medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Proteomic aging clock (PAC) predicts age-related outcomes in middle-aged and older adults

Chia-Ling Kuo, PhD[1,2,3]*, Zhiduo Chen, MS[3], Peiran Liu, PhD[2], Luke C. Pilling, PhD[4], Janice L. Atkins, PhD[4], Richard H. Fortinsky, PhD[3], George A. Kuchel, MD[3], Breno S. Diniz, MD, PhD[1,3]

[1]Department of Public Health Sciences, University of Connecticut Health Center, Farmington CT, USA

[2]The Cato T. Laurencin Institute for Regenerative Engineering, University of Connecticut Health Center, Farmington, CT, USA

[3]UConn Center on Aging, University of Connecticut Health Center, Farmington, CT, USA

[4]Epidemiology and Public Health Group, Department of Clinical and Biomedical Sciences, University of Exeter, Exeter, UK

*Corresponding author:

Chia-Ling Kuo, PhD

The Cato T. Laurencin Institute for Regenerative Engineering, University of Connecticut

195 Farmington Ave, Suite 2080

Connecticut, CT 06032

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Summary

Background

Beyond mere prognostication, optimal biomarkers of aging provide insights into qualitative and quantitative features of biological aging and might therefore offer useful information for the testing and ultimately clinical use of gerotherapeutics. Using a large proteomic cohort in the UK Biobank (UKB), we aimed to develop a proteomic aging clock for all-cause mortality risk as a proxy of biological age (BA).

Methods

Participants in the UK Biobank Pharma Proteomics Project (UKB PPP) were included with ages between 39 and 70 years (n=53,021). Data were split into a training (70%, n=37,115) and a test set (30%, n=15,906), including 2,923 plasma proteins assessed using the Olink Explore 3072 assay®. We developed a proteomic aging clock (PAC) for all-cause mortality risk as a surrogate of BA using a combination of Least Absolute Shrinkage and Selection Operator (LASSO) penalized Cox regression and Gompertz proportional hazards models. The validation for PAC included assessing its age-adjusted associations with, and predictions for all-cause mortality and 18 incident diseases, and head-to-head comparisons with two biological age measures (PhenoAge and BioAge) and leukocyte telomere length (LTL). Additionally, a functional analysis was performed to identify gene sets and tissues enriched with genes associated with BA deviation, based on different BA measures.

Findings

The Spearman correlation between PAC proteomic age and chronological age was 0.76. 10.9% of the combined training and test samples died during a mean follow-up of 13.3 years (SD=2.2), with the mean age at death 70.1 years (SD=8.1). PAC proteomic age, after controlling

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

for age and other covariates, showed stronger associations than PhenoAge, BioAge, and LTL, with mortality and multiple incident diseases in the test set sample and in disease-free participants, such as mortality, heart failure, pneumonia, delirium, Chronic Obstructive Pulmonary Disease (COPD), and dementia. Additionally, PAC proteomic age showed higher predictive power for the conditions above compared to chronological age, PhenoAge, and BioAge, based on Harrell's C-statistics. Proteins associated with PAC proteomic age deviation (from chronological age) are enriched in various hallmarks of biological aging, including immunoinflammatory responses, cellular senescence, extracellular matrix remodeling, cellular response to stressors, and vascular biology.

Interpretation

PAC showed robust age-adjusted associations and predictions for all-cause mortality and the onset of various diseases. The diverse hallmark gene sets linked with PAC proteomic age deviation highlight the potential efficacy of geroscience-guided interventions. Further validation is essential to ascertain the use of PAC across different settings.

Funding

Access to UK Biobank data was granted under application no. 92647 "Research to Inform the Field of Precision Gerontology" (PI: Richard H. Fortinsky), funded by the Claude D. Pepper Older American Independence Centers (OAIC) program: P30AG067988 (MPIs: George A. Kuchel and Richard H. Fortinsky). CLK, BSD, RHF, and GAK are partially supported by P30AG067988. JLA has a UK National Institute for Health and Care Research (NIHR) Advanced Fellowship (NIHR301844).

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Research in context

Proteomic aging clocks have been developed using the BA surrogate of chronological age, but their validation remains limited. None had been developed using the surrogate of all-cause mortality risk, which is believed more sensitive to changes in biological aging processes. We relied on two recent comprehensive review papers[1,2] and expanded our search on PubMed and Google Scholar for English articles, using keywords including “biomarkers of aging”, “biological age predictors”, “biological age deviation”, “accelerated biological aging”, “methylation clocks”, “proteomic clocks”, “UK Biobank biomarkers”, “UK Biobank Pharma Proteomics Project”, “PhenoAge”, “BioAge”, “telomere length”, “SASP index”, and “composite aging biomarkers”.

Added value of this study

To the best of our knowledge, PAC is the first proteomic aging clock developed for all-cause mortality risk as a surrogate of BA, using the largest dataset of proteins and individuals in the world. Our results expand previous findings by showing that PAC age acceleration strongly predicts not only all-cause mortality but also several incident disease outcomes, with a follow-up exceeding a decade and a substantial sample size to ensure adequate statistical power.

Implications of all the available evidence

Due to its associations and predictive value for all-cause mortality and multiple incident diseases, PAC has the potential to serve as a valuable tool for assessing the effects of geoscience-guided interventions. It facilitates the evaluation of risk for multiple conditions in a disease-free population; thereby, contributing to the prevention of initial diseases, which vary among individuals and may subsequently lead to additional comorbidities.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Introduction

Aging-related outcomes often require years to observe, posing challenges for short-term studies. Biomarkers of aging facilitate the testing of geroscience-guided interventions. The geroscience hypothesis[3] posits that targeting aging may prevent or delay the onset of multiple diseases, where chronological age is a major risk factor.

BA acceleration, indicated by BA surpassing chronological age, has been linked with various adverse health outcomes[1,2]. Several biomarkers have been developed to fulfill such goals using various data such as biochemistry markers, omics data (transcriptomics, proteomics, methylome), and electronic health records[1,2]. Due to its prognostic value in predicting age-related conditions and sensitivity to interventions, BA measures have become increasingly popular as an intermediate phenotype in randomized controlled trials[4,5].

First-generation BA measures[6–9] were initially developed using chronological age as a surrogate of BA. These measures are now being surpassed by second-generation BA measures, which use an age-related outcome or the pace of aging to improve predictions of morbidity and mortality[10–13]. While epigenetic clocks are widely recognized, DNA methylation itself does not contain downstream biological information. In contrast, proteins provide direct links to aging-related pathology, making them more relevant for disease prognosis in the clinical context[14].

