

RESEARCH

Open Access



# Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing

Qian Zhang<sup>1</sup>, Costanza L. Vallergera<sup>1</sup>, Rosie M. Walker<sup>2</sup>, Tian Lin<sup>1</sup>, Anjali K. Henders<sup>1</sup>, Grant W. Montgomery<sup>1</sup>, Ji He<sup>3</sup>, Dongsheng Fan<sup>3</sup>, Javed Fowdar<sup>4</sup>, Martin Kennedy<sup>5</sup>, Toni Pitcher<sup>6,7</sup>, John Pearson<sup>5</sup>, Glenda Halliday<sup>8</sup>, John B. Kwok<sup>8</sup>, Ian Hickie<sup>8</sup>, Simon Lewis<sup>8</sup>, Tim Anderson<sup>6,7</sup>, Peter A. Silburn<sup>9</sup>, George D. Mellick<sup>4</sup>, Sarah E. Harris<sup>2,10</sup>, Paul Redmond<sup>10</sup>, Alison D. Murray<sup>11</sup>, David J. Porteous<sup>2,10</sup>, Christopher S. Haley<sup>12</sup>, Kathryn L. Evans<sup>2</sup>, Andrew M. McIntosh<sup>2,10,13</sup>, Jian Yang<sup>1</sup>, Jacob Gratten<sup>1,9</sup>, Riccardo E. Marioni<sup>2,10</sup>, Naomi R. Wray<sup>1,9</sup>, Ian J. Deary<sup>10,14</sup>, Allan F. McRae<sup>1†</sup> and Peter M. Visscher<sup>1\*†</sup>

## Abstract

**Background:** DNA methylation changes with age. Chronological age predictors built from DNA methylation are termed ‘epigenetic clocks’. The deviation of predicted age from the actual age (‘age acceleration residual’, AAR) has been reported to be associated with death. However, it is currently unclear how a better prediction of chronological age affects such association.

**Methods:** In this study, we build multiple predictors based on training DNA methylation samples selected from 13,661 samples (13,402 from blood and 259 from saliva). We use the Lothian Birth Cohorts of 1921 (LBC1921) and 1936 (LBC1936) to examine whether the association between AAR (from these predictors) and death is affected by (1) improving prediction accuracy of an age predictor as its training sample size increases (from 335 to 12,710) and (2) additionally correcting for confounders (i.e., cellular compositions). In addition, we investigated the performance of our predictor in non-blood tissues.

**Results:** We found that in principle, a near-perfect age predictor could be developed when the training sample size is sufficiently large. The association between AAR and mortality attenuates as prediction accuracy increases. AAR from our best predictor (based on Elastic Net, <https://github.com/qzhang314/DNAm-based-age-predictor>) exhibits no association with mortality in both LBC1921 (hazard ratio = 1.08, 95% CI 0.91–1.27) and LBC1936 (hazard ratio = 1.00, 95% CI 0.79–1.28). Predictors based on small sample size are prone to confounding by cellular compositions relative to those from large sample size. We observed comparable performance of our predictor in non-blood tissues with a multi-tissue-based predictor.

**Conclusions:** This study indicates that the epigenetic clock can be improved by increasing the training sample size and that its association with mortality attenuates with increased prediction of chronological age.

**Keywords:** DNA methylation, Age prediction, Epigenetic clock, Ageing, Mortality

\* Correspondence: [peter.visscher@uq.edu.au](mailto:peter.visscher@uq.edu.au)

†Allan F. McRae and Peter M. Visscher jointly supervised the study.

<sup>1</sup>Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia

Full list of author information is available at the end of the article



## Background

Ageing is a major risk for diseases and mortality [1, 2]. Chronological age has been widely used as a marker of ageing due to ease and accuracy of measurement [1]. However, it is not necessarily a good predictor of biological ageing since individuals with the same chronological age can vary in health, especially in later life [3]. Therefore, researchers have attempted to search for biomarkers of ageing that can predict functional capability at a later age [4, 5]. In 2013, Hannum et al. and Horvath built chronological age predictors (termed 'epigenetic clock') based on DNA methylation [6, 7]. Subsequently, a number of other DNA methylation-based 'clocks' have been developed [8–11]. These clocks utilize age acceleration residuals (AAR, defined as the residuals from regressing predicted age on chronological age) as a biomarker of ageing [7]. Individuals with positive AAR are considered to be biologically older than their chronological age, and vice versa.

A number of studies have identified associations between AAR and mortality, obesity and other health-related traits [12–15]. However, a better way of predicting these health-related traits is developing a predictor with the target trait as a reference [8, 10, 11, 16]. For example, a mitotic clock has been built for tracking the proliferation of cancer [8, 16]. DNAPhenoAge [10] and DNAmGrimAge [11] predictors were developed to predict a composite phenotype (composed of chronological age and clinical markers). Both of these predictors show stronger associations with lifespan and mortality than Horvath's age predictor [7]. By definition, AAR from a perfect chronological age predictor should be zero. Therefore, no information on inter-individual variation in biological age can be identified based on such a predictor [17]. Nevertheless, whether we can develop a perfect chronological age predictor based on DNA methylation is unknown. Besides, whether the reported associations between AAR and health-related traits (e.g., mortality) are inflated (e.g., by potential confounders) and/or affected by the prediction accuracy of 'epigenetic clock' has not been investigated.

