly, men age significantly faster overall and in brain, liver, and intestine (Figure 6B). We then investigated how initial biological age impacts biological aging rates. Individuals with lower initial biological ages have higher age accelerations (Figure 6C), a trend observed for most organs (Figures S4J–S4R). This likely results from regression to the mean. Nonetheless, individuals with higher age accelerations accumulate more diseases; each 0.1 increase in yearly log(mortality hazard) increase is associated with on average 5.7 extra International Classification of Diseases, 10th edition (ICD-10) disease annotations (Figure 6D).

## Organ-specific models provide insights into organ-specific diseases in independent datasets

We assessed our feature-reduced aging models in the independent Filbin et al. dataset with longitudinal data for 305 COVID-19-positive and 78 COVID-19-negative individuals.[33] Both the conventional and the lung-specific feature-reduced mortality-based models correlate with chronological age (r: 0.50 and 0.08; Figures S5A and S5B). The conventional model predicts significantly lower mortality hazards in COVID-19-positive versus COVID-19-negative individuals (Figure 7A; difference in average log(mortality hazard) deviation: 0.05, *p-value* = $7*10^{-14}$). This result is surprising but not unheard of. For example, Cao et al.[34] also report lower PhenoAge-based epigenetic ages in non-severe patients with COVID-19 versus controls. Conversely, the lung-specific model gives the opposite result (Figure 7B; difference in average log(mortality hazard) deviation: 0.59, *p-value* = 0.04). However, while the conventional model predicts COVID-19 severity well (Hommel-adjusted *p-value* = $3*10^{-13}$, r: 0.14; Figure 7C), the lung model shows no significant correlation with COVID-19 severity (Hommel-adjusted *p-value* = 0.2, r: 0.003; Figure 7D). The lung is indeed an obvious viral entry organ, making the lung model an accurate predictor for COVID-19 status. However, as severe COVID-19 presents systemically, disease severity is better assessed with a conventional model. Indeed, all other organ-specific models also strongly predict COVID-19 severity (Figures S5C–S5K). The lung-specific and conventional model's associations with COVID-19 status are not significantly driven by any single protein (Figures 7E–7G), nor is the non-significant association of the lung model with severity. However, the association of the conventional model with severity is mainly driven by immune-response markers like IL-6, IL-7 receptor (IL-7R), and matrix metalloproteinase-8 (MMP8).

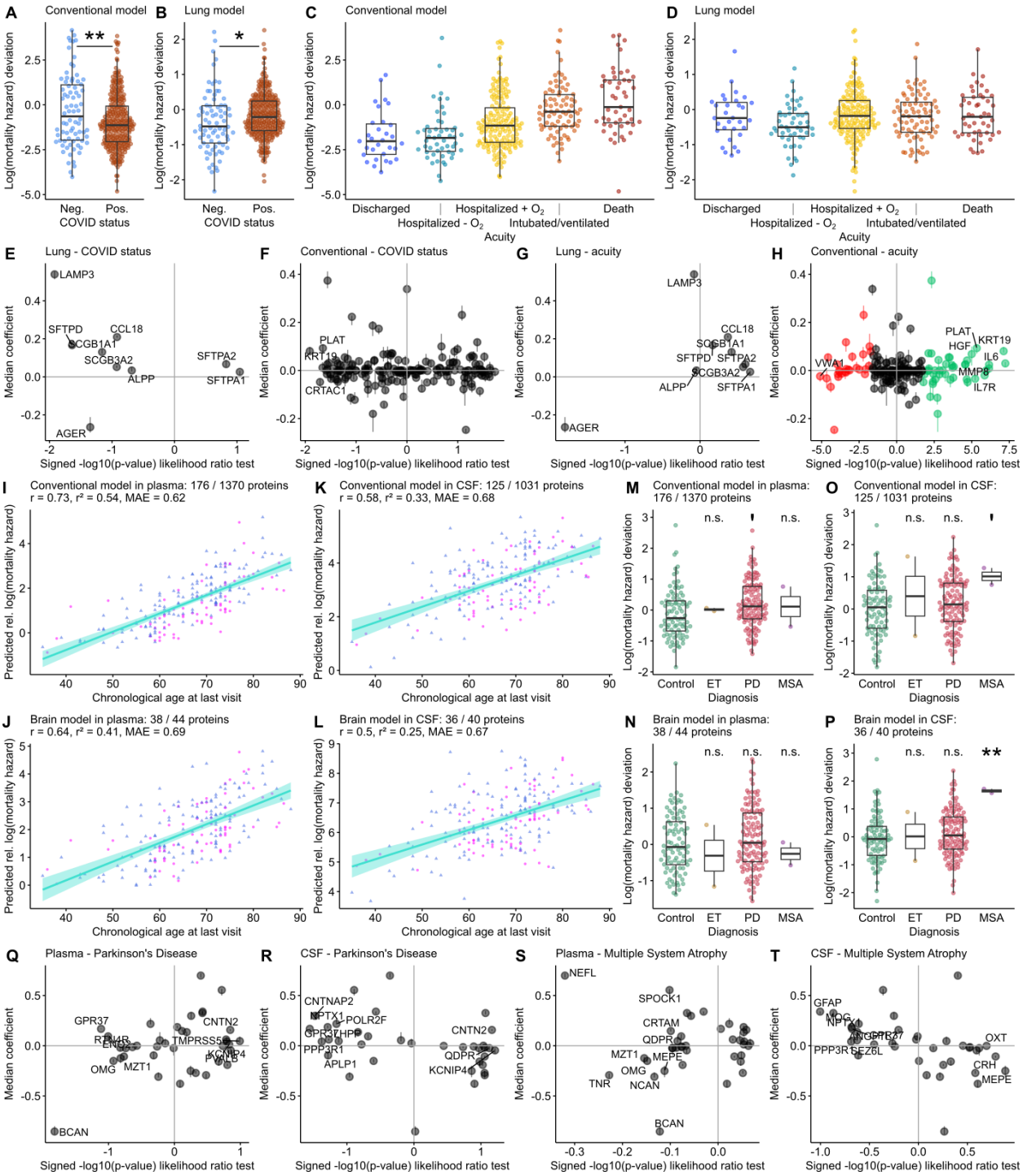



**Figure 7.** Organ-specific feature-reduced aging models reveal the relationship between aging and disease

**A.** Conventional log(mortality hazard) deviation is significantly higher in COVID-19-negative versus positive individuals at day 0 (*p-value* = 0.004, t test).

**B.** Lung log(mortality hazard) deviation is significantly higher in COVID-19-positive versus COVID-19-negative individuals (*p-value* = 0.04, t test).

**C.** Increased conventional log(mortality hazard) deviation is significantly associated with increased COVID-19 infection severity (*p-value* = $3*10^{-13}$, r = 0.14, ordinary least squares (OLS) regression).

**D.** Lung log(mortality hazard) deviation is not significantly associated with COVID-19 infection severity (*p-value* = 0.23, r = 0.003, OLS regression).

**E–H.** The horizontal axis shows the signed $-\log_{10}(p$ value) from likelihood ratio tests comparing the explanatory power of the full model predictions versus predictions where one protein was removed. Outcomes are COVID-19 status (E and F) or acuity (G and H). Predictors are chronological age and the log(mortality hazards) predicted by lung-specific (E and G) or conventional (F and H) mortality-based models on full versus incomplete data. Green proteins have significantly better full-data predictions, excluding red proteins significantly improves predictions. The vertical axis shows each protein's median coefficient in the full model predictions. Vertical whiskers denote the coefficient's range across the 5-fold.

**I and J.** Relative log(mortality hazard) predicted by the conventional (I) and brain (J) aging models correlate positively with chronological age in plasma samples taken at each individual's last visit. Non-zero protein coefficients are shown as a fraction of total proteins on which the models are trained that are available in the dataset. Magenta, women; blue, men. Robust regression lines with 95% confidence bands are added.

**K and L.** Same as for (I) and (J) but applied to CSF.

**M.** In plasma, the conventional model does not detect significant log(mortality hazard) differences at last visit after correcting for chronological age, sex, and their interaction between essential tremor (ET), Parkinson's disease (PD), and multiple systems atrophy (MSA), compared with controls. Non-zero protein coefficients are shown as a fraction of total proteins on which the models are trained that are available in the dataset (single-step adjusted *p-value* = 1, 0.07, and 0.8, respectively).

**N.** Same as in (M), for the brain model. Single-step adjusted *p-value* = 0.9, 0.8, and 1.

**O.** Same as in (M), but in CSF. Single-step adjusted *p-value* = 1, 0.8, and 0.1.

**P.** Same as in (O), for the brain model. Brain-specific log(mortality hazard) in MSA is significantly increased compared with controls (single-step adjusted *p-value* = 0.006). ET and PD are not significantly different versus controls (single-step adjusted *p-value* = 1 and 0.5).

**A–D and M–P.** Thick black lines represent median values, the boxes span 50% of the data, and the black whiskers span 1.5 times the interquartile range or the range of the data, whichever is smaller.

**Q–T.** Same as in (E)–(H), but for the brain-specific mortality-based model applied to plasma (Q and S) and CSF (R and T), with outcomes PD (Q and R) or MSA (S and T) and covariates chronological age, sex, and their interaction.

Data in (A)–(H) are from Filbin et al.,[33] and data in (I)–(T) are from Dammer et al.[35] **A and B.** $n_{negative}$ = 305, $n_{positive}$ = 78. **C and D.** $n$ = 383. I and L. $n$ = 212. **M and P.** $n_{control}$ = 90, $n_{ET}$ = 2, $n_{ET}$ = 118, $n_{MSA}$ = 2.

We further focused on plasma and cerebrospinal fluid (CSF) samples from Dammer et al., wherein 1,398 and 1,054 proteins were respectively measured with the Olink platform.[35] Of 212 participants, 90 were controls, 2 had essential tremor (ET), 118 had Parkinson's disease (PD), and 2 had multiple system atrophy (MSA). Our feature-reduced conventional mortality-based aging model predicts chronological age better than the brain-specific model in plasma (Figures 7I and 7J; r: 0.73 and 0.64) and in CSF (Figures 7K and 7L; r: 0.58 and 0.50).

