nature medicine

Review article

https://doi.org/10.1038/s41591-023-02784-9

Validation of biomarkers of aging

Received: 7 September 2023

Accepted: 19 December 2023

Published online: 14 February 2024

Check for updates

Mahdi Moqri[1,2,3,32], Chiara Herzog[4,32], Jesse R. Poganik[1,32], Kejun Ying[1,5], Jamie N. Justice[6], Daniel W. Belsky[7], Albert T. Higgins-Chen[8], Brian H. Chen[9], Alan A. Cohen[10], Georg Fuellen[11,12], Sara Hägg[13], Riccardo E. Marioni[14], Martin Widschwendter[4,15,16], Kristen Fortney[17], Peter O. Fedichev[18], Alex Zhavoronkov[19], Nir Barzilai[20], Jessica Lasky-Su[21], Douglas P. Kiel[22], Brian K. Kennedy[23,24], Steven Cummings[25,26], P. Eline Slagboom[27], Eric Verdin[28], Andrea B. Maier[24,29], Vittorio Sebastiano[3], Michael P. Snyder[2], Vadim N. Gladyshev[1] ✉, Steve Horvath[30] ✉ & Luigi Ferrucci[31] ✉

The search for biomarkers that quantify biological aging (particularly 'omic'-based biomarkers) has intensified in recent years. Such biomarkers could predict aging-related outcomes and could serve as surrogate endpoints for the evaluation of interventions promoting healthy aging and longevity. However, no consensus exists on how biomarkers of aging should be validated before their translation to the clinic. Here, we review current efforts to evaluate the predictive validity of omic biomarkers of aging in population studies, discuss challenges in comparability and generalizability and provide recommendations to facilitate future validation of biomarkers of aging. Finally, we discuss how systematic validation can accelerate clinical translation of biomarkers of aging and their use in gerotherapeutic clinical trials.

Aging is the strongest risk factor for most chronic diseases, physical and cognitive impairment and death. Despite this, our approach to understanding and treating aging-associated diseases has largely overlooked the biology underlying the aging process. The geroscience hypothesis posits that targeting aging itself has the potential to forestall multiple aging-associated disease processes simultaneously. As the aging population continues to grow across the globe, the promise of therapeutic targeting of aging to extend healthy lifespan has come into ever-sharper focus. To achieve this goal, there is growing interest in biomarkers that can quantitatively assess biological age and may ultimately serve as surrogate endpoints for aging-associated outcomes in clinical studies.

Many existing biomarkers of aging were initially developed to predict chronological age, although it was found that the deviation between their predicted age and the true chronological age ('AgeDev') was associated with age-related outcomes and diseases. More recent biomarkers of aging focus instead on prediction of biological age (that is, the level of age-dependent biological changes, such as molecular and cellular damage accumulation and its consequences at a certain point in time) and/or health outcomes rather than chronological age. Of note, in practical use, biological age is often summarized as a number (in units of time), just like chronological age. Regardless of the development strategy, most current biomarkers of aging predict aging-related outcomes and identify factors associated with (the pace of) aging in retrospective epidemiological studies[1–7]. In addition, they have started to provide clues on the biological mechanisms of aging. Despite these advances, the validity and usefulness of biomarkers of aging is still not widely acknowledged by biomedical scientists[1]. In contrast to biomarkers of various specific diseases, there are currently no recommended guidelines for standardizing development, measurement or validation of biomarkers of aging by regulatory bodies such as the Food and Drug Administration or the European Medicines Agency.

Validation is the multistep process by which the characteristics of biomarkers are defined, including the conditions under which they prove reliable and accurate and their ability to predict relevant outcomes[8–10]. In the context of aging biomarkers, this process requires a wide range of expertise in areas such as the biological mechanisms of aging (including conserved pathways and mechanisms in model systems and in humans, the design and construction of composite biomarkers, the design, execution and analysis of epidemiological studies that collect and store biological specimens and assess age-related predictors and outcomes in representative populations (including biobanks and cohorts), and the validation of biomarkers across

A full list of affiliations appears at the end of the paper. ✉e-mail: vgladyshev@rics.bwh.harvard.edu; shorvath@altoslabs.com; ferruccilu@grc.nia.nih.gov

https://doi.org/10.1038/s41591-023-02784-9

multiple, diverse population samples). Thus, collaboration between basic scientists and clinical investigators is essential for successfully navigating this process.

We previously proposed a consensus framework for classification and evaluation of aging biomarkers[1]. Now, we address biomarker validation as the next step in the clinical translation process. First, we review current efforts to validate predictive biomarkers of aging using population-based cohort studies and discuss challenges encountered during this process. We primarily focus on biomarkers that are blood-based, composite (as opposed to single-molecule markers) and based on 'omic' assays. Blood is widely accessible, obtained in a minimally invasive manner, and in constant contact with other tissues, potentially providing information about the biological age of the entire organism (although this is still under active exploration[11–13]). Composite panels of biomarkers are more likely to capture systemic effects of the complex aging process than single biomarkers[12–16], and those based on rapidly advancing high-throughput omic technologies and artificial intelligence (AI) are expected to substantially advance the performance and translational value of next-generation biomarkers of aging[15]. To facilitate and enhance rigor in the validation process[17], we provide guidelines for standardization and harmonization of biomarkers across populations with unique characteristics, and we make recommendations on the metrics that should be used to report their predictive performance.

Current status of validation efforts

Ideally, a biomarker measure should be robust against random and systematic sources of variability arising from technical and pre-analytical sources or application to different populations. Also, extensive information should be available on covariates to be considered to optimize their performance. We briefly outline some important types of conceptual and technical considerations and terminology important for biomarker validation in Box 1. Overall, a comprehensive process that encompasses multiple types of validation is desirable to establish reliability, accuracy and clinical utility of a biomarker of aging.

To date, predictive validation of aging biomarkers (for their association with age-related outcomes) has mostly relied on data previously collected in observational cohort studies. This process is currently the most active area of research in the aging biomarker validation space as an important prerequisite to further validation and ultimate clinical use. Cohort studies typically collect samples and clinical data on health and functional status at multiple points in time and allow assessment of association and predictivity of biomarkers for multiple health outcomes across different populations as well as the identification of relevant covariates. We focus on cross-population validation (that is, validation in more than one cohort) because it is the most robust approach for validation of blood-based biomarkers of aging in retrospective observational studies. To contextualize recommendations outlined in later sections, we first outline the current state of biomarker validation efforts (including different data sources) and discuss challenges to progression in this field.

Application of different data sources and study designs

The development of early biomarkers of aging was facilitated by open-access availability of large datasets (such as those stored within the Gene Expression Omnibus[18]), many of which are derived from cross-sectional studies. Cross-sectional studies provide a snapshot in time of variable measurements and corresponding phenotypic data (Fig. 1a). Such studies identified many biomarkers that correlate with chronological age. These include several soluble biomarkers of inflammation (for example, interleukin-6 (IL-6) or C-reactive protein) or hormonal status (such as fasting insulin and dehydroepiandrosterone sulfate). Early 'first-generation' epigenetic biomarkers were also used to predict chronological age. However, cross-sectional age associations can be biased by secular trends and selective attrition to study participation, which can preclude assessment of the predictive value of the marker in relation to future age-relevant outcomes. Furthermore, cross-sectional studies do not allow assessment of within-individual changes in response to interventions (sensitivity to change), a key requirement for the use of biomarkers of aging in clinical trials.

In contrast to cross-sectional studies, longitudinal studies collect biological measures (omics or other biomarkers), phenotypes (clinical characteristics) and adverse age-related health outcomes serially over time in the same individuals (Fig. 1b). Most longitudinal studies also include data on genetic variants, and, through Mendelian randomization studies, they may help determine whether specific biomarkers are causally related to health outcomes or rather reflect the activation of mechanisms aimed at counteracting the pathologic processes that lead to those adverse health outcomes (generally defined as 'resilience' mechanisms)[19]. Most studies collect longitudinal information on participant demographics (for example, age, sex), physiological measurements (for example, body mass index, blood pressure) and routine laboratory results (for example, complete blood count or hemograms or blood biochemistry) and may additionally collect data on mortality and cause of death as well as other aging-associated outcomes including multimorbidity, performance-based measures of physical and cognitive function, and frailty. Measures of disability in activities of daily living and instrumental activities of daily living provide information on a participant's level of independence but also health deterioration over time.

Analytically, biomarkers are often considered at one point in time and related prospectively to future outcomes, such as

BOX 1

Types of biomarker validation relevant to biomarkers of aging

Biological validation evaluates the extent to which the measurement reflects fundamental knowledge about the biology of aging. Biomarkers can be particularly insightful if they lie within a pathway that is causal to, rather than merely associated with, aging.

Cross-species validation involves assessing the functionality of a biomarker in multiple species. If a pathway associated with a biomarker is phylogenetically conserved, it is more likely to be connected with aging as a universal phenomenon[1,77].

