Future Medicine

Version: 15th January 2019

Title Page Template

- **Article title:** Comparison of DNA methylation clocks in Black South African men

- **Short running title:** Methylation clocks in Africans

- **Author names:**

Héléne Toinét Cronjé[1,2*], Cornelie Nienaber-Rousseau[1,6], Josine L. Min[3,4], Fiona R. Green[5], Hannah Rebecca Elliott[3,4], Marlien Pieters[1,6]

- **Author affiliations:**

[1]Centre of Excellence for Nutrition, North-West University, Potchefstroom, South Africa

[2]Section of Epidemiology, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

[3]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

[4]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

[5]Formerly, School of Biosciences and Medicine, Faculty of Health and Medical Sciences, University of Surrey, Guildford, UK

[6]Medical Research Council Unit for Hypertension and Cardiovascular Disease, North-West University, Potchefstroom, South Africa

- **Author for correspondence:** Héléne Toinét Cronjé (toinet.cronje@sund.ku.dk)

- **Author Contributions:**

HTC, MP, CNR and HRE conceptualized the study. Funding was acquired by MP and FRG. HTC performed the data analysis and wrote the manuscript. MP supervised the data analysis and interpreted the results with JM, HRE and HTC. All authors contributed to the critical review and editing of the manuscript.

- **Acknowledgements:**

The authors would like to thank all those who participated in the PURE-SA-NW study and those who made the PURE-SA-NW study possible, including the fieldworkers, researchers and staff of both the PURE-SA-NW (Africa Unit for Transdisciplinary Health Research, Faculty of Health Sciences, NWU, Potchefstroom, South Africa) and PURE International (S Yusuf and the PURE project office staff at the Population Health Research Institute, Hamilton Health Sciences and McMaster University, Ontario, Canada) teams. We also thank the staff of the Bristol bioresource laboratories (Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK) who generated the DNA methylation data.

- **Financial disclosure:**

The PURE-SA-NW study was funded by the North-West University, South African National Research Foundation (SANRF), Population Health Research Institute, South African Medical Research Council (SAMRC), the North West Province Health Department, and the South African Netherlands Partnerships in Development. Grants from the SANRF, Academy of Medical Sciences UK (Newton Fund Advanced Fellowship Grant [AMS-NAF1-Pieters to MP and FRG]) and the SAMRC funded the DNAm analysis. HTC was supported by the SANRF [SFH106264, MND 121094] and is currently supported by the Novo Nordisk Fonden Challenge Programme: Harnessing the Power of Big Data to Address the Societal Challenge of Aging [NNF17OC0027812]. HRE works in the Medical Research Council Integrative Epidemiology Unit at the University of Bristol, which is supported by the Medical Research Council and the University of Bristol (MC_UU_00011/5). Funding bodies were not involved in the design of the study,

collection, analysis or interpretation of the data or in writing of this manuscript. Opinions expressed and conclusions arrived at are those of the authors and are not to be attributed to the funding sources.

- **Information pertaining to writing assistance:** No

- **Ethical disclosure:** Ethical approval for the 2015 data collection of the PURE-SA-NW study was granted by the Health Research Ethics Committee of the North-West University (NWU-00016-10-A1, NWU-00119-17-A1). All participants provided written informed consent, including consent for genetic/epigenetic analysis.

- **Data sharing statement:** The data that support the findings of this study are available upon reasonable request and with the permission of the Health Research Ethics Committee of the North-West University and the principal investigator of the PURE-SA-NW study, Prof. I.M. Kruger (lanthe.kruger@nwu.ac.za) at the North-West University's Africa Unit for Transdisciplinary Health Research.

Comparison of DNA methylation clocks in Black South African men

Abstract

Aims: DNA methylation (DNAm) clocks are widely used to estimate biological age, although limited data are available on non-European ethnicities. Here, we characterize the behavior of five DNAm clocks in 120 older Black South African men.

Methods: We investigate the accuracy in age estimation by the Horvath, Hannum and skin and blood clocks and the relative age-related mortality risk and predicted time to death portrayed by the PhenoAge and GrimAge estimators, respectively.

Results: We confirm the tendency of DNAm clocks to underestimate the biological age of older individuals. GrimAge more accurately characterized biological decline compared to the PhenoAge in this cohort, owing to the unique inclusion of smoking-related damage in the GrimAge estimate.

Conclusion: Each clock provides a different fraction of information regarding the aging body. It is essential to continue studying under-represented population groups to ensure methylation-derived indicators are robust and useful in all populations.

Keywords: GrimAge, PhenoAge, phenotypic age, biological age, smoking

Introduction

As global life expectancy continues to increase, the chronic disease burden expands and the need for a better understanding of how to promote healthy aging is emphasized [1]. Chronological age is an integral component of frailty, non-communicable disease risk and mortality [2]. Although easily accessible and standardized, chronological age as a biomarker is limited by its inability to portray changes in biological functionality accurately over the life span, especially in later life [3]. A group of peers, for example, may be the same chronological age, while exhibiting a spectrum of age-related deterioration [3]. For this reason, extensive efforts have been made to develop markers that are able to reflect biological aging better than years since birth do [3, 4]. Ultimately, the availability of such markers could allow for improved targeted intervention through the identification of high-risk, functionally declining individuals before clinical symptoms appear [4].

DNA methylation (DNAm) refers to attachment of a methyl group to a DNA base. DNAm changes accumulate with age [5, 6] and are thought to mediate the effects of environmental risk factors on disease [7]. DNAm levels at specific cytosine-phosphate-guanine sites (referred to as clock CpGs), can also be used to predict chronological age [4, 8-10]. These predictors are termed "epigenetic clocks" and quantify (in years) a biological age estimate [4]. Residuals from the regression of epigenetic age on chronological age are defined as DNAmAge acceleration (DNAmAgeAccel). Positive DNAmAgeAccel (biological age is projected to be older than chronological age) has been associated with diabetes [11], cancer [12], cardiovascular disease [13] and all-cause mortality [10, 12, 14], although causality still has to be established [4, 15].