In the present study, to investigate whether a perfect DNA methylation-based age predictor is theoretically available, we estimated the proportion of variance of age that could be explained by DNA methylation using a mixed linear model. We developed age predictors based on training sets with various sample sizes using Elastic Net [18] and Best Linear Unbiased Prediction (BLUP) [19]. We calculated AAR based on these age predictors and examined whether its association with mortality is affected by the prediction accuracy and potential confounders. We further investigated the performance of our predictor in samples from tissues other than blood.

## Methods

### Study population

We collected 14 data cohorts with 13,661 samples (13,402 from blood and 259 from saliva) in the age range of 2 to 104 years measured on the DNA methylation HumanMethylation450 chips and Illumina EPIC (850 K) arrays (Table 1). Eight of these cohorts were from the public domain (GEO database) and six datasets from the investigators. The six datasets include Lothian Birth Cohort (LBC) 1921/1936, Brisbane Systems Genomics Study (BSGS), Systems Genomic of Parkinson's Disease consortium (SGPD), Motor Neuron Disease cohort (MND), and Generation Scotland (GS). Details of samples in BSGS and LBC cohorts can be found in Powell et al. [22] and Deary et al. [20, 21]. GS is a population- and family-based cohort recruited through the National Health Service (NHS) Scotland general practitioner research network [24, 25]. The SGPD cohort is from a collaborative research project on systems genomics of Parkinson's disease. Similarly, the MND cohort is from a systems genomics study of Motor Neuron Disease in Chinese subjects (see descriptions in Benyamin et al. [23]). DNA methylation beta value at each probe was used for analysis.

After quality control, we obtained a set of 319,607 probes (called the No Pruned set) for each sample (Additional file 1). The effective number of independent methylation probes was previously reported to be around 200 [30], indicating a dense correlation structure. Therefore, we generated a pruned probe set (128,405 probes) (Additional file 1) and compared its performance in age prediction with that based on No Pruned set. Two cohorts were identified to be outliers in the principal components analysis (PCA) using probes from the No Pruned set (Additional file 1: Figure S1). However, the prediction accuracy in both of these cohorts is not low, and thus, we kept them in the subsequent analysis (Additional file 1).

Most of the training samples of our age predictors are from the blood. To test the performance of our age predictors in non-blood tissues, we downloaded 13 additional cohorts (from GEO database) with samples from tissues other than the blood (Additional file 2: Table S1).

### Estimating the proportion of variance of chronological age explained by DNA methylation

The GS and SGPD samples were used in estimating the proportion of variance of chronological age explained by DNA methylation. Among the 5101 samples in the GS cohort, a subset of 2586 unrelated individuals, with a genetic relationship coefficient below 0.05 and with no shared nuclear family environment, were considered for the analysis. Meanwhile, we selected 1299 unrelated (genetic relationship coefficient < 0.05) individuals with available age information in SGPD. We estimated the

**Table 1** Description of DNA methylation cohorts

Cohort <sup>1</sup>	Sample size <sup>2</sup>	Number of samples with valid age	Mean age (SD)	Age range	Source	Disease
LBC1921 [20, 21]	692	692	82.3 (4.3)	[77.8, 90.6]	Blood	Not available
LBC1936 [20, 21]	2326	2326	72.4 (2.8)	[67.7, 77.7]	Blood	Not available
BSGS [22]	614	614	21.4 (14.1)	[9.9, 74.9]	Blood	Not available
SGPD	1962	1556	67.2 (9.5)	[23.0, 104.0]	Blood	Parkinson's disease 988, control 974
MND [23]	695	600	45.2 (15.0)	[17.0, 76.0]	Blood	Motor neuron disease (MND) 497, control 198
GS [24, 25]	5101	5100	48.5(14.0)	[18.0, 94.5]	Blood	Not available
GSE72775 [26]	335	335	70.2 (10.3)	[36.5, 90.5]	Blood	Not available
GSE78874 [26]	259	259	68.8(9.7)	[36.0, 88.0]	Saliva	Not available
GSE72773 [26]	310	310	65.6 (13.9)	[35.1, 91.9]	Blood	Not available
GSE72777 [26]	46	46	14.7 (10.4)	[2.2, 35.0]	Blood	Not available
GSE41169 [27]	95	95	31.6 (10.3)	[18.0, 65.0]	Blood	Schizophrenia 62, control 33
GSE40279 [6]	656	656	64.0 (14.7)	[19.0, 101.0]	Blood	Not available
GSE42861 [28]	689	689	51.9 (11.8)	[18.0, 70.0]	Blood	Rheumatoid arthritis 354, control 335
GSE53740 [29]	384	383	67.8(9.6)	[34.0, 93.0]	Blood	Alzheimer's disease 15, corticobasal degeneration 1, frontotemporal dementia (FTD) 121, FTD/MND 7, progressive supranuclear palsy 43, control 193, unknown 4

<sup>1</sup>LBC Lothian Birth Cohort, BSGS Brisbane Systems Genomics Study, SGPD Systems Genomic of Parkinson's Disease consortium, MND Motor Neuron Disease cohort, GS Generation Scotland. Cohorts with prefix GSE are from the GEO database

<sup>2</sup>The number of samples in each cohort. Some samples in LBC were measured from the same individual but at different chronological age

proportion of variance of age explained when fitting all probes simultaneously by the restricted maximum likelihood (REML) method implemented in OSCA [31], which is analogous to estimating heritability from SNP data [32].

$$Y = Xu + e$$

where  $Y$  is an  $n \times 1$  vector of phenotype values (here chronological age) with  $n$  being the sample size.  $X$  is an  $n \times m$  matrix of standardized DNA methylation measures of all  $m$  probes,  $u$  is an  $m \times 1$  vector of the joint effects of all probes on the phenotype, and  $e$  is an  $n \times 1$  vector of residuals. Both  $u$  and  $e$  are random effects with  $u \sim N(0, I\sigma_u^2)$  and  $e \sim N(0, I\sigma_e^2)$ ,  $\sigma_u^2$  and  $\sigma_e^2$  can be estimated by REML. The proportion of variance of chronological age explained by all DNA methylation probes is defined as:

$$\rho^2 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

$\rho^2 = 0$  means that DNA methylation is not associated with phenotypic differences between individuals;  $\rho^2 = 1$  means all the variation in the phenotype can be

explained by the joint effects of DNA methylation from all probes.