Predictive validation involves unbiased testing of the performance of the predictive model underlying the biomarker to predict a future aging-associated outcome. For instance, HRs or time to event may be evaluated. Ideally, a true external predictive validation is carried out using independent data that were not used to train the model (often using machine learning or statistical methods). In the context of aging biomarkers, most predictive validation has been performed using retrospective analysis, but future studies should consider performing predictive validation by tracking aging-associated outcomes prospectively.

Analytical validation assesses the accuracy and reliability of the methods used to measure the biomarker, including sample collection and storage methods, analytical assays and covariates considered. This process aims to establish standard measurement practices and determine the precision, sensitivity, specificity and reproducibility of the assay.

Clinical validation aims to determine the clinical utility of a biomarker, that is, whether using that biomarker in a given setting allows for a better understanding of the ongoing disease or process that may contribute to better health outcomes. For instance, clinical validation of an aging biomarker may involve establishing that the biomarker has better predictive power for aging-associated outcomes than does chronological age.

https://doi.org/10.1038/s41591-023-02784-9

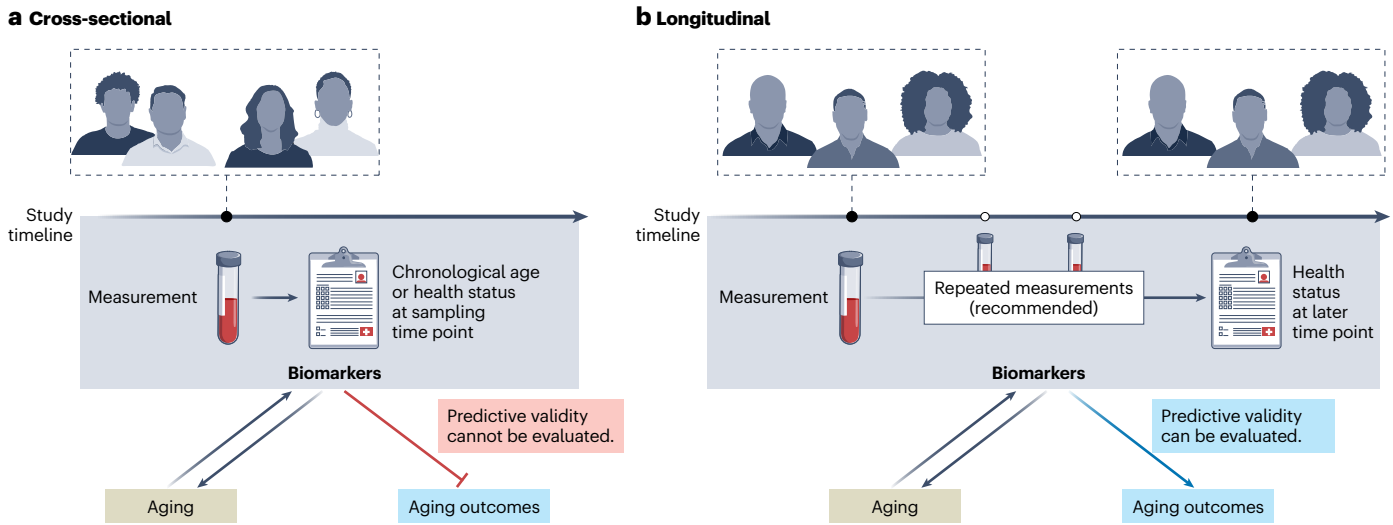

Fig. 1 | Different approaches to cohort study design in the context of biomarkers of aging. Biomarkers of aging are commonly validated using cross-sectional or longitudinal study designs. **a**, Cross-sectional studies involve measurement of biomarkers and chronological age or aging-related outcome data at a single time point. These data can only support association of these measures at that time point. **b**, Longitudinal designs, on the other hand, allow for assessment of predictive validity of biomarkers measured at one time point and future aging-related outcomes.

disease onset, change in physical and cognitive function over time or mortality. However, a more informative approach is to consider repeated measures obtained from the same participants at regular intervals. This approach allows the study of the relationship between biomarkers and the time trajectories of clinical outcomes, which provide the best approximation of the 'pace of aging'[20]. Therefore, longitudinal cohort data can uniquely support the development and validation of biomarkers of aging, such as prospective validation against multiple different outcomes and across independent populations. Additional approaches focus on resilience, healthy aging[21,22] or other aging-related outcomes[23–25]. Moreover, outcomes related to healthcare resource utilization, such as the rate of hospital admissions and use of emergency rooms, may also be highly relevant. The prioritization of aging-associated outcomes and information on (functional) aging trajectories separate from mortality could make such biomarkers even more appealing for translation to clinical studies.

Many cohort studies establish biobanks that safely store biospecimens that can then be accessed in the future to test new hypotheses or employ newly available technology for analysis. Biobanks are invaluable resources for biomarker research (particularly when it comes to testing and validation), especially if linked clinical and/or omic data and follow-up samples and/or data are available. In addition to the samples collected as part of a standard cohort study with specific research questions, large-scale, general-purpose biobanks also exist and can be useful for biomarker development. For example, the UK Biobank contains in-depth genetic and health information and holds biological samples from half a million UK participants. Multiple studies have already evaluated omic-based predictors of various aging-related outcomes in the UK Biobank[26–28]. With the decreasing costs of measuring biomarkers, this and other biobanks are currently expanding their range of available omics data[29]. The Finnish FinnGen cohort (n = ~500,000)[30], BioBank Japan (n = ~260,000)[31] and the Mass General Brigham Biobank (n = ~135,000)[32] have also recently generated large multiomics datasets, which are expected to be used to validate multiple biomarkers for aging-related outcomes. Some repositories are taking steps to organize their data in well-documented and accessible databases: for instance, the US National Institute on Aging has launched complementary translational longevity initiatives to generate large-scale, cross-species, multiomics datasets.

The current state of cross-population validation studies

Even with existing cohort studies and biobanks, systematic cross-population validation remains relatively limited. Nevertheless, several biomarkers of aging have been tested across multiple cohorts, with the most commonly examined outcome being all-cause mortality. Although there are issues surrounding mortality as an endpoint, it has the advantage of being clearly defined.

A representative list of studies validating blood-based composite biomarkers for prediction of future mortality is shown in Table 1. Many of these studies were conducted by researchers who developed the biomarker and validated that specific biomarker across multiple cohorts or by researchers who used biomarkers developed by others and compared multiple biomarkers within one cohort, with differences between the two approaches illustrated in Fig. 2. In addition, a representative list of biomarkers of aging that have been tested across multiple cohorts is shown in Supplementary Table 1 (cohorts are described in Supplementary Table 2). Our intention here is not to systematically review previous studies or perform a meta-analysis; rather, the studies considered were selected to illustrate the challenges of validating biomarkers of aging in a reliable, comparable and generalizable manner.

In studies validating blood-based biomarkers, most reported hazard ratios (HRs) for prediction of mortality risk are in the moderate range; however, a few studies have reported impressive metrics that render those biomarkers good potential candidates for use in preclinical and clinical studies. For example, Huan et al.[33,34] and Deelen et al.[35] reported increased mortality risk (HRs of 1.85 and 2.73) for their epigenetic and metabolomic biomarkers, respectively. Nevertheless, these values should be interpreted with caution because they rely on different units of measure; they will need to be substantiated by independent validation in a different cohort, and their performance needs to be compared with other biomarkers using consistent reporting measures. Thus far, relatively few studies have compared individual (composite) biomarkers across multiple cohorts or multiple biomarkers across the same cohort using standardized and equivalent measurement units that make them fully comparable[33,36,37]. We argue that studies featuring systematic and comprehensive benchmarking of diverse biomarkers of aging across many large cohorts with extended follow up (>10 years) are needed to substantially advance the field (Fig. 2a).

https://doi.org/10.1038/s41591-023-02784-9

Table 1 | Validation studies of blood-based composite biomarkers based on future mortality

