



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### DNA methylation age of blood predicts all-cause mortality in later life

**Citation for published version:**

Marioni, RE, Shah, S, McRae, AF, Chen, BH, Colicino, E, Harris, SE, Gibson, J, Henders, AK, Redmond, P, Cox, SR, Pattie, A, Corley, J, Murphy, L, Martin, NG, Montgomery, GW, Feinberg, AP, Fallin, M, Multhaup, ML, Jaffe, AE, Joehanes, R, Schwartz, J, Just, AC, Lunetta, KL, Murabito, JM, Starr, JM, Horvath, S, Baccarelli, AA, Levy, D, Visscher, PM, Wray, NR & Deary, IJ 2015, 'DNA methylation age of blood predicts all-cause mortality in later life' *Genome Biology*, vol. 16, no. 1, 25. DOI: 10.1186/s13059-015-0584-6

**Digital Object Identifier (DOI):**

[10.1186/s13059-015-0584-6](https://doi.org/10.1186/s13059-015-0584-6)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Published In:**

Genome Biology

**Publisher Rights Statement:**

provisional pdf

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



This Provisional PDF corresponds to the article as it appeared upon acceptance. Fully formatted PDF and full text (HTML) versions will be made available soon.

## DNA methylation age of blood predicts all-cause mortality in later life

*Genome Biology*

doi:10.1186/s13059-015-0584-6

Riccardo E Marioni (riccardo.marioni@ed.ac.uk)  
Sonia Shah (s.shah1@uq.edu.au)  
Allan F McRae (a.mcrae@uq.edu.au)  
Brian H Chen (brian.chen@nih.gov)  
Elena Colicino (ecolicin@hsph.harvard.edu)  
Sarah E Harris (Sarah.Harris@igmm.ed.ac.uk)  
Jude Gibson (jgibson6@staffmail.ed.ac.uk)  
Anjali K Henders (Anjali.Henders@qimrberghofer.edu.au)  
Paul Redmond (paul.redmond@ed.ac.uk)  
Simon R Cox (simon.cox@ed.ac.uk)  
Alison Pattie (Alison.Pattie@ed.ac.uk)  
Janie Corley (janie.corley@ed.ac.uk)  
Lee Murphy (Lee.Murphy@ed.ac.uk)  
Nicholas G Martin (Nick.Martin@qimrberghofer.edu.au)  
Grant W Montgomery (Grant.Montgomery@qimrberghofer.edu.au)  
Andrew P Feinberg (afeinberg@jhu.edu)  
M Daniele Fallin (dfallin@jhu.edu)  
Michael L Multhaup (mmulthaup@gmail.com)  
Andrew E Jaffe (andrew.jaffe@libd.org)  
Roby Joehanes (robjoehanes@hsl.harvard.edu)  
Joel Schwartz (jschwartz@hsph.harvard.edu)  
Allan C Just (acjust@hsph.harvard.edu)  
Kathryn L Lunetta (klunetta@bu.edu)  
Joanne M Murabito (murabito@bu.edu)  
John M Starr (jstarr@staffmail.ed.ac.uk)  
Steve Horvath (shorvath@mednet.ucla.edu)  
Andrea A Baccarelli (abaccare@hsph.harvard.edu)  
Daniel Levy (levyd@nhlbi.nih.gov)  
Peter M Visscher (peter.visscher@uq.edu.au)  
Naomi R Wray (naomi.wray@uq.edu.au)  
Ian J Deary (i.deary@ed.ac.uk)

Published online: 30 January 2015

**ISSN** 1465-6906

**Article type** Research

**Submission date** 10 July 2014

**Acceptance date** 12 January 2015

**Article URL** <http://dx.doi.org/10.1186/s13059-015-0584-6>

Like all articles in BMC journals, this peer-reviewed article can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in BMC journals are listed in PubMed and archived at PubMed Central.

For information about publishing your research in BMC journals or any BioMed Central journal, go to <http://www.biomedcentral.com/info/authors/>

© 2015 Marioni *et al.*; licensee BioMed Central

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

# DNA methylation age of blood predicts all-cause mortality in later life

Riccardo E Marioni<sup>1,2,3,†</sup>  
Email: riccardo.marioni@ed.ac.uk

Sonia Shah<sup>3,4,†</sup>  
Email: s.shah1@uq.edu.au

Allan F McRae<sup>3,4,†</sup>  
Email: a.mcrae@uq.edu.au

Brian H Chen<sup>5,6,†</sup>  
Email: brian.chen@nih.gov

Elena Colicino<sup>7,†</sup>  
Email: ecolicin@hsph.harvard.edu

Sarah E Harris<sup>1,2</sup>  
Email: Sarah.Harris@igmm.ed.ac.uk

Jude Gibson<sup>8</sup>  
Email: jgibson6@staffmail.ed.ac.uk

Anjali K Henders<sup>9</sup>  
Email: Anjali.Henders@qimrberghofer.edu.au

Paul Redmond<sup>10</sup>  
Email: paul.redmond@ed.ac.uk

Simon R Cox<sup>1,10</sup>  
Email: simon.cox@ed.ac.uk

Alison Pattie<sup>10</sup>  
Email: Alison.Pattie@ed.ac.uk

Janie Corley<sup>10</sup>  
Email: janie.corley@ed.ac.uk

Lee Murphy<sup>8</sup>  
Email: Lee.Murphy@ed.ac.uk

Nicholas G Martin<sup>9</sup>  
Email: Nick.Martin@qimrberghofer.edu.au

Grant W Montgomery<sup>9</sup>  
Email: Grant.Montgomery@qimrberghofer.edu.au

Andrew P Feinberg<sup>11,12</sup>  
Email: afeinberg@jhu.edu

M Daniele Fallin<sup>11,13</sup>  
Email: dfallin@jhu.edu

Michael L Multhaup<sup>11</sup>  
Email: mmulthaup@gmail.com

Andrew E Jaffe<sup>13,14</sup>  
Email: andrew.jaffe@libd.org

Roby Joehanes<sup>5,15,16</sup>  
Email: robyjoehanes@hsl.harvard.edu

Joel Schwartz<sup>7,17</sup>  
Email: jschwartz@hsph.harvard.edu

Allan C Just<sup>7</sup>  
Email: acjust@hsph.harvard.edu

Kathryn L Lunetta<sup>5,18</sup>  
Email: klunetta@bu.edu

Joanne M Murabito<sup>5,19</sup>  
Email: murabito@bu.edu

John M Starr<sup>1,20</sup>  
Email: jstarr@staffmail.ed.ac.uk

Steve Horvath<sup>21,22,†</sup>  
Email: shorvath@mednet.ucla.edu

Andrea A Baccarelli<sup>7,17,†</sup>  
Email: abaccare@hsph.harvard.edu

Daniel Levy<sup>5,6,†</sup>  
Email: levyd@nhlbi.nih.gov

Peter M Visscher<sup>1,3,4,†</sup>  
Email: peter.visscher@uq.edu.au

Naomi R Wray<sup>4\*,†</sup>  
\* Corresponding author  
Email: naomi.wray@uq.edu.au

Ian J Deary<sup>1,10\*,†</sup>  
\* Corresponding author  
Email: i.deary@ed.ac.uk

- <sup>1</sup> Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK
- <sup>2</sup> Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK
- <sup>3</sup> Queensland Brain Institute, The University of Queensland, Brisbane 4072, QLD, Australia
- <sup>4</sup> University of Queensland Diamantina Institute, Translational Research Institute, The University of Queensland, Brisbane 4072, QLD, Australia
- <sup>5</sup> The NHLBI's Framingham Heart Study, Framingham, MA 01702, USA
- <sup>6</sup> Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, MD 01702, USA
- <sup>7</sup> Department of Environmental Health, Harvard School of Public Health, Boston, MA 02115, USA
- <sup>8</sup> Wellcome Trust Clinical Research Facility, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK
- <sup>9</sup> Queensland Institute of Medical Research Berghofer Medical Research Institute, Brisbane 4029, QLD, Australia
- <sup>10</sup> Department of Psychology, University of Edinburgh, Edinburgh, EH8 9JZ, UK
- <sup>11</sup> Center for Epigenetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
- <sup>12</sup> Departments of Medicine, Molecular Biology/Genetics, Oncology, and Biostatistics, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA
- <sup>13</sup> Department of Mental Health, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD 21205, USA
- <sup>14</sup> Lieber Institute for Brain Development, Baltimore, MD 21205, USA
- <sup>15</sup> Harvard Medical School, Boston, MA 02115, USA
- <sup>16</sup> Hebrew Senior Life, Boston, MA 02131, USA
- <sup>17</sup> Department of Epidemiology, Harvard School of Public Health, 665 Huntington Ave, Boston, MA 02115, USA
- <sup>18</sup> Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA

<sup>19</sup> Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA

<sup>20</sup> Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh EH8 9JZ, UK

<sup>21</sup> Human Genetics, Gonda Research Center, David Geffen School of Medicine, 695 Charles E. Young Drive South, Box 708822, Los Angeles, CA 90095-7088, USA

<sup>22</sup> Biostatistics, School of Public Health, University of California Los Angeles, Los Angeles, CA 90095, USA

† Equal contributors.

