

Undulating changes in human plasma proteome profiles across the lifespan

Benoit Lehallier^{1,2,3*}, David Gate^{1,2,3,4}, Nicholas Schaum⁵, Tibor Nanasi^{1,2,3,6}, Song Eun Lee^{1,2,3,4}, Hanadie Yousef^{1,2,3,4}, Patricia Moran Losada^{1,2,3}, Daniela Berdnik^{1,2,3,4}, Andreas Keller^{1,2,3,4}, Joe Verghese^{8,9}, Sanish Sathyan^{8,9}, Claudio Franceschi^{10,11}, Sofiya Milman^{8,12}, Nir Barzilai^{8,12} and Tony Wyss-Coray^{1,2,3,4*}

Aging is a predominant risk factor for several chronic diseases that limit healthspan¹. Mechanisms of aging are thus increasingly recognized as potential therapeutic targets. Blood from young mice reverses aspects of aging and disease across multiple tissues^{2–10}, which supports a hypothesis that age-related molecular changes in blood could provide new insights into age-related disease biology. We measured 2,925 plasma proteins from 4,263 young adults to nonagenarians (18–95 years old) and developed a new bioinformatics approach that uncovered marked non-linear alterations in the human plasma proteome with age. Waves of changes in the proteome in the fourth, seventh and eighth decades of life reflected distinct biological pathways and revealed differential associations with the genome and proteome of age-related diseases and phenotypic traits. This new approach to the study of aging led to the identification of unexpected signatures and pathways that might offer potential targets for age-related diseases.

Aging underlies declining organ function and is the primary risk factor for several diseases¹. Thus, a deeper understanding of aging is likely to provide insights into mechanisms of disease and to facilitate the development of new antiaging therapeutics. A growing number of investigators have applied genomic, transcriptomic and proteomic assays (collectively referred to as ‘omics’) to studies of aging¹¹. Human genetic studies have uncovered relatively few modifiers of aging, yet other omics modalities, which measure more dynamic gene modifications or products, have provided valuable insights. For example, the transcriptome varies greatly during aging across tissues and organisms¹², pointing to evolutionarily conserved, fundamental roles of developmental and inflammatory pathways¹³. The protein composition of cells, bodily fluids and tissues changes similarly with age and provides insights into complex biological processes, as proteins are often direct regulators of cellular pathways. In particular, blood, which contains proteins from nearly every cell and tissue, has been analyzed to discover biomarkers and gain insights into disease biology. Accordingly, organismal aging results in proteomic changes in blood that reflect aspects of aging of different cell types and tissues.

Perhaps the strongest evidence that blood can be used to study aging comes from experiments employing heterochronic parabiosis, which is a surgically induced state that connects the circulatory systems of young and old mice. These studies show that multiple tissues, including muscle, liver, heart, pancreas, kidney, bone and brain, can be rejuvenated in old mice^{2–10}. Plasma (the soluble fraction of blood) from old mice is sufficient to accelerate brain aging after infusion into young mice⁹, and young plasma can reverse aspects of brain aging^{10,14}. Together, these studies support the notion that the plasma proteome harbors key regulators of aging. Identifying such protein signatures may help in understanding mechanisms of organismal aging. However, plasma proteomic changes with age have not been thoroughly exploited and require new tools to derive insights into the biology of aging. In this study, we carried out a deep proteomic analysis of plasma from young adults to nonagenarians. Using new analysis tools, we discovered changes in protein expression across the lifespan and linked these changes to biological pathways and disease.

Results

Linear modeling links the plasma proteome to functional aging and identifies a conserved aging signature. We analyzed plasma isolated from EDTA-treated blood acquired by venipuncture from 4,263 healthy individuals aged 18–95 years from the INTERVAL¹⁵ and LonGenity¹⁶ cohorts (Fig. 1a and Extended Data Fig. 1). Currently, one of the most advanced tools for the measurement of plasma proteins is the single-stranded oligonucleotides known as aptamers^{17,18}, which bind to targets with high affinity and specificity. To generate a proteomic dataset of the human lifespan, we used the SomaScan aptamer technology, which is capable of quantifying thousands of proteins (Supplementary Tables 1 and 2) with high precision within and between runs¹⁹ (Supplementary Table 3). The INTERVAL and LonGenity datasets analyzed here can be interrogated with an interactive web interface (https://t2c-stanford.shinyapps.io/aging_plasma_proteome/).

Because females have a longer average lifespan than males²⁰, we assessed whether sex and aging proteomes are interconnected

¹Department of Neurology and Neurological Sciences, Stanford University, Stanford, CA, USA. ²Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA, USA. ³Paul F. Glenn Center for the Biology of Aging, Stanford University, Stanford, CA, USA. ⁴Department of Veterans Affairs, VA Palo Alto Health Care System, Palo Alto, CA, USA. ⁵Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, CA, USA.

⁶Institute of Cognitive Neuroscience and Psychology, Hungarian Academy of Sciences Research Centre for Natural Sciences, Budapest, Hungary.

⁷Clinical Bioinformatics, Saarland University, Saarbrücken, Germany. ⁸Institute for Aging Research, Department of Medicine, Albert Einstein College of Medicine, Bronx, NY, USA. ⁹Department of Neurology, Albert Einstein College of Medicine, Bronx, NY, USA. ¹⁰Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, Bologna, Italy. ¹¹Department of Applied Mathematics, National Research Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod, Russia. ¹²Department of Genetics, Albert Einstein College of Medicine, Bronx, NY, USA. *e-mail: lehallib@stanford.edu; t2c@stanford.edu

(Fig. 1b–d). The proteins most strongly changed with sex included the well-known follicle-stimulating hormone (CGA FSHB), human chorionic gonadotropin (CGA CGB) and prostate-specific antigen (KLK3). With age, the most prominent overall changes with respect to fold change and statistical significance included sclerostin (SOST), ADP ribosylation factor interacting protein 2 (ARFIP2) and growth differentiation factor 15 (GDF15), in addition to several proteins that differed with sex, such as CGA FSHB. The proteins most strongly associated with age also changed significantly with sex (Fig. 1d): 895 of the 1,379 proteins altered with age were significantly different between the sexes ($q < 0.05$; Supplementary Table 4). These results are aligned with several studies that demonstrated that males and females age differently²¹. To determine whether these findings are representative of the general population, we compared changes identified in this study with findings from four independent cohorts from the US and Europe ($n = 171$; age range, 21–107 years; Extended Data Fig. 1d) and with an independent study²². Although these independent cohorts used an older version of the SomaScan assay measuring only a subset of the current proteins (1,305 proteins; Supplementary Table 2), we observed high consistency of the aging and sex proteomes across cohorts (Extended Data Fig. 2).

To establish the biological relevance of these changes, we queried the Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome databases and measured enrichment of proteins in pathways using sliding enrichment pathway analysis (SEPA) (Supplementary Tables 5 and 6). The heat maps produced by SEPA first illustrated the relationship between the top 100 proteins and the biological pathways they represent; second, the heat maps emphasized how a restricted list of top aging-related proteins revealed biological pathways that would have escaped common pathway mining modalities (Fig. 1e). SEPA indicated that incremental lists of proteins are needed to determine the biological functions of sex-related proteins and pointed to expected differences in hormonal metabolism and activity. Conversely, an extensive list of aging proteins contained enrichment for blood-related pathways, such as heparin and glycosaminoglycan binding, as recently reported²².

To determine whether the plasma proteome can predict biological age and serve as a ‘proteomic clock’, we used 2,817 randomly selected individuals to fine-tune a predictive model that was tested on the remaining 1,446 individuals (Fig. 1f). We identified a sex-independent plasma proteomic clock consisting of 373 proteins (Supplementary Table 7), which was highly accurate in predicting age in the discovery, validation and four independent cohorts ($r = 0.93\text{--}0.97$; Fig. 1g and Extended Data Fig. 3a,b). Remarkably, individuals who were predicted to be younger than their chronological age performed better on cognitive and physical tests (Fig. 1h and Supplementary Table 8). Although a reduced model comprising only nine proteins predicted age with good accuracy (Extended Data Fig. 3c and Supplementary Table 7), a combination of different

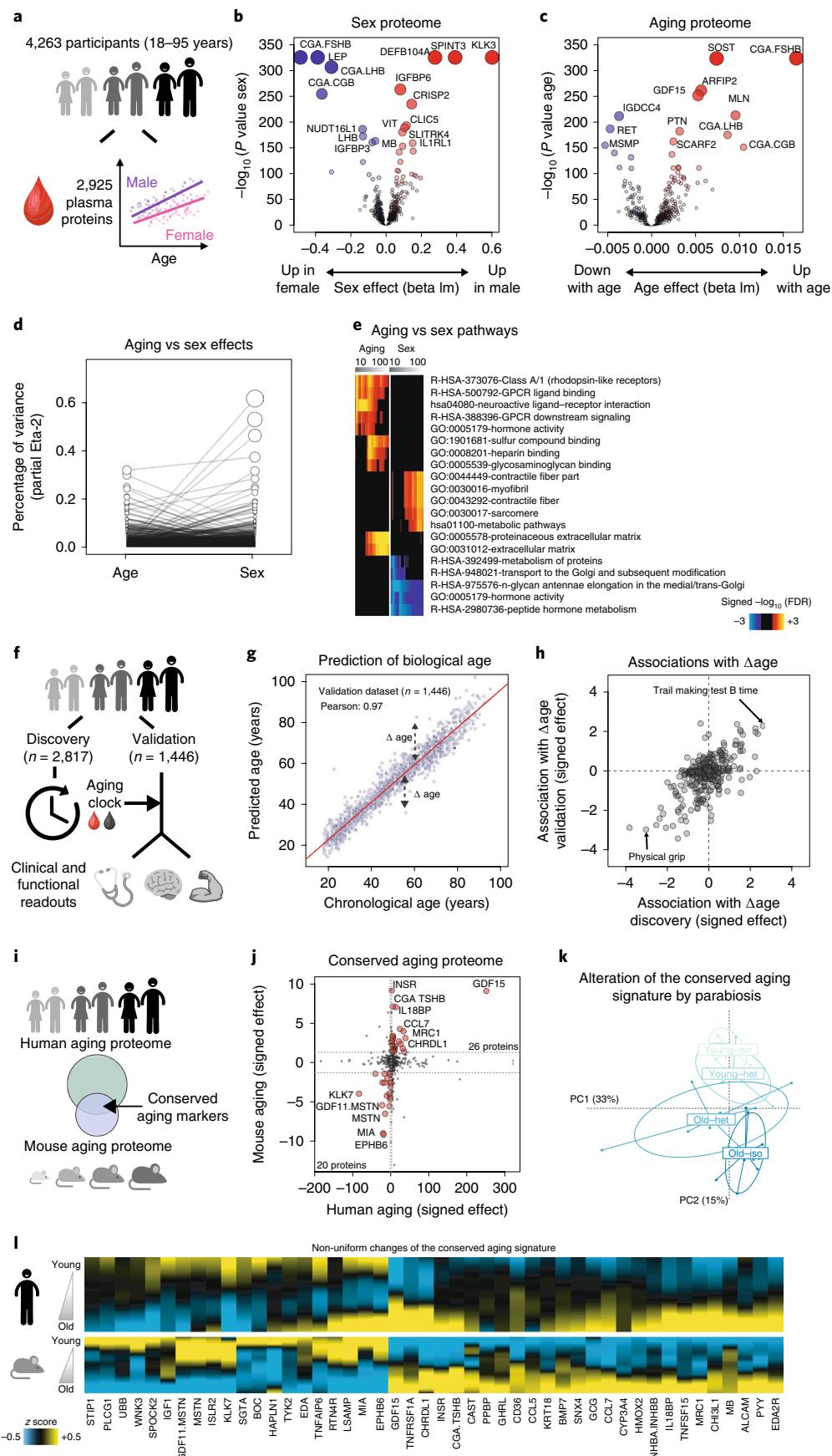
sets of proteins may be required to model changes in a large set of clinical and functional parameters (Extended Data Fig. 3d).