Study	Biomarker	Validation set	Validation sample size	Events (%)	Adjusted HR	HR adjusted for (in addition to age)	HR unit/grouping
Levine et al.[78]	Phenotypic age	NHANES IV	6,209	1,052 (16.9%)	1.09	–	Per unit increase (year)
Sebastiani et al.[79]	'Agglomerative algorithm'	FHS	2,734	657 (24.0%)	0.68–1.23 (varying)	Sex	Cluster 1 versus clusters 2–26
Mamoshina et al.[80]	BloodAge	NHANES	2,768	873 (31%)	1.67	Sex	AgeDev <−5 versus >+5 years
Levine et al.[78]	PhenoAge	WHI, FHS, JHS, NAS	8,965	2,074 (22.8%)	1.05 (1.04–1.05)	Ethnicity	Per unit increase (year)
Lu et al.[25]	GrimAge	FHS, WHI, JHS, InCHIANTI	7,375	1,848 (25.1%)	1.10 (1.09–1.12)	Ethnicity	Per unit increase (year)
Lu et al.[4]	GrimAge2	FHS, WHI, JHS, InCHIANTI, NAS	10,065	3,900 (39%)	1.10 (1.09–1.10)	Sex, ethnicity	Per unit increase (year)
Zhang et al.[81]	DNAmRS	KORA	1,727	61 (3.5%)	10.95 (3.09–38.84)	Sex	Scores 0–1 versus 5+
Belsky et al.[20]	DunedinPACE	NAS	771	354 (45.9%)	1.26 (1.14–1.40)	Sex	Per s.d. increase
		FHS	2,471	575 (23.3%)	1.65 (1.51–1.79)	Sex	Per s.d. increase
Bernabeu et al.[82]	bAge	LBC, FHS, WHI	4,125	1,653 (40.1%)	1.52 (1.44–1.59)	Sex	Per s.d. increase
Deelen et al.[35]	MetaboHealth score	FINRISK 1997	7,603	1,213 (16.0%)	2.73 (2.60–2.86)	Sex	Per unit increase (scores −2 to 3)
van den Akker et al.[83]	MetaboAge score	LLS_SIBS	811	793 (97.7%)	1.25 (1.14–1.37)	Sex	Per unit increase (year)
Balasubramanian et al.[6]	M-metabo-score	WHI-HT	1,355	685 (50.6%)	1.95 (1.46–2.62)	Clinical and lifestyle risk factors	Highest versus lowest quartile
Tanaka et al.[5]	'Proteomic signature'	InCHIANTI	997	504 (50.6)	1.03 (1.02–1.04)	Sex, study site	Per unit increase (year)
Huan et al.[33]	'Integrative biomarker'	ARIC	969	331 (34.1%)	1.85 (1.44–2.37)	Sex, clinical factors	Per s.d. increase
Li et al.[84]	DNAmAge (Horvath)	SATSA	387	240 (62%)	1.17 (1.01–1.36)	Sex, education, lifestyle risk factors	Per s.d. increase
	DNAmAge (Hannum)				1.17 (0.98–1.40)		
	PhenoAge				1.26 (1.08–1.47)		
	GrimAge				1.39 (1.11–1.75)		
Hillary et al.[85]	EEAA	GS	2,578	57 (2.2%)	1.39 (1.12–1.72)	Sex	Per s.d. increase
	PhenoAge				1.38 (1.11–1.73)		
	GrimAge				1.70 (1.35–2.14)		
	DunedinPoAm				1.69 (1.30–2.18)		
	DNAm estimate of telomere length				0.81 (0.63–1.04)		
McCrory et al.[86]	DNAmAge (Horvath)	TILDA	490	45 (7.2%)	1.03 (0.74–1.44)	Sex, lifestyle risk factors	Per s.d. increase
	DNAmAge (Hannum)				0.92 (0.67–1.28)		
	PhenoAge				1.13 (0.81–1.57)		
	GrimAge				1.91 (1.23–2.96)		
Föhr et al.[87]	DNAmAge (Horvath)	FITSA	413	156 (35.2%)	1.05 (0.89–1.23)	Family relatedness, lifestyle risk factors	Per s.d. increase
	GrimAge				1.31 (1.08–1.59)		
Wang et al.[37]	IEAA	NAS	737	337 (45.7%)	1.08 (0.92–1.28)	Clinical and lifestyle risk factors	Per s.d. increase
	EEAA				1.10 (0.93–1.3)		
	PhenoAge				1.17 (0.98–1.41)		
	GrimAge				1.56 (1.24–1.96)		
	DNAmRS				1.37 (1.06–1.78)		

Studies with ≥10-year follow up are listed. Hazard ratios (HRs) derived from Cox proportional-hazard regression for several biomarkers of aging against all-cause mortality are shown. We aimed to include the most representative adjustment model for studies that report multiple models and report the HR as the most frequently used metric for assessing performance of a biomarker with regard to time-to-event analysis. Note that reported HRs are not directly comparable because they refer to different units of measure of the predictor. In addition, many factors (for example, population characteristics and data preprocessing) may influence predictive performances. Abbreviations: AgeDev, age deviation; EEAA, extrinsic epigenetic age acceleration; IEAA, intrinsic epigenetic age acceleration; s.d., standard deviation. Cohorts: NHANES, National Health and Nutrition Examination Survey; FHS, Framingham Heart Study; FINRISK, National FINRISK Study (Finland) for non-communicable disease intervention; WHI, Women's Health Initiative; WHI-HT, Women's Health Initiative-Hormone Therapy; JHS, Jackson Heart Study; NAS, Normative Aging Study; InCHIANTI, Invecchiare in Chianti; KORA, Cooperative Health Research in the Region Augsburg; LBC, Lothian Birth Cohort; LLS_SIBS, Leiden Longevity Sibling Study; ARIC, Atherosclerosis Risk in Communities Study; SATSA, Swedish Adoption/Twin Study of Aging; GS, Generation Scotland; TILDA, the Irish Longitudinal Study on Ageing; FITSA, the Finnish Twin Study on Ageing.

https://doi.org/10.1038/s41591-023-02784-9

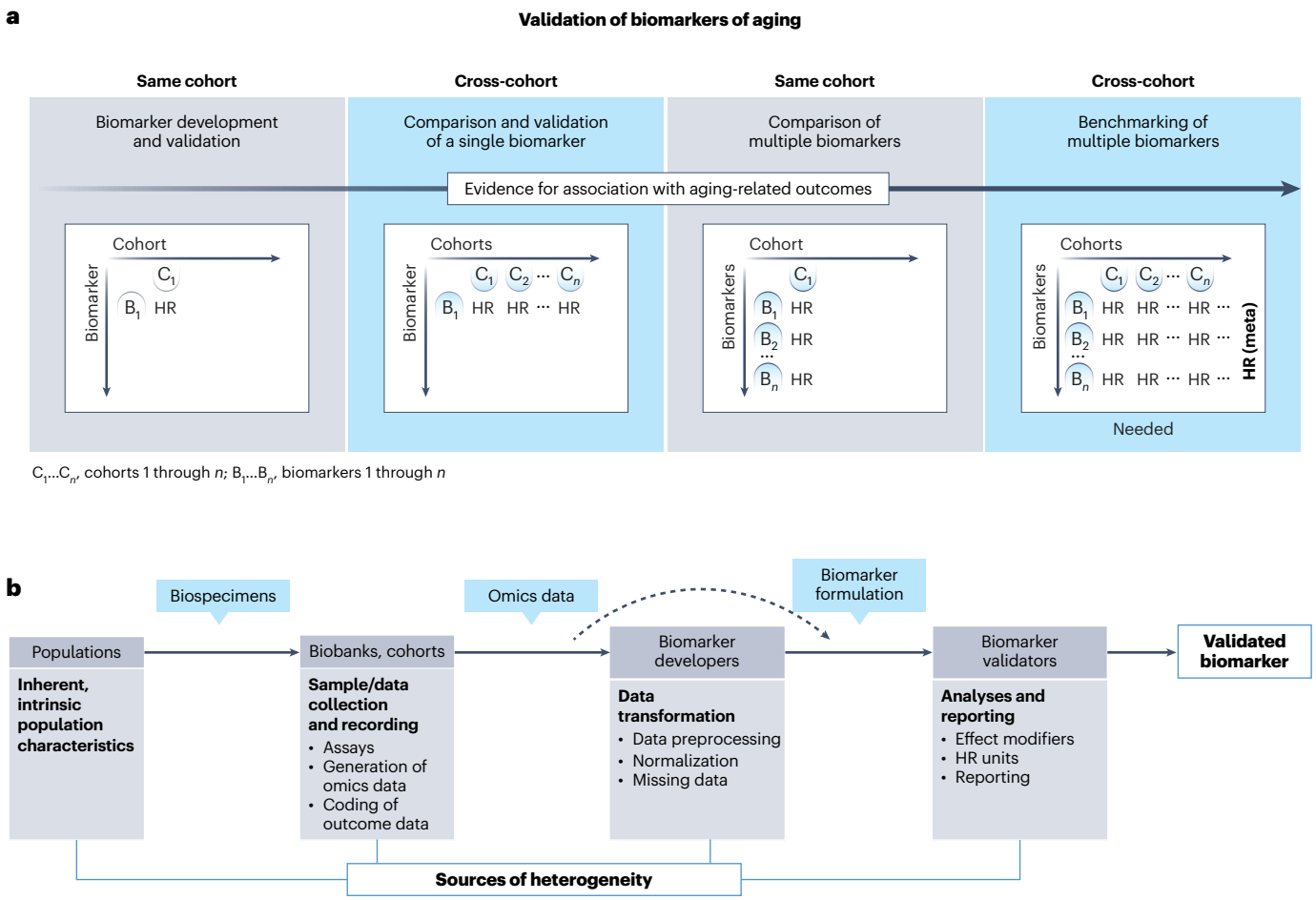

Fig. 2 | Validation of biomarkers of aging with different numbers of cohorts or biomarkers. a, Most existing biomarkers have been developed using data from a single cohort, and some have been validated in a second external cohort. Analysis of multiple biomarkers across multiple cohorts allows for a meta-analysis comparison. **b**, Biomarker validation studies need to consider different sources of variation, such as heterogeneity in population characteristics, sample collection, data preprocessing, analyses and reporting.