## Abstract

### Background

DNA methylation levels change with age. Recent studies have identified biomarkers of chronological age based on DNA methylation levels. It is not yet known whether DNA methylation age captures aspects of biological age.

### Results

Here we test whether differences between people's chronological ages and estimated ages, DNA methylation age, predict all-cause mortality in later life. The difference between DNA methylation age and chronological age, ( $\Delta_{\text{age}}$ ), was calculated in four longitudinal cohorts of older people. Meta-analysis of proportional hazards models from the four cohorts was used to determine the association between  $\Delta_{\text{age}}$  and mortality. A 5-year higher  $\Delta_{\text{age}}$  is associated with a 21% higher mortality risk, adjusting for age and sex. After further adjustments for childhood IQ, education, social class, hypertension, diabetes, cardiovascular disease, and *APOE* e4 status, there is a 16% increased mortality risk for those with a 5-year higher  $\Delta_{\text{age}}$ . A pedigree-based heritability analysis of  $\Delta_{\text{age}}$  was conducted in a separate cohort. The heritability of  $\Delta_{\text{age}}$  was 0.43.

### Conclusions

DNA methylation-derived measures of accelerated ageing are heritable traits that predict mortality independently of health status, lifestyle factors, and known genetic factors.

## Background

DNA sequence variants and epigenetic marks that are associated with changes in gene expression contribute to interindividual variation in complex phenotypes. Epigenetic mechanisms such as DNA methylation, characterised by the addition of a methyl group to a cytosine nucleotide primarily at cytosine-phosphate-guanine (CpG) sites, play essential roles during development, acting through the regulation of gene expression [1]. Unlike genomic

variants, such as single nucleotide polymorphisms (SNPs), levels of DNA methylation vary across the life course [2-6]. DNA methylation levels are influenced by lifestyle and environmental factors [7], as well as by genetic variation [8,9].

Age-related changes in DNA methylation are also well documented, and two recent studies used methylation measures from multiple CpG sites across the genome to predict chronological age in humans [10,11]. Hannum et al. created an age predictor based on a single cohort in which DNA methylation was measured in whole blood [10]. Horvath developed an age predictor using DNA methylation data from multiple studies (including the Hannum dataset) and multiple tissues [11]. In both studies, the difference between methylation-predicted age and chronological age (i.e.,  $\Delta_{\text{age}}$ ) was put forth as an index of disproportionate ‘biological’ ageing and was hypothesized to be associated with risk for age-related diseases and mortality [10,11]. Weidner et al. [12] proposed an age predictor based on three CpGs taken from a methylation array with fewer total CpG sites than the Hannum and Horvath models (27 k probes versus 450 k probes). To date, however, no study has tested whether DNA methylation-based  $\Delta_{\text{age}}$  or other genome-wide DNA methylation biomarkers are significant predictors of all-cause mortality.

Here, we tested the association of two DNA methylation measures of  $\Delta_{\text{age}}$  (using the Hannum and Horvath predictors) with all-cause mortality in four cohorts: the Lothian Birth Cohorts of 1921, and 1936 [13-15], the Framingham Heart Study [16,17], and the Normative Aging Study [18,19]. In addition, we estimated the heritability of  $\Delta_{\text{age}}$  using the Brisbane Systems Genetics Study (BSGS) [20].

## Results

The association between  $\Delta_{\text{age}}$  (DNA methylation-predicted age minus chronological age) and mortality was examined in four cohorts: Lothian Birth Cohort 1921, LBC1921, (N = 446,  $n_{\text{deaths}} = 292$ ), Lothian Birth Cohort 1936, LBC1936, (N = 920,  $n_{\text{deaths}} = 106$ ), the Framingham Heart Study, FHS, (N = 2,635,  $n_{\text{deaths}} = 238$ ), and the Normative Aging Study, NAS, (N = 657,  $n_{\text{deaths}} = 226$ ). The mean ages of the cohorts were 79.1 (SD 0.6), 69.5 (SD 0.8), 66.3 (SD 8.9), and 72.9 (SD 6.9) years, respectively. The Hannum predicted values were higher than the participants’ chronological ages by a mean of 2 to 6 years (SDs ~5 years) across the four cohorts. The Horvath predicted values were lower than the chronological ages in LBC1921 and LBC1936 participants by 4 to 5 years (SD ~6 years) but very similar to chronological age in the FHS (-0.60 years; SD 5.2) and the NAS (0.6 years; SD 5.8). A third predictor, based on the Weidner predictor was also examined, although it had a low correlation with chronological age (LBC1921: Pearson R = 0.02; LBC1936: Pearson R = -0.03; FHS: Pearson R = 0.25; NAS: Pearson R = 0.43) and very large absolute median differences (LBC1921: 29.9 years, LBC1936: 19.8 years, FHS: 12.6 years, NAS: 18.4 years) so was not examined further. A full description of the cohorts is provided in Table 1 and Additional file 1. Combining information from these studies, the correlation between chronological age and predicted age was 0.83 for the Hannum measure and 0.75 for the Horvath measure (Figure 1). The correlation between the Hannum and Horvath predictors was 0.77.

**Table 1 Summary details of the four analysis cohorts**

	LBC1921	LBC1936	FHS	NAS
N	446	920	2635	657
n deaths	292	106	238	226
Time to death (years)	7.2 (3.5)	4.4 (2.2)	6.0 (1.2)	10.5 (3.3)
Age (years)	79.1 (0.6)	69.5 (0.8)	66.3 (8.9)	72.9 (6.9)
Sex (male)	176 (40%)	465 (51%)	1200 (46%)	657 (100%)
Hannum methylation age (years)	85.0 (5.6)	75.8 (5.0)	68.2 (8.7)	77.6 (6.7)
Hannum $\Delta_{age}$ (years)	5.9 (5.6)	6.2 (5.1)	1.9 (4.8)	4.6 (4.8)
Hannum median error (years)	5.5	6.4	1.9	4.6
Horvath methylation age (years)	73.7 (6.2)	66.0 (5.8)	65.7 (8.3)	73.5 (7.4)
Horvath $\Delta_{age}$ (years)	-5.4 (6.3)	-3.6 (5.8)	-0.6 (5.2)	0.6 (5.8)
Horvath median error (years)	6.0	4.7	0.69	3.4
Tissue sample	whole blood	whole blood	whole blood	buffy coat
Methylation array	Illumina 450 k	Illumina 450 k	Illumina 450 k	Illumina 450 k

LBC: Lothian Birth Cohort, FHS: Framingham Heart Study, NAS: Normative Aging Study.

**Figure 1 Plot of predicted methylation age against chronological age and plot of Hannum versus Horvath predicted methylation age.** LBC: Lothian Birth Cohort, NAS: Normative Aging Study, FHS: Framingham Heart Study.  $r$  = Pearson correlation coefficient. \*To prevent the potential identification of individual participants, only FHS data points with chronological ages between 45 and 85, and NAS data points between ages 56 and 100 are displayed.

