

Predicting biological age based on the BBMRI-NL ¹H-NMR metabolomics repository

E.B. van den Akker^{1,3}, S. Trompet^{4,5}, J.J.H. Barkey Wolf¹, M. Beekman¹, H.E.D. Suchiman¹, J. Deelen^{1,6}, F.W. Asselbergs⁷⁻¹⁰, BBMRI-NL¹¹, E. Boersma¹², D. Cats¹, P.M. Elders^{13,14}, J.M. Geleijnse¹⁵, M.A. Ikram¹⁶⁻¹⁸, M. Kloppenburg^{19,20}, H. Mei^{1,21}, I. Meulenbelt¹, S.P. Mooijaart⁴, R.G.H.H. Nelissen²², M.G. Netea^{23,24}, B.W.J.H. Penninx^{14,25}, M. Slofstra²⁶, C.D.A. Stehouwer^{27,28}, M.A. Swertz²⁶, C.E. Teunissen²⁹, G.M. Terwindt³⁰, L.M. 't Hart^{1,14,31-33}, A.M.J.M. van den Maagdenberg³⁴, P. van der Harst³⁵, I.C.C. van der Horst³⁶, C.J.H. van der Kallen^{27,28}, M.M.J. van Greevenbroek^{27,28}, W.E. van Spil³⁷, C. Wijmenga²⁶, A. Zhernakova²⁶, A.H. Zwinderman³⁸, N. Sattar³⁹, J.W. Jukema⁵, C.M. van Duijn¹⁶, D.I. Boomsma^{14,40}, M.J.T. Reinders^{2,3}, P. Eline Slagboom^{1,6}

1. Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands
2. Leiden Computational Biology Center, Leiden University Medical Center, Leiden, The Netherlands
3. Department of Pattern Recognition and Bioinformatics, Delft University of Technology, Delft, The Netherlands
4. Department of Internal Medicine, Division of Gerontology and Geriatrics, Leiden University Medical Centre, Leiden, The Netherlands
5. Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands
6. Max Planck Institute for Biology of Ageing, Cologne, Germany
7. Department of Cardiology, Division of Heart and Lungs, University Medical Center Utrecht, Utrecht, The Netherlands
8. Durrer Center for Cardiovascular Research, Netherlands Heart Institute, Utrecht, the Netherlands
9. Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, United Kingdom

10. Farr Institute of Health Informatics Research and Institute of Health Informatics, University College
London, London, United Kingdom
11. BBMRI-NL: <https://www.bbmri.nl>; See Consortium Banner in Table S1
12. Thorax centre, Erasmus Medical Centre, Rotterdam, the Netherlands
13. Department of General Practice and Elderly Care Medicine, VU University Medical Center, Amsterdam,
The Netherlands
14. Amsterdam Public Health research institute, VU University Medical Center, Amsterdam, The
Netherlands
15. Division of Human Nutrition and Health, Wageningen University, Wageningen, The Netherlands
16. Department of Epidemiology, Erasmus University Medical Center Rotterdam, Rotterdam, The
Netherlands
17. Department of Radiology, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands
18. Department of Neurology, Erasmus University Medical Center Rotterdam, Rotterdam, The
Netherlands
19. Department of Rheumatology, Leiden University Medical Center, The Netherlands
20. Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands
21. Sequencing Analysis Support Core, Leiden University Medical Center, Leiden, The Netherlands
22. Department of Orthopaedics, Leiden University Medical Center, Leiden, The Netherlands
23. Department of Internal Medicine, Radboud Center for Infectious Diseases, Radboud University
Medical Center, Nijmegen, Netherlands
24. Department for Genomics & Immunoregulation, Life and Medical Sciences Institute (LIMES), University
of Bonn, Bonn, Germany
25. Department of Psychiatry, VU University Medical Center, Amsterdam, The Netherlands.
26. Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen,
The Netherlands
27. Department of Internal Medicine, Maastricht University Medical Center (MUMC+), Maastricht, the
Netherlands
28. School for Cardiovascular Diseases (CARIM), Maastricht University, Maastricht, the Netherlands