Challenges for validation of biomarkers of aging

Despite ongoing progress, comparing the predictive strength of biomarkers of aging remains challenging. Even for a well-defined outcome such as mortality, studies evaluating predictive performance of omic biomarkers have provided heterogeneous results. Potential reasons for this inconsistency include different study populations with different characteristics; differences in recording, formatting and coding of molecular and outcome data; differences in preprocessing and biomarker formulation; and different approaches to validation analyses and reporting (Fig. 2b and Table 2). In the following sections, we focus on each one of these problems.

Population-specific characteristics

Predictive performance of a biomarker of aging may vary by characteristics of the underlying population, including age demographics, ethnicity, health and disease status or physical and cognitive function. For instance, in a population with high exposure to pollution or environmental contaminants, cancer biomarkers will appear to be highly predictive of all-cause mortality even if they are not so in the general population. A related challenge is the lack of participant diversity in many large cohort studies and biobanks that suffer from heavy over-representation of people with European ancestry and predominantly white participants. Results from these studies may not apply to non-white, ethnically diverse individuals, which limits their external validity. Exceptions featuring more diverse populations or a focus on minority populations exist, such as the Jackson Heart Study (https://www.jacksonheartstudy.org/) or Healthy Aging in Neighborhoods of Diversity across the Life Span[38], but many more studies are needed to understand similarities and differences in biomarkers of aging across diverse populations. Notably, aging biomarkers that are reproducible across population groups likely reflect fundamental mechanisms of aging biology; such biomarkers would be broadly useful for both clinical and basic research applications.

Molecular and outcome data

Cohort studies are generally designed to address specific sets of scientific questions. Therefore, each cohort or biobank features unique content, collected and recorded in a unique manner to address these questions. Even studies carrying out similar analyses may use different approaches. For example, epigenetic data could be collected using different microarray assays (27K, 450K, 850K) or by isolating DNA with different methods, which produce slightly different estimates even for epigenetic targets shared between platforms[39]. Similarly, metabolomics or proteomics data could be collected from plasma or serum, leading to different data distributions[35], measured using different technologies (for example, mass spectrometry or aptamer-based assays) or tagged using different nomenclature[40]. Unfortunately, no harmonization standards currently exist for molecular data and aging-associated outcomes for the purpose of validating biomarkers of aging. Existing programs and consortia, such as Reference Set of Metabolite Names

(RefMet[41]), Consortium of Metabolomics Studies[42], Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE[43]), Common Infrastructure for National Cohorts in Europe, Canada, and Africa[44], Trans Omics for Precision Medicine program (TOPMed[45]), UK Longitudinal Linkage Collaboration, University Biobank Limburg (UBiLim[46]), Biobank Standardisation and Harmonisation for Research Excellence in the European Union (BioSHaRE-EU[47]) and Biobanking and Biomolecular Resources Research Infrastructure Netherlands (BBMRI-NL), are developing standards for organization of specific data types to facilitate large-scale collaborations in other fields, but no such efforts have been initiated in geroscience.

Biomarkers constructed with time to mortality as a reference outcome also tend to predict chronic diseases and functional and cognitive outcomes independent of chronological age, suggesting that they capture a dimension related to overall health[1,36]. However, the direct use of non-mortality aging-related outcomes, such as multimorbidity, poor mobility and frailty, may better capture information on the pace of aging and may be more useful for clinical applications. These could include internationally recognized scoring systems of multimorbidity, frailty, disability, cognition[48] or quality of life measures as well as more health-focused metrics (such as vitality, resilience and healthspan), although no consensus yet exists on how to quantify the latter two[49]. Beyond cohort-specific challenges, access to cohort data remains a general ongoing issue: applying for access to many government-funded datasets often requires lengthy paperwork and review processes and can often span several months or years.

Biomarker procedures and formulations

Statistical and machine learning models used to identify or learn the relation between biomarkers and aging outcomes are still in early stages of development and validation, and many modeling challenges remain to be addressed. For example, many existing models assume a linear relation between biomarkers of aging and the likelihood of aging outcomes throughout the lifespan, while recent studies have discovered multiple examples of non-linearity[5,12,50]. Technical considerations surrounding data preparation also pose challenges. For instance, recent work has demonstrated that calculating principal components from CpG-level data as input for biological age prediction can improve test–retest reliability of epigenetic biomarkers[51]. These and other unique transformations of individual measurements make cross-comparison of composite biomarkers challenging. Moreover, biomarkers or their components may be sensitive to underlying sample composition. For example, there is evidence that age-related methylation varies across different circulating immune cells[52]. Therefore, comparative or validation studies should always carefully adjust for the proportions of different types of circulating cells. Studies may also treat missing data or repeated measurements for biomarkers or outcomes differently, potentially influencing power or skewing performance estimates. This issue is particularly important for proteomic assays that tend to generate many values 'below the threshold of detection' that may not be random but rather convey important information. Finally, there is currently no guidance on how to best integrate longitudinal repeated measures from the same individual and whether trajectories or unique values should be considered.

Study design and reporting

Several aspects of study design, such as follow-up time, number of events and bias in mortality reporting may introduce variability across studies. Differences in statistical approaches are also a notable source of variation. For example, different validation studies often account for distinct potential effect modifiers by controlling, adjusting or stratifying for them. These factors are expected to affect the magnitude of the relationship between the omic biomarkers and aging-associated outcomes, representing another challenge for comparison of biomarkers in (and across) validation studies. Additionally, studies can report performance metrics such as HRs in different ways (for example, per standard deviation (s.d.), compared to a reference group, or per-unit increase) using different adjustment strategies for covariates (Table 1). In Cox proportional-hazard regression models, biomarkers can be coded as continuous variables (standardized or not) or as ordinal variables that capture quantiles of biomarker level or even as time-dependent covariates. The former approach provides information on risk estimates per one-unit difference in biomarker level (for example, per s.d.), while the second considers one level (typically the lowest quantile) as a reference group. These inconsistencies, which also plague other fields, have hindered reliable cross-comparison, benchmarking and meta-analysis of evaluated biomarkers of aging.

Recommendations for validation of biomarkers of aging

Cross-population validation of multiple biomarkers across several cohorts on a large scale is necessary but challenging and will require considerable coordinated effort and increased funding. Based on the current state of the field and the challenges outlined above, we provide the following recommendations (summarized in Table 2) for benchmarking and reporting of validation studies, grouped by target stakeholders.

Recommendations for biomarker developers

Before composite or algorithmic biomarkers are validated across populations, the underlying statistical or machine learning models capturing the biological relation between biomarkers and outcomes need to be verified. It is important to examine the extent to which an association could be reasonably attributed to the underlying biology. We recommend that biomarker developers verify that the statistical assumptions of their models reflect the expected biological phenomena to the extent of our current knowledge. For example, as recent studies continue to reveal unique age-dependent epigenetic changes during different phases of life, it is becoming clear that non-linear or piecewise epigenetic biomarkers might represent the whole human lifespan more accurately than those that assume a linear relationship with age across the life course[53,54].

Successful validation of biomarkers requires full transparency of the methods used for their development, computational preprocessing and analysis, and verification of their predictive validity in multiple independent populations. Hence, preprocessing pipelines should follow best-practice guidelines that ultimately enable data harmonization[55]. For example, the treatment of missing or repeated measurements (for example, using imputation or machine learning methods[56,57]), data normalization and quality control influence predictive performance results; therefore, it is important to establish and follow standards and best practices for these steps[58]. Similarly, fully specified computational procedures (formulations) for composite biomarkers should be made available publicly (as recommended for all omic tests by the US National Academy of Science, Institute of Medicine[17]) to allow for computation of biomarker scores independently by other researchers, without the need to upload or transfer data to biomarker developers. In addition, biomarker formulation should allow for simple implementation across new datasets. Indeed, most omic biomarkers could be formulated in standardized mathematical terms (see harmonization efforts by ClockBase for epigenetic biomarkers[20] and MiMIR for metabolomic biomarkers[59]) and standardized software packages, which enable streamlined calculation of various biomarkers including blood biochemistry (for example, BioAge[60]) and epigenetics (for example, Biolearn at https://bio-learn.github.io (ref. 61) and methylCIPHER[62]). We believe that such a process of validation and implementation would provide even stronger results if it were undertaken according to guidelines that are widely discussed and adopted by the scientific community.

https://doi.org/10.1038/s41591-023-02784-9

Table 2 | Challenges and associated recommendations for validation of biomarkers of aging

