Proteomic aging signatures predict disease risk and mortality across diverse populations

In a large human population study of proteomic aging, we developed a proteomics-based age clock for UK Biobank participants and validated its accuracy in the China Kadoorie Biobank and FinnGen. Proteomic aging is associated with mortality, risk of 18 chronic diseases and numerous age-related traits, including cognitive function.

This is a summary of:
Argentieri, M. A. et al. Proteomic aging clock predicts mortality and risk of common age-related diseases in diverse populations. *Nat. Med.* https://doi.org/10.1038/10.1038/s41591-024-03164-7 (2024).

Published online: 08 August 2024

The problem

Chronological age is the main determinant of most common chronic diseases but is an imperfect surrogate for aging, which is the driver of age-related multimorbidity and mortality. Aging can be estimated more precisely through the use of '-omics' data to capture the biological functioning of a person relative to an expected level of functioning for a person of a given chronological age[1]. Although the most common biological aging 'clocks' use DNA methylation[1,2], protein levels may provide a more direct mechanistic and functional insight into aging biology. Moreover, the proteome is the most common target for drug development. However, previous proteomics age clock studies[3,4] have not been validated independently across populations with diverse genetic and geographic backgrounds. Also, thus far, none have been developed in large or well-powered general population samples that allow association testing across a wide spectrum of age-related disorders, multimorbidity and mortality.

The solution

We developed a machine learning model that uses blood proteomics information to estimate a proteomics age clock for a large sample of participants from the UK Biobank (UKB) (n = 45,441; age range, 40–70 years). We further validated this model in two biobanks across the world: the China Kadoorie Biobank (CKB) (n = 3,977; age range, 30–80 years) and FinnGen (n = 1,990; age range, 20–80 years). These biobanks represent geographically and genetically distinct populations that differ from the UKB in age range and morbidity profiles. We systematically assessed the influence of the proteomic age gap (defined as the difference between protein-predicted age and chronological age) on 27 aging-related phenotypes related to biological, functional and cognitive function; all-cause mortality; and incidence of 26 common age-related diseases.

We identified 204 proteins that accurately predict chronological age, and we further identified a set of 20 aging-related proteins that capture 91% of the age-prediction accuracy of the larger model. We demonstrated that our proteomics age clock showed age-prediction accuracy in independent participants from China and Finland similar to its performance for the UK Biobank. We found that proteomic aging was associated with the incidence of 18 major chronic diseases (including diseases of the heart, liver, kidneys and lungs, diabetes, neurodegeneration and cancer), multimorbidity and all-cause mortality (Fig. 1). Proteomic aging was also associated with age-related measures of biological, physical and cognitive function, including telomere length, frailty index and several cognitive tests.

The implications

We have provided some of the largest and most comprehensive evidence thus far demonstrating that proteomic aging is a common biological signature related to numerous age-related functional traits, morbidities and mortality. We have also provided some of the first evidence that a proteomics-based age clock can be highly generalizable across human populations of diverse genetic ancestries, age ranges and morbidity profiles. Multimorbidity is an important problem in clinical and population health that has a major impact on the cost of healthcare. Our proteomics clock provides insight into the pathways that form the biological basis of multimorbidity.

Limitations of our work include the fact that we had data available on the expression of only 2,987 proteins, which leaves room for future improvements as larger proteomics panels become available in biobank datasets. We also did not have repeat measurements of proteins in enough people to assess protein changes over time in them. Finally, combining proteomics with DNA methylation will boost research further in terms of drug development (such as for cancer) and elucidation of the role of the environment on aging.

In the near future, we envision that proteomics age clocks will be used to study the relationship between genetics and the environment in aging, yielding novel insights into the drivers of aging and multimorbidity across the lifespan. An important avenue will also be to use proteomics clocks as a biomarker for the effectiveness of preventive interventions that target aging and multimorbidity. Furthermore, proteomics clocks may be used to accelerate drug development and clinical trials through the identification of patients at high and low risk (for example, less than 1% of those in the bottom decile of proteomic aging developed Alzheimer's disease over the ensuing 10–15 years).

M. Austin Argentieri[1] & Cornelia M. van Duijn[2]
[1]Massachusetts General Hospital, Boston, MA, USA. [2]University of Oxford, Oxford, UK.