29. Neurochemistry Laboratory, Clinical Chemistry Department, Amsterdam University Medical Center, Amsterdam Neuroscience, the Netherlands
30. Department of Neurology, Leiden University Medical Center, Leiden, the Netherlands
31. Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, the Netherlands
32. Department of Epidemiology and Biostatistics, Amsterdam University Medical Center, Amsterdam, the Netherlands
33. Department of General practice, Amsterdam University Medical Center, Amsterdam, the Netherlands
34. Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
35. Department of Cardiology, University Medical Center Groningen, Groningen, The Netherlands
36. Department of Critical Care, University Medical Center Groningen, Groningen, The Netherlands.
37. UMC Utrecht, Department of Rheumatology & Clinical Immunology, Utrecht, The Netherlands
38. Department of Clinical Epidemiology, Biostatistics, and Bioinformatics, Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands
39. Institute of Cardiovascular and Medical Sciences, Cardiovascular Research Centre, University of Glasgow, Glasgow, UK
40. Netherlands Twin Register, Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands

Materials & Correspondence

Erik B. van den Akker, PhD; Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; Einthovenweg 20, 2333 ZC, Leiden, The Netherlands; Tel: +31 (0)71 526 85 57; Fax: +31 (0)71 526 82 80; E-mail: e.b.van_den_akker@lumc.nl; website: <http://www.molepi.nl>

ABSTRACT

The blood metabolome incorporates cues from the environment as well as the host's genetic background, potentially offering a holistic view of an individual's health status. We have compiled a vast resource of ^1H -NMR metabolomics and phenotypic data encompassing over 25,000 samples derived from 26 community and hospital-based cohorts. Using this resource, we constructed a metabolomics-based age predictor (metaboAge) to calculate an individual's biological age. Exploration in independent cohorts demonstrates that being judged older by one's metabolome, as compared to one's chronological age, confers an increased risk on future cardiovascular disease, mortality and functionality in older individuals. A web-based tool for calculating metaboAge (metaboage.researchlumc.nl) allows easy incorporation in other epidemiological studies. Access to data can be requested at bbmri.nl/samples-images-data. In summary, we present a vast resource of metabolomics data and illustrate its merit by constructing a metabolomics-based score for biological age that captures aspects of current and future cardio-metabolic health.

MAIN

Chronological age is an important risk factor for virtually all types of common disease, including diabetes mellitus type 2, cardiovascular disease and many forms of cancer [1]. Moreover, chronological age is often employed as an important criterion on which clinical treatment decisions in older adults are based. Yet, especially in the elderly, chronological age is a poor representative of an individual's intrinsic biological age, including the susceptibility to disease and resilience to treatment [2]. Hence, novel biomarkers are required that give additional information about the disparity between chronological and biological age, i.e. whether individuals are biologically older and potentially more vulnerable than their peers.

A range of multi-marker algorithms has been developed to serve as indicators of biological age. Examples are those based on physiological deterioration of organ systems from the second [3] or third [4] decade onwards, or those based on combined health deficits in later life, the so-called 'frailty' indices [5,6]. Others have exploited large quantities of highly standardized molecular data, e.g. DNA methylation data, to train so-called 'clock' algorithms [7-9] that allow one to calculate an omics-based age. The difference between an individual's actual chronological age and the estimated 'methylation age' was for instance shown to associate with mortality [10]. Interestingly, when compared, each of these omics-based biological age indicators appeared to mark unique aspects of ageing [11,12], giving ample incentive for the development of other, possibly complementary omics-based indicators of biological age. However, the requisite of a large body of data initially required for training similar omics age predictors is currently limiting the application of other molecular domains.

RESULTS

The BBMRI-NL resource

We present a novel, well standardized ^1H -NMR blood-based metabolomics data set encompassing over 25,000 samples collected by the Dutch Biobanking and BioMolecular resources and Research Infrastructure derived from 26 community and hospital based cohorts (**Figure 1 and Table S2**, available upon request at BBMRI-NL: bbmri.nl/samples-images-data, along with routine clinical variables). We have employed these data to construct a metabolomics-based clock (predictions made available as web resource (metaboage.researchlumc.nl, see **Online Methods** for instructions), and show that the difference between chronological age and 'metabolomic age' captures aspects of cardio-metabolic health.

Deriving a metabolomics-based score for biological age

A metabolomics predictor for chronological age was trained and evaluated (see **DocS1**) employing 56 out of 226 most reliable and independent [13] metabolomic variables (see **Table S3 and DocS2**), derived from 24 cohorts (**Figure 1**). Two biobanks missing a metabolomic variable were omitted (see **Online Material**). In addition, PROSPER and LLS_SIBS were left out from training the metabolomic age predictor and used to independently explore the predictive value of the obtained indicator of biological age. With use of the data of the remaining 22 biobanks comprising 18,716 samples (9,680 males and 10,036 females), a linear model was trained with the 56 metabolomic variables to estimate chronological age (see **Online Material**). A 5-Fold-Cross-Validation (5FCV, see **Online Material**) scheme was employed for randomly splitting the data in training (80%, 15,208

samples) and test (20%, 3,802 samples) sets for an unbiased training and evaluation of the models. The age-independent part of the difference between the estimated metabolomic age and chronological age (**Figure 1C**), hereafter referred to as $\Delta\text{metaboAge}$, may reflect for each individual the disparity between their biological and chronological age. Consequently, a high $\Delta\text{metaboAge}$ indicates a relatively ‘old’ blood metabolome for a given chronological age.

