nature portfolio

Peer Review File

Homeostasis landscape of the human blood proteome in health and disease

Corresponding Author: Dr Jinzhuo Wang

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
In this study the authors present a longitudinal analysis of a variety of proteomic markers. They show that similar to common clinical laboratory measures, many of these markers are homeostatically regulated, and that deviations from this regulation are associated with negative outcomes. This work is quite interesting, and potentially high value. I have a few major comments/concerns that I'd like to see addressed though:
1. Some aspects of the methods are quite unclear. Two examples: 'free of major chronic diseases' (referring to the cohort) is not clearly defined. Was this done by screening? exclusions based on ICD codes? Or something else. Similarly, the process for 'visit-specific batch correction' is not clearly described - which is a key component.
2. The authors show that proteomic deviations are associated with negative outcomes. How much of this is just recapitulating signals that would exist in more easily measured clinical values? I.e., do patients who show proteomic deviations also show deviations in more easily measured clinical markers (blood counts, metabolic + renal markers, etc.)? If making a clinical claim about risk, it is important to show that this is a novel aspect of patient risk. This is a relevant point for most of the results in Fig 5-6 - how much of this is novel risk stratification, versus a complex way of capturing already known risk, based on common markers?
3. The authors do not adjust for multiple comparisons in a clear way - this is important, given the extremely wide number of proteomic measures. Particularly for Fig 6
4. The PHI metric is an odd choice for capturing stability of measures. In laboratory medicine, stability of clinical markers is traditionally measured through the index of individuality - the ratio of average intra-patient variance in a marker, to inter-patient variance. The authors have four time points of data per patient, so they can easily calculate the IOI for each marker. It would be valuable to show that the PHI is strongly correlated with the IOI, as given PHI is a correlation-based metric, it is more sensitive to outliers, etc. (though I do note the authors have some outlier control - its just hard to interpret without seeing clear plots of the raw data underlying PHI calculation).
5. C-indices in Figure 5 are not particularly impressive. For context it would be good to present these against a baseline model, comprising age + sex - to see what signal is being added (ideally also against age + sex + basic lab measures, as a better control).

Minor:
- Why is Fig 6G a line plot? The data isn't linked in that way
- Providing some context for how large a 1-unit increase in proteomic age gap is, would be valuable. What % of patients had a 1unit increase?
- Given issues of multiple comparisons, how can the authors be sure the hazard ratios for individual proteomic markers are not just due to random variance?

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

The authors address a relevant and timely research question, and the manuscript is generally well written. However, it is currently not possible to assess the strength of the evidence because key information on the datasets, measurements, and statistical analyses is missing. In addition, some important results are lacking and Tables S1–S19 were missing from the submission.

1. The CMCS study includes over 27,000 participants, but the authors report using data from only 2,621 participants. A flow chart is needed to show step-by-step how participants were excluded from the original cohort to obtain the subpopulation used in this study. In addition, the authors should provide a comparison between included and excluded participants to allow readers to assess how representative this subcohort is of the original study population.

2. The true sample size for the analysis of the proteomic homeostasis index (PHI) should correspond to the number of participants with two or more proteomic assessments over time. I could not identify this number from the manuscript, and it seems unlikely that 2,621 corresponds to this group. Table 1 shows that the sample size was 2,621 at baseline, 1,056 at the first follow-up, 1,048 at the second follow-up, and 733 at the third follow-up. Proteomic profiling was performed on 5,458 samples, implying a mean of approximately two samples per participant. The analysis of personalized homeostatic failure used three consecutive proteomic measurements, suggesting that a sufficient number of participants had at least three repeated measurements. However, given the average of two repeats, the reported sample size of 2,621 likely includes individuals with only one proteomic assessment; such individuals cannot be included in analyses of stability. Please provide a figure or table clearly showing the distribution of participants by number of proteomic assessments and the visits from which these assessments originate.

3. In addition to CMCS, UK Biobank data were used. Please indicate in the captions of the figures which cohort the results are based on and provide the number of participants included in each analysis. In addition, the manuscript does not describe the UK Biobank population or its measurements – please add these.

4. The measurement of proteins remains insufficiently described. There is no information on how the EDTA plasma samples were handled and stored, although this is a critical issue for sample quality.

5. The authors note that “for each protein analysis, outliers were identified and removed using a ±3 standard deviation (SD) threshold from the mean for each protein at each visit” (lines 405–406). Please report the range of proportions of removed data to allow the reader to assess the extent of outliers in the proteomic data. In addition, there is little quality control information such as SomaScan normalization procedures, probe validation, or other QC metrics. Please add information on the quality of the proteomic measurements (=key exposure variables in this study). For example: what was the scale factor acceptance criterion per plate, how many plates passed these criteria, what was the calibrator percent-in-tails, what proportion exceeded the alert criterion, and how many plates passed the QC percent-in-tails acceptance criteria at each visit? This information is important because measurement error directly affects estimates of stability.

6. Minor issue: it is unclear when the samples were analysed.

7. Figure 5 shows associations of PHI with CHD, stroke, COPD, liver disease, renal failure, type 2 diabetes, and arthritis. However, the Methods section does not describe how these outcomes were ascertained or whether both prevalent and incident cases were included. I could not find even a list of these diseases in the Methods section. Please add sections describing how these outcomes were defined and measured.

8. The statistical analysis section lacks important information. Proteins were log-transformed—what base was used, and did this transformation adequately address skewness? In the analysis of outcomes, only the analysis of all-cause mortality is described. Information on how associations with chronic diseases were analysed is missing. The authors report using Cox proportional hazards models, but it is unclear whether proportional hazards assumptions were tested. Please report these tests for both mortality and chronic disease outcomes.

9. PHI was defined as the Pearson correlation of age- and sex-adjusted protein levels between two visits within the same individuals. Were the log-transformed proteins sufficiently normally distributed to justify the use of Pearson correlation? Would rank-based correlation coefficients be more appropriate for age- and sex-specific percentile data?

10. Missing data: no information is provided on the proportion of missing values in variables or how these were handled (e.g., imputation, exclusion, or other approaches).

11. More detail is needed on the ageing clock analyses. UK Biobank does not include SomaScan proteomics; its proteomic profiling is based on the Olink platform (approximately 3,000 proteins). Unlike SomaScan, this technology is not aptamer-based, and the correspondence between SomaScan and Olink protein measurements is not particularly high. The manuscript does not describe how measurements were harmonised across platforms.

12. Lines 472–473: “ProtAge20-replaced, where the three unstable proteins were substituted with three high PHI and age-associated proteins (WFDC2, AGER, CA14).” How were these three proteins selected? It is unclear by which method and from which pool of proteins the replacement proteins were chosen, and whether multiple testing was accounted for. These analyses need to be described and justified in greater detail.

13. Tables S1–S19 were missing from the submission, although they appear essential for evaluating the results. In addition, protein-specific results for intracellular, extracellular membrane, and secreted plasma proteins were not available. The

authors state that “analysis results for all 10,776 proteins are available in the supplementary materials” (lines 519–520), but I could not locate these results in the submitted files.

14. Statements on clinical implications should be tempered. The predictive performance indicated by the C-index in Figure 5 for death and chronic disease was low (<0.7) or very low (<0.6), suggesting limited clinical utility. In addition, SomaScan proteomics measure relative protein abundance rather than absolute concentrations (e.g., mmol/L) used in clinical practice; therefore, clinical thresholds cannot be directly derived from these semi-quantitative measurements. Consequently, translation into clinical practice is less direct than suggested in the manuscript.

15. Other textual points: the statement “This study provides the first lifespan-level, high-resolution atlas of the human blood proteome, spanning two decades and thousands of individuals… ranging from lifelong stability to pronounced volatility” (lines 289–290) appears overstated. The study does not provide lifespan-level results or evidence of lifelong stability; it covers the period from midlife to older age.

16. It would be worth noting as a limitation that the reproducibility/generalisability of these findings has not yet been confirmed in an independent cohort study.

In conclusion, this is a potentially interesting study. However, without clearer reporting of the data, methods, and results, the strength of the evidence cannot currently be assessed. I would be happy to review a revised version that addresses the points raised above.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
The authors revised manuscript is greatly improved from their original submission. I have no further concerns

(Remarks on code availability)
I did not have capacity to review the code in detail, but from a brief evaluation, it looks appropriate.

Reviewer #3

(Remarks to the Author)
The authors have carefully addressed all my comments. I have only two minor points:

Response 1: Extended Data Figure 1A. Please add to the second box the number of participants recruited from the Beijing communities.

Response 2: The authors note that 1,298 participants had two or more proteomic assessments, constituting the valid sample for the PHI analysis. Please clarify which analyses included data from the 1,323 participants with only one proteomic assessment (i.e. 2,621 – 1,298). If none, then the effective sample size appears to be 1,298 rather than the stated 2,621, and the manuscript should be revised accordingly.

Mika Kivimaki

(Remarks on code availability)

Point-to-Point Response

The 20-year longitudinal landscape of the human blood proteome: Stability, individual specificity, and clinical applications

Nature Health | Manuscript Number: NATHEALTH-A03049

To All Reviewers:

We sincerely thank all reviewers for the thoughtful and invaluable feedback, which has greatly helped us improve the quality of our manuscript. We also appreciate your recognition of the timeliness, novelty, and effectiveness of our approach. In the following, we aim to resolve the raised concerns point by point. A revised version of the manuscript has been uploaded, with all changes highlighted in red. We hope these revisions could address your concerns.

Response to Reviewer 2

Summary
In this study the authors present a longitudinal analysis of a variety of proteomic markers. They show that similar to common clinical laboratory measures, many of these markers are homeostatically regulated, and that deviations from this regulation are associated with negative outcomes. This work is quite interesting, and potentially high value. I have a few major comments/concerns that I'd like to see addressed though.

Response
We sincerely appreciate your constructive feedback and encouraging remarks on our manuscript. Your insightful comments and suggestions have been invaluable in refining our study, and we have carefully incorporated them into the revised manuscript, as detailed in the following point-to-point response.

Comment 1
Some aspects of the methods are quite unclear. Two examples: 'free of major chronic diseases' (referring to the cohort) is not clearly defined. Was this done by screening? exclusions based on ICD codes? Or something else. Similarly, the process for 'visit-specific batch correction' is not clearly described - which is a key component.

Response 1
We thank the reviewer for noting areas where methodological clarity could be improved. We first addressed the two explicitly raised concerns:

We refined the description of "free of major chronic diseases" and the related cohort design in Methods ("Study Population and Cohort Design"):

"At enrolment, trained medical professionals administered a standardized medical history questionnaire adapted from the WHO-MONICA protocol to ascertain pre-existing conditions. Participants were included in the CMCS if they were considered free of major chronic diseases, defined as no prior diagnosis or hospitalization for conditions including cardiovascular disease (myocardial infarction, stroke), diabetes, chronic obstructive pulmonary disease, chronic kidney disease, liver cirrhosis, and any malignancy.

For the present proteomics project, participants were selected from the parent CMCS cohort according to the following criteria: (i) availability of an archived blood sample from the 2002 visit, which served as the proteomic baseline, and (ii) complete follow-up information on cardiovascular and cerebrovascular health events and mortality from 2002 onward. This selection yielded 2,621 participants (aged 40-76 years in 2002). […] A step-by-step participant flow diagram is provided in **Extended data figure 1A**, and a comparison of baseline characteristics between the proteomics subcohort and the parent CMCS cohort is shown in **Extended data figure 1B**."