### Building age predictors

We generated 65 training sets from the 14 cohorts. Each training set has a certain number (ranging between 1 and 13) of cohorts sampled from the 14 cohorts, and the unselected cohorts were used as test set(s). For each number, we repeated random sampling five times (Additional file 2: Figure S1). For example, there will be five training sets composed of ten cohorts, and the ten cohorts in each training set were sampled from the 14 cohorts randomly. In total, 65 (13 × 5) training sets were generated.

Based on each training set, we built our predictors using two methods, namely Elastic Net and BLUP. Both of them are based on a linear regression:

$$Y = X\beta + e$$

where  $Y$  is an  $n \times 1$  vector of chronological age with  $n$  being the sample size.  $X$  is an  $n \times m$  matrix of DNA methylation measures of all  $m$  probes, whereby  $X_{ij}$  is the DNA methylation of individual  $i$  at probe  $j$ , and  $e$  is the Gaussian error. The two methods differ in how

they select probes that are associated with age and how their effects are estimated (Additional file 1). BLUP would perform better than Elastic Net when there are many predictors (probes), all with non-zero effects on the target trait and effects drawn from a normal distribution. However, this method needs a large sample size to estimate small effect sizes. It is not always the case that there are many predictors associated with a trait.

We implemented two estimates to evaluate the performance of our age predictors: (1) correlation between predicted age and chronological age in the test data set and (2) root mean square error (RMSE) of the predicted age in the test data set. Correlation indicates the strength of a linear relationship between the predicted age and chronological age, and RMSE reveals the variation of the difference between predicted and chronological age.

The relationship between chronological age and DNA methylation could be nonlinear [33]. We selected eight DNA methylation cohorts with a sample size larger than 600 to evaluate the impact of data transformation in age prediction: LBC1921, LBC1936, GS, BSGS, SGPD, MND, GSE40279, and GSE42861. For each cohort, we randomly selected 70% of the samples as a training set and the remaining 30% were used as the test set. Only 50,000 randomly selected probes were used for computational efficiency. Power parameter  $\lambda$  (ranges from 0.1 to 2 with 0.05 as the interval) was used to transform the original beta value of DNA methylation BV to  $BV^\lambda$ . Only BLUP was used for age prediction because of its low bias. DNA methylation  $M$  value, arcsine square root transformed methylation beta value, and log transformed methylation beta value were also used to compare to raw DNA methylation beta value in prediction accuracy.

#### Association between age acceleration residual and mortality

We detected the association between age acceleration residual (AAR) and mortality by using the Cox proportional hazards regression model with age at sample collection, sex plate, array, position on the array and hybridization date as the covariates (all treated as fixed effect factors). Samples from the updated data in Marioni et al. [12]: LBC1921 (wave one,  $N = 436$ ,  $N_{\text{deaths}} = 386$ ) and LBC1936 (wave one,  $N = 906$ ,  $N_{\text{deaths}} = 214$ ) were used in this analysis. AAR was calculated based on age predictors excluding LBC1921/LBC1936 as part of the training set (sample size ranges from 335 to 12,710). Cox models were performed utilizing the 'survival' library in R [34]. We applied a sensitivity analysis by additionally including the measured cell count of each white blood cell type (basophils, eosinophils, monocytes, lymphocytes, and neutrophils) as covariates in the Cox

model. The change of test statistics of AAR before and after fitting these covariates was quantified.

Variation in cellular compositions is known to be associated with DNA methylation [35]. We examined whether AAR-associated CpG sites were enriched in the probes that show heterogeneity in DNA methylation across cell types (72,393 cellular heterogeneity probes) [36] using the Fisher exact test. We calculated AAR for samples from LBC1936 wave one using predictors without LBC1936 in the training set. Based on AAR from each predictor, we estimated its association with DNA methylation at each CpG site. AAR-associated CpG sites were defined as the probes with  $P$  value smaller than Bonferroni-corrected  $P$  value threshold ( $P = 0.05/319,607$ ).