Associations of metaboAge with cardio-metabolic risk factors

In subsequent analyses we explored which aspects of biological age are marked by $\Delta\text{metaboAge}$. First, we investigated whether $\Delta\text{metaboAge}$ correlates with established clinical risk factors for cardio-metabolic disease using phenotypic data available within the BBMRI-NL resource (**Figure 1B**). Meta-analyses across biobanks showed that a positive $\Delta\text{metaboAge}$ corresponded with a poor cardio-metabolic health, as represented by higher body mass index (BMI), higher serum levels of C-reactive protein (hsCRP) and not unsurprisingly, higher cholesterol and triglycerides. In addition, use of blood pressure lowering medication, but not lipid lowering medication, is associated with a higher $\Delta\text{metaboAge}$ (**Figure 2A**). These associations remained significant when further adjusted for sex and BMI (see **Table S4**).

Associations of metaboAge with current and future cardio-metabolic disease

Next, we investigated whether $\Delta\text{metaboAge}$ marks current and future clinical metabolic disease endpoints. Participants with current metabolic syndrome or diabetes mellitus type 2 (T2D) were consistently estimated older as compared to their healthy counterparts of similar age (**Figure 2B**), with T2D remaining significant when also adjusting for sex and BMI (See **Table S5 in Online Material**). The predictive value of $\Delta\text{metaboAge}$ for future cognitive and

cardio-metabolic disease was tested in the PROSPER study, a multi-center clinical trial investigating the efficacy of lipid lowering medication for elderly patients (70 – 82 years) at risk of cardiovascular events followed for a median follow-up time of 3.3 years [14] (**Table S2**). While at most marginal correlations were observed between $\Delta\text{metaboAge}$ and measures of cognitive decline at baseline (**Table S6**) or during follow-up (**Table S7**), patients with a positive $\Delta\text{metaboAge}$ were shown to be at risk of future coronary and cardiovascular events independent of sex, BMI, smoking status, T2D status, anti-hypertensive medication and pravastatin treatment (**Figure 2C**). Using the same model, we found patients with a positive $\Delta\text{metaboAge}$ to be at increased risk of heart failure hospitalization and vascular and all-cause mortality (**Figure 2C**).

Associations of metaboAge with mortality and functionality in the oldest old

Finally we evaluated whether $\Delta\text{metaboAge}$ marks biological ageing near the extremes of human life span. We examined participants of the LLS-SIBS [15], aged 89 years and older and followed during a median follow-up time of 12.4 years for all-cause mortality (**Table S2**). At baseline, a positive $\Delta\text{metaboAge}$ correlated with lower Instrumental Activities of Daily Living (IADL, $p=2.0 \times 10^{-16}$), a measure of physical independence. Moreover, a positive $\Delta\text{metaboAge}$ also marked nonagenarians at an increased risk of all-cause mortality (**Figure 2D**) during ten years of follow up, even when adjusting for IADL (**Table S8**).

DISCUSSION

We present a rich resource of ^1H -NMR serum metabolomics and routine serum measurements encompassing over 25,000 samples, derived from 26 community and hospital based cohorts (download data access request at bbmri.nl/samples-images-data). Using this resource, we have constructed a score reflecting an individual's biological age, called metaboAge, and demonstrate that the excess of metaboAge over chronological age ($\Delta\text{metaboAge}$) confers an increased risk for future cardiovascular disease, mortality up to the highest ages and functionality among older adults. Lastly, we have made a web-based tool available at metaboage.researchlumc.nl facilitating an easy incorporation of $\Delta\text{metaboAge}$ scores in future epidemiological studies.

We evaluated the applicability of $\Delta\text{metaboAge}$ as a biomarker for current and future cardio-metabolic health and disease as the same metabolomics platform has previously been successfully employed to predict outcomes for cardiovascular disease [13] and type 2 diabetes [16]. In line with these papers, we observed that higher $\Delta\text{metaboAge}$ indicates various aspects of current and future cardio-metabolic health, including significant associations with body mass index ($p = 2.59 \times 10^{-33}$), C-reactive protein ($p = 1.76 \times 10^{-07}$), current Type 2 diabetes ($p = 5.10 \times 10^{-17}$), future cardiovascular events ($p = 2.64 \times 10^{-04}$) and vascular mortality ($p = 8.56 \times 10^{-07}$). Hence, $\Delta\text{metaboAge}$ can be readily explored, also in studies lacking cardio-metabolic risk factors or endpoints, as a surrogate marker to capture some aspects of current or future cardio-metabolic health.