As most biological pathways that change with age are evolutionarily conserved²³, we next identified aging-related proteins conserved between mice and humans. We analyzed mouse plasma ($n = 110$; age, 1–30 months) using SomaScan, which reliably measures hundreds of non-human proteins and has proven useful in mouse studies^{6,24} (Fig. 1i and Supplementary Table 9). In mice, 172 proteins changed with age (out of 1,305 proteins measured; Supplementary Table 10; $q < 0.05$), and 46 proteins overlapped with human aging-related proteins (Fig. 1j). Remarkably, many of these proteins were also modulated by heterochronic parabiosis: young mice exposed to old plasma (young heterochronic mice) showed a relatively older plasma signature, whereas aged mice exposed to young plasma (old heterochronic mice) showed a younger signature (Fig. 1k). Altogether, standard linear modeling of the plasma proteome during the human lifespan revealed established aging pathways, possibly indicating accelerated and decelerated aging in humans and mice. Intriguingly, changes to the conserved aging-related proteins did not occur simultaneously (Fig. 1l). Thus, the chronology of aging in the plasma proteome requires further investigation.

Clustering protein trajectories reveals undulation of the aging plasma proteome. Although standard linear modeling showed prominent changes in plasma protein composition, the undulating behavior of the 46 conserved proteins (Fig. 1l), and, more globally, the 2,925 plasma proteins as a group when they were visualized as z-scored changes across the lifespan, was striking (Fig. 2a,b). These undulating patterns were detected in independent human cohorts and in mice (Extended Data Fig. 4), suggesting that they are robust and conserved.

To reduce the complexity of the proteome, we grouped proteins with similar trajectories using unsupervised hierarchical clustering (Fig. 2c) and identified eight clusters of protein trajectories changing with age, which ranged in size from 8 to 1,415 proteins (Supplementary Table 11). In addition to linear patterns (clusters 1 and 5), several non-linear trajectories were evident, including step-wise, logarithmic and exponential trajectories (clusters 2, 3, 4, 6, 7 and 8) (Fig. 2d). Notably, these cluster trajectories were similarly detectable in independent cohorts (Extended Data Fig. 5). Of the eight clusters analyzed, six were enriched for specific biological pathways ($q < 0.05$; Extended Data Fig. 6 and Supplementary Table 12), suggesting distinct, yet orchestrated, changes in biological processes during the lifespan. For example, proteins present in blood microparticles consistently decreased with age (cluster 5), and other blood-related pathways, such as heparin and glycosaminoglycan binding, increased in a two-step manner (cluster 4), whereas levels of proteins involved in axon guidance and EPH–ephrin signaling remained constant until age 60 before rising exponentially (cluster 6) (Fig. 2d). Altogether, most plasma proteome changes across the lifespan were non-linear.

Fig. 1 | Linear modeling links the plasma proteome to functional aging and identifies a conserved aging signature. **a**, Schematic representation of analysis of the plasma proteome. **b,c**, Volcano plots representing changes of the plasma proteome ($n = 4,263$) with sex (**b**) and age (**c**). Linear models (l.m.), adjusted for age, sex and subcohort, were tested using the F-test. **d**, Relative percentage of variance explained by age and sex. Values for each plasma protein are connected by edges. **e**, Pathways associated with sex and age identified by SEPA ($n = 4,263$). Proteins upregulated and downregulated were analyzed separately. The top ten pathways per condition are represented. Enrichment was tested using Fisher’s exact test (GO) and the hypergeometric test (Reactome and KEGG). **f**, Schematic representation of biological age modeling using the plasma proteome. **g**, Prediction of age in the validation cohort ($n = 1,446$) using 373 plasma proteins. The Pearson correlation coefficient between chronological and predicted age is given. **h**, Association between delta age (difference between predicted age and chronological age) and functional readouts in old. Top associations in both the discovery and validation datasets are represented. **i**, Schematic representation of the comparison between the human and mouse aging proteomes. **j**, Conserved markers of aging. Both human and mouse aging effects are signed by the beta age of the corresponding linear analysis. Forty-six plasma proteins change in the same direction in mice and humans (red dots) and define a conserved aging signature. **k**, Alteration of the conserved aging signature by parabiosis. Normed principal-component analysis was used to characterize changes of the conserved aging signature when mice were exposed to young or old blood. **l**, Age-related changes of the conserved aging signature. Plasma protein levels were z scored, and aging trajectories were estimated by LOESS.



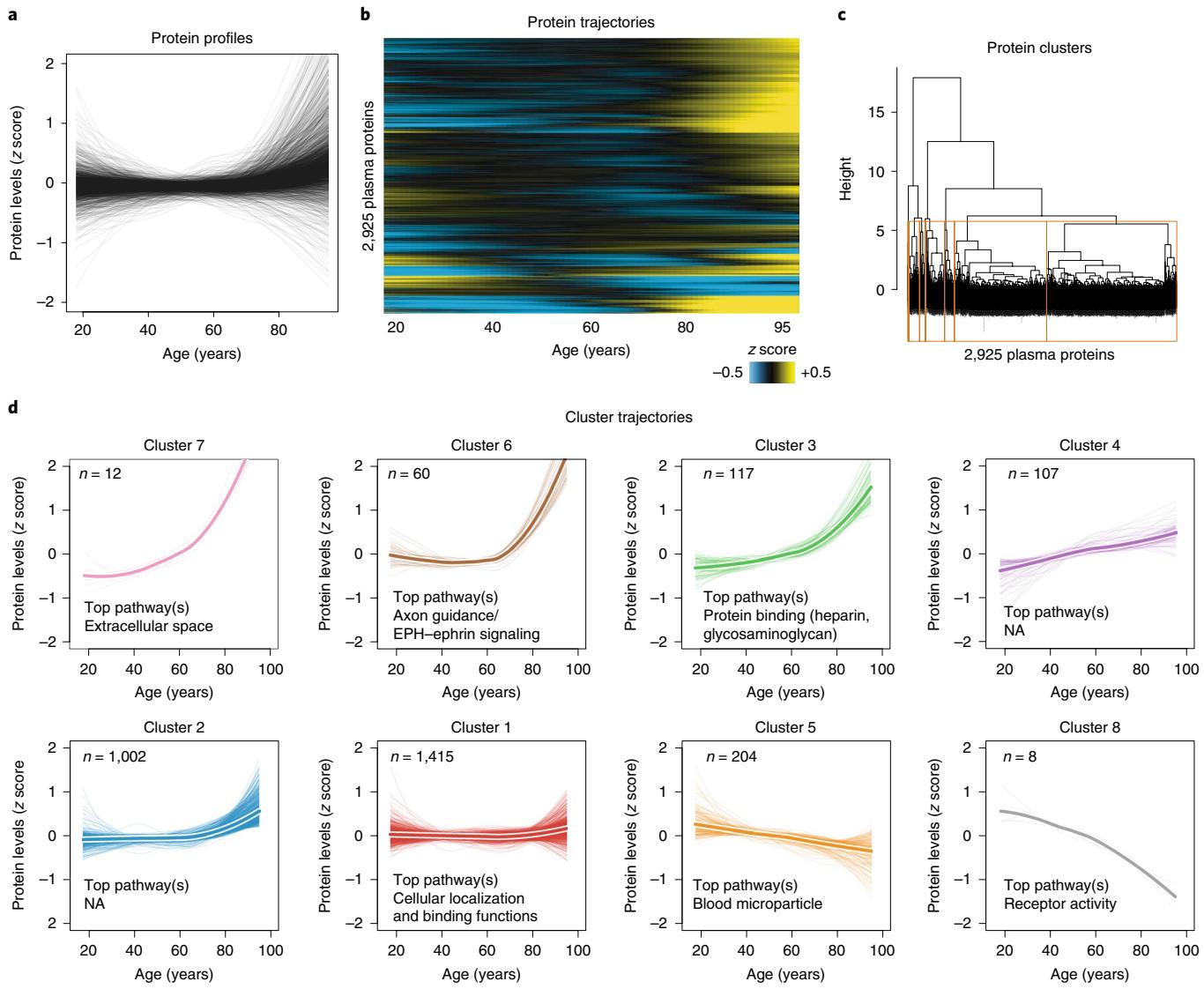


Fig. 2 | Clustering of protein trajectories identifies linear and non-linear changes during aging. **a**, Protein trajectories during aging. Plasma protein levels were z scored, and trajectories of the 2,925 plasma proteins were estimated by LOESS. **b**, Trajectories are represented in two dimensions by a heat map, and unsupervised hierarchical clustering was used to group plasma proteins with similar trajectories. **c**, Hierarchical clustering dendrogram. The eight clusters identified are represented by orange boxes. **d**, Protein trajectories of the eight identified clusters. Clusters are grouped by the similarity of global trajectories, with the thicker lines representing the average trajectory for each cluster. The number of proteins and the most significant enriched pathways are presented for each cluster. Pathway enrichment was tested using the GO, Reactome and KEGG databases. The top 20 pathways for each cluster are listed in Supplementary Table 12.

Quantification of proteomic changes across the lifespan uncovers waves of aging-related proteins. To quantitatively understand the proteomic changes occurring throughout life, we developed the software tool differential expression-sliding window analysis (DE-SWAN) (Fig. 3a). This algorithm analyzes protein levels within a window of 20 years and compares two groups in parcels of 10 years (e.g., 35–45 years compared to 45–55 years), while sliding the window in increments of 1 year from young to old. Using DE-SWAN, we detected changes at particular stages of life and determined the sequential effects of aging on the plasma proteome (while also controlling for the effect of confounding factors). This approach identified hundreds of proteins changing in waves throughout aging (Fig. 3b). Summing the number of differentially expressed proteins at each age uncovered three crests at ages 34, 60 and 78 (Fig. 3c, Extended Data Fig. 7a and Supplementary Table 13). These crests disappeared when the ages of individuals were permuted

(Extended Data Fig. 7b) but were still detectable using different statistical models (e.g., smaller or larger sliding windows) (Extended Data Fig. 7c, Supplementary Fig. 1 and Supplementary Table 13), indicating the robustness of these age-related waves.

Intriguingly, the three age-related crests were largely composed of different proteins (Fig. 3d and Supplementary Table 14), but a few proteins were among the top ten differentially expressed in each crest, such as *GDF15*, which was consistent with its pronounced increase across the lifespan (Fig. 3a). Other proteins, such as chordin-like protein 1 (*CHRDL1*) or matrix metallopeptidase 12 (*MMP12*), were significantly changed only at the last two crests, reflecting their exponential increase with age. Overlap between proteins changing at age 34, 60 and 78 years was statistically significant ($P < 0.05$) but limited (Fig. 3e), and most proteins changing in old age were not identified by linear modeling (Fig. 3f). This prompted us to use SEPA to determine whether these waves reflected distinct

biological processes. Strikingly, we observed a prominent shift in multiple biological pathways with age (Fig. 3g). At young age (34 years), we observed a downregulation of proteins involved in structural pathways, such as the extracellular matrix. These changes were reversed in middle and old age (60 and 78 years, respectively). At age 60 years, we found a prominent role of hormonal activity, binding functions and blood pathways. At age 78 years, key processes still included blood pathways but also bone morphogenetic protein signaling, which is involved in numerous cellular functions²⁵. Pathways changing with age by linear modeling overlapped most strongly with the crests at age 34 and 60 years (Fig. 3g), indicating that dramatic changes occurring in the elderly might be masked in linear modeling by more subtle changes at earlier ages. Altogether, these results showed that aging is a dynamic, non-linear process characterized by waves of changes in plasma proteins that reflect complex shifts in biological processes.

Proteins linked to age-related diseases are enriched in distinct waves of aging. The plasma proteome is sensitive to the physiological state of an individual but is also genetically influenced²⁶. To deconvolute complexity between the genome, proteome and physiology, we asked whether the top aging-related proteins change owing to genetic polymorphisms or whether they are among the top predictors of disease or phenotypic traits. More specifically, we sought to determine whether proteins that comprised the three waves of aging were uniquely linked to the genome or proteome of age-related diseases and traits (Fig. 4a). We used the ranked lists of the top proteins identified by DE-SWAN at each of the three crests (Fig. 3c and Supplementary Table 14) and summed the number of proteins linked to the genome and proteome of specific diseases and traits separately for each wave (i.e., the cumulative sum) (Fig. 4b–i and Extended Data Fig. 8). First, we mined the genomic atlas of the human plasma proteome²⁶ (Fig. 4b) and discovered that the aging proteome is also genetically determined (Fig. 4c and Extended Data Fig. 8). However, the rank of proteins determined by trans-association appeared more random with aging (Fig. 4c), suggesting that other sources drive the aging plasma proteome. We then tested whether the waves of aging proteins were differentially linked with changes in cognitive and physical functions identified in Fig. 1h. Interestingly, the proteome associated with these traits overlapped with the proteome defining middle and old age, when these functions decline the most (Fig. 4d,e). Finally, we used public datasets and summary statistics from SomaScan proteomic studies focused on age-related diseases, including Alzheimer's disease (AD)²⁷, Down syndrome (DS)²⁸ and cardiovascular disease (CVD)²⁹. A plasma proteomic study predicting body mass index (BMI)³⁰ was used as a control because weight gain varies widely with age (Supplementary Fig. 2). As expected, the proteome linked to BMI was not selectively enriched for proteins defining waves of aging (Fig. 4f). Conversely,

CVD-associated proteins were strongly enriched in waves of proteins defining middle and old age compared to young age (Fig. 4g). This enrichment corresponded to an increased incidence of CVD after 55 years of age³¹. Finally, AD- and DS-associated proteins overlapped with the top proteins defining middle and old age but not those defining young age (Fig. 4h,i). The fact that the proteome defining these two diseases also changed in old individuals of a separate disease-free cohort supports the notion of accelerated aging in DS and AD^{32,33}. Altogether, these results show that waves of proteomic aging are differentially linked to the genomic and proteomic traits of various diseases.