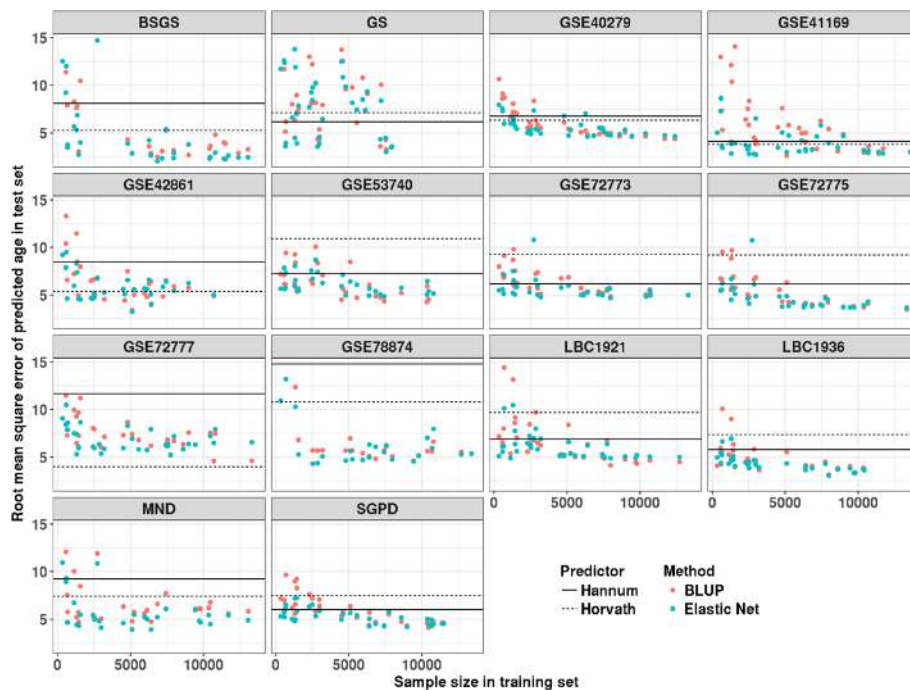
## Results

### Estimation of variation in age from using all probes

The proportion of variance of age explained by DNA methylation was close to 1 in both cohorts based on REML analysis (GS: proportion explained = 1, SE = 0.0036; SGPD: proportion explained = 0.99, SE = 0.058) ('Methods' section), suggesting that variation of chronological age between individuals could be entirely explained by the joint effect of DNA methylation from all CpG sites. It indicates that a perfect age predictor can in principle be developed based on DNA methylation data if all probe associations are estimated without error. To demonstrate that this result is not caused by overestimation, we undertook a permutation test using the same cohorts. We shuffled the ages across individuals and found that DNA methylation did not explain any significant amount of variation in GS (proportion explained = 0, SE = 0.0030) and SGPD (proportion explained = 0.0079, SE = 0.013).

### Age predictors with different prediction accuracy

Based on each training set (65 in total), we built our predictors using BLUP and Elastic Net ('Methods' section). Results on the test sets show that both methods have a decrease of RMSE (Fig. 1) and an increase of correlation (Additional file 2: Figure S2) when the training sample size increased. The smallest RMSE based on Elastic Net was 2.04 years, which is lower than that based on Hannum's and Horvath's age predictors (Additional file 2: Figure S3). This method gave better results with RMSE relative to BLUP for small training sample size, although the difference with BLUP became smaller with increased sample size (Additional file 2: Figure S4). The imperfect prediction performance (RMSE = 2.04) of the predictor in this study could be caused by an insufficient number of training samples and/or different batch effects between the training and test dataset. Analogous to estimation and prediction of complex traits using SNPs, prediction accuracy is expected to be less than the total variance explained



**Fig. 1** The relationship between training sample size and predictor error measured at the square root of the mean squared error (RMSE) in test data sets. Each point represents the RMSE of the test result based on predictors with different sample size and methods. Points with RMSE larger than 15 are excluded. Prediction results from Horvath are marked as black dash line, and the black solid line represents prediction result from Hannum's age predictor

by all features. They are the same when effect sizes are estimated without error.

Apart from the sample size, we found increasing the age range of training set and the age similarity between training and test set can improve the prediction accuracy (Additional file 2: Table S2 & S3). No steady improvement could be achieved by using transformed beta value ('Methods' section, Additional file 2: Figure S5 & S6). In addition, higher RMSE and lower correlation can always be observed for prediction results based pruned set than full probe set (Additional file 2: Figure S7). The overlap (and correlation) is small between 514 probes in our predictor (selected by Elastic Net, based on 13,566 training samples) and that in the age predictors of Hannum (30 in common) and Horvath (11 in common) (Additional file 2: Figure S8). Probes in these two predictors were found to be redundant for age prediction (Additional file 2: Figure S9), and better prediction accuracy can still be observed after removing these probes (Additional file 2: Figure S10).

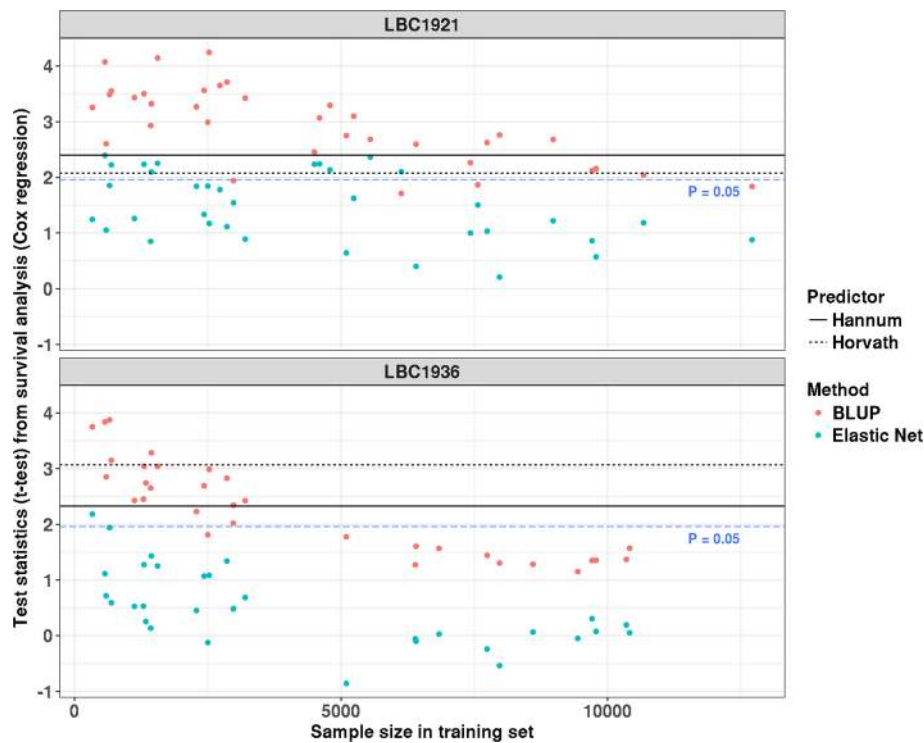
#### Association between AAR and mortality