Ideally, biomarkers of biological age are broadly applicable and are thus indicative of one or several of the five health domains as defined by Lara *et al.* [17]. Whereas we showed

that $\Delta\text{metaboAge}$ is indicative of classical biomarkers belonging to the physiological (cardiovascular health), immune (hsCRP), and physical capability domain (IADL), we were unable to establish significant correlations with classical biomarkers of the cognitive or endocrine domain. This was either because we lacked the classical biomarkers, as for the endocrine domain, or that $\Delta\text{metaboAge}$ did not correlate with the available classical biomarkers, as for the cognitive domain. Of note is that a measure not available to us, general cognitive ability, has recently been reported to associate with several metabolite measurements of this platform in a large epidemiological study [18]. Hence, we expect that future large-scale metabolomics studies using the Nightingale platform, e.g. the UK Biobank, will shed more light on other aspects of biological age indicated by $\Delta\text{metaboAge}$.

While the blood metabolome can be readily assessed using ^1H -NMR metabolomics at high throughput, high reproducibility and low costs, no ^1H -NMR metabolomics clock has to date been made available. We have applied the clock paradigm popularized by the work of Horvath *et al.* [8] to derive such a metabolomics-based predictor of age. Similarly, we have shown that our clock associates with various clinical endpoints including mortality. While clock algorithms have become increasingly popular as a means to perform sample stratification, an important limitation of the clock paradigm remains that it is hard to trace back why such scores reflect aspects of current and future disease, let alone for which disease applications a particular score is most suitable. Hence, newly proposed scores inevitably require additional empirical evidence in other epidemiological cohorts to support its added value. To accommodate future research with $\Delta\text{metaboAge}$, we have made a web-based tool available at metaboage.researchlumc.nl.

In summary, we present a rich resource of ^1H -NMR serum metabolomics and routine serum measurements encompassing over 25,000 samples (download data access request at bbmri.nl/samples-images-data). Moreover, we illustrate the merit of such a resource by presenting $\Delta\text{metaboAge}$, a novel metabolomics-based indicator of biological age capturing aspects of current and future cardio-metabolic health (predictions available at: metaboage.researchlumc.nl).

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

ACKNOWLEDGMENTS

See **Table S9**.

AUTHOR CONTRIBUTIONS

EBvdA and PES conceived and wrote the manuscript. EBvdA, ST and JJHBW performed the analyses. MJTR verified and supervised the analyses. PES, CMvD en DIB coordinated the BBMRI-NL consortium. MB, JD, HEDS coordinated sample collection and data acquisition. HM, MS, JJHBW, DC en MAS set up the database and catalogue. PES, FWA, EB, PME, JMG, MAI, MK, IM, SPM, RGHHN, MGN, BWJHP, CDAS, CET, GMT, LM'tH, AMJMvdM, PvdH, ICCvdH, CJHvdK, MMJvG WEvS, CW, AZ, AHZ, NS, JWJ, CMvD en DIB made phenotypic and metabolomics measurements available. All authors discussed the results and contributed to the final manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

1. Rae MJ, Butler RN, Campisi J, *et al.* The demographic and biomedical case for late-life interventions in aging. *Sci Transl Med.* 2010 Jul 14;2(40):40cm21.
2. Conroy SP, Westendorp RGJ and Witham MD. Hypertension treatment for older people-navigating between Scylla and Charybdis. *Age Ageing.* 2018 May 21. doi: 10.1093/ageing/afy053. [Epub ahead of print]
3. Mitnitski AB, Graham AJ, Mogilner AJ, *et al.* Frailty, fitness and late-life mortality in relation to chronological and biological age. *BMC geriatrics* 2002 2: 1-8, DOI: 10.1186/1471-2318-2-1
4. Levine ME and Crimmins EM. A comparison of methods for assessing mortality risk. *Am J Hum Biol.* 2014 Nov-Dec;26(6):768-76.
5. Rockwood K, Fox RA, Stolee P, *et al.* Frailty in elderly people: an evolving concept. *CMAJ.* 1994 Feb 15; 150(4): 489–495. PMID: PMC1486322
6. Belsky DW, Caspi A, Houts R, *et al.* Quantification of biological aging in young adults. *Proc Natl Acad Sci U S A.* 2015 Jul 28;112(30):E4104-10. PMID: 26150497 PMID: PMC4522793 DOI: 10.1073/pnas.1506264112.
7. Hannum G, Guinney J, Zhao L, *et al.* Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Mol Cell* 2013 49, 359–367, PMID: 23177740, DOI: 10.1016/j.molcel.2012.10.016
8. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013;14(10):R115. PMID: 24138928 PMID: PMC4015143 DOI: 10.1186/gb-2013-14-10-r115.
9. Peters MJ, Joehanes R, Pilling LC, *et al.* The transcriptional landscape of age in human peripheral blood. *Nat Commun.* 2015 Oct 22;6:8570. PMID: 26490707 PMID: PMC4639797 DOI: 10.1038/ncomms9570.