Discussion

Our analysis of the plasma proteome reveals complex, non-linear changes over the human lifespan. Although linear analysis provides information about the aging plasma proteome, modeling non-linear protein trajectories is necessary to fully appreciate these undulating changes.

It is well known that men and women age differently²¹, yet we were surprised that two-thirds of proteins that changed with age also changed with sex (895 of 1,379 proteins). This supports the National Institutes of Health (NIH) policy on the inclusion of women in clinical research and using sex as a biological variable in experiments. Nevertheless, a unique proteomic clock can be used to predict age in men and women, and deviations from this plasma proteomic clock are correlated with changes in clinical and functional parameters (Fig. 1f–h). This panel of 373 proteins can be used to assess the relative health of an individual and to measure healthspan, analogous to epigenetic clocks based on DNA methylation patterns³⁴. More large-scale plasma proteomic studies are required to establish the validity and utility of this clock and whether specific protein subsets are more appropriate to reflect particular clinical and functional parameters.

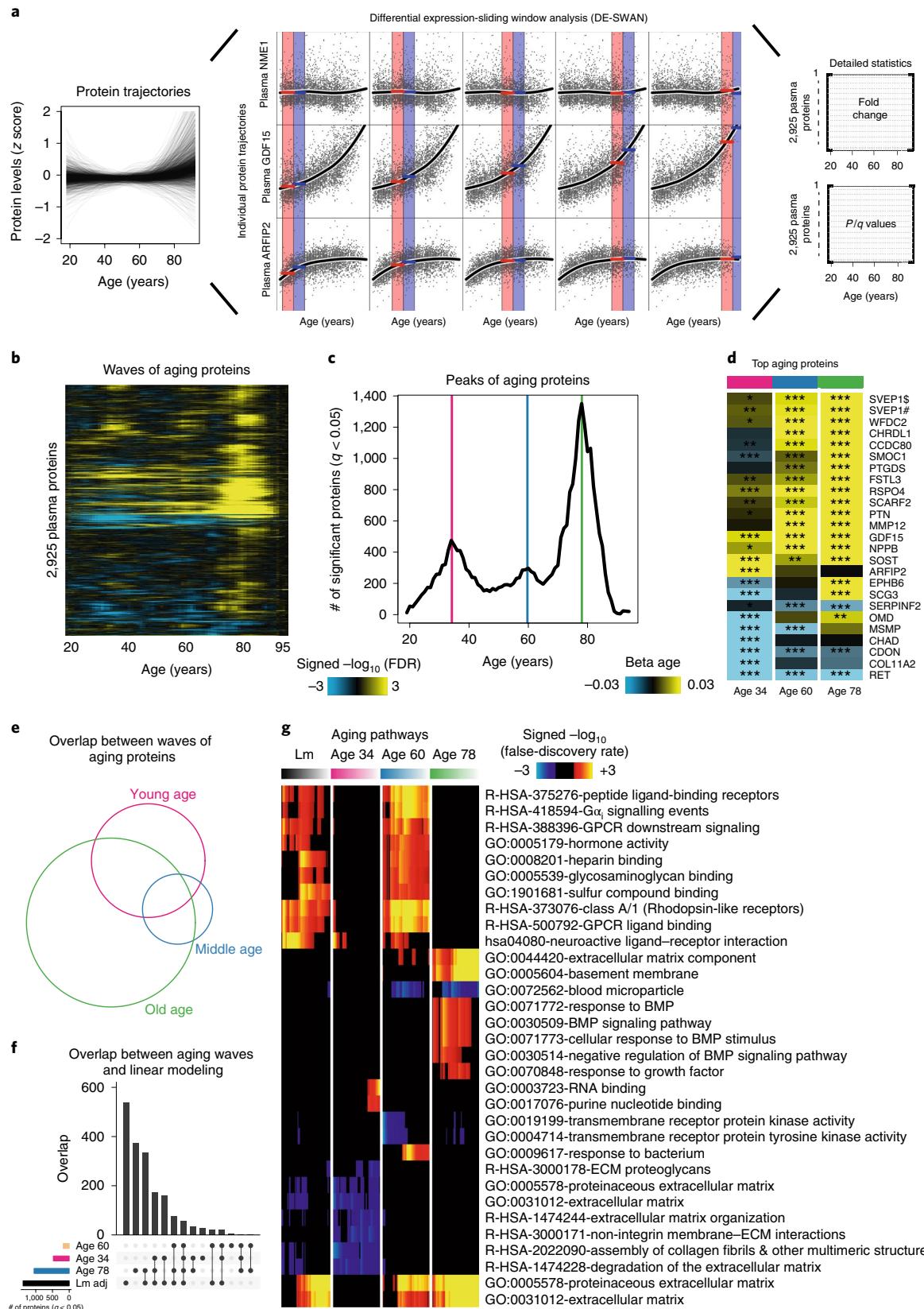
Blood is a sensitive marker of functional aging that also plays an active role in aging. Several studies have shown that soluble factors from young mouse blood reverse aspects of aging^{2–10,35}. Here we describe a 46-protein aging signature that was conserved in humans and mice, containing known aging-related proteins such as *GDF15*³⁶ and *IGF1-INSR*³⁷ but also less investigated ones (Fig. 1l). This conserved signature may allow deeper investigation of translational aging interventions in mice, such as heterochronic parabiosis, which partially reverses age-related changes of these proteins (Fig. 1k).

By deep mining the aging plasma proteome, we identified undulating changes during the human lifespan. These changes were the result of clusters of proteins moving in distinct patterns, culminating in the emergence of three waves of aging. Unexpectedly, we found that these clusters were often part of shared biological pathways, particularly cellular signaling (Fig. 2). In addition,

Fig. 3 | Sliding window analysis distinguishes waves of aging plasma factors. **a**, DE-SWAN. DE-SWAN compares protein levels between groups of individuals in parcels of 10 years (e.g., 30–40 years compared to 40–50 years). DE-SWAN identifies linear and non-linear changes during aging. Examples are shown of DE-SWAN for three proteins and five age windows. Red and blue rectangles show the two parcels, and the red and blue lines represent the mean within each parcel. DE-SWAN provides statistics for each age window and each plasma protein, allowing detailed analysis of plasma proteomic changes during aging. **b**, Waves of aging plasma proteins characterized by DE-SWAN ($n=4,263$). Within each window, $-\log_{10}$ (P values) and $-\log_{10}$ (q values) were estimated by linear modeling adjusted for age and sex, and significance was tested using the F -test. Local changes attributable to age were signed on the basis of the corresponding beta age. **c**, Number of plasma proteins differentially expressed during aging. Three local peaks at the ages of 34, 60 and 78 years were identified by DE-SWAN. **d**, Top ten plasma proteins identified by DE-SWAN at age 34, 60 and 78 years ($n=4,263$). Linear models adjusted for age, sex and subcohort were used, and significance was tested using the F -test. Blue and yellow represent local decrease and increase, respectively. # and \$ indicate different SOMAmers targeting the same protein. * $q < 0.05$, ** $q < 0.01$, *** $q < 0.001$. **e**, Intersections between waves of aging proteins ($n=4,263$; $q < 0.05$). Linear models adjusted for age, sex and subcohort were used, and significance was tested using the F -test. **f**, Intersections between linear modeling and the aging waves ($n=4,263$; $q < 0.05$). Linear models adjusted for age, sex and subcohort were used, and significance was tested using the F -test. **g**, Visualization of pathways significantly enriched for aging-related proteins identified by linear modeling and DE-SWAN at age 34, 60 and 78 years ($n=4,263$). Proteins upregulated and downregulated were analyzed separately. The top ten pathways per condition are represented. Enrichment was tested using Fisher's exact test (GO) and the hypergeometric test (Reactome and KEGG).

we provided biological relevance for the three main waves of aging-related proteins (Fig. 3), which are characterized by key biological pathways with little overlap. In comparison, linear modeling failed to identify changes occurring late in the eighth decade of

life (Fig. 3g). We conclude that linear modeling of aging based on omics data does not capture the complexity of biological aging across the lifespan. Thus, DE-SWAN will be invaluable for analyzing longitudinal datasets with linear and non-linear quantifiable



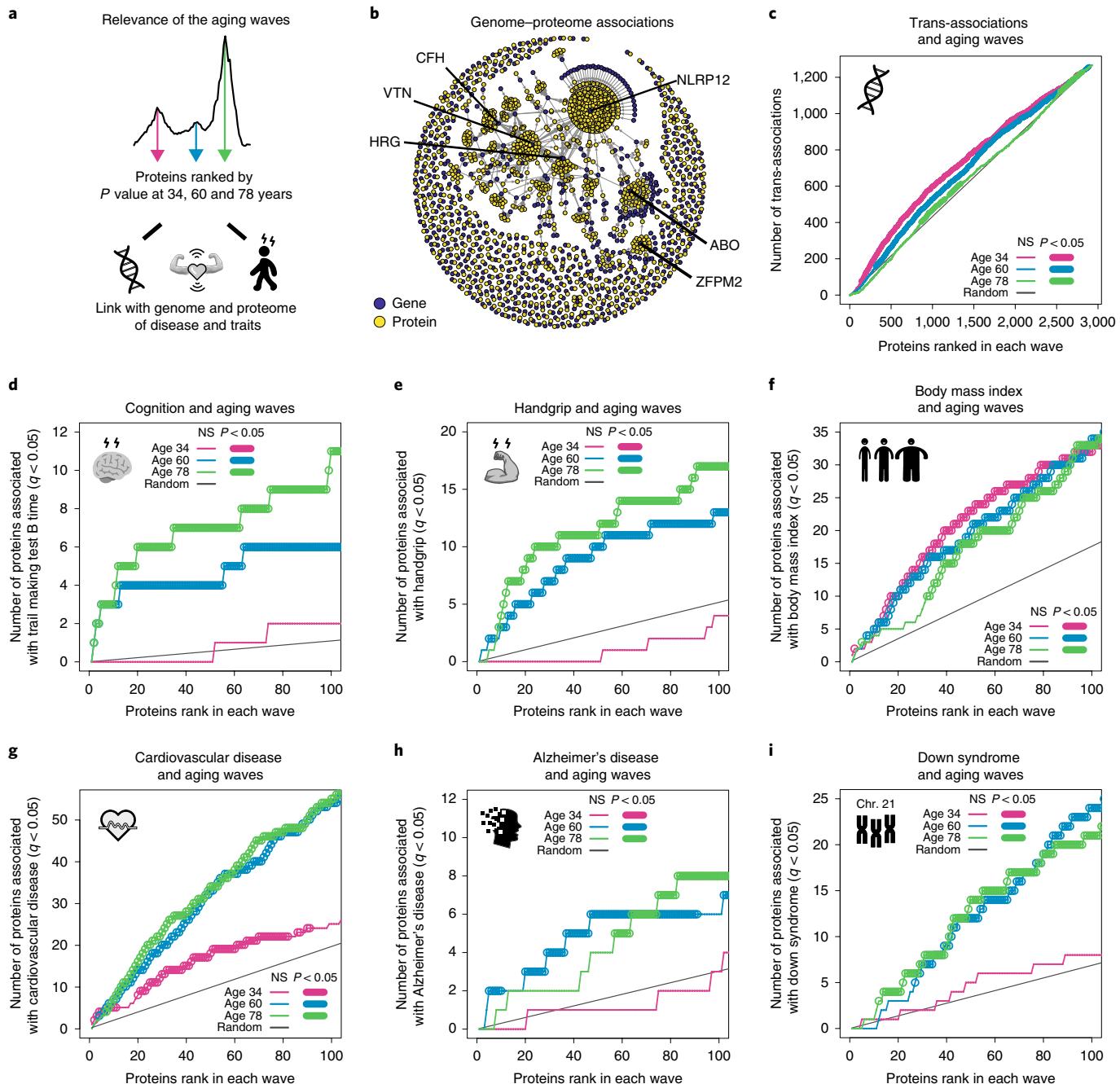


Fig. 4 | Waves of aging-related proteins are differentially linked to the genomes and proteomes of diseases and traits. **a**, Relevance of the aging waves. Schematic representation of analysis. The proteins changing at 34, 60 and 78 years were ranked by P value and were associated with the genome and proteome of diseases and traits. **b**, Association between the genome and the proteome. The network was created using the protein quantitative trait locus associations identified by Sun et al.²⁶. **c**, Enrichment for trans-association in the waves of aging-related proteins identified by DE-SWAN. Aging-related proteins at age 34, 60 and 78 years were ranked on the basis of P value, and the cumulative number of trans-associations was enumerated. One-sided permutation tests (1×10^5 permutations) were used to assess significance. **d,e**, Enrichment for proteins involved in cognitive (**d**) and physical (**e**) performance in the waves of aging-related proteins. **f-i**, Enrichment for disease-associated proteins in the waves of aging proteins, including for BMI (**f**), CVD (**g**), AD (**h**) and DS (**i**).

changes and for integrating non-linear changes in the analysis of omics datasets.

