A proteomic signature of healthspan

Chia-Ling Kuo, PhD[1,2,3†*], Peiran Liu[†], PhD[2], Zhiduo Chen, MS[3], 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[1,3,5]

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

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

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

[4]Department of Clinical and Biomedical Sciences, University of Exeter, Exeter, UK

[5]Department of Psychiatry, University of Connecticut Health Center, Farmington Connecticut, USA

[†]Co-first authors

[*]Corresponding:

Chia-Ling Kuo

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

195 Farmington Ave, Suite 2080

Connecticut, CT 06032

Abstract

The focus of aging research has shifted from increasing lifespan to enhancing healthspan to reduce the time spent living with disability. Despite significant efforts to develop biomarkers of aging, few studies have focused on biomarkers of healthspan. We developed a proteomics-based signature of healthspan (healthspan proteomic score (HPS)) using data from the UK Biobank Pharma Proteomics Project (53,018 individuals and 2920 proteins). A lower HPS was associated with higher mortality risk and several age-related conditions, such as COPD, diabetes, heart failure, cancer, myocardial infarction, dementia, and stroke. HPS showed superior predictive accuracy for these outcomes compared to chronological age and biological age measures. Proteins associated with HPS were enriched in hallmark pathways such as immune response, inflammation, cellular signaling, and metabolic regulation. Our findings demonstrate the validity of HPS, making it a valuable tool for assessing healthspan and as a potential surrogate marker in geroscience-guided studies.

Main

Over recent decades, global life expectancy has significantly increased, especially in developing countries. However, the corresponding healthy life expectancy, defined as years lived in full health since birth, has failed to keep up the pace[1]. The gap between healthspan and lifespan highlights a substantial challenge posed by chronic diseases among older adults[2], emphasizing the need for novel approaches to narrow the gap and prolong disease-free longevity[3]. The geroscience approach is a potential solution, addressing the hypothesis that targeting the hallmarks of aging, rather than individual diseases of aging, may prevent or delay the onset of multiple age-related conditions[4].

Unlike lifespan, which has a universal definition, there is no consensus on the definition of healthspan[5,6]. Previous research has suggested characterizing healthy aging in five domains: physical capability, cognitive function, physiological and musculoskeletal, endocrine, and immune functions[5,7]. For practical purposes, healthspan typically refers to the period of life spent in good health, free from the chronic diseases and disabilities of aging[6]. Studies aiming to evaluate the effects of interventions on healthspan are challenging due to the need for long follow-up lengths and large sample sizes of healthy individuals to observe the outcomes of interest. Thus, developing surrogate biomarkers that can predict healthspan is crucial for improving the feasibility of clinical trials to test interventions to prolong healthspan and lifespan.

Composite biomarkers incorporating multiple measures are more robust in predicting age-related outcomes than single biomarkers[8–10]. Several composite biomarkers for predicting lifespan or mortality have been developed using clinical biomarkers[11,12] or omics data[11,13–15]. However, to date, no composite biomarker measures have been developed based on a healthspan definition. To mitigate this gap, we developed a proteomics-based healthspan biomarker

(healthspan proteomic score, HPS) using chronological age and expression data of 2,920 proteins at the UK Biobank (UKB) baseline/recruitment (2006-10). In line with previous studies[16,17], we defined healthspan as the number of years from birth without a major chronic medical condition, including cancer (excluding non-melanoma skin cancer), diabetes (type I diabetes, type II diabetes, and malnutrition-related diabetes), heart failure, myocardial infarction (MI), stroke, chronic obstructive pulmonary disease (COPD), dementia, or death. Participants from the UKB-PPP were followed up during a mean follow-up of 13.5 years. Using an independent internal cohort, HPS was validated against chronological age and biological age measures using clinical biomarkers or proteins. In addition, we investigated the biological processes associated with healthspan through gene set enrichment analysis.

Results

UKB PPP participants free from conditions in the healthspan condition at baseline

Of 502,269 active UKB participants, 53,018 were included in the UKB Pharma Proteomics Project (PPP) and had proteomic data available, as released by the UKB. Of those included in the UKB PPP, 43,119 (81.3%) participants had not been diagnosed with any condition included in the healthspan definition at baseline (**Supplementary Table 1**). Their mean age was 56 years (SD=8.2, range 39 to 70), and the majority were of European descent (93.8%) and female (55.8%). 34.3% of participants held a college or university degree, and the mean level of material deprivation, as measured by the Townsend deprivation index, was lower than the UK population average (-1.3 vs. 0). Additionally, 43.4% were previous or current smokers, and the mean BMI was 27.1 (SD=4.6). The prevalence of hypertension and hypercholesterolemia was 23.2% and 11.2%, respectively (**Supplementary Table 2**).

Baseline characteristics of UKB PPP participants ending healthspan during follow-up

After a mean follow-up of 13.5 years (from baseline to death or last follow-up, whichever occurred first), 12,427 (28.8%) developed at least one condition in the healthspan definition, and the first diagnosis was at the mean age of 66.7 years (SD=8.1). After excluding participants with multiple first-diagnosed conditions in the healthspan definition (n=1,214, 9.8%), cancer was the most common first-diagnosed condition (n=4,882, 44%), followed by MI (n=2,036, 18.2%), diabetes (n=1,455, 13.0%), and then COPD (n=1,131, 10.1%). Compared to participants who developed any condition in the healthspan definition during follow-up, participants who remained healthy were younger and more likely to be female or non-white. They had better socioeconomic status, healthier lifestyles, and a lower prevalence of hypertension and hypercholesterolemia at baseline (**Supplementary Table 2**).

Healthspan Proteomic Score (HPS) development using UKB PPP participants free from conditions in the healthspan definition at baseline

Participants who did not have any conditions in the healthspan definition at baseline (n=43,119) were randomly split into training (70%, n=30,184) and test sets (30%, n=12,935), where 2,296 (7.6%) and 989 (7.6%) participants died during follow-up, respectively (**Supplementary Figure 1**). The training and test samples shared similar participant characteristics at baseline (**Supplementary Table 3**).

We developed the HPS using the training set data, including 2,920 proteins and chronological age. In the variable selection, we selected chronological age and 86 proteins based on a Least Absolute Shrinkage and Selection Operator (LASSO) penalized Cox regression model for achieving a more parsimonious model with a close to the optimal deviance for predicting healthspan, following the one-standard-error rule[18] (lambda 0.009277) (**Supplementary Table**

4). Then, chronological age and the selected proteins were used to fit a Gompertz model (**Supplementary Table 4**), where we calculated the 10-year risk of reaching the end of healthspan (r_h), which was used to derive the healthspan proteomic score (HPS) by 1-r_h. The lower the HPS, the higher the risk of ending healthspan.

HPS and the number of healthspan conditions at baseline among UKB PPP participants

HPS followed a left-skewed distribution in UKB PPP participants free from the healthspan conditions (**Figure 1A**). The distribution shifted to the left and showed a right tail as the number of healthspan conditions increased. The median HPS in those who had developed one or more conditions in the healthspan definition (0.70 [1 condition], 0.49 [2 conditions], 0.36 [3 conditions], 0.17 [4 conditions], 0.07 [5 conditions]) was significantly lower than that (0.84) in those without any condition included in the healthspan definition at baseline ($p<0.001$). Among participants with a chronic medical condition at baseline (n=8,136), those with cancer showed the highest median HPS (0.78), and those with MI (0.63), diabetes (0.62), or heart failure (0.62) showed the lowest median HPS (**Figure 1B**).

HPS and participant characteristics at baseline in the test set

Among the participants in the test set without any condition in the healthspan definition at baseline, the median HPS was significantly lower in males (vs. females), older adults (≥60 years vs. <60 years), former or current smokers (vs. non-smokers), obese participants (BMI≥30 vs. BMI<30), and participants diagnosed with hypertension or hypercholesterolemia before or at baseline ($p<0.001$, **Supplementary Figure 2**).

HPS and chronological age, biological age measures, as well as aging traits at baseline in the test set

HPS showed moderately negative correlations with chronological age (Spearman correlation *r*= -0.73) and previous composite biological age measures (**Supplementary Figure 3**): proteomic aging clock (PAC) trained to predict mortality[15] (*r*= -0.87), PhenoAge[11] (*r*= -0.79) and BioAge[19] (*r*= -0.74) to predict mortality and chronological age, respectively. Moreover, HPS had a low negative correlation with a 49-item frailty index[20] (*r*= -0.21), based on a wide range of self-reported health deficits, and a low positive correlation with leukocyte telomere length[21] (*r*= 0.21) (**Supplementary Figure 3**). In addition, HPS was negatively correlated with BMI (*r*= -0.32), systolic blood pressure (*r*= -0.37), and reaction time based on a cognitive function test (*r*= -0.26) (**Supplementary Figure 3**). HPS showed a positive correlation with usual walking pace (*r*= 0.23) and a minimal correlation with maximal grip strength (*r*= -0.01) (**Supplementary Figure 3**). After regressing out the effect of chronological age, HPS continued to show a moderate correlation with PAC proteomic age (*r*= -0.65) but had lower correlations with PhenoAge (*r*= -0.39) and BioAge (*r*= -0.10) (**Supplementary Figure 4**).