Previous studies have identified circulating proteins associated with chronological age[14] and chronic diseases[15]. A protein-based score has shown significant improvements in risk classification even after accounting for common risk factors. Several proteomic clocks have been developed for chronological age[9,16,17]. Recently, a proteomic composite, known as the senescence-associated secretory phenotype (SASP) index[18], was developed incorporating 22 pre-selected SASP proteins and indicates a phenotypic manifestation of cellular senescence, a

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

hallmark of biological aging[19]. Elevated SASP index levels have been associated with aging outcomes in older adults with major depression, such as cognitive impairment, increased medical burden–particularly in cardiovascular disease–and compromised brain health (including neuroinflammation and cortical atrophy)[20]. Consistently, other composite SASP biomarkers have shown high predictive value for all-cause mortality in healthy older adults[21].

Despite their relevance, these proteomic clocks exhibit significant limitations. They were mostly trained to predict chronological age (i.e., first-generation clocks) and trained in a small sample with a small set of proteins, which may be pre-selected to reflect a specific biological aging process (e.g., cellular senescence). In this study, we aimed to develop a proteomic clock, referred to as the proteomic aging clock (PAC) to predict all-cause mortality. Data included 2,923 plasma proteins, assessed using the Olink Explore 3072 assay®, from a UK Biobank (UKB) cohort of 53,021 participants aged between 39 and 70 years. PAC underwent validation for both its associations and predictions of all-cause mortality and age-related conditions, during a follow-up exceeding a decade. Biological insights into PAC proteomic age acceleration were revealed through gene set analysis and gene property analysis for tissue specificity.

Methods

UK Biobank Pharma Proteomics Project

The UK Biobank (UKB) recruited more than 500,000 participants with ages between 40 and 70 years during 2006 and 2010[22]. At recruitment (baseline), participants completed online questionnaires, cognitive function tests, verbal interviews, and physical measurements. Additionally, blood samples were collected for future biological assays. Disease diagnoses and death status since then have been updated through linkages to electronic health records.

Participants who supplied blood samples at baseline were selected for inclusion in the UK Biobank Pharma Proteomics Project (UKB-PPP)[23]. Of the included samples (n=53,021), the

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

majority (n=46,792, 88.3%) were a random sample from the UKB baseline cohort. Others (n=6,229, 11.7%) included participants who attended the first imaging visit and COVID-19 repeat imaging study and/or those selected by the consortium of 13 biopharmaceutical companies for research interests.

Data

All-cause mortality risk was used as an indicator for BA. Death data were provided by the UK National Health Service (NHS) England, NHS Central Register, and National Records of Scotland. Participants with no recorded date of death were assumed to remain alive until the censoring date 11/30/2022.

The normalized proteomic expression (NPX) data encompassed 2,923 proteins (**Supplementary Table 1**). Three proteins with high rates of missing data were removed from the analysis: GLIPR1 (99.7%), NPM1 (74.0%), and PCOLCE (63.6%). For the remaining proteins, we applied a *k*-nearest neighbors approach to impute missing proteomic data (*k*=10). Participants with complete NPX, chronological age, and mortality data were included in the PAC development.

Data splitting

The included samples (n=53,021) were split into a training set (70%, n=37,115) and a test set (30%, n=15,906). In the training set, 4,034 participants (10.9%) died at the mean age of 70.1 years (SD=8.1) over a mean follow-up of 13.3 years (SD=2.2). Within the test set, 1,731 participants (10.9%) died, with the mean age at death 70.1 years (SD=8.1) during a mean follow-up of 13.3 years (SD=2.2).

A baseline summary for participants in the training and test sets versus others in the UKB baseline cohort is presented in **Supplementary Table 2**. Data were extracted using the field IDs in **Supplementary Table 3**. The training and test samples showed comparable baseline

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

characteristics to the rest of the UKB baseline cohort. The disease prevalence was slightly higher within the UKB PPP than the rest of the baseline cohort, which is expected, due to the enrichment of diseases in the UKB-PPP samples[23].

PAC development

The NPX data and chronological age at baseline in the training set were used to train a Least Absolute Shrinkage and Selection Operator (LASSO) penalized Cox regression model for the risk of all-cause mortality. The selected proteomic signatures and chronological age were used to fit Gompertz proportional hazards models and formulate PAC to estimate the proteomic age based on the input data (**Supplementary Methods**).

PAC validation

We evaluated PAC proteomic age for correlations with chronological age, BioAge, PhenoAge, short LTL, physiological or cognitive measures (**Supplementary Table 3**), a 49-item frailty[25], and disease-associated biomarkers (**Supplementary Table 3**) – all measured at baseline – using the test set data. Next, we tested if PAC proteomic age deviation was linked with mortality and incident diseases (hypertension, myocardial infarction, heart failure, stroke, type 2 diabetes, COPD, pneumonia, chronic kidney disease, dementia, delirium, Parkinson's disease, any cancer excluding non-melanoma skin cancer, and common cancers including breast cancer [females only], prostate cancer [males only], lung cancer, and colorectal cancer).

Participants free of the disease at baseline (2006-2010) were followed up until the first disease diagnosis, death (censoring date 11/30/2022), or the last follow-up date (censoring dates: 11/30/2022 [England], 7/31/2021 [Scotland], 2/28/2018 [Wales]) depending on which occurred first. First diagnosis dates were identified using the UKB cancer registries data and first occurrence data, which linked data from different sources based on 3-character ICD-10 codes: longitudinal primary care (45% of the UK Biobank cohort), hospital inpatient, death registry

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

data, and self-reported medical conditions at baseline (**Supplementary Table S3**). Data on the covariates were collected by UKB through online surveys, physical measurements, and linkages to electronic health records (**Supplementary Table 3**).

Associations of PAC proteomic age deviation with all-cause mortality and incident diseases

Using the test set data, Cox regression models for all-cause mortality and Fine-Gray subdistribution hazard models to account for the effect of death on the risk of incident diseases were fitted adjusting for different sets of covariates at baseline: 1) age-adjusted: age only, 2) partially adjusted: sociodemographic factors (age, self-reported sex, ethnicity [White, Black, Asian, Other], and education [from none to college or university degree], Townsend deprivation index [higher values indicating higher levels of material deprivation]) and lifestyle factors (smoking status [current, former, never], body mass index [BMI]), and 3) fully adjusted models: covariates in the partially adjusted model and pre-existing diseases (hypertension, myocardial infarction, heart failure, stroke, type 2 diabetes, COPD, pneumonia, chronic kidney disease, dementia, delirium, Parkinson's disease, any cancer excluding non-melanoma skin cancer).

The p-values from the age-adjusted, partially adjusted, and fully adjusted models for all-cause mortality and incident diseases were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate (FDR) method.