We elucidated the "visit-specific batch correction" in Methods ("Proteomics assays and data preprocess"):

"To address systematic technical variability or visit batch effects across the four time points likely arising from factors such as blood storage duration, we adopted the normalization strategy proposed by Oh et al.[5], who originally corrected for cohort-level batch effects by aligning age trajectories across different cohorts. Similarly, we treated each visit as a distinct batch and aimed to align the age-proteome trajectories across visits. Specifically, for each protein, we modeled log-transformed levels as a function of age and visit, with visit treated as a categorical variable, then subtracted the estimated visit-specific intercepts to remove technical offsets, thereby harmonizing the age-dependent trends across all visits without altering the underlying biological variance. As shown in **Figure S4**, we illustrate the efficacy of this approach using GDF15 and LPA, where the post-correction plots demonstrate that the previously divergent age slopes observed at different visits are now consistent across the entire longitudinal span, confirming successful removal of visit-specific technical artifacts while preserving the true biological aging signal."

Beyond these explicitly raised concerns, we have comprehensively enhanced the Methods section by adding detailed descriptions of several critical analytical steps. We have added a detailed description of the UK Biobank validation cohort, including its study design, participant characteristics, and the UKB-PPP sub-cohort. We have defined clinical outcomes by specifying their corresponding ICD-10 codes and describing how first-event dates were ascertained from integrated multi-source health

records, as well as the standardized procedure for excluding prevalent cases at baseline. We have comprehensively elaborated the complete proteomic data preprocessing pipeline, covering SomaScan intra-plate and inter-plate normalization, rigorous probe quality control thresholds.

We sincerely appreciate your constructive comments, which have significantly improved the clarity and rigor of our methodological descriptions and strengthened the overall quality of the manuscript.

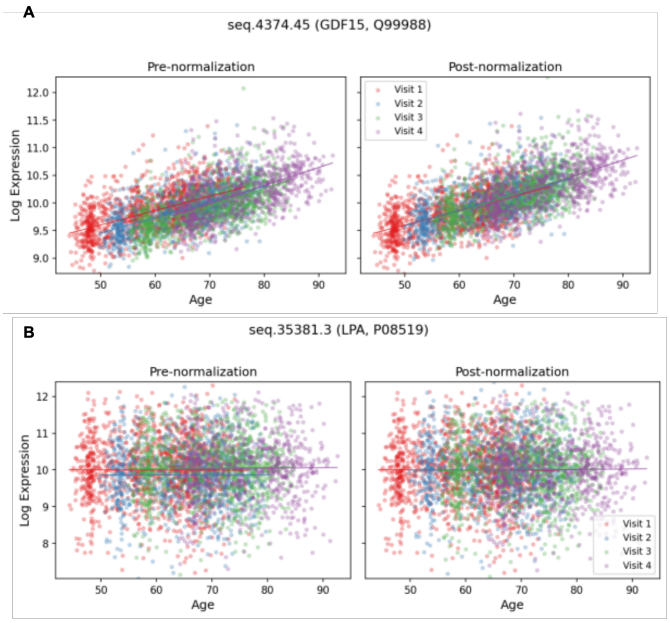


Figure S4. Visit-specific batch correction. Scatter plots show log-transformed protein levels versus age for GDF15 (A) and LPA (B) before (left) and after (right) correction. Colors denote the four visits (2002, 2007, 2012, 2022). After removing visit-specific effects via the linear model, the systematic offsets between time points are substantially reduced, revealing consistent age relationships across all visits.

Reference:

Oh, H.SH., Rutledge, J., Nachun, D. et al. Organ aging signatures in the plasma proteome track health and disease. Nature 624, 164–172 (2023).

Comment 2

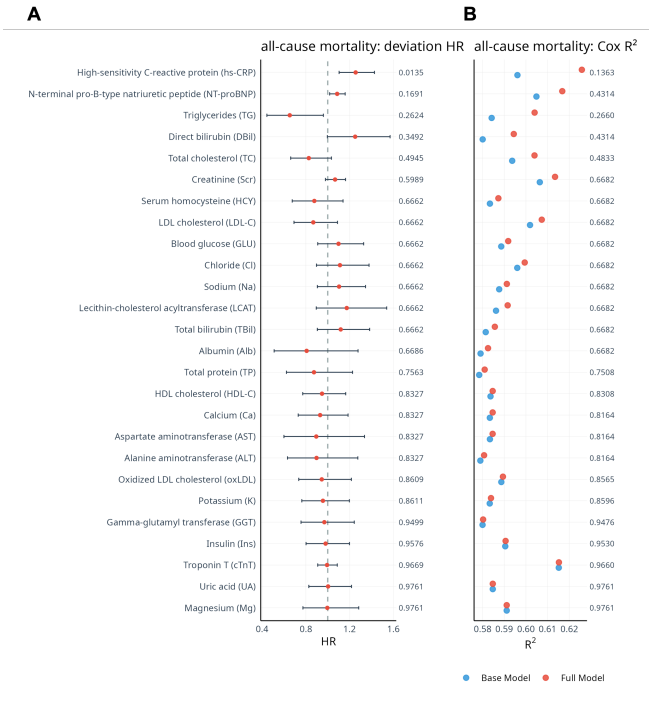
The authors show that proteomic deviations are associated with negative outcomes. How much of this is just recapitulating signals that would exist in more easily measured clinical values? I.e., do patients who show proteomic deviations also show deviations in more easily measured clinical markers (blood counts, metabolic + renal markers, etc.)? If making a clinical claim about risk, it is important to show that this is a novel aspect of patient risk. This is a relevant point for most of the results in Fig 5-6 - how much of this is novel risk stratification, versus a complex way of capturing already known risk, based on common markers?

Response 2

We thank the reviewer for this essential point regarding the novelty of the proteomic risk signal relative to standard clinical markers. To address it, we applied our personalized deviation framework to 26 longitudinally measured clinical laboratory markers (including renal, hepatic, metabolic, and hematological parameters) using the identical analytical pipeline employed for the proteomic markers. The results are presented in new Extended data figure 5, and we expanded the Results ("Personalized homeostatic failure" subsection) as follows:

"To assess whether this proteomic signal merely recapitulated information available from routine clinical laboratory tests, we applied the identical deviation framework to 26 longitudinally measured standard laboratory markers. **Only high-sensitivity C-reactive protein (hs-CRP) showed a significant association with mortality (BH-FDR adjusted $p < 0.05$)**; none of the remaining 25 markers (including renal, hepatic, metabolic, and hematological parameters) independently predicted mortality (BH-FDR adjusted $p > 0.05$; Extended data figure 5). Even for hs-CRP, the hazard ratio was modest, and this marker was measured in our cohort as part of a cardiovascular research assay, not as a component of a standard basic clinical chemistry panel."

These results demonstrate that the proteomic homeostatic deviation signals identified in our study can provide prognostic information that is not captured by routine clinical markers. We sincerely thank the reviewer for this constructive suggestion, which has helped us more clearly highlight the unique contribution of our work through this direct comparative analysis (Extended data figure 5).



Extended data figure 5. Prognostic performance of routine clinical marker

deviations for all-cause mortality. (A) Forest plot of hazard ratios (HRs) and 95% confidence intervals for associations between personalized deviations of 26 longitudinally measured routine clinical markers and all-cause mortality. HRs were estimated using Cox proportional hazards models adjusted for age and sex, with Benjamini–Hochberg false discovery rate (BH-FDR) correction applied across all 26 markers. The dashed vertical line denotes HR = 1 (null association). **(B)** Comparison of Cox model R^2 between the base model (blue; age and sex only) and the full model (red; base model plus the personalized deviation of each marker). P-values were derived from likelihood ratio tests (LRT). Only high-sensitivity C-reactive protein (hs-CRP) showed a significant independent association after BH-FDR correction; none of the remaining 25 routine clinical markers independently predicted mortality.

Comment 3

The authors do not adjust for multiple comparisons in a clear way - this is important, given the extremely wide number of proteomic measures. Particularly for Fig 6.

Response 3

We thank the reviewer for raising this important point. We have revised Figure 6 and all corresponding analyses to explicitly incorporate the Benjamini–Hochberg false discovery rate (BH-FDR) procedure. The Cox model p-values from the deviation analysis were corrected across all 571 proteins with PHI > 0.7 (the set tested a priori for individual-specific stability). After BH-FDR correction, 21 proteins retained a significant independent association with all-cause mortality (BH-FDR adjusted $p < 0.05$). These changes are described in Results ("Personalized homeostatic failure subsection") and Figure 6 Legend:

"To control for multiplicity, we applied the Benjamini-Hochberg False Discovery Rate adjusted p-values (BH-FDR adjusted p) across the 571 proteins. 21 proteins showed a significant independent association between their personalized deviation and all-cause mortality (BH-FDR adjusted $p < 0.05$; Figure 6F)."

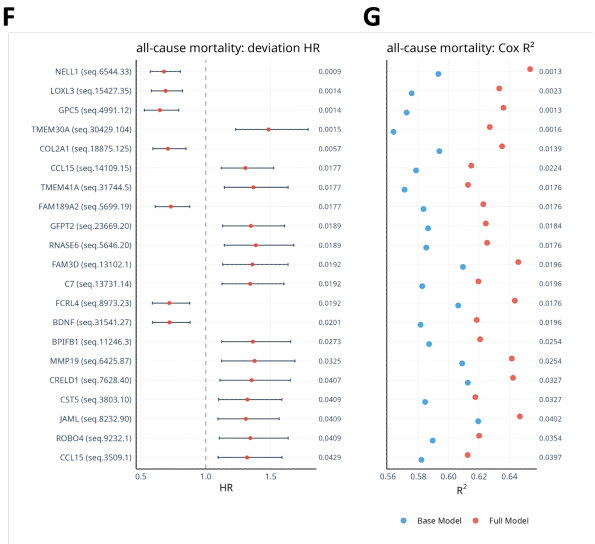


Figure 6. (F) Forest plot of hazard ratios for associations between personalized proteomic deviations and all-cause mortality from 2012 onward. The Benjamini–Hochberg false discovery rate (BH-FDR) correction was applied across all 571 proteins with PHI > 0.7; the 21 proteins shown are those that retained significance after correction (BH-FDR adjusted $p < 0.05$). (G) Comparison of Cox model R^2 for all-cause mortality between the base model (excluding the deviation term) and the full model (including the deviation term) for each of the 21 significant proteins. Improvement in model fit was assessed by likelihood ratio tests, with BH-FDR correction applied across all 571 proteins with PHI > 0.7.

Comment 4

The PHI metric is an odd choice for capturing stability of measures. In laboratory medicine, stability of clinical markers is traditionally measured through the index of individuality - the ratio of average intra-patient variance in a marker, to inter-patient variance. The authors have four time points of data per patient, so they can easily calculate the IOI for each marker. It would be valuable to show that the PHI is strongly correlated with the IOI, as given PHI is a correlation-based metric, it is more sensitive to outliers, etc. (though I do note the authors have some outlier control - its just hard to interpret without seeing clear plots of the raw data underlying PHI calculation).

Response 4

We thank the reviewer for this insightful and constructive suggestion. To address this point, we have calculated the Index of Individuality (IOI), the variance-based metric for marker stability in laboratory medicine, using all four time points of data per participant, and directly compared it with our pairwise correlation-based Protein Homeostatic Index (PHI).

The IOI is defined as the ratio of the average intra-individual variance to the inter-individual variance of a marker. By this definition, more stably regulated proteins exhibit lower IOI values, as intra-individual fluctuations are small relative to population-level differences. In contrast, higher PHI values indicate greater longitudinal stability of individual-specific protein levels.

As expected, we observed a strong inverse monotonic relationship between PHI and IOI across all 10,776 proteins (Spearman correlation coefficient = -0.865, $p < 0.001$; Figure S5A). This demonstrates that PHI, which only requires two repeated measurements per individual, captures essentially the same stability signal as IOI, which depends on three or more time points. Notably, both pairwise correlation-based approaches (Cohen et al., Nat Med 2021) and multi-time-point variance-based IOI methods (Foy et al., Nature 2025) have been previously used to quantify the stability of routine clinical markers in landmark studies, confirming the validity of both frameworks.