Challenge	Target stakeholder	Recommendation	Example
1. Population-specific characteristics	Data maintainers	1. Adopt data-sharing mechanisms that enable timely and broad access to enable validations using many populations.	Provide transparent information on available data as well as data-access review processes including expected review time.
	Validation study teams	2. Include multiple diverse populations.	Validate biomarkers across multiple diverse cohorts and report stratified analyses.
2. Molecular and outcome data	Data maintainers	3. Follow FAIR[67] data principles, provide a detailed metadata and data dictionary and use standard data formats.	Ensure that data are FAIR[67] by providing appropriate documentation and guidance: for example, use Gene Expression Omnibus data format for gene expression (transcriptomics) and epigenetics data.
	Biomarker developers	4. Verify assumptions of statistical/machine learning models used to identify/learn the relation between biomarkers and aging outcomes.	Consider non-linear or piecewise models for biomarkers with established non-linear relation with aging outcomes.
	Validation study teams	5. Standardize and harmonize individual biomarker measurements and aging outcomes in different datasets.	Extend biomarker-standardization programs (for example, RefMet) and consortium data-harmonization efforts by CHARGE, TOPMed, UBiLim, BBMRI and BioSHaRE-EU toward developing assay-agnostic and generalizable biomarkers.
		6. Consider aging outcomes beyond mortality.	Consider multimorbidity, frailty, disability, quality of life measures or health-focused metrics, such as vitality, resilience and healthspan.
3. Biomarker procedures and formulations	Biomarker developers	7. Improve transparency of sample preparation, data processing and biomarker formulation.	Provide fully specified computational procedures, including details on the normalization method and treatment of missing measurements.
	Validation study teams	8. Carry out post hoc harmonization of composite biomarker formulations	Develop/extend packages or solutions for computation of multiple biomarkers, for example, methylCIPHER, BioAge, ClockBase, MiMIR and Biolearn.
4. Comparability of validation studies	Biomarker developers	9. Improve interpretability, generalizability and robustness.	Incorporate various omics data from diverse populations during biomarker development.
	Validation study teams	10. Account for potential effect modifiers by controlling, adjusting for or stratifying based on them.	Report HRs for chronological age- and sex-adjusted models.
		11. Standardize the validation process including: a. Biomarker formulation, b. Minimal requirements, c. Statistical analysis and d. Performance metrics.	'Lock down' predefined biomarker formulation and conduct validation studies according to defined standards, for example, minimum follow-up time for aging outcome, Cox proportional-hazard regression and HRs per s.d. and absolute unit increase.
		12. Report the results comprehensively and appropriately.	Follow established guidelines for reporting of observational studies, such as STROBE.

To support future validation studies, we recommend that developers consider methods and data sources that improve the likelihood of future generalizability, cross-population validity and potential clinical validity of their biomarkers. Currently, epigenetic markers are the most commonly proposed and investigated type of composite biomarker[1]. We recommend developing methods to address many widely acknowledged challenges with these and other biomarkers, including interpretability[2,19] and technical robustness[20,51]. Studies involving longitudinal sample collection may be particularly useful for this purpose, as the resulting data can enhance our understanding of the dynamic properties of these biomarkers. In addition, we recommend the use of other complementary omics data, including metabolomics, proteomics, transcriptomics and lipidomics, the biological interpretation of which is often less complex, to develop biomarkers capable of capturing aspects of aging that may not be best reflected by epigenetic data[63]. As the costs of many omic assays are decreasing and cohorts and biobanks are increasingly incorporating multiple data modalities (Table 3), we expect multiomic biomarkers to become common in the near future, emphasizing the need for accessible and standardized approaches.

Recommendations for data maintainers

Successful validation of biomarkers depends on access to data and harmonization of aging-related phenotypic and molecular data across relevant cohorts. Procedures for easy data sharing that enable more timely and broad access while maintaining privacy of individual human data (such as the National Heart, Lung, and Blood Institute BioData Catalyst[64]) should be widely adopted. Data repositories can and should provide transparent information on available data and data formats as well as data-access criteria and review processes, including expected review time based on historical statistics. In addition, synthetic datasets (with the same data structures and distributions), data safe havens (that is, secure storage and computing for sensitive data) and federated access (unified central access) would ideally be provided to facilitate broader access to the data. Providing one or more of the above should be incentivized by funding agencies and other financial supporters. For sensitive data with controlled access, federated analysis, in which data remain decentralized on host institution servers but are made available for analysis in a privacy-preserving manner[65], may offer a suitable compromise, especially using cloud-based methods. Rather than requesting transfer of sensitive data, individuals aiming to validate a biomarker could provide the formulation of the biomarker to data owners and/or conduct their analysis in a secure environment, with access to only summary or synthetic data.

Many initiatives (for example, RefMet, CHARGE, TOPMed, UBiLim, BBMRI, BioSHaRE-EU) have taken steps to standardize biomarker nomenclatures or cohort or biobank data to facilitate cross-population

studies, often following rigorous guidelines for retrospective data harmonization (such as Maelstrom[66]). While these post hoc efforts are needed to improve existing data, cohort or biobank data maintainers may facilitate the process by following best practices in recording and reporting biomarkers and aging outcome measures from the inception of their biobanks or cohort studies. In particular, data owners should aim for alignment with FAIR data principles (ensuring that data are findable, accessible, interoperable and reusable[67]), provide machine-readable metadata and data dictionaries that allow for harmonization and make available records of data structure in data description publications. Especially for older or ongoing longitudinal studies with long follow up, the above steps represent a considerable challenge that requires increased support from the aging research community. Success with the above efforts would increase data utility, particularly for federated learning and analysis across populations, which require standardized data.

Recommendations for cross-population validation study teams

Biomarkers of aging should be evaluated across multiple diverse populations to account for differences across genetic ancestries, sex, geographic contexts, environmental or lifestyle factors, life stages and health or disease states. This step is critical because even seemingly established biomarkers may not be valid in all human populations. For example, the *APOE4* allele is the strongest risk factor for Alzheimer's disease risk in white populations, but the association is substantially weaker in African American and Hispanic populations[68]. Moreover, in Tsimane horticulturalists (a subsistence population in Brazil), *APOE4* appears to be protective against cognitive decline[69]. As mentioned above, many existing composite biomarkers of aging have been trained in cohorts of predominantly white, European ancestry. A similar bias in the design of genetic studies has resulted in the development of polygenic risk scores that have diminished predictive accuracy in populations with non-European ancestry[70].

While many composite biomarkers of aging have shown some evidence of comparable predictive accuracy across genetic ancestry populations[25,71,72], establishing diverse cohorts with non-European ancestries to validate new composite biomarkers of aging remains a priority. Other key axes for cross-validation could include climate zones, country or continent and exposure levels to various chemical or biological risk factors. This will require efforts to establish resources and research capacity in various geographical regions and minority populations. In addition to commonly used cohorts, many other cohort studies or biobanks (many of which are focused on recruiting diverse or minority populations) may be suitable for validation studies of biomarkers of aging. Some of these have already added or are in the process of adding (multi)omics data (Table 3), which will help to further improve development or validation of biomarkers of aging.

Efforts by developers to standardize various aspects of biomarker validation, including biomarker formulation and statistical analyses (described above), will allow for a reliable comparison across studies. For instance, biomarker formulations should be established 'a priori' and not be further modified during validation (in other words, formulation 'lock down'[17]). Additionally, statistical analysis outputs, such as hazard ratios, should be reported for unadjusted, chronological age- and sex-adjusted and fully adjusted models, permitting a broader cross-comparison of studies. Studies may additionally account for other factors mentioned above, including sample composition. To ensure comparability of performance, the community needs to take steps to agree on the minimal set of covariates to be included in the analysis and the use of stratified analyses by subgroups, such as age, sex and/or race. Finally, reporting HRs per s.d. and the absolute unit differences in biomarker levels (for example, 1 s.d. and one unit of increase in the biomarker) allows for easier comparison of different biomarkers and meta-analysis. While perfect standardization may not be realistically achievable, moving in the direction of standardization will at least enable the qualitative assessment of the extent to which results in different populations converge.

Correct reporting of study results is vital to enable cross-population validation. We recommend that investigators follow established guidelines for reporting of observational studies, such as Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)[73], to enhance transparency and reproducibility of findings. All populations should be sufficiently described, either in summary or individually, when including multiple cohorts in one study. When focusing on mortality as a key aging-associated outcome, studies should report all-cause mortality based on reliable information and, when possible, cause-specific mortality, which may differ based on underlying population characteristics. Several multiomics datasets offer information on aging-related outcomes separate from mortality (Table 3) that may be used instead of or in addition to mortality. Analysis in subgroups with certain (chronic) conditions that lead to accelerated aging (for example, infection with human immunodeficiency virus[74]) will inform on whether these changes in aging biomarker levels are associated with increased morbidity or mortality or whether bespoke biomarkers may be required in those groups of individuals to predict clinical outcomes. At a minimum, results should be stratified and reported separately by age group and sex, given the clear sexual dimorphism in aging. Additionally, extended reporting of stratified analyses by various demographics (for example, ethnicity, country or pre-existing health status) is recommended to evaluate generalizability, as models that perform well across distinct strata are more likely to have good external validity[75]. Reporting highly stratified results will also facilitate meta-analyses.

Outlook

The past decade has seen substantial progress in the development of new blood-based biomarkers of aging. Despite the tremendous promise of these tools for use in trials for longevity interventions, major roadblocks persist in translating them to clinical use. Our article highlights challenges encountered in the validation of these biomarkers and proposes efforts to overcome these barriers. Addressing even relatively simple challenges, such as standardization of effect size reporting, stands to greatly benefit the comparison and validation of biomarkers of aging. We anticipate that studies benchmarking multiple biomarkers of aging, especially those using different technologies (for example, metabolomics, proteomics, epigenetics and multiomic approaches), across multiple populations will provide a more comprehensive understanding of their performance and robustness. Performing such large-scale comparative studies is a key priority to progress in this field but will require increasing cooperation between research groups and the creation of incentives for transparent sharing of biomarker formulations and data. Efforts toward harmonization will require endorsement across diverse stakeholder groups, including biomarker developers, data owners and epidemiological researchers, and may ultimately enable the goal of identifying the most promising biomarker candidates for clinical prioritization.