Based on samples from wave one of both LBC1921 and LBC1936, we observed a decrease of the test statistics ( $z$ -statistic) for the association between AAR and death (from the Cox regression) with increasing sample size in training data set (Fig. 2). For AAR calculated from the

predictor with the largest training sample size, it was not associated with the mortality in either LBC1921 or LBC1936 using BLUP (LBC1921: hazard ratio = 1.20, 95% CI 0.99–1.46,  $P$  value = 0.066; LBC1936: hazard ratio = 1.25, 95% CI 0.95–1.64,  $P$  = 0.12) or Elastic Net (LBC1921: hazard ratio = 1.08, 95% CI 0.91–1.27,  $P$  = 0.38; LBC1936: hazard ratio = 1.00, 95% CI 0.79–1.28,  $P$  = 0.96) (Table 2). In contrast, results based on the age predictors of Hannum and Horvath were significant ( $P$  < 0.05, Table 2).

AAR-associated CpG sites from age predictors of Hannum (odds ratio = 3.85, 95% CI 3.71–4.00,  $P$  <  $2.2 \times 10^{-16}$ ) and Horvath (odds ratio = 2.53, 95% CI 2.45–2.61,  $P$  <  $2.2 \times 10^{-16}$ ) were found to be enriched in probes that show heterogeneity in DNA methylation across cell types ('Methods' section), suggesting AAR from these two predictors is affected by the cellular compositions. A decrease of the odds ratio from the enrichment test was observed with the increase of training sample size for both Elastic Net and BLUP-based age predictors (Fig. 3). No significant enrichment (Elastic Net: odds ratio = 0.78, 95% CI 0.47–1.23,  $P$  = 0.33; BLUP: odds ratio = 1, 95% CI 0.82–1.21,  $P$  = 1.00) was found for the age predictors based on the largest training sample size (Table 3).

Apart from AAR, cellular compositions are also related to mortality [37], which suggests it could be a confounder in the association between AAR and mortality. We re-ran



**Fig. 2** Relationship between the training sample size and the test statistics (*t* test) from the association between age acceleration residual (AAR) and mortality. Each point represents the test statistic from the survival analysis based on the predicted ages from predictors with different training sample sizes

the analysis based on AAR adjusting for measured white blood cell (WBC) counts (basophils, eosinophils, monocytes, lymphocytes, and neutrophils) ('Methods' section). A decrease of the test statistics (from the Cox regression) after correcting for the WBC counts was observed, especially when the training sample size is small (Additional file 2: Figure S11). After adjustment for WBC, none of the associations remained significant ( $P < 0.05$ ) except for the association in LBC1936 based on the predictor of Horvath ( $P = 0.032$ ). Nevertheless, the significance of this association did not pass the Bonferroni-corrected  $P$  value threshold ( $P = 0.05/4$ ) (Table 2).

#### Age prediction in non-blood tissues

The majority of our samples are from the blood, and we observed a significant improvement in the prediction results for the samples from saliva when more blood samples were included in the training set (Fig. 1, Additional file 2: Figure S5). This increase is expected since samples from saliva were reported to exhibit more than 80% contamination by immune cells [38]. We also quantified the performance of our predictor in other non-blood tissues based on samples from 13 additional data sets (Additional file 2: Table S1). We compared the performance of our predictor (based on Elastic Net) with Horvath's age predictor (based on Elastic Net) in these cohorts. Horvath's age predictor is

a pan-tissue epigenetic clock (training samples were from 51 tissues and cell types). It has a good tolerance for tissue specificity since DNA methylation on the selected CpG sites by his predictor was related to age across the tissues and cell types in his training dataset. We found that our predictor has better performance in samples from the endometrium and saliva, in terms of the Pearson correlation between predicted age and chronological age (Fig. 4a). On the other hand, Horvath's age predictor outperformed our predictor in samples from the brain. Their performance in other tissues (breast, liver, adipose, and muscle) was similar, even though training samples in our predictor are not from these tissues. A similar pattern was observed when comparing the RMSE between these two predictors (Fig. 4b).

#### Discussion

We investigated the relationship between the prediction accuracy of a DNA methylation-based age predictor ('epigenetic clock') and its application as a biomarker of ageing. Age predictors with various prediction performance were built based on datasets with different sample sizes (ranging from  $N = 335$  to 13,566). We used Cox regression to detect the association between age acceleration residuals (AAR, from different age predictors) and mortality based on samples from LBC1921 and LBC1936. We