We have integrated these results and additional discussion into Discussion Section:

"PHI also provides a practical framework for assessing longitudinal biomarker stability in population cohorts with limited repeated sampling. Previous studies have characterized stable, individual-specific molecular or clinical traits through repeated-measurement concordance in longitudinal laboratory-test data (Cohen et al., 2021). Variance-based individuality frameworks, such as the index of individuality, have also been used to quantify the extent to which within-person variation is small relative to between-person variation (Foy et al., 2025). Consistent with these concepts, PHI showed strong concordance with both Spearman-based PHI and IOI-derived stability measures in our data (Supplementary Figs. 5 and 6), indicating that the pairwise PHI captures a stability signal consistent with multi-visit variance-based estimates. Beyond this concordance, the pairwise correlation-based approach underlying PHI only requires two repeated measurements per individual, whereas IOI calculation depends on three or more time points. This substantially lower sampling requirement makes PHI far more broadly applicable to the vast majority of population biobanks that have only sparse longitudinal sampling. Furthermore, the pairwise construction of PHI enables direct evaluation of stability across different time intervals (e.g., 5-year vs. 20-year), allowing for additional explorations of how protein stability changes."

We thank the reviewer for this constructive suggestion. The IOI comparison strengthened the methodological rigor and interpretability of our analysis by benchmarking PHI against an established variance-based stability metric. These added results clarify that PHI captures a highly consistent stability signal while retaining its practical advantage for sparsely sampled longitudinal cohorts, thereby improving the clarity and generalizability of our study.

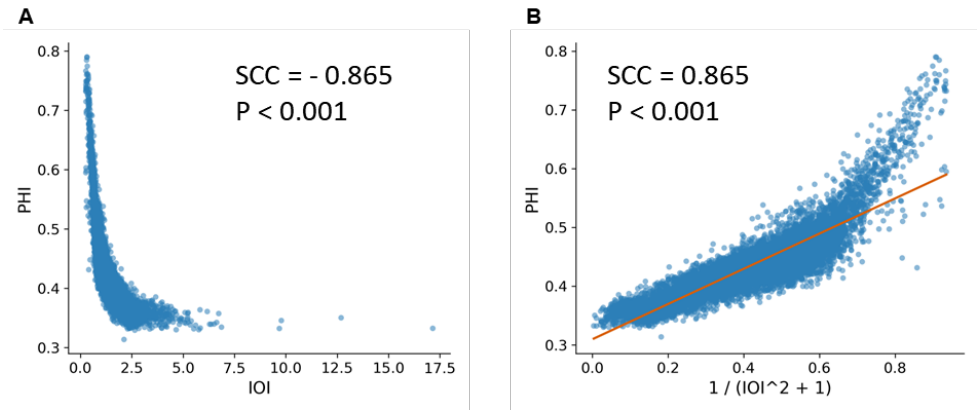


Figure S5. Comparison of the PHI and the index of individuality (IOI). (A) Scatter plot of the 20-year PHI (2002–2022) against raw IOI for all 10,776 assayed protein markers. The strong inverse monotonic relationship (SCC=-0.865, p<0.001) demonstrates that proteins with high PHI have low IOI, consistent with both metrics identifying proteins whose intra-individual variability is small relative to inter-individual differences. (B) Scatter plot showing the relationship between PHI calculated from

the 2002–2022 visit pair and the transformed IOI metric $1/(1+IOI^2)$ estimated using all four longitudinal visits. Each point represents one protein. The orange line indicates the linear regression fit. A strong positive association was observed across proteins (SCC=0.865, $p<0.001$), supporting the consistency between correlation-based PHI and variance-ratio-based IOI.

References:
Foy, B.H., Petherbridge, R., Roth, M.T. et al. Haematological setpoints are a stable and patient-specific deep phenotype. Nature 637, 430–438 (2025).

Cohen, N.M., Schwartzman, O., Jaschek, R. et al. Personalized lab test models to quantify disease potentials in healthy individuals. Nat Med 27, 1582–1591 (2021).

Comment 5
C-indices in Figure 5 are not particularly impressive. For context it would be good to present these against a baseline model, comprising age + sex - to see what signal is being added (ideally also against age + sex + basic lab measures, as a better control).

Response 5
Thank you for this helpful suggestion. We fully agree that benchmarking against a baseline model is critical to contextualize the incremental prognostic value of our proteomic aging signals.

Following your recommendation, we have revised Figure 5 and the corresponding Results section to add the requested age+sex baseline model. Our core revision logic is to explicitly test whether our proteomic age gaps add predictive value beyond basic demographic factors: for each of our PHI-optimized clock variants, we compared the C-index of the full model (age+sex+z-scored proteomic age gap) against the age+sex-only baseline, to clearly demonstrate the independent signal contributed by the proteomic age gap. This directly addresses the concern about the context of the C-index values, as it shows that even with age and sex already accounted for, our optimized proteomic age gap consistently improved predictive performance across all eight outcomes. These are described in updated Results and Figure legends:

"Prognostic performance was quantified using the C-index. For each clock variant, we evaluated whether adding the z-scored proteomic age gap to a baseline model comprising chronological age and sex improved discrimination. Across all eight outcomes, inclusion of the proteomic age gap consistently increased the C-index over the age+sex baseline, with the ProtAge20-replaced (M3) achieving the best overall performance (Figure 5A). Correspondingly, hazard ratios per 1‑SD increase in the age gap were highest for M3 across most outcomes (Figure 5B). These results show that incorporating PHI as an explicit criterion for marker selection enhances the robustness and clinical utility of sparse proteomic aging clocks, and

that blood protein-based biomarker temporal stability should be considered alongside cross-sectional association strength when designing biomarker panels."

Furthermore, we want to clarify that these proteomic aging clocks are designed to quantify aging. Because they are broad indicators of aging instead of targeted and direct disease predictive tools, their C-index values are naturally more modest. To remain tempered, avoid overstating the clinical implications of these models, and present a balanced view of our study's contributions, we have added a new limitation section:

"Limitations of clinical implications: First, the predictive performance of our modified proteomic aging clock (ProtAge20-replaced), as measured by C-index, is modest for individual clinical outcomes. It is critical to emphasize that ProtAge20-replaced was not designed as a direct clinical diagnostic or prognostic tool for specific diseases. Rather, it was developed as an improved quantitative measure of biological aging, and its association with age-related outcomes serves primarily to validate that it captures biologically meaningful aging signals more accurately than the original clock. The primary goal of this analysis was to demonstrate the principle that incorporating temporal stability (PHI) into biomarker panel design enhances the robustness of biological age estimation, not to develop a clinically deployable disease predictor. Second, the SomaScan 11K platform measures relative protein abundance rather than the absolute concentrations (e.g., mmol/L) used in routine clinical practice. While relative quantification is sufficient for evaluating longitudinal protein stability, comparing inter-individual differences, and performing population-level risk stratification, it cannot directly generate standardized clinical thresholds that are required for point-of-care decision-making. Translation of these findings into clinical practice will require the development of independent, absolute-quantification assays for the most promising PHI-informed biomarker panels, as well as rigorous clinical validation in diverse patient populations. Taken together, the clinical implications of this work lie not in providing immediately actionable clinical tools, but in establishing temporal stability (quantified by PHI) as a fundamental biological property that should be explicitly considered alongside effect size and disease association strength when designing blood-based proteomic biomarkers for precision medicine. This principle applies broadly to all biomarker development efforts, regardless of the proteomic platform used."

We are deeply grateful to the reviewer for highlighting this important distinction. By updating our analytical presentation to include the baseline model and tempering our discussion on clinical translation, the manuscript is now significantly more rigorous, balanced, and objective.

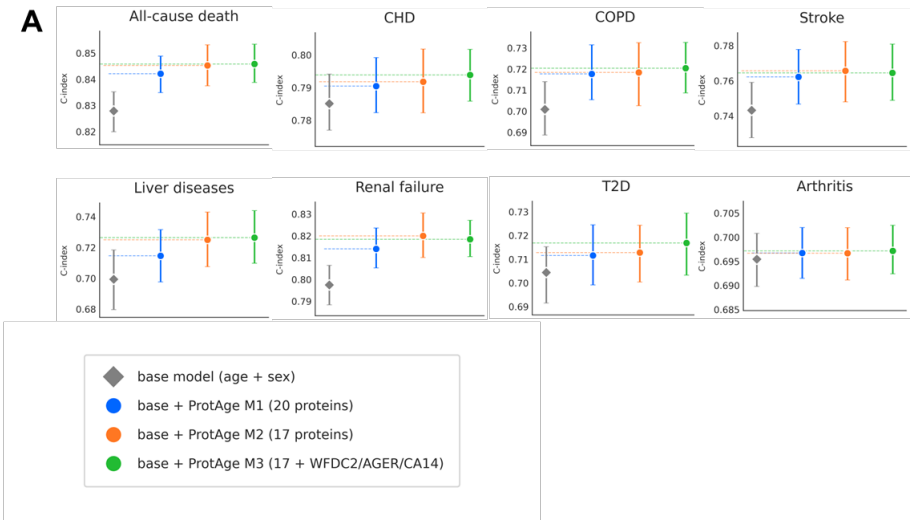


Figure 5. Optimization of proteomic aging clock by selecting proteins with high PHI. (A) C-indices for incident age-related outcomes. Four models are compared: age + sex only (base model); age + sex + proteomic age gap from the original ProtAge20 clock (M1); age + sex + proteomic age gap from ProtAge20-removed (M2, excluded the three low-PHI markers); and age + sex + proteomic age gap from ProtAge20-replaced (M3, replaced by three high-stability, age-associated proteins). All models were evaluated using Cox proportional hazards regression. Error bars denote 95% confidence intervals.

Minor:

Comment 6

Why is Fig 6G a line plot? The data isn't linked in that way.

Response 6

We apologize for the confusion. Each protein analysis is independent and not linked. We have removed the lines in the revised Figure 6G.

Comment 7

Providing some context for how large a 1-unit increase in proteomic age gap is, would be valuable. What % of patients had a 1unit increase?

Response 7

We thank the reviewer for this helpful suggestion to contextualize the effect size of the proteomic age gap, which will greatly improve the interpretability of our results. To clarify: we z-scored the proteomic age gap values across the entire cohort prior to analysis. Thus, the 1-unit increase in age gap referenced in our hazard ratio estimates corresponds to a 1-standard-deviation difference in this measure. Accordingly, approximately 16% of participants had an age gap of at least 1 unit, illustrating that this effect size corresponds to a prevalent, interpretable difference between individuals. We have added this clarification to Figure 5 legend:

"As the age gaps were z-scored, a 1-unit increase corresponds to 1 standard

deviation above the population mean; approximately 16% of participants had an age gap ≥1 unit."

Comment 8
Given issues of multiple comparisons, how can the authors be sure the hazard ratios for individual proteomic markers are not just due to random variance?

Response 8
We sincerely thank the reviewer for raising this critical statistical concern, which is essential for the reliability of our study findings. Regarding the risk of type I error inflation and false-positive results caused by multiple comparisons, we have presented a multi-dimensional, full-process strict quality control and correction strategy during the study design and statistical analysis stages, which systematically eliminates the interference of random variance on the hazard ratio results of individual proteomic markers. The details are as follows:

For the association analysis between personalized homeostatic baseline deviation and all-cause mortality, we included a total of 571 proteins with high Protein Homeostatic Index (PHI, PHI>0.7) for testing, which meant 571 independent Cox regression analyses were performed. To control the false-positive risk caused by multiple comparisons, we pre-specified the Benjamini-Hochberg False Discovery Rate (BH-FDR) correction method, which is the gold standard for controlling multiple comparison bias in large-scale omics research, and has a better balance between test power and false-positive control than the traditional Bonferroni correction. After BH-FDR correction, 21 proteins retained a significant independent association between their personalized deviation and all-cause mortality (BH-FDR adjusted $p < 0.05$, see the subsection "*Personalized homeostatic failure"* in the Results section, Figure 6F and Supplementary Table S19 of the manuscript). This step controls the false-positive results caused by random variance from the source of the association analysis. We further verified the proportional hazards assumption (the core premise of the Cox proportional hazards regression model) using global Schoenfeld residual tests. For each Cox model demonstrating a significant protein-outcome association, the global Schoenfeld test p-value was > 0.05. This further confirms that the model fitting results were robust and reliable and excludes false-positive associations caused by model specification bias or random fluctuations.