To date, multiple DNAmAge clocks have been developed, with some variation in composition and outcome (Summarized in Table 1). The Horvath [8], Hannum [16] and skin and blood (SB, [10]) clocks are widely used, because of their ability to robustly predict either the chronological age of unknown donors or biological age discrepancies in various tissues in a single individual. Because

these age predictor clocks were developed with chronological age as the sole outcome of interest, they often fail to capture the inter-individual methylation differences that discern biological decline above that of advancing age itself [3, 4]. The more recent development of mortality predictor clocks; PhenoAge [17] and GrimAge [18], address this limitation. Instead of relying solely on chronological age, these models incorporate a composite outcome of aging-related clinical measurements that differentiate between healthy and unhealthy aging (Table 1). The mortality predictor clocks are particularly important in the context of health research, because they have been developed using longitudinal data and well-defined mortality outcomes and are, therefore, able to generate useful biological age biomarkers. This gives cohorts without such data, such as the PURE-SA-NW study, an opportunity to use methylation-derived biomarkers supported by causal models to indicate the health risks of the study populations.

Table 1. Descriptive characteristics of widely used epigenetic clocks

Name	Estimates	DNAm data description*	Clock computation
Horvath [8]	DNAmAge which reflects chronological age in the majority of healthy tissues. Where DNAmAge and CA differ DNAmAgeAccel is calculated as a measure of faster/slower relative biological aging.	27K/450K probes analyzed in 51 tissues and cell types.	ENR of CA on DNAm returned **353** CA-associated CpGs.
Hannum [16]		450K probes tested on whole blood.	ENR of CA on DNAm returned **71** CA-associated CpGs.
Skin and Blood [10]		450K/EPIC probes tested on peripheral tissues used for *ex vivo* studies.	ENR of CA on returned **391** CA- associated CpGs.
PhenoAge [17]	Biological age based on physiological dysregulation that relates to mortality risk.	450K/EPIC probes tested on whole blood.	ENR of Phenotypic age (PA) on DNAm returned **513** PA-associated CpGs.
GrimAge [18]	Biological age calculated as an estimation of time-to-death.	450K/EPIC probes tested on whole blood.	ENR of 7 protein markers and smoking pack-years on DNAm deliver 8 DNAm surrogates that amount to a collective biomarker of **1030** CpGs.

CA: Chronological age; DNAm: DNA methylation; ENR: Elastic net regression; PA: Weighted average of albumin, alkaline phosphatase, creatinine, C-reactive protein, serum glucose, mean cell volume, lymphocyte percentage, red

cell distribution width, white blood cell counts and chronological age. *Naming of two platform denotes the use of only overlapping probes; 27K/450K/EPIC refers to the various Illumina Infinium Methylation BeadChips.

While these clocks are widely used and studied, a number of limitations in their use have been identified. Firstly, a systematic underestimation of Horvath and Hannum DNAmAge in older adults has been reported [19], necessitating further investigation into their use in older populations. Secondly, although ethnic differences in the behavior of the epigenetic clocks have been reported [17, 20, 21], most of the current literature represent data obtained from individuals of European ancestry only, with very limited information being available on other ethnic groups. In addition to genetic confounding, environmental differences between populations also influence the physiological features these models are built on, particularly when lifestyle factors are incorporated such as in the case of the GrimAge clock where the damaging effects of smoking are captured. For this reason, replication in various cohorts should continually be pursued and clocks should be refined, when previously unmeasured confounding becomes known.

We, therefore, compared the behavior of five epigenetic clocks [8] in a group of black South African men between the ages of 45 and 88. We investigated the accuracy in age estimation by the Horvath, Hannum and SB clocks (and the potential issue of underestimation in older adults) and the relative age-related mortality risk and predicted time to death portrayed by the PhenoAge and GrimAge estimators, respectively, in a non-European population with a particularly high prevalence of smoking. Because many datasets of existing cohorts possess DNAm data, mostly generated using the Illumina platforms, understanding the multitude of uses of such DNAm data is essential. For this reason we restrict our analysis to the most frequently used DNAm clocks in the epigenetic literature [15, 18, 20, 22, 23] developed using the Illumina platform and available on the online Methylation age calculator ([8, 10], https://dnamage.genetics.ucla.edu/new).

Methods

Study population

The international Prospective Urban and Rural Epidemiology (PURE) study comprises cohorts from 27 countries tracking participants over a period of 20 years. In this manuscript we report cross-sectional data from the North-West arm of the PURE study in South Africa (PURE-SA-NW). A subset of ostensibly healthy black South African men (n = 120) with available peripheral blood samples were randomly selected for this study, provided they tested negative for the human immunodeficiency virus at the time of data collection (2015). Eligibility was restricted to reduce confounding by sex and CD4-T cell count. Additional information on the international cohort [24] and this sub-study [25] has been published.

Data collection

Participants reported their current smoking and alcohol use via interview. If applicable, they reported the frequency and quantity of intake, age at the start of use and previous attempts at abstinence. The Cobas Integra 400 (Roche Diagnostics®, Indianapolis, IN, USA) was used to quantify serum albumin, creatinine and high-sensitivity C-reactive protein. Peripheral blood collected in fluoride tubes were used to measure glucose concentrations using the same device. Serum akaline phosphatase was quantified using the Cobas Integra 400 plus (Roche®, Basel, Switzerland).

Peripheral blood samples were used for DNA extraction. Genome-wide methylation data were generated by the Bristol Bioresource Laboratory (Bristol Medical School, University of Bristol, Bristol, UK) using the standard protocol of the Illumina Infinium MethylationEPIC BeadChip (Illumina®, San Diego, CA, USA) platform. Quality control, sample filtering and functional normalization were done using the *meffil* [26] package in R 3.4.3 [27]. A detailed description of the DNA extraction, quality control, methylation quantification and data processing protocols has been published [25]. Lymphocyte proportions were methylation-derived [25, 28].

Cell counts and DNAmAge

The IDOL optimised L-DMR library for whole blood samples [28] was used to estimate the distribution of B-, CD4-T, CD8-T, neutrophil, monocyte and natural killer cells. For the estimation of DNAmAge and DNAmAgeAccel, a widely used online calculator was used ([8], https://dnamage.genetics.ucla.edu/new). The following age estimates were used as provided in the calculator output: DNAmAge, DNAmAgeHannum, EEAA, DNAmAgeSkinBloodClock, DNAmPhenoAge, DNAmGrimAge, BloodskinAA, PhenoAA and GrimAA. Intrinsic epigenetic age acceleration (IEAA) was calculated using the residuals from a linear model where DNAmAge was used as the outcome and chronological age, the calculator's plasmaBlast, CD8pCD28nCD45Ran and CD8.naive and IDOL-estimated CD4-T, natural killer cells, monocytes and neutrophils as predictors. As opposed to the extrinsic epigenetic age acceleration estimate which includes leukocyte cell distribution in its algorithm, IEAA represents a cell-count-adjusted age acceleration (AA) estimate. Note that the Illumina Infinium MethylationEPIC array used to quantify methylation data for the current study excludes 19 of the 353 Horvath and six of the 71 Hannum clock CpGs [8, 16]. The absence of these CpGs has previously been reported not to compromise the accuracy of these clocks [29]. All further cell-count adjustments were performed using cell estimates from the IDOL package.