EXPERT OPINION

"This manuscript is certainly a landmark paper, given its scale and statistical power. It thus sets a new standard and should convince the last holdouts that large-scale plasma proteomics is the future of human health and disease assessment and prediction". **Tony Wyss-Coray, Stanford University, Stanford, CA, USA.**

FIGURE

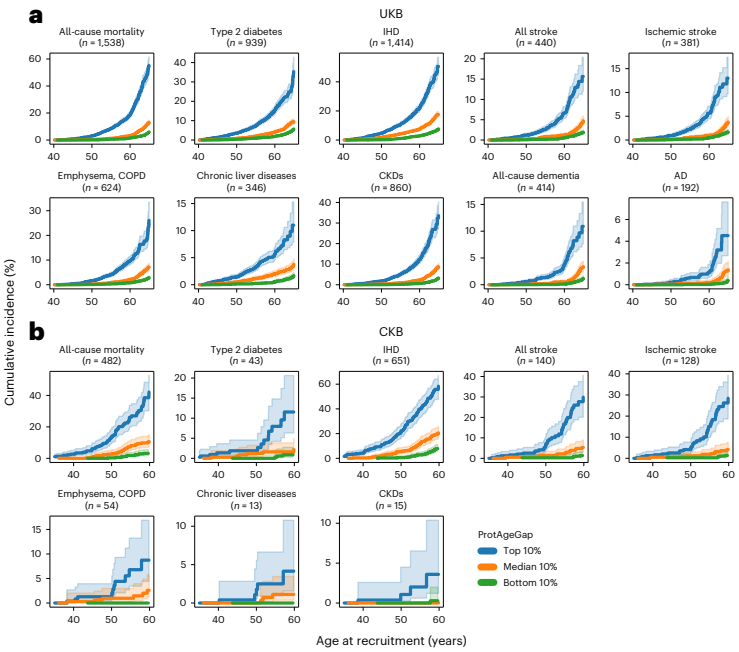

Fig. 1 | Proteomics age clock stratifies people into divergent age-specific mortality and disease risk trajectories. Cumulative incidence of various diseases and mortality (above plots) by decile of proteomic age gap (ProtAgeGap) (in years) in the UKB (**a**) and CKB (**b**). Incidence rates are shown for the subsequent 11–16 years (UKB) or 11–14 years (CKB) of follow-up after recruitment for each given age at recruitment. All plots show the cumulative density of events at a given timepoint on the basis of the Kaplan–Meier survival function, with 95% confidence intervals (lighter shading).

BEHIND THE PAPER

Our team has conducted aging-related research together for years, focusing on mapping the environmental and biological underpinnings of human aging. The challenges of this research include developing re-normalization techniques for Olink proteomics data generated in separate cohorts, and developing models that are generalizable across diverse datasets. To tackle these challenges, we performed systematic benchmarking of machine learning model architectures to identify the model types with the greatest generalizability to new populations. Our work lays the foundation for a new generation of large-scale human population research focused on elucidating aging biology through leveraging the power of biobank datasets. We welcome future collaborations with worldwide biobanks, cohort studies and other genomic and multi-omics studies to improve our understanding of human aging. **M.A.A. & C.M.v.D.**

REFERENCES

1. Rutledge, J., Oh, H. & Wyss-Coray, T. Measuring biological age using omics data. *Nat. Rev. Genet.* **23**, 715–727 (2022). **This review article outlines research thus far on using '-omics' data to create biological age clocks.**
2. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* **19**, 371–384 (2018). **This review article presents research thus far that used DNA methylation data to create biological age clocks.**
3. Lehallier, B. et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nat. Med.* **25**, 1843–1850 (2019). **This paper reports non-linear patterns of protein expression across age groups and develops a proteomics age clock.**
4. Oh, H. S. et al. Organ aging signatures in the plasma proteome track health and disease. *Nature* **624**, 164–172 (2023). **This paper reports the development of organ-specific proteomics age clocks and their associations with select diseases.**

FROM THE EDITOR

"Biological aging is typically monitored via epigenetic clocks. This study goes in a different direction by building a biological aging clock based on protein levels in the blood and comprehensively tests its performance in three different biobanks, showing how this clock can predict the incidence of a wide variety of chronic diseases and age-related functional traits in populations of diverse ancestries." **Editorial Team, *Nature Medicine*.**