PAC versus other BA measures in associations of biological age deviation with all-cause mortality and incident diseases

Using the test set data, the associations of PAC proteomic age with all-cause mortality and incident diseases were compared with those of other BA measures, namely BioAge, PhenoAge, and leukocyte telomere length (LTL), adjusting for the full set of covariates. BioAge[8] was trained for chronological age, while PhenoAge[10] was trained for all-cause mortality, both using routine clinical biomarkers from blood samples in the National Health and Nutrition Survey (NHANES) III (detailed in **Supplementary Method**s). Further validation of both

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

measures was performed in additional cohorts, including UKB[26,27], confirming their robustness. LTL was assessed using a multiplex qPCR technique as T/S ratio, which compares the telomere amplification product (T) to that of a single-copy gene (S), adjusting for technical parameters[28]. The rank-based inverse normal transformation was applied to each BA measure to convert the data to z-scores to standardize the scales of different BA measures. Short LTL by reversing the signs of LTL was compared with other BA measures so the expected association direction was consistent across measures.

For sensitivity analysis, the associations above were investigated in participants without any pre-existing diseases at baseline. Due to a significant number of participants having one or more pre-existing diseases at baseline and disease-free participants having lower incidence of chronic diseases, we included both training and test samples to maintain necessary power. The primary fully adjusted models were simplified to the partially adjusted models as none of the included participants had developed any of the diseases.

PAC versus other BA measures in predictions for all-cause mortality and incident diseases

Harrell's C-statistic, a concordance probability within the range from 0.5 to 1, compares individuals in a pair that the individual who has a shorter time to a disease also has higher risk for the disease based on the model. Harrell's C-statistic serves as a standard output to quantify discriminative power for Cox regression models, yet it demands an extended computation time for Fine-Gray subdistribution hazard models. Although we used Fine-Gray subdistribution hazard models to link BA deviation with incident diseases to account for the competing event of death, corresponding Cox regression models, which censored individuals who died before disease diagnosis yielded similar associations (results not shown). Without losing the generalization of our findings, we opted for Cox regression models to assess the predictions of

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

PAC proteomic age against other BA estimates for all-cause mortality and incident diseases using the test set data.

Functional analysis

To unravel the biological processes underlying BA deviation, proteins after the inverse normal transformation were associated with PAC proteomic age, BioAge, PhenoAge or short LTL in the fully adjusted linear regression models using the test set data. Significant proteins (Bonferron-corrected $p<0.05$) were carried forward to perform a gene set analysis and a gene property analysis for tissue specificity using the Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA version 1.6.0)[29]. Similar analyses were conducted for BioAge, PhenoAge, and short LTL.

In the gene set analysis, genes associated with BA deviation were compared with the background genes (20,260 protein-coding genes) for the presence in a hallmark gene set using a hypergeometric test. Enriched hallmark gene sets with at least five genes overlapped with the input genes were identified at the Bonferroni-corrected level of 5% (50 hallmark gene sets in total).

In the gene property analysis for tissue specificity, the input genes were compared with the background genes (protein-coding genes with mean normalized $\log_2$ expression value > 1 in at least one of 30 general tissues) using a hypergeometric test for the presence in a tissue-specific differentially expressed gene set (genes with p-value ≤ 0.05 after Bonferroni correction and absolute log fold change ≥ 0.58 in GTEx v8). Bonferroni-corrected p-values smaller than 5% were considered statistically significant.

Results

Development of the proteomic aging clock (PAC)

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Using the training set data, a Least Absolute Shrinkage and Selection Operator (LASSO) penalized Cox regression model was applied to 2,920 proteins and chronological age at baseline to predict the time-to-event outcome of death. Chronological age and 128 proteins remained (lambda 0.004543) and were carried forward to fit a Gompertz model (**Supplementary Table 4**). Additionally, another Gompertz model was fitted to predict death solely using chronological age. We calculated the PAC proteomic age based on the shape and rate parameters, and the regression coefficients associated with individual proteins from the models above (**Supplementary Table 5**). The mean PAC proteomic age was 53.4 years, 3.4 years younger than the mean chronological age in the training set. In the test set sample, the mean PAC proteomic age and chronological age were 53.4 and 56.9 years, respectively.

Correlations between chronological age, PAC proteomic age, PhenoAge, BioAge, LTL, plus a selection of aging phenotypes at baseline

10,451 participants had complete data for chronological age, PAC proteomic age, BioAge, PhenoAge, and LTL in the test set. The Spearman correlation (r) between PAC proteomic age and chronological age was 0.77, lower than the correlations of BioAge (r=0.98) and PhenoAge (r=0.87) with chronological age (**Supplementary Figure 1**). Consistently, short LTL demonstrated weak correlations with chronological age and other BA measures (r≈0.2) (**Supplementary Figure 1**). Single physiological or cognitive measures, frailty, and disease-associated biomarkers showed a weak to moderate association with chronological age, similarly with PAC proteomic age, BioAge, and PhenoAge (**Supplementary Figures 1 and 2**).

Associations of PAC proteomic age acceleration with all-cause mortality and incident diseases

The PAC proteomic age acceleration showed significant associations with all-cause mortality and various incident diseases in the test set sample adjusting for chronological age only (age adjusted model). These associations were attenuated in the partially adjusted and fully

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

adjusted models, though remaining statistically significant (**Fig. 1; Supplementary Table 6**). For instance, the HR for all-cause mortality was 1.097 per year increase in PAC proteomic age (95% CI 1.091 to 1.103, $p_{adj}=3.83\times10^{-232}$) in the fully adjusted model, versus 1.104 in the age-adjusted model, and 1.102 in the partially adjusted model. For sensitivity analysis, we included an indicator for participants selected by the UKB-PPP consortium in the fully adjusted models. The results were very similar to the fully adjusted model results, likely attributable to the overlap in effects between the selection and baseline disease states.

PAC versus other BA measures in associations with all-cause mortality and incident diseases

PAC outperformed PhenoAge, BioAge, and short LTL for most outcomes, showing the strongest associations with all-cause mortality, heart failure, pneumonia, delirium, COPD, dementia, lung cancer, myocardial infarction, osteoporosis, Parkinson's disease, any cancer, and colorectal cancer (**Fig. 2)**. In contrast, the associations with type 2 diabetes and chronic kidney disease were strongest with PhenoAge. BioAge showed the strongest associations with stroke and hypertension (**Fig. 2**). Similar associations were observed in the age-adjusted and partially adjusted models (**Supplementary Figures 3 and 4)**. Sensitivity analyses, including only disease-free participants at baseline (n=18,882), showed similar associations of BA acceleration with all-cause mortality and incident diseases, highlighting the robustness of our findings (**Fig. 3**). Interestingly, the associations of PAC proteomic age acceleration with lung cancer and dementia were stronger in the disease-free participants than in the test set sample.