Response to Reviewer 3

Summary
The authors address a relevant and timely research question, and the manuscript is generally well written. However, it is currently not possible to assess the strength of the evidence because key information on the datasets, measurements, and statistical analyses is missing. In addition, some important results are lacking and Tables S1–S19 were missing from the submission.

Response

We appreciate your recognition of our work and your comprehensive constructive feedback. We have taken your concerns very seriously and systematically supplemented all missing details to strengthen the evidence base of our findings and improve reproducibility. We have also resolved the issue of missing supplementary tables with this revision.

Comment 1

The CMCS study includes over 27,000 participants, but the authors report using data from only 2,621 participants. A flow chart is needed to show step-by-step how participants were excluded from the original cohort to obtain the subpopulation used in this study. In addition, the authors should provide a comparison between included and excluded participants to allow readers to assess how representative this subcohort is of the original study population.

Response 1

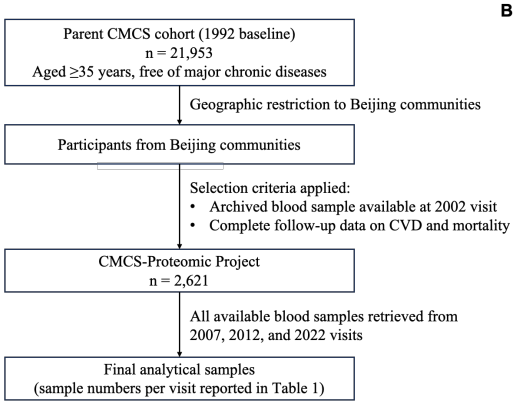
We thank the reviewer for this critical suggestion to enhance the transparency and interpretability of our study design. This proteomic sub-study, CMCS Proteomics Project, is a nested cohort derived from the parent Chinese Multi-provincial Cohort Study (CMCS). The parent CMCS was initiated in 1992 and enrolled 21,953 participants aged >=35 years who were free of major chronic diseases at baseline. During the longitudinal follow-up, a subset of participants from communities in Beijing provided blood samples at the 2002, 2007, 2012, and 2022 visits; these samples were archived and stored long-term in a dedicated biobank.

In 2024–2025, when we designed the current proteomics project, we applied the following selection criteria to define the analytical cohort. We selected participants who (i) had an available archived blood sample from the 2002 visit that could serve as the proteomic baseline, and (ii) had complete follow-up information on cardiovascular and cerebrovascular health events and mortality from 2002 onward, consistent with the original purpose of the CMCS to establish an internationally compatible research model for cardiovascular disease epidemiology in China. This systematic selection procedure yielded 2,621 participants. For all these individuals, we additionally retrieved and subjected to proteomic profiling any further remaining blood samples that were available from the 2007, 2012, and 2022 follow-up visits. We have now added a detailed participant flow diagram (Extended data figure 1A, Methods subsection "Study Population and Cohort Design") to show this step-by-step how participants were selected from the original cohort to obtain the subpopulation used in this study.

To assess the representativeness of the CMCS-PP subcohort, we have conducted a comprehensive comparison of key baseline characteristics between included participants (n=2,621) and the entire parent CMCS cohort (n=21,953). No significant

differences were observed in most factors between the CMCS-PP subcohort and the parent CMCS cohort, indicating that our analytical population is generally representative of the original study population. The full comparison table has been added as Extended data figure 1B (Methods subsection "Study Population and Cohort Design") in the revised manuscript.

A



B

Characteristic	Original CMCS (n=21,953)	CMCS-PP (n=2,621)
Age at 1992 survey, years	48.3 ± 8.9	47.4 ± 8.1
Males, n (%)	11,366 (51.8)	1,171 (44.7)
Smoking, n (%)	6,313 (28.8)	437 (16.7)
Systolic BP, mmHg	123.9 ± 20.4	121.7 ± 17.6
Diastolic BP, mmHg	80.2 ± 12.0	78.1 ± 10.8
LDL-C, mmol/L	2.7 ± 1.0	2.7 ± 0.8
HDL-C, mmol/L	1.4 ± 0.4	1.4 ± 0.4
Total cholesterol, mmol/L	4.8 ± 1.0	4.7 ± 0.9
Triglyceride, mmol/L	1.2 [0.9, 1.6]	1.0 [0.7, 1.5]
Fasting blood glucose, mmol/L	5.2 ± 1.4	5.4 ± 0.9
BMI, kg/m²	24.2 ± 3.3	24.0 ± 3.2

Extended data figure 1. Participant selection and baseline characteristics comparison for the CMCS Proteomics Project. A, Participant flow diagram showing the derivation of the CMCS Proteomics Project analytical cohort from the parent Chinese Multi-provincial Cohort Study (CMCS). The parent CMCS enrolled 21,953 participants aged ≥35 years free of major chronic diseases at baseline in 1992. During longitudinal follow-up, a subset of participants from Beijing communities provided blood samples at the 2002, 2007, 2012, and 2022 visits, which were archived in a dedicated biobank. For the current proteomics project (designed 2024–2025), participants were selected based on: (i) availability of archived blood sample from the 2002 visit (proteomic baseline), and (ii) complete follow-up information on cardiovascular health events and mortality from 2002 onward. This systematic selection yielded 2,621 participants for proteomic profiling. The diagram shows step-by-step exclusions with numbers and reasons at each stage. B, Baseline characteristics comparison between CMCS Proteomics Project participants (n=2,621) and the original CMCS cohort (n=21,953). Data are presented as mean ± standard deviation for continuous variables.

Comment 2

The true sample size for the analysis of the proteomic homeostasis index (PHI) should correspond to the number of participants with two or more proteomic assessments over time. I could not identify this number from the manuscript, and it seems unlikely that 2,621 corresponds to this group. Table 1 shows that the sample size was 2,621 at baseline, 1,056 at the first follow-up, 1,048 at the second follow-up, and 733 at the third follow-up. Proteomic profiling was performed on 5,458 samples, implying a mean of approximately two samples per participant. The analysis of personalized homeostatic failure used three consecutive proteomic measurements, suggesting that a sufficient number of participants had at least three repeated measurements. However, given the average of two repeats, the reported sample size

of 2,621 likely includes individuals with only one proteomic assessment; such individuals cannot be included in analyses of stability. Please provide a figure or table clearly showing the distribution of participants by number of proteomic assessments and the visits from which these assessments originate.

Response 2

We thank the reviewer for this important clarification regarding the longitudinal analytical samples. We have added new table (Table 2) and expanded the Methods (subsection: "Study Population and Cohort Design") to explicitly details the number of participants stratified by their pattern of proteomic assessments across the four study visits (2002, 2007, 2012, 2022):

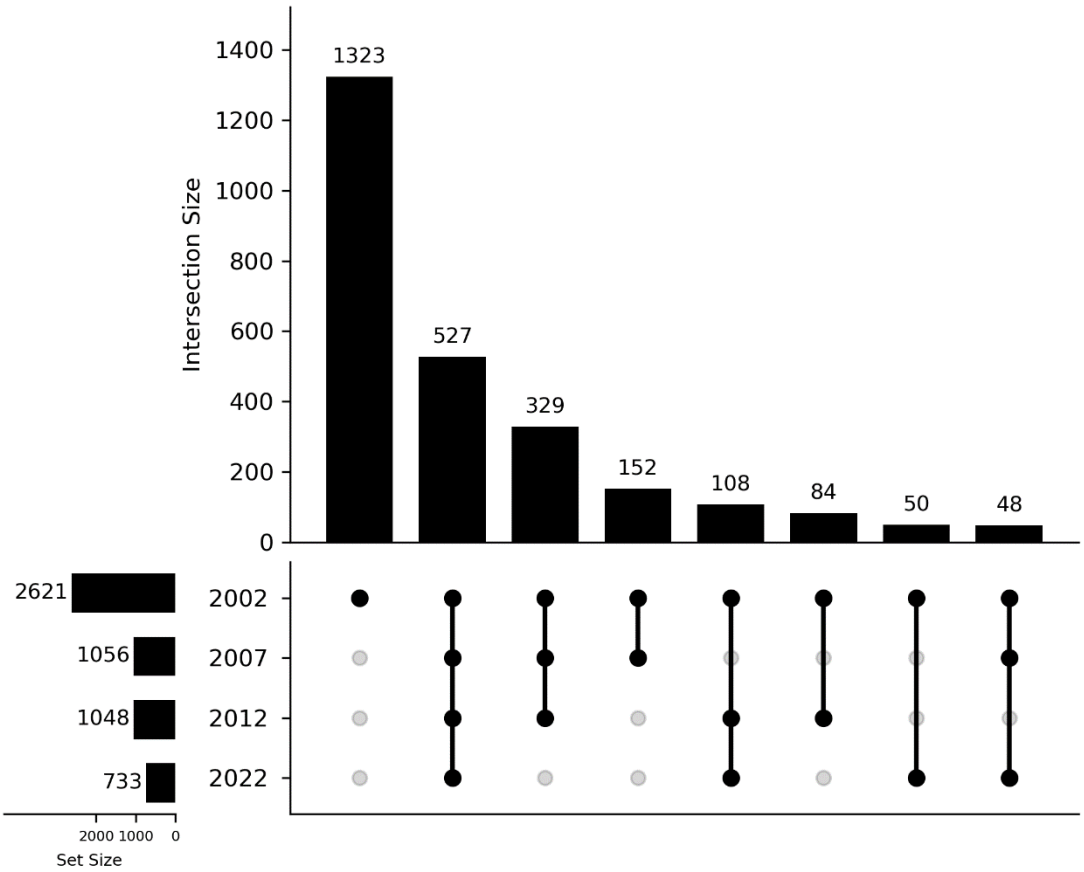
"The distribution of participants by number and timing of proteomic assessments is detailed in Table 2. Of the 2,621 participants, 1,298 participants had two or more proteomic assessments; this group constitutes the valid sample for the Protein Homeostatic Index (PHI) analysis. 1,323 provided proteomic data at baseline (2002) only and were therefore excluded from longitudinal stability analyses. For the personalized homeostatic failure analysis, which required three consecutive measurements (2002, 2007, 2012), 856 participants were eligible (329 with data from exactly these three visits, and 527 with data from all four visits)."

We also revised the introduction to explicitly clarify the number of participants involved in the stability calculation: "we performed a large-scale, high-coverage longitudinal proteomic study of 2,621 participants, each sampled at four time points (2002, 2007, 2012, 2022), profiling 10,776 blood protein markers using the SomaScan 11K platform (Fig. 1A, Table 1). Of these, 1,298 participants had two or more proteomic assessments and were included in analyses quantifying the temporal stability of protein levels (Table 2). "

# of protein profile	2002	2007	2012	2022	Total
1 measurement	2621				2621
2 measurements		1056	192	50	1298
3 measurements			859	156	1012
4 measurements				527	527
Total	2621	1056	1048	733	5458

Table 2. Distribution of participants by number and timing of proteomic assessments. Rows indicate the number of proteomic assessments a participant contributed; columns correspond to the year of their last assessment. Each cell reports the number of participants who had that specific number of assessments with the final measurement occurring in the given year. For example, 1,056 participants had exactly two measurements, and their last sample was collected in 2007; the 192 participants with last two

measurements falling in 2012 are shown in the next column. The four-assessment group (n=527) completed all visits (2002, 2007, 2012, 2022).