### Methylation age acceleration predicts mortality

In the meta-analysed results across the four cohorts, a five year higher Hannum  $\Delta_{age}$  was associated with a 21% (95% CI [1.14, 1.29],  $P < 0.0001$ ) greater mortality risk after adjustment for chronological age and sex (Figure 2). The corresponding increase in mortality risk for the Horvath  $\Delta_{age}$  was 11% (95% CI [1.05, 1.18],  $P = 0.0003$ ). Kaplan-Meier survival curves for the Horvath and Hannum  $\Delta_{age}$  (split into highest versus lowest quartile, for descriptive purposes only) in these models are presented in Figure 3 for the LBC1921 sample, which was the study with the greatest number of deaths. The plot illustrates the higher mortality rate for those with higher  $\Delta_{age}$ .

**Figure 2 Meta-analysis results of  $\Delta_{age}$  versus mortality.** LBC: Lothian Birth Cohort, NAS: Normative Aging Study, FHS: Framingham Heart Study. HR: Hazard Ratio, CI: confidence interval, W: fixed effect weight. The basic adjusted models controlled for chronological age, sex (NAS had only male participants), and laboratory batch (FHS only). The fully adjusted models controlled for chronological age, sex, smoking, education, childhood IQ (LBC1921 and LBC1936 only), social class (LBC1921 and LBC1936 only), *APOE* (LBC1921, LBC1936, and NAS only), cardiovascular disease, high blood pressure, and diabetes.

**Figure 3 Survival probability by quartiles of  $\Delta_{age}$  in LBC1921 adjusted for sex, and chronological age.** LBC: Lothian Birth Cohort.

### Association between methylation age indicators and white blood cell counts

It is well known that blood cell types have distinct methylation profiles [21,22]. A sensitivity analysis adjusting for measured white blood cell counts (basophils, eosinophils, monocytes,



lymphocytes, and neutrophils) resulted in mostly minor differences to the results (Additional file 2).

In addition to the five white blood cell types, we also examined the association between estimated naive T cell abundance and  $\Delta_{\text{age}}$ . However, to prevent spurious correlations between  $\Delta_{\text{age}}$ , which by definition correlates negatively with age, and cell counts, we used age acceleration defined as the residuals from a regression of predicted age on chronological age. There were stronger associations with these measures and the Hannum predictor (naive CD4+ T cells, average correlation  $r = -0.35$ , naive CD8+ T cells, average  $r = -0.34$ ) compared to the Horvath predictor ( $r = -0.20$ , and  $-0.20$ , respectively). After adjustment for naive T cells, both predictors were still significantly associated with mortality. Naive T cell count was also associated with mortality in addition to the Hannum predictor (Additional file 3). Chronological age had a significant negative relationship with the abundance of naive T cells (on average  $r = -0.12$  for CD8+ and  $r = -0.10$  for CD4+ T cells, Additional file 3).

The moderately strong correlation between naive T cell abundance and the Hannum predictor suggests that the latter keeps track of the age related decline of certain T cell populations. This makes sense in light of the facts that a) the Hannum predictor was constructed on the basis of DNA methylation data from whole blood and b) that naive T cells diminish with age due to age related thymic involution. In contrast to the Hannum predictor, the Horvath predictor exhibits a weaker relationship with naive T cell abundance, which probably reflects the fact that it was constructed on a range of different tissues and cell types.

## Adjusting for demographic variables and risk factors

Sensitivity analyses were performed to control for potentially confounding variables: smoking, education, childhood IQ (LBC1921 and LBC1936 only), social class (LBC1921 and LBC1936 only), *APOE* (LBC1921, LBC1936, and NAS only), cardiovascular disease, high blood pressure, and diabetes. When entered together in a fully adjusted model (Figure 2) the meta-analysed hazards ratio (HR) per 5 year increment was 1.16 (95% CI [1.08, 1.25],  $P = <0.0001$ ) for the fully adjusted Hannum  $\Delta_{\text{age}}$  and 1.09 (95% CI [1.02, 1.15],  $P = 0.0069$ ) for the fully adjusted Horvath  $\Delta_{\text{age}}$ . In the LBC datasets, which were the only datasets to contain information on all of the covariates, inclusion of the covariates one at a time made very little difference to basic, age- and sex-adjusted results (Additional file 4).

Separate age-adjusted Cox models for men and women are presented in Additional file 5; there was no notable difference in the relation of  $\Delta_{\text{age}}$  to survival by sex. Sensitivity analyses that excluded deaths within the first two years of follow-up made negligible differences to the effect size and significance of the  $\Delta_{\text{age}}$  associations (Additional file 6).

We tested the associations between  $\Delta_{\text{age}}$  and several key covariates (Additional file 7). With the exception of sex, where women had significantly lower  $\Delta_{\text{age}}$  estimates than men, there were no consistent associations with the covariates. There was some evidence for an association between the Hannum (but not Horvath)  $\Delta_{\text{age}}$  and childhood IQ and social class, although these covariates were assessed in only the two LBC cohorts.

## Heritability of methylation $\Delta_{\text{age}}$

Using data from the BSGS cohort, the estimated heritability for the Horvath and Hannum  $\Delta_{\text{age}}$  was 0.43 (SE 0.07,  $P = 9 \times 10^{-13}$ ) and 0.42 (SE 0.07,  $P = 4 \times 10^{-10}$ ), respectively, indicating that

approximately 40% of inter-individual differences in  $\Delta_{\text{age}}$  are due to genetic factors. The contribution to heritability broken down across relationship classes is given in Figure 4, and is consistent with an additive genetic model of inheritance.

---

**Figure 4 Heritability of methylation  $\Delta_{\text{age}}$ .** (A) Intra-class correlation of Hannum and Horvath  $\Delta_{\text{age}}$  across relationship class (B) heritability of Hannum and Horvath  $\Delta_{\text{age}}$  in BSGS data. Both plots show estimates with standard errors. \*Pseudo-independent pairs.  $r$  represents the degree of relatedness.

---

## Discussion

The difference between DNA methylation predicted age and chronological age ( $\Delta_{\text{age}}$ ) using two sets of epigenetic markers [10,11] is a heritable trait that is associated with an increased risk of mortality in four independent cohorts of older individuals. This association is independent of life-course predictors of ageing and death such as possession of the e4 allele of *APOE*, education, childhood IQ, social class, diabetes, high blood pressure, and cardiovascular disease. Moreover, there was no clear association between these factors and  $\Delta_{\text{age}}$ .

A strength of our study is that we evaluate two distinct epigenetic biomarkers of ageing,  $\Delta_{\text{age}}$ , one from DNA methylation in whole blood, and one based on results across multiple tissues. The associations between  $\Delta_{\text{age}}$  and mortality were stronger for the blood-based predictor but the two measures became comparable after adjusting for naive CD8 T cell abundances (Additional file 3). It is well known that immunosenescence is accompanied by diminishing naive T cells due to thymic involution and that the ageing immune system is a predictor of human longevity [23,24]. While it is possible that the association between accelerated epigenetic age and all cause mortality is mediated by changes in blood cell composition, we think that this is unlikely for the following reasons. First, the Horvath measure only exhibited a weak relationship (average  $r = -0.20$ ) with naive T cell abundance which reflects its definition and applicability to most tissues and cell types. Second, both measures remained significant predictors of mortality in multivariate regression models that adjusted for naive T cell abundances.

Given the similarity of the findings for the two epigenetic biomarkers of ageing, it is a credible hypothesis that  $\Delta_{\text{age}}$  is an epigenetic biomarker of the pace of biological aging throughout life.

Nearly all biological and lifestyle factors that we studied did not materially influence either the Hannum or Horvath  $\Delta_{\text{age}}$  across the four cohorts. Specifically, vascular insults, as measured by diabetes and hypertension, which are linked to cognitive decline, dementia, and death [25-27], were not associated with  $\Delta_{\text{age}}$ . There were sex differences, with men having higher  $\Delta_{\text{age}}$  than women, which is consistent with previous findings [10]. Childhood IQ and social class were modestly associated with  $\Delta_{\text{age}}$  in the two Lothian Birth Cohorts; an increased IQ in childhood and a less deprived social class were associated with lower  $\Delta_{\text{age}}$  in later life, although these were driven almost entirely by LBC1936. These correlated variables have been repeatedly associated with a range of health inequalities, including mortality [28], and it is possible that  $\Delta_{\text{age}}$  might offer insight into the mechanisms by which they are linked to health outcomes.