Statistical analysis

Data normality was evaluated using the Shapiro-Wilks test. We report the mean and standard deviation of the chronological age, DNAmAge and DNAmAA estimates in this study population. We also report Pearson's correlation coefficients between chronological age and each of the DNAmAge estimates. Figures 1 and 2 were compiled with the *ggplot2* and *BlandAltmanLeh* packages. In Table 3, chronological age, lymphocyte fraction and logarithmically transformed akaline phosphatase, serum albumin, creatinine, high-sensitivity C-reactive protein and glucose

concentrations were correlated with PhenoAge and PhenoAA. We also compared the means of these clinical components of phenotypic age between current and never smokers using t-tests.

Type III analysis of variance models in the *car* package were applied to linear regression objects to quantify the differences in outcome means between smokers and non-smokers, reported in Table 4. Adjusted group means and standard errors were extracted using the *effects* package. Two models were run for each outcome. First, a model adjusting for chronological age, body mass index (BMI), education and white blood cell (WBC) counts, and then, a second model, additionally adjusting for alcohol consumption with the former variables. Covariates were chosen based on current literature to ease comparability [30]. R version 3.5.0 was used for all analyses [27].

Results

In this study we evaluate the behavior of five epigenetic clocks in 120 black men from the PURE-SA-NW study aged 45 to 88 years. Approximately half (61 and 58) of the 120 participants were current smokers and alcohol consumers, respectively, and 48 participants (40% of the cohort) were using both substances at the time of data collection. The majority of the study population were of normal weight, with only 21% classified as overweight (BMI ≥25 – 29.9 kg/m^2) and 8% as obese (BMI ≥30 kg/m^2). Fifty-five percent of the study population had undergone 1–7 years of schooling, while 23% had undergone 8–12 years. The remainder of the population had received no formal education.

Comparison of biological age and age acceleration estimates with chronological age

In Table 2 we report the means and standard deviations of three age predictor and two mortality predictor DNAmAge and DNAmAgeAccel estimates. The association of each DNAmAge estimate with chronological age is also reported. Figures 1 and 2 depict scatterplots comparing the age estimations by the age predictor and mortality predictor clocks, respectively. On the left, chronological age is plotted against the difference between each clock's estimated age and chronological age (linear regression line of best fit depicted in blue). On the right, Bland Altman plots depict the relationship between the mean of each estimated and chronological age against the difference between each estimated and chronological age. The age- and mortality predictor clocks are discussed separately in the following sections, because of the integral differences in their aims and outcomes.

Table 2. Descriptive characteristics of age, DNAmAge and DNAmAgeAccel in the study sample

Age measure (years)	**Mean ± SD**	**Correlation with chronological age**	
		r	**P**
Chronological Age	63 ± 10		
Age predictor clocks			
HorvathAge	59 ± 8	0.58	2.3E–12
HannumAge	47 ± 8	0.64	2.2E–15
SBAge	54 ± 8	0.70	7.4E–19
IEAA	0 ± 6.4		
EEAA	0 ± 7.6		
SBAA	0 ± 5.5		
Mortality predictor clocks			
PhenoAge	47 ± 9	0.51	2.7E–09
GrimAge	64 ± 9	0.80	1.9E–28
PhenoAA	0 ± 7.5		
GrimAA	0 ± 5.3		

AA: Age acceleration; EEAA: Extrinsic epigenetic age acceleration; IEAA: Intrinsic epigenetic age acceleration; SBAge: Skin and blood age; SBAA: Skin and blood age acceleration; SD: Standard deviation