HPS at baseline and risks of ending healthspan, mortality, and age-related conditions during follow-up in the test set

HPS was significantly associated with the time from baseline to the diagnosis of a first healthspan condition in the test set after adjusting for chronological age and other covariates at baseline (**Supplementary Table 1):** sex, ethnicity, education, Townsend deprivation index, BMI, smoking status, hypertension, hypercholesterolemia, and UKB PPP consortium selection for specific disease or ancestry interests (false discovery rate [FDR]-adjusted p-value, $p_{FDR\text{-}adj}=1.22\times10^{-81}$). Using the HPS range of 0.75 to 1 as the reference (with the lowest risk of ending healthspan in the next 10 years among the four equal-interval groups of HPS), the risk of ending healthspan increased as HPS decreased, with an additional 2,000 cases per 100,000 person-years

in participants with an HPS range from 0.5 to 0.75, 6,070 cases and 14,000 cases per 100,000 person-years in those with the ranges from 0.25 to 0.5 and from 0 to 0.25, respectively (**Figure 2**).

Lower HPS was significantly associated with conditions in the healthspan definition ($p_{FDR\text{-}adj}<0.05$), particularly mortality, incident COPD, and diabetes (**Figure 2**). Among specific cancer diagnoses, a lower HPS was significantly associated with lung cancer and prostate cancer ($p_{FDR\text{-}adj}<0.05$), but there was no significant association with colorectal cancer or breast cancer in females ($p_{FDR\text{-}adj}>0.05$) (**Figure 2**). In addition, a lower HPS was associated with other medical conditions that were not included in the healthspan definition, such as pneumonia, chronic kidney disease, delirium, and osteoarthritis ($p_{FDR\text{-}adj}<0.05$) (**Supplementary Figure 5**).

For sensitivity analysis, we dichotomized the HPS using the cut-off of 0.73 ("Low: HPS<0.73 vs. High: HPS≥0.73"). This threshold corresponded to the first quartile of HPS in the UKB PPP participants free from the healthspan conditions, effectively separating participants with and without healthspan conditions at baseline (**Figure 1A**). A low HPS score was associated with higher risks of conditions in the healthspan definition and other medical conditions (**Supplementary Figures S6 and S7**). Interestingly, the association between low HPS and healthspan was stronger in younger adults (<60 vs. ≥60 years), previous or current smokers (vs. never smokers), and obese participants (BMI≥30 vs. <30) (**Supplementary Figure S8**). Previous or current smokers also showed a stronger association with mortality (**Supplementary Figure S9**).

HPS at baseline and a second healthspan condition or mortality during follow-up among UKB PPP participants with a specific healthspan condition at baseline

We further investigated how HPS at baseline is associated with the development of a second healthspan condition and mortality during follow-up in those with a specific healthspan condition at baseline, considering the most prevalent healthspan conditions at baseline: cancer, MI, diabetes, and COPD. HPS was strongly associated with a second healthspan condition and mortality during follow-up, regardless of the pre-existing healthspan condition (**Figure 3**).

Joint analysis of HPS and PAC proteomic age groups at baseline and the risks of conditions in the healthspan definition and other medical conditions during follow-up in the test set

PAC is a recently developed proteomics-based clock to predict mortality[15]. Both HPS and PAC proteomic age strongly predicted conditions in the healthspan definition and other medical conditions (**Figure 4**). However, HPS was better at predicting diabetes, while PAC proteomic age was better at predicting dementia and osteoporosis. Compared to HPS and PAC proteomic age, PhenoAge and chronological age were less predictive across conditions, except chronic kidney disease and osteoporosis, which were best predicted by PhenoAge and chronological age, respectively (**Figure 4**).

Despite their high predictive power for mortality and healthspan, HPS and PAC showed distinct predictive profiles for chronic diseases, suggesting that different biological processes underlie their measurements. We hypothesized that low HPS and high PAC significantly increased the risks of adverse health outcomes than low HPS or high PAC alone. To test the hypothesis, we compared the groups of HPS (high or low) and PAC (high or low) for the risks of ending healthspan, mortality, and medical conditions during follow-up. We categorized HPS and PAC at baseline using the cutoffs 0.73 and 58, respectively, corresponding to the first and third quartiles among UKB PPP participants. Either cutoff effectively separated participants with and

without conditions included in the healthspan definition, as illustrated in **Figure 1A** and **Supplementary Figure 10**.

Among the participants in the test set with complete covariate data, 8,783 participants were in the high HPS and low PAC group (*better biological health*), who were compared with 704 in the low HPS and low PAC group (*intermediate biological health*), 762 in the high HPS and high PAC group (*intermediate biological health*), and 2,349 in the low HPS and high PAC group (*worse biological health*). The low HPS and high PAC showed significantly higher risk than the additive risk associated with low HPS alone and high PAC alone for mortality and end of healthspan (FDR-adjusted interaction $p=1.20\times10^{-5}$ [mortality], 0.002 [end of healthspan]) (**Figure 5**). Similar association patterns across the HPS and PAC groups were found with conditions in the healthspan definition such as heart failure, MI, stroke, and COPD (**Figure 5**), and other medical conditions, including pneumonia and chronic kidney disease (**Supplementary Figure 11**).

Gene set enrichment analysis of proteins associated with HPS

Of 2,920 proteins analyzed, 1,398 showed significant associations (Bonferroni-corrected $p<0.05$) with HPS after adjusting for covariates at baseline (age, sex, ethnicity, education, Townsend deprivation index, BMI, smoking status, hypertension, and hypercholesterolemia, and UKB PPP consortium selection) (**Supplementary Table 5**). Those significantly upregulated or downregulated proteins in the low HPS group are highlighted in **Figure 6A** (mean difference between the low and high HPS group after inverse normal transformation greater than 0.4 SD or smaller than 0.4 SD, chosen to label a maximal number of genes without losing clarity). The significant proteins (Bonferroni-corrected $p<0.05$) showed enrichment in 26 hallmark gene sets,

including immune response, inflammation, cellular signaling, and metabolic regulation (**Figure 6B**).

Discussion

Unlike previous biological age predictors trained to predict chronological age, mortality[22] or organ/disease-specific predictors[23–26], HPS was developed based on a healthspan definition to evaluate the risk of developing a major chronic disease or mortality. HPS outperformed chronological age and other biological age measures in predicting healthspan and various chronic diseases. Subgroup analyses revealed that individuals at higher risk for adverse health outcomes—such as males, older adults, current or former smokers, obese individuals, and those with hypertension or hypercholesterolemia—had lower HPS values, indicating a less healthy systemic biological status in these groups. Additionally, we discovered that the biological processes associated with HPS align with the hallmarks of biological aging, particularly immune response, inflammation, cellular signaling, and metabolic regulation. Overall, our findings demonstrate that HPS has strong clinical, predictive, and biological validity, making it a valuable tool for assessing healthspan in humans and potentially guiding the development of geroscience-based interventions in the future.

Several biological age measures have been developed to predict mortality and the onset of age-related diseases. These measures may reflect different aspects of biological aging due to data variations, study populations, and the methods used for their development[8,27]. We compared HPS with several well-established biological age measures. HPS and PAC showed a strong correlation (r = -0.87), even after controlling for chronological age (r = -0.65), indicating a significant overlap in their measurements. The correlations of HPS with other biological age measures, such as PhenoAge and BioAge, were less pronounced, and those with leukocyte

telomere length and a 49-item frailty index were weak to moderate. Although these measures are strongly associated with mortality and disability, they vary significantly in their ability to predict specific chronic medical conditions. While their measurements may overlap, they also capture distinct facets of the complex biology of aging.

HPS may offer several advantages over other biological age measures, particularly in the context of geroscience-guided interventions. It was trained on relatively young individuals who had not developed chronic medical conditions at baseline and were followed until a first healthspan condition. HPS strongly predicts healthspan, mortality, and the onset of medical conditions in participants free from the conditions in the healthspan definition. Additionally, HPS predicts comorbidities and mortality in those with a specific chronic medical condition. By monitoring the effects of interventions on biological health, HPS can shorten trial lengths by serving as a surrogate measure for healthspan or chronic medical conditions. Therefore, incorporating HPS into clinical trials could enhance the evaluation of interventions to improve overall biological health and prevent chronic diseases.