Extended data figure 6. Sample intersection and follow-up frequency. Upset plot displays the number of participants who provided blood samples at various combinations of the four visits. The vertical bars represent the intersection size (number of participants), while the connected dots below indicate the specific visits included in each intersection. A total of 527 participants provided samples at all four visits, with varying degrees of overlap across the 20-year study period.

Comment 3
In addition to CMCS, UK Biobank data were used. Please indicate in the captions of the figures which cohort the results are based on and provide the number of participants included in each analysis. In addition, the manuscript does not describe the UK Biobank population or its measurements - please add these.

Response 3

We greatly appreciate your constructive comments, which have helped us further clarify the cohort information and relevant details of our analyses. For the adjustments to the figure captions, we have fully revised the caption of Figure 5 as follows:

"All age-related outcome analyses were performed in the UK Biobank Pharma Proteomics Project (UKB-PPP) cohort (n=45,441), with exclusion of prevalent cases for each disease-specific outcome. Outcomes include all-cause mortality, coronary heart disease (CHD), stroke, chronic obstructive pulmonary disease (COPD), liver diseases, renal failure, type 2 diabetes (T2D), and arthritis."

For the description of the UK Biobank population and its measurements, we have added a new subsection titled 'UK Biobank validation cohort' in the Methods section to fully supplement this missing information. This subsection comprehensively covers the basic demographic profile of the UK Biobank cohort, its multi-stage follow-up design, comprehensive phenotypic data resources, as well as the detailed plasma proteomic measurement platforms and data characteristics. The relevant content of this newly added subsection is as follows:

"The UK Biobank (UKB) is a population-based prospective cohort study that recruited ~500,000 participants aged 40-69 years across the United Kingdom between 2006 and 2010. This dataset contains detailed personal information, medical records, and comprehensive phenotypic data covering demographics, lifestyle factors, physiological and biochemical indicators, imaging results, and long-term follow-up health outcomes. A subset of participants attended subsequent visits: ~20,337 completed a first repeat assessment (2012–2013), ~85,000 attended a first imaging visit (from 2014, ongoing), and ~9,000 attended a second imaging visit (from 2019, ongoing). All participants provided informed consent. Detailed phenotypic information is available at https://biobank.ndph.ox.ac.uk/showcase/. Plasma proteomic data were generated by the UKB Pharma Proteomics Project (UKB-PPP) using the high-throughput affinity-based Olink Explore 3072 platform, which simultaneously quantifies 2,923 proteins in plasma samples. Data were available for ~46,000 randomly selected samples. Additionally, proteomic data were available for ~1,200 participants from the COVID-19 repeat imaging study. For these participants, proteomic profiling was performed using an earlier version of the Olink platform targeting ~1,500 proteins, with measurements collected at three time points: baseline, the first imaging visit (2014 onwards), and the second imaging visit (2019 onwards)."

All the above revisions have been fully implemented in the revised manuscript. We sincerely thank the reviewer for these valuable suggestions, which not only clarifies data sources and analytical details, but also standardizes the presentation of research findings, and further strengthened the persuasiveness and logical rigor of this work.

Comment 4

The measurement of proteins remains insufficiently described. There is no information on how the EDTA plasma samples were handled and stored, although this is a critical issue for sample quality.

Response 4

We thank the reviewer for highlighting this omission. We have now expanded the description of sample handling and storage in the revised Methods section. Venous blood samples were drawn from the antecubital vein after an overnight fast (≥8 h) and collected in the lavender-top tube(s) containing Ethylenediaminetetraacetic acid as anticoagulant in the morning after fasting for at least 8 hours. Aliquots were stored at -80°C for subsequent analysis. Of these, 50 μl serum aliquots were prepared and stored on dry ice for shipment to Accuramed Inc. Lab (Shanghai, China) for protein quantification using the SomaScan® 11K Assay certified by SomaLogic Operating Co., Inc (Colorado, United States, https://somalogic.com/).

Comment 5

The authors note that “for each protein analysis, outliers were identified and removed using a ±3 standard deviation (SD) threshold from the mean for each protein at each visit” (lines 405–406). Please report the range of proportions of removed data to allow the reader to assess the extent of outliers in the proteomic data. In addition, there is little quality control information such as SomaScan normalization procedures, probe validation, or other QC metrics. Please add information on the quality of the proteomic measurements (=key exposure variables in this study). For example: what was the scale factor acceptance criterion per plate, how many plates passed these criteria, what was the calibrator percent-in-tails, what proportion exceeded the alert criterion, and how many plates passed the QC percent-in-tails acceptance criteria at each visit? This information is important because measurement error directly affects estimates of stability.

Response 5

We appreciate the reviewer’s attention to data quality control. Below we provide the details on outlier removal and SomaScan quality control metrics.

The range of proportions of removed data: the median proportion of removed data was 1.2%, and the maximum across all proteins was 3.9% (Figure S3A), indicating that extreme outliers were rare.

All SomaScan assays strictly followed the manufacturer’s standard operating procedures. At the plate level, intra- and inter-plate variation was first corrected using built-in calibrator replicates, and data were normalized to a predefined population reference via adaptive maximum likelihood calibration. Principal component analysis of all study samples, calibrator controls, technical QC replicates, and buffer blanks revealed tight, completely separated clusters with no detectable separation by plate or longitudinal visit (Figure S3B). Every assay plate across all visits satisfied the scale factor acceptance criterion (each plate's median signal ratio falling within the 0.8 - 1.2 range), confirming negligible plate-to-plate technical variation and the absence of confounding batch effects. Probe-level quality control was performed according to SomaLogic's official definitions. The Calibrator Percent-In-Tails (the

proportion of probes with a calibrator ratio outside 0.6 - 1.4) was 0.34%, far below the ≤10% acceptance threshold. The overall QC Percent-In-Tails (the proportion outside 0.8–1.2) was 2.53%, drastically better than the ≤15% requirement, and all plates at every visit passed this QC percent-in-tails criterion. Correspondingly, 97.47% of the 10,776 human protein probes passed the ColCheck alert; the 2.53% of flagged probes exactly match the QC percent-in-tails. To ensure full transparency, we retained all probes and provide their complete QC status and flags in Table S20. The median coefficient of variation computed from technical replicates across all plates was 4.10% (IQR 3.21% - 5.50%), well exceeding the manufacturer's specification that>90% of analytes have a CV <20%, demonstrating excellent measurement reproducibility.

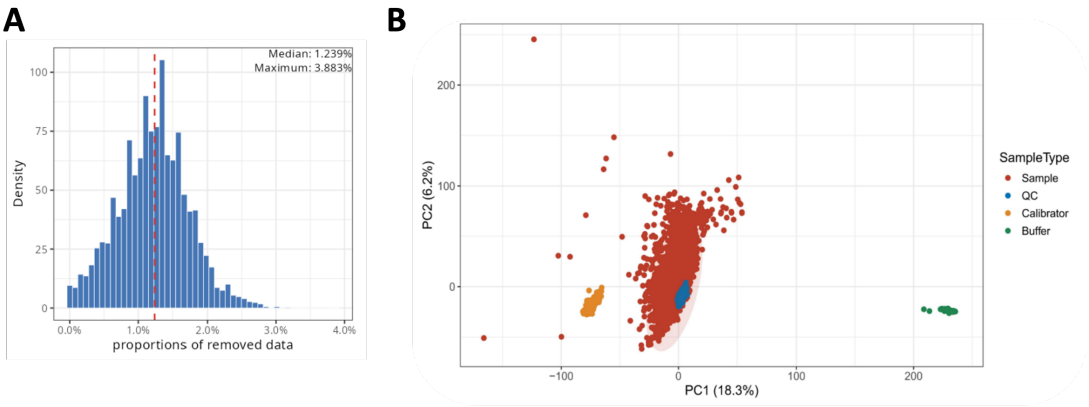


Figure S3. Distribution of outlier removal proportions and principal component analysis of SomaScan assay samples. A. For each of the 10,776 protein markers at each longitudinal visit, outliers were defined as values exceeding ±3 standard deviations from the mean. The histogram shows the distribution of the per-protein proportion of removed data points (median = 1.2%; maximum = 3.9%). n = 10,776 proteins. B. Principal component analysis (PCA) of study samples, calibrator controls, technical QC replicates, and buffer blanks.

Comment 6
Minor issue: it is unclear when the samples were analysed.

Response 6
We apologize for the lack of clarity and have now explicitly stated in Methods ("Proteomics assays and data preprocess") that all 5,458 samples were analysed from April 2024 to February 2025.

Comment 7
Figure 5 shows associations of PHI with CHD, stroke, COPD, liver disease, renal failure, type 2 diabetes, and arthritis. However, the Methods section does not describe how these outcomes were ascertained or whether both prevalent and incident cases were included. I could not find even a list of these diseases in the

Methods section. Please add sections describing how these outcomes were defined and measured.

Response 7

We thank the reviewer for this essential point regarding the clarity of our outcome definitions and case ascertainment for the disease association analyses presented in Figure 5. To address it, we supplemented a Methods section named "Ascertainment of Outcomes", clarifying that we used UK Biobank first-occurrence data (Category 1712) for outcome ascertainment, excluded prevalent cases and only included incident cases and event-free censored participants, and added the complete list of the 7 diseases with their corresponding ICD-10 codes. We extended the following Methods section to ensure that all steps of our outcome assessment are fully documented and accessible to readers:

"All-cause mortality was ascertained using the linked national death registry data (UK Biobank field 40000). All disease outcomes were ascertained using UK Biobank first-occurrence data (Category 1712), which maps health records from primary care, hospital inpatient, death registry, and self-reported assessments to standardized ICD-10 codes and captures the earliest recorded date of each disease across these sources. Specifically, the included diseases and their corresponding ICD-10 codes were as follows: coronary heart disease (I20–I25), stroke (G45, I60–I64), chronic obstructive pulmonary disease (J40–J44, J47), liver diseases (C22, E83, E88, I85, K70–K76), renal failure (N17–N19), type 2 diabetes (E11), and arthritis (M05–M08, M10–M13). In Cox proportional hazards models examining associations between biological age gap and health outcomes, prevalent cases were excluded for each outcome independently. Specifically, participants diagnosed with the corresponding disease at or before baseline (defined as event time ≤ baseline age, whereby the duration would be ≤ 0) were removed from that specific outcome's analysis. Only incident cases (who developed the outcome after baseline) were included."

Comment 8

The statistical analysis section lacks important information. Proteins were log-transformed—what base was used, and did this transformation adequately address skewness? In the analysis of outcomes, only the analysis of all-cause mortality is described. Information on how associations with chronic diseases were analysed is missing. The authors report using Cox proportional hazards models, but it is unclear whether proportional hazards assumptions were tested. Please report these tests for both mortality and chronic disease outcomes.

Response 8

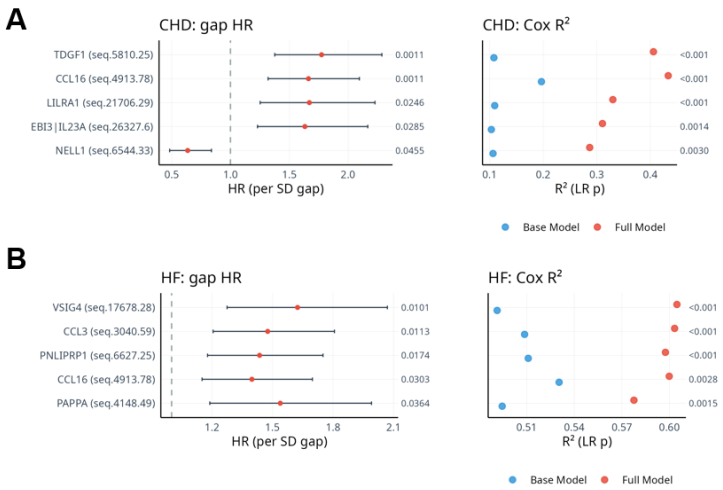
We thank the reviewer for these constructive suggestions. We address each point below.