10. Marioni RE, Shah S, McRae AF, *et al.* DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol.* 2015 Jan 30;16:25. doi: 10.1186/s13059-015-0584-6.
11. Belsky DW, Moffitt TE, Cohen AA, *et al.* Eleven Telomere, Epigenetic Clock, and Biomarker-Composite Quantifications of Biological Aging: Do They Measure the Same Thing? *Am J Epidemiol.* 2017 Nov 15. doi: 10.1093/aje/kwx346. [Epub ahead of print]
12. Jylhävä J, Pedersen NL, Hagg S. Biological Age Predictors. *EBioMedicine* 2017 DOI: 10.1016/j.ebiom.2017.03.046
13. Wurtz P, Havulinna AS, Soininen P, *et al.* Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* 2015;131:774-85.
14. Shepherd, J., Blauw, G.J., Murphy, *et al.* PROspective Study of Pravastatin in the Elderly at Risk. PROspective Study of Pravastatin in the Elderly at Risk. Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial. *Lancet.* 2002;360:1623–1630.
15. Schoenmaker M, de Craen AJ, de Meijer PH, *et al.* Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet.* 2006 Jan;14(1):79-84.
16. Wurtz P, Wang Q, Kangas AJ, *et al.* Metabolic signatures of adiposity in young adults: Mendelian randomization analysis and effects of weight change. *PLoS Med* 2014;11:e1001765.
17. Lara J, Cooper R, Nissan J, *et al.* A proposed panel of biomarkers of healthy ageing. *BMC Med* 2015; 15;13:222.
18. Van der Lee SJ, Teunissen CE, Pool R, *et al.* Circulating metabolites and general cognitive ability and dementia: Evidence from 11 cohort studies. *Alzheimers Dement* 2018 14(6):707-722

FIGURE LEGENDS

FIGURE 1: *BBMRI-NL a vast ¹H-NMR metabolomics resource enabling approaches for personalized medicine*

A: Cohorts in the Dutch Biobanking and Biomolecular resources Research Infrastructure (BBMRI-NL), totaling 25,253 samples display interlinked age distributions robustly covering the complete adult lifespan from 18 till 85 years. While, VUNTR and CODAM (grey) were omitted for training the age predictor due to incomplete data (**Online Methods**), LLS_SIBS and PROSPER (boxed) were held out to independently evaluate the merit of age predictions as surrogate biomarkers for clinical endpoints. **B:** Additional omics data (orange) and phenotypic variables (blue) available within the BBMRI-NL resource. **C:** Flow chart of the analyses: a predictor for chronological age is trained on BBMRI-NL metabolomics data. The age-independent part of differences between predicted age and chronological age, termed $\Delta\text{metaboAge}$, is associated with endpoints. **D:** 5-Fold-Cross Validation (5FCV) is performed to assess the accuracy of the age predictor. Predictions on the test set of a representative fold are depicted, with $\Delta\text{metaboAge}$ exemplified in orange.

FIGURE 2: *Associations of $\Delta\text{metaboAge}$ with (risk factors of) cardio-metabolic disease risk and all-cause mortality*

Associations with **A:** Association of cardio-metabolic risk factors with $\Delta\text{metaboAge}$ in BBMRI-NL. **B:** Association of prevalent cardio-metabolic disease with $\Delta\text{metaboAge}$ in BBMRI-NL. **C:** Association of incident cardio-metabolic disease with $\Delta\text{metaboAge}$ in PROSPER. **D:** Association of mortality with $\Delta\text{metaboAge}$ in LLS_SIBS adjusted for age and sex. A Kaplan-Meijer curve illustrates the difference in mortality between quintiles with the highest (blue; estimated ≥ 6.9 years older) and lowest (red; estimated ≥ 7.3 years younger) $\Delta\text{metaboAge}$. β 's are reported as increase in $\Delta\text{metaboAge}$ per unit of increase in the risk factor (A) or disease status (B). HR's reported as increased risk per 10-years of $\Delta\text{metaboAge}$. *log-transformed; CI: confidence interval; med.: medication. P-values are in bold when significant after correction for multiple testing (Bonferroni).