PAC versus other BA measures in predictions for all-cause mortality and incident diseases

Using the test set data only, we compared the C-statistics for all-cause mortality and incident diseases of four models: 1) chronological age only (**M-Age**), 2) PAC proteomic age only (**M-PAC**), 3) BioAge only (**M-BioAge**), and 4) PhenoAge only (**M-PhenoAge**). M-PAC

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

outperformed other models based on C-statistics, particularly all-cause mortality, COPD, pneumonia, and heart failure (**Fig. 4**). The M-PAC C-statistics for dementia and delirium were the highest across diseases and models but not significantly different from those of M-Age (**Fig. 4**). Models with multiple BA measures showed minimally improved C statistics (**Supplementary Fig. 5**).

Functional analysis

1,001 significant proteins coded by 1,008 genes were selected for significant associations with PAC proteomic age deviation to initiate a functional analysis by FUMA (**Supplementary Table 7**). Genes associated with PAC proteomic age deviation were enriched in 25 hallmark gene sets (Bonferroni-corrected $p<0.05$) (**Fig. 5**, **Supplementary Table 11**).

The hallmark gene sets associated with PAC proteomic age deviation include a wide range of biological processes and signaling pathways, particularly epithelial-mesenchymal transition, coagulation, inflammatory response, allograft rejection, IL-6-JAK-STAT3 signaling, complement, and IL2-STAT5 signaling (**Fig. 5**). Additionally, genes associated with PAC proteomic age deviation were overrepresented in the differentially expressed genes in multiple tissues, topped by lung and adipose tissues (**Supplementary Figure 6**). The findings above were consistent across BA measures (**Supplementary Figures 6 and 7**), suggesting the presence of conserved biology underlying the aging processes.

Discussion

We developed a proteomic clock (PAC) for all-cause mortality as a surrogate of BA. PAC proteomic age acceleration was robustly associated with all-cause mortality and age-related diseases after controlling for sociodemographic, lifestyle factors, and pre-existing diseases, in the test set sample and disease-free participants. PAC proteomic age showed, in general, better performance in predicting all-cause mortality and incident diseases than chronological age and

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

other BA measures. Proteins associated with PAC proteomic age deviation were enriched in various hallmarks of biological aging, including immunoinflammatory responses, cellular senescence, extracellular matrix remodeling, cellular response to stressors, and vascular biology. These findings demonstrate the robustness of this proteomic aging clock to predict different adverse health outcomes and reflect the current understanding of the perturbations in multiple biological pathways in the aging process.

Previous studies evaluated the proteomic correlates of chronological aging using different proteomic platforms and assays (e.g., SOMAscan assays® or mass spectrometry)[9,16,17]. Proteins associated with chronological age showed significant associations with age-related clinical outcomes: walking speed, grip strength, frailty, multimorbidity, and all-cause mortality. A recent preprint[30] used 1,459 proteins from the initial release of UKB PPP (n=31,581) and data from China Kadoorie Biobank (n=1,418) to train a proteomic clock (ProtAge) to predict chronological age. The trained ProtAge was associated with all-cause mortality and several diseases, e.g., Alzheimer's disease. However, there are marked differences between the two studies. First, we used a larger set of proteins to train a proteomic clock (2,920 vs. 1,459 proteins), providing a broader coverage of the human proteome. Second, we trained a proteomic clock to predict mortality instead of chronological age –a shift from a "first-generation" to a "second-generation" clock. Lastly, we went beyond the ProtAge analyses and also reported C-statistics to show the predictive power of PAC versus other BA measures. Although the PAC and ProtAge are not directly comparable, PAC consistently showed high predictive power for multiple aging outcomes.

PAC proteomic age acceleration showed the strongest associations with mortality risk and several diseases outcomes (e.g., heart failure, pneumonia, delirium, COPD, and dementia).

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

On the other hand, BioAge showed the strongest association with hypertension and stroke, whereas PhenoAge showed the strongest associations with type 2 diabetes and chronic kidney disease. This pattern of associations remained similar among individuals who had no medical comorbidity at baseline, except the associations of PAC proteomic age acceleration with lung cancer and dementia became stronger. Different BA measures may be implemented depending on the study context or the outcomes of interest. Notably, PAC is particularly valuable in identifying high-risk individuals years before the earliest manifestations of chronic conditions.

PAC proteomic age acceleration showed the strongest associations with dementia and delirium compared to other BA measures, although the gains in predictive power for these conditions were significantly diminished. A parallel pattern was observed for BioAge with hypertension and stroke, as well as for PhenoAge with type 2 diabetes and chronic kidney disease. Incorporating proteins into PAC that are associated with the variation in mortality risk after controlling for covariates, may improve disease prediction. Further investigation is needed to test this hypothesis rigorously.

We found that genes associated with proteomic age deviation are enriched in various hallmarks of biological aging, including immunoinflammatory responses, cellular senescence, extracellular matrix remodeling, cellular response to stressors, and vascular biology. Additionally, several hallmark gene sets are conserved across BA measures, including inflammatory response, allograft rejection, IL-6-JAK-STAT3 signaling, IL2-STAT5 signaling, TNG alpha signaling via NF-κB, and apoptosis. Our findings suggest that regardless of the BA measures used, our findings indicate consistent manifestations of biological processes and pathways in BA acceleration. The multifaced biological influence on aging phenotypes

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

reinforces the potential for geroscience-guided interventions to target multiple age-related outcomes.

This study has limitations that need to be considered when interpreting our findings. First, we did not exclude deaths unrelated to biological aging, such as those resulting from accidents. However, such incidents are rare in the UKB cohort and unlikely to impact our findings significantly. Second, we could not compare PAC with commonly used epigenetic clocks since UKB does not have data on DNA methylation. However, PhenoAge was used to train DNAm PhenoAge, thus providing an indirect comparison between PAC and DNAm PhenoAge clock. Third, although we used independent samples for PAC development and its validation, the test set shares substantial homogeneity with the training set. Although our findings are robust to disease states, additional validation is needed in other populations and using proteomic data from alternative platforms.

In conclusion, we have developed a novel proteomic aging clock termed PAC, which demonstrated robust associations and predictions for mortality and the onset of various diseases. The diverse hallmark gene sets linked with PAC proteomic age deviation highlight the potential efficacy of geroscience-guided interventions. Further validation is essential to ascertain the use of PAC across different settings.

Contributors

CLK and BSD designed the study. CLK, LCP, and JLA processed the data while CLK conducted the statistical analyses. The initial manuscript was drafted by CLK and BSD, with contributions from ZC and PL. All the authors reviewed and approved the final version.