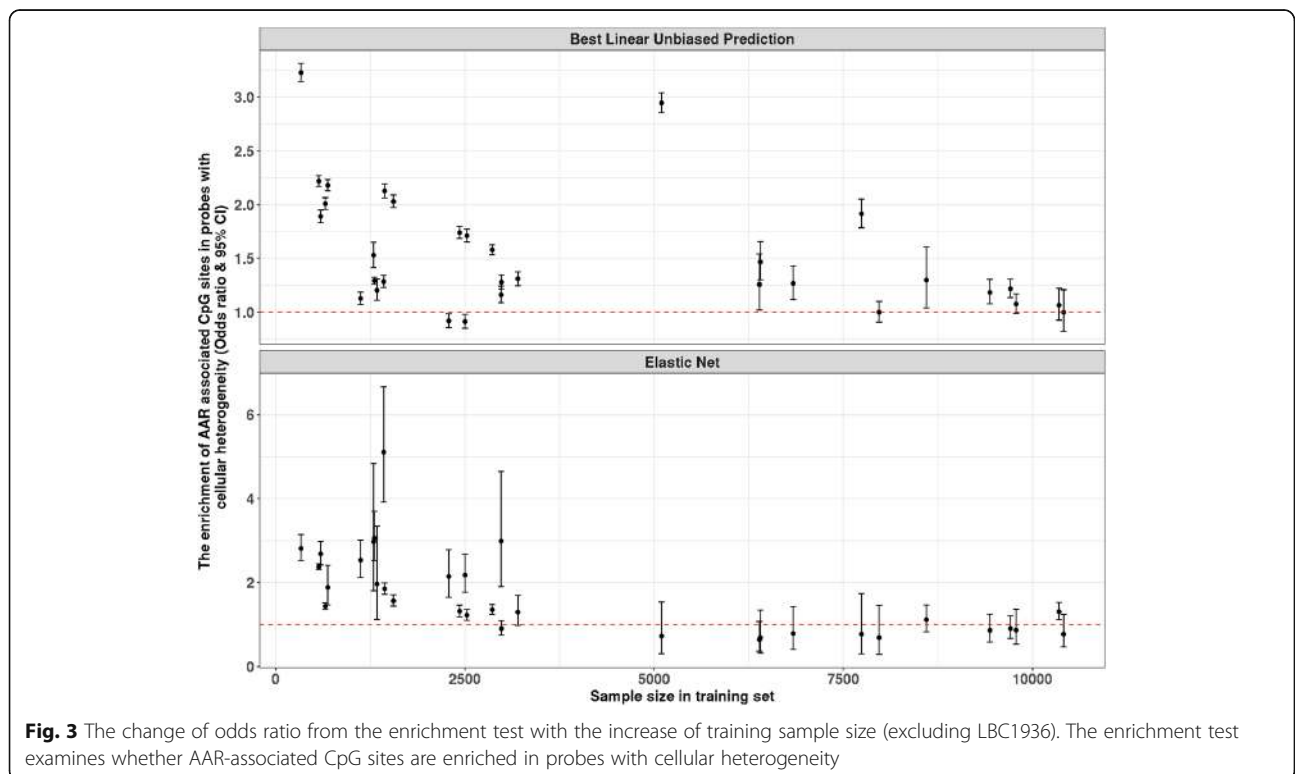
**Table 2** Summary details of two LBC cohorts and the relationship between all-cause mortality and predicted age from different methods (before and after cell count correction)

	LBC1921 wave one	LBC1936 wave one
<i>N</i>	436	906
<i>N</i> <sub>deaths</sub>	386	214
Chronological age, mean (SD) <sup>1</sup>	79.1 (0.6)	69.5 (0.8)
Before cell count correction		
Hannum, mean (SD)	80.3 (6.2)	71.3 (5.7)
Hannum, hazard ratio ( <i>P</i> value, 95% CI) <sup>2</sup>	1.12 (0.016, 1.02–1.23)	1.18 (0.020, 1.02–1.37)
Horvath, mean (SD)	73.8 (6.9)	66.1 (6.4)
Horvath, hazard ratio ( <i>P</i> value, 95% CI)	1.09 (0.038, 1.00–1.20)	1.19 (0.0022, 1.06–1.32)
Elastic Net, mean (SD) <sup>3</sup>	77.4 (3.6)	72.5 (3.2)
Elastic Net, hazard ratio ( <i>P</i> value, 95% CI)	1.08 (0.38, 0.91–1.27)	1.00 (0.96, 0.79–1.28)
BLUP, mean (SD) <sup>3</sup>	77.3 (3.3)	72.5 (2.8)
BLUP, hazard ratio ( <i>P</i> value, 95% CI)	1.20 (0.066, 0.99–1.46)	1.25 (0.12, 0.95–1.64)
After cell count correction		
Hannum, hazard ratio ( <i>P</i> value, 95% CI)	1.10 (0.057, 1.00–1.21)	1.11 (0.15, 0.96–1.29)
Horvath, hazard ratio ( <i>P</i> value, 95% CI)	1.07 (0.13, 0.98–1.17)	1.14 (0.032, 1.01–1.28)
Elastic Net, hazard ratio ( <i>P</i> value, 95% CI) <sup>3</sup>	1.07 (0.39, 0.91–1.27)	1.03 (0.79, 0.82–1.31)
BLUP, hazard ratio ( <i>P</i> value, 95% CI) <sup>3</sup>	1.21 (0.05, 1.00–1.48)	1.21 (0.17, 0.92–1.60)

<sup>1</sup>Mean (predicted) age and its standard deviation

<sup>2</sup>Hazard ratio, *P* value, and 95% confidence interval from the survival analysis based on the predicted age. Hazard ratios were expressed per 5 years of methylation age acceleration

<sup>3</sup>Both results of Elastic Net and BLUP were based on the age predictor with the largest training sample size (sample size = 10,411 for LBC1936 and sample size = 12,710 for LBC1921)



**Fig. 3** The change of odds ratio from the enrichment test with the increase of training sample size (excluding LBC1936). The enrichment test examines whether AAR-associated CpG sites are enriched in probes with cellular heterogeneity