We then compared our feature-reduced mortality-based conventional model trained on the UK Biobank plasma to 1st-generation conventional (Figures S6A–S6C) and brain-specific (Figures S6D–S6F) models trained directly on the CSF data from Dammer et al. (Table S6).[35] The 1st-generation conventional CSF-based model outperforms the feature-reduced mortality-based model (r: 0.72 and 0.58, respectively; Figures 7K and S6A). However, the brain-specific mortality-based model outperforms the 1st-generation brain-specific CSF-based model at predicting chronological age (r: 0.42 and 0.50, respectively; Figures 7L and S6D). The brain-specific CSF-based model, but not the conventional model, is able to differentiate MSA from controls (single-step adjusted *p-value* = $6*10^{-5}$; Figures S6C and S6F). This model's ability to differentiate PD from controls narrowly fails to pass our significance threshold (single-step adjusted *p-value* = 0.05). In plasma, neither the conventional nor the brain-specific mortality-based model find significant differences in mortality hazard between ET, PD, and MSA versus controls (Figures 7M and 7N). In CSF, the conventional mortality-based model again fails to detect significant differences between ET, PD, and MSA versus controls (Figure 7O). However, our brain-specific mortality-based model trained on UK Biobank plasma better discerns MSA (Figure 7P; single-step-adjusted *p-value* = 0.006), though these results will need to be verified in samples with larger numbers of MSA cases. These models' associations are not significantly driven by any single protein (Figures 7Q–7T).

Finally, we used the data from Damsky et al. to assess our feature-reduced mortality-based models in the context of sarcoidosis, a disease characterized by the growth of inflammatory granulomas throughout the body, especially in the lungs and lymph nodes.[36] In line with this, we find a significant increase in mortality hazard for patients with sarcoidosis versus healthy controls as predicted by the immune and lung models (Hommel-adjusted *p-value* = 0.003 and 0.005) but not the conventional model nor other organ-specific models (Figure S7).

## Discussion

The question of whether aging is a disease has long been debated, with a primary focus on whether aging is a normal, natural, physiological process, or a pathology.[37] Our study turns the table by asking whether chronic diseases are manifestations of aging. Our data on organ-specific aging patterns and their connections with diseases suggest that chronic age-related diseases represent accelerated aging of respective organismal systems or subsystems.

Previously, accelerated age-associated alterations within the heart and arteries have been proposed as a type of cardiovascular disease.[38] These authors argued for incorporating the concept of cardiovascular aging as a disease into clinical medicine. Similarly, many age-related changes in lungs, such as decreased lung function, increased gas trapping, loss of lung elastic recoil, and enlargement of the distal air spaces, are also characteristic of COPD, prompting researchers to consider COPD as an accelerated lung aging condition.[39] Type 2 diabetes has also been considered as a condition of accelerated aging.[40] However, viewing chronic disease as an accelerated aging condition does not necessarily mean that faster aging of an organ always results in a disease, nor does it imply that accelerated organ aging may be a pathogenic mechanism in all individuals with chronic disease. Indeed, not all very old organisms develop diseases in all organs. Disease may be better understood as an extreme form of aging within a system, which can occur at any age but is more likely to occur later in life. In some instances, such extreme aging may affect only a subsystem within an organ, compromising its function, or span functions of interacting organs.

Accelerated organ aging may be compensated by interactions with younger organs and systems within the organism, although such compensation may decrease with aging. Moreover, "gravitation" of a system with accelerated aging toward other systems within the organism with lower biological ages would lead to a state of the accelerated aging system that differs from normal aging.

In our study, we leveraged large-scale plasma proteomics data from over 53,000 UK Biobank participants to develop chronological (1$^{st}$-generation), mortality-based, longitudinal, and organ-specific aging models. Our findings indicate that chronic diseases typically manifest as accelerated organ- and system-specific aging phenomena.

The UK Biobank's rich proteomics data, with measurements of almost 3,000 proteins in the plasma of over 50,000 individuals,[12] enables developing highly accurate aging biomarkers.[6,7] Recent studies have demonstrated that epigenetic clocks trained on different tissues may yield varied outcomes, suggesting that organs may age at different rates.[9] Building on the findings of Oh et al.,[10] who demonstrated that the secreted proteomes of different organs could age differently between individuals, organ-specific aging models have garnered

increasing attention in aging research. Our study advances this concept by utilizing the extensive UK Biobank dataset to develop robust organ-specific aging models with superior statistical power and generalizability.

We initially developed conventional aging models, trained on either chronological age or mortality. Both models, especially the mortality-based model, demonstrated substantial predictive capability for future mortality, with the mortality-based model outperforming most, if not all, existing, validated aging biomarkers.[23] Moreover, both models had consistently high hazard ratios for a broad spectrum of severe age-related diseases.

We further built upon the work of Oh et al.[10] and developed organ-specific models of aging. These researchers showed that accelerated organ aging correlates with increased mortality risk and demonstrated a link between organ-specific diseases and faster aging of those organs. To avoid bias toward particular diseases, we trained both organ-specific models that predict chronological age, as in Oh et al.,[10] and organ-specific models that predict mortality, all of which exhibit modest to high correlations with chronological age in out-of-fold predictions. Our findings indicate that organs and systems exhibit distinct characteristic aging signatures in the plasma proteome, which are effectively captured by organ-specific aging models. Notably, despite being trained solely on chronological age or mortality, our organ-specific models were predictive of organ-specific diseases.

For instance, the mortality-based liver aging model showed the highest hazard ratios for liver cirrhosis/fibrosis, followed by kidney failure and heart failure. Kidney and heart failure often arise in patients with liver cirrhosis,[41] with up to 50% of hospitalized patients with cirrhosis developing acute renal failure.[42] The lung aging model was the best predictor for COPD, and, after correcting for conventional mortality-based age, the kidney-based model most strongly associated with renal failure and type 2 diabetes. The association between kidney aging and type 2 diabetes makes sense as the kidney is the primary target of microvascular damage in diabetes, and about half of all patients with type 2 diabetes develop chronic kidney disease.[42]

While conventional models may generate better overall associations with various diseases, organ-specific aging models regularly outperform them in predicting organ-specific diseases. For instance, only the brain-specific model in CSF, but not the conventional model, can distinguish between patients with MSA and controls (Figures 7M–7P). Similarly, only the immune-specific and lung-specific aging models, but not the conventional model nor other organ-specific models, can distinguish patients with sarcoidosis from healthy controls (Figure S7).

Interestingly, dementia is associated with accelerated brain aging but corresponds to decelerated aging in other organs after correcting for conventional mortality-based biological age (Figure 4E). This may be due to dementia typically developing later than most other chronic diseases, becoming prevalent primarily in very old individuals, who are expected to age slower in order to reach the age of onset. Conversely, individuals with accelerated aging of all organ systems or uniform organ aging would likely succumb primarily to diseases that manifest earlier in life. Taken together, these results suggest that age-related chronic diseases may arise from accelerated aging processes within specific organs or systems rather than from a uniform, systemic aging process and that our organ-specific aging models can predict risks of age-related organ-specific diseases that are not fully captured by conventional aging biomarkers.

Our analyses demonstrated strong associations between many lifestyle factors, such as diet, occupation, and medication use, with biological age predicted by our models. Teaching professionals tended to exhibit lower biological ages, while those in routine, lower-paid jobs showed accelerated aging across multiple organs. Similarly, healthy dietary patterns, such as vegetable consumption, were associated with lower biological ages, whereas unhealthy foods, like processed meats and sweetened beverages, were linked to increased biological aging.

We also observed some organ-specific associations of behavior with biological age. For example, since smoking directly affects the lungs, the lung-specific model outperforms all other organ-specific models and the conventional model in distinguishing current smokers from non-smokers. Similarly, as alcohol is absorbed through the intestine, the intestine-specific model shows the highest increase in predicted mortality with increased alcohol consumption. Although observational studies on alcohol consumption are confounded by many factors, it is nevertheless interesting to speculate that the vasodilating and damage buildup clearing properties of alcohol may contribute to lower biological ages in the artery-specific model. There is indeed still no consensus regarding the effects of moderate alcohol consumption on cardiovascular disease specifically.[43] Similarly, after correcting for disease status, vasodilating medications such as metformin and simvastatin are no longer associated with increased arterial aging.

Our models' interpretability allows investigation into which proteins drive accelerated organ-specific aging. For instance, the contributions of cadherins CDHR2 and CDHR5 to the association of alcohol intake with intestinal aging suggest that alcohol-induced accelerated intestinal aging may be partially driven by intestinal epithelial damage.[32] Similarly, the conventional mortality-based model's association with COVID-19 acuity relies heavily on

immune-response markers such as IL-6, IL-7R, and MMP8, indicating that severe COVID-19's pro-aging features mainly lie in the immune space.[33,34,44,45]

This approach also allows studying the differences between the effects of chronological and accelerated aging on specific organs. For example, circulating levels of the club cell marker SCGB1A1 significantly increase with age, while patients with COPD have significantly reduced levels of circulating SCGB1A1, in line with the literature.[46,47] This finding highlights that even if COPD shows plenty of hallmarks of accelerated lung aging,[48] some aging signatures will inevitably be different between COPD lungs and old non-COPD lungs.