The joint analysis of HPS and PAC revealed intriguing patterns of associations with healthspan, mortality, and specific chronic medical conditions. Most individuals (70%) were classified in the better biological health group (i.e., high HPS and low PAC). Meanwhile, 19% were in the worse biological health group (i.e., low HPS and high PAC), and a small group, comprising 12% of the sample, were in the intermediate biological health groups (i.e., those with low HPS and low PAC or high HPS and high PAC). As expected, individuals in the worse biological health group showed the highest risk of adverse health outcomes. Those in the intermediate biological health group also had an elevated risk, although the associations were not as strong as those in the worse biological health group. Our findings imply that there is a group

of "clinically" healthy individuals who already exhibit significant systemic biological changes, indicating a higher risk of disease development. This group may benefit the most from geroscience-guided interventions, particularly those in transitional stages of biological health (i.e., intermediate biological health groups) who are at higher risks of diseases but could also respond well to such interventions. Prior studies have evaluated proteomic markers associated with chronological age[28–30] and longevity, using various definitions of longevity, such as parental age at death above 95 years[31] or survival to 90 years[32]. These studies identified robust proteomic signatures related to chronological age and longevity. Interestingly, a few markers are common across these studies, including ours, such as GDF-15, pleiotrophin, and chordin-like protein 1. Despite differences in proteomic platforms, proteome coverage, statistical modeling, and outcomes (e.g., chronological age, longevity, and healthspan), there appears to be a common proteomic signature that can capture multiple processes involved in healthy aging. Although still elusive, this common signature offers novel opportunities to develop interventions targeting multiple aging-related conditions.

Several limitations should be noted when interpreting our results:

1. Diagnostic Lag: Disease diagnosis often occurs after the actual onset of pathological processes. Consequently, some participants included in the HPS training set might have already experienced biological changes related to specific chronic medical conditions, which could influence the selection of proteins in the HPS model. To address this issue, we used multi-source data to determine the first diagnosis date, which may reduce the diagnostic lag. Additionally, this diagnostic lag likely occurred randomly, which is expected to drive associations toward the null, meaning our results may be more conservative than the actual associations[33].

2. Healthspan definition: Given that there is no consensus on the healthspan definition, predictors trained using different definitions may differ from HPS in assessing biological aging. We adopted a practical healthspan definition using electronic health records, which differs from the definition used by Walter et al.[34]. They defined healthspan as the number of years from birth to the first occurrence of conditions in our healthspan definition plus hip fracture. Additionally, a critique of both definitions is that, except for dementia, late-onset diseases are underrepresented, meaning those who remain disease-free may not necessarily be in good health[35]. Including additional late-onset diseases is unlikely to significantly change our results, as the UKB is a relatively young cohort, and our goal was to train a predictor for early interventions. It is important to note that although our analyses focused on the conditions used to define healthspan, other diseases could also impact healthspan. Therefore, our results do not necessarily reflect a proteomic signature of the complete absence of disease. However, we showed that HPS was a strong predictor of other age-related chronic medical conditions not included in the healthspan definition, indicating that it can be used to assess the risk for multiple chronic conditions.

3. Comparison with Epigenetic Clocks: Although epigenetic clocks are the most popular biological age predictors, we could not include any for comparison with HPS due to the lack of epigenetic data in the UKB. However, PhenoAge was used to train DNAm PhenoAge, thus providing an indirect comparison between HPS and DNAm PhenoAge clock[11].

4. Validation with External Cohorts*:* We were unable to validate HPS using external cohorts because no large cohort uses the same proteomic assay as the UKB. Although other cohorts have proteomic data available from assays such as different versions of the Somalogic assay, significant differences in proteome coverage and analytical characteristics prevent us from using these data to validate our models externally.

5. HPS vs. organ/disease-specific clocks: Despite its robustness, HPS was developed as a tool to test the geroscience hypothesis and may serve as a systemic gauge for monitoring biological health. Recent organ- or disease-specific clocks[23–26] may have stronger predictive power for particular diseases of interest. However, HPS may be more useful in studies aimed at monitoring the development of multiple diseases simultaneously.

6. Population Diversity*:* The UKB has limited racial and ethnic diversity, being predominantly composed of individuals of Caucasian ancestry. Therefore, our models may not be generalized to more diverse populations.

In conclusion, we have developed a novel healthspan proteomic score termed HPS, which demonstrated robust associations and predictions for healthspan, mortality, and the onset of various diseases. The proteins associated with HPS were enriched in several processes related to the hallmarks of biological aging. HPS can be useful for gauging an individual's biological health and monitoring the impact of geroscience-guided interventions, serving as a surrogate marker of healthspan.

Online Methods

The UKB is a volunteer community cohort in the United Kingdom that recruited over 500,000 participants aged 40 to 70 between 2006 and 2010[36]. During baseline assessments, data were collected on sociodemographic and lifestyle factors, environmental exposures, health and medical history, self-reported medications, cognitive function, physical activity, physical measurements, and biological samples for future assays. Over the years, various biological data have been generated, including plasma proteomic data from a nested cohort, as part of the UKB PPP[37]. The baseline cohort participants have been followed up through linkages with electronic health records for disease diagnoses and deaths[38].

The UKB PPP participants (initially n=54,219) were primarily a random sample (n=46,595, 85.9%) from the UKB baseline cohort. Additionally, the cohort included individuals selected by the UKB PPP consortium of 13 biopharmaceutical companies with specific disease or ancestry interests (n=6,376) and participants attending the COVID-19 repeat imaging study (n=1,268)[37]. This study included proteomic data from 53,018 active UKB PPP participants after quality control. Protein concentrations of 2,923 proteins were measured using the Olink Explore 3072 platform and normalized through a two-step process involving within-batch and across-batch intensity normalization. [39]. The normalized protein expression (NPX) data were used throughout the present project. Three proteins, GLIPR1 (99.7%), NPM1 (74.0%), and PCOLCE (63.6%), were excluded due to high missing rates. The median individual missing rate was 0.5% (25th percentile 0.1% and 75th percentile 7.5%). However,7.2% of the included samples only had complete proteomic data, necessary for the HPS development methods. To address this issue, we imputed the missing proteomic data using the *k*-nearest neighbors approach (*k*=10) with the R package 'multiUS'[40].

Healthspan was defined as the number of years from birth free of cancer (excluding non-melanoma skin cancer), diabetes (including type I diabetes, type II diabetes, and malnutrition-related diabetes), heart failure, MI, stroke, COPD, dementia, or death, in line with previous studies[16,17]. The UKB acquired the date of death data through linkages with national death registries and derived the first diagnosis dates for conditions in the healthspan definition and other medical conditions based on ICD-10 codes (**Supplementary Table 1**), which were used to link multi-source data, including primary care records, hospital inpatient data, cancer and death registries records, and self-reported medical condition codes.

Participants without any condition in the healthspan definition at baseline were randomly split into a training set and a test set in a seven-to-three ratio (**Supplementary Figure 1**). The training set data, including chronological age and NPX of 2920 proteins, was related to time from baseline to a first healthspan condition (censored at death or last follow-up of hospital inpatient data [main source], whichever occurred first) using a LASSO Cox regression model. The death censoring date was 2022/11/30. The inpatient censoring dates were 2022/11/30 for participants attending baseline assessment centers in England, 2021/7/31 for participants in Scotland, and 2018/2/28 for those in Wales. Proteins with non-zero regression coefficients in the fitted model were carried forward to fit a Gompertz regression model. Based on the cumulative density probability of this model, the risk of ending healthspan within 10 years from baseline can be derived (r), with 1-r representing the likelihood of remaining healthy, referred to as the healthspan proteomic score (HPS).

We tested the validity of HPS using participants in the test set data by examining:

- Spearman correlations of HPS with chronological age, biological age measures, as well as aging traits at baseline, and between HPS, PAC, and PhenoAge after removing the effect of chronological age in linear regression models, where the composite measures of PAC[15], PhenoAge[11], BioAge[19], and a 49-item frailty index[20] were derived using data fields listed in **Supplementary Table 1.**
- Associations of HPS at baseline with healthspan, mortality, and medical conditions during follow-up (detailed in **Supplementary Table 1**).
- Associations of HPS at baseline with the conditions above during follow-up in subgroups by age (≥60, <60), sex (male, female), smoking status (previous or current smokers,

never smokers), BMI (≥30, <30), hypertension (yes/no), and hypercholesterolemia (yes/no).