**Table 3** Enrichment test on the AAR-associated CpG sites from different methods based upon samples from LBC1936 wave one

	Number of significant associations ( $P < 0.05/319,607$ )	$\lambda_{\text{median}}$	Number of CpG sites showing significant cellular heterogeneity	Odds ratio ( $P$ value, 95% CI) <sup>1</sup>
Hannum	12,015	3.6	4958	3.85 ( $P < 2.2 \times 10^{-16}$ , 3.71–4.00)
Horvath	18,847	5.4	5955	2.53 ( $P < 2.2 \times 10^{-16}$ , 2.45–2.61)
Elastic Net <sup>2</sup>	159	2.1	21	0.78 ( $P = 0.33$ , 0.47–1.23)
BLUP <sup>2</sup>	793	2.6	130	1.00 ( $P = 1.0$ , 0.82–1.21)

<sup>1</sup>The odds ratio for the enrichment of EWAS significant CpG sites in the probe set showing significant cellular heterogeneity

<sup>2</sup>Both results of Elastic Net and BLUP were based on the age predictor with the largest training sample size (training set without LBC1936, sample size = 10,411)

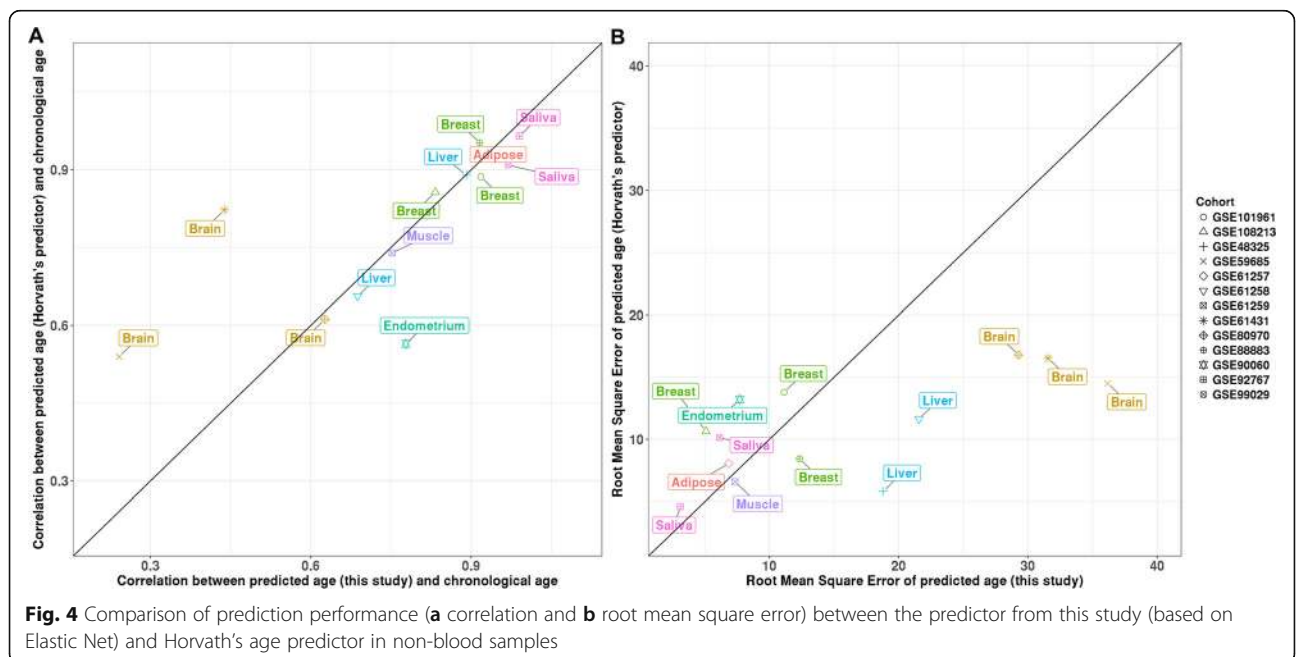
observed a decrease in the significance of the association between AAR and mortality with the improvement of the age predictor. No significant ( $P < 0.05$ ) associations were found based on the age predictor with the largest training sample size (Table 2), suggesting the improved prediction of chronological age from DNA methylation limits it as a biomarker of ageing. Our age predictor from the largest training set of 13,566 individuals is available online [39] (see Declarations).

Among the 13,661 samples from 14 cohorts in this study, 2088 were known to have a disease. DNA methylation at a few CpG sites might be different between these samples and others. Such probes would not be selected into a chronological age predictor when they are not related to age. Even if they are age-related, these probes would be weighted less (or still unselected) since their associations with the disease could affect their prediction accuracy on chronological age. Assigning these probes small effect sizes (or removing them) would

decrease the prediction error in the training process. Our results from randomization training and test sets show that our age predictors do not appear to be biased by the disease status.

Notwithstanding the highly correlated pattern of DNA methylation across the genome, we observed a decline of prediction accuracy when using a correlation pruned probe set, so that including more probes in the training model is beneficial, especially when the training sample size is small (Additional file 2: Figure S7). The improvement of prediction accuracy could be explained by the decrease of noise effect (such as batch effects) of DNA methylation in age prediction since using more probes can reduce the unexpected impact of the noise. It could also be caused by the existence of many probes with a small correlation with age, and the cumulative effect of these may be lost when using a pruned set of probes.