## Conclusions

There is continued interest in identifying new risk factors, environmental, genetic, and epigenetic that can improve our ability to predict disease and mortality. Epidemiological studies have identified numerous measures from across the human life-course that are associated with an increased risk of mortality. These include health factors such as cardiovascular disease, diabetes and hypertension [27], genetic factors such as presence of the *APOE* e4 allele [29], lifestyle variables such as smoking [30] and education [31], behavioural traits such as cognitive ability [31,32], the personality trait of conscientiousness [33], and candidate biomarkers of age such as telomere length [34,35]. Here, we report on an epigenetic biomarker that is predictive of human mortality, after accounting for known risk factors. We found that two heritable DNA methylation-based measures of the difference between epigenetic age and chronological age are significant predictors of mortality in our meta-analysis of four independent cohorts of older people. Individual genetic or environmental exposures that drive the associations are not yet known, but they appear not to be clearly linked to classic life-course risk factors. The difference between DNA methylation age and chronological age predicts mortality risk over and above a combination of smoking, education, childhood IQ, social class, *APOE* genotype, cardiovascular disease, high blood pressure, and diabetes. It may therefore be possible to think of DNA methylation predicted age as an 'epigenetic clock' [11] that measures biological age and runs alongside, but not always in parallel with chronological age, and may inform life expectancy predictions. Our results imply that epigenetic marks, such as gene methylation, are like other complex traits: influenced by both genetic and environmental factors and associated with major health-related outcomes.

## Materials and methods

### Cohort descriptions

#### *The Lothian Birth Cohort 1921*

Data were from the Lothian Birth Cohort 1921 (LBC1921), which is the basis of a longitudinal study of ageing [13,15]. Participants were born in 1921 and most completed a cognitive ability test at about age 11 years in the Scottish Mental Survey 1932 (SMS1932) [36]. The SMS1932 was administered nationwide to almost all 1921-born children who attended school in Scotland on June 1<sup>st</sup> 1932. The cognitive test was the Moray House Test No. 12, which provides a measure of general cognitive ability and has a scoring range between 0 and 76. The LBC1921 study attempted to follow up individuals who might have completed the SMS1932 and resided at about age 79 years in the Lothian region (Edinburgh and its surrounding areas) of Scotland; 550 people (n = 234, 43% men) were successfully traced and participated in the study from age 79 years. To date, there have been four additional follow-up waves at average ages of 83, 87, 90, and 92 years. The cohort has been deeply phenotyped during the later-life waves, including blood biomarkers, cognitive testing, and psycho-social, lifestyle, and health measures [13]. Genome wide single nucleotide polymorphisms and exome chip data are also available. DNA methylation measured in subjects at an average age of 79 (n = 514) was used for analyses in this report.

### ***Lothian Birth Cohort 1936***

The methylation mortality survival analysis was investigated in a second study, the Lothian Birth Cohort 1936 (LBC1936) [13,14]. All participants were born in 1936. Most had taken part in the Scottish Mental Survey 1947 at mean age 11 years as part of national testing of almost all children born in 1936 who attended Scottish schools on June 4th 1947 [37]. The cognitive test administered was the same Moray House Test No. 12 used in the SMS1932. A total of 1,091 participants (n = 548, 50% men) who were living in the Lothian area of Scotland were re-contacted in later life. Extensive phenotyping has also been carried out in this study, with data collection waves at three time points [13]. Genome wide single nucleotide polymorphisms and exome chip data are also available. DNA methylation was measured in 1,004 subjects at Wave 1 (mean age 70 years). To date, there have been two additional follow-up waves at average ages of 73, and 76.

### ***The Framingham Heart Study***

Framingham Heart Study (FHS) is a community based longitudinal study of participants living in and near Framingham, MA, at the start of the study in 1948 [16]. The Offspring cohort was comprised of the children and spouses of the original FHS participants, as described previously [17]. Briefly, enrollment for the Offspring cohort began in 1971 (n = 5,124), and in-person evaluations occurred approximately every 4 to 8 years thereafter. The current analysis was limited to participants from the Offspring cohort who survived until the 8<sup>th</sup> examination cycle (2005–2008) and consented to genetics research. DNA methylation data of peripheral blood samples collected at the 8<sup>th</sup> examination cycle were available in 2,741 participants.

### ***The Normative Aging Study***

The US Department of Veterans Affairs (VA) Normative Aging Study (NAS) is an ongoing longitudinal cohort established in 1963, which included men who were 21–80 years of age and free of known chronic medical conditions at entry [18,19]. Participants were subsequently invited to medical examinations every three to five years. At each visit, participants provided information on medical history, lifestyle, and demographic factors, and underwent a physical examination and laboratory tests. DNA samples were collected from between 1999–2007 from the 675 active participants and used for DNA methylation analysis. We excluded 18 participants who were not of European descent or had missing information on race, leaving a total of 657 individuals.

### ***Brisbane Systems Genetics Study***

The Brisbane Systems Genetic Study (BSGS) [20] is a cohort comprising adolescent monozygotic (MZ) and dizygotic (DZ) twins, their siblings and their parents. They were originally recruited into an ongoing study of the genetic and environmental factors influencing cognition and pigmented nevi. DNA methylation was measured on 614 individuals from 117 families of European descent. Families consist of adolescent monozygotic (MZ; n = 67 pairs) and dizygotic (DZ; n = 111 pairs) twins, their siblings (n = 119), and their parents (n = 139). Children have a mean age of 14 years (range 9–23) and parents 47 years (range 33–75).

## **Ethics**

### ***LBC consent***

Following informed consent, venesected whole blood was collected for DNA extraction in both LBC1921 and LBC1936. Ethics permission for the LBC1921 was obtained from the Lothian Research Ethics Committee (Wave 1: LREC/1998/4/183). Ethics permission for the LBC1936 was obtained from the Multi-Centre Research Ethics Committee for Scotland (Wave 1: MREC/01/0/56), the Lothian Research Ethics Committee (Wave 1: LREC/2003/2/29). Written informed consent was obtained from all subjects.

### ***FHS consent***

All participants provided written informed consent at the time of each examination visit. The study protocol was approved by the Institutional Review Board at Boston University Medical Center (Boston, MA).

### ***NAS consent***

The NAS study was approved by the Institutional Review Boards (IRBs) of the participating institutions. Participants have provided written informed consent at each visit.

### ***BSGS consent***

The BSGS study was approved by the Queensland Institute for Medical Research Human Research Ethics Committee. All participants gave informed written consent.

## **DNA Methylation measurement**

In all cohorts, bisulphite converted DNA samples were hybridised to the 12 sample Illumina HumanMethylation450BeadChips [38] using the Infinium HD Methylation protocol and Tecan robotics (Illumina, San Diego, CA, USA).

### ***LBC1921 and LBC1936 DNA methylation***

DNA was extracted from 514 whole blood samples in LBC1921 and from 1,004 samples in LBC1936. Samples were extracted at MRC Technology, Western General Hospital, Edinburgh (LBC1921) and the Wellcome Trust Clinical Research Facility (WTCRF), Western General Hospital, Edinburgh (LBC1936), using standard methods. Methylation typing of 485,512 probes was performed at the WTCRF. Raw intensity data were background-corrected and methylation beta-values generated using the R minfi package [39]. Quality control analysis was performed to remove probes with a low (<95%) detection rate at  $p < 0.01$ . Manual inspection of the array control probe signals was used to identify and remove low quality samples (e.g. samples with inadequate hybridisation, bisulfite conversion, nucleotide extension or staining signal). The Illumina-recommended threshold was used to eliminate samples with a low call rate (samples with <450,000 probes detected at  $p < 0.01$ ). Since the LBC samples had previously been genotyped using the Illumina 610-Quadv1 genotyping platform, genotypes derived from the 65 SNP control probes on the methylation array using the watermelon package [40] were compared to those obtained from the

genotyping array to ensure sample integrity. Samples with a low match of genotypes with SNP control probes, which could indicate sample contamination or mix-up, were excluded (n = 9). Moreover, 8 subjects whose predicted sex, based on XY probes, did not match reported sex were also excluded.