We applied our organ-specific models in two independent datasets. In the COVID-19 dataset from Filbin et al.,[33] the lung-specific model accurately differentiated COVID-19-positive from COVID-19-negative individuals, reflecting the lung as the primary SARS-CoV-2 entry point. Conversely, the conventional model and all other organ-specific models better predicted disease severity, underscoring the systemic nature of severe COVID-19 infection.[44,45]

In the neurodegenerative disease dataset from Dammer et al.,[35] brain-specific models outperformed conventional models in predicting chronological age when applied to CSF data. Distinguishing patients with PD and controls proved challenging. Only a brain-specific 1$^{st}$-generation model trained on the CSF data itself achieved significant differentiation between PD and controls (single-step-adjusted *p-value* = 0.01), but our mortality-based conventional and brain-specific models trained on UK Biobank plasma data failed to distinguish PD from controls. Conversely, our UK Biobank plasma-trained mortality-based brain model outperformed the CSF-trained 1$^{st}$-generation models in predicting MSA. This difference in predictive power may reflect our models' training on a mortality outcome. People with PD have a near-normal life expectancy, while MSA is a progressive, fatal neurodegenerative disorder with a median survival of only 8.6 years after diagnosis.[49] Although these results need confirmation in a dataset with more MSA cases, this observation underscores that aging models trained on mortality are more adept at predicting diseases with higher mortality risks. Similarly, validation in the dataset of Damsky et al.[36] demonstrates that patients with sarcoidosis, who experience growth of inflammatory granulomas throughout the body, but mainly in lungs and lymph nodes, are predicted to be significantly older than healthy controls by the immune- and lung-specific models but not by others. This again underscores the ability of our organ-specific models to detect organ-specific diseases and reinforces the link between disease and organ-specific aging.

## Limitations of the study

Our study has several limitations. First, designating a protein as organ-specific based on four times higher gene expression in that organ may not optimally indicate organ-specificity as

mRNA-protein correlations are typically below 40%.[50] However, this low correlation may partially stem from accuracy and reproducibility limitations in mRNA and protein measurement techniques.[51] We reasoned that mRNA levels better indicate which organ produces a protein, rather than where it may be taken up. Nonetheless, our manuscript serves as a proof of concept, and other organ-specificity metrics such as AdaTiSS[52] and Tau[53] could also be considered. Additionally, 4,944 out of 44,952 participants included when training the full models, and 14 out of 985 participants with longitudinal data were included due to their diseases' relevance to the UKB-PPP Consortium. This could introduce a minor bias toward less healthy participants. The UK Biobank's massive sample size grants this study exceptional statistical power, enabling us to detect even minute differences with high confidence—differences that smaller-scale studies may overlook. While these effects are likely genuine, assuming no systematic bias, they may be very small in magnitude. Therefore, we consistently provide the estimated effect sizes, allowing readers to assess both the likelihood of an effect being real and its practical significance. Furthermore, the UK Biobank data are observational. Hence, while our correction for various confounders, including chronological age, sex, their interaction, Townsend deprivation index, UK Biobank assessment center, IPAQ activity level, and smoking status, increases our confidence in the associations between organ-specific aging and various environmental factors, one cannot claim causality due to unknown and potentially unmeasured confounders. Additionally, since our models were trained on UK Biobank, they are theoretically only fully generalizable toward UK Biobank participants. Despite this, they generalize well toward other populations, as demonstrated by our validation in various cohorts.[17,33,35,36] Organ-specific models could also be built based on the transcriptome. Integrating proteome and transcriptome levels into (organ-specific) aging models has the potential to enhance predictive accuracy that may lead to a synergistic effect and improve the reliability and efficacy of such models. Exploring this avenue could provide valuable insights into the interplay between the transcriptome and the proteome. Finally, our study used established aging outcomes: chronological age (1$^{st}$-generation models) and mortality (2$^{nd}$-generation models). However, as the exact causes of biological aging, and especially organ-specific aging, remain largely unknown, other organ-specific phenotypes or biological measurements could also be considered when training organ-specific aging models to even better reflect organ-specific aging.

Overall, we demonstrated the effectiveness of organ-specific aging models in capturing information beyond conventional aging models and predicting organ-specific disease. Based on these models, aging impacts different organs and systems at varying rates. We further demonstrate that different lifestyles, including diet, professional occupation, and medication intake, are associated with varying rates of aging across different organs. This suggests that future anti-aging interventions should not only target organismal aging, but

also the oldest organs, calling for a more personalized approach toward anti-aging interventions. The concept of chronic disease as accelerated aging of specific systems offers insights into the ongoing debate on whether aging should be considered a disease. Viewing aging of a system or subsystem as a disease state suggests that aging of an entire organism could be viewed as whole-organismal disease. However, this analogy is incomplete. Accelerated aging at the system level occurs non-autonomously in the background of interactions with younger systems, whereas whole-organism aging occurs within a relatively isolated system.

## Resource availability

### Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Vadim N. Gladyshev (vgladyshev@bwh.harvard.edu).

### Materials availability

This study did not generate new, unique reagents.

### Data and code availability

The existing UK Biobank data[11] reported in this study cannot be deposited in a public repository because they are sensitive personal-level data. To request access, all bona fide researchers who wish to conduct health-related research can apply for access to UK Biobank through UK Biobank's access management system at https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access.

The existing MESA study[17] data reported in this study cannot be deposited in a public repository because they are sensitive personal-level data. To request access, researchers need to submit a request for access to dbGaP. Afterward, they will be able to access the data through dbGaP identifier dbGaP: phs001416.v3.p1.

This paper analyzes the existing, publicly available Filbin et al.[33] data, accessible at Mendeley Data (https://doi.org/10.17632/nf853r8xsj.2.

This paper analyzes the existing, publicly available Damsky et al.[36] data, accessible at the Gene Expression Omnibus (GEO) under identifier GEO: GSE169148.

The existing Dammer et al.[35] data reported in this study cannot be deposited in a public repository because they are sensitive personal-level data. To request access, researchers need to create a Synapse account and agree to its Terms and Conditions. Afterward, they

will be able to download the data from https://www.synapse.org/#!Synapse:syn30549757/files/.

In addition, processed datasets derived from these data have been deposited at Mendeley Data (https://doi.org/10.17632/kj6y3zsz4d.1) and are publicly available as of the date of publication.

The coefficients from our full and feature-reduced conventional and organ-specific 1st-generation and mortality-based models are available in Tables S1 and S5, respectively.

All original code has been deposited at Mendeley Data and is publicly available at https://doi.org/10.17632/kj6y3zsz4d.1

as of the date of publication. Additional information required to reanalyze the data reported in this paper, including instructions for applying the models and recreating the figures, can be found on GitHub at www.github.com/ludgergoeminne/organAging

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.


## Acknowledgments

This work was funded by the National Institute on Aging and Hevolution Foundation. The research has been conducted using the UK Biobank Resource under application number 21988. We would like to thank Dr. Peter Fedichev and Dr. Timothy Pyrkov for providing us estimates of phenotypic age and locomotor age.


## Author contributions

L.J.E.G. performed analyses, made the figures and tables, and wrote the manuscript. A.V. implemented the K-fold validations, leave-one-out validations, and accelerated failure time models. A.E. formatted the Anatomical Therapeutic Chemical codes and mapped them to UK Biobank medications. A.T. performed gene set enrichment analyses and correlation analyses of existing transcriptomic data. M.A.A. performed all comparisons of our study with the Oh et al.[10] study and GrimAge. K.Y. and M.M. performed validation analyses. V.N.G. conceived the study, supervised the work, provided funding, and wrote the manuscript. All authors approved the manuscript and gave critical comments.

## Declaration of interests

L.J.E.G. and V.N.G. are the inventors on a U.S. Patent Application related to this work.

# STAR★Methods

## Key resources table

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| **Deposited data** | | |
| Transcriptomic signatures of chronological aging | Tyshkovskiy et al.,[28] Tikhonov et al.[54] | https://age-meta.com/ |
| Transcriptomic signatures of mortality | Tyshkovskiy et al.[4] | Unpublished |
| UK Biobank data | Sudlow et al.[11] | https://www.ukbiobank.ac.uk/ |
| SomaScan plasma proteomics data of the Multi-Ethnic Study of Atherosclerosis (MESA) | Bild et al.[17] | dbGaP: phs001416.v3.p1 |
| Olink Explore 1536 plasma proteomics data of COVID-19-positive and -negative patients | Filbin et al.[33] | https://doi.org/10.17632/nf853r8xsj.2 |
| Olink Explore 1536 plasma and CFS proteomics data of controls, ET, PD, and MSA patients | Dammer et al.[35] | https://www.synapse.org/#!Synapse:syn30549757/files/ |

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Olink Explore 1536 plasma proteomics data of healthy controls and sarcoidosis patients treated with or without tofacitinib | Damsky et al.[36] | GEO: GSE169148 |
| Model coefficients of the full 1st-generation models | This paper | Table S1A |
| Model coefficients of the full mortality-based models | This paper | Table S1C |
| Model coefficients of the feature-reduced 1st-generation models | This paper | Table S5A |
| Model coefficients of the feature-reduced mortality-based models | This paper | Table S5C |
| Standard deviations of the full and feature-reduced models | This paper | Table S3 |
| Data underlying all plots (summarized for not publicly available individual-level data) | This paper | Data S1 |
| R and Python code, intermediary results | This paper | Mendeley Data: https://doi.org/10.17632/kj6y3zsz4d.1 |
| **Software and algorithms** | | |

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Programming environment: RStudio Pro 2022.12.0 Build 353.pro20 "Elsbeth Geranium" | RStudio Workbench | https://dailies.rstudio.com/version/2022.12.0+353.pro20/ |
| Software package: R version 4.1.2 | R Core Team[55] | https://cran.r-project.org/ |
| Summarizing individual-level data for figure tables: kMeansEqual R package version 0.02 | This paper | https://github.com/ludgergoeminne/kMeansEqual |
| Ordinary linear regression, t-tests: stats R package version 4.1.2 | R Core Team[55] | https://cran.r-project.org/ |
| Functional enrichment: fgsea R package version 1.20.0 | Korotkevich et al.[56] | https://bioconductor.org/packages/release/bioc/html/fgsea.html |
| Cox proportional-hazards models: survival R package version 1.3-28.1 | Therneau and Grambsch[57] | https://cran.r-project.org/web/packages/survival/index.html |
| Generalized additive models: mgcv R package version 1.8-42 | Wood[58] | https://cran.r-project.org/web/packages/mgcv/index.html |
| Mixed models: lme4 R package version 1.1-33 | Bates et al.[59] | https://cran.r-project.org/web/packages/lme4/index.html |
| Robust linear regression: MASS R package version 7.3-60 | Venables and Ripley[60] | https://cran.r-project.org/web/packages/MASS/index.html |