Raw protein abundance values were natural log-transformed (base e), as the assay's raw data follow an expected log-normal distribution. Downstream, we applied age-

and sex-specific percentile transformation for PHI calculation and z-score standardization for Cox regression analyses, ensuring appropriate distributional properties for each modeling step.

We apologize for the incomplete description of chronic disease analyses in the original submission. We have now fully expanded this section in the Results (subsection "Personalized homeostatic failure") and added corresponding quantitative results in Extended data figure 4 and Table S19.We extended our analysis to two major chronic cardiovascular outcomes: incident coronary heart disease (CHD) and heart failure (HF), which are the primary prespecified outcomes of the CMCS cohort design. For each outcome, we excluded individuals with a prevalent diagnosis before the 2012 baseline visit and used identical Cox proportional hazards models as those employed for all-cause mortality. For incident CHD: We identified 5 proteins whose personalized deviations were significantly associated with first-event CHD after BH-FDR correction (TDGF1, CCL16, LILRA1, EBI3|IL23A, NELL1; all adjusted $p < 0.05$; Extended data figure 4A). Inclusion of the deviation term significantly improved model fit for all 5 proteins compared to the base model (all likelihood ratio test $p < 0.05$; Extended data figure 4A). For incident HF: We identified 5 proteins with significant associations with first-event HF (VSIG4, CCL3, PNLIPRP1, CCL16, PAPPA; all adjusted $p < 0.05$; Extended data figure 4B, left). Again, all models showed significant improvement in fit upon addition of the deviation term (all likelihood ratio test $p < 0.05$; Extended data figure 4B, right). CCL16 emerged as the only protein shared between CHD and HF, while NELL1 was shared between all-cause mortality and CHD.

We acknowledge the importance of validating the proportional hazards assumption for Cox models. For all Cox models fitted across the three outcomes (all-cause mortality, CHD, HF), we tested the proportional hazards assumption using global Schoenfeld residual tests. For each Cox model demonstrating a significant protein-outcome association, the global Schoenfeld test p-value was > 0.05. This confirms that the proportional hazards assumption was well satisfied across the reported cox models.



Extended data figure 4. Personalized proteomic deviations and incident CHD and HF. Analyses include 571 proteins with Protein Homeostatic Index (PHI) > 0.7. Personalized proteomic deviation is defined as the residual between observed 2012 protein levels and levels predicted from individual-specific 2002 and 2007 measurements. A, Left: Forest plot of hazard ratios (HRs) and 95% confidence intervals (CIs) for incident coronary heart disease (CHD) per 1 standard deviation (SD) increase in personalized deviation. Analyses include 856 individuals free of CHD at 2012 baseline. Cox models were adjusted for age, sex, 2002/2007 protein levels, and predicted 2012 levels. P-values (right) are Benjamini-Hochberg false discovery rate (BH-FDR) corrected. Right: Cox R^2 comparison between base models (blue circles, no deviation term) and full models (red circles, with deviation term). P-values are from likelihood ratio (LR) tests. B, Left: Forest plot of HRs and 95% CIs for incident heart failure (HF) per 1 SD increase in personalized deviation. Analyses include 856 individuals free of HF at 2012 baseline. Adjustments and multiple testing correction are identical to CHD. Right: Cox R^2 comparison between base and full models for HF outcomes. P-values are from LR tests.

Comment 9

PHI was defined as the Pearson correlation of age- and sex-adjusted protein levels between two visits within the same individuals. Were the log-transformed proteins sufficiently normally distributed to justify the use of Pearson correlation? Would rank-based correlation coefficients be more appropriate for age- and sex-specific percentile data?

Response 9

Thank you for this excellent and thoughtful methodological comment. We fully share your concern about the appropriateness of correlation metrics for percentile-normalized data and have conducted comprehensive validation analyses to address this question rigorously. Our results demonstrate that our PHI calculation is methodologically robust and that our core findings are entirely independent of the choice of correlation metric.

First, as you suggested, we recalculated PHI using the Spearman rank correlation coefficient (SCC). Strikingly, the Pearson-based and Spearman-based PHI estimates were virtually identical across all 10,776 proteins (SCC=0.999; Supplementary Figure 6). This near-perfect concordance is not coincidental: the age- and sex-specific percentile normalization we applied transforms protein levels into relative ranks within each demographic stratum. For rank-transformed data, Pearson correlation and Spearman rank correlation are mathematically nearly equivalent, which explains the observed perfect alignment. This result directly confirms that our use of Pearson correlation on percentile-normalized data yields results indistinguishable from a rank-based approach.

To further validate the PHI framework using an entirely orthogonal method that does not rely on pairwise correlation at all, we also calculated the Index of Individuality (IOI), the variance-based metric for marker stability in laboratory medicine, using multiple time points of data per participant, and directly compared it with our pairwise correlation-based Protein Homeostatic Index (PHI). The IOI is defined as the ratio of the average intra-individual variance to the inter-individual variance of a marker. By this definition, more stably regulated proteins exhibit lower IOI values, as intra-individual fluctuations are small relative to population-level differences. In contrast, higher PHI values indicate greater longitudinal stability of individual-specific protein levels.

As expected, we observed a strong inverse monotonic relationship between PHI and IOI across all 10,776 proteins (Spearman correlation coefficient = -0.865, $p < 0.001$; Supplementary Fig. 5A). This demonstrates that PHI, which only requires two repeated measurements per individual, captures essentially the same stability signal as IOI, which depends on three or more time points. This provided another evidence that PHI captures the right underlying stability signal. Notably, both pairwise correlation-based approaches (Cohen et al., Nat Med 2021) and multi-time-point variance-based IOI methods (Foy et al., Nature 2025) have been previously used to quantify the stability of routine clinical markers in landmark studies.

We have integrated these important methodological validation results into Discussion section reads:

"PHI also provides a practical framework for assessing longitudinal biomarker stability in population cohorts with limited repeated sampling. Previous studies have characterized stable, individual-specific molecular or clinical traits through repeated-measurement concordance in longitudinal laboratory-test data (Cohen et al., 2021). Variance-based individuality frameworks, such as the index of individuality, have also been used to quantify the extent to which within-person variation is small relative to between-person variation (Foy et al., 2025). Consistent with these concepts, PHI showed strong concordance with both Spearman-based PHI and IOI-derived stability measures in our data (Supplementary Figs. 5 and 6), indicating that the pairwise PHI

captures a stability signal consistent with multi-visit variance-based estimates. Beyond this concordance, the pairwise correlation-based approach underlying PHI only requires two repeated measurements per individual, whereas IOI calculation depends on three or more time points. This substantially lower sampling requirement makes PHI far more broadly applicable to the vast majority of population biobanks that have only sparse longitudinal sampling. Furthermore, the pairwise construction of PHI enables direct evaluation of stability across different time intervals, allowing for additional explorations of how protein stability changes."

These revisions strengthen the methodological foundation of our work and provide clear evidence that our PHI framework is both statistically valid. We greatly appreciate the reviewer's insight that prompted these important additional analyses.

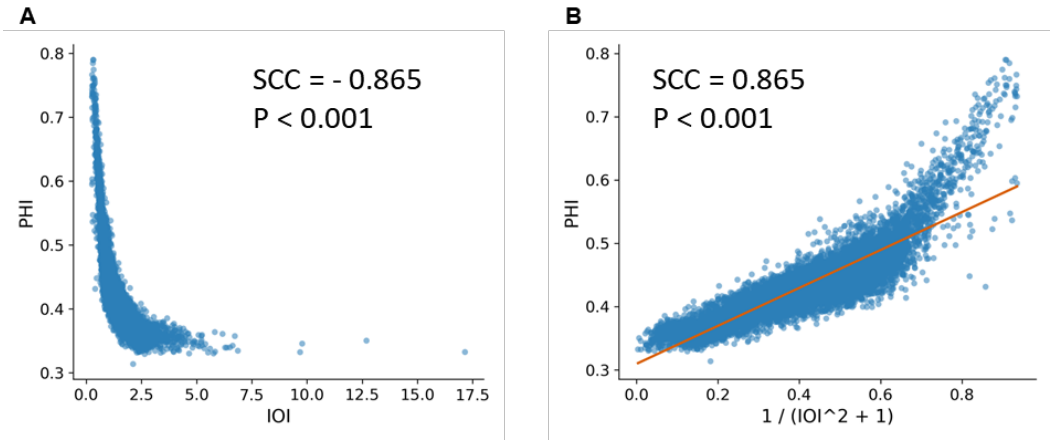


Figure S5. Comparison of the PHI and the index of individuality (IOI). (A) Scatter plot of the 20-year PHI (2002–2022) against raw IOI for all 10,776 assayed protein markers. The strong inverse monotonic relationship (SCC=-0.865, p<0.001) demonstrates that proteins with high PHI have low IOI, consistent with both metrics identifying proteins whose intra-individual variability is small relative to inter-individual differences. (B) Scatter plot showing the relationship between PHI calculated from the 2002–2022 visit pair and the transformed IOI metric $1/(1+IOI^2)$ estimated using all four longitudinal visits. Each point represents one protein. The orange line indicates the linear regression fit. A strong positive association was observed across proteins (SCC=0.865, p<0.001), supporting the consistency between correlation-based PHI and variance-ratio-based IOI.

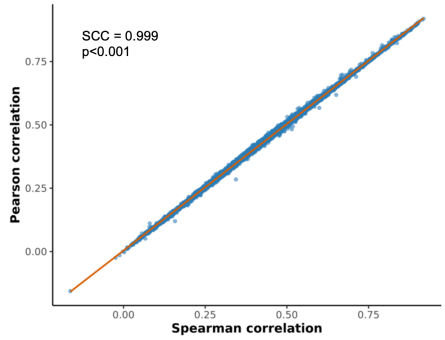


Figure S6. Concordance between Pearson- and Spearman-based PHI estimates. Each point represents one protein. Pearson- and Spearman-based PHI estimates were highly concordant across proteins (SCC=0.999, R^2 = 0.999, p<0.001), indicating that the results were robust to the choice of correlation metric.

References:

Foy, B.H., Petherbridge, R., Roth, M.T. et al. Haematological setpoints are a stable and patient-specific deep phenotype. Nature 637, 430–438 (2025).

Cohen, N.M., Schwartzman, O., Jaschek, R. et al. Personalized lab test models to quantify disease potentials in healthy individuals. Nat Med 27, 1582–1591 (2021).

Comment 10
Missing data: no information is provided on the proportion of missing values in variables or how these were handled (e.g., imputation, exclusion, or other approaches).

Response 10
We thank the reviewer for noting this omission. We have added missing-data handling procedures to Methods.

For the Olink Explore platform (UK Biobank validation cohort), the proximity extension assay can yield missing values. We obtained Olink data from UK Biobank Record Table 1072 and applied the following pipeline, now described in Methods ("UK Biobank Validation Cohort"):

"To address missing values in the proteomic data, strict quality control filtering and validated imputation were applied. First, participants with >50% missing protein values and proteins with >10% missing rates across all participants were excluded, retaining 2,916 of the initial 2,923 proteins. The remaining sparse missing values were imputed via the k-nearest neighbors (KNN) method. To avoid data leakage, the dataset was randomly split into 5 mutually exclusive cross-validation folds; for each fold, a KNN model (k=10) was trained on 80% of the data and applied to the held-out fold for missing value estimation."