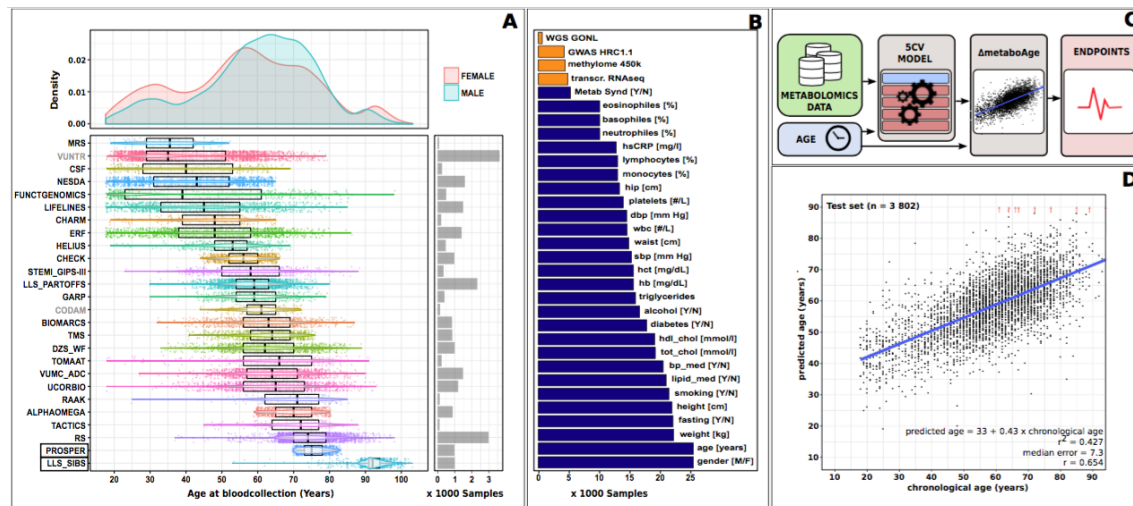


FIGURE 1

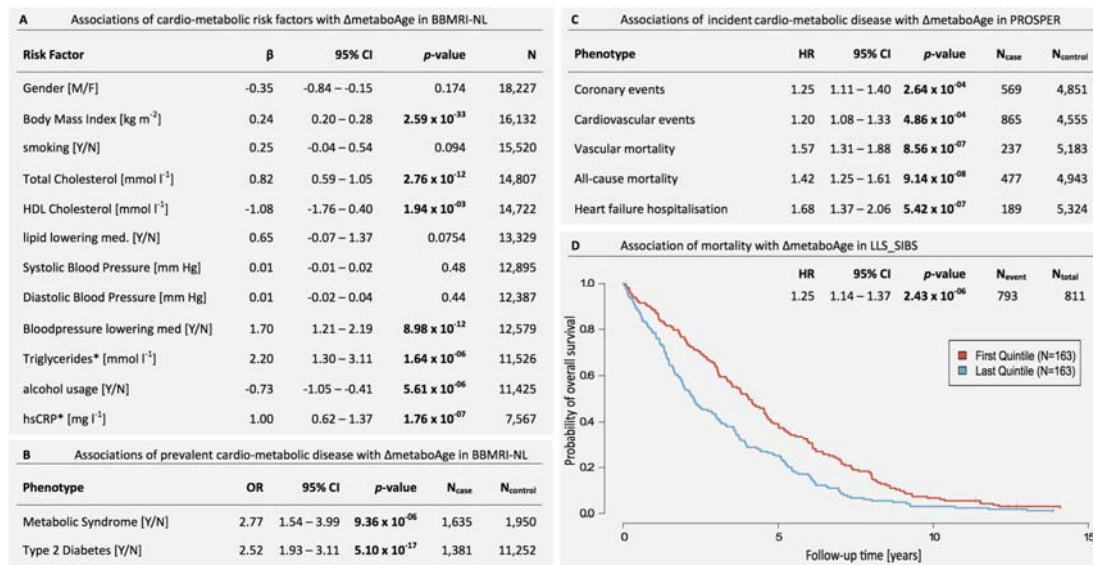


FIGURE 2

ONLINE METHODS

STUDY SAMPLE

The study sample was selected from the following 26 Dutch biobanks: MRS, VUNTR, CSF, NESDA, FUNCTGENOMICS, LIFELINES, HARM, ERF, HELIUS, CHECK, STEMI_GIPS-III, LLS_PARTOFFS, GARP, CODAM, BIOMARCS, TMS, DCS, TOMAAT, VUMC-ADC, UCORBIO, RAAK, ALPHAOMEGA, TACTICS, ERGO, PROSPER and LLS_SIBS, jointly participating in the Dutch Biobanking and BioMolecular resources and Research infrastructure (BBMRI-NL: <http://www.bbmri.nl>). Protocols for these studies were approved by the ethics committees at all involved institutes, and informed consent was obtained from all participants. Individuals with missing phenotypic data on age or sex (92) or aged under 18 years (5) were excluded, leaving 23,590 individuals in the current study. A description of the included cohorts is provided in **Table S2** in the **Supplementary Appendix**.

METABOLOMICS

Metabolite concentrations were measured in EDTA plasma samples using the commercially available high-throughput proton nuclear magnetic resonance ($^1\text{H-NMR}$) metabolomics platform of *Nightingale Health Ltd.* (Helsinki, Finland). In addition to the quantification of routine lipids, lipoproteins, fatty acid composition, amino acids, ketone bodies and gluconeogenesis-related metabolites, this platform also reports on average sizes of lipoprotein particle subclasses VLDL, LDL, and HDL and their respective lipid content ratios. In total, 226 metabolomic variables are reported for measurements in EDTA plasma, described in more detail elsewhere [1], of which 62 were analyzed in the current study (See **Table S3** in the **Supplementary Appendix** for a full list). The selection of 62 metabolomic variables was based on previous publications [2] and omits derived or highly correlating metabolomic variables from the dataset to prevent instability in the age predictor.

DATA PREPROCESSING