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Color palettes: MetBrewer R package version 0.2.0 | Mills[61] | https://github.com/BlakeRMills/MetBrewer |
| Multiple comparison testing: multcomp R package version 1.4-23 | Hothorn et al.[62] | https://cran.r-project.org/web/packages/multcomp/index.html |
| Non-nested log-likelihood ratio tests of ordinary linear regression models: nonnest2 R package version 0.5-7 | Merkle and You[63] | https://cran.r-project.org/web/packages/nonnest2/index.html |
| Non-nested log-likelihood ratio tests of Cox proportional-hazards models: nonnestcox R package version 0.0.0.9000 | Hielscher[64] | https://github.com/thomashielscher/nonnestcox |
| k-nearest neighbors imputation: impute R package version 1.68.0 | Hastie et al.[65] | https://bioconductor.org/packages/release/bioc/html/impute.htr |
| Software package: Python version 3.12.3 | Python programming language | https://www.python.org/downloads/release/python-3123/ |
| Model training: scikit-learn Python library version 1.3.2 | scikit-learn | https://scikit-learn.org/stable/ |
| Figure editing: Inkscape version 1.3.2 | Inkscape | https://inkscape.org/ |

## Method details

### Human cohorts

UK Biobank recruited ~500,000 people aged between 40-69 years old from 2006 until 2010 from across the United Kingdom. This dataset contains detailed personal information and medical records, amongst others. Recently, Eldjarn et al.[12] used the high-throughput affinity-based proteomics Olink Explore 3072 platform to measure the levels of 2,923 proteins simultaneously in the plasma of 53,014 UK Biobank participants. They also measured the levels of 1,463 proteins in the plasma of 1,172 participants at the imaging visit (2014+) and for 1,123 participants at the first repeat imaging visit (2019+).

The Olink proteomic datasets of Filbin et al., Dammer et al., and Damsky et al. were used for validation purposes.[33,35,36] These studies measured 1,429,[33] 1,398 (Dammer et al.,[35] plasma), 1,054 (Dammer et al.,[35] CSF), and 1,463[36] unique proteins, respectively, with the Olink Explore 1536 platform. After filtering (see below), these numbers reduced to 1,400, 1,370, 1,031, and 1,463, respectively. The Filbin et al. dataset contains 305 patients with COVID-19 and 78 COVID-negative controls for whom the Olink plasma proteins were measured at days 0, 3, and 7 and the acuity levels were recorded days at 0, 3, 7, and 28.[33] The Dammer et al. dataset contains 184 participants at first visit (visit month 0), and 118 patients with PD, 2 patients with ET, 2 patients with MSA, and 90 controls ($n$ = 212 participants) at last visit.[35] The sarcoidosis dataset of Damsky et al. contains 11 healthy controls, 9 untreated patients with sarcoidosis, and 11 patients with sarcoidosis treated with tofacitinib.[36] The Multi-Ethnic Study of Atherosclerosis (MESA) study of Bild et al. contains SomaScan data for 755 proteins measured in 921 participants at two visits each.[17]

## Quantification and statistical analysis

### UK Biobank data processing

Grip strength of the strongest hand was defined as the highest grip strength between the left (UK Biobank field 46) and the right (UK Biobank field 47) hands, and grip strength of the weakest hand was defined as the lowest grip strength between both hands. Mother's and father's age of death were derived from fields 3526 and 1807, respectively. Participants with unknown ages of death for the parents, or those who replied "Yes", "Prefer not to answer", or "Do not know" to the question "Were you adopted as a child?" (field 1767) were excluded from the analyses related to parental age of death.

Mortality was recorded as the time of death, as stored in the death records (UK Biobank field 40000). Participants for whom the primary cause of death (UK Biobank field 40001) is part of Chapter XX (External causes of morbidity and mortality) were considered lost to follow-up at their moment of death. Heart failure was defined as the first reported occurrence of heart failure (field 131354). Dementia was defined as the first reported occurrence of any of

Alzheimer's disease (field 130836), vascular dementia (field 130838), other dementia (field 130840), or unspecified dementia (130842). Kidney failure was defined as the first reported occurrence of any of acute renal failure (field 132030), chronic renal failure (field 132032), or unspecified renal failure (field 132034). Stroke was defined as the first reported occurrence of stroke, not specified as haemorrhage or infarction (field 131368). Myocardial infarction was defined as the first occurrence of acute (field 131298) or subsequent myocardial infarction (field 131300). Chronic obstructive pulmonary disease (COPD) was defined as the first reported occurrence of other chronic obstructive pulmonary disease (field 131492). Type 2 diabetes was defined as the first reported occurrence of non-insulin-dependent diabetes mellitus (field 130708). Liver cirrhosis/fibrosis was defined as the first reported occurrence of fibrosis and cirrhosis of liver (field 131666). Hypertension was defined as the first reported occurrence of any of essential (primary) hypertension (field 131286), or secondary hypertension (field 131294). The first cancer occurrence was based on field 40005. Food intake was based on the total weight of each food group yesterday (UK Biobank category 100118).[66,67] This data comes from online 24-hour recall dietary questionnaires that were e-mailed by UK Biobank to participants after their first visit. Job codings were obtained from job coding (field 22601) and, for each occupation, were assigned to that occupation if they have ever worked in it for at least 6 months for at least 15 hours per week. Smoking status was obtained from UK Biobank field 20116, collected at first visit.

Treatment/medication codes in the UK Biobank (field 20003) were streamlined and categorized through the following process. First, we applied mapping data generated by Wu et al.[68] to map UK biobank medications taken by at least 10 subjects to their respective Anatomical Therapeutic Chemical (ATC) codes. ATC provides a classification system that enabled us to hierarchically categorize each UK medication into 5 distinct levels of classification.[69] We downloaded the ATC categories from BioPortal, formatted them so each row corresponded to each of 5 classifications for a given drug, and then merged each drug with the UK biobank codes, creating a final table of medications with integrated UK biobank drugs and ATC classifications.

**Effects of individual proteins on chronological age and mortality**

The effects of the individual plasma proteins on chronological age after correction for sex were assessed using ordinary linear regression. More specifically, for each protein $p$, we constructed the following model:

$$\text{age}_i = \beta^0 + x_i^p \beta^p + x_i^{\text{sex}} \beta^{\text{sex}} + \varepsilon_i \text{ (Equation 1)}$$

Herein, $\text{age}_i$ is the chronological age of the $i$th individual, $\beta^0$ is a fixed intercept, $x_i^p$ is the median-centered Olink Normalized Protein Expression (NPX) value for protein $p$ and

participant $i$, as provided by UK Biobank Record Table 1072. $\beta^{p}$ is the effect of protein $p$, $x_i^{sex}$ is a dummy variable equal to 1 if participant $i$ is male, and 0 if participant $i$ is female. $\beta^{\text{sex}}$ is the effect of male sex, and $\varepsilon_i$ is a random error term.

The effects of the individual plasma proteins on mortality were assessed in a similar way with Cox proportional hazards models with the time difference between death and the first visit as an outcome for the deceased subjects, and the time difference between the latest recorded death and the first visit as the outcome for the surviving subjects. Here, we included main effects for chronological age and sex, and the interaction between chronological age and sex as covariates in the model.

Gene set enrichment analysis (GSEA) was performed on Z values, using the adaptive multilevel splitting Monte Carlo approach implemented in the fgseaMultilevel function from the fgsea R package v 1.20.0. Gene sets were obtained from MSigDB and included gene ontology: biological process (GObp), KEGG, Reactome, and Hallmarks.

The gene expression signatures of chronological aging and mortality in mouse and human blood, human tissues (multi-tissue), mouse and rat tissues (multi-tissue), and mouse, rat and human tissues (multi-tissue) were obtained from Tyshkovskiy et al.[4,28] Correlation analysis was performed at the level of individual genes and proteins using slope estimates and at the level of enriched pathways using normalized enrichment scores (NES) derived from GSEA.

**Aging models**

The Olink proteomics data was obtained from UK Biobank Record Table 1072. Participants with > 49.9% missing values and proteins with > 10% missing values were removed from the analysis, retaining 44,952 out of 53,014 participants and 2,916 out of 2,923 proteins. The UK Biobank data (44,952 individuals) was then randomly split into five non-overlapping test sets ($n$ = 8990 for the first four folds and $n$ = 8992 for the fifth fold; K-fold validation with K = 5). For each of these five test datasets, the remaining ~80% of the data were used as training datasets. These 5 training and 5 test proteomics datasets were then each imputed using the k-nearest neighbors algorithm, with the default value of k = 10, as implemented in the impute.knn function from the impute R package v 1.68.0, and the data was imported into Python. Chronological aging (1$^{st}$-generation) models were trained using elastic net regression on the plasma proteomics data, with chronological age as the outcome. Mortality-based models were trained using Cox proportional hazards elastic net regression with the time difference between death and the first visit as an outcome for the deceased subjects, and the time difference between the latest recorded death in the whole dataset and the first visit as the outcome for the surviving subjects. The Elastic Net regularization

technique has previously been shown to be effective[70] and combines the properties of both LASSO (Least Absolute Shrinkage and Selection Operator) and ridge regression, allowing for simultaneous feature selection and shrinkage by linearly combining the L1 and L2 penalties. Subjects that died due to external causes were modeled as if they were lost to follow up at their moment of death. Both models were optimized for the mean absolute error (MAE) of the residuals. When results are shown for training and test datasets, we always used the models generated in the first K-fold, unless explicitly stated otherwise. The coefficients for these models are available in Table S1. The results from the first fold were also used for all validation analyses on external datasets. Any missing protein values in the external datasets were excluded when calculating predicted ages or predicted log(mortality hazards). For the SomaScan data in the MESA dataset,[17] we first $\log_2$ transformed the SomaScan data, performed mean centering, and rescaled with the standard deviations obtained from the UK Biobank Olink data (given in Table S3). All results shown for the full UK Biobank dataset are based on out-of-fold predictions, unless explicitly stated otherwise. This means that each outcome is calculated with the model trained on the four other folds, excluding the fold with the participant in question. Age deviations for all 1$^{st}$-generation and mortality-based models are defined as the residuals from an ordinary linear regression of predicted biological age on chronological age.