- Predictions for the conditions above during follow-up against chronological age, PAC, and PhenoAge
- The combined effect of low HPS (HPS<0.73) and high PAC (PAC≥58) versus each alone for associations with the conditions above during follow-up.
- In addition, we tested the validity of HPS in those with a specific healthspan condition at baseline for associations with a second healthspan condition and mortality.

The association analysis was conducted using Aalen's additive hazard models[41] adjusting for baseline covariates (sex, ethnicity, education, Townsend deprivation index, BMI, smoking status, hypertension, hypercholesterolemia, and UKB PPP consortium selection, **Supplementary Table 1**), which facilitated testing for an additive interaction effect between HPS and PAC[42]. Throughout the association analysis, p-values were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate method[43]. Harrell's C statistic[44] under a Cox regression model was used to assess the predictive accuracy of HPS versus chronological age, PAC, and PhenoAge for ending healthspan, mortality, and medical conditions during follow-up.

To understand the measurement of HPS, we conducted a gene set enrichment analysis for proteins associated with low HPS (HPS<0.73) after adjusting for baseline covariates. Prior to the association analysis, the inverse normal transformation was applied to each protein to normalize the distribution and unify the scale into z-scores. Proteins significant at the Bonferroni-corrected level were entered into the gene set analysis implemented in the Functional Mapping and Annotation of Genome-Wide Association Studied (FUMA version 1.5.2). Genes associated with HPS 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).

Author Contributions

CLK and BSD designed the study, and CLK and PL processed and analyzed the data. CLK and BSD drafted the initial manuscript, which included contributions from PL. All the authors reviewed and approved the final version.

Data Availability Statements

Data access is granted upon application to the UK Biobank. The R code (Liu & Kuo n.d.) for computing HPS can be obtained from the GitHub repository at https://github.com/kuo-lab-uchc/HPS.

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.

Funding Information

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).

Conflict of Interest Statement

We have no conflicting interests to disclose.

References

1. GHE: Life expectancy and healthy life expectancy [Internet]. [cited 2024 May 21]. Available from: https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghe-life-expectancy-and-healthy-life-expectancy

2. Garmany A, Yamada S, Terzic A. Longevity leap: mind the healthspan gap. NPJ Regen Med. 2021 Sep 23;6(1):57. PMCID: PMC8460831

3. Khan M, Al Saud H, Sierra F, Perez V, Greene W, Al Asiry S, Pathai S, Torres M. Global Healthspan Summit 2023: closing the gap between healthspan and lifespan. Nat Aging. 2024 Apr;4(4):445–448. PMID: 38486031

4. Kennedy BK, Berger SL, Brunet A, Campisi J, Cuervo AM, Epel ES, Franceschi C, Lithgow GJ, Morimoto RI, Pessin JE, Rando TA, Richardson A, Schadt EE, Wyss-Coray T, Sierra F. Geroscience: linking aging to chronic disease. Cell. 2014 Nov 6;159(4):709–713. PMCID: PMC4852871

5. Fuellen G, Jansen L, Cohen AA, Luyten W, Gogol M, Simm A, Saul N, Cirulli F, Berry A, Antal P, Köhling R, Wouters B, Möller S. Health and Aging: Unifying Concepts, Scores, Biomarkers and Pathways. Aging Dis. 2019 Aug;10(4):883–900. PMCID: PMC6675520

6. Kaeberlein M. How healthy is the healthspan concept? Geroscience. 2018 Aug;40(4):361–364. PMCID: PMC6136295

7. Lara J, Cooper R, Nissan J, Ginty AT, Khaw KT, Deary IJ, Lord JM, Kuh D, Mathers JC. A proposed panel of biomarkers of healthy ageing. BMC Med. 2015 Sep 15;13:222. PMCID: PMC4572626

8. Belsky DW, Moffitt TE, Cohen AA, Corcoran DL, Levine ME, Prinz JA, Schaefer J, Sugden K, Williams B, Poulton R, Caspi A. Eleven Telomere, Epigenetic Clock, and Biomarker-Composite Quantifications of Biological Aging: Do They Measure the Same Thing? Am J Epidemiol. 2018 Jun 1;187(6):1220–1230. PMCID: PMC6248475

9. Blodgett JM, Pérez-Zepeda MU, Godin J, Kehler DS, Andrew MK, Kirkland S, Rockwood K, Theou O. Prognostic accuracy of 70 individual frailty biomarkers in predicting mortality in the Canadian Longitudinal Study on Aging. Geroscience. 2024 Jun;46(3):3061–3069. PMCID: PMC11009196

10. Liu Z, Kuo PL, Horvath S, Crimmins E, Ferrucci L, Levine M. A new aging measure captures morbidity and mortality risk across diverse subpopulations from NHANES IV: A cohort study. Basu S, editor. PLoS Med. 2018 Dec 31;15(12):e1002718.

11. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, Hou L, Baccarelli AA, Stewart JD, Li Y, Whitsel EA, Wilson JG, Reiner AP, Aviv A, Lohman K, Liu Y, Ferrucci L, Horvath S. An epigenetic biomarker of aging for lifespan and healthspan. Aging (Albany NY). 2018 Apr 18;10(4):573–591. PMCID: PMC5940111

12. Yang J, Lu J, Miao J, Li J, Zhu M, Dai J, Ma H, Jin G, Hang D. Development and validation of a blood biomarker score for predicting mortality risk in the general population. J Transl Med. 2023 Jul 15;21(1):471. PMCID: PMC10349520

13. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, Hou L, Baccarelli AA, Li Y, Stewart JD, Whitsel EA, Assimes TL, Ferrucci L, Horvath S. DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging. 2019 Jan 21;11(2):303–327.

14. Deelen J, Kettunen J, Fischer K, van der Spek A, Trompet S, Kastenmüller G, Boyd A, Zierer J, van den Akker EB, Ala-Korpela M, Amin N, Demirkan A, Ghanbari M, van Heemst D, Ikram MA, van Klinken JB, Mooijaart SP, Peters A, Salomaa V, Sattar N, Spector TD, Tiemeier H, Verhoeven A, Waldenberger M, Würtz P, Davey Smith G, Metspalu A, Perola M, Menni C, Geleijnse JM, Drenos F, Beekman M, Jukema JW, van Duijn CM, Slagboom PE. A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. Nat Commun. 2019 Aug 20;10(1):3346. PMCID: PMC6702196

15. Kuo CL, Chen Z, Liu P, Pilling LC, Atkins JL, Fortinsky RH, Kuchel GA, Diniz BS. Proteomic aging clock (PAC) predicts age-related outcomes in middle-aged and older adults. Aging Cell. 2024 May 15;e14195. PMID: 38747160

16. Li X, Ploner A, Wang Y, Zhan Y, Pedersen NL, Magnusson PK, Jylhävä J, Hägg S. Clinical biomarkers and associations with healthspan and lifespan: Evidence from observational and genetic data. EBioMedicine. 2021 Apr;66:103318. PMCID: PMC8047464

17. Zenin A, Tsepilov Y, Sharapov S, Getmantsev E, Menshikov LI, Fedichev PO, Aulchenko Y. Identification of 12 genetic loci associated with human healthspan. Commun Biol. 2019;2:41. PMCID: PMC6353874

18. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33(1):1–22. PMCID: PMC2929880

19. Levine ME. Modeling the rate of senescence: can estimated biological age predict mortality more accurately than chronological age? J Gerontol A Biol Sci Med Sci. 2013 Jun;68(6):667–674. PMCID: PMC3660119

20. Williams DM, Jylhävä J, Pedersen NL, Hägg S. A Frailty Index for UK Biobank Participants. J Gerontol A Biol Sci Med Sci. 2019 Mar 14;74(4):582–587. PMCID: PMC6417451

21. Codd V, Denniff M, Swinfield C, Warner SC, Papakonstantinou M, Sheth S, Nanus DE, Budgeon CA, Musicha C, Bountziouka V, Wang Q, Bramley R, Allara E, Kaptoge S, Stoma S, Jiang T, Butterworth AS, Wood AM, Di Angelantonio E, Thompson JR, Danesh JN, Nelson CP, Samani NJ. Measurement and initial characterization of leukocyte telomere length in 474,074 participants in UK Biobank. Nat Aging. 2022 Feb;2(2):170–179.