### ***FHS DNA Methylation***

Peripheral blood samples were collected at the 8<sup>th</sup> examination samples (2005–2008). Genomic DNA was extracted from buffy coat using the Gentra Puregene DNA extraction kit (Qiagen) and bisulfite converted using EZ DNA Methylation kit (Zymo Research Corporation). DNA methylation quantification was conducted in two laboratory batches. Methylation beta values were generated using the Bioconductor minfi package with background correction. Sample exclusion criteria included poor SNP matching of control positions, missing rate >1%, outliers from multi-dimensional scaling (MDS), and sex mismatch. Probes were excluded if missing rate >20%. In total, 2,635 samples and 443,304 CpG probes remained for analysis.

### ***NAS DNA Methylation***

DNA was extracted from buffy coat using the QIAamp DNA Blood Kit (QIAGEN, Valencia, CA). 500 ng of DNA was used to perform bisulfite conversion using the EZ-96 DNA Methylation Kit (Zymo Research, Orange, CA). To limit chip and plate effects, a two-stage age-stratified algorithm was used to randomise samples and ensure similar age distributions across chips and plates; we randomised 12 samples – which were sampled across all the age quartiles – to each chip, then chips were randomised to plates (each housing eight chips). Quality control analysis was performed to remove samples where >1% of probes had a detection p-value > 0.05. The remaining samples were preprocessed using the Illumina-type background correction without normalization as reimplemented in the Bioconductor minfi package, which was used to generate methylation beta values [39]. All 485,512 CpG and CpH probes were in the working set.

### ***BSGS DNA Methylation***

DNA was extracted from peripheral blood lymphocytes by the salt precipitation method [41] from samples that were time matched to sample collection of PAXgene tubes for gene expression studies in the Brisbane Systems Genetics Study [20]. Bisulphite converted DNA samples were hybridised to the 12 sample Illumina HumanMethylation450 BeadChips using the Infinium HD Methylation protocol and Tecan robotics (Illumina, San Diego, CA, USA). Samples were randomly placed with respect to the chip they were measured on and to the position on that chip in order to avoid any confounding with family. Box-plots of the red and green intensity levels and their ratio were used to ensure that no chip position was under- or over-exposed, with any outlying samples repeated. Similarly, the proportion of probes with detection p-value less than 0.01 was examined to confirm strong binding of the sample to the array. Raw intensity values were background corrected using the Genome Studio software performing normalization to internal controls and background subtraction.

## **Mortality ascertainment**

### ***LBC mortality ascertainment***

For both LBC1921 and LBC1936, mortality status was obtained via data linkage from the National Health Service Central Register, provided by the General Register Office for Scotland (now National Records of Scotland). Participant deaths and cause of death are routinely flagged to the research team on approximately a 12-weekly basis.

### ***FHS mortality ascertainment***

Deaths that occurred prior to January 1, 2013 were ascertained using multiple strategies, including routine contact with participants for health history updates, surveillance at the local hospital and in obituaries of the local newspaper, and queries to the National Death Index. We requested death certificates, hospital and nursing home records prior to death, and autopsy reports. When cause of death was undeterminable, the next of kin were interviewed. The date and cause of death were reviewed by an endpoint panel of 3 investigators.

### ***NAS mortality ascertainment***

Regular mailings to study participants have been used to maintain vital-status information, and official death certificates were obtained for decedents from the appropriate state health department. Death certificates were reviewed by a physician, and cause of death coded by an experienced research nurse using ICD-9. Both participant deaths and cause of death are routinely updated by the research team and last update available was December 31<sup>st</sup>, 2013.

## **Covariate measurement**

### ***LBC covariates***

Mortality-associated variables assessed in LBC1921 and LBC1936 were used as covariates in the statistical models: educational attainment, age-11 cognitive ability, *APOE* e4 status (carriers versus non-carriers), smoking status, and the presence or absence of diabetes, high blood pressure, or cardiovascular disease. Age-11 cognitive ability (age-11 IQ) was measured in 1932 for LBC1921 and in 1947 for LBC1936 using the Moray House Test Number 12, described above. All other variables were measured at the late-life baseline waves (age 79 for LBC1921 and age 70 for LBC1936). *APOE* was genotyped from venous blood using PCR amplification of a 227-bp fragment of the *APOE* gene, which contains the two single nucleotide polymorphisms that are used to define the e2, e3, and e4 alleles [42] in LBC1921, and by TaqMan technology (Applied Biosystems, Carlsbad, CA, USA) in LBC1936. Subjects were then categorised by the presence or absence of the e4 allele. Social class was based on the most prestigious occupation held by the participant prior to retirement. It was grouped into five categories in LBC1921 and six categories in LBC1936, where Class III was split into manual and non-manual professions [43,44]. It was treated as a continuous variable with lower values representing the more prestigious classes. The other variables were determined via self-report: number of years of education (measured as a continuous variable), diabetes (yes/no), high blood pressure (yes/no), cardiovascular disease (yes/no), and categorical smoking status (current/ex-smoker, never smoked).

Given the known influence of blood cell count on methylation [21], we adjusted for five types of white blood cell count (basophils, monocytes, lymphocytes, eosinophils, and neutrophils) that were measured at on the same blood that was analysed for methylation. These data were collected and processed the same day; technical details are reported in McIlhagger et al. [45].

### ***FHS covariates***

At the 8<sup>th</sup> in-person examination visit participants completed a questionnaire that inquired about their education, occupation, smoking status, and disease status. Highest levels of educational attainment was assessed by 8 categories—no schooling, grades 1–8, grades 9–11, completed high school or GED, some college but no degree, technical school certificate, associate degree, Bachelor’s degree, graduate or professional degree. Smoking status was dichotomized as current/past smokers and those who reported to never have smoked. Diabetes was defined as having fasting blood glucose  $\geq 126$  mg/dl or current treatment for diabetes. Hypertension was defined as having systolic blood pressure  $\geq 140$  mmHg, diastolic blood pressure  $\geq 90$  mmHg, or current treatment for hypertension. Cardiovascular disease was determined by a panel of 3 physicians, who reviewed participants' medical records, laboratory findings, and clinic exam notes.

### ***NAS covariates***

At each in-person examination visit, participants completed a questionnaire that enquired about their smoking status, education, diabetes (self-reported diagnosis and/or use of diabetes medications), and diagnosis of coronary heart disease (validated on medical records, ECG, and physician exams). High blood pressure was defined as antihypertensive medication use or SBP  $\geq 140$  mmHg or DBP  $\geq 90$  mmHg at study visit. *APOE*-e4 allele status was assessed through genotyping on a Sequenom MassArray MALDI-TOF mass spectrometer.

### ***Estimated naive T cell abundance***

In LBC1921, LBC1936, FHS, and NAS, we considered the abundance of defined different subtypes of T cells: Naïve T cells were defined as RA+ IL-7 Receptor + cells. Central Memory T cells = RA negative IL7 Receptor positive Effector memory = RA negative IL7Receptor negative. To estimate the naïve T cells in our cohort studies, we used a prediction method that was developed on an independent data set. The predictor of T cell counts (e.g., naïve CD4 T cell count) was found by applying a penalized regression model (elastic net) to regress T cell counts (dependent variable) on a subset of CpGs reported in Supplemental Table 3 from Zhang et al. [46]. By applying this resulting penalized regression model to our data, we arrived at predicted T cell counts.

### ***Data availability***

LBC methylation data have been submitted to the European Genome-phenome Archive under accession number EGAS00001000910; phenotypic data are available at dbGaP under the accession number phs000821.v1.p1. The FHS and NAS data are available at dbGaP under the accession numbers phs000724.v2.p9 phs000853.v1.p1, respectively. BSGS methylation data are available from the NCBI Gene Expression Omnibus under accession number GSE56105.