Declaration of interests

We have no conflicting interests to disclose.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

Data sharing

Data access is granted upon application to the UK Biobank. The R code for computing PAC proteomic age can be obtained from the GitHub repository at https://github.com/kuo-lab-uchc/PAC.

Acknowledgments

Access to UK Biobank data was granted under application no. 92647 “Research to Inform the Field of Precision Gerontology” (PI: Richard H. Fortinsky). This research used data assets made available by National Safe Haven as part of the Data and Connectivity National Core Study, led by Health Data Research UK in partnership with the Office for National Statistics and funded by UK Research and Innovation (research which commenced between 1 October 2020–31 March 2021 grant ref MC_PC_20029; 1 April 2021–30 September 2022 grant ref MC_PC_20058). This research also used data provided by patients and collected by the NHS as part of their care and support. Copyright © (year), NHS England. Re-used with the permission of the NHS England [and/or UK Biobank]. All rights reserved.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

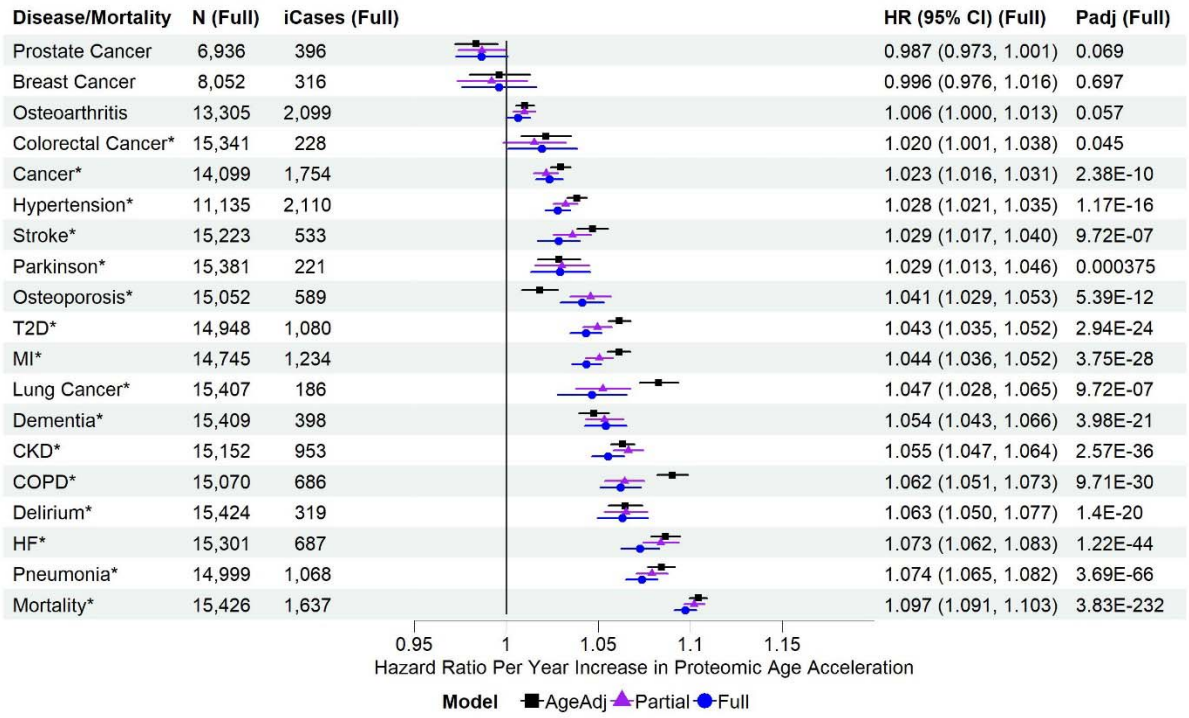

Figure 1. Associations of PAC proteomic age acceleration with all-cause mortality and incident diseases in the test set sample. MI: myocardial infarction; **T2D**: type 2 diabetes; **COPD**: chronic obstructive pulmonary disease; **CKD**: chronic kidney disease; **HF**: heart failure. **N (Full)**: sample size with complete data for the fully adjusted model, after excluding participants diagnosed with the disease at or prior to baseline. **iCases (Full)**: number of incident cases of N samples. Cox regression model for all-cause mortality and Fine-Gray sub-distribution hazard models to account for the effect of death on the risk for incident diseases, adjusting for different sets of covariates at baseline (age adjusted, partially adjusted, and fully adjusted models). **AgeAdj**: chronological age; **Partial**: chronological age, sex, ethnicity, education, Townsend deprivation index, smoking status, and body mass index; **Full**: covariates in the partially adjusted model, and pre-existing diseases (hypertension, myocardial infarction, heart failure, stroke, type 2 diabetes, COPD, pneumonia, chronic kidney disease, any cancer excluding non-melanoma skin cancer, dementia, and Parkinson's disease, without delirium as there were only 2 samples with a history of delirium at baseline in the test set sample). **P_{adj} (Full)**: p-values adjusted for multiple testing (tests based on age adjusted, partially adjusted, and fully adjusted models for all-cause mortality and incident diseases). Disease/all-cause mortality highlighted with asterisk (*) if $p_{adj} < 0.05$.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