Sources of variation of the plasma proteome can be diverse and under genetic control^{26,38}. Intriguingly, we observed that the relative importance of trans-associations decreased with aging (Fig. 4c), which led us to investigate sources of variance with a focus on disease-associated proteomes. Proteins comprising the waves in middle and old age differentially overlapped with proteins associated

with cognitive and physical impairments. These proteins also discriminated patients from age-matched controls in AD, DS and CVD (Fig. 4g-i), suggesting that the characteristic plasma proteins of aging are amplified in these age-related diseases. Using an AD- and DS-free aging cohort, we provided evidence of accelerated aging for these two diseases^{32,33}. Further investigation of these proteins is warranted to determine whether these associations indicate aging biomarkers and/or causal mechanisms of disease. Nonetheless, these

results suggest that variance within the aging plasma proteome slowly transitions from hard coding factors (i.e., genomic) to soft coding factors (e.g., diseases, environmental factors and resulting changes in cognitive and physiological functions).

The undulating nature of the aging plasma proteome and its interactions with diseases should be considered when developing proteomic signatures for diagnostic purposes. Indeed, disease proteomes overlap significantly with the waves of aging proteins (Supplementary Table 15). Accounting for heterogeneous and complex changes to the plasma proteome during life will likely improve the sensitivity and specificity of prognostic and diagnostic tests. Moreover, these results are pertinent when considering the use of blood or blood products to treat aging and age-related diseases³⁹. Specifically, identifying plasma proteins that promote or antagonize aging at different stages of life could lead to more targeted therapeutics and/or preventative therapies. Such reliable tests and treatments are urgently needed for several diseases, and, in the future, we hope to describe plasma proteome changes that predict subjects transitioning to disease. Of particular interest are studies of AD, for which blood-based biomarkers are unavailable, and clinical symptoms are believed to occur up to two decades after disease onset.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-019-0673-2>.

Received: 18 January 2019; Accepted: 30 October 2019;

Published online: 5 December 2019

References

1. Harman, D. The aging process: major risk factor for disease and death. *Proc. Natl Acad. Sci. USA* **88**, 5360–5363 (1991).
2. Baht, G. S. et al. Exposure to a youthful circulation rejuvenates bone repair through modulation of β-catenin. *Nat. Commun.* **6**, 7131 (2015).
3. Conboy, I. M. et al. Rejuvenation of aged progenitor cells by exposure to a young systemic environment. *Nature* **433**, 760–764 (2005).
4. Huang, Q. et al. A young blood environment decreases aging of senile mice kidneys. *J. Gerontol. A Biol. Sci. Med. Sci.* **73**, 421–428 (2018).
5. Katsimpardi, L. et al. Vascular and neurogenic rejuvenation of the aging mouse brain by young systemic factors. *Science* **344**, 630–634 (2014).
6. Loffredo, F. S. et al. Growth differentiation factor 11 is a circulating factor that reverses age-related cardiac hypertrophy. *Cell* **153**, 828–839 (2013).
7. Salpeter, S. J. et al. Systemic regulation of the age-related decline of pancreatic beta-cell replication. *Diabetes* **62**, 2843–2848 (2013).
8. Sinha, M. et al. Restoring systemic GDF11 levels reverses age-related dysfunction in mouse skeletal muscle. *Science* **344**, 649–652 (2014).
9. Villeda, S. A. et al. The ageing systemic milieu negatively regulates neurogenesis and cognitive function. *Nature* **477**, 90–94 (2011).
10. Villeda, S. A. et al. Young blood reverses age-related impairments in cognitive function and synaptic plasticity in mice. *Nat. Med.* **20**, 659–663 (2014).
11. Valdes, A. M., Glass, D. & Spector, T. D. Omics technologies and the study of human ageing. *Nat. Rev. Genet.* **14**, 601–607 (2013).
12. Stegeman, R. & Weake, V. M. Transcriptional signatures of aging. *J. Mol. Biol.* **429**, 2427–2437 (2017).
13. Aramillo Irizar, P. et al. Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. *Nat. Commun.* **9**, 327 (2018).
14. Castellano, J. M. et al. Human umbilical cord plasma proteins revitalise hippocampal function in aged mice. *Nature* **544**, 488–492 (2017).
15. Di Angelantonio, E. et al. Efficiency and safety of varying the frequency of whole blood donation (INTERVAL): a randomised trial of 45 000 donors. *Lancet* **390**, 2360–2371 (2017).
16. Gubbi, S. et al. Effect of exceptional parental longevity and lifestyle factors on prevalence of cardiovascular disease in offspring. *Am. J. Cardiol.* **120**, 2170–2175 (2017).
17. Zhou, J. & Rossi, J. Aptamers as targeted therapeutics: current potential and challenges. *Nat. Rev. Drug Discov.* **16**, 440 (2017).
18. Emilsson, V. et al. Co-regulatory networks of human serum proteins link genetics to disease. *Science* **361**, 769–773 (2018).
19. Gold, L. et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* **5**, e15004 (2010).
20. Austad, S. N. & Fischer, K. E. Sex differences in lifespan. *Cell Metab.* **23**, 1022–1033 (2016).
21. Ostan, R. et al. Gender, aging and longevity in humans: an update of an intriguing/neglected scenario paving the way to a gender-specific medicine. *Clin. Sci.* **130**, 1711–1725 (2016).
22. Tanaka, T. et al. Plasma proteomic signature of age in healthy humans. *Aging Cell* **17**, e12799 (2018).
23. Cohen, A. A. Aging across the tree of life: the importance of a comparative perspective for the use of animal models in aging. *Biochim. Biophys. Acta. Mol. Basis Dis.* **1864**, 2680–2689 (2018).
24. Guiraud, S. et al. Identification of serum protein biomarkers for utrophin based DMD therapy. *Sci. Rep.* **7**, 43697 (2017).
25. Wang, R. N. et al. Bone morphogenetic protein (BMP) signaling in development and human diseases. *Genes Dis.* **1**, 87–105 (2014).
26. Sun, B. B. et al. Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
27. Sattler, M. et al. Alzheimer's disease biomarker discovery using SOMAscan multiplexed protein technology. *Alzheimers Dement.* **10**, 724–734 (2014).
28. Sullivan, K. D. et al. Trisomy 21 causes changes in the circulating proteome indicative of chronic autoinflammation. *Sci. Rep.* **7**, 14818 (2017).
29. Ganz, P. et al. Development and validation of a protein-based risk score for cardiovascular outcomes among patients with stable coronary heart disease. *JAMA* **315**, 2532–2541 (2016).
30. Carayol, J. et al. Protein quantitative trait locus study in obesity during weight-loss identifies a leptin regulator. *Nat. Commun.* **8**, 2084 (2017).
31. Go, A. S. et al. Heart disease and stroke statistics—2013 update: a report from the American Heart Association. *Circulation* **127**, e6–e245 (2013).
32. Franceschi, C., Garagnani, P., Parini, P., Giuliani, C. & Santoro, A. Inflammaging: a new immune-metabolic viewpoint for age-related diseases. *Nat. Rev. Endocrinol.* **14**, 576–590 (2018).
33. Franceschi, C. et al. The continuum of aging and age-related diseases: common mechanisms but different rates. *Front. Med.* **5**, 61 (2018).
34. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* **19**, 371–384 (2018).
35. Castellano, J. M., Kirby, E. D. & Wyss-Coray, T. Blood-borne revitalization of the aged brain. *JAMA Neurol.* **72**, 1191–1194 (2015).
36. Wiklund, F. E. et al. Macrophage inhibitory cytokine-1 (MIC-1/GDF15): a new marker of all-cause mortality. *Aging Cell* **9**, 1057–1064 (2010).
37. Cohen, E. & Dillin, A. The insulin paradox: aging, proteotoxicity and neurodegeneration. *Nat. Rev. Neurosci.* **9**, 759–767 (2008).
38. Suhre, K. et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* **8**, 14357 (2017).
39. Sha, S. J. et al. Safety, tolerability, and feasibility of young plasma infusion in the plasma for Alzheimer symptom amelioration study: a randomized clinical trial. *JAMA Neurol.* **76**, 35–40 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

Methods

Plasma proteomics measurements. The SomaScan platform was used to quantify relative levels of protein involved in several processes, such as intercellular signaling, extracellular proteolysis and metabolism. This platform was established to identify biomarker signatures of diseases and conditions, including cardiovascular risk²⁹, cancer⁴⁰ and neurodegenerative diseases²⁷. The SomaScan platform is based on modified single-stranded DNA aptamers (SOMAmer reagents) binding to specific protein targets. Assay details were previously described¹⁹. Different versions of the SomaScan assay were used in the LonGenity, INTERVAL and four independent human cohorts. These versions contained 5,284, 4,034 and 1,305 aptamers, respectively.

Of the 4,034 aptamers measured in the INTERVAL cohort, 3,283 were contained in the publicly available dataset (European Genome–Phenome Archive EGAS00001002555). Our study focused on 2,925 aptamers with identical SeqID and SeqIDVersion in both INTERVAL and LonGenity cohorts (Supplementary Table 1). Of the 2,925 aptamers, 888 were measured in the four independent cohorts and in mice (Supplementary Table 2).

Human cohort characteristics. *INTERVAL cohort.* Participants in the INTERVAL randomized controlled trial (ISRCTN24760606) were recruited with the active collaboration of the National Health Service (NHS) Blood and Transplant (<http://www.nhsbt.nhs.uk>), which has supported field work and other elements of the trial. DNA extraction and genotyping were co-funded by the National Institute for Health Research (NIHR), the NIHR BioResource (<http://bioresource.nihr.ac.uk/>) and the NIHR Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust. The INTERVAL study was funded by NHS Blood and Transplant (11-01-GEN). The academic coordinating center for INTERVAL was supported by core funding from the NIHR Blood and Transplant Research Unit in Donor Health and Genomics (NIHR BTRU-2014-10024), the UK Medical Research Council (MR/L003120/1), the British Heart Foundation (RG/13/13/30194) and the NIHR Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust. Proteomic assays were funded by the academic coordinating center for INTERVAL and Merck Research Laboratories (Merck & Co.). A complete list of the investigators and contributors to the INTERVAL trial was previously reported¹⁵. The academic coordinating center would like to thank blood donor center staff and blood donors for participating in the INTERVAL trial. For more information, see the Nature Research Reporting Summary.

Proteomics measurements from 3,301 human plasma samples (1,685 males and 1,616 females) from two different subcohorts were used for this study. Age ranged from 18 to 76 years with a median age of 45 years (first quartile = 31; third quartile = 55). Sample selection, processing and preparation were detailed previously²⁶.