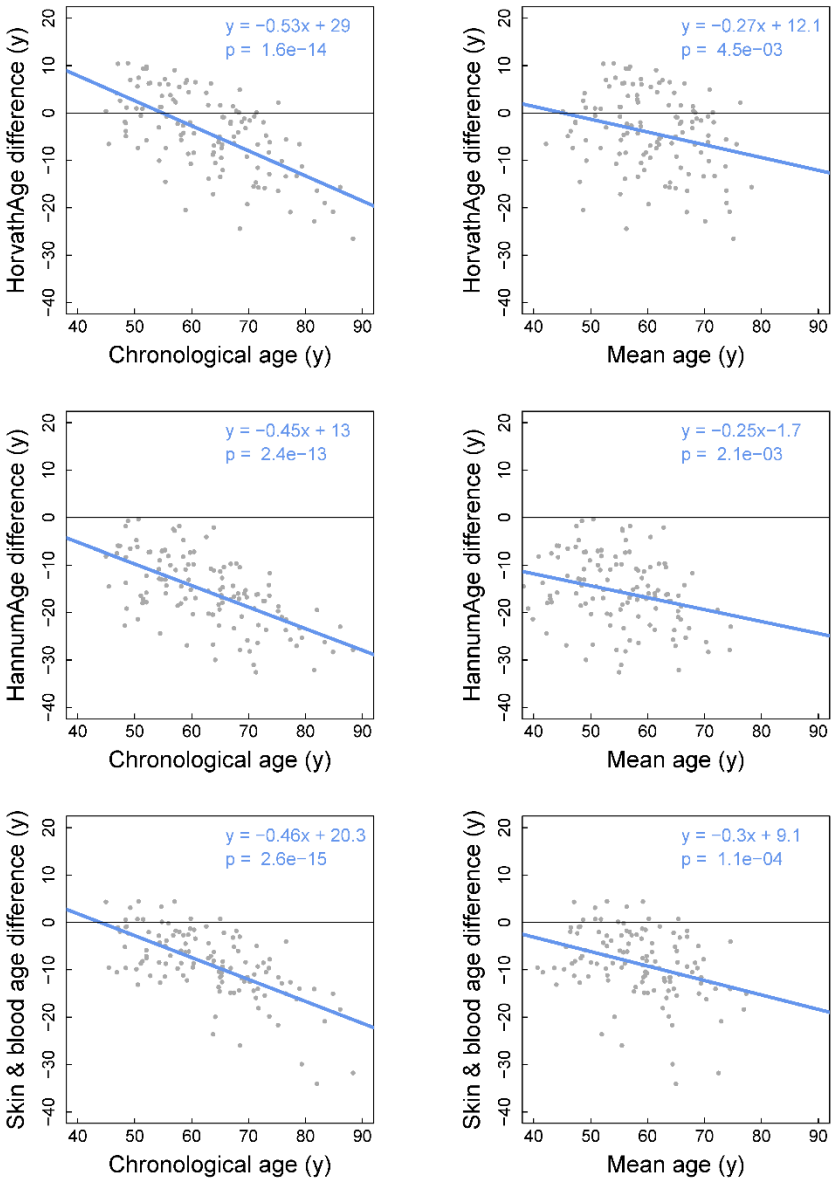

Figure 1 Scatterplots illustrating the relative difference in biological vs chronological age by three age predictor clock DNAmAge estimates.

Left: Scatterplots depicting the relationship between chronological age and the age difference when subtracting chronological age from the DNAmAge estimates. Right: Bland Altman plots depicting the mean of each DNAmAge and chronological age against the difference between each DNAmAge and chronological age. The line of best fit from a linear regression model is formulated as y = mx + c in the top right corner and depicted in blue on the plot. The p-value represents the statistical significance of the linear regression model.

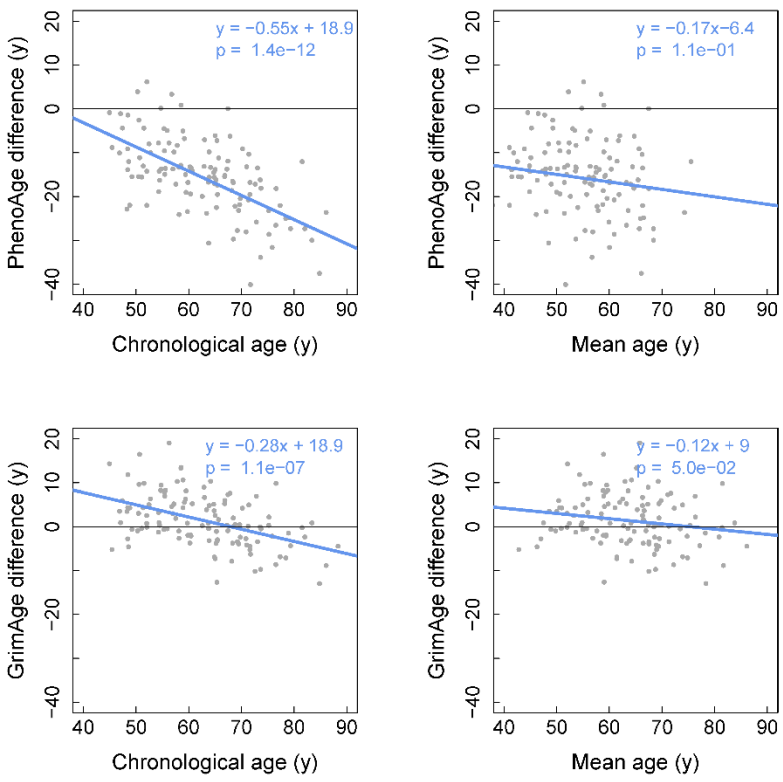

Figure 2 Scatterplots illustrating the relative difference in biological vs chronological age by two mortality predictor clock DNAmAge estimates.

Left: Scatterplots depicting the relationship between chronological age and the age difference when subtracting chronological age from the DNAmAge estimates. Right: Bland Altman plots depicting the mean of each DNAmAge and chronological age against the difference between each DNAmAge and chronological age. The line of best fit from a linear regression model is formulated as y = mx + c in the top right corner and depicted in blue on the plot. The p-value represents the statistical significance of the linear regression model.

Age predictor clocks

SBAge presented with the highest correlation with chronological age, although HorvathAge gave the closest estimate of mean age (Table 2). HorvathAge correlated with HannumAge and SBAge with a correlation coefficient of 0.62 (p = 7.7E–14) and 0.64 (p = 2.1E–5), respectively. HannumAge and SBAge associated with each other more strongly (r = 0.71, p = 6.2E–20). In terms of AA measures, Horvath's IEAA correlated with Hannum's EEAA with a correlation coefficient of 0.31 (p = 6.0E–

4), while equally strong correlations were observed between the SBAA and the IEAA (p = 3.8E–06) and EEAA 0.41 (p = 2.8E–06), respectively.

All three clocks tended to underestimate chronological age in a systematic fashion with this deviation increasing with age (Figure 1). The trend of underestimation was stronger for HorvathAge than HannumAge and SBAge, although the HannumAges were consistently lower than the other two. The Horvath clock was most accurate at a chronological age of 55 years (the age at which the regression line intercepted the y-axis, Figure 1). The Hannum and SB clocks seemed to be most accurate for individuals younger than those represented in our cohort, at the chronological ages of 29 and 44 years, respectively. Additional adjustment for WBC count variation did not significantly alter the degree of underestimation.

Mortality predictor clocks

A seventeen-year difference was observed between the biological age estimation of the PhenoAge and GrimAge clocks (47 vs 64 years, Table 2). Biological age and AA estimates from the PhenoAge and GrimAge clocks correlated with a coefficient of only 0.51 (p = 4.2E–09) and 0.21 (p = 0.02), respectively. Similar to the Hannum and SB clock estimates, the PhenoAge clock likely performs best at ages outside of the PURE-SA-NW age range, as shown in the simulation on PURE-SA-NW data where the y-axis intercept is 34.3 years (Figure 2). For the GrimAge clock, however, optimal estimation is at age 67.4 years, which is the highest optimal age of all the tested clocks and is also closest to the mean age of our cohort here. Additional adjustment for WBC count variation did not significantly alter the degree of underestimation.