Robust regression lines with 95% confidence bands shown in figures are based on robust linear regression (M estimation with Huber weights), as implemented by default in the rlm function from the MASS R package. Curves shown in Figures S4J–S4R are obtained from generalized additive models as implemented in the mgcv R package, whereby the log(mortality hazard) deviation was modeled as a penalized thin-plate regression spline.

Extreme agers as shown in Figure 1J were defined as follows. We normalized the age deviations of the conventional mortality-based model by dividing them by the standard deviation of all age deviations. We then converted these normalized age deviations to p-values, assuming them to follow a standard normal distribution under the null hypothesis of no age deviation. We then converted the p-values to q-values using the Benjamini-Hochberg false discovery rate (BH-FDR) method. Any participant with a q-value < 0.05 was considered to have a significant age deviation and therefore an extreme ager. We identified 359 of such individuals. First occurrences of ICD-10 disease codes were based on UK Biobank Category 1712, which contains information from primary care data, hospital inpatient data, death register records and self-reported medical condition codes mapped to 3-character ICD-10 codes.

**Annotation of proteins**

To identify organ-enriched proteins, we incorporated information related to tissue transcription from the Genotype-Tissue Expression (GTEx) project,[29] in an approach similar to what has been done in Oh et al.[10] We grouped the GTEx tissues into organs according to Supplementary Table 2 from these authors. A protein was then designated as organ-enriched if the average GTEx gene counts for an organ were at least four times higher than all other organs. We decided to make use of transcription data and not proteomics because transcription can inform about the organ of origin of a protein (see the "limitations of the study" section for a broader discussion).

Organ-specific aging models were trained only on the subset of proteins enriched in a certain organ, in the same way as explained above for the conventional models. Finally, we only retained those organ-specific aging models for which the correlation coefficient (r) was consistently larger than 30% for both the chronological aging model and the mortality-based aging model, in both training and test sets (Figures 3B and 3C).

For the analysis in Figures 2G and 2H, we downloaded the human proteome from Uniprot on February 27, 2024. Proteins were considered secreted if they were annotated as belonging to the extracellular space in Gene Ontology cellular component (GOcc), or annotated as secreted in Uniprot's subcellular location. Proteins were considered to belong to the extracellular membrane if they were not secreted and annotated as belonging to the plasma membrane in Gene Ontology cellular component (GOcc), or annotated as cell membrane in Uniprot's subcellular location. All other proteins were considered to be intracellular. In Figures 2I and 2J, secreted proteins are the same as for the analysis in Figures 2G and 2H, and non-secreted proteins are all other proteins (i.e. intracellular and extracellular membrane proteins).

**Associations with environmental factors**

Associations with environmental factors as shown in Figures 5A–5F were performed with the following ordinary linear regression models:

$$\begin{aligned}\mathrm{BioAge}_i = \beta^0 &+ x_i^{\mathrm{age}}\beta^{\mathrm{age}} + x_i^{\mathrm{sex}}\beta^{\mathrm{sex}} + x_i^{\mathrm{age}}x_i^{\mathrm{sex}}\beta^{\mathrm{age:sex}} + x_i^{\mathrm{predictor}}\beta^{\mathrm{predictor}} \\ &+ x_i^{\mathrm{center}}\beta^{\mathrm{center}} + x_i^{\mathrm{deprivation}}\beta^{\mathrm{deprivation}} + x_i^{\mathrm{activity}}\beta^{\mathrm{activity}} \\ &+ x_i^{\mathrm{smoking}}\beta^{\mathrm{smoking}} + \varepsilon_i \quad \text{(Equation 2)}\end{aligned}$$

Herein, $\mathrm{BioAge}_i$ denotes the predicted relative log(mortality hazard) as calculated by the mortality-based aging models for the $i$th participant. $\beta^0$ is the intercept. Throughout the manuscript, we use the abbreviation "log" for the natural logarithm (with base *e*) and "$\log_{10}$" for the Briggs logarithm (with base 10). $x_i^{\mathrm{age}}$ is the chronological age. $\beta^{\mathrm{age}}$ is the effect of chronological age, $x_i^{sex}$ is a dummy variable equal to 1 if participant $i$ is male, and 0 if

participant $i$ is female. $\beta^{\text{sex}}$ is the effect of male sex, $\beta^{\text{age:sex}}$ is the effect of the age-sex interaction.

$x_i^{\text{predictor}}$ is a dummy variable or continuous variable that denotes the predictor of interest (food, profession, medication, smoking or drinking), $\beta^{\text{predictor}}$ is the effect of the predictor of interest.

$x_i^{\text{center}}$ is a set of dummy variables to denote the UK Biobank assessment center (field 54), $x_i^{\text{deprivation}}$ is the Townsend deprivation index (field 22189), $x_i^{\text{activity}}$ is a set of dummy variables to denote the IPAQ activity group ("low", "moderate" or "high", field 22032), $x_i^{\text{smoking}}$ is a set of dummy variables to denote the smoking status ("never, "previous", "current" or "prefer not to answer", field 20116). The corresponding $\beta$ values are their respective effects, and $\varepsilon_i$ is a random error term.

For the analyses with foods as the predictors of interest, participants for who the reported diet was not typical (field 100020) or not credible (field 100026) were excluded from the analysis.

For the analyses with smoking as the predictor of interest, we only included current and never smokers, excluding previous smokers and people who prefer not to answer, and we obviously did not include the redundant $x_i^{\text{smoking}}\beta^{\text{smoking}}$ term.

For the metformin analysis in Figure 5D, we additionally included a confounder based on UK Biobank field 130708 indicating the type 2 diabetes status (either no type 2 diabetes, type 2 diabetes before the first visit, or type 2 diabetes diagnosed later on after the first visit).

For the simvastatin analysis in Figure 5D, we additionally included a confounder indicating whether the participants reported to regularly take lipid-modifying agents at first visit (based on UK Biobank field 20003), as a proxy for whether high cholesterol was diagnosed.

**Effects of individual proteins on organ-specific aging models**

The interpretability of elastic net approaches allows us to assess the effect of each protein on the outcome. For the analysis in Figures 4F–4I, we fit the following Cox proportional-hazards model for each combination of outcome and organ-specific model $j$:

$$\begin{aligned}\mathrm{H}(\mathrm{t})^{\text{outcome}_j} = \mathrm{H}_0(\mathrm{t})^{\text{outcome}_j} \\ + \exp\big(x^{age}\beta^{age} + x^{sex}\beta^{sex} + x^{age}x^{sex}\beta^{age:sex} + x^{BioAge_j}\beta^{BioAge_j} \\ + \varepsilon\big)\ (\text{Equation 3})\end{aligned}$$

Herein, $\mathrm{H}(\mathrm{t})^{\text{outcome}_j}$ is the hazard for outcome $j$, $\mathrm{H}_0(\mathrm{t})^{\text{outcome}_j}$ is the baseline hazard for outcome $j$. $x^{\text{age}}$ is the chronological age. $\beta^{\text{age}}$ is the effect of chronological age, $x_i^{sex}$ is a

dummy variable equal to 1 if the participant is male, and 0 if the participant is female. $\beta^{\text{sex}}$ is the effect of male sex, $\beta^{\text{age:sex}}$ is the effect of the age-sex interaction, $x^{\text{BioAge}_j}$ is the log(mortality hazard) predicted by organ-specific mortality-based model $j$, and $\varepsilon$ is a random error term. For each protein $p$ in model $j$, we then fit the exact same Cox proportional-hazards model described in Equation 3, in which the $x^{\text{BioAge}_j}$ is replaced by the log(mortality hazard) predicted by organ-specific model $j$ on the data from all proteins except protein $p$. We then perform a partial likelihood-ratio test for non-nested Cox proportional-hazards models, as implemented in the plrtest function from the nonnestcox R package to the model given in Equation 3 with the latter model.[64,71]

For the analysis in Figures 5G–5J, we used the same ordinary linear regression models as specified in Eq. 2, only now including $\text{BioAge}$ as a predictor and the predictor of interest from that model as an outcome. We then used a similar approach as above, in which we re-calculate $\text{BioAge}$, each time excluding one protein. When comparing both models, we used Vuong's 1989 non-nested likelihood ratio test, as implemented in the vuongtest function from the nonnest2 R package.[63,72]

The analysis in Figures 7E–7H is similar (COVID status as an outcome was coded as either 0 (negative) or 1 (positive); acuity was coded numerically from 0 (discharged) to 4 (dead). For the analysis in Figures 7Q–7T, controls were coded as 0 and PD, respectively MSA, as 1.

**Longitudinal analyses and external validation**

UK Biobank has unique follow-up plasma proteomics data for 1,161 participants at their third visits, and for 1,113 participants at their fourth visits. For the longitudinal analyses, we only considered the 1,463 proteins that were measured with the Olink Explore 1536 platform in the samples taken at the third and fourth visits. For model training, we only retained the 52,341 participants at first visit with < 49.9% missing values over these 1,463 proteins. Similarly, we only retained the subset of 1,451 proteins with < 10% missing values at each of the three visits (first, third and fourth visits). We then re-trained all mortality-based models using these 1,451 proteins ("feature-reduced" models). Here, the number of subjects in the test sets equals 10,468 for the first four folds and 10,469 for the fifth fold. The coefficients for these models are available in Table S5.

The Filbin et al.,[33] Dammer et al.[35] and Damsky et al.[36] datasets used the Olink Explore 1536 platform (as opposed to the Olink Explore 3072 platform that was used for the samples from the initial visit in UK Biobank). We therefore applied the first fold of our feature-reduced models trained in UK Biobank to validate our findings in these datasets. As the Filbin et al.[33] dataset does not contain exact ages, we replaced each given age category with the median chronological age in that category and arbitrarily imputed 85 as the chronological age for the

80+ age category for all analyses involving chronological age, including the calculation of the age deviations.