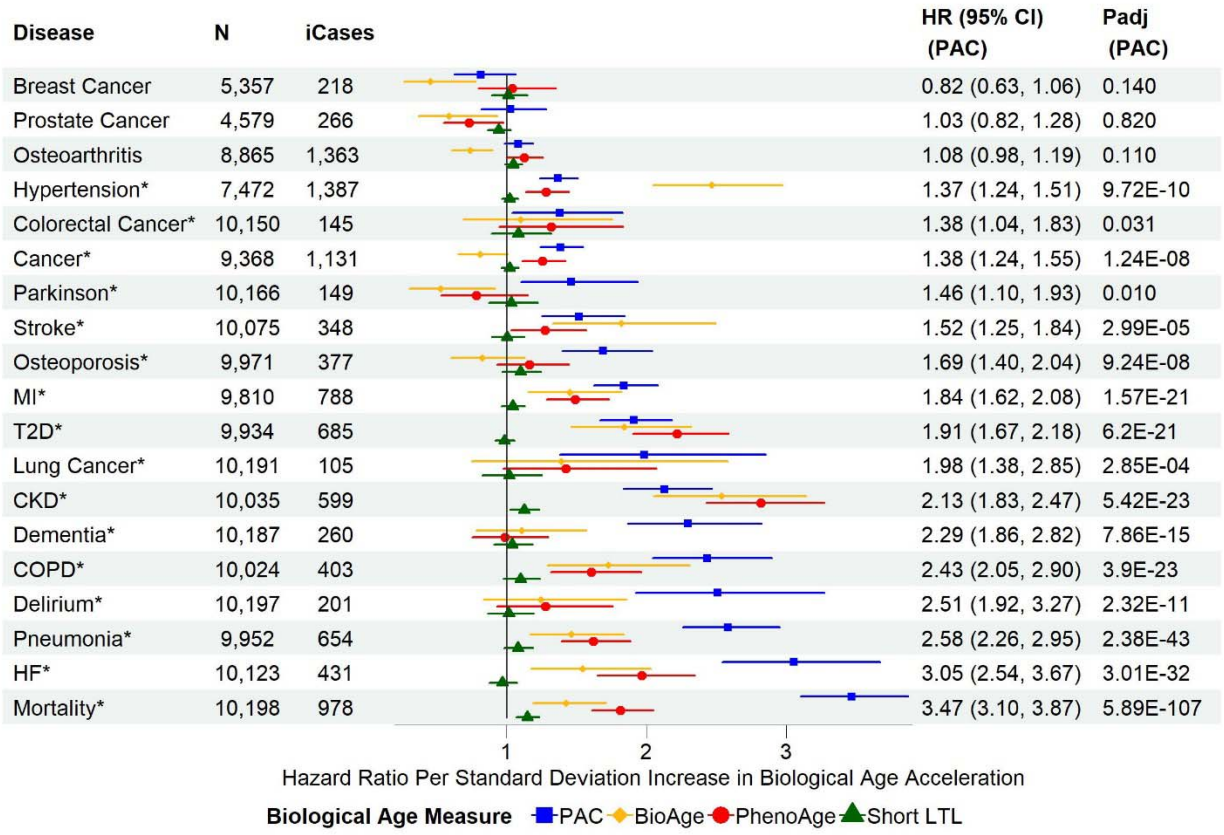

Figure 2. Associations of biological age acceleration based on different biological age measures with all-cause mortality and incident diseases in the test set sample using the fully adjusted models. MI: myocardial infarction; **T2D**: type 2 diabetes; **COPD**: chronic obstructive pulmonary disease; **CKD**: chronic kidney disease; **HF**: heart failure. **N**: sample size with complete data for the fully adjusted models of PAC, BioAge, PhenoAge, and short LTL, after excluding participants diagnosed with the disease at or prior to baseline. **iCases**: number of incident cases of N samples. Cox regression model for all-cause mortality and Fine-Gray subdistribution hazard models to account for the effect of death for the risk of incident diseases. The full covariate adjustment included chronological age, sex, ethnicity, education, Townsend deprivation index, smoking status, body mass index, and pre-existing diseases (hypertension, myocardial infarction, heart failure, stroke, type 2 diabetes, COPD, pneumonia, chronic kidney disease, any cancer excluding non-melanoma skin cancer, dementia, and Parkinson's disease). Delirium was not included as there were only two samples with a history of delirium at baseline. $\mathbf{P_{adj}}$: p-values adjusted for multiple testing per BA measure (tests based on age-adjusted, partially adjusted, and fully adjusted models for all-cause mortality and incident diseases)

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

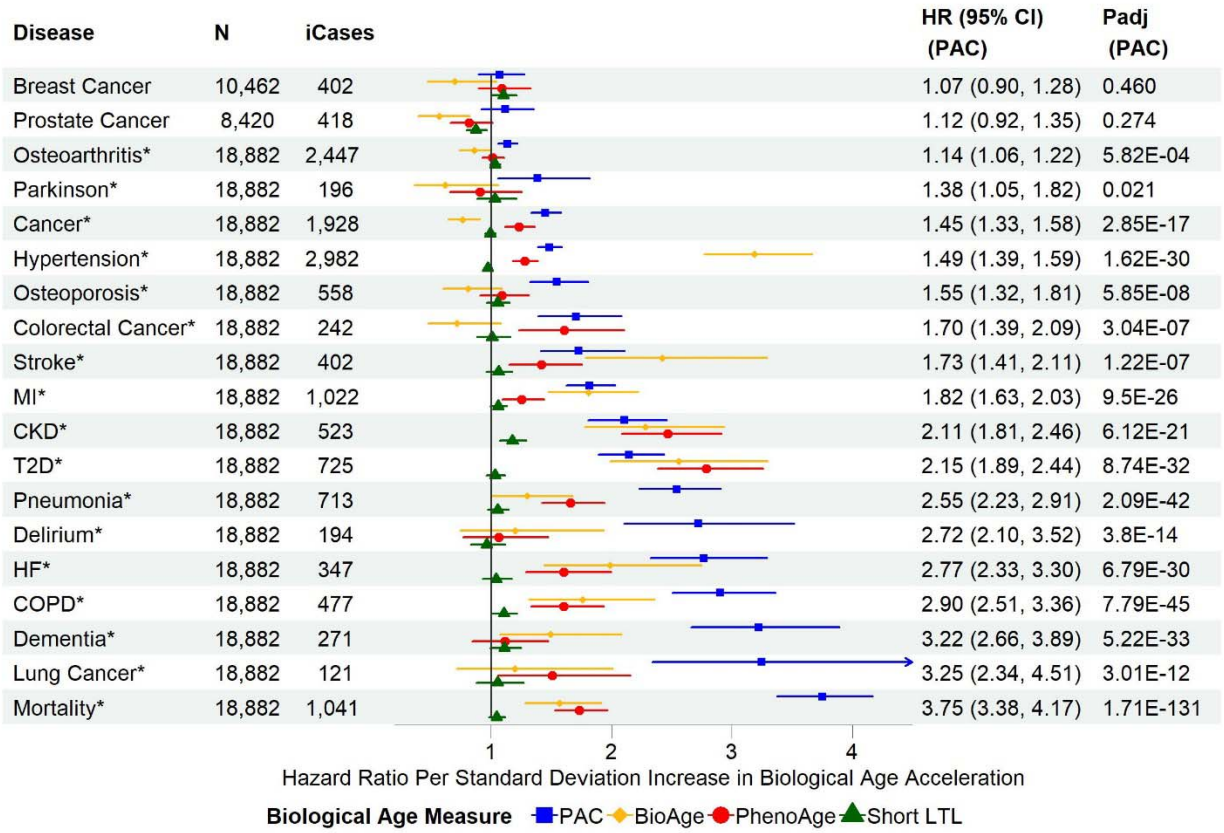

Figure 3. Associations of biological age acceleration based on different biological age measures with all-cause mortality and incident diseases using disease-free participants in the combined training and test set, and fully adjusted models. MI: myocardial infarction; **T2D**: type 2 diabetes; **COPD**: chronic obstructive pulmonary disease; **CKD**: chronic kidney disease; HF: heart failure. **N**: sample size with complete data for partially adjusted models of PAC, BioAge, PhenoAge, and short LTL, after excluding participants diagnosed with any of the diseases at or prior to baseline. **iCases**: number of incident cases of N samples. Cox regression model for all-cause mortality and Fine-Gray subdistribution hazard models to account for the effect of death on the risk for incident diseases. The partial covariate adjustment included chronological age, sex, ethnicity, education, Townsend deprivation index, smoking status, and body mass index. **Padj**: p-values adjusted for multiple testing per biological age measure (n=19)