We further recommend that future work should aim to incorporate more clinically relevant and potentially actionable outcomes instead of or in addition to mortality. Many cohort studies provide alternative health outcome data (Table 3) that may support this goal, including data on specific chronic diseases, multimorbidity and organ-specific physiological integrity as well as physical and cognitive function. However, agreement on the standard definition and operationalization of these outcomes would be highly desirable and ensure true comparability. Prospective studies that develop individual, longitudinal profiles of biomarkers will provide a resource that is urgently needed in the field. Such studies will be critical, particularly with respect to assessing whether biomarkers are sensitive to physiological changes such

https://doi.org/10.1038/s41591-023-02784-9

Table 3 | Cohorts or biobanks with existing multiomics data

Study characteristics					Omics				Follow-up outcomes			
Abbreviation	**Cohort**	**Location**	**Description/note**	**Total *n***	**Methylation**	**Metabolome**	**Proteome**	**Other**	**Time to death**	**Cause of death**	**Multimorbidity**	**Other**
UKB	UK Biobank	UK	Largest multiomics dataset to date	>500,000		✓	✓	Transcriptome	✓	✓	✓	Cognitive function, accelerometer
LIFELINES	Lifelines	The Netherlands	Multigenerational European cohort	>167,000	✓	✓	✓	Microbiome	✓	✓	✓	Cognitive function
MGB	Mass General Brigham Biobank	USA	Leading hospital biobank	>135,000	✓	✓	✓	Lipidome	✓	✓	✓	
RS	Rotterdam Study	The Netherlands	Three cohorts focusing on mid-to-late life	>15,000	✓	✓	✓	Microbiome, lipidome	✓	✓	✓	
MESA	Multi-Ethnic Study of Atherosclerosis	USA	Multi-ethnic population followed since 2000	>6,000	✓	✓	✓	Transcriptome	✓	✓		
CARDIA	Coronary Artery Risk Development in Young Adults	USA	Population of mixed ethnicities followed from young adulthood (18–30)	>5,000	✓	✓	✓		✓	✓	✓	Cognitive function
HABC	Health, Aging and Body Composition Study	USA	Focus on functional decline and body composition	>3,000		✓	✓		✓	✓	✓	Cognitive function, physical function, MRI
BLSA	Baltimore Longitudinal Study of Aging	USA	Comprehensive characterization of aging phenotypes	>3,000	✓	✓	✓	Microbiome	✓	✓	✓	Cognitive function, physical function
AASK	African American Study of Kidney Disease and Hypertension	USA	African American population followed over time	691		✓	✓		✓	✓		
iPOP	Integrated Personal Omics Profiling	USA	Deep longitudinal profiling	>100	✓	✓	✓	Microbiome, lipidome	✓	✓	✓	

MRI, magnetic resonance imaging.

as those induced by longevity interventions and gerotherapeutics or other preventive measures.

We anticipate that the ideal biomarkers of aging shall have moderate to strong associations with chronological age and predict multiple aging-related outcomes beyond mortality, such as functional decline, frailty, chronic diseases and disability and (multi)morbidity. They should be sensitive to upstream factors thought to influence aging such as stress, adverse events, environment, genetics and lifestyle, and they should mediate the relationship between these factors and aging outcomes. They should do so in many diverse populations and should do so relatively similarly across populations. Biomarkers of aging that meet these requirements should be prioritized for validation as screening and diagnostic biomarkers and eventually as surrogate endpoints in clinical trials. While a clear roadmap to realize this long-term goal does not yet exist, harmonization and standardization of biomarkers and population data across the field will greatly enhance our ability to identify, characterize and validate the most promising biomarker candidates.

As biomarkers of aging move toward clinical implementation, several key questions remain to be addressed. First, there is no widespread agreement on the extent to which biological age may be captured by a single biomarker. Further validation of aging biomarkers through their use in clinical and epidemiological studies will help establish whether a single biomarker or multiple complementary biomarkers may be most useful. A looming question is whether or how biomarkers of aging should be integrated into the current disease-centric and disease-specific approach to healthcare. A shift toward holistic prevention, in line with the geroscience hypothesis, has the potential to substantially change public health and expand the portion of life free of diseases and disability but will require endorsement across diverse stakeholder groups, particularly in the clinical realm. Next, the clinical utility of biomarkers of aging remains to be validated using prospective clinical trials to demonstrate that they can indeed improve how patients feel, function and survive. Finally, while we focused on blood-based biomarkers, more studies investigating aging across various organ systems are warranted to enhance our understanding and the clinical potential of biomarkers of aging.

Conclusion

The translation of the science of aging to clinical applications holds substantial promises for the improvement of healthcare and the expansion of health expectancy, with the potential to both reduce healthcare expenditure and improve population health[76]. An important prerequisite to accomplish this goal is the availability of robust biomarkers of aging, which requires a process of validation to advance these biomarkers into clinically valuable and actionable tools. It is our hope that the challenges we highlight and the recommendations we offer will aid in advancing biomarkers of aging into clinical tools that empower the action of health planners and healthcare providers.

References

1. Moqri, M. et al. Biomarkers of aging for the identification and evaluation of longevity interventions. *Cell* **186**, 3758–3775 (2023).
2. Moqri, M. et al. PRC2 clock: a universal epigenetic biomarker of aging and rejuvenation. Preprint at *bioRxiv* https://doi.org/10.1101/2022.06.03.494609 (2022).
3. Poganik, J. R. et al. Biological age is increased by stress and restored upon recovery. *Cell Metab.* **35**, 807–820 (2023).
4. Lu, A. T. et al. DNA methylation GrimAge version 2. *Aging* **14**, 9484–9549 (2022).
5. Tanaka, T. et al. Plasma proteomic biomarker signature of age predicts health and life span. *eLife* **9**, e61073 (2020).
6. Balasubramanian, R. et al. Metabolomic profiles associated with all-cause mortality in the Women's Health Initiative. *Int. J. Epidemiol.* **49**, 289–300 (2020).
7. Sarkar, T. J. et al. Transient non-integrative expression of nuclear reprogramming factors promotes multifaceted amelioration of aging in human cells. *Nat. Commun.* **11**, 1545 (2020).
8. Lee, J. W. et al. Fit-for-purpose method development and validation for successful biomarker measurement. *Pharm. Res.* **23**, 312–328 (2006).
9. Wagner, J. A. Overview of biomarkers and surrogate endpoints in drug development. *Dis. Markers* **18**, 41–46 (2002).
10. Hunter, D. J. et al. A pathway and approach to biomarker validation and qualification for osteoarthritis clinical trials. *Curr. Drug Targets* **11**, 536–545 (2010).
11. Bortz, J. et al. Biological age estimation using circulating blood biomarkers. *Commun. Biol.* **6**, 1089 (2023).
12. Lehallier, B. et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nat. Med.* **25**, 1843–1850 (2019).
13. Ahadi, S. et al. Personal aging markers and ageotypes revealed by deep longitudinal profiling. *Nat. Med.* **26**, 83–90 (2020).
14. Alpert, A. et al. A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. *Nat. Med.* **25**, 487–495 (2019).
15. Rutledge, J., Oh, H. & Wyss-Coray, T. Measuring biological age using omics data. *Nat. Rev. Genet.* **23**, 715–727 (2022).
16. Kudryashova, K. S., Burka, K., Kulaga, A. Y., Vorobyeva, N. S. & Kennedy, B. K. Aging biomarkers: from functional tests to multi-omics approaches. *Proteomics* **20**, 1900408 (2020).
17. Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials; Board on Health Care Services; Board on Health Sciences Policy; & Institute of Medicine. *Evolution of Translational Omics: Lessons Learned and the Path Forward* (eds Micheel, C. M. et al.) (National Academies Press, 2012).
18. Ying, K. et al. ClockBase: a comprehensive platform for biological age profiling in human and mouse. Preprint at *bioRxiv* https://doi.org/10.1101/2023.02.28.530532 (2023).
19. Ying, K. et al. Causality-enriched epigenetic age uncouples damage and adaptation. *Nat. Aging* https://doi.org/10.1038/s43587-023-00557-0 (2024).
20. Belsky, D. W. et al. DunedinPACE, a DNA methylation biomarker of the pace of aging. *eLife* **11**, e73420 (2022).
21. Bautmans, I. et al. WHO working definition of vitality capacity for healthy longevity monitoring. *Lancet Healthy Longev.* **3**, e789–e796 (2022).
22. Lara, J. et al. A proposed panel of biomarkers of healthy ageing. *BMC Med.* **13**, 222 (2015).
23. Crimmins, E. M., Thyagarajan, B., Levine, M. E., Weir, D. R. & Faul, J. Associations of age, sex, race/ethnicity, and education with 13 epigenetic clocks in a nationally representative U.S. sample: the Health and Retirement Study. *J. Gerontol. A Biol. Sci. Med. Sci.* **76**, 1117–1123 (2021).
24. Faul, J. D. et al. Epigenetic-based age acceleration in a representative sample of older Americans: associations with aging-related morbidity and mortality. *Proc. Natl Acad. Sci. USA* **120**, e2215840120 (2023).
25. Lu, A. T. et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* **11**, 303–327 (2019).
26. Gadd, D. A. et al. Blood protein levels predict leading incident diseases and mortality in UK Biobank. Preprint at *medRxiv* https://doi.org/10.1101/2023.05.01.23288879 (2023).
27. Tian, Y. E. et al. Heterogeneous aging across multiple organ systems and prediction of chronic disease and mortality. *Nat. Med.* **29**, 1221–1231 (2023).
28. Buergel, T. et al. Metabolomic profiles predict individual multidisease outcomes. *Nat. Med.* **28**, 2309–2320 (2022).