The 1st-generation aging models trained directly on CSF from the Dammer et al.[35] dataset as shown in Figure S6 were only trained on the 1,031 proteins that were reliably detected in CSF. The coefficients for these models are given in Table S6. We used leave-one-out cross-validation (LOO CV) with k-nearest neighbors (kNN) imputation (k = 10) to train these models. We used leave-one-out-predictions to calculate the biological ages at first visit (*n* = 184 participants). For the observations at later timepoints, we predicted biological ages based on the first of the 184 LOO models.

For the longitudinal analysis, rates of aging were calculated for the 985 participants with Olink proteomics data available for the first, third, and fourth visit, based on the following mixed-effects model:

$$\mathrm{BioAge}_{ij} = \beta^0 + x_{ij}^{\Delta\mathrm{age}}\beta^{\mathrm{slope}} + u_i^{\mathrm{intercept}} + u_i^{\mathrm{slope}} + \varepsilon_{ij} \quad \text{(Equation 4)}$$

Herein, $\mathrm{BioAge}_{ij}$ denotes the predicted relative log(mortality hazard) as calculated by the mortality-based aging models for the $i$th participant on the $j$th visit, $\beta^0$ is a fixed intercept, $x_{ij}^{\Delta\mathrm{age}}$ is the difference in chronological age between the $j$th visit and the first visit for the $i$th participant. The fixed effect $\beta^{\mathrm{slope}}$ is the overall change in predicted relative log(mortality hazard) per year of chronological age. $u_i^{\mathrm{intercept}}$ is a random effect that allows for participant-specific intercepts, $u_i^{\mathrm{slope}}$ is a random effect that allows for participant-specific changes in predicted relative log(mortality hazard) per year of chronological age. $\varepsilon_{ij}$ is a random error term. All random effects and the random error term are assumed to follow normal distributions. Mixed-effects models were implemented with the lmer function from lme4 R package. The longitudinal slope for each individual $i$ was calculated as $\hat{\beta}^{\mathrm{slope}} + \hat{u}_i^{\mathrm{slope}}$ (with hats indicating the estimators for the corresponding model parameters).

Hazard ratios from the Oh et al. study for the LonGenity cohort were obtained from their Supplementary Table 12.[10] Hazard ratios for the Lu et al. study were obtained from their Supplementary Table 5.[22] They were standardized by dividing the regression coefficient and its confidence intervals on the log-scale by the standard deviation of their age gaps for the LonGenity cohort (Supplementary Table 8 from Oh et al.[10]) and the standard deviation of their AgeAccelGrim phenotypes (Supplementary Table 9 from Lu et al.[22]), respectively, to calculate standardized regression coefficients. Standardized hazard ratios and their confidence intervals were then calculated by taking the exponential function of standardized effect sizes.

**Statistical testing**

Unless indicated otherwise, we used the equal-variance two-sided two-sample t-test, as implemented in the R stats package, when comparing two groups without covariates. If covariates were included, we used ordinary least squares linear regression. For time-to-event analyses, we used Cox proportional-hazards models.

**Multiple testing corrections**

We controlled for multiple testing wherever appropriate. When performing a small number of tests simultaneously (<10) within the same model framework, we controlled the family-wise error rate at 5% using the default single-step method, as implemented in the glht function of the multcomp R package. When performing a small number of tests simultaneously (<10) outside of the same model framework, we controlled the family-wise error rate at 5% using the Hommel procedure. For larger numbers of comparisons (≥10), we controlled the false discovery rate (FDR) at 5% with the Benjamini-Hochberg FDR method.

**Reproducing figures**

Excel files containing the values that were used to create the graphs in the paper can be found in Data S1. Note that it is not permitted to share individual-level data from the UK Biobank. For figures based on individual-level data, we therefore used a nearly-equal k-means clustering algorithm (k = 10) as implemented in the kMeansEqual R package to group individuals in a sex-specific way and provide cluster averages instead.

Similarly, the Dammer et al.[35] dataset is also not publicly available, as researchers need Synapse account and agree to its Terms and Conditions. For this smaller dataset, we used the same approach with k = 5. For violin plots related to UK Biobank and Dammer et al., we provide the minimum and maximum, as well as 0.1, 1, 5, 10, 20, 25, 30, 40, 50, 60, 70, 75, 80, 90, 95, 99, and 99.9 percentiles. For the ET and MSA categories in the Dammer et al.[35] dataset, which each contain only two patients, we only provide the 50% quantile (i.e. the median).

**Software**

All analyses have been carried out using freely available software packages in Python and R. Aging models were trained using the scikit-learn library version 1.3.2 using Python version 3.12.3. All other analyses were performed in RStudio Pro 2022.12.0 Build 353.pro20 "Elsbeth Geranium" running R version 4.1.2 (2021-11-01) on a 64-bit CentOS Linux 7 server. When needed, manual editing of figures was done in Inkscape v. 1.3.2. Some color palettes used in this publication were obtained from or inspired by the MetBrewer R package.

## Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.cmet.2024.10.005.

**Document S1.** Figures S1–S7.

**Table S1.** Model coefficients and hyperparameters for the full conventional and organ-specific aging models, related to STAR Methods. (A) Model coefficients for the 5 folds of the 1st-generation conventional and organ-specific models. Proteins that are not specific for the corresponding organ have empty values.

(B) Model hyperparameters for the 5 folds of the 1st-generation conventional and organ-specific models.

(C) Model coefficients for the 5 folds of the mortality-based conventional and organ-specific models. Proteins that are not specific for the corresponding organ have empty values.

(D) Model hyperparameters for the 5 folds of the mortality-based conventional and organ-specific models.

**Table S2.** Correlation and accuracy measures of biological ages predicted by conventional and organ-specific models, related to Figures 1, 3, and 6. (A) Correlation and accuracy measures of biological ages predicted by the full conventional and organ-specific 1st-generation and mortality-based models.

(B) Correlation and accuracy measures of biological ages predicted by the feature-reduced conventional and organ-specific 1st-generation and mortality-based models.

r: correlation coefficient, $r^2$: coefficient of determination, MAE: mean absolute error of the residuals, MSE: mean squared error, pval: p-value.

**Table S3.** Standard deviations of the Olink NPX values in the datasets used to train the full models and the feature-reduced models, related to STAR Methods. Empty values indicate that the protein was not considered for model building.

**Table S4.** Effects on chronological age and mortality for 2,923 proteins and their associated pathways in 53,016 UK Biobank participants, related to Figure 2. (A) Effects on chronological age after correction for sex.

(B) Gene set enrichment analysis using MSigDB pathways on the Z-values in Table S4A.

(C) Effects on mortality after correction for sex, chronological age, and the interaction between sex and chronological age.

(D) Gene set enrichment analysis using MSigDB pathways on the Z-values in Table S4C.

ln: natural logarithm, pval: enrichment p-value, padj: Benjamini-Hochberg-adjusted p-value, log2err: the expected error for the standard deviation of the $\log_2$(p-value), ES: enrichment score, NES: normalized enrichment score, size: size of the pathway after removing genes not present in the input, leadingEdge: vector with indexes of leading edge genes that drive the enrichment, zval: Z value.

**Table S5.** Model coefficients and hyperparameters for the feature-reduced conventional and organ-specific aging models, related to STAR Methods. (A) Model coefficients for the 5 folds of the feature-reduced 1st-generation conventional and organ-specific models. Proteins that are not specific for the corresponding organ have empty values.

(B) Model hyperparameters for the 5 folds of the feature-reduced 1st-generation conventional and organ-specific models.

(C) Model coefficients for the 5 folds of the feature-reduced mortality-based conventional and organ-specific models. Proteins that are not specific for the corresponding organ have empty values.

(D) Model hyperparameters for the 5 folds of the feature-reduced mortality-based conventional and organ-specific models.

**Table S6.** Model coefficients for the 1st-generation conventional and organ-specific models trained on 1,031 proteins measured in the CSF in the Dammer et al. (2022) dataset, related to STAR Methods. Model coefficients for the 1st-generation conventional and organ-specific models trained on 1,031 proteins measured in the CSF in the Dammer et al. (2022) dataset. Proteins that were not reliably measured in the CSF or are not specific for the corresponding organ have empty values.

**Data S1.** Unprocessed data underlying the display items in the manuscript, related to Figures 1–7 and S1–S7.

**Document S2.** Article plus supplemental information.

## References


1. V.N. Gladyshev, S.B. Kritchevsky, S.G. Clarke, A.M. Cuervo, O. Fiehn, J.P. de Magalhães, T. Mau, M. Maes, R.L. Moritz, L.J. Niedernhofer, *et al.* Molecular damage in aging. Nat. Aging, 1 (2021), pp. 1096-1106, 10.1038/s43587-021-00150-3

2. S. Horvath. DNA methylation age of human tissues and cell types. Genome Biol., 14 (2013), Article R115, 10.1186/GB-2013-14-10-R115/COMMENTS

3. M.J. Peters, R. Joehanes, L.C. Pilling, C. Schurmann, K.N. Conneely, J. Powell, E. Reinmaa, G.L. Sutphin, A. Zhernakova, K. Schramm, *et al.* The transcriptional landscape of age in human peripheral blood. Nat. Commun., 6 (2015), Article 8570, 10.1038/ncomms9570

4. A. Tyshkovskiy, D. Kholdina, K. Ying, M. Davitadze, A. Molière, Y. Tongu, T. Kasahara, L.M. Kats, A. Vladimirova, A. Moldakozhayev, *et al.* Transcriptomic Hallmarks of Mortality Reveal Universal and Specific Mechanisms of Aging, Chronic Disease, and Rejuvenation. Preprint at bioRxiv (2024), 10.1101/2024.07.04.601982