## Statistical analyses

Two measures of DNA methylation age ( $m_{\text{age}}$ ) were calculated. The Horvath [11]  $m_{\text{age}}$  uses 353 probes common to the Illumina 27 K and 450 K Methylation arrays using data from a range of tissues and cell types. The Hannum [10]  $m_{\text{age}}$  is based on 71 methylation probes from the Illumina 450 K Methylation array derived as the best predictors of age using data generated from whole blood. Of the Hannum age predictor probes, 70, 71, and 71 were included in the LBC, NAS, and FHS data, respectively.  $m_{\text{age}}$  was calculated as the sum of the beta values multiplied by the reported effect sizes for the Hannum predictor. For the Horvath predictor,  $m_{\text{age}}$  was determined in all cohorts using the online calculator (<http://labs.genetics.ucla.edu/horvath/dnamage/>). A third predictor, based on the three probes highlighted in the Weidner et al. paper [12], was also examined although, due to its poorer predictive accuracy, it was not included for the main analyses. To account for technical variability in the measurement of the methylation CpGs in the LBC studies,  $m_{\text{age}}$  was adjusted for plate, array, position on the chip, and hybridisation date (all treated as fixed effect factors) using linear regression. In a sensitivity analysis, additional adjustments were made for white blood cell counts (the number of basophils, monocytes, lymphocytes, eosinophils, and neutrophils per volume of blood) or DNA methylation-estimated cell counts, as described elsewhere [21]. The residuals from these models were added to the mean predicted methylation age to give the new, adjusted measure of  $m_{\text{age}}$ . The two methylation age predictors contained 6 overlapping probes. A methylation-based age acceleration index ( $\Delta_{\text{age}}$ ) was calculated for all subjects, defined as the adjusted methylation age in years minus chronological age at sample collection in years ( $\Delta_{\text{age}} = m_{\text{age}} - \text{chronological age}$ ).

Cox proportional hazards regression models were used to test the association between the Horvath and Hannum measures of  $\Delta_{\text{age}}$  and mortality, adjusting for age at sample collection, and sex. Cox models in FHS further adjusted for laboratory batch (fixed effect) and used a robust variance estimator to account for familial relatedness. Hazard ratios for  $\Delta_{\text{age}}$  were expressed per 5 years of methylation age acceleration. Schoenfeld residuals were examined to test the proportional hazards assumption. Sensitivity analyses, also using Cox proportional hazards regression, excluded deaths within the first 2 years of follow-up to eliminate the potential influences of (fatal) acute illness on the methylation measurements. Analyses to account for possible confounders/mediators included potential life-course predictors of mortality: age-11 IQ (LBC only), education in years, social class (LBC only), *APOE* e4 carrier status (LBC and NAS), smoking status, and self-reported diabetes, high blood pressure, and cardiovascular disease. A fully adjusted model was tested, in which all variables were entered together. Chronological age- and sex-adjusted linear regression models were used to explore the relationship between  $\Delta_{\text{age}}$  and the additional covariates; e.g., does methylation age acceleration depend on smoking or diabetes?

The results from the individual cohorts were meta-analysed using the 'meta' package in R [47]. The cohorts were weighted based on the standard errors of the log hazard ratios. There was no evidence of cohort heterogeneity in the primary Cox model analyses according to the DerSimonian-Laird estimator of between-study variance so fixed effects models were considered.

All analyses were performed in the statistical software R [48] with the Cox models utilising the survival library [49].

Finally, we calculated the heritability of  $\Delta_{\text{age}}$  in the BSGS cohort. As  $m_{\text{age}}$  was a better predictor of chronological age in the adult compared to adolescent samples, the difference between methylation age and chronological age was firstly standardised within generations (parents and offspring). Regression models were fitted to methylation age removing the effects of age and sex. Additionally, the regression on the adolescent samples included  $\text{age}^2$  to account for the non-linearity between chronological and methylation age [11]. The residuals from these regressions were standardised to have a variance of 1 before combining the generations. See Additional file 8 for a graphical representation of the correction performed.

For each probe, the Intra Class Correlation of  $\Delta_{\text{age}}$  for the various relative pairs was calculated using ANOVA as follows:

$$ICC = \frac{MS_B - MS_W}{MS_B + MS_W}$$

where  $MS_B$  is the Mean Square Between pairs and  $MS_W$  is the Mean Square Within. The confidence intervals were based on the number of pseudo-independent relative pair for each relationship.

The heritability for each probe was estimated by partitioning its variance into additive genetic ( $V_a$ ) and environmental ( $V_e$ ) component by fitting a linear mixed model of the form

$$\mathbf{y} = \mu + \mathbf{Za} + \mathbf{e}$$

where  $\mathbf{y}$  is the vector of adjusted methylation age,  $\mathbf{a}$  is the additive genetic effects and  $\mathbf{e}$  is the unique environmental effects (residuals). The model was fitted using QTDT [50].

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

Contributed equally: REM, SS, AFM, BHC, EC & SH, AAB, DL, PMV, NRW, IJD. Supervised Research: GM, KLL, JS, SH, AAB, DL, PMV, NRW, IJD, JMM. Conceived and designed the experiments: GM, JS, SH, AAB, DL, PMV, NRW, IJD. Performed the experiments: SEH, JG, LM, GM, APF, MLM, MDF, AEJ. Performed the statistical analyses: REM, SS, AFM, BHC, EC, MLM, AEJ, SH. Analysed the data: REM, SS, AFM, BHC, EC. Contributed reagents/materials/analysis tools: SEH, JG, AKH, PR, SRC, AP, LM, NGM, ACJ, KLL, JMM, JMS, APF, MLM, RJ, KLL, MDF, AEJ. Wrote the paper: REM, SS, AFM, BHC, EC, JMM, SH, AAB, DL, PMV, NRW, IJD. All authors read and approved the final manuscript.

## Acknowledgements

We thank the cohort participants and team members who contributed to these studies. This work was supported by numerous funding bodies. Phenotype collection in the Lothian Birth Cohort 1921 was supported by the UK's Biotechnology and Biological Sciences Research Council (BBSRC), The Royal Society and The Chief Scientist Office of the Scottish Government. Phenotype collection in the Lothian Birth Cohort 1936 was supported by Age UK (The Disconnected Mind project). Methylation typing was supported by the Centre for Cognitive Ageing and Cognitive Epidemiology (Pilot Fund award), Age UK, The Wellcome Trust Institutional Strategic Support Fund, The University of Edinburgh, and The University of Queensland. REM, SEH, SRC, JMS, PMV, and IJD are members of the University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology (CCACE). CCACE is supported by funding from the BBSRC, the Medical Research Council (MRC), and the University of Edinburgh as part of the cross-council Lifelong Health and Wellbeing initiative (MR/K026992/1). Research reported in this publication was supported by National Health and Medical Research Council (NHMRC) project grants 613608, APP496667, APP1010374 and APP1046880. NHMRC Fellowships to GWM, PMV, and NRW (613602) and Australia Research Council (ARC) Future Fellowship to NRW (FT0991360). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NHMRC or ARC. The Framingham Heart Study is funded by National Institutes of Health contract N01-HC-25195. The laboratory work for this investigation was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health. The analytical component of this project was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, and the Center for Information Technology, National Institutes of Health, Bethesda, MD. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster (<http://biowulf.nih.gov>) and Helix Systems (<http://helix.nih.gov>) at the National Institutes of Health, Bethesda, Md. JMM and KLL were supported by R01AG029451. The present work on the US Department of Veterans Affairs (VA) Normative Aging Study has been supported by funding from the U.S. National Institute of Environmental Health Sciences (NIEHS) (R01ES015172, R01ES021733). The VA Normative Aging Study is supported by the Cooperative Studies Program/ERIC, US Department of Veterans Affairs, and is a research component of the Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC). Additional support to the VA Normative Aging Study was provided by the US Department of Agriculture, Agricultural Research Service (contract 53-K06-510). The views expressed in this paper are those of the authors and do not necessarily represent the views of the US Department of Veterans Affairs. We thank Stuart J Ritchie for his helpful comments and suggestions on the initial draft of the manuscript.