22. Moqri M, Herzog C, Poganik JR, Biomarkers of Aging Consortium, Justice J, Belsky DW, Higgins-Chen A, Moskalev A, Fuellen G, Cohen AA, Bautmans I, Widschwendter M, Ding J, Fleming A, Mannick J, Han JDJ, Zhavoronkov A, Barzilai N, Kaeberlein M, Cummings S, Kennedy BK, Ferrucci L, Horvath S, Verdin E, Maier AB, Snyder MP, Sebastiano V, Gladyshev VN. Biomarkers of aging for the identification and evaluation of longevity interventions. Cell. 2023 Aug 31;186(18):3758–3775. PMID: 37657418

23. Oh HSH, Rutledge J, Nachun D, Pálovics R, Abiose O, Moran-Losada P, Channappa D, Urey DY, Kim K, Sung YJ, Wang L, Timsina J, Western D, Liu M, Kohlfeld P, Budde J, Wilson EN, Guen Y, Maurer TM, Haney M, Yang AC, He Z, Greicius MD, Andreasson KI, Sathyan S, Weiss EF, Milman S, Barzilai N, Cruchaga C, Wagner AD, Mormino E, Lehallier B, Henderson VW, Longo FM, Montgomery SB, Wyss-Coray T. Organ aging signatures in the plasma proteome track health and disease. Nature. 2023 Dec 7;624(7990):164–172.

24. Goeminne LJE, Eames A, Tyshkovskiy A, Argentieri MA, Ying K, Moqri M, Gladyshev VN. Plasma-based organ-specific aging and mortality models unveil diseases as accelerated aging of organismal systems [Internet]. 2024 [cited 2024 May 24]. Available from: http://medrxiv.org/lookup/doi/10.1101/2024.04.08.24305469

25. Sehgal R, Meer M, Shadyab AH, Casanova R, Manson JE, Bhatti P, Crimmins EM, Assimes TL, Whitsel EA, Higgins-Chen AT, Levine M. Systems Age: A single blood methylation test to quantify aging heterogeneity across 11 physiological systems [Internet]. 2023 [cited 2024 Apr 4]. Available from: http://biorxiv.org/lookup/doi/10.1101/2023.07.13.548904

26. You J, Guo Y, Zhang Y, Kang JJ, Wang LB, Feng JF, Cheng W, Yu JT. Plasma proteomic profiles predict individual future health risk. Nat Commun. 2023 Nov 28;14(1):7817.

27. Ferrucci L, Gonzalez-Freire M, Fabbri E, Simonsick E, Tanaka T, Moore Z, Salimi S, Sierra F, de Cabo R. Measuring biological aging in humans: A quest. Aging Cell. 2020 Feb;19(2):e13080. PMCID: PMC6996955

28. Menni C, Kiddle SJ, Mangino M, Viñuela A, Psatha M, Steves C, Sattlecker M, Buil A, Newhouse S, Nelson S, Williams S, Voyle N, Soininen H, Kloszewska I, Mecocci P, Tsolaki M, Vellas B, Lovestone S, Spector TD, Dobson R, Valdes AM. Circulating Proteomic Signatures of Chronological Age. J Gerontol A Biol Sci Med Sci. 2015 Jul;70(7):809–816. PMCID: PMC4469006

29. Tanaka T, Biancotto A, Moaddel R, Moore AZ, Gonzalez-Freire M, Aon MA, Candia J, Zhang P, Cheung F, Fantoni G, CHI consortium, Semba RD, Ferrucci L. Plasma proteomic signature of age in healthy humans. Aging Cell. 2018 Oct;17(5):e12799. PMCID: PMC6156492

30. Lehallier B, Gate D, Schaum N, Nanasi T, Lee SE, Yousef H, Moran Losada P, Berdnik D, Keller A, Verghese J, Sathyan S, Franceschi C, Milman S, Barzilai N, Wyss-Coray T. Undulating changes in human plasma proteome profiles across the lifespan. Nat Med. 2019 Dec;25(12):1843–1850. PMCID: PMC7062043

31. Sathyan S, Ayers E, Gao T, Weiss EF, Milman S, Verghese J, Barzilai N. Plasma proteomic profile of age, health span, and all-cause mortality in older adults. Aging Cell. 2020 Nov;19(11):e13250. PMCID: PMC7681045

32. Liu X, Axelsson GT, Newman AB, Psaty BM, Boudreau RM, Wu C, Arnold AM, Aspelund T, Austin TR, Gardin JM, Siggeirsdottir K, Tracy RP, Gerszten RE, Launer LJ, Jennings LL, Gudnason V, Sanders JL, Odden MC. Plasma proteomic signature of human longevity. Aging Cell. 2024 Mar 5;e14136. PMID: 38440820

33. Loken E, Gelman A. Measurement error and the replication crisis. Science. 2017 Feb 10;355(6325):584–585. PMID: 28183939

34. Walter S, Atzmon G, Demerath EW, Garcia ME, Kaplan RC, Kumari M, Lunetta KL, Milaneschi Y, Tanaka T, Tranah GJ, Völker U, Yu L, Arnold A, Benjamin EJ, Biffar R, Buchman AS, Boerwinkle E, Couper D, De Jager PL, Evans DA, Harris TB, Hoffmann W, Hofman A, Karasik D, Kiel DP, Kocher T, Kuningas M, Launer LJ, Lohman KK, Lutsey PL, Mackenbach J, Marciante K, Psaty BM, Reiman EM, Rotter JI, Seshadri S, Shardell MD, Smith AV, van Duijn C, Walston J, Zillikens MC, Bandinelli S, Baumeister SE, Bennett DA, Ferrucci L, Gudnason V, Kivimaki M, Liu Y, Murabito JM, Newman AB, Tiemeier H, Franceschini N. A genome-wide association study of aging. Neurobiol Aging. 2011 Nov;32(11):2109.e15–28. PMCID: PMC3193030

35. Deelen J. Targeting multimorbidity: Using healthspan and lifespan to identify biomarkers of ageing that pinpoint shared disease mechanisms. EBioMedicine. 2021 May;67:103364. PMCID: PMC8114127

36. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015 Mar;12(3):e1001779. PMCID: PMC4380465

37. Sun BB, Chiou J, Traylor M, Benner C, Hsu YH, Richardson TG, Surendran P, Mahajan A, Robins C, Vasquez-Grinnell SG, Hou L, Kvikstad EM, Burren OS, Davitte J, Ferber KL, Gillies CE, Hedman ÅK, Hu S, Lin T, Mikkilineni R, Pendergrass RK, Pickering C, Prins B, Baird D, Chen CY, Ward LD, Deaton AM, Welsh S, Willis CM, Lehner N, Arnold M, Wörheide MA, Suhre K, Kastenmüller G, Sethi A, Cule M, Raj A, Alnylam Human Genetics, AstraZeneca Genomics Initiative, Biogen Biobank Team, Bristol Myers Squibb, Genentech Human Genetics, GlaxoSmithKline Genomic Sciences, Pfizer Integrative Biology, Population Analytics of Janssen Data Sciences, Regeneron Genetics Center, Kang HM, Burkitt-Gray L, Melamud E, Black MH, Fauman EB, Howson JMM, Kang HM, McCarthy MI, Nioi P, Petrovski S, Scott RA, Smith EN, Szalma S, Waterworth DM, Mitnaul LJ, Szustakowski JD, Gibson BW, Miller MR, Whelan CD. Plasma

proteomic associations with genetics and health in the UK Biobank. Nature. 2023 Oct 12;622(7982):329–338.

38. Allen NE, Lacey B, Lawlor DA, Pell JP, Gallacher J, Smeeth L, Elliott P, Matthews PM, Lyons RA, Whetton AD, Lucassen A, Hurles ME, Chapman M, Roddam AW, Fitzpatrick NK, Hansell AL, Hardy R, Marioni RE, O'Donnell VB, Williams J, Lindgren CM, Effingham M, Sellors J, Danesh J, Collins R. Prospective study design and data analysis in UK Biobank. Sci Transl Med. 2024 Jan 10;16(729):eadf4428. PMID: 38198570

39. Sun B, Ferber K, Lin T, Whelan C. UK Biobank Pharma Proteomics Project: Olink quality control summary [Internet]. [cited 2024 Jun 2]. Available from: https://biobank.ndph.ox.ac.uk/ukb/refer.cgi?id=4658

40. Torgo L. Data Mining with R [Internet]. 0 ed. Chapman and Hall/CRC; 2011 [cited 2023 Aug 18]. Available from: https://www.taylorfrancis.com/books/9781439876404

41. Aalen O. A Model for Nonparametric Regression Analysis of Counting Processes. In: Klonecki W, Kozek A, Rosiński J, editors. Mathematical Statistics and Probability Theory [Internet]. New York, NY: Springer New York; 1980 [cited 2024 Jun 3]. p. 1–25. Available from: http://link.springer.com/10.1007/978-1-4615-7397-5_1

42. Rod NH, Lange T, Andersen I, Marott JL, Diderichsen F. Additive Interaction in Survival Analysis: Use of the Additive Hazards Model. Epidemiology. 2012 Sep;23(5):733–737.

43. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological). 1995 Jan;57(1):289–300.

44. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA. 1982 May 14;247(18):2543–2546. PMID: 7069920

Figure Legends

Figure 1 Healthspan Proteomic Score (HPS) distributions in UKB PPP participants, separated by the number of healthspan conditions at baseline (Figure 1A), and within those with a specific healthspan condition only at baseline (Figure 1B).

Figure 2 Risk differences comparing healthspan proteomic score (HPS) of (0.5, 0.75], (0.25, 0.5], or (0, 0.25) to HPS (0.75, 1] at baseline for ending healthspan or developing conditions in the healthspan definition using disease-free participants in the test set.

Figure 3 Risk differences comparing healthspan proteomic score (HPS) of (0.5, 0.75], (0.25, 0.5], or (0, 0.25) to HPS (0.75, 1] at baseline for a second healthspan condition (Figure 5a), or mortality (Figure 5b), in participants initially presenting with a healthspan condition at baseline.

Figure 4 Predictive power of healthspan proteomic score (HPS) versus proteomic aging clock (PAC) and PhenoAge assessed by C-statistics for time to the end of healthspan, incident diseases, and mortality using the test set data.

Figure 5 Risk differences comparing more adverse HPS and PAC groups to high HPS and low PAC group for mortality, end of healthspan, and incident chronic medical conditions in the healthspan definition in the test set.

Figure 6 HPS-associated proteins and enriched hallmark gene sets.

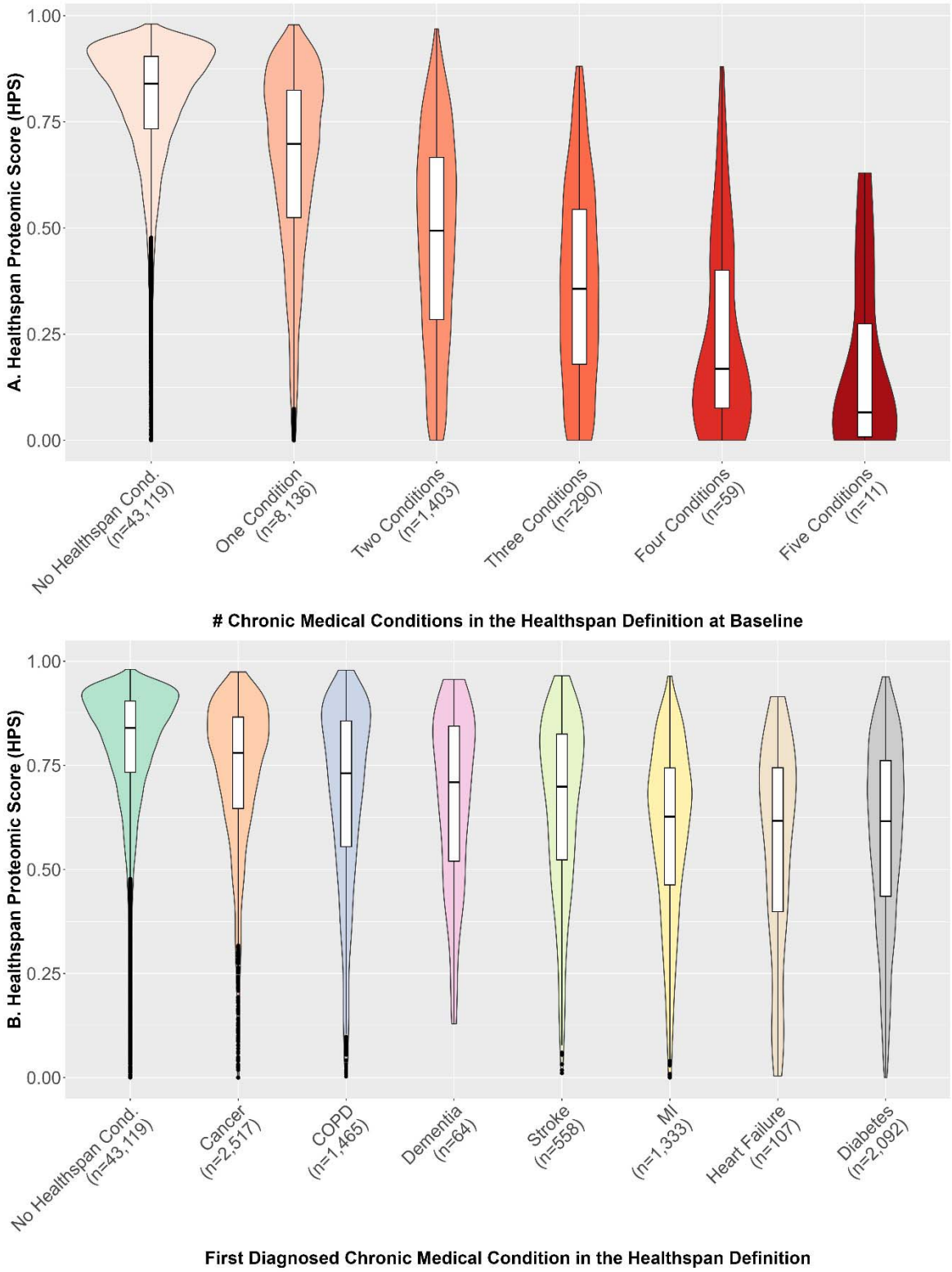

Figure 1 Healthspan Proteomic Score (HPS) distributions in UKB PPP participants, separated by the number of healthspan conditions at baseline (Figure 1A), and within those with a specific healthspan condition only at baseline (Figure 1B).