Cohorts reporting on all 62 metabolomic variables were included, i.e. omitting VUNTR (N=3,564) and CODAM (N=150) missing glutamine and acetoacetate respectively. Metabolomic variables measured at low success rates (<98%) or that frequently failed to reach the detection limit (<99%) were excluded (3-hydroxybutyrate, XXL_VLDL_L, XL_VLDL_L, L_VLDL_L, XL_HDL_L and L_HDL_L). Outlier samples were removed by allowing at most 1 missing value per sample (removing 200), 1 zero per sample (removing 64) and no metabolite levels more than 5 standard deviations away from the overall mean per metabolomics variable (removing 442). Obtained data (19,014 samples x 56 metabolomic variables) was scaled and the remaining 467 missing data points (0.046% of the data) were imputed using nipals of the R package pcaMethods. For more details, See **Supplemental Material Doc S2**.

STATISTICS

Constructing the metabolomics-based age predictor:

The metabolomics-based age predictor was trained using a linear model:

$$\text{Chronological Age} = \beta_0 + \beta_1 \times \text{met}_1 + \beta_2 \times \text{met}_2 + \dots + \beta_{56} \times \text{met}_{56} + \varepsilon \quad [1]$$

Two evaluation procedures were employed for getting an unbiased estimate of the model performance under two different scenarios. A 5-Fold-Cross-Validation (5FCV) was applied splitting the samples at random in a partition for model training, i.e. 'TRAIN' consisting of 80% of the samples, and an independent partition for model evaluation, i.e. 'TEST' consisting of 20% of the samples. Samples were drawn while preserving overall age-distribution using the function createFolds of R package caret. Secondly, a Leave-One-Biobank-Out-Validation (LOBOV) was applied by holding out the samples belonging to one particular cohort to create 'TRAIN' and evaluating these in the left-out samples, i.e. 'TEST'.

By lack of a single performance measure reporting faithfully under all conditions, we report two complementary measures, as was previously proposed for the methylation age predictor [3]. First, we report the Pearson correlation between chronological age and predicted age (r). Second, we report the median absolute difference between chronological and predicted age ('median error'). The range and median of both measures are reported across folds or cohorts for respectively the 5CV and LOBOV settings. For more details, See **Supplemental Material Doc S1**.