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

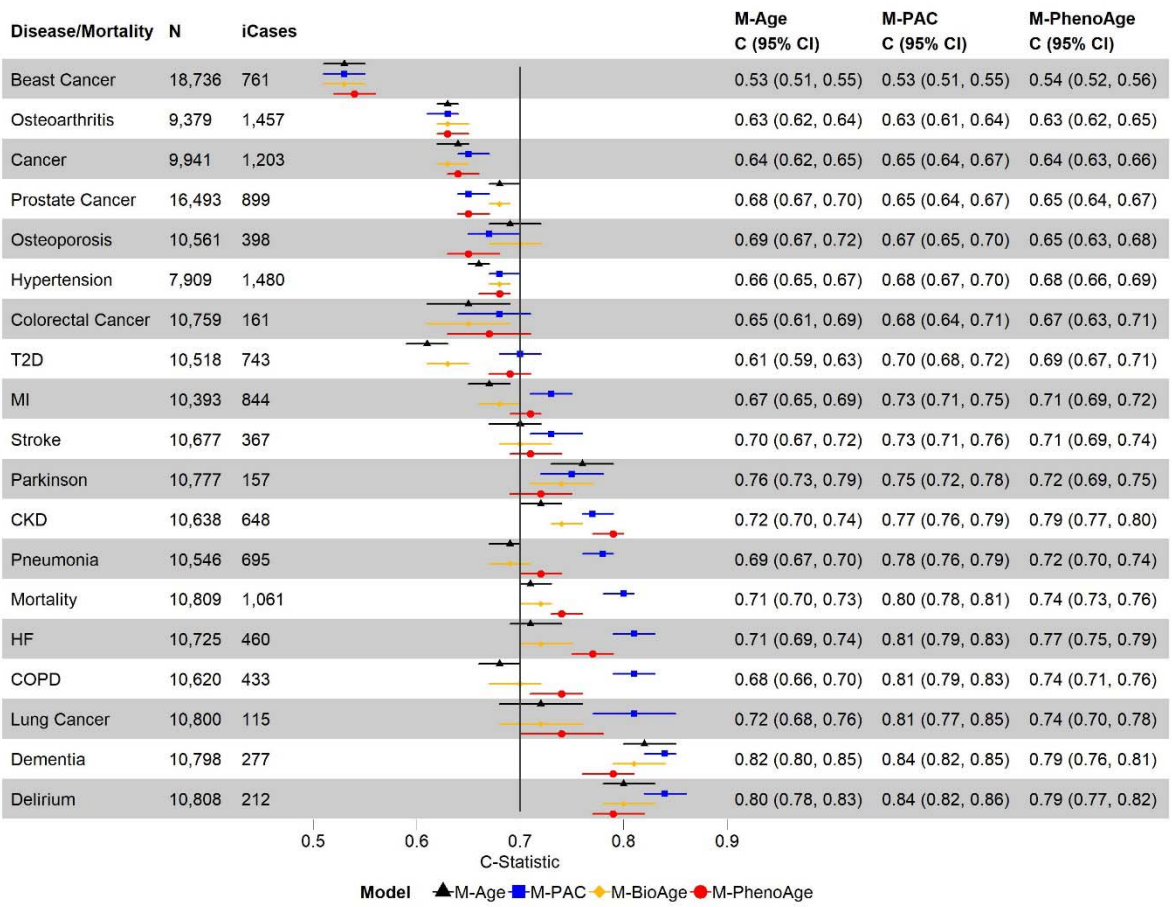

Figure 4. Discriminative power of biological age based on different measures for all-cause mortality and incident diseases using Cox models and the test set data: 1) model with chronological age only (M-Age), 2) model with PAC proteomic age only (M-PAC), 3) model with BioAge only (M-BioAge), and 4) model with PhenoAge only (M-PhenoAge). **MI**: myocardial infarction; **T2D**: type 2 diabetes; **COPD**: chronic obstructive pulmonary disease; **CKD**: chronic kidney disease; **HF**: heart failure. **N**: sample size with complete data for chronological age, PAC proteomic age, BioAge, and PhenoAge, after excluding participants diagnosed with the disease at or prior to baseline. **iCases**: number of incident cases of N samples.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