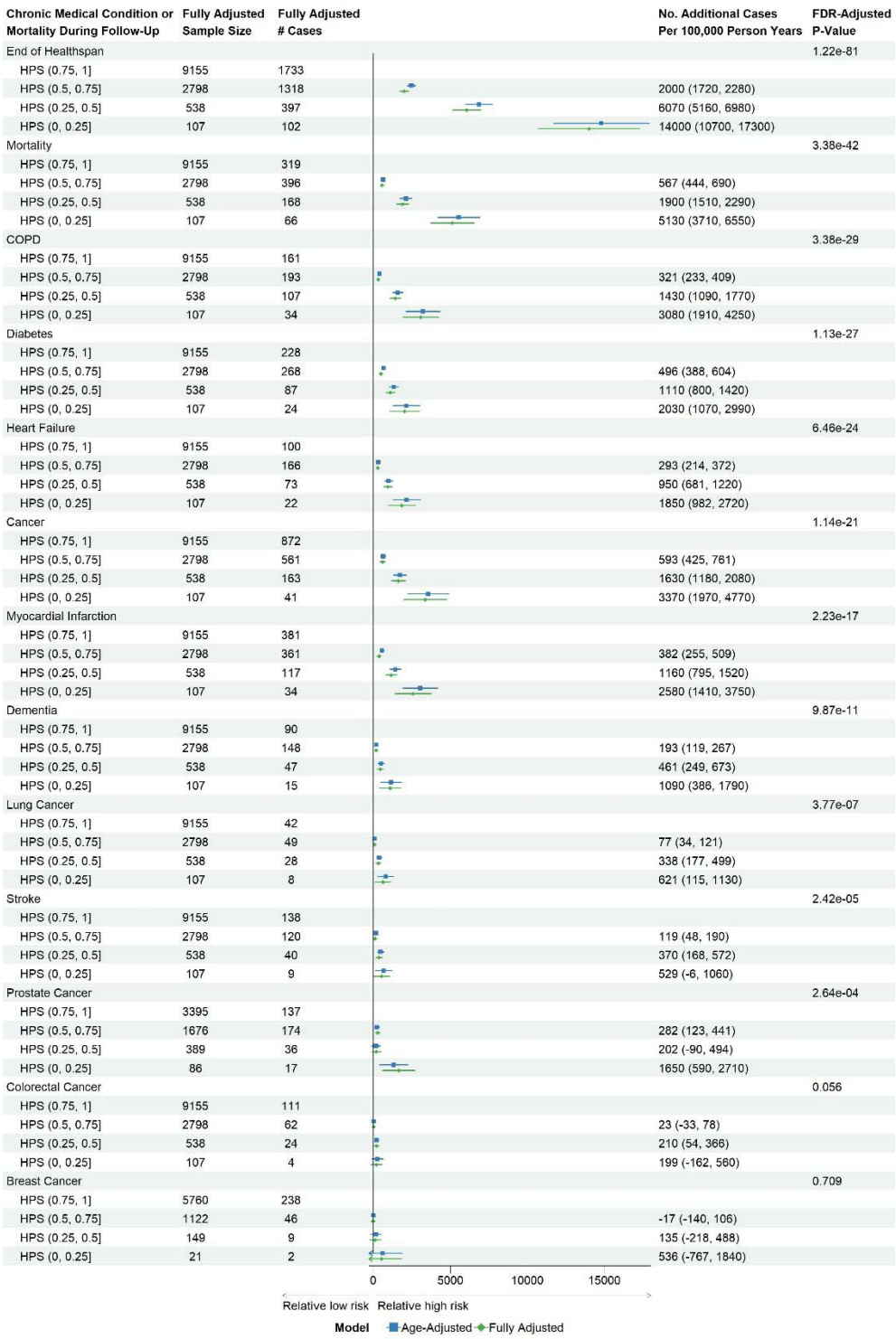

Figure 2 Risk differences comparing healthspan proteomic score (HPS) of (0.5, 0.75], (0.25, 0.5], or (0, 0.25) to HPS (0.75, 1] at baseline for ending healthspan or developing conditions in the healthspan definition using disease-free participants in the test set. Fully adjusted covariates: age, sex, ethnicity, BMI, education, Townsend deprivation index, smoking status, hypertension, hypercholesterolemia, and UKB PPP consortium selection.

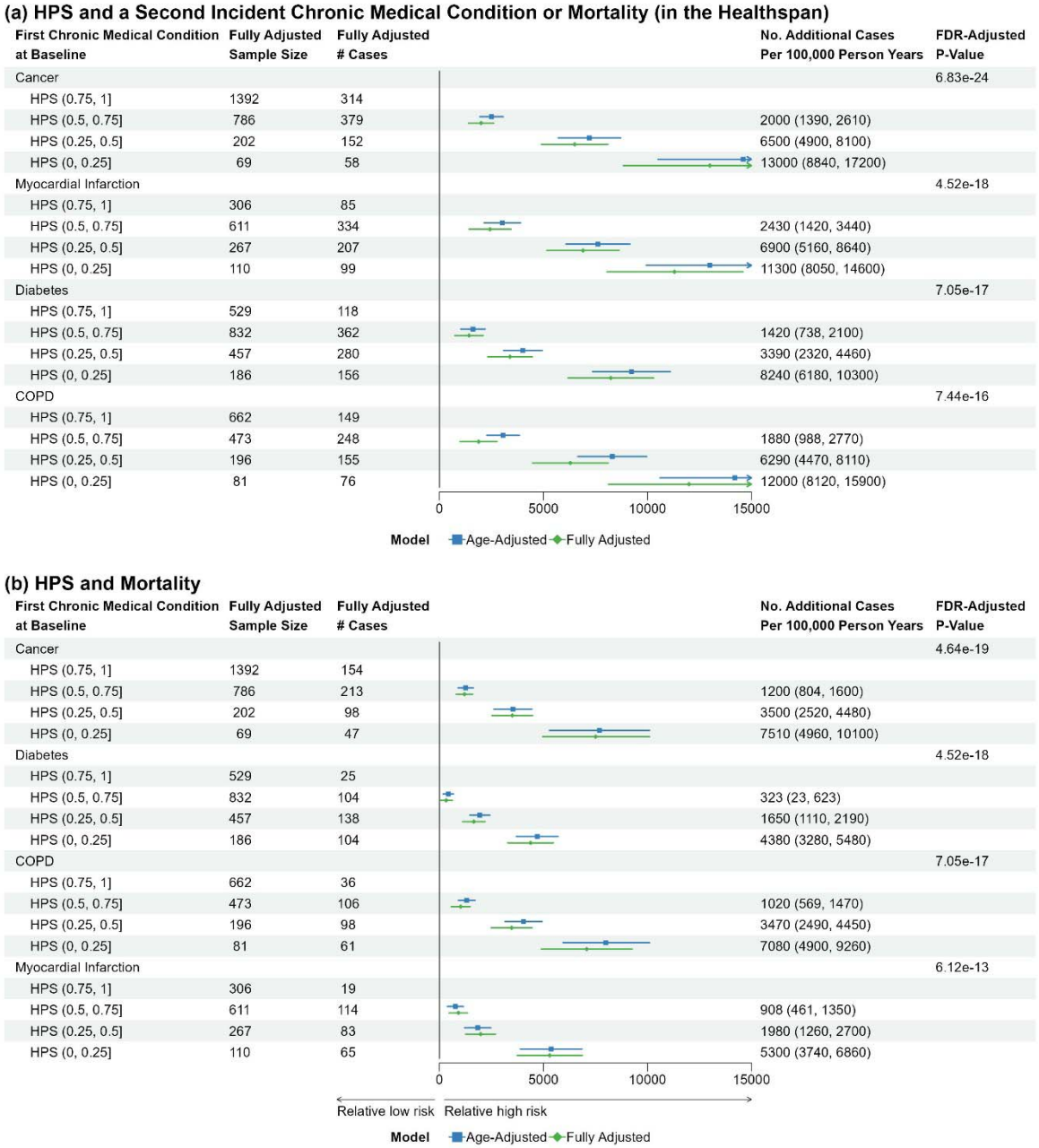

Figure 3 Risk differences comparing healthspan proteomic score (HPS) of (0.5, 0.75], (0.25, 0.5], or (0, 0.25) to HPS (0.75, 1] at baseline for a second healthspan condition (Figure 5a), or mortality (Figure 5b), in participants initially presenting with a healthspan condition at baseline. Fully adjusted baseline covariates: age, sex, ethnicity, BMI, education, Townsend, smoking, hypertension, hypercholesterolemia, and UKB PPP consortium selection.