29. Raghavachari, N., Wilmot, B. & Dutta, C. Optimizing translational research for exceptional health and life span: a systematic narrative of studies to identify translatable therapeutic target(s) for exceptional health span in humans. *J. Gerontol. A Biol. Sci. Med. Sci.* **77**, 2272–2280 (2022).
30. Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
31. Nagai, A. et al. Overview of the BioBank Japan Project: study design and profile. *J. Epidemiol.* **27**, S2–S8 (2017).
32. Boutin, N. T. et al. The evolution of a large biobank at Mass General Brigham. *J. Pers. Med.* **12**, 1323 (2022).
33. Huan, T. et al. Integrative analysis of clinical and epigenetic biomarkers of mortality. *Aging Cell* **21**, e13608 (2022).
34. Eiriksdottir, T. et al. Predicting the probability of death using proteomics. *Commun. Biol.* **4**, 758 (2021).
35. Deelen, J. et al. A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. *Nat. Commun.* **10**, 3346 (2019).
36. Kuiper, L. M. et al. Epigenetic and metabolomic biomarkers for biological age: a comparative analysis of mortality and frailty risk. *J. Gerontol. A Biol. Sci. Med. Sci.* **78**, 1753–1762 (2023).
37. Wang, C. et al. DNA methylation-based biomarkers of age acceleration and all-cause death, myocardial infarction, stroke, and cancer in two cohorts: the NAS, and KORA F4. *eBioMedicine* **63**, 103151 (2021).
38. Evans, M. K. et al. Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS): overcoming barriers to implementing a longitudinal, epidemiologic, urban study of health, race, and socioeconomic status. *Ethn. Dis.* **20**, 267–275 (2010).
39. Olivia, S. et al. Comparison of DNA methylation measured by Illumina 450K and EPIC BeadChips in blood of newborns and 14-year-old children. *Epigenetics* **13**, 655–664 (2018).
40. Koistinen, V. et al. Towards a Rosetta stone for metabolomics: recommendations to overcome inconsistent metabolite nomenclature. *Nat. Metab.* **5**, 351–354 (2023).
41. Fahy, E. & Subramaniam, S. RefMet: a reference nomenclature for metabolomics. *Nat. Methods* **17**, 1173–1174 (2020).
42. Yu, B. et al. The Consortium of Metabolomics Studies (COMETS): metabolomics in 47 prospective cohort studies. *Am. J. Epidemiol.* **188**, 991–1012 (2019).
43. Psaty, B. M. et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ. Cardiovasc. Genet.* **2**, 73–80 (2009).
44. Keane, T. & Glass, L. CINECA: Common Infrastructure for National Cohorts in Europe, Canada, and Africa — Kick Off Report. *Zenodo* https://doi.org/10.5281/zenodo.3908145 (2019).
45. Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
46. Linsen, L. et al. Raising to the challenge: building a federated biobank to accelerate translational research—the University Biobank Limburg. *Front. Med.* **6**, 224 (2019).
47. Kaye, J. et al. Access governance for biobanks: the case of the BioSHaRE-EU cohorts. *Biopreserv. Biobank.* **14**, 201–206 (2016).
48. Cummings, S. R. & Kritchevsky, S. B. Endpoints for geroscience clinical trials: health outcomes, biomarkers, and biologic age. *GeroScience* **44**, 2925–2931 (2022).
49. Kaeberlein, M. How healthy is the healthspan concept? *GeroScience* **40**, 361–364 (2018).
50. Okada, D., Cheng, J. H., Zheng, C., Kumaki, T. & Yamada, R. Data-driven identification and classification of nonlinear aging patterns reveals the landscape of associations between DNA methylation and aging. *Hum. Genomics* **17**, 8 (2023).
51. Higgins-Chen, A. T. et al. A computational solution for bolstering reliability of epigenetic clocks: implications for clinical trials and longitudinal tracking. *Nat. Aging* **2**, 644–661 (2022).
52. Tomusiak, A. et al. Development of a novel epigenetic clock resistant to changes in immune cell composition. Preprint at *bioRxiv* https://doi.org/10.1101/2023.03.01.530561 (2023).
53. Fang, F. et al. Evaluation of pediatric epigenetic clocks across multiple tissues. *Clin. Epigenetics* **15**, 142 (2023).
54. Johnson, N. D. et al. Non-linear patterns in age-related DNA methylation may reflect CD4$^+$ T cell differentiation. *Epigenetics* **12**, 492–503 (2017).
55. Zhou, W. et al. SeSAMe: Reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. *Nucleic Acids Res.* **46**, 20 (2018).
56. Emmanuel, T. et al. A survey on missing data in machine learning. *J. Big Data* **8**, 140 (2021).
57. Li, P., Stuart, E. A. & Allison, D. B. Multiple imputation: a flexible tool for handling missing data. *JAMA* **314**, 1966–1967 (2015).
58. Paul, Y. et al. Considerations for normalization of DNA methylation data by Illumina 450K BeadChip assay in population studies. *Epigenetics* **8**, 1141–1152 (2013).
59. Bizzarri, D., Reinders, M. J. T., Beekman, M., Slagboom, P. E. & van den Akker, E. B. MiMIR: R-shiny application to infer risk factors and endpoints from Nightingale Health's ^{1}H-NMR metabolomics data. *Bioinformatics* **38**, 3847–3849 (2022).
60. Kwon, D. & Belsky, D. W. A toolkit for quantification of biological age from blood chemistry and organ function test data: BioAge. *GeroScience* **43**, 2795–2808 (2021).
61. Ying, K. et al. Biolearn, an open-source library for biomarkers of aging. Preprint at *bioRxiv* https://doi.org/10.1101/2023.12.02.569722 (2023).
62. Thrush, K. L., Higgins-Chen, A. T., Liu, Z. & Levine, M. E. R methylCIPHER: a methylation clock investigational package for hypothesis-driven evaluation & research. Preprint at *bioRxiv* https://doi.org/10.1101/2022.07.13.499978 (2022).
63. Pun, F. W. et al. Hallmarks of aging-based dual-purpose disease and age-associated targets predicted using PandaOmics AI-powered discovery engine. *Aging* **14**, 2475–2506 (2022).
64. Manning, A. K. et al. NHLBI BioData Catalyst and the future of cloud computing. *Genet. Epidemiol.* **45**, 774–775 (2021).
65. Li, T., Sahu, A. K., Talwalkar, A. & Smith, V. Federated learning: challenges, methods, and future directions. *IEEE Signal Process. Mag.* **37**, 50–60 (2020).
66. Fortier, I. et al. Maelstrom Research guidelines for rigorous retrospective data harmonization. *Int. J. Epidemiol.* **46**, 103–105 (2017).
67. Wilkinson, M. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
68. Tang, M. X. et al. The *APOE*-ε4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics. *JAMA* **279**, 751–755 (1998).
69. Benjamin, C. T. et al. Apolipoprotein E4 is associated with improved cognitive function in Amazonian forager–horticulturalists with a high parasite burden. *FASEB J.* **31**, 1251–1767 (2017).
70. Privé, F. et al. Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. *Am. J. Hum. Genet.* **109**, 12–23 (2022).
71. Graf, G. H. et al. Testing Black–white disparities in biological aging among older adults in the United States: analysis of DNA-methylation and blood-chemistry methods. *Am. J. Epidemiol.* **191**, 613–625 (2022).
72. Liu, Z. et al. A new aging measure captures morbidity and mortality risk across diverse subpopulations from NHANES IV: a cohort study. *PLoS Med.* **15**, e1002718 (2018).