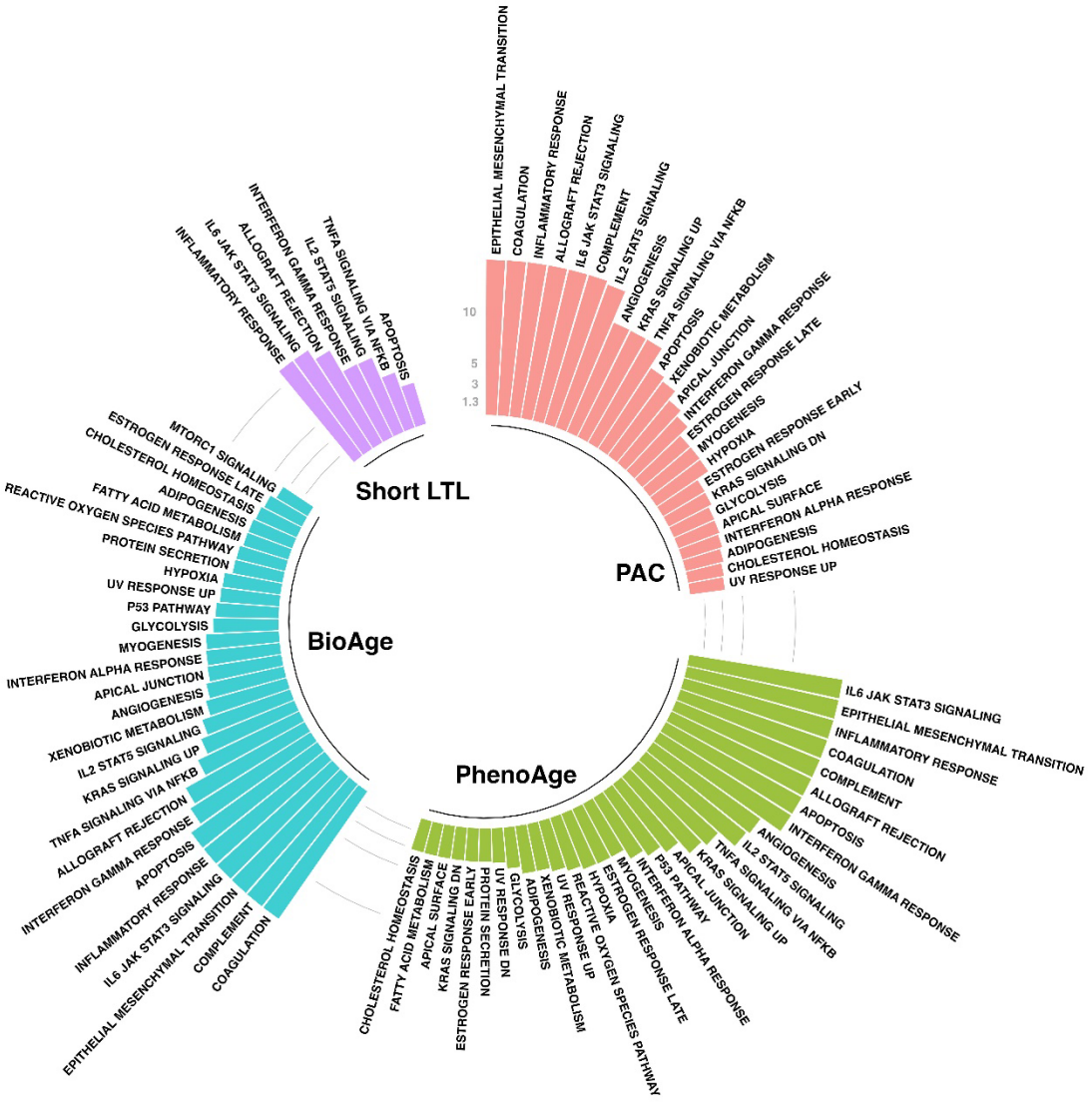

Figure 5. Hallmark gene sets enriched with genes associated with BA deviation based on different BA measures. The colored bars represent -log10(p) for different biological age measures after Bonferroni correction (n=50) and those greater than 15 are truncated at 15.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

References

1. Rutledge, J., Oh, H. & Wyss-Coray, T. Measuring biological age using omics data. *Nat Rev Genet* **23**, 715–727 (2022).
2. Moqri, M. *et al.* Biomarkers of aging for the identification and evaluation of longevity interventions. *Cell* **186**, 3758–3775 (2023).
3. Kennedy, B. K. *et al.* Geroscience: linking aging to chronic disease. *Cell* **159**, 709–713 (2014).
4. Waziry, R. *et al.* Effect of long-term caloric restriction on DNA methylation measures of biological aging in healthy adults from the CALERIE trial. *Nat Aging* **3**, 248–257 (2023).
5. Lohman, T. *et al.* HIGH INTENSITY interval training reduces transcriptomic age: A randomized controlled trial. *Aging Cell* **22**, e13841 (2023).
6. Hannum, G. *et al.* Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Molecular Cell* **49**, 359–367 (2013).
7. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biology* **14**, R115 (2013).
8. Levine, M. E. Modeling the rate of senescence: can estimated biological age predict mortality more accurately than chronological age? *J Gerontol A Biol Sci Med Sci* **68**, 667–674 (2013).
9. Tanaka, T. *et al.* Plasma proteomic signature of age in healthy humans. *Aging Cell* **17**, e12799 (2018).
10. Levine, M. E. *et al.* An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)* **10**, 573–591 (2018).
11. Lu, A. T. *et al.* DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* **11**, 303–327 (2019).

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

12. McGreevy, K. M. *et al.* DNAmFitAge: biological age indicator incorporating physical fitness. *Aging* **15**, 3904–3938 (2023).

13. Belsky, D. W. *et al.* DunedinPACE, a DNA methylation biomarker of the pace of aging. *eLife* **11**, e73420 (2022).

14. Moaddel, R. *et al.* Proteomics in aging research: A roadmap to clinical, translational research. *Aging Cell* **20**, e13325 (2021).

15. Gadd, D. A. *et al. Blood protein levels predict leading incident diseases and mortality in UK Biobank*. http://medrxiv.org/lookup/doi/10.1101/2023.05.01.23288879 (2023) doi:10.1101/2023.05.01.23288879.

16. Sayed, N. *et al.* An inflammatory aging clock (iAge) based on deep learning tracks multimorbidity, immunosenescence, frailty and cardiovascular aging. *Nat Aging* **1**, 598–615 (2021).

17. Sathyan, S. *et al.* Plasma proteomic profile of age, health span, and all-cause mortality in older adults. *Aging Cell* **19**, e13250 (2020).

18. Diniz, B. S. *et al.* Enhanced Molecular Aging in Late-Life Depression: the Senescent-Associated Secretory Phenotype. *Am J Geriatr Psychiatry* **25**, 64–72 (2017).

19. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).

20. Seitz-Holland, J. *et al.* Major depression, physical health and molecular senescence markers abnormalities. *Nat. Mental Health* **1**, 200–209 (2023).

21. St Sauver, J. L. *et al.* Biomarkers of cellular senescence and risk of death in humans. *Aging Cell* e14006 (2023) doi:10.1111/acel.14006.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.19.23300228; this version posted December 20, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC 4.0 International license .

22. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).

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

24. UK Biobank Showcase. *UK Biobank Showcase* https://biobank.ndph.ox.ac.uk/showcase/.

25. Williams, D. M., Jylhävä, J., Pedersen, N. L. & Hägg, S. A Frailty Index for UK Biobank Participants. *J Gerontol A Biol Sci Med Sci* **74**, 582–587 (2019).

26. Liu, Z. *et al.* A new aging measure captures morbidity and mortality risk across diverse subpopulations from NHANES IV: A cohort study. *PLoS Med* **15**, e1002718 (2018).

27. Kuo, C.-L., Pilling, L. C., Liu, Z., Atkins, J. L. & Levine, M. E. Genetic associations for two biological age measures point to distinct aging phenotypes. *Aging Cell* **20**, e13376 (2021).

28. Codd, V. *et al.* Measurement and initial characterization of leukocyte telomere length in 474,074 participants in UK Biobank. *Nat Aging* **2**, 170–179 (2022).

29. Watanabe, K., Taskesen, E., Van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).

30. Argentieri, M. A. *et al. Proteomic aging clock predicts mortality and risk of common age-related diseases in diverse populations*. http://medrxiv.org/lookup/doi/10.1101/2023.09.13.23295486 (2023) doi:10.1101/2023.09.13.23295486.