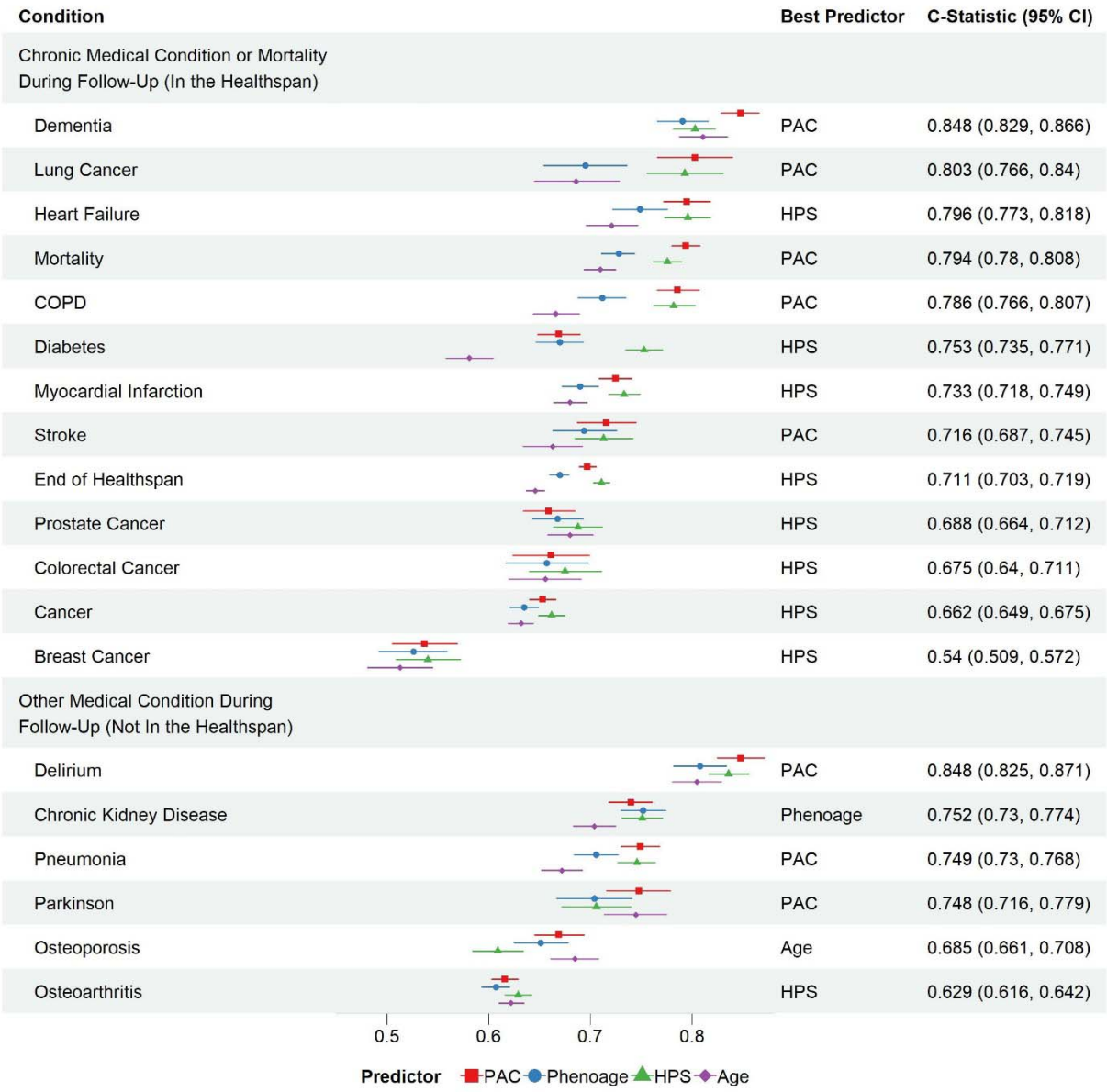

Figure 4 Predictive power of healthspan proteomic score (HPS) versus proteomic aging clock (PAC) and PhenoAge assessed by C-statistics for time to the end of healthspan, incident diseases, and mortality using the test set data.

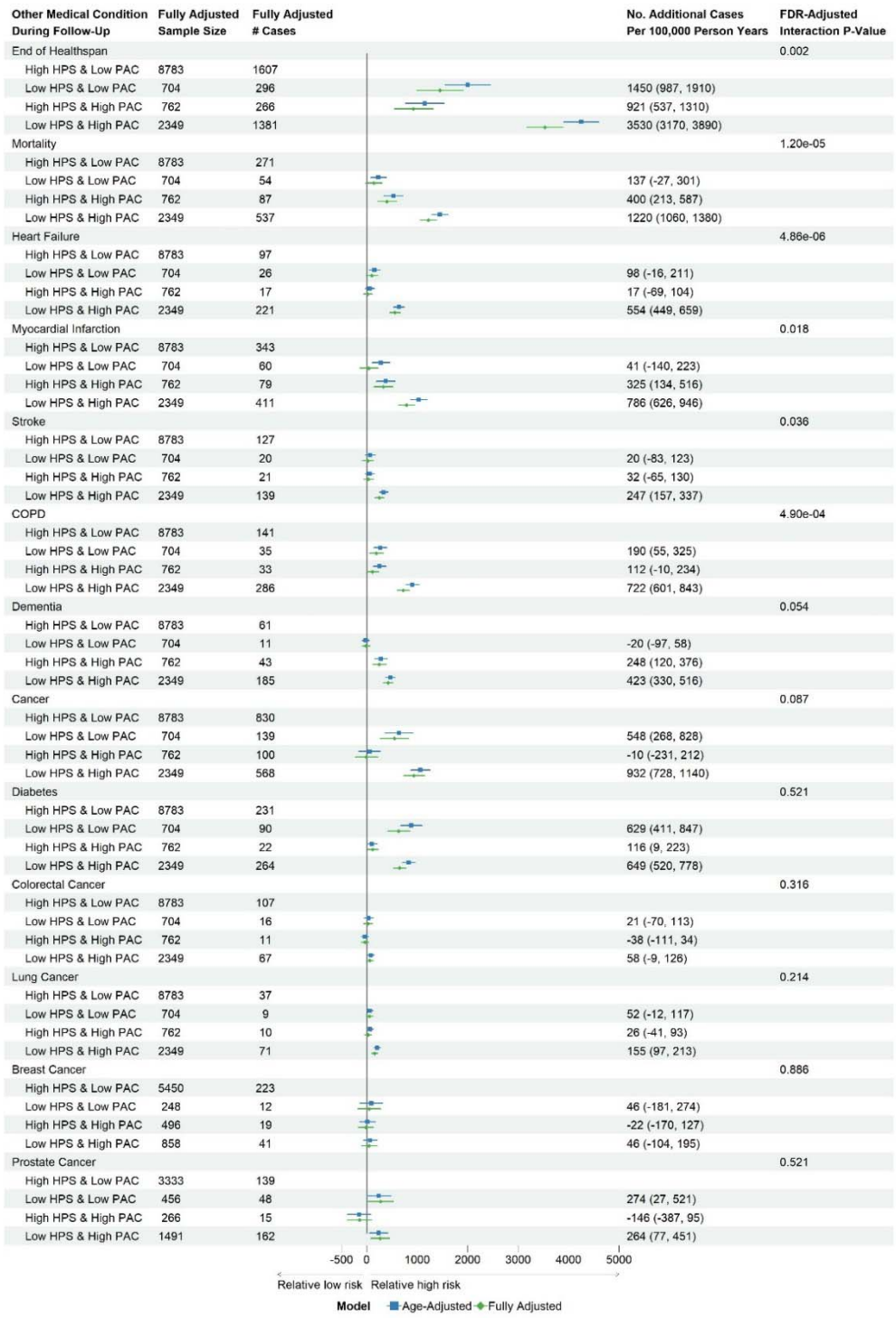

Figure 5 Risk differences comparing more adverse HPS and PAC groups to high HPS and low PAC group for mortality, end of healthspan, and incident chronic medical conditions in the healthspan definition in the test set. Fully adjusted covariates: age, sex, ethnicity, BMI, education, Townsend, smoking, hypertension, hypercholesterolemia, UKB PPP consortium selection, and the conditions in the definition of healthspan.

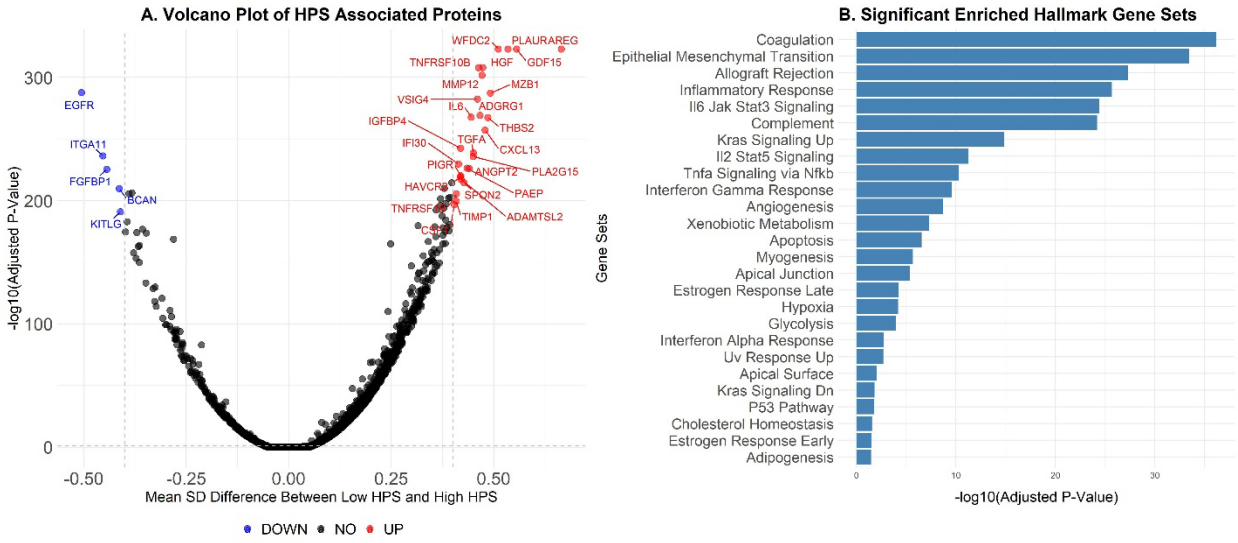

Figure 6 HPS-associated proteins and enriched hallmark gene sets. A. Each point represents a protein, with the x-axis indicating the mean standard deviation (SD) difference in expression after inverse normal transformation (beta value) between the low and high HPS groups and the y-axis representing the statistical significance (-log10 adjusted p-value for multiple testing using Bonferroni correction). *Blue points* indicate significantly downregulated proteins in the low HPS group (beta < -0.4 and adjusted p-value < 0.05). *Red points* indicate significantly upregulated proteins in the low HPS group (beta > 0.4 and adjusted p-value < 0.05). *Black points* indicate proteins that are not significantly differentially expressed. **B**. The x-axis represents the significant gene sets, ordered by adjusted p-values for multiple testing, and the y-axis shows the -log10 of the adjusted p-value for each significant gene set (Bonferroni-corrected p-value <0.05).