<u>Role of smoking in the mortality predictor clocks</u>

Because more than half of the PURE-SA-NW study population were current smokers, we hypothesized that the difference in the contribution of smoking to the PhenoAge and GrimAge

estimates might explain the large discrepancy between the performances of these two clocks. PhenoAges of 47.4 ± 9 and 46.8 ± 9 years, respectively were observed when never and current smokers were compared (p = 0.71). No differences between the PhenoAA of smokers (–0.35 ± 7.96) and non-smokers (0.49 ± 7.23) were observed (p = 0.55) either. In contrast, GrimAges of non-smokers were on average five years younger than that of smokers (61.7 ± 9 vs 66.7 ± 9, p = 0.003) while a four and a half year difference was observed in AA comparisons (–2.39 ± 4.54 vs 2.20 ± 5.12, p = 1.2E–06).

To explore this further, we evaluated the association of smoking, PhenoAge and PhenoAA with seven of the 10 clinical components (protein markers) of phenotypic age that were available for the PURE-SA-NW study cohort (Table 3). We also investigated the association of smoking with each of the eight methylation surrogate markers that comprise the GrimAge estimate (Table 4).

Table 3. Comparison of seven clinical components of phenotypic age between current and never smokers and each component's association with PhenoAge and chronological age

Clinical components of phenotypic age	**Never smoker** (n = 56)	**Current smoker** (n = 61)	**t-test p**	**PhenoAge correlation**		**PhenoAA correlation**	
	Geometric mean ± SD			**r**	**p**	**r**	**p**
Age (years)	63 ± 10	63 ± 10	0.78	0.51	**2.7E–09**	0.00	0.96
Albumin (g/L)	42.1 ± 1.19	44.2 ± 1.17	0.10	–0.15	0.10	–0.11	0.23
ALP (U/L)	83.9 ± 1.45	85.1 ± 1.39	0.82	0.04	0.63	0.04	0.69
Creatinine (umol/L)	0.08 ± 0.03	0.07 ± 0.02	**0.04**	0.10	0.30	–0.06	0.54
CRP (mg/dL)	3.62 ± 4.23	3.00 ± 3.37	0.45	0.39	**1.1E–05**	**0.36**	**9.0E–05**
Glucose (mmol/L)	5.21 ± 1.20	4.99 ± 1.16	0.19	–0.03	0.72	–0.07	0.45
Lymphocyte (%)	37.2 ± 9.30	36.6 ± 10.5	0.75	–0.31	**4.8E–04**	**–0.33**	**2.9E–04**

ALP: Alkaline phosphatase, CRP: C-reactive protein. All components were log transformed apart from age and lymphocyte %. Lymphocyte estimates are methylation-derived. Phenotypic age components for which no data are available are mean cell volume, WBC counts, and red cell distribution width. p <0.05 highlighted in bold.

Apart from Albumin, none of the available clinical components of phenotypic age differed significantly between smokers and non-smokers. PhenoAge correlated with chronological age, C-reactive protein and lymphocyte percentage, while PhenoAA, correlated only with the latter two (Table 3).

Table 4. Adjusted group means of aging-related phenotypes for current vs never smokers

Outcome	**Model**	**Smoking status** Group mean ± SE		
		Never (n = 56)	**Current** (n = 61)	**p**
GrimAge	1	62.2 ± 0.68	66.6 ± 0.65	2.0E–05
	2	62.0 ± 0.79	66.7 ± 0.76	3.5E–04
GrimAA	1	–2.27 ± 0.68	2.19 ± 0.65	2.0E–05
	2	–2.4 ± 0.79	2.31 ± 0.76	3.5E–04
DNAmPackY	1	19.9 ± 1.6	35.1 ± 1.6	4.2E–09
	2	19.5 ± 1.9	35.5 ± 1.8	6.5E–07
DNAmLeptin	1	6448 ± 541	8514 ± 520	1.1E–02
	2	6450 ± 634	8512 ± 604	4.5E–02
DNAmPAI-1	1	16657 ± 503	18757 ± 483	5.5E–03
	2	16956 ± 586	18479 ± 559	1.1E–01

Group N: Never = 56, Current = 61. AA: Age acceleration; PackY: Smoking pack years. Variables prefaced by DNAm are the methylation-derived surrogates of the following component as used for the GrimAge estimate. Model 1: Outcome ~ smoking status + chronological age + body mass index + education + WBC counts; Model 2: Outcome ~ smoking status + chronological age + body mass index + education + WBC counts + alcohol consumption status.

Apart from the smoking pack-years methylation component (DNAmPackY), DNAmPAI-1 and DNAmLeptin also differed between current and never smokers (Table 4). Because alcohol

consumption often coincides with smoking habits, we performed a sensitivity analysis additionally adjusting for alcohol use (Model 2). This did not significantly alter the associations observed with smoking status, apart from an attenuating effect on DNAmPAI-1.

As a sensitivity test, we investigated potential bias in our self-reported smoking status data by analyzing the association between the continuous methylation % of cg05575921 with each of the PhenoAge and GrimAge constituents. Previous literature has suggested cg05575921 is a less biased estimate of smoking behavior than self-report [31]. Again, only GrimAge and GrimAA and not PhenoAge or PhenoAA associated with cg05575921. Similar to self-reported smoking comparisons, the only PhenoAge component associated with cg05575921 was creatinine and for the GrimAge components, associations with smoking pack years and DNAmPAI-1 and DNAmB2M were observed. Table 4 only includes results at $p < 0.05$. For the full Table, consult the supplementary material (Table S1).

Discussion

This study compared five epigenetic clocks previously proven to be highly effective in reaching their respective aims of chronological age prediction (age predictor clocks) and prediction of mortality-related functional decline (mortality predictor clocks). Our results echoes previous findings of discrepancies between the predicted biological ages generated by the different clocks. We also confirm the tendency of all the tested clocks to underestimate the biological age of older individuals. We demonstrate that the GrimAge best fits the aging of a study sample of continental African men with a high prevalence of smoking and alcohol use.

Comparison of biological age and age acceleration estimates with chronological age

Age predictor clocks

The average chronological ages predicted by the three age predictor clocks spanned 12 years (47–59 years) and were all younger than the average calendar chronological ages reported by the PURE-SA-NW participants (63 ± 10 years). The ages predicted by these clocks are known to differ within the same dataset, in part because of the limited overlap in represented CpGs and the likelihood that each clock (and the clock CpGs it contains) captures varying degrees of cell count, environmental and ethnic influences or confounding [4, 23, 32]. The SB clock shares 45 loci with the Hannum and 60 with the Horvath clocks [10], while the latter two share only five loci [8, 16]. This larger overlap and the similarity of the target tissue on which these clocks were trained (whole blood), is reflected in the stronger correlation between the PURE-SA-NW estimated Hannum and SB age estimates than their respective correlations with HorvathAge.