73. Vandenbroucke, J. P. et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med.* **4**, e297 (2007).
74. Rodés, B., Cadiñanos, J., Esteban-Cantos, A., Rodríguez-Centeno, J. & Arribas, J. R. Ageing with HIV: challenges and biomarkers. *eBioMedicine* **77**, 103896 (2022).
75. Cohen, A. A. et al. Detection of a novel, integrative aging process suggests complex physiological integration. *PLoS ONE* **10**, e0116489 (2015).
76. Ferrucci, L. & Kohanski, R. Better care for older patients with complex multimorbidity and frailty: a call to action. *Lancet Healthy Longev.* **3**, e581–e583 (2022).
77. Baker, G. T. & Sprott, R. L. Biomarkers of aging. *Exp. Gerontol.* **23**, 223–239 (1988).
78. Levine, M. E. et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging* **10**, 573–591 (2018).
79. Sebastiani, P. et al. Biomarker signatures of aging. *Aging Cell* **16**, 329–338 (2017).
80. Mamoshina, P. et al. Population specific biomarkers of human aging: a big data study using South Korean, Canadian, and Eastern European patient populations. *J. Gerontol. A Biol. Sci. Med. Sci.* **73**, 1482–1490 (2018).
81. Zhang, Y. et al. DNA methylation signatures in peripheral blood strongly predict all-cause mortality. *Nat. Commun.* **8**, 14617 (2017).
82. Bernabeu, E. et al. Refining epigenetic prediction of chronological and biological age. *Genome Med.* **15**, 12 (2023).
83. van den Akker, E. B. et al. Metabolic age based on the BBMRI-NL ^{1}H-NMR metabolomics repository as biomarker of age-related disease. *Circ. Genom. Precis. Med.* **13**, 541–547 (2020).
84. Li, X. et al. Longitudinal trajectories, correlations and mortality associations of nine biological ages across 20-years follow-up. *eLife* **9**, e51507 (2020).
85. Hillary, R. F. et al. Epigenetic measures of ageing predict the prevalence and incidence of leading causes of death and disease burden. *Clin. Epigenet.* **12**, 115 (2020).
86. McCrory, C. et al. GrimAge outperforms other epigenetic clocks in the prediction of age-related clinical phenotypes and all-cause mortality. *J. Gerontol. A Biol. Sci. Med. Sci.* **76**, 741–749 (2021).
87. Tiina, F. et al. Does the epigenetic clock GrimAge predict mortality independent of genetic influences: an 18 year follow-up study in older female twin pairs. *Clin. Epigenetics* **13**, 128 (2021).

Acknowledgements

This work was supported in part by the Intramural Research Program of the National Institute on Aging, NIH, and grants from the National Institute on Aging and Hevolution Foundation. D.P.K. was supported by a grant from the National Institute of Arthritis and Musculoskeletal and Skin Diseases (R01AR041398). We are very grateful to D.M. Wilson III for many suggestions during the writing of this paper. We express our gratitude to all members of the Biomarkers of Aging Consortium Roadmap Group (https://www.agingconsortium.org/) for their fruitful discussions that helped to define the scope and direction of this work. The list of authors reflects Roadmap Group members who directly contributed to or refined the manuscript.

Competing interests

M.M., V.S., M.P.S. and V.N.G. have filed a patent on measuring cellular aging. C.H. is also affiliated with the Institute for Biomedical Aging Research, Universität Innsbruck, Austria and is an honorary research fellow at the Department of Women's Cancer, EGA Institute for Women's Health, University College London. C.H. is a shareholder of Sola Diagnostics and is named as an inventor on a patent on an epigenetic clock indicative of breast cancer risk. J.N.J. is also affiliated with the Sticht Center for Healthy Aging and Alzheimer's Prevention, Wake Forest University School of Medicine and the XPRIZE Foundation. J.N.J. serves on the advisory board for the American Federation for Aging Research's Finding Aging Biomarkers by Searching Existing Trials Initiative and the editorial board of the *Journals of Gerontology Series A Biological Sciences*, *eLife* and *Experimental Gerontology*. D.W.B. is also affiliated with the Child Brain Development Network, Canadian Institute for Advanced Research and SocioMed Research Nucleus and Universidad Mayor. D.W.B. is an inventor of DunedinPACE, a Duke University and University of Otago invention licensed to TruDiagnostic. A.T.H.-C. is an inventor of epigenetic clocks that are the subject of a provisional patent and have been licensed to TruDiagnostic. A.T.H.-C. has also received consulting fees from TruDiagnostic and FOXO Biosciences. B.H.C. owns stock in Illumina, the manufacturer of the DNA methylation arrays used in epigenetic biomarkers of aging, and is listed as a co-inventor on filed patents on commercial applications of epigenetic prediction models. A.A.C. is a founder, president and majority shareholder at Oken Health. R.E.M. has received a speaker fee from Illumina and is an advisor to the Epigenetic Clock Development Foundation and Optima Partners. M.W. is also affiliated with the Institute for Biomedical Aging Research, Universität Innsbruck. M.W. is a shareholder of Sola Diagnostics and is named as an inventor on a patent on an epigenetic clock indicative of breast cancer risk. K.F. is the CEO of BioAge Labs. P.O.F. is an employee and stakeholder of Gero. A.Z. is the founder and the CEO of Insilico Medicine, a clinical-stage generative AI and robotics biotechnology company specializing in aging research. N.B. is the scientific director of the American Federation for Aging Research, is on the board of the executive committee of the Longevity Biotech Association and is advisor on the board of the Academy for Health and Lifespan Research. D.P.K. has received a grant from Solarea Bio and royalties from Wolters Kluwer. D.P.K. sits on the scientific advisory boards of Solarea Bio, Pfizer, Radius Health and Reneo and has participated in the data safety monitoring board for the AgNovos Healthcare treatment trial. E.V. is a scientific cofounder of Napa Therapeutics and BHB Therapeutics, serves on the scientific advisory board of Seneque and is a named co-inventor on a patent relating to an epigenetic clock robust to cell composition changes. A.B.M. declares herself chief medical officer of NU and co-founder of Chi Longevity. V.S. is a cofounder, SAB chair and head of research of Turn Biotechnologies. M.P.S. is a cofounder and scientific advisor of Personalis, SensOmics, Qbio, January AI, Fodsel, Filtricine, Protos, RTHM, iollo, Marble Therapeutics, Crosshair Therapeutics and Mirvie. He is a scientific advisor of Jupiter, Neuvivo, Swaza and Mitrix. S. Horvath is a founder of the nonprofit Epigenetic Clock Development Foundation that licenses patents surrounding epigenetic clocks. The Regents of the University of California is the sole owner of a patent application directed at GrimAge and other epigenetic clocks for which S. Horvath is a named inventor.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41591-023-02784-9.

Correspondence and requests for materials should be addressed to Vadim N. Gladyshev, Steve Horvath or Luigi Ferrucci.

Peer review information *Nature Medicine* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Karen O'Leary, in collaboration with the *Nature Medicine* team.

[1]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [2]Department of Genetics, School of Medicine, Stanford University, Stanford, CA, USA. [3]Department of Obstetrics and Gynecology, School of Medicine, Stanford University, Stanford, CA, USA. [4]European Translational Oncology Prevention and Screening Institute, Universität Innsbruck, Innsbruck, Austria. [5]T.H. Chan School of Public Health, Harvard University, Boston, MA, USA. [6]Department of Internal Medicine, Section on Gerontology and Geriatric Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, USA. [7]Department of Epidemiology, Butler Columbia Aging Center, Mailman School of Public Health, Columbia University, New York, NY, USA. [8]Department of Psychiatry, Yale University, New Haven, CT, USA. [9]Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, San Diego, CA, USA. [10]Department of Environmental Health Sciences, Butler Columbia Aging Center, Mailman School of Public Health, Columbia University, New York, NY, USA. [11]Institute for Biostatistics and Informatics in Medicine and Ageing Research, Rostock University Medical Center, Rostock, Germany. [12]School of Medicine, University College Dublin, Dublin, Ireland. [13]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [14]Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK. [15]Department of Women's Cancer, EGA Institute for Women's Health, University College London, London, UK. [16]Department of Women's and Children's Health, Division of Obstetrics and Gynaecology, Karolinska Institutet, Stockholm, Sweden. [17]BioAge Labs, Richmond, CA, USA. [18]Gero PTE, Singapore, Singapore. [19]Insilico Medicine Hong Kong, Hong Kong SAR, China. [20]Institute for Aging Research, Albert Einstein College of Medicine, Bronx, NY, USA. [21]Department of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [22]Musculoskeletal Research Center, Hinda and Arthur Marcus Institute for Aging Research and Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA. [23]Healthy Longevity Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [24]Healthy Longevity Translational Research Program, Yong Loo Lin School of Medicine, National University of Singapore, Centre for Healthy Longevity, @AgeSingapore, National University Health System, Singapore, Singapore. [25]San Francisco Coordinating Center, California Pacific Medical Center Research Institute, San Francisco, CA, USA. [26]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA. [27]Section of Molecular Epidemiology, Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands. [28]Buck Institute for Research on Aging, Novato, CA, USA. [29]Department of Human Movement Sciences, @AgeAmsterdam, Amsterdam Movement Sciences, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [30]Altos Labs, San Diego, CA, USA. [31]National Institute on Aging, Baltimore, MD, USA. [32]These authors contributed equally: Mahdi Moqri, Chiara Herzog, Jesse R. Poganik. ✉e-mail: vgladyshev@rics.bwh.harvard.edu; shorvath@altoslabs.com; ferruccilu@grc.nia.nih.gov