Associations of Δ metaboAge with other phenotypes and endpoints

To investigate whether age predictions might serve as surrogate biomarkers for biological age and age-related disease, we define Δ age as the predicted age minus the actual chronological age, i.e. the model residuals. For direct comparisons, we define Δ metaboAge as the age-independent part of Δ age, obtained by regressing Δ age onto chronological age. For associations with phenotypes we prefer to adjust for chronological age directly, e.g.:

$$\Delta\text{age} = \beta_0 + \beta_1 \times \text{chronological age} + \beta_2 \times \text{phenotype} + \epsilon \quad [2]$$

Associations of Δ metaboAge with phenotypes were performed for each cohort separately and meta-analyzed across cohorts on the estimated coefficients (β_2) as implemented in the function `rma` of R package `metaphor`. While analyses in BBMRI-NL were adjusted for chronological age and optionally sex and BMI, a more extensive model was evaluated in PROSPER to include also the canonical risk factors of cardiovascular disease, i.e. sex, BMI, current smoking status [YES/NO], diabetes mellitus type 2 status [YES/NO], hypertension status [YES/NO] and study covariates, i.e. pravastatin treatment [YES/NO]. Associations with cardio-metabolic phenotypes in BBMRI-NL were summarized using a random effect meta-analysis and β 's were reported as increase in Δ metaboAge per unit of increase in the risk factor or

disease status. The associations with cognitive tests were summarized across the three sub-cohorts of PROSPER using a fixed effect meta-analysis.

A Cox proportional hazards model with age at sampling as the time scale was used to test for the association between $\Delta\text{metaboAge}$ and incident endpoints in PROSPER and LLS-SIBS. While we only adjusted for sex and optionally IADL in the LLS_SIBS, we evaluated an extended model in PROSPER representing the aforementioned canonical risk factors for cardiovascular disease and study covariates. HR's are reported as increased risk per 10-years of $\Delta\text{metaboAge}$.

A linear mixed model was employed to test for the association between differences in time-course trajectories of cognitive tests with $\Delta\text{metaboAge}$:

$$\text{TEST} = \beta_0 + \beta_1 \times \Delta\text{metaboAge} + \beta_2 \times \text{tp} + \beta_3 \times \Delta\text{metaboAge} \times \text{tp} + \dots + \varepsilon \quad [3]$$

Associations of $\Delta\text{metaboAge}$ with cognitive test scores (one of: Stroop, Letter Digit Test (LTD), Picture Learning Test, immediate (PLTi) and delayed (PLTd)) were performed for each sub-cohort of PROSPER separately and fixed-effect meta-analyzed on the estimated coefficients of the interaction term (β_3) of $\Delta\text{metaboAge} \times \text{tp}$ (timepoint) and were adjusted for the aforementioned canonical risk factors.

metaboAge web-based tool

Metabolic age predictions can be obtained at <https://bbmri.nl.researchlumc.nl/metaboage>. The web-based tool requires the original tab-delimited files as supplied by Nightingale Health and will return metabolic age predictions. Data is transferred to the web-based tool using an encrypted connection (HTTPS), and the data will be stored only the duration of the computation and directly deleted thereafter. In case you would like to contribute your metabolomics data and participate in BBMRI-NL, please contact the corresponding author.

Data requests

Please visit bbmri.nl/samples-images-data and fill out and sign the data access request and code of conduct forms to request the data in this manuscript. Applications compliant with ethical and legal legislations will be reviewed by the BBMRI-NL board for overlap with other ongoing projects before access is granted.

REFERENCES

1. Soininen P, Kangas AJ, Würtz P, *et al.* Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet.* 2015 Feb;8(1):192-206.
2. Wurtz P, Havulinna AS, Soininen P, *et al.* Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* 2015;131:774-85.
3. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013;14(10):R115.

List of Supplemental Material

Table S1: BBMRI-NL Consortium banner.

Table S2: List of cohort descriptions.

Table S3: List of 56 metabolomic variables used to train the model.

Table S4: Associations with cardio-metabolic risk factors with $\Delta\text{metaboAge}$ in BBMRI-NL adjusted for sex and BMI.

Table S5: Associations of T2D status with $\Delta\text{metaboAge}$ in BBMRI-NL adjusted for sex and BMI.

Table S6: Associations of cognitive measures with $\Delta\text{metaboAge}$ in PROSPER adjusted for sex, BMI, current smoking status [YES/NO], prevalence of T2D [YES/NO] and hypertension status [YES/NO] and pravastatin treatment [YES/NO].

Table S7: Associations of change in cognitive measures with $\Delta\text{metaboAge}$ in PROSPER adjusted for sex, BMI, current smoking status [YES/NO], prevalence of T2D [YES/NO] and hypertension status [YES/NO] and pravastatin treatment [YES/NO].

Table S8: Associations of mortality with $\Delta\text{metaboAge}$ in LLS_SIBS, adjusted for sex and IADL.

Table S9: Acknowledgements.

Doc S1: Model training.

Doc S2: Quality control.