Correlations between the ages estimated by the age predictor clocks and chronological ages in the PURE-SA-NW cohort were comparatively weaker than those previously reported. In the SBAge validation analysis conducted by Horvath et al. [10], chronological age and predicted age correlated at a coefficient of 0.96 for HorvathAge, 0.97 for HannumAge and 0.98 for SBAge [10]. We

postulate the weaker correlations observed in our data than to Horvath's validation data [10] are likely due to the wider range of ages represented in the validation than the current study (19 to 82 [10] compared to 45 to 88 years).

In agreement with the SBAge having the strongest correlation with chronological age, SBAA also displayed the least amount of AA variance, followed by the IEAA and then EEAA measures. These differences in AA variation probably occurred because of the varying role of WBC counts within the different DNAmAge algorithms from which each AA was derived. Based on the fact that WBC composition changes with age, the EEAA incorporates WBC changes in a weighted manner by aggregating the HannumAge estimate with plasmablasts, naïve cytotoxic T-cells and exhausted cytotoxic T-cells [16, 33]. This can, however, introduce inter-individual variation, becauseWBC composition is influenced by factors apart from aging itself, such as sex, medication use, disease and ethnicity and genetic factors [33-37]. For the IEAA, however, inter-individual variation is reduced by adjustment for cell counts. The adjustment is specifically made to optimize the performance of the multi-tissue predictor (HorvathAge) when it is applied to blood samples [8, 33]. The SBAA consequently outperforms both these estimates, as it was initially developed for blood samples and, therefore, needs no additional WBC count adjustments, nor does it introduce potential additional variance by incorporating differential WBC composition [10].

The SB clock was developed to improve the accuracy of age estimation of the Horvath clock (by limiting training data to a smaller number of cell types) and Hannum clock (by using bigger training datasets and increasing represented loci). In our cohort, the superiority of the SB clock is confirmed in its higher correlation with chronological age than the other clocks and its improvement in AA variation, in samples derived from whole blood.

Mortality predictor clocks

The PhenoAge clock estimated the PURE-SA-NW cohort at a much lower age-related mortality risk than suggested by both their chronological age and by the time to death estimate reflected by their GrimAges. Although the GrimAge and PhenoAge estimates cannot be directly compared in terms of age estimation, and were not designed to reflect chronological age, the ideal would be that these clocks reveal a comparable and complementary estimation of biological decline.

The role of chronological age in the development of the PhenoAge and GrimAge clocks is critical in untangling their behavior in this investigation. For the PhenoAge algorithm, chronological age was incorporated as one of ten clinical markers associated with the risk of mortality, which were aggregated to represent phenotypic age. Chronological age is, therefore, used in a similar manner as used by the age predictor clocks, the difference being its aggregation with other clinical markers, rather than being the single outcome measure.

The GrimAge clock, however incorporated chronological age as an adjustment variable in the CpG selection models. Age was selected together with sex, smoking pack-years and seven protein markers to best predict time to death. First a penalized regression, adjusted for sex and chronological age, was applied to select CpGs associated with each of the GrimAge proteins and smoking pack-years. After all the selected CpGs had been combined to form the GrimAge marker, linear transformation based on forcing the GrimAge mean and variance to match chronological age was applied.

Based on these protocols, a weaker correlation between chronological age and PhenoAge than to HorvathAge, HannumAge and SBAge, was expected, because of the diluted contribution of chronological age in the prediction models. However, even compared to external data, the PhenoAge clock performed particularly poorly in the PURE-SA-NW cohort [17, 32]. For example, the correlations between PhenoAge and chronological age reported for four independent validation

cohorts, investigated by Levine et al. in the PhenoAge development manuscript [17], were much stronger than what was observed for the PURE-SA-NW participants (r = 0.51 in PURE-SA-NW versus 0.66, 0.69, 0.78 and 0. 89 in the respective validation cohorts [17]). For the GrimAge, however, the relationship we observed with chronological age (r = 0.81) was strikingly similar to those reported for the same validation cohorts mentioned above (reported here in the same order, r = 0.79, 0.80, 0.82 and 0.89 [18]). The reduced accuracy of the PhenoAge compared with the GrimAge-chronological age correlation in the PURE-SA-NW compared to other cohorts, potentially reflects environmental or ethnic confounding in the PURE-SA-NW group that is not captured by the design of the PhenoAge clock, but is captured by the GrimAge clock. In agreement with this hypothesis of ethnic variability, Levine et al., [17] reported notable differences in PhenoAge between Hispanics, non-Hispanic blacks and non-Hispanic whites (p = 5.1E–05). When investigating the standard deviations of the PhenoAA compared to GrimAA, Lu et al. [18] found an overall larger variance for the PhenoAA, as well as larger inter-ethnic differences, with the largest variation reported for African American cohorts, followed by white and then Hispanic groups. Our observation of larger variation in the PhenoAA than the GrimAA is in agreement with that of both Lu et al. [18] and a more recent comparison by Zhao et al. [32] in an independent African American study cohort. Zhao et al. reported PhenoAge estimates that were on average 13 years younger than chronological age (44 vs 57 years) compared to much more congruent behavior observed in the GrimAge estimate (54 years), analogous to our findings. Because of the limited literature, it remains difficult to confirm whether the comparatively weaker performance of the PhenoAge clock is the result of population-specific ethnic or environmental confounding, or a general limitation of the PhenoAge clock.

Population-specific confounding could result from either different associations between the clinical markers used in the design of the clock and mortality risk in the study population investigated or differing CpG-phenotypic age associations. Although we were unable to investigate the role of

ethnic confounding in the PURE-SA-NW cohort, the premise of genetic confounding related to PhenoAge was recently strengthened by evidence revealing that at least 50 of the clock CpGs comprising the PhenoAge algorithm capture ancestry-specific data [22]. Environmental confounding, however, was one hypothesis we were able to test. We hypothesized that the PhenoAge clock may not have accurately captured the age-acceleration associated with smoking behavior in the PURE-SA-NW cohort, resulting in severe biological age underestimation. Although the PhenoAge algorithm was not developed to encapsulate smoking habits specifically, previous research, including the PhenoAge validation analysis [17], reported that this age estimate was able to discriminate between smokers and non-smokers [17, 32]. In the PURE-SA-NW cohort, however, we found that neither PhenoAge nor the clinical components of phenotypic age encapsulated the biological effects of smoking in the PURE-SA-NW study population. The fact that our study sample only included men should not have affected this association [15, 32].