LonGenity cohort. LonGenity is an ongoing longitudinal study initiated in 2008 and designed to identify biological factors that contribute to healthy aging¹⁶. The LonGenity study enrolls older adults of Ashkenazi Jewish descent with age 65–94 years at baseline. Approximately 50% of the cohort consists of offspring of parents with exceptional longevity, defined as having at least one parent who survived to 95 years of age. The other half of the cohort includes offspring of parents with usual survival, defined as not having a parental history of exceptional longevity. Proteomics measurements from 1,030 human plasma samples (457 males and 573 females) collected at baseline in LonGenity participants were used for this study. Age ranged from 61 to 95 years with a median age of 74 years (first quartile = 69; third quartile = 80). LonGenity participants are thoroughly characterized demographically and phenotypically at annual visits that include collection of medical history and physical and neurocognitive assessments. Sixty-eight individuals without clinical and functional data were excluded from the analysis. The LonGenity study was approved by the institutional review board (IRB) at the Albert Einstein College of Medicine.

Four additional independent cohorts. One hundred seventy-one human plasma samples (84 males and 87 females) were obtained from four different cohorts from the US and Europe (VASeattle, PRIN06, PRIN09 and GEHA). Sample selection, processing and preparation of the VASeattle cohort were detailed previously⁴¹. Participants from the PRIN06, PRIN09 and GEHA⁴² cohorts were enrolled by multiple Italian study centers. Participants were mainly of European ancestry. Age ranged from 21 to 107 years with a median age of 70 years (first quartile = 58; third quartile = 89). Written informed consent was obtained from each subject. The IRB determined that our research did not meet the definition of human subject research per Stanford's Human Research Protection Program policy, and no IRB approval was required for this study.

For these cohorts, all samples were stored at -80°C , and 150 μl aliquots of plasma were sent on dry ice to SomaLogic. Plasma samples were analyzed in three different batches: 24 samples in 2015, 70 samples in 2016 and 77 samples in 2017. In addition to these 171 plasma samples, 12 additional aliquots from 4 of these samples were measured in the different batches to estimate intra- and inter-assay variability (Supplementary Table 3). Data for 1,305 SOMAmer probes were obtained. No sample or probe data were excluded. HybNorm.plateScale.medNorm

files provided by SomaLogic were bridged to data from the first batch of samples using calibrators.

Normalization of INTERVAL and LonGenity datasets. Relative fluorescence units (RFUs) of each plasma protein were \log_{10} transformed. We normalized the levels of each protein within each subcohort on the basis of the average of the subjects in the 60- to 70-year range. Supplementary Figure 3 shows representative normalization examples. Note that this normalization is needed when fitting aging trajectories (Fig. 2) but does not affect the results when 'subcohort' is included as a covariate in the modeling.

The data from the four independent cohorts were \log_{10} transformed and bridged together using the SomaLogic procedure on the basis of calibrators. However, the number of samples in the 60- to 70-year range was too small to reliably bridge these data to the INTERVAL and LonGenity cohorts.

Linear changes in the aging plasma proteome. To determine the effect of age and sex at the protein level, we used the following linear model:

$$\text{Protein level} \sim \alpha + \beta_1 \text{age} + \beta_2 \text{sex} + \beta_3 \text{subcohort} + \epsilon$$

The type II sum of squares was calculated using the ANOVA function of the R car package⁴³. This sum of squares type tests for each main effect after the other main effects. q values were estimated using the Benjamini–Hochberg approach⁴⁴. It should be noted that the age range differed between cohorts. If the adjustment for cohort effect decreases the number of false positives, it could also alter the true-positive rate. In the four independent cohorts, the 'subcohort' covariate also accounted for batch effect, as samples from different cohorts were measured in different batches (except for PRIN06 and GEHA, which were measured together).

To determine the relative proportion of variance explained by age and sex, we calculated the partial Eta² as follows:

$$\text{Partial Eta}^2 = \frac{\text{Sum of squares}_{\text{effect}}}{(\text{Sum of squares}_{\text{effect}} + \text{Sum of squares}_{\text{error}})}$$

Validation of the aging-related proteins. To provide confidence in the reproducibility of the protein assays, we compared our findings with the associations with age reported by Tanaka et al.²². To this end, we merged our results with those from Tanaka et al. using 'Somalid'. Of note, Tanaka et al. used the same version of the SomaScan platform that we used for the four independent cohorts (1,305 proteins).

SEPA. To determine the biological meaning of a group of plasma proteins, we ranked the top 100 proteins on the basis of the product of $-\log_{10}(P \text{ values})$ and beta age (or beta sex) and queried three of the most comprehensive biological annotation and pathway databases: GO⁴⁵, KEGG⁴⁶ and Reactome⁴⁷. Using these databases, we tested enrichment for pathways in the top 10 to top 100 proteins in increments of 1 protein. The 2,925 proteins measured in this study cover 90% of the human GO, Reactome and KEGG terms containing more than eight genes (Supplementary Figure 4).

To analyze each incremental list of proteins, we used the R topGO package⁴⁸ for GO analysis and the R clusterProfiler package⁴⁹ for KEGG and Reactome analyses. As input for SEPA, we used gene symbols provided by SomaLogic (Supplementary Table 1). The 2,925 proteins measured by SomaScan served as the background set of proteins against which to test for over-representation. Because several individual proteins (33 of 2,925) were mapped to multiple gene symbols, we kept only the first gene symbol provided by SomaLogic to prevent false-positive enrichment. For KEGG and Reactome analysis, clusterProfiler requires EntrezID as input. Therefore, we mapped gene symbols to EntrezID using the org.Hs.eg.db package⁵⁰. Again, to avoid false-positive enrichment, only the first EntrezID was used when gene symbols were mapped to multiple EntrezID. q values were estimated using the Benjamini–Hochberg approach⁴⁴ for the different databases taken separately. For GO analysis, q values were calculated for the three GO classes (molecular function, cellular component and biological process) independently. To identify the most biologically meaningful terms and pathways, we reported only those with 20–500 proteins measured by the SomaScan assay. In addition, we focused on pathways that were consistently highly significant ($q < 0.05$ for at least 20 different incremental lists of proteins) and kept the top ten pathways per condition (e.g., for each wave of aging proteins). Ranking was performed on the basis of the minimum false-discovery rate across the incremental lists of proteins. SEPA can be viewed as an extension of the gene set enrichment analysis approach⁵¹, with more control for true and false positives.

Validation of the aging signature in mice. Male and virgin female C57BL/6JN mice were shipped from the National Institute on Aging colony at Charles River (housed at 67–73 °F) to the Veterinary Medical Unit (VMU; housed at 68–76 °F) at the VA Palo Alto (VA). At both locations, mice were housed on a 12-h light/dark cycle and provided food and water ad libitum. The diet at Charles River was NIH-31; the diet at the VA VMU was Teklad 2918. Littermates were not recorded or tracked. Mice that were 18 months old and younger were housed at the VA VMU for no longer than 2 weeks before being killed; mice older than 18 months

were housed at the VA VMU until they reached the experimental age. After anaesthetization with 2.5% vol/vol Avertin, blood was drawn by cardiac puncture. All animal care and procedures were carried out in accordance with institutional guidelines approved by the VA Palo Alto Committee on Animal Research.

Heterochronic parabiosis was conducted as previously described^{3,10,52} with 3- and 18-month-old mice. Briefly, incisions in the flank were made through the skin and peritoneal cavity of both sets of mice, and adjacent peritoneal cavities were sutured together. Adjacent knee and elbow joints were then sutured together to facilitate coordinated locomotion. Skin was then stapled together using surgical autoclips (9 mm; Clay Adams), and mice were placed under heat lamps to recover from anesthesia. Each individual mouse was injected subcutaneously with Baytril antibiotic (5 µg g⁻¹) and buprenorphine (0.05–0.1 µg ml⁻¹ in phosphate-buffered saline) for pain management and 0.9% (wt/vol) NaCl for hydration. Mice were monitored and administered drugs and saline over the next week as previously described.

EDTA-treated plasma was isolated by centrifugation at 1,000 g for 10 min at 4 °C before aliquotting and storing at –80°C. A total of 110 plasma samples were analyzed, and aliquots of 150 µl of plasma were sent on dry ice to SomaLogic. Samples were sent in two different batches: 29 samples in 2016 and 81 samples in 2018. Data for 1,305 SOMAmer probes were obtained, and no sample or probe data were excluded. RFUs of each plasma protein were log₁₀ transformed.

The SomaScan assay was developed and validated for human fluids but has been successfully used in mouse research^{2,4}. To understand how similar mouse and human sequences are, we downloaded all homologies between mouse and human along with sequence identifiers for each species (HOM_MouseHumanSequence.rpt) from Mouse Genome Informatics (<http://www.informatics.jax.org/>) as plain text files. Then, the protein reference sequences for both organisms were extracted from UniProt (<https://www.uniprot.org/>). On these matched sequence pairs, for each protein we computed a global pairwise sequence alignment. The alignments were calculated by using the R 'Biostings' library⁵³. The average identity was 0.85, supporting the use of the SomaScan assay with mouse plasma.

To determine the effect of age and sex at the protein level, we used the 81 samples from 1 month to 30 months. To this end, we fitted the following linear model:

$$\text{Protein level} \sim \alpha + \beta_1 \text{age} + \beta_2 \text{sex} + \epsilon$$

The type II sum of squares was calculated, and *q* values were estimated using the Benjamini–Hochberg approach.

To characterize the effects of young and old blood on the aging plasma proteome, normed scaled principal-component analysis was performed using the R ade4 package⁵⁴.

Prediction of human biological age using the plasma proteome. To determine whether the plasma proteome could predict biological age, we used glmnet⁵⁵ and fitted a LASSO model (alpha = 1; 100 lambda tested; 'lambda.min' as the shrinkage variable was estimated after tenfold cross-validation). Input variables consisted of *z*-scaled log₁₀-transformed RFUs and sex information. Two-thirds (*n* = 2,817) of the INTERVAL and LonGenity cohorts were used for training the model, and the remaining 1,446 samples were used as a validation. In addition, the 171 samples from the four independent cohorts were used to further assess the robustness of the predictive model.

To estimate whether a subset of proteins in the aging clock could provide similar predictive results, we used a two-step approach that we described previously⁵⁶. One hundred models (100 lambda) including 0–373 proteins were created in step 1, and we estimated the accuracy of each of these models on the discovery and validation datasets, separately. Broken-stick regression was used to determine the best compromise between the number of variables and prediction accuracy.

Associations between delta age and clinical and functional variables in old age. We used the individuals from the LonGenity cohort to identify associations between deviations from the proteomic clock (delta age = predicted age – chronological age) and 334 clinical and functional variables (Supplementary Table 8). To this end, we tested the following linear model:

$$\text{Variable of interest} \sim \alpha + \beta_1 \text{age} + \beta_2 \text{age} + \beta_3 \text{sex} + \epsilon$$

For binary outcomes, logistic regression was used. This analysis was separately performed in the discovery (*n* = 2,817) and validation (*n* = 1,446) cohorts. Type II sum of squares were calculated using the ANOVA function of the R car package⁴³.

Clustering of protein trajectories. To estimate protein trajectories during aging, plasma protein levels were *z* scored, and locally estimated scatterplot smoothing (LOESS) regression was fitted for each plasma factor. To group proteins with similar trajectories, pairwise differences between LOESS estimates were calculated on the basis of the Euclidian distance, and hierarchical clustering was performed using the complete method. To understand the biological functions of each cluster, we queried the Reactome, KEGG and GO databases, as described above.

DE-SWAN. To identify and quantify linear and non-linear changes of the plasma proteome during aging, we developed the DE-SWAN approach. Considering a vector *l* of *k* unique ages, we iteratively used *l_k* as the center of a 20-year

window and compared protein levels of individuals in parcels below and above *l_k* (i.e., [*l_k* – 10y; *l_k*; *l_k* + 10y]). To test for differential expression, we used the following linear model:

$$\text{Protein level} \sim \alpha + \beta_1 \text{age}_{\text{Low}/\text{High}} + \beta_2 \text{sex} + \epsilon$$

with age binarized according to the parcels. For each *l_k*, *q* values were estimated using the Benjamini–Hochberg correction. The type II sum of squares was calculated using the ANOVA function of the R car package⁴³.

To assess the robustness and relevance of DE-SWAN results, we tested multiple parcel widths (5, 10, 15 and 20 years). In addition, we used multiple *q*-value thresholds and compared these results with those obtained by chance. To this end, we randomly permuted the phenotypes of the individuals and applied DE-SWAN to this new dataset. To keep the data structure, age and sex were permuted together. In addition, we analyzed the INTERVAL and LonGenity cohorts separately (Supplementary Fig. 1). Finally, we tested the same linear model when adjusting for subcohort. This led to a loss of statistical power when the age range of the INTERVAL and LonGenity cohorts overlapped, but the three waves of aging-related proteins remained and the ranks of the top proteins were nearly identical (Supplementary Fig. 1). We used the model adjusted for subcohort when trying to understand the waves of aging proteins (Figs. 3d–g and 4). The significance levels of the intersections between aging plasma protein signatures identified by linear modeling and DE-SWAN at different ages were determined using the R SuperExactTest package⁵⁷.