## References

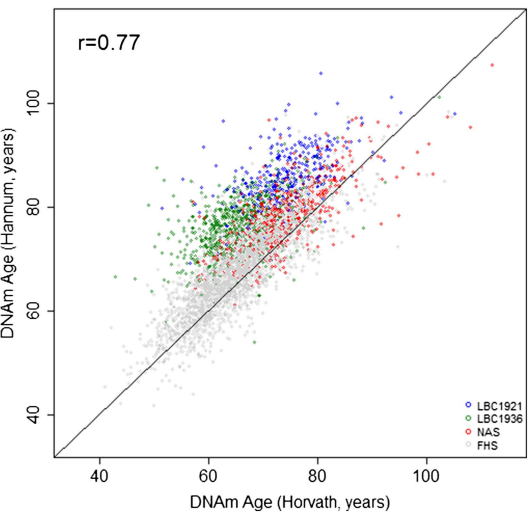
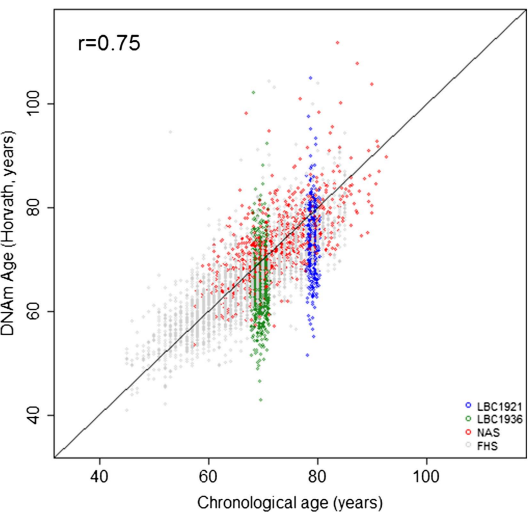
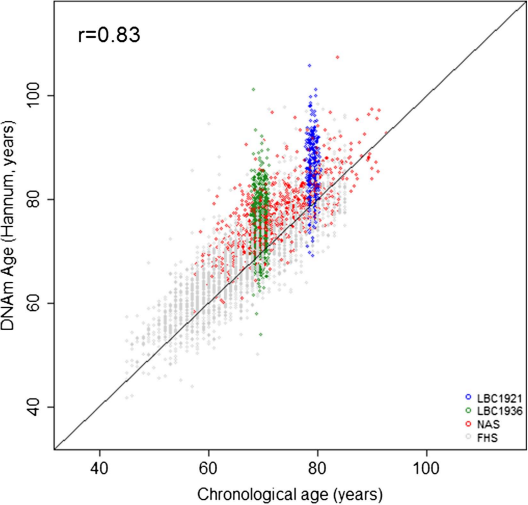
1. Thornburg KL, Shannon J, Thuillier P, Turker MS. In utero life and epigenetic predisposition for disease. *Adv Genet.* 2010;71:57–78.
2. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet.* 2012;8:e1002629.

3. Bollati V, Schwartz J, Wright R, Litonjua A, Tarantini L, Suh H, et al. Decline in genomic DNA methylation through aging in a cohort of elderly subjects. *Mech Ageing Dev.* 2009;130:234–9.
4. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, et al. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet.* 2009;5:e1000602.
5. Fraga MF, Esteller M. Epigenetics and aging: the targets and the marks. *Trends Genet.* 2007;23:413–8.
6. Rodriguez-Rodero S, Fernandez-Morera JL, Fernandez AF, Menendez-Torre E, Fraga MF. Epigenetic regulation of aging. *Discov Med.* 2010;10:225–33.
7. Mathers JC, Strathdee G, Relton CL. Induction of epigenetic alterations by dietary and other environmental factors. *Adv Genetics.* 2010;71:3–39.
8. Bell JT, Spector TD. DNA methylation studies using twins: what are they telling us? *Genome Biol.* 2012;13:172.
9. Kilpinen H, Dermitzakis ET. Genetic and epigenetic contribution to complex traits. *Hum Mol Genet.* 2012;21:R24–8.
10. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell.* 2013;49:359–67.
11. Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013;14:R115.
12. Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P, et al. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol.* 2014;15:R24.
13. Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol.* 2012;41:1576–84.
14. Deary IJ, Gow AJ, Taylor MD, Corley J, Brett C, Wilson V, et al. The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatr.* 2007;7:28.
15. Deary IJ, Whiteman MC, Starr JM, Whalley LJ, Fox HC. The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. *J Pers Soc Psychol.* 2004;86:130–47.
16. Dawber TR, Meadors GF, Moore Jr FE. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health.* 1951;41:279–81.
17. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol.* 1979;110:281–90.

18. Bell B, Rose CL, Damon A. The Veterans Administration longitudinal study of healthy aging. *The Gerontologist*. 1966;6:179–84.
19. Spiro 3rd A, Vokonas P. The VA Normative Aging Study. In: Markides KS, editor. *Encyclopedia of health and aging*. Thousand Oaks, CA: Sage; 2007. p. 421–2.
20. Powell JE, Henders AK, McRae AF, Caracella A, Smith S, Wright MJ, et al. The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. *PLoS One*. 2012;7:e35430.
21. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86.
22. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol*. 2014;15:R31.
23. Larbi A, Franceschi C, Mazzatti D, Solana R, Wikby A, Pawelec G. Aging of the immune system as a prognostic factor for human longevity. *Physiology (Bethesda)*. 2008;23:64–74.
24. Fagnoni FF, Vescovini R, Passeri G, Bologna G, Pedrazzoni M, Lavagetto G, et al. Shortage of circulating naive CD8(+) T cells provides new insights on immunodeficiency in aging. *Blood*. 2000;95:2860–8.
25. Biessels GJ, Strachan MW, Visseren FLJ, Kappelle LJ, Whitmer RA. Dementia and cognitive decline in type 2 diabetes and prediabetic stages: towards targeted interventions. *The Lancet Diabetes and Endocrinology*. 2014;2:246–55.
26. Qiu C, Winblad B, Fratiglioni L. The age-dependent relation of blood pressure to cognitive function and dementia. *Lancet Neurol*. 2005;4:487–99.
27. World Health Organization. *Global Health Risks: Mortality and burden of disease attributable to selected major risks*. 2009.
28. Deary IJ, Weiss A, Batty GD. Intelligence and personality as predictors of illness and death: How researchers in differential psychology and chronic disease epidemiology are collaborating to understand and address health inequalities. *Psychol Sci Publ Interest*. 2010;11:53–79.
29. Schächter F, Faure-Delanef L, Guénot F, Rouger H, Froguel P, Lesueur-Ginot L, et al. Genetic associations with human longevity at the APOE and ACE loci. *Nat Genet*. 1994;6:29–32.
30. Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male British doctors. *BMJ*. 2004;328:1519.
31. Batty GD, Kivimaki M, Deary IJ. Intelligence, education, and mortality. *BMJ*. 2010;340:c563.

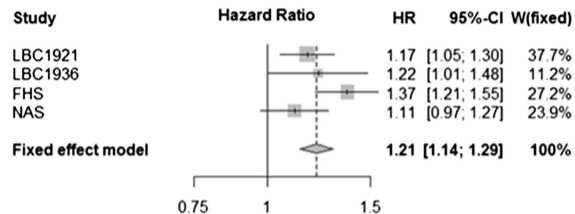
32. Calvin CM, Deary IJ, Fenton C, Roberts BA, Der G, Leckenby N, et al. Intelligence in youth and all-cause-mortality: systematic review with meta-analysis. *Int J Epidemiol.* 2011;40:626–44.
33. Kern ML, Friedman HS. Do conscientious individuals live longer? A quantitative review. *Health Psychol.* 2008;27:505–12.
34. Bendix L, Thinggaard M, Fenger M, Kolvraa S, Avlund K, Linneberg A, et al. Longitudinal Changes in Leukocyte Telomere Length and Mortality in Humans. *J Gerontol A Biol Sci Med Sci.* 2014;69:231–9.
35. Cawthon RM, Smith KR, O'Brien E, Sivatchenko A, Kerber RA. Association between telomere length in blood and mortality in people aged 60 years or older. *Lancet.* 2003;361:393–5.
36. Scottish Council for Research in Education. *The Intelligence of Scottish Children: a National Survey of an Age-Group.* London: University of London Press; 1933.
37. Scottish Council for Research in Education. *The Trend of Scottish Intelligence.* London: University of London Press; 1949.
38. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. *Genomics.* 2011;98:288–95.
39. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA Methylation microarrays. *Bioinformatics* 2014; <http://dx.doi.org/10.1093/bioinformatics/btu049>, In press.
40. Schalkwyk LC, Pidsley R, Wong CCY. waterMelon: Illumina 450 methylation array normalization and metrics. R package version 1.2.2. 2013. Accessed Sept 2013.
41. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 1988;16:1215.
42. Wenham PR, Price WH, Blandell G. Apolipoprotein E genotyping by one-stage PCR. *Lancet.* 1991;337:1158–9.
43. Office of Population Censuses and Surveys. *Classification of Occupations 1980.* London: HMSO; 1980.
44. General Register Office. *Census 1951: Classification of occupations.* London, England: Her Majesty's Stationery Office; 1956.
45. McIlhagger R, Gow AJ, Brett CE, Corley J, Taylor M, Deary IJ, et al. Differences in the haematological profile of healthy 70 year old men and women: normal ranges with confirmatory factor analysis. *BMC Blood Disord.* 2010;10:4.