For the SomaScan 11K platform (CMCS cohort), the aptamer-based assay generates continuous fluorescence-intensity values for all probes and therefore produces no structurally missing values in the delivered dataset.

Comment 11
More detail is needed on the ageing clock analyses. UK Biobank does not include SomaScan proteomics; its proteomic profiling is based on the Olink platform (approximately 3,000 proteins). Unlike SomaScan, this technology is not aptamer-based, and the correspondence between SomaScan and Olink protein measurements is not particularly high. The manuscript does not describe how measurements were harmonised across platforms.

Response 11
We thank the reviewer for this insightful and constructive comment, which has enabled us to clarify our analytical workflow and further strengthen the methodological robustness of our ageing clock analyses. **No predictive models or regression coefficients were transferred across proteomic platforms.** All proteomic aging clocks were exclusively developed, trained, and validated within the Olink Explore 3072-based UK Biobank Pharma Proteomics Project (UKB-PPP) data. The longitudinal SomaScan 11K data from the CMCS cohort served a single, distinct purpose: to characterize the Protein Homeostatic Index (PHI), an intrinsic **biological property** quantifying how stably each circulating protein is maintained within individuals over time. PHI was then used as a selection criterion to refine the Olink-derived aging clock by excluding proteins with high intra-individual volatility (low PHI) and prioritizing those with robust longitudinal stability (high PHI).

The reviewer rightly notes that the measurement correspondence between SomaScan and Olink is generally modest. Multiple independent studies have consistently demonstrated this: Eldjarn et al. (Nature, 2023) directly compared SomaScan and Olink measurements in ~1,500 Icelandic individuals with paired-platform profiling and found modest overall correlation; critically, they demonstrated that proteins sharing the same UniProt ID on the two platforms can exhibit divergent genomic associations (e.g., differing cis-pQTLs) and disease associations, underscoring that biological properties derived from one platform cannot be naively transferred to the other through identifier matching alone. Wang et al. (Nat Commun, 2025) reported a median inter-platform correlation of 0.29 across 2,168 proteins profiled on both platforms in ~4,000 Chinese adults; and Oh et al. (Nat Med, 2025). recently documented that proteomic aging models trained on one platform transfer poorly to the other. A fundamental explanation for these discrepancies is that assays mapped to the same UniProt ID on the two platforms may nonetheless target distinct molecular entities (different protein isoforms, post-translationally modified variants, or epitopes) such that UniProt ID-based matching alone is insufficient for reliable cross-platform inference. Therefore, before transferring any biological property (such as PHI) across platforms, it is essential to empirically verify that the corresponding

assays on each platform measure comparable molecular species.

Given these well-documented platform differences, the central question is whether SomaScan-derived PHI values can be legitimately applied to guide protein selection in the Olink-based aging clock. We addressed this through inter-platform measurement consistency (IMC) filtering. We leveraged the paired-platform data from Wang et al. (Nat Commun, 2025), in which 2,168 proteins were profiled using both Olink and SomaScan in the same 3,976 individuals. Even when two assays share the same UniProt ID, they may recognize different epitopes or isoforms, yielding discordant measurements. We therefore used the empirical inter-platform correlation for each protein as a direct indicator of whether both platforms measure comparable molecular species. SomaScan-derived PHI was transferred to the corresponding Olink assay only for proteins with an IMC (Spearman correlation coefficient) above the population median of 0.29. Proteins below this threshold, where the two platforms likely target distinct molecular entities despite sharing the same UniProt ID, were excluded from cross-platform PHI transfer. We have added a new Methods subsection ("Cross-platform integration of PHI for aging clock refinement") describing this framework in the revised manuscript.

We are grateful to the reviewer for raising this constructive and important point. As large-scale proteomic biobanks increasingly adopt either SomaScan or Olink as their primary profiling platform, the question of how and under what conditions biological properties derived from one platform can be transferred to the other has become a pressing methodological challenge for the field. Our IMC-based filtering framework offers a principled, empirically grounded approach to this problem, and we believe the lessons learned here will be broadly informative for future cross-platform proteomic studies.

References:

Eldjarn, G.H., Ferkingstad, E., Lund, S.H. et al. Large-scale plasma proteomics comparisons through genetics and disease associations. Nature 622, 348–358 (2023).

Wang, B., Pozarickij, A., Mazidi, M. et al. Comparative studies of 2168 plasma proteins measured by two affinity-based platforms in 4000 Chinese adults. Nat Commun 16, 1869 (2025).

Oh, H.SH., Le Guen, Y., Rappoport, N. et al. Plasma proteomics links brain and immune system aging with healthspan and longevity. Nat Med 31, 2703–2711 (2025).

Comment 12

Lines 472–473: “ProtAge20-replaced, where the three unstable proteins were substituted with three high PHI and age-associated proteins (WFDC2, AGER, CA14).” How were these three proteins selected? It is unclear by which method and

from which pool of proteins the replacement proteins were chosen, and whether multiple testing was accounted for. These analyses need to be described and justified in greater detail.

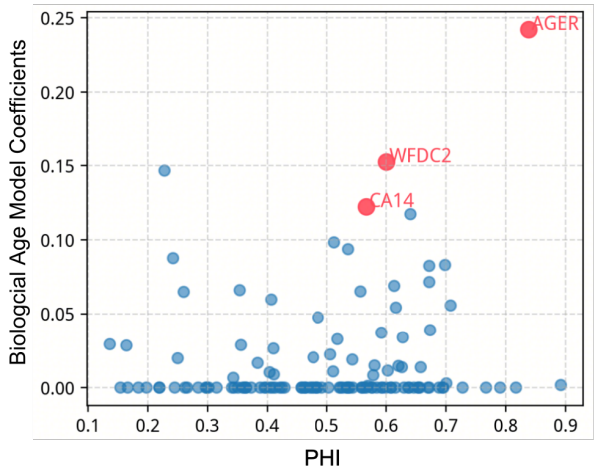
Response 12

We thank the reviewer for pointing out this unclear description regarding the selection of replacement proteins. The original ProtAge20 clock was derived by Argentieri et al. (Nat Med, 2024) from a set of 204 proteins that were pre-selected for their ability to predict chronological age in the UK Biobank. To identify replacement candidates, we considered the same pool of 204 proteins to maintain comparable biological coverage. Within that pool, we chose candidates with PHI > 0.5 (indicating long-term temporal stability) and the largest absolute coefficients (reflecting age-predictive capacity) from a published biological age clock (Goeminne et al., Cell Metab, 2024), yielding WFDC2, AGER, and CA14 (Extended data figure 3). We have now clarified the selection procedure in the revised manuscript:

"These replacement candidates were selected from the same pool of 204 age-associated proteins used to derive the original ProtAge20 clock, with inclusion criteria of PHI > 0.5 and the largest absolute coefficients from a validated mortality-linked biological age clock (Extended data figure 3). This strategy primarily serves to demonstrate the applicability of PHI as a stability metric for sparse panel construction."

We also acknowledge that this procedure (considering only coefficient magnitude and PHI) is not an optimized selection strategy. It was designed solely to demonstrate that incorporating PHI as a stability filter can improve panel robustness. As this was a methodological validation rather than an unbiased discovery analysis for novel markers, multiple testing correction was not required for this targeted selection step. We have revised the Discussion to state this explicitly:

" Importantly, the aging clock framework presented here should be viewed as a proof-of-concept application illustrating the utility of PHI-informed biomarker prioritization, rather than as a finalized clinical predictor. The panel optimization strategy employed in this study was designed to illustrate the conceptual value of temporal stability in sparse panel construction and does not represent a definitive optimization framework. Future efforts integrating PHI with other protein properties, such as assay reproducibility, cross-population applicability, and clinical accessibility, may further improve the design of proteomic markers for applications in precision medicine."



Extended data figure 3. Selection of replacement proteins for the PHI-informed aging clock. Each point represents one protein from a pre-defined pool of 204 age-predictive proteins originally used to develop the ProtAge20 clock (Argentieri et al., Nat Med, 2024) in the UK Biobank. All proteins are plotted by their PHI (x-axis) against the absolute coefficient from a published biological age clock (Goeminne et al., Cell Metab, 2024; y-axis). The three selected replacement proteins (WFDC2, AGER, and CA14) are highlighted in red and labeled; they were prioritized from the 204-protein pool by satisfying the threshold of PHI > 0.5 and exhibiting the largest absolute clock coefficients among eligible candidates.

Reference:
Goeminne, L.J.E., et al. Plasma protein-based organ-specific aging and mortality models unveil diseases as accelerated aging of organismal systems. Cell Metab 37, 205-222 (2025).

Argentieri, M.A., et al. Proteomic aging clock predicts mortality and risk of common age-related diseases in diverse populations. Nat Med 30, 2450-2460 (2024).

Comment 13
Tables S1–S19 were missing from the submission, although they appear essential for evaluating the results. In addition, protein-specific results for intracellular, extracellular membrane, and secreted plasma proteins were not available. The authors state that “analysis results for all 10,776 proteins are available in the supplementary materials” (lines 519–520), but I could not locate these results in the submitted files.

Response 13
We sincerely apologize for the omission of Tables S1–S19 during the initial submission. We have now uploaded the complete supplementary table file with the revised submission as “Revised-Table S1-20.xlsx”. To facilitate review, we have also provided a Google Drive link to the same file (https://docs.google.com/spreadsheets/d/1HO3E5NaTu5jxlu7_H22fhp8T94OLRf-

d/edit?usp=drive_link&ouid=102497683760550802626&rtpof=true&sd=true).

The revised supplementary table file contains original Tables S1–S19 and a new Table S20. In particular, the protein-level results for all 10,776 assayed human protein markers are provided in Tables S1–S19, including the PHI estimates across visit intervals and demographic strata, genetic and exposomic association results, subcellular localization annotations, organ-enrichment analyses, and personalized deviation/Cox regression results. The newly added Table S20 provides SomaScan assay quality-control information, including probe-level QC status and flags. All captions are provided in the revised manuscript.

Comment 14
Statements on clinical implications should be tempered. The predictive performance indicated by the C-index in Figure 5 for death and chronic disease was low (<0.7) or very low (<0.6), suggesting limited clinical utility. In addition, SomaScan proteomics measure relative protein abundance rather than absolute concentrations (e.g., mmol/L) used in clinical practice; therefore, clinical thresholds cannot be directly derived from these semi-quantitative measurements. Consequently, translation into clinical practice is less direct than suggested in the manuscript.

Response 14
Thank you for this constructive and insightful comment. We fully agree that the clinical implications of our findings should be appropriately tempered, and we have added a "Limitations of Clinical Implications" subsection to the new Limitations section to address these important points clearly and explicitly. The added text reads as follows:

"Limitations of clinical implications: First, the predictive performance of our modified proteomic aging clock (ProtAge20-replaced), as measured by C-index, is modest for individual clinical outcomes. It is critical to emphasize that ProtAge20-replaced was **not designed as a direct clinical diagnostic or prognostic tool for specific diseases**. Rather, it was developed as an improved quantitative measure of biological aging, and its association with age-related outcomes serves primarily to validate that it captures biologically meaningful aging signals more accurately than the original clock. The primary goal of this analysis was to demonstrate the principle that incorporating temporal stability (PHI) into biomarker panel design enhances the robustness of biological age estimation, not to develop a clinically deployable disease predictor. Second, the SomaScan 11K platform measures relative protein abundance rather than the absolute concentrations (e.g., mmol/L) used in routine clinical practice. While relative quantification is sufficient for evaluating longitudinal protein stability, comparing inter-individual differences, and performing population-level risk stratification, it cannot directly generate standardized clinical thresholds that are required for point-of-care decision-making. Translation of these findings into clinical practice will require the development of independent, absolute-quantification assays for the most promising PHI-informed biomarker panels, as well as rigorous clinical

validation in diverse patient populations. Taken together, the clinical implications of this work lie not in providing immediately actionable clinical tools, but in establishing temporal stability (quantified by PHI) as a fundamental biological property that should be explicitly considered alongside effect size and disease association strength when designing blood-based proteomic biomarkers for precision medicine. This principle applies broadly to all biomarker development efforts, regardless of the proteomic platform used."