Relationships between the aging waves and the genome and the proteome of diseases and traits. To quantify the overlap between proteins changing with age at different stages of life and the genome and the proteome of diseases and traits, we ranked DE-SWAN results on the basis of *P* values and created a *k*-ranked list of aging-related proteins, *L_k*. To reflect the degree to which the genome or proteome is linked to the waves of aging plasma proteins, we walked down *L_k* and counted the number of proteins associated with the genome or specific proteome. When different versions of the SomaScan platform were used, we walked down *L_k* until reaching the top 100 proteins measured in both studies.

To identify specific genetic variants associated with the aging plasma proteome, we mined the summary statistics generated by Sun et al.²⁶, who found 1,927 associations with 1,104 plasma proteins. The Qgraph⁵⁸ R package was used to create a network between the genome and the 2,925 proteins analyzed in this study.

To determine whether the aging proteome was associated with the proteome of clinical and functional variables, we used the individuals from the LonGenity cohort and tested the following linear model for the top variables identified in Supplementary Table 8:

$$\text{Protein level} \sim \alpha + \beta_1 \text{age} + \beta_2 \text{sex} + \beta_3 \text{variable of interest} + \epsilon$$

The type II sum of squares was calculated using the ANOVA function of the R car package⁴³.

To determine whether the aging proteome was associated with disease proteomes, we integrated data and results from previous proteomic studies using the SomaScan platform. We re-analyzed one AD dataset publicly available by AddNeuroMed²⁷ and used summary statistics from published studies focused on CVD²⁹, DS²⁸ and BMI³⁰.

AddNeuroMed is a European multi-center study in which the AD proteome was quantified in plasma samples from 681 control individuals and individuals with mild cognitive impairment (MCI) and AD using a previous version of the SomaScan assay. Files used were downloaded from the Synapse portal in March 2016 (syn5367752) and included measurements of 1,016 plasma proteins from 931 samples. We limited our analysis to the 645 samples available at visit 1 (191 control, 165 MCI and 289 AD). Raw data were log₁₀ transformed. Four samples (two control and two AD) were considered as outliers on the basis of visual inspection of the results of a principal-component analysis and filtered out.

To identify plasma proteins associated with AD, we used linear models with diagnosis, age, sex and center as covariates:

$$\text{Protein level} \sim \alpha + \beta_1 \text{diagnosis} + \beta_2 \text{age} + \beta_3 \text{sex} + \beta_4 \text{center} + \epsilon$$

The type II sum of squares was calculated using the ANOVA function of the R car package⁴³.

To determine whether aging plasma proteins were involved in other disease signatures, we identified three studies using the SomaScan platform in large human cohorts providing detailed summary statistics. Carayol et al. mined the plasma proteome to obtain new insights into the molecular mechanisms of obesity. Of 1,129 proteins measured, they identified 192 plasma proteins significantly associated with BMI (*P* < 0.05 after Bonferroni correction). Summary statistics we used were obtained from Supplementary Data 1 of their publication³⁰. Ganz et al. derived and validated a nine-protein risk score to predict risk of cardiovascular outcomes³⁹. In addition to these 9 proteins, 191 other proteins were significantly associated with cardiovascular risk (*P* < 0.05 after Bonferroni correction). Summary statistics for these 200 proteins are available in their eTable 4, and the 1,130 proteins measured in this study are listed in their eTable 1. Finally, Sullivan et al. used an

extended version of the SomaScan platform to study DS and identified a large number of dysregulated proteins²⁸. We used the results for the discovery cohort (sheet A of Supplementary File 1) in which 258 of 3,586 proteins were reported to be associated with DS ($P < 0.05$ after Bonferroni correction).

Because detailed protein information was not available for all studies, we used either gene symbols or uniprotID to merge disease proteomes characterized in published studies with the aging proteome identified in this study. When multiple P values were reported for the same gene symbol (or a combination of gene symbols), only the most significant P value was retained.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

We created a searchable web interface to mine the human INTERVAL and LonGenity datasets: https://tvc-stanford.shinyapps.io/aging_plasma_proteome/. The independent human cohorts and mouse protein data are available in Supplementary Tables 16 and 17. The INTERVAL data are available through the European Genome–Phenome Archive under accession EGAS00001002555.

Code availability

An R package for DE-SWAN is available in GitHub: <http://lehallib.github.io/DEswan/>.

References

40. Mehan, M. R. et al. Protein signature of lung cancer tissues. *PLoS One* **7**, e35157 (2012).
41. Britschgi, M. et al. Modeling of pathological traits in Alzheimer's disease based on systemic extracellular signaling proteome. *Mol. Cell Proteomics* **10**, M111 008862 (2011).
42. Franceschi, C. et al. Genetics of healthy aging in Europe: the EU-integrated project GEHA (GEnetics of Healthy Aging). *Ann. NY Acad. Sci.* **1100**, 21–45 (2007).
43. Fox, J. & Weisberg, S. *An R Companion to Applied Regression* (SAGE Publications, 2011).
44. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* **57**, 289–300 (1995).
45. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
46. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
47. Croft, D. et al. The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42**, D472–D477 (2014).
48. Alexa, A. & Rahnenfuhrer, J. topGO: enrichment analysis for Gene Ontology. <https://doi.org/10.18129/B9.bioc.topGO> (2016).
49. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
50. Carlson, M. org.Hs.eg.db: genome wide annotation for human. <https://doi.org/10.18129/B9.bioc.Hs.eg.db> (2017).
51. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
52. Castellano, J. M. et al. In vivo assessment of behavioral recovery and circulatory exchange in the peritoneal parabiosis model. *Sci. Rep.* **6**, 29015 (2016).
53. Pagès, H., Aboyoun, P., Gentleman, R. & DebRoy, S. Biostrings: efficient manipulation of biological strings. <https://doi.org/10.18129/B9.bioc.Biostrings> (2019).
54. Dray, S. & Dufour, A. B. The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.* **22**, 1–20 (2007).
55. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
56. Lehallier, B. et al. Combined plasma and cerebrospinal fluid signature for the prediction of midterm progression from mild cognitive impairment to Alzheimer disease. *JAMA Neurol.* **73**, 203–212 (2016).
57. Wang, M., Zhao, Y. & Zhang, B. Efficient test and visualization of multi-set intersections. *Sci. Rep.* **5**, 16923 (2015).
58. Epskamp, S., Cramer, A., Waldorp, L., Schmittmann, V. & Borsboom, D. qgraph: network visualizations of relationships in psychometric data. *J. Stat. Softw.* **48**, 1–18 (2012).

Acknowledgements

We thank the members of the Wyss-Coray laboratory for feedback and support. We thank the clinical staff for human blood and plasma collection/coordination. We thank A. Butterworth for his help in getting access to the INTERVAL proteomics data. The AddNeuroMed data are from a public–private partnership supported by EFPIA companies and the European Union Sixth Framework program priority FP6-2004-LIFESCIHEALTH-5. Clinical leads responsible for data collection were I. Kloszewska (Lodz), S. Lovestone (London), P. Mecocci (Perugia), H. Soininen (Kuopio), M. Tsolaki (Thessaloniki) and B. Vellas (Toulouse); imaging leads were A. Simmons (London), L.O. Wahlund (Stockholm) and C. Spenger (Zurich); and bioinformatics leads were R. Dobson (London) and S. Newhouse (London). This work was supported by National Institutes of Health National Institute on Aging (NIA) F32 1F32AG055255 01A1 (D.G.), Hungarian Brain Research Program Grant No. 2017-1.2.1-NKP-2017-00002 (T.N.), the Fulbright Foreign Student Program (T.N.), the Cure Alzheimer's Fund (T.W.-C.), Nan Fung Life Sciences (T.W.-C.), the NOMIS Foundation (T.W.-C.), the Stanford Brain Rejuvenation Project (an initiative of the Stanford Wu Tsai Neurosciences Institute), the Paul F. Glenn Center for Aging Research (T.W.-C.), NIA R01 AG04503 and DP1 AG053015 (T.W.-C.) and the NIA-funded Stanford Alzheimer's Disease Research Center P50AG047366, NIA K23AG051148 (S.M.), R01AG061155 (S.M.), the American Federation for Aging Research (S.M.), R01AG044829 (J.V. and N.B.), NIA R01AG057909 (N.B.), the Nathan Shock Center of Excellence for the Basic Biology of Aging P30AG038072 (N.B.) and the Glenn Center for the Biology of Human Aging (N.B.).

Author contributions

B.L. and T.W.-C. planned the study. D.B., C.F., S.M., J.V., S.S. and N.B. provided human plasma samples. N.S., S.E.L. and H.Y. performed the mouse experiments. B.L. analyzed the data, with contributions from T.N. and A.K. P.M.L. developed the searchable web interface (shiny app). B.L., D.G. and T.W.-C. wrote the manuscript. A.K., C.F., S.M., J.V., S.S., N.B. and T.W.-C. supervised the study. All authors edited and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

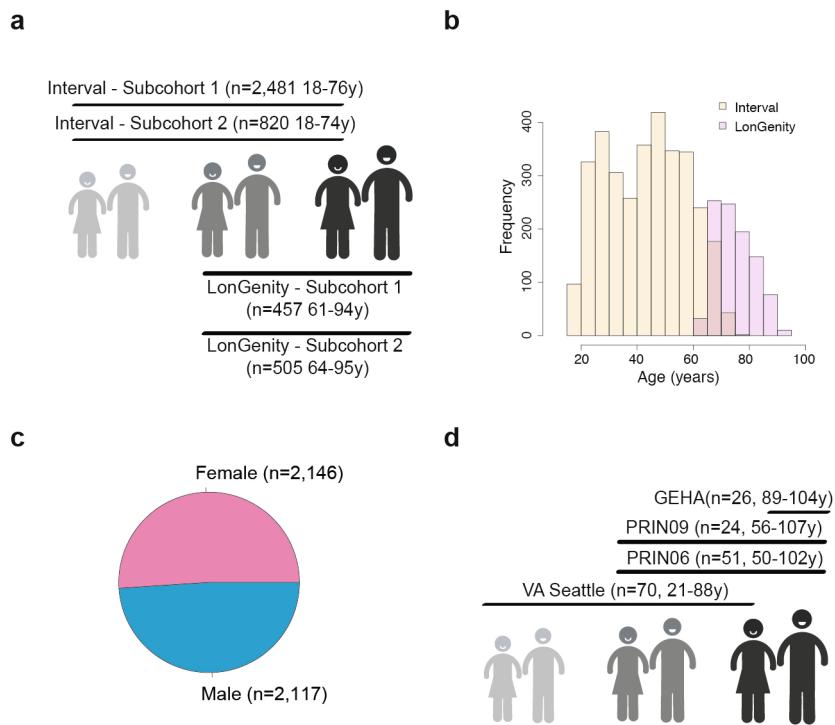
Extended data is available for this paper at <https://doi.org/10.1038/s41591-019-0673-2>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41591-019-0673-2>.

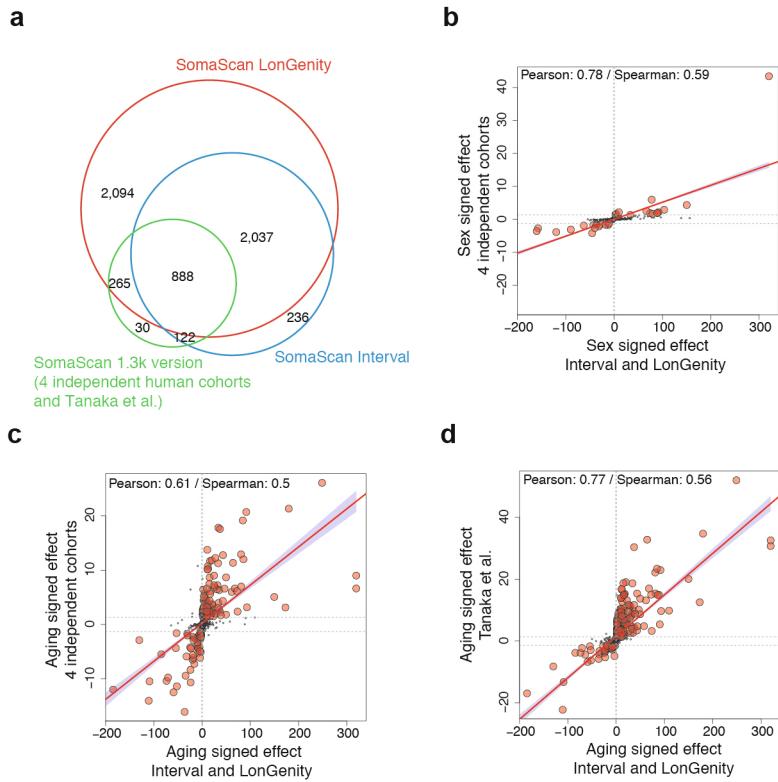
Correspondence and requests for materials should be addressed to B.L. or T.W.-C.