5. B. Lehallier, D. Gate, N. Schaum, T. Nanasi, S.E. Lee, H. Yousef, P. Moran Losada, D. Berdnik, A. Keller, J. Verghese, *et al.* Undulating changes in human plasma proteome profiles across the lifespan. Nat. Med., 25 (2019), pp. 1843-1850, 10.1038/s41591-019-0673-2

6. M.A. Argentieri, S. Xiao, D. Bennett, L. Winchester, A.J. Nevado-Holgado, U. Ghose, A. Albukhari, P. Yao, M. Mazidi, J. Lv, *et al.* Proteomic aging clock predicts mortality and risk of common age-related diseases in diverse populations. Nat. Med., 30 (2024), pp. 2450-2460, 10.1038/s41591-024-03164-7

7. C.-L. Kuo, Z. Chen, P. Liu, L.C. Pilling, J.L. Atkins, R.H. Fortinsky, G.A. Kuchel, S. Diniz. Proteomic aging clock (PAC) predicts age-related outcomes in middle-aged and older adults. Aging Cell, 23 (2024), Article e14195, 10.1111/ACEL.14195

8. M. Moqri, C. Herzog, J.R. Poganik, Biomarkers of Aging Consortium, J. Justice, D.W. Belsky, A. Higgins-Chen, A. Moskalev, G. Fuellen, A.A. Cohen, *et al.* Biomarkers of aging for the identification and evaluation of longevity interventions. Cell, 186 (2023), pp. 3758-3775, 10.1016/J.CELL.2023.08.003

9. F. Prattichizzo, C. Frigé, V. Pellegrini, L. Scisciola, A. Santoro, D. Monti, M.R. Rippo, M. Ivanchenko, F. Olivieri, C. Franceschi. Organ-specific biological clocks: Ageotyping for personalized anti-aging medicine. Ageing Res. Rev., 96 (2024), Article 102253, 10.1016/J.ARR.2024.102253

10. H.S.H. Oh, J. Rutledge, D. Nachun, R. Pálovics, O. Abiose, P. Moran-Losada, D. Channappa, D.Y. Urey, K. Kim, Y.J. Sung, *et al.* Organ aging signatures in the plasma proteome track health and disease. Nature, 624 (2023), pp. 164-172, 10.1038/s41586-023-06802-1

11. C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, *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 Med., 12 (2015), Article e1001779, 10.1371/JOURNAL.PMED.1001779

12. G.H. Eldjarn, E. Ferkingstad, S.H. Lund, H. Helgason, O.T. Magnusson, K. Gunnarsdottir, T.A. Olafsdottir, B.V. Halldorsson, P.I. Olason, F. Zink, *et al.* Large-scale plasma proteomics comparisons through genetics and disease associations. Nature, 622 (2023), pp. 348-358, 10.1038/s41586-023-06563-x

13. M.E. Levine, A.T. Lu, A. Quach, B.H. Chen, T.L. Assimes, S. Bandinelli, L. Hou, A.A. Baccarelli, J.D. Stewart, Y. Li, *et al.* An epigenetic biomarker of aging for lifespan and healthspan. Aging, 10 (2018), pp. 573-591, 10.18632/AGING.101414

14. M. Gögele, C. Pattaro, C. Fuchsberger, C. Minelli, P.P. Pramstaller, M. Wjst. Heritability Analysis of Life Span in a Semi-isolated Population Followed Across Four Centuries Reveals the Presence of Pleiotropy Between Life Span and Reproduction. J. Gerontol. A Biol. Sci. Med. Sci., 66 (2011), pp. 26-37, 10.1093/GERONA/GLQ163

15. J. Kaplanis, A. Gordon, T. Shor, O. Weissbrod, D. Geiger, M. Wahl, M. Gershovits, B. Markus, M. Sheikh, M. Gymrek, *et al.* Quantitative analysis of population-scale family trees with millions of relatives. Science, 360 (2018), pp. 171-175, 10.1126/science.aam9309

16. J.G. Ruby, K.M. Wright, K.A. Rand, A. Kermany, K. Noto, D. Curtis, N. Varner, D. Garrigan, D. Slinkov, I. Dorfman, *et al.* Estimates of the Heritability of Human Longevity Are Substantially Inflated due to Assortative Mating. Genetics, 210 (2018), pp. 1109-1124, 10.1534/GENETICS.118.301613

17. D.E. Bild, D.A. Bluemke, G.L. Burke, R. Detrano, A.V. Diez Roux, A.R. Folsom, P. Greenland, D.R. Jacob, R. Kronmal, K. Liu, *et al.* Multi-Ethnic Study of Atherosclerosis: Objectives and Design. Am. J. Epidemiol., 156 (2002), pp. 871-881, 10.1093/AJE/KWF113

18. L. Gold, D. Ayers, J. Bertino, C. Bock, A. Bock, E.N. Brody, J. Carter, A.B. Dalby, B.E. Eaton, T. Fitzwater, *et al.* Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. PLoS One, 5 (2010), Article e15004, 10.1371/JOURNAL.PONE.0015004

19. E.M. Crimmins, H. Shim, Y.S. Zhang, J.K. Kim. Differences between Men and Women in Mortality and the Health Dimensions of the Morbidity Process. Clin. Chem., 65 (2019), pp. 135-145, 10.1373/CLINCHEM.2018.288332

20. G. Hannum, J. Guinney, L. Zhao, L. Zhang, G. Hughes, S.V. Sadda, B. Klotzle, M. Bibikova, J.B. Fan, Y. Gao, *et al.* Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. Mol. Cell, 49 (2013), pp. 359-367, 10.1016/j.molcel.2012.10.016

21. R.F. Hillary, A.J. Stevenson, D.L. McCartney, A. Campbell, R.M. Walker, D.M. Howard, C.W. Ritchie, S. Horvath, C. Hayward, A.M. McIntosh, *et al.* Epigenetic measures of ageing predict the prevalence and incidence of leading causes of death and disease burden. Clin. Epigenetics, 12 (2020), p. 115, 10.1186/s13148-020-00905-6

22. A.T. Lu, A. Quach, J.G. Wilson, A.P. Reiner, A. Aviv, K. Raj, L. Hou, A.A. Baccarelli, Y. Li, J.D. Stewart, *et al.* DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging, 11 (2019), pp. 303-327, 10.18632/AGING.101684

23. M. Moqri, C. Herzog, J.R. Poganik, K. Ying, J.N. Justice, D.W. Belsky, A.T. Higgins-Chen, B.H. Chen, A.A. Cohen, G. Fuellen, *et al.* Validation of biomarkers of aging. Nat. Med., 30 (2024), pp. 360-372, 10.1038/S41591-023-02784-9

24. J.S. Moon, L.J.E. Goeminne, J.T. Kim, J.W. Tian, S.H. Kim, H.T. Nga, S.G. Kang, B.E. Kang, J.S. Byun, Y.S. Lee, *et al.* Growth differentiation factor 15 protects against the aging-mediated systemic inflammatory response in humans and mice. Aging Cell, 19 (2020), Article e13195, 10.1111/ACEL.13195

25. J. You, Y. Guo, Y. Zhang, J.J. Kang, L.B. Wang, J.F. Feng, W. Cheng, J.T. Yu. Plasma proteomic profiles predict individual future health risk. Nat. Commun., 14 (2023), Article 7817, 10.1038/S41467-023-43575-7

26. S.A. Kaeser, B. Lehallier, M. Thinggaard, L.M. Häsler, A. Apel, C. Bergmann, D. Berdnik, B. Jeune, K. Christensen, S. Grönke, *et al.* A neuronal blood marker is associated with mortality in old age. Nat. Aging, 1 (2021), pp. 218-225, 10.1038/s43587-021-00028-4

27. Y. Guo, J. You, Y. Zhang, W.S. Liu, Y.Y. Huang, Y.R. Zhang, W. Zhang, Q. Dong, J.F. Feng, W. Cheng, *et al.* Plasma proteomic profiles predict future dementia in healthy adults. Nat. Aging, 4 (2024), pp. 247-260, 10.1038/s43587-023-00565-0

28. A. Tyshkovskiy, S. Ma, A.V. Shindyapina, S. Tikhonov, S.G. Lee, P. Bozaykut, J.P. Castro, A. Seluanov, N.J. Schork, V. Gorbunova, *et al.* Distinct longevity mechanisms across and within species and their association with aging. Cell, 186 (2023), pp. 2929-2949.e20, 10.1016/J.CELL.2023.05.002

29. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat. Genet., 45 (2013), pp. 580-585, 10.1038/ng.2653

30. T.V. Pyrkov, E. Getmantsev, B. Zhurov, K. Avchaciov, M. Pyatnitskiy, L. Menshikov, K. Khodova, A.V. Gudkov, P.O. Fedichev. Quantitative characterization of biological age and frailty based on locomotor activity records. Aging, 10 (2018), pp. 2973-2990, 10.18632/AGING.101603

31. S.A. Burgard, K.Y. Lin. Bad Jobs, Bad Health? How Work and Working Conditions Contribute to Health Disparities. Am. Behav. Sci., 57 (2013), pp. 1105-1127, 10.1177/0002764213487347

32. C.H. Kuo, L.L. Wu, H.P. Chen, J. Yu, C.Y. Wu. Direct effects of alcohol on gut-epithelial barrier: Unraveling the disruption of physical and chemical barrier of the gut-epithelial barrier that compromises the host–microbiota interface upon alcohol exposure. J. Gastroenterol. Hepatol., 39 (2024), pp. 1247-1255, 10.1111/JGH.16539

33. M.R. Filbin, A. Mehta, A.M. Schneider, K.R. Kays, J.R. Guess, M. Gentili, B.G. Fenyves, N.C. Charland, A.L.K. Gonye, I. Gushterova, *et al.* Longitudinal proteomic analysis of severe COVID-19 reveals survival-associated signatures, tissue-specific cell death, and

cell-cell interactions. Cell Rep. Med., 2 (2021), Article 100287, 10.1016/J.XCRM.2021.100287