This revision clarifies the distinction between our research tool for biological age quantification and a ready-to-use clinical diagnostic, acknowledges the technical limitations of relative proteomic quantification for direct clinical translation, and reframes our core clinical contribution as establishing PHI as a critical criterion for future biomarker panel design. We believe this modification strengthens the manuscript by providing a more accurate and balanced assessment of the clinical relevance of our work.

Comment 15
Other textual points: the statement “This study provides the first lifespan-level, high-resolution atlas of the human blood proteome, spanning two decades and thousands of individuals… ranging from lifelong stability to pronounced volatility” (lines 289–290) appears overstated. The study does not provide lifespan-level results or evidence of lifelong stability; it covers the period from midlife to older age.

Response 15
Thank you for this accurate correction. We fully agree that the original wording was overstated and apologize for this imprecision. In the revised manuscript, we have replaced the inaccurate phrase "lifespan-level" with "mid-to-late-life". We have revised "ranging from lifelong stability to pronounced volatility" to "ranging from long-term stability to pronounced volatility". We greatly appreciate the reviewer’s attention to this detail, which has helped us improve the accuracy and scientific rigor of our work.

Comment 16
It would be worth noting as a limitation that the reproducibility/generalisability of these findings has not yet been confirmed in an independent cohort study.

Response 16
Thank you for raising this critically important point regarding the reproducibility and generalizability of our findings. We fully agree that independent cohort validation is a cornerstone of rigorous molecular epidemiology, and we have addressed this comment comprehensively by both detailing our existing validation efforts and adding a dedicated subsection to the Limitations section to explicitly acknowledge the remaining gaps.

As you correctly note, a fully independent replication of our proteome-wide PHI atlas would ideally require a separate cohort with longitudinal SomaScan 11K profiling. To our knowledge, such a resource is not currently available to the scientific community. We have therefore taken a rigorous, pragmatic approach to probe the generalizability of the PHI concept using the best available independent data: the longitudinal Olink 1.5K subset of the UK Biobank (n = 1,268 participants with three repeated measurements spanning approximately 10 years).

A key challenge in cross-platform validation is that assays mapping to the same gene symbol on different proteomic platforms often target distinct molecular isoforms, post-translational modifications, or epitopes. A simple gene-level match does not guarantee biochemical comparability, and naive cross-platform comparison can lead to misleading conclusions. To address this, we stratified proteins by their empirically measured inter-platform measurement consistency (IMC) from a large paired-platform study (Wang et al., Nat Commun, 2025), which profiled 2,168 proteins using both SomaScan and Olink assays in the same 3,976 individuals.

As shown in the revised Extended data figure 2, we observed a strong and highly significant correlation between SomaScan-CMCS-based PHI and Olink-UKB-based PHI for the high-consistency protein set (IMC > 0.29, n = 782; Spearman correlation coefficient = 0.701, $p < 0.001$). In contrast, the low-consistency protein set (IMC < 0.29, n = 462) showed negligible correlation (Spearman correlation coefficient = 0.119, $p = 0.01$). This pattern is exactly what would be expected: when two platforms measure biochemically comparable molecular species, the temporal stability signal captured by PHI transfers robustly both across cohort and platform; when they measure distinct molecular entities, PHI values are inherently platform-specific and cannot be directly compared.

To ensure full transparency about the limitations of our validation, we have added the following dedicated subsection to the Limitations section of the Discussion:

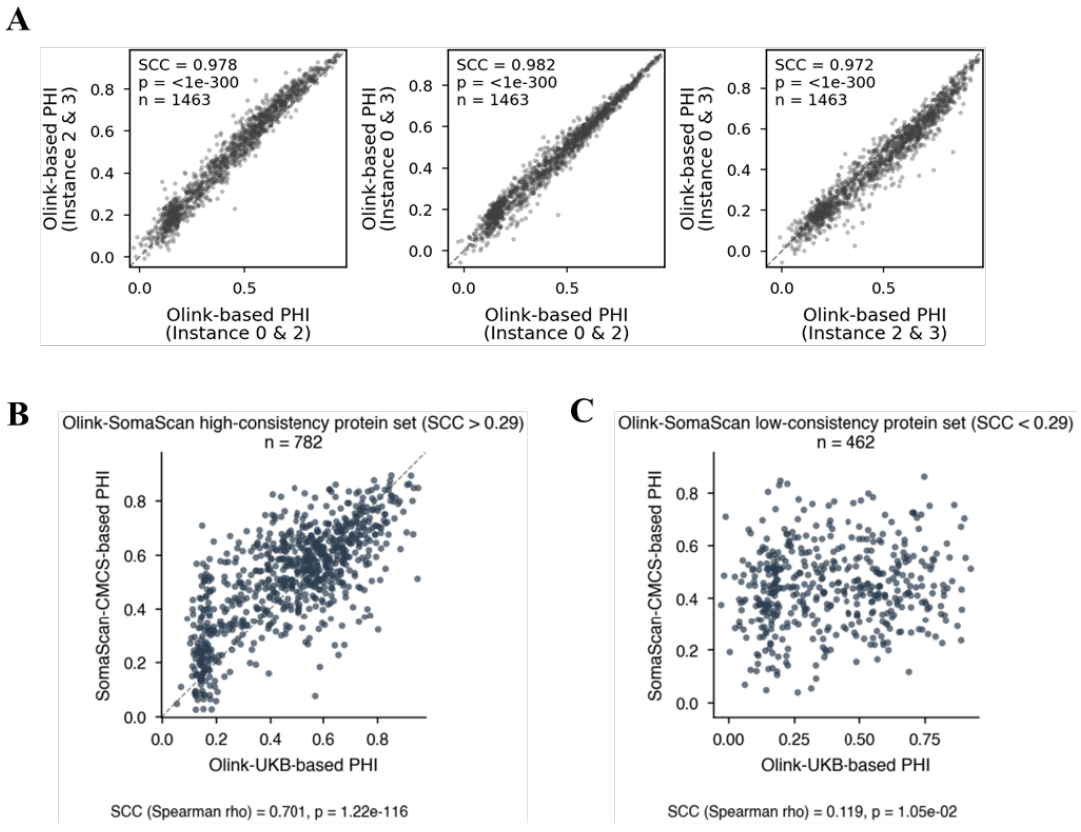
"Limitations of Independent Cohort Validation of PHI: Although we have completed an independent validation of the PHI framework and analysis of its cross-platform transferability in the UK Biobank cohort, this validation has notable coverage limitations. Longitudinal Olink data in UK Biobank only covers approximately 1,500 proteins, and only 782 of these proteins exhibit sufficient molecular measurement consistency between the SomaScan and Olink platforms to support meaningful cross-platform comparison of PHI. In this study, we systematically calculated PHI values for 10,776 proteins in the CMCS cohort, constructing the most comprehensive atlas of long-term human blood protein homeostatic indices available globally to date. However, PHI values for more than 93% of these proteins still have not yet been validated in independent cohorts.

Importantly, the observed discordance in PHI values for proteins with low inter-platform measurement consistency (IMC) is not a failure of validation, but an

expected and biologically interpretable result. Assays targeting the same gene symbol on SomaScan and Olink often recognize distinct protein isoforms, post-translationally modified variants, or epitopes, such that they measure non-identical molecular entities. For these proteins, PHI values are inherently platform-specific and cannot be naively transferred across technologies.

Future fully independent longitudinal cohort studies using the SomaScan 11K platform are urgently needed to comprehensively validate the reproducibility and generalizability of the proteome-wide PHI atlas reported in this study. Such studies will also help disentangle platform-specific technical effects from true biological variation in protein temporal stability."

This revision clearly articulates both the strength of our partial validation (demonstrating that the PHI concept is generalizable when molecular measurements are comparable) and the critical remaining limitation (the vast majority of our 10,776 PHI values have not yet been independently validated). We believe this balanced assessment significantly strengthens the scientific rigor of our manuscript.



Extended data figure 2. Validation of PHI using an independent longitudinal Olink 1.5K profiling subcohort from UK Biobank. A, Scatter density plots comparing Olink-based PHI values calculated from different pairs of longitudinal visits in the UK Biobank COVID-19 repeat imaging subcohort (n = 1,268 participants with three proteomic assessments). PHI was

computed for three visit pairs: Instance 0 (baseline) & Instance 2 (first imaging visit, 2014+), Instance 0 & Instance 3 (second imaging visit, 2019+), and Instance 2 & Instance 3. Spearman correlation coefficients (SCC) and two-sided P values are shown for each comparison, demonstrating near-perfect consistency of PHI estimates across different time intervals. B, Scatter plot comparing PHI values derived from the SomaScan-CMCS cohort (x-axis) and the Olink-UKB cohort (y-axis) for proteins with high inter-platform measurement consistency (IMC; SCC > 0.29, n = 782 proteins). The strong positive correlation (SCC = 0.701, $P = 1.22 \times 10^{-116}$) indicates that PHI signals are highly reproducible across platforms and cohorts when assays measure comparable molecular species. The dashed grey line represents the diagonal (y = x). C, Scatter plot comparing PHI values for proteins with low inter-platform measurement consistency (IMC; SCC ≤ 0.29, n = 462 proteins). The negligible correlation (SCC = 0.119, $P = 1.05 \times 10^{-2}$) demonstrates that PHI values are not transferable across platforms when assays target distinct molecular entities (e.g., different protein isoforms or post-translationally modified variants), highlighting the risks of naive cross-platform extrapolation.

References:

Wang, B., Pozarickij, A., Mazidi, M. et al. Comparative studies of 2168 plasma proteins measured by two affinity-based platforms in 4000 Chinese adults. Nat Commun 16, 1869 (2025).

Comment

In conclusion, this is a potentially interesting study. However, without clearer reporting of the data, methods, and results, the strength of the evidence cannot currently be assessed. I would be happy to review a revised version that addresses the points raised above.

Response

We sincerely thank the reviewer for acknowledging the potential interest of our study. We greatly appreciate your thoughtful and constructive comments, all of which have helped us improve the clarity and rigor of our work.

In response to the concerns you raised, we have now provided a point-by-point response to each of the 16 comments in our revised manuscript. We have carefully clarified the reporting of data, methods, and results throughout the paper and the supplementary materials.

We hope that the revised version now presents the evidence with sufficient strength and transparency. We would be very grateful if you would review this revised manuscript, and we believe it has been substantially improved based on your valuable suggestions.

Reviewer #1:
None

Reviewer #2:
Remarks to the Author:
The authors revised manuscript is greatly improved from their original submission. I have no further concerns

Response: Thank you for your kind words.

Remarks on code availability:
I did not have capacity to review the code in detail, but from a brief evaluation, it looks appropriate.

Response: Thank you for your kind words.

Reviewer #3:
Remarks to the Author:
The authors have carefully addressed all my comments. I have only two minor points:

Response 1: Extended Data Figure 1A. Please add to the second box the number of participants recruited from the Beijing communities.

Response: Thank you for this suggestion. We have added to the second box the number of participants (n = 3,131) recruited from the Beijing communities.

Response 2: The authors note that 1,298 participants had two or more proteomic assessments, constituting the valid sample for the PHI analysis. Please clarify which analyses included data from the 1,323 participants with only one proteomic assessment (i.e. 2,621 − 1,298). If none, then the effective sample size appears to be 1,298 rather than the stated 2,621, and the manuscript should be revised accordingly.

Response: Thank you. We have revised the abstract and introduction: 1,298 were effective samples.