Peer review information Brett Bennedetti and Jennifer Sargent were the primary editors on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

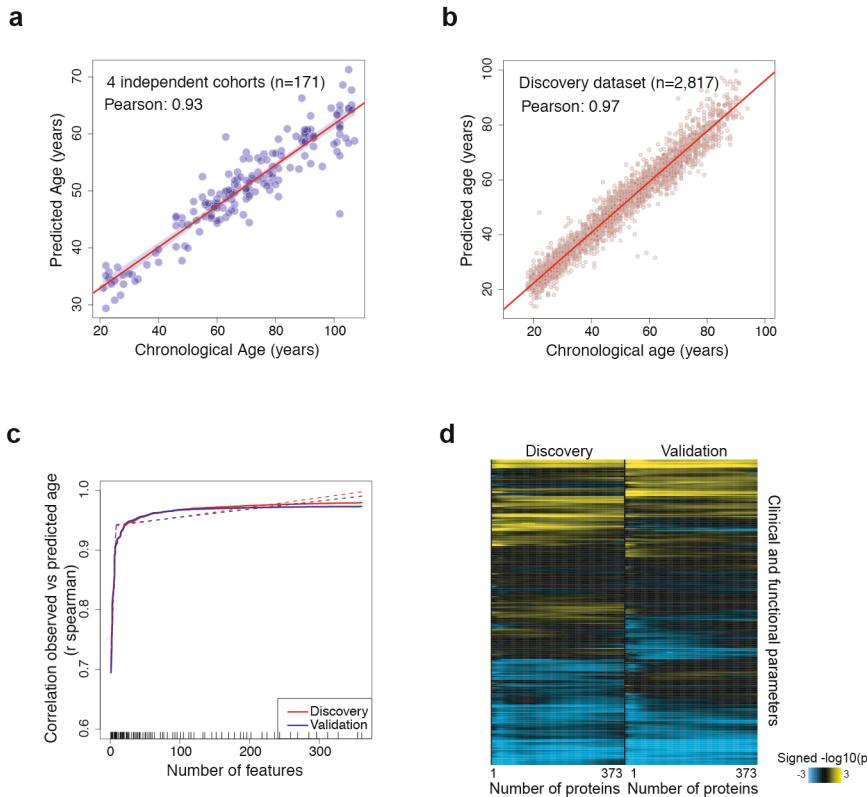
Reprints and permissions information is available at www.nature.com/reprints.



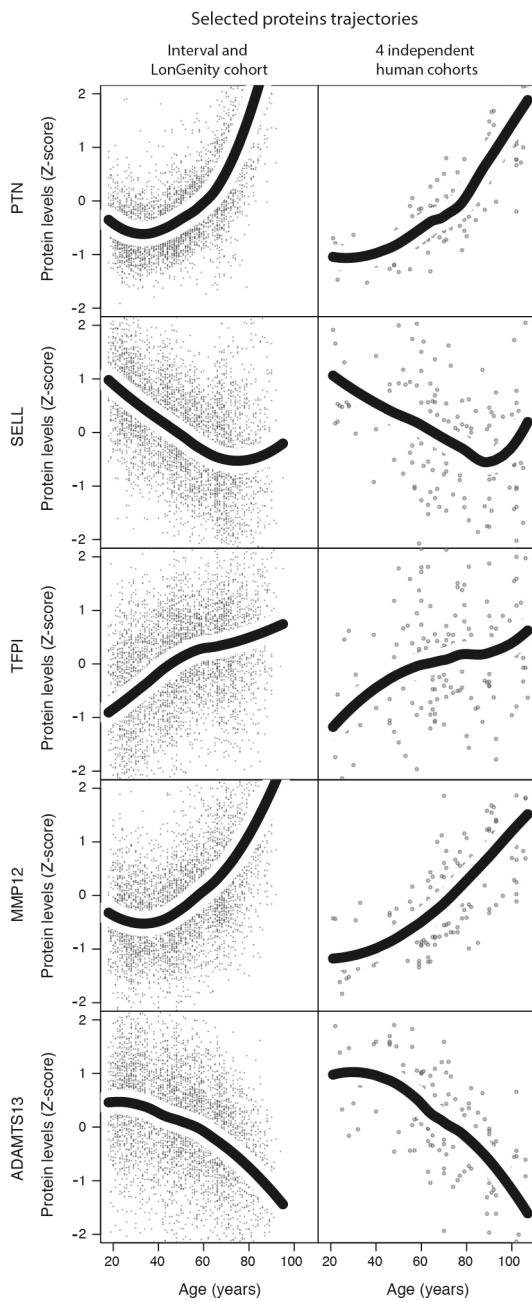
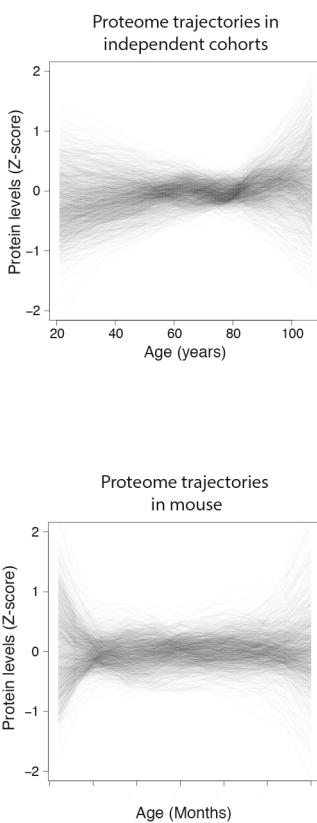
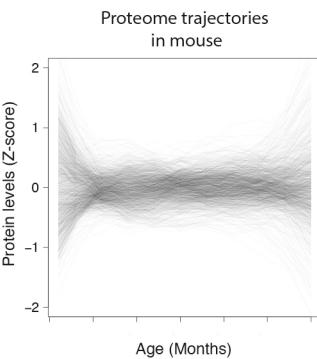
Extended Data Fig. 1 | Sample demographics. Age (**a, b**), cohort (**a, b**) and sex distributions (**c**) of the 4,263 subjects from the INTERVAL and LonGenity cohorts. (**d**) Age and cohort distributions of the 171 subjects from the 4 independent cohorts.



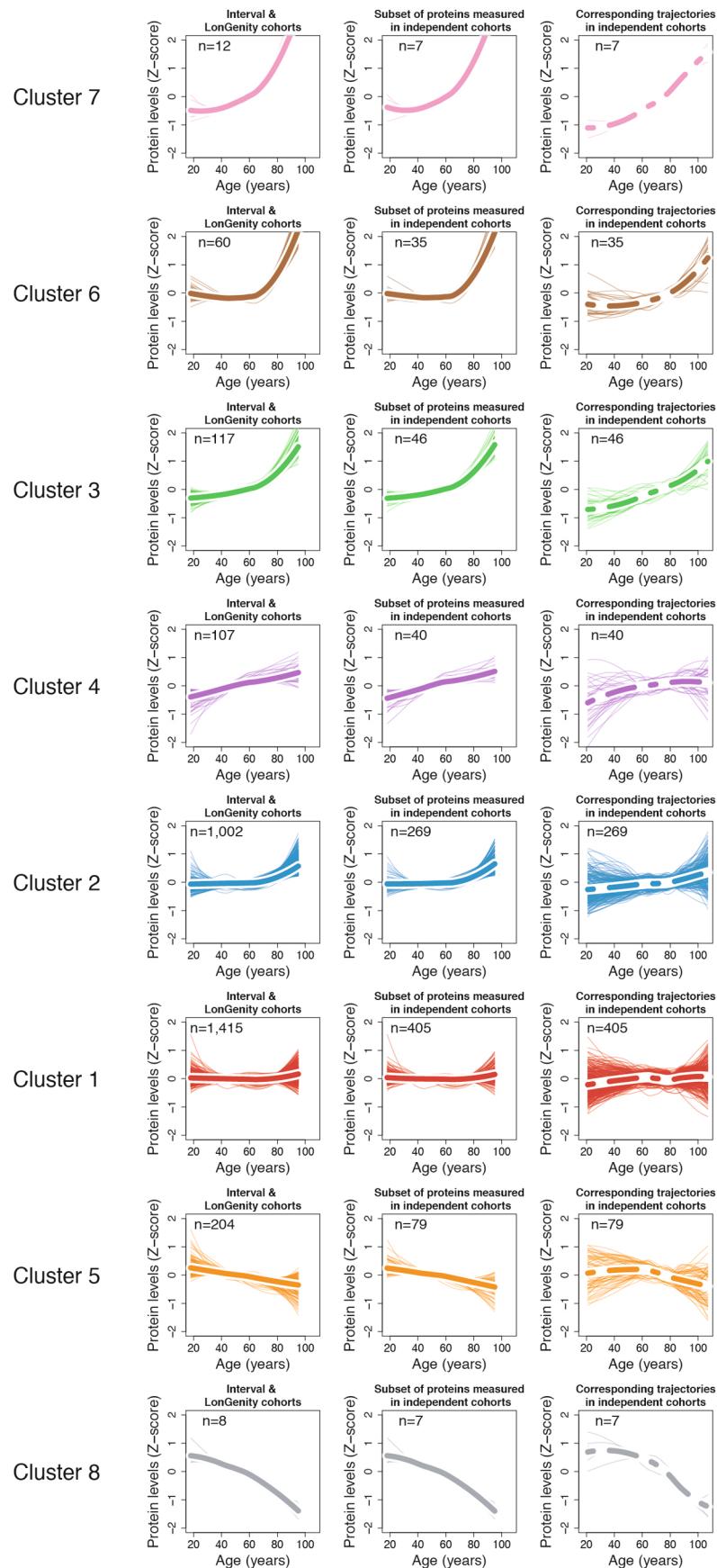
Extended Data Fig. 2 | Comparing age and sex effects in independent cohorts. (a) Age and sex effects in the INTERVAL and LonGenity studies ($n=4,263$) were compared to age and sex effects in 4 independent cohorts analyzed together ($n=171$) and to age effect from Tanaka et al. ($n=240$, 2018). The aging plasma proteome was measured with the SomaScan assay in these cohorts and 888 proteins were measured in all studies (b) Scatter plot representing the signed $-\log_{10}(q \text{ value})$ of the sex effect in the INTERVAL/LonGenity cohorts (x axis, $n=4,263$) vs the 4 independent cohorts (y axis, $n=171$). Similar analysis for the age effect in the 4 independent cohorts (c, $n=171$) and in Tanaka et al study (d, $n=240$).