The integral role of smoking in the GrimAge clock, on the other hand, was confirmed by the associations observed between smoking status and multiple methylation components of the GrimAge clock in our cohort. Our findings reflected those from Zhao et al.[32] in which smoking status associated not only with DNAmPackY, but also with other such as the adrenomedullin, beta-2 microglobulin, growth differentiation factor 15 (DNAmB2M), Cystatin C, DNAmPAI-1 and tissue inhibitor metalloproteinase methylation components. Our data furthermore demonstrated that it is primarily smoking rather than combined smoking and alcohol use that is captured by GrimAge. The only methylation component that additional adjustment for alcohol consumption affected was DNAmPAI-1. This is in line with Zhao's findings of alcohol consumption having the strongest influence on DNAmPAI-1 [32].

Strengths and limitations

Because we tested the clocks in an older population in apparently good health, it allowed us to investigate the largely disease-independent underestimation of biological age in older individuals. Literature in this area is currently lacking. We were also able to contribute to the limited ethnic diversity represented in the epigenetic aging literature by investigating a group of continental Africans. We were, however, limited by our sample size and our study sample had a relatively narrow age range. We are therefore unable to extrapolate our findings to women or to younger individuals, or those with specific disease diagnoses. Furthermore, we were only able to investigate seven of the ten clinical phenotypic age components, which may have presented individual differences in smoking status groups. The fact that smoking-related differences were not observed in the aggregate PhenoAge marker does, however, suggest that it may be unlikely. Lastly, the lack of longitudinal and mortality data limits our ability to test the PhenoAge and GrimAge clocks' accuracy in predicting age-related mortality risk and time to death, respectively. In a recent study, Li et al. [32] investigated a range of biological aging measures as predictors of mortality in a Swedish longitudinal study. Methylation measures included were Horvath, Hannum, PhenoAge and GrimAge. Alongside Frailty Index, Horvath and GrimAge were independently predictive of mortality risk, confirming their value as measures of healthy aging and indicating that they may capture different components of health. Use of methylation-based measures combined with other measures of biological age (e.g. Frailty Index as used by Li et al. [32]) may, therefore, offer more optimized tools to identify individuals within populations who may most benefit from health interventions.

Conclusions

Epigenetic clocks provide unique possibilities to understand healthy vs accelerated aging better and consequently could ultimately improve the quality of life of a global population with an ever-increasing lifespan. In the PURE-SA-NW cohort, although all the tested clocks to underestimated the biological age of older individuals, the GrimAge clock was particularly useful as it was the only

clock that incorporated the damaging effects of smoking in biological aging. Our data illustrates that each clock provides a different fraction of information regarding the aging body. Although the differences in the physiological features these models capture strengthens the usefulness of specific clocks for particular investigations, it also makes their accuracy sensitive to heterogeneity in exposure between communities. For this reason, replication in various cohorts is important to reduce the weight of these influences.

Future perspectives

Research on the characterization of epigenetic age and ageing is advancing at a rapid pace. Future research should focus on: i) increased ethnic and environmental diversity in the data that underlie the algorithms designed for prediction in order to better predict morbidity and mortality in multiple population groups and; ii) the inclusion of methylation-based aging biomarkers (e.g. Brenner age and the Pace of Aging measure) that provide estimates according to scale-based rather than time-based risk. Because variation in lifestyle and ethnicity may influence the physiological features methylation clocks are built on, increased diversity will enable the developers of aging algorithms to improve their accuracy and ultimately reduce the weight of these influences on generally used clocks. Apart from optimizing the precision of methylation-based biomarkers, increased efforts should be made to integrate methylation-based and methylation-independent measures of biological age (e.g. Frailty Index) to broaden our understanding of the 'aging' phenotype and improve population-based risk stratification.

Summary

1. Accelerated biological *vs* chronological aging is associated with a variety of chronic diseases and mortality;

2. DNA methylation (DNAm) clocks are widely used to estimate biological age, although only limited data are available on non-European ethnicities;
3. This study characterizes three age-predictor (Horvath, Hannum and skin and blood (SB)) and two mortality-predictor (PhenoAge and GrimAge) DNAm clocks in Black South African men, aged 45 to 88 years;
4. Of the age-predictor clocks, SBAge had the strongest correlation with chronological age and the least variation in DNAmAge acceleration;
5. Our findings confirm the tendency of all the tested clocks to underestimate the biological age of older individuals,
6. The Horvath, Hannum and SB clocks best estimated chronological age at 55, 29 and 44 years, respectively;
7. The PhenoAge clock estimated our cohort at a much lower age-related mortality risk than suggested by both their chronological age and the time to death reflected by their GrimAges;
8. GrimAge provided superior characterization of age-related biological decline than PhenoAge;
9. The superior performance of the GrimAge *vs* PhenoAge in this population is likely because of its incorporation of smoking-related biological decline, as more than half of the study sample were current smokers.

References

1. Wang H, Naghavi M, Allen C *et al*. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The lancet* 388(10053), 1459-1544 (2016).
2. Beard HPJR, Bloom DE. Towards a comprehensive public health response to population ageing. *Lancet (London, England)* 385(9968), 658 (2015).
3. Jylhava J, Pedersen NL, Hagg S. Biological Age Predictors. *EBioMedicine* 21 29-36 (2017).
4. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics* 19(6), 371 (2018).

5. Martin GM. Epigenetic drift in aging identical twins. *Proceedings of the National Academy of Sciences of the United States of America* 102(30), 10413-10414 (2005).

6. Wang Y, Karlsson R, Lampa E *et al*. Epigenetic influences on aging: A longitudinal genome-wide methylation study in old Swedish twins. *Epigenetics* (just-accepted), (2018).

7. Teschendorff AE, Relton CL. Statistical and integrative system-level analysis of DNA methylation data. *Nature Reviews Genetics* 19(3), 129 (2018).