46. Zhang Y, Maksimovic J, Naselli G, Qian J, Chopin M, Blewitt ME, et al. Genome-wide DNA methylation analysis identifies hypomethylated genes regulated by FOXP3 in human regulatory T cells. *Blood*. 2013;122:2823–36.
47. Schwarzer G. meta: Meta-Analysis with R. R package version 2.16. 2014. Accessed Jan 2014.
48. R Core Team. R: A Language and Environment for Statistical Computing. 2013. Accessed Sep 2013.
49. Therneau TM. A Package for Survival Analysis in S. R package version 2.37-7. 2014. Accessed Jan 2014.
50. Abecasis GR, Cardon LR, Cookson WO. A general test of association for quantitative traits in nuclear families. *Am J Hum Genet*. 2000;66:279–92.

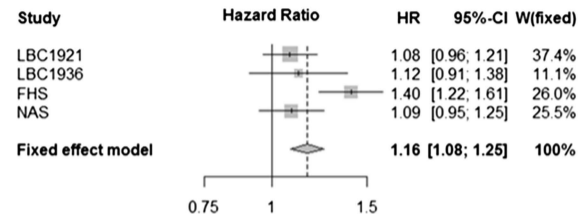




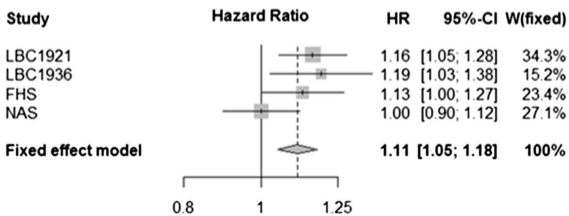
## Basic Adjusted Hannum $\Delta_{age}$ and Mortality



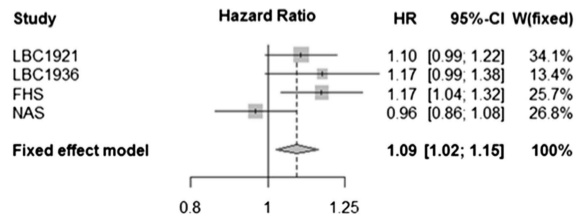
## Fully Adjusted Hannum $\Delta_{age}$ and Mortality

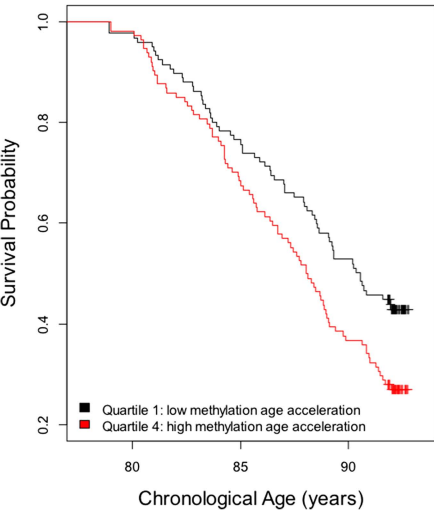
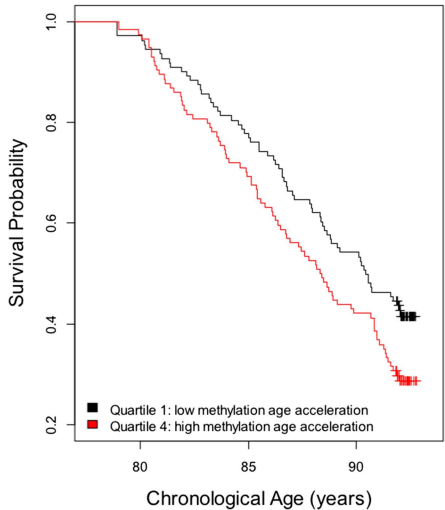


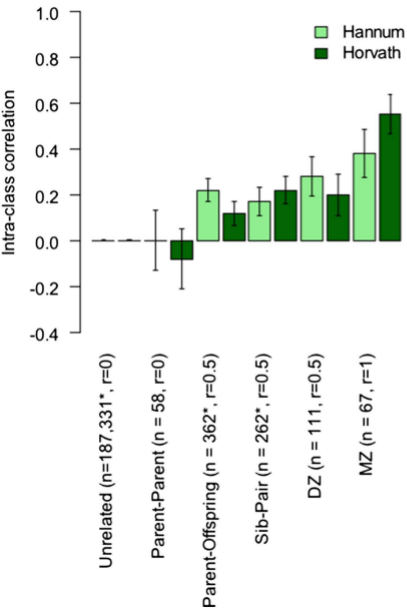
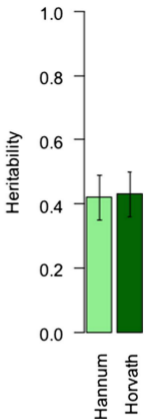
## Basic Adjusted Horvath $\Delta_{age}$ and Mortality



## Fully Adjusted Horvath $\Delta_{age}$ and Mortality



**LBC1921 Hannum Survival Curves****LBC1921 Horvath Survival Curves**

**A****B**

## Additional files provided with this submission:

**Additional file 1.** Contains a table with summary information for the additional covariate data in the four cohorts (17k)

<http://genomebiology.com/supplementary/s13059-015-0584-6-s1.docx>

**Additional file 2.** Contains a table with the white blood cell-adjusted associations of Horvath and Hannum  $\Delta$ age with mortality (17k)

<http://genomebiology.com/supplementary/s13059-015-0584-6-s2.docx>

**Additional file 3.** Presents the results from the analyses that accounted for differences in naive T cell abundance. It contains a table and cohort specific figures that assess the association between naive T cell abundance and age acceleration, chronological age, and mortality. Cox model output is also included to show the association between methylation age acceleration and mortality after adjusting for naive T cells (858k)

<http://genomebiology.com/supplementary/s13059-015-0584-6-s3.docx>

**Additional file 4.** Contains a table of the associations of  $\Delta$ age (per 5 years) with mortality in LBC1921 and LBC1936 after individual adjustment for covariates. The basic adjustment model controls for age and sex. A separate Cox model adjusting for age, sex, and a single covariate was analysed along with a saturated model that included age, sex, and all covariates together (16k)

<http://genomebiology.com/supplementary/s13059-015-0584-6-s4.docx>

**Additional file 5.** Contains a figure with the meta-analysis results of sex-stratified, age-adjusted models of  $\Delta$ age against mortality (212k)

<http://genomebiology.com/supplementary/s13059-015-0584-6-s5.docx>

**Additional file 6.** Contains a figure with the meta-analysis results of age- and sex-adjusted  $\Delta$ age against mortality with a 2 year time lag (37k)

<http://genomebiology.com/supplementary/s13059-015-0584-6-s6.docx>

**Additional file 7.** Contains a table with the associations of  $\Delta$ age with known mortality risk factors. Separate linear regression analyses were performed for each covariate. All models adjusted for sex except for NAS, which only had male participants. Analysis of FHS data were adjusted for laboratory batch and family structure (18k)

<http://genomebiology.com/supplementary/s13059-015-0584-6-s7.docx>

**Additional file 8.** Contains a figure illustrating the model fitting in BSGS to create residuals from a regression of predicted age against chronological age. A quadratic chronological age term was included in the adolescent model along with chronological age and sex (red line). A linear model adjusting for age and sex was included in the adult model (blue line) (37k)

<http://genomebiology.com/supplementary/s13059-015-0584-6-s8.docx>