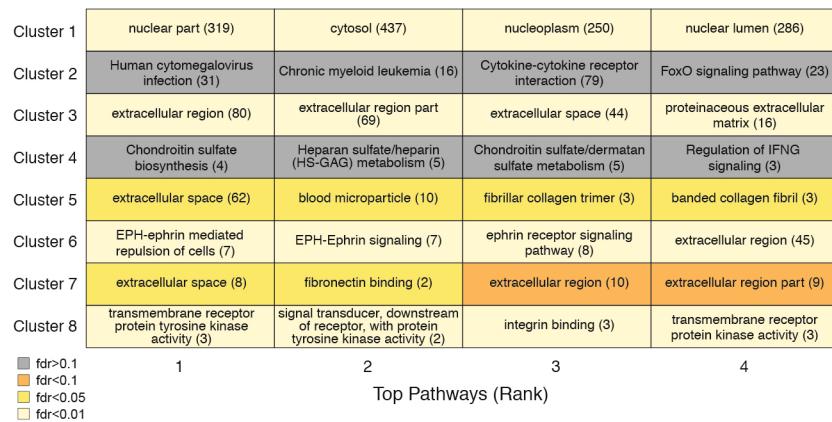
Extended Data Fig. 3 | Deeper investigation of the aging proteomic clock. (a) Prediction of age in the 4 independent cohorts ($n=171$) using the proteomic clock. Only 141 proteins out of the 373 constituting the clock were measured in these samples. (b) Prediction of age in the discovery cohort ($n=2,817$) using the 373 plasma markers. (c) Feature reduction of the aging model in the Discovery and Validation cohorts to estimate whether a subset of the aging signature can provide similar results to the 373 aging proteins. Dashed lines represent a broken stick model and indicate the best compromise between number of variables and prediction accuracy. (d) Heatmap representing the associations between delta age and 334 clinical and functional variables. For quantitative traits, linear models adjusted for delta age, age and sex were used and significance was tested using F-test. For binary outcomes, binomial generalized linear models adjusted for delta age, age and sex were used and significance was tested using likelihood ratio chi-square test. As in (c) the analysis was performed for the top 2 to top 373 variables predicting age. The non-uniformity in the heatmaps suggests that specific subsets of proteins may best predict certain clinical and functional parameters.

a**b****c**

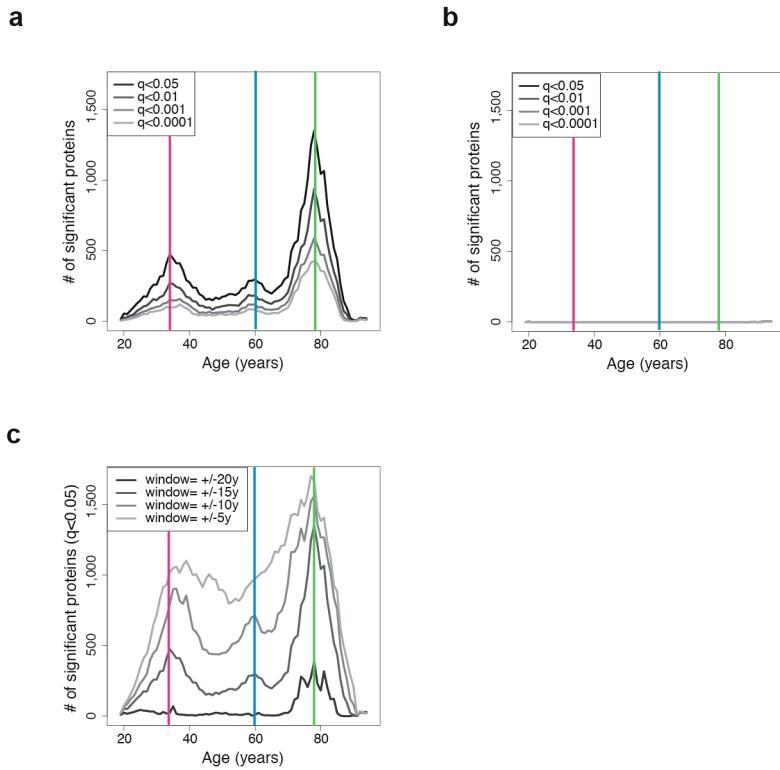
Extended Data Fig. 4 | Proteins and proteome undulations in independent human cohorts and in mouse. (a) Trajectories of 5 selected proteins based on the INTERVAL and LonGenity cohorts ($n=4,263$, left) and 4 independent human cohorts ($n=171$, right). Trajectories were estimated using LOESS regression. Undulation of the 1,305 plasma proteins measured in 4 independent cohorts (b, $n=171$) and in mouse (c, $n=81$). Plasma proteins levels were z-scored and LOESS regression was fitted for each plasma factor.



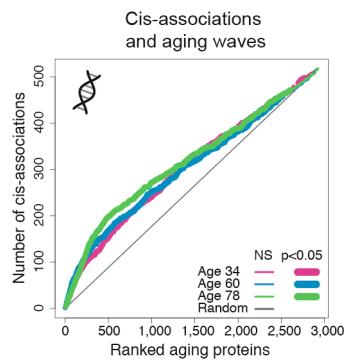
Extended Data Fig. 5 | Cluster trajectories in independent cohorts. Protein trajectories for the 8 clusters identified in the INTERVAL and LonGenity cohorts (left column). Thicker lines represent the average trajectory for each cluster. Cluster trajectories for the subset of proteins measured in the 4 independent cohorts (middle column). Corresponding cluster trajectories in 4 independent cohorts (right column).



Extended Data Fig. 6 | Pathways in clusters. Pathway enrichment was tested using GO, Reactome and KEGG databases ($n=4,263$). Enrichment was tested using Fisher's exact test (GO) and hypergeometric test (Reactome and KEGG). The top 4 pathways for each cluster are shown. Pathway IDs and number of plasma proteins associated are represented in the table.



Extended Data Fig. 7 | DE-SWAN age effect for multiple q-values cutoffs, windows size and after phenotypes permutations. Different Q-value cutoffs are represented in (a). Similar analysis with different after phenotype permutations (b) and different windows size in (c). The 3 local peaks identified at age 34, 60 and 78 are indicated by colored vertical lines.



Extended Data Fig. 8 | Cis-associations and aging waves. Enrichment for cis-association in the waves of aging proteins identified by DE-SWAN. Aging proteins were ranked based on p-values at age 34, 60 and 78 and the cumulative number of cis-associations was counted. One-sided permutation tests ($1e+5$ permutations) were used to assess significance.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

No software was used for data collection

Data analysis

R version 3.6.1 with packages: DEswan (0.0.0.9001), car (3.0-3), topGO (2.36), clusterProfiler(3.12.0) , org.Hs.eg.db (3.8.2) , NHANES (2.1.0), ade4 (1.7-13), SuperExactTest (1.0.7), qgraph (1.6.3). Details and references can be found within text in the relevant Methods sections.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

We created a searchable web interface to mine the human INTERVAL and LonGenity datasets (https://twc-stanford.shinyapps.io/aging_plasma_proteome/). The Human independent cohorts and mouse protein data are available in Supplementary Tables 16 and 17. The INTERVAL data is available through the European Genome-Phenome Archive (<https://ega-archive.org/studies/EGAS00001002555>).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Methods, "Human cohorts characteristics" and "Validation of the aging signature in mice" subsections.
 Human cohorts. This is a discovery study, not focusing on specific effect-size. Power calculation are not applicable for our study but we are well-powered (4000+ subjects) and we identified hundreds of highly significant changes after adjustment for multiple comparisons. Sample sizes of the Human cohorts are listed in Extended Data 1)
 Mouse cohorts. No statistical methods were used to predetermine sample size. Sample size was determined based on the number of animals used in prior experiments conducted in the Wyss-Coray lab (Villeda et al., Nature 2011; Villeda et al., Nature Medicine 2014, Yousef et al., Nature Medicine 2019). Again, we are well powered (110 mice) and we identified hundreds of highly significant changes. Sample sizes of the mouse cohorts are listed in Supplementary Table 9.

Data exclusions

Methods, "Human cohorts characteristics" and "Validation of the aging signature in mice" subsections.
 For the 4 independent cohorts and the mouse data. "Data for 1305 SOMAmer probes were obtained and no sample or probe data were excluded"
 For the Interval cohort, we used the dataset publicly available without excluding samples / proteins
 For the LonGenity cohort. "Sixty-eight subjects without clinical and functional data were excluded from the analysis." By excluding these samples, the same subjects (and not different subsets) were used in the whole paper, making the interpretation of the results more straightforward.
 For the Addneuromed data, no exclusion criteria were pre-established but "we filtered out 4 addneuromed samples based on visual inspection of the results of a Principal Component Analysis (PCA)".

Replication

All attempts at replication were successful (See Extended Data 2, 3a, 4).

Randomization

Human cohorts. Within each subcohort, the age was balanced and age distribution between cohorts overlapped to assess cohort and batch effect. Sex was balanced.
 Mouse cohorts. Number of male mice per group was balanced. Female mice were available only at 3m, 12, 18 and 21 months.

Blinding

Somalogic measured proteins levels without any information about the samples.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| | |
|-------------------------------------|---|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Animals and other organisms |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Human research participants |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Clinical data |

Methods

| | |
|-------------------------------------|---|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Animals and other organisms

Policy information about [studies involving animals; ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

Methods, "Validation of the aging signature in mice" subsection. A total of 110 male and virgin female C57BL/6JN mice were used. Mouse groups are summarized in ST9.
 In the aging cohort, 6 1 months old (mo), 10 3mo, 6 6mo, 6 9mo, 10 12mo, 6 15mo, 10 18mo, 10 21 mo, 5 24 mo, 6 27mo and 6 30mo were used.
 In the parabiosis cohort, 11 4mo and 18 19mo were used.

Wild animals

This study did not involve wild animals

Field-collected samples

This study did not involve samples collected from the field.

Ethics oversight

Methods, "Validation of the aging signature in mice" subsection. All animal care and procedures were carried out in accordance with institutional guidelines approved by the VA Palo Alto Committee on Animal Research

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

Participants were healthy blood donors from the Interval, LonGenity, VASeattle, GEHA, PRIN06 and PRIN09 cohorts. For Interval, age ranged from 18 to 76 years with a median age of 45 (1st Quartile=31, 3rd Quartile=55). For LonGenity, age ranged from 61 to 95 years with a median age of 74 (1st Quartile=69, 3rd Quartile=80). For the 4 independent cohorts (combined). Age ranged from 21 to 107 years with a median of 70 years (1st Quartile=58, 3rd Quartile=89).

Recruitment

Cohorts characteristics are summarized in Extended data 1
 Participants in the INTERVAL randomized controlled trial were recruited with the active collaboration of the National Health Service Blood and Transplant England (www.nhsbt.nhs.uk), which has supported field work and other elements of the trial. As described by Sun et al. (Nature, 558, pages73–79, 2018), 50,000 participants were enrolled in the randomized trial of varying blood donation intervals. People with a history of major diseases (e.g. myocardial infarction, stroke, cancer, HIV, and hepatitis B or C) or with recent illness or infection were excluded. For proteomics measurements, subjects were randomly selected. LonGenity is an ongoing longitudinal study initiated in 2008, designed to identify biological factors that contribute to healthy aging. The LonGenity study enrolls older adults of Ashkenazi Jewish descent, age 65–94 years at baseline. Approximately 50% of the cohort consists of offspring of parents with exceptional longevity (OPEL), defined by having at least one parent that survived to 95 years of age. Subjects were randomly selected within each cohort.
 Samples from the 4 independent cohort were obtained from cohorts in US (VASeattle) and Europe (GEHA, PRIN06 and PRIN09). GEHA cohort (Franceschi et al. Ann N Y Acad Sci 2007) and VASeattle samples (Britschgi, M., et al. Mol Cell Proteomics 2011) used were described previously. Subjects from the PRIN06 and PRIN09 cohorts were enrolled by multiple Italian study centers. Subjects were randomly selected within each cohort.

Ethics oversight

All participants from the Interval cohort gave informed consent before joining the study and the National Research Ethics Service approved this study (11/EE/0538)
 The LonGenity study was approved by the Institutional Review Board (IRB) at the Albert Einstein College of Medicine. The IRB has determined that our research using VASeattle, GEHA, PRIN06 and PRIN09 cohorts does not meet the definition of human subject research per STANFORD's HRPP policy because
 1) we are not obtaining or receiving private individually identifiable information
 2) data or specimens were not collected specifically for this study
 3) no direct intervention or interaction.
 For these reasons, this part of the study did not require approval from the IRB

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

ISRCTN24760606

Study protocol

Study protocol is described by Moore et al (Trials. 17;15:363, 2014, <https://www.ncbi.nlm.nih.gov/pubmed/25230735>)

Data collection

According to Moore et al (Trials. 17;15:363, 2014), INTERVAL is a randomised trial of whole blood donors enrolled from all 25 static centres of NHS Blood and Transplant. Recruitment of about 50,000 male and female donors started in June 2012 and was completed in June 2014.

Outcomes

According to Moore et al (Trials. 17;15:363, 2014), the primary outcome is the number of blood donations made. Multiple secondary outcome were investigated. The most important are (i) donor quality of life (assessed using the Short Form Health Survey) and (ii) the number of 'deferrals' due to low haemoglobin (and other factors), iron status, cognitive function, physical activity, and donor attitudes.

Several papers published the results of this clinical trial

2016 results in: <https://www.ncbi.nlm.nih.gov/pubmed/27645285>2017 results in: <https://www.ncbi.nlm.nih.gov/pubmed/28941948>2019 extension study results in: <https://www.ncbi.nlm.nih.gov/pubmed/31383583>