34. X. Cao, W. Li, T. Wang, D. Ran, V. Davalos, L. Planas-Serra, A. Pujol, M. Esteller, X. Wang, H. Yu. Accelerated biological aging in COVID-19 patients. Nat. Commun., 13 (2022), p. 2135, 10.1038/s41467-022-29801-8

35. E.B. Dammer, L. Ping, D.M. Duong, E.S. Modeste, N.T. Seyfried, J.J. Lah, A.I. Levey, E.C.B. Johnson. Multi-platform proteomic analysis of Alzheimer's disease cerebrospinal fluid and plasma reveals network biomarkers associated with proteostasis and the matrisome. Alzheimers Res. Ther., 14 (2022), p. 174, 10.1186/s13195-022-01113-5

36. W. Damsky, A. Wang, D.J. Kim, B.D. Young, K. Singh, M.J. Murphy, J. Daccache, A. Clark, R. Ayasun, C. Ryu, *et al.* Inhibition of type 1 immunity with tofacitinib is associated with marked improvement in longstanding sarcoidosis Nat. Commun., 13 (2022), Article 3140, 10.1038/s41467-022-30615-x

37. T.V. Gladyshev, V.N. Gladyshev. A Disease or Not a Disease? Aging As a Pathology. Trends Mol. Med., 22 (2016), pp. 995-996, 10.1016/j.molmed.2016.09.009

38. E.G. Lakatta. So! What's aging? Is cardiovascular aging a disease? J. Mol. Cell. Cardiol., 83 (2015), pp. 1-13, 10.1016/J.YJMCC.2015.04.005

39. W. MacNee. Is chronic obstructive pulmonary disease an accelerated aging disease? Ann. Am. Thorac. Soc., 13 (Suppl 5) (2016), pp. S429-S437, 10.1513/AnnalsATS.201602-124AW

40. O. Kuss, J. Baumert, C. Schmidt, T. Tönnies. Mortality of type 2 diabetes in Germany: additional insights from Gompertz models. Acta Diabetol., 61 (2024), pp. 765-771, 10.1007/s00592-024-02237-w

41. A. Xanthopoulos, R.C. Starling, T. Kitai, F. Triposkiadis. Heart Failure and Liver Disease: Cardiohepatic Interactions. JACC Heart Fail., 7 (2019), pp. 87-97, 10.1016/J.JCHF.2018.10.007

42. M.K. Nadim, G. Garcia-Tsao. Acute Kidney Injury in Patients with Cirrhosis. N. Engl. J. Med., 388 (2023), pp. 733-745, 10.1056/NEJMra2215289

43. A.G. Hoek, S. van Oort, K.J. Mukamal, J.W.J. Beulens. Alcohol Consumption and Cardiovascular Disease Risk: Placing New Data in Context. Curr. Atheroscler. Rep., 24 (2022), pp. 51-59, 10.1007/S11883-022-00992-1

44. E. La Porta, P. Baiardi, L. Fassina, A. Faragli, S. Perna, F. Tovagliari, I. Tallone, G. Talamo, G. Secondo, G. Mazzarello, *et al*. The role of kidney dysfunction in COVID-19 and the influence of age. Sci. Rep., 12 (2022), Article 8650, 10.1038/s41598-022-12652-0

45. M. Mavrikaki, J.D. Lee, I.H. Solomon, F.J. Slack. Severe COVID-19 is associated with molecular signatures of aging in the human brain. Nat. Aging, 2 (2022), pp. 1130-1137, 10.1038/s43587-022-00321-w

46. A.M. Bernard, H.A. Roels, J.P. Buchet, R.R. Lauwerys. Serum Clara Cell Protein: An Indicator of Bronchial Cell Dysfunction Caused by Tobacco Smoking. Environ. Res., 66 (1994), pp. 96-104, 10.1006/ENRS.1994.1047

47. V.V. Veerapaneni, S. Upadhyay, T.A. Thimraj, J.B. Siddaiah, C.S. Krishnarao, K.S. Lokesh, R. Thimmulappa, L. Palmberg, K. Ganguly, M.P. Anand. Circulating secretoglobin family 1a member 1 (Scgb1a1) levels as a marker of biomass smoke induced chronic obstructive pulmonary disease. Toxics, 9 (2021), p. 208, 10.3390/toxics9090208

48. C.A. Brandsma, M. de Vries, R. Costa, R.R. Woldhuis, M. Königshoff, W. Timens. Lung ageing and COPD: is there a role for ageing in abnormal tissue repair? Eur. Respir. Rev., 26 (2017), Article 170073, 10.1183/16000617.0073-2017

49. J.J. Figueroa, W. Singer, A. Parsaik, E.E. Benarroch, J.E. Ahlskog, R.D. Fealey, J.E. Parisi, P. Sandroni, J. Mandrekar, V. Iodice, *et al*. Multiple System Atrophy: Prognostic Indicators of Survival. Mov. Disord., 29 (2014), pp. 1151-1157, 10.1002/MDS.25927

50. B.C. Orsburn. Better than flipping a coin? Organ specific plasma proteins are not confidently identified by gene expression data. Preprint at OSF (2024), 10.31219/OSF.IO/XYZQW

51. S.R. Upadhya, C.J. Ryan. Experimental reproducibility limits the correlation between mRNA and protein abundances in tumor proteomic profiles. Cell Rep. Methods, 2 (2022), Article 100288, 10.1016/J.CRMETH.2022.100288

52. M. Wang, L. Jiang, M.P. Snyder. AdaTiSS: a novel data-Adaptive robust method for identifying Tissue Specificity Scores. Bioinformatics, 37 (2021), pp. 4469-4476, 10.1093/BIOINFORMATICS/BTAB460

53. I. Yanai, H. Benjamin, M. Shmoish, V. Chalifa-Caspi, M. Shklar, R. Ophir, A. Bar-Even, S. Horn-Saban, M. Safran, E. Domany, *et al.* Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics, 21 (2005), pp. 650-659, 10.1093/BIOINFORMATICS/BTI042

54. S. Tikhonov, M. Batin, V.N. Gladyshev, S.E. Dmitriev, A. Tyshkovskiy. AgeMeta: Quantitative Gene Expression Database of Mammalian Aging. Biochemistry (Mosc.), 89 (2024), pp. 313-321, 10.1134/S000629792402010X

55. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing (2021)

56. G. Korotkevich, V. Sukhov, N. Budin, B. Shpak, M.N. Artyomov, A. Sergushichev. Fast gene set enrichment analysis. Preprint at bioRxiv (2021), 10.1101/060012

57. T.M. Therneau, P.M. Grambsch. Modeling Survival Data: Extending the Cox Model. Springer (2000), 10.1007/978-1-4757-3294-8

58. S.N. Wood. Thin Plate Regression Splines. J. R. Stat. Soc. B, 65 (2003), pp. 95-114, 10.1111/1467-9868.00374

59. D. Bates, M. Mächler, B.M. Bolker, S.C. Walker. Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw., 67 (2015), pp. 1-48, 10.18637/JSS.V067.I01

60. W.N. Venables, B.D. Ripley. Modern Applied Statistics with S. Springer (2002), 10.1007/978-0-387-21706-2

61. B.R. Mills. MetBrewer: Color Palettes Inspired by Works at the Metropolitan Museum of Art. R package, https://cran.rstudio.com/web/packages/MetBrewer/MetBrewer.pdf (2022)

62. T. Hothorn, F. Bretz, P. Westfall. Simultaneous Inference in General Parametric Models. Biom. J., 50 (2008), pp. 346-363, 10.1002/BIMJ.200810425

63. E. Merkle, D. You. nonnest2: Tests of Non-Nested Models. R package, https://cran.r-project.org/web/packages/nonnest2/nonnest2.pdf (2024)

64. T. Hielscher. nonnestcox: Comparing nonnested CoxPH models. R package, https://github.com/thomashielscher/nonnestcox/ (2024)

65. T. Hastie, R. Tibshirani, B. Narasimhan, G. Chu. impute: Imputation for microarray data. R package, https://bioconductor.org/packages/release/bioc/html/impute.html (2024)

66. A. Perez-Cornago, Z. Pollard, H. Young, M. van Uden, C. Andrews, C. Piernas, T.J. Key, A. Mulligan, M. Lentjes. Description of the updated nutrition calculation of the Oxford WebQ questionnaire and comparison with the previous version among 207,144 participants in UK Biobank. Eur. J. Nutr., 60 (2021), pp. 4019-4030, 10.1007/s00394-021-02558-4

67. C. Piernas, A. Perez-Cornago, M. Gao, H. Young, Z. Pollard, A. Mulligan, M. Lentjes, J. Carter, K. Bradbury, T.J. Key, *et al.* Describing a new food group classification system for UK biobank: analysis of food groups and sources of macro- and micronutrients in 208,200 participants. Eur. J. Nutr., 60 (2021), pp. 2879-2890, 10.1007/s00394-021-02535-x

68. Y. Wu, E.M. Byrne, Z. Zheng, K.E. Kemper, L. Yengo, A.J. Mallett, J. Yang, P.M. Visscher, N.R. Wray. Genome-wide association study of medication-use and associated disease in the UK Biobank. Nat. Commun., 10 (2019), Article 1891, 10.1038/s41467-019-09572-5

69. NCBO BioPortal. Anatomical Therapeutic Chemical Classification - Summary, version 2023AB, https://bioportal.bioontology.org/ontologies/ATC (2023)

70. K. Watanabe, T. Wilmanski, C. Diener, J.C. Earls, A. Zimmer, B. Lincoln, J.J. Hadlock, J.C. Lovejoy, S.M. Gibbons, A.T. Magis, *et al.* Multiomic signatures of body mass index identify heterogeneous health phenotypes and responses to a lifestyle intervention. Nat. Med., 29 (2023), pp. 996-1008, 10.1038/s41591-023-02248-0

71. J.P. Fine. Comparing Nonnested Cox Models. Biometrika, 89 (2002), pp. 635-647, https://www.jstor.org/stable/4140607

72. Q.H. Vuong. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. Econometrica, 57 (1989), p. 307, 10.2307/1912557