8. Horvath S. DNA methylation age of human tissues and cell types. *Genome biology* 14(10), 3156 (2013).

**This is the original Horvath clock conceptualization, development and application manuscript.

9. Marioni R, Suderman MJ, Chen BH *et al*. Tracking the epigenetic clock across the human life course. *Journals of Gerontology, Series A* doi:10.1093/gerona/gly060 1-5 (2018).

10. Horvath S, Oshima J, Martin GM *et al*. Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging* 10(7), 1758 (2018).

**This is the original Skin and Blood clock conceptualization, development and application manuscript.

11. Grant CD, Jafari N, Hou L *et al*. A longitudinal study of DNA methylation as a potential mediator of age-related diabetes risk. *Geroscience* 39(5-6), 475-489 (2017).

12. Perna L, Zhang Y, Mons U, Holleczek B, Saum K-U, Brenner H. Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. *Clinical epigenetics* 8(1), 64 (2016).

13. Lind L, Ingelsson E, Sundström J, Siegbahn A, Lampa E. Methylation-based estimated biological age and cardiovascular disease. *European journal of clinical investigation* 48(2), (2018).

14. Marioni RE, Shah S, Mcrae AF *et al*. DNA methylation age of blood predicts all-cause mortality in later life. *Genome biology* 16(1), 25 (2015).

15. Fransquet PD, Wrigglesworth J, Woods RL, Ernst ME, Ryan J. The epigenetic clock as a predictor of disease and mortality risk: a systematic review and meta-analysis. *Clinical epigenetics* 11(1), 62 (2019).

16. Hannum G, Guinney J, Zhao L *et al*. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular cell* 49(2), 359-367 (2013).

**This is the original Hannum clock conceptualization, development and application manuscript.

17. Levine ME, Lu AT, Quach A *et al*. An epigenetic biomarker of aging for lifespan and healthspan. *Aging* 10(4), 573 (2018).

**This is the original PhenoAge clock conceptualization, development and application manuscript.

18. Lu AT, Quach A, Wilson JG *et al*. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* 11(2), 303 (2019).

**This is the original GrimAge clock conceptualization, development and application manuscript.

19. El Khoury LY, Gorrie-Stone T, Smart M *et al*. Systematic underestimation of the epigenetic clock and age acceleration in older subjects. *Genome Biology* 20(1), 283 (2019).

*This manuscript highlighted, for the first time, the systematic underestimation of biological ages when DNAm clocks are applied to older cohorts.

20. Tajuddin SM, Hernandez DG, Chen BH *et al*. Novel age-associated DNA methylation changes and epigenetic age acceleration in middle-aged African Americans and whites. *Clinical Epigenetics* 11(1), 119 (2019).

21. Liu Z, Chen BH, Assimes TL, Ferrucci L, Horvath S, Levine ME. The role of epigenetic aging in education and racial/ethnic mortality disparities among older US Women. *Psychoneuroendocrinology* 104 18-24 (2019).

22. Philibert R, Beach SRH, Lei M-K *et al*. Array-Based Epigenetic Aging Indices May Be Racially Biased. *Genes* 11(6), 685 (2020).

*This manuscript emphasizes the importance of keeping the influence of genetic variation in the accuracy of methylation-derived aging algorithms in mind in epigenetic epidemiology.

23. Ryan J, Wrigglesworth J, Loong J, Fransquet PD, Woods RL. A Systematic Review and Meta-analysis of Environmental, Lifestyle, and Health Factors Associated With DNA Methylation Age. *The Journals of Gerontology: Series A* doi:10.1093/gerona/glz099 (2019).

24. Teo K, Chow CK, Vaz M, Rangarajan S, Yusuf S. The Prospective Urban Rural Epidemiology (PURE) study: examining the impact of societal influences on chronic noncommunicable diseases in low-, middle-, and high-income countries. *Am. Heart J.* 158(1), 1-7. e1 (2009).

25. Cronjé HT, Elliott HR, Nienaber-Rousseau C, Pieters M. Replication and expansion of epigenome-wide association literature in a black South African population. *Clinical Epigenetics* 12(1), 6 (2020).

26. Min JL, Hemani G, Davey Smith G, Relton C, Suderman M. Meffil: efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics* 34(23), 3983-3989 (2018).

27. R Core Team. R: A language and environment for statistical computing. (2017).

28. Salas LA, Koestler DC, Butler RA *et al*. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome biology* 19(1), 64-64 (2018).

29. Mcewen LM, Jones MJ, Lin DTS *et al*. Systematic evaluation of DNA methylation age estimation with common preprocessing methods and the Infinium MethylationEPIC BeadChip array. *Clinical Epigenetics* 10(1), 123 (2018).

30. Fiorito G, Mccrory C, Robinson O *et al*. Socioeconomic position, lifestyle habits and biomarkers of epigenetic aging: a multi-cohort analysis. *Aging* 11(7), 2045 (2019).

31. Philibert R, Dogan M, Beach SR, Mills JA, Long JD. AHRR methylation predicts smoking status and smoking intensity in both saliva and blood DNA. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 183(1), 51-60 (2020).

32. Zhao W, Ammous F, Ratliff S *et al*. Education and Lifestyle Factors Are Associated with DNA Methylation Clocks in Older African Americans. *International journal of environmental research and public health* 16(17), 3141 (2019).

33. Quach A, Levine ME, Tanaka T *et al*. Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging* 9(2), 419-446 (2017).

34. Azab B, Camacho-Rivera M, Taioli E. Average values and racial differences of neutrophil lymphocyte ratio among a nationally representative sample of United States subjects. *PloS one* 9(11), e112361 (2014).

35. Forget P, Khalifa C, Defour J-P, Latinne D, Van Pel M-C, De Kock M. What is the normal value of the neutrophil-to-lymphocyte ratio? *BMC Research Notes* 10(1), 12 (2017).

36. Kelsey KT, Wiencke JK. Immunomethylomics: A Novel Cancer Risk Prediction Tool. *Ann Am Thorac Soc* 15(Suppl 2), S76-s80 (2018).

37. Astle WJ, Elding H, Jiang T *et al*. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* 167(5), 1415-1429. e1419 (2016).

38. Li X, Ploner A, Wang Y *et al*. Longitudinal trajectories, correlations and mortality associations of nine biological ages across 20-years follow-up. *eLife* 9 e51507 (2020).

*This manuscript highlights the opportunity and usefulness of combining various biological age estimators (including those that derive from physical, telomere and methylation measurements) to better predict adverse outcomes and provide a more holistic understanding of aging.