The AAR-associated probes from the age predictors of Horvath and Hannum were enriched in CpG sites





showing DNA methylation heterogeneity across cell types, suggesting AAR from these predictors is affected by the variation in cellular composition. The sensitivity analysis confirmed that no significant ( $P < 0.05/4$ ) associations were observed after adjusting for white blood cell counts (Table 2). This demonstrates that the difference in the cellular makeup of the samples in our test sets is a confounder in the association between AAR from the Hannum/Horvath age predictors and mortality. This result was not consistent with what has been reported by the previous study. Chen et al. demonstrated that AAR still predicts mortality after adjusting for blood cell counts in a large meta-analysis. [15] However, it should be noted that their blood cell counts were estimated based on DNA methylation, but not measured by the experiment. Hence, the actual cellular composition could still affect the association between AAR and mortality. It is also worth noting that the insignificant result in this study could be caused by lack of power. The association between AAR and mortality was merely examined in two cohorts (LBC1921 and LBC1936). More datasets with measured white blood cell counts are needed to increase detection power.

Our results show that improving the prediction accuracy of an age predictor would reduce the effect of confounders and thereby attenuate the association between AAR and death (Fig. 2). This decrease could be caused by the loss of biological age-associated CpGs in an improved epigenetic clock. It should be noted that building a biological age predictor is difficult since there is no clear definition of biological age. Nevertheless, one of the essential features of biological age is its ability to indicate the different ageing rates between individuals with the same chronological age. A previous study has reported a number of CpG sites that show variation in the longitudinal changing rates between individuals [40]. Utilizing these probes to build a biological age clock might be useful. An alternative approach is developing a predictor for biological age-related traits (e.g., life expectancy [41]), but not for biological age itself.

Although most of the samples in our age predictor are from the blood, it showed good out-of-sample prediction performance in samples from non-blood tissues. Compared with Horvath's age predictor, we observed larger correlations (between predicted age and chronological age) and smaller RMSE in samples from the saliva and endometrium, but lower correlations and larger RMSE in samples from the brain. These smaller correlations (and larger RMSE) are expected since a large proportion (23.4%) of training samples in Horvath's age predictor are from the brain. Moreover, these two predictors have similar performance in other tissues. The CpG sites in our age predictor were selected based on their associations with chronological age in blood samples. And

Horvath's age predictor used CpG sites with DNA methylation associated with chronological age across tissues and cell types. The comparable performance of these two predictors implies that most of the age-associated DNA methylation sites in the blood also change along with age in non-blood tissues.

## Conclusions

Our results have several implications for the utility of DNA methylation patterns of age as biomarkers of ageing. From the REML analysis on the SGPD and GS cohorts, we estimated that almost 100% of the variation in chronological age in those samples could be effectively captured by all the DNA methylation probes on the arrays. This implies that a near-perfect predictor of chronological age can be built based on a very large training set. Our results showing that larger sample sizes lead to a more accurate prediction is consistent with this implication. The association between AAR and mortality is confounded by the variation in cellular composition (i.e., white blood cell counts), especially when AAR is from of an age predictor ('epigenetic clock') with low performance. Overall, these results suggest that caution is warranted when interpreting estimates from these epigenetic clocks as an indicator of mortality or lifespan.

## Additional files

**Additional file 1:** Quality Control steps for DNA methylation and details of the prediction methods used in this study. **Figure S1.** Principal component (PC) 1 v.s. PC 2. Two PCs are from the PCA analysis on the samples from 14 selected cohorts. (DOCX 342 kb)

**Additional file 2:** **Figure S1.** The generation of a training set based on selected cohorts. **Figure S2.** The correlation between the predicted age and chronological age in the test data set. **Figure S3.** The comparison between chronological age and predicted age based on different studies. **Figure S4.** The difference between BLUP and Elastic Net with the increase of sample size in training data set. **Figure S5.** Improvement of age prediction based on power transformed DNA methylation. **Figure S6.** Comparison of age prediction based on arcsine square root transformed DNA methylation, log transformation, DNA methylation M value, and DNA methylation beta value. **Figure S7.** The comparison between the full probe and pruned probe sets. **Figure S8.** The correlations of DNA methylation between probes selected by Elastic Net (based on 13,566 training samples) in this study and those in Horvath's and Hannum's age predictors. **Figure S9.** The comparison of prediction accuracy before and after removing probes from probes in the Hannum and Horvath predictors. **Figure S10.** The prediction accuracy of predictors without DNA methylation probes in Hannum's and Horvath's Age predictors. **Figure S11.** Relationship between the training sample size and the change of test statistics before and after correcting for the cell counts. **Table S1.** Description of 13 DNA methylation cohorts with non-blood samples. **Table S2.** The contributions of three factors (training sample size, the absolute mean age difference between training data set and test data set, and standard deviation of age from training data set) on the prediction accuracy (RMSE). **Table S3.** The contributions of three factors (training sample size, the absolute mean age difference between training data set and test data set, and standard deviation of age from training data set) on the prediction accuracy (correlation). (DOCX 2106 kb)

### Abbreviations

AAR: Age acceleration residual; LBC: Lothian Birth Cohort; BLUP: Best linear unbiased prediction; GEO: Gene Expression Omnibus; BSGS: Brisbane Systems Genomics Study; SGPD: Systems Genomic of Parkinson's Disease consortium; MND: Motor Neuron Disease cohort; GS: Generation Scotland; PCA: Principal components analysis; REML: Restricted maximum likelihood; RMSE: Root mean square error; WBC: White blood cell

### Acknowledgements

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 the 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 Gertrude Winifred Gear Fund, The University of Edinburgh, and The University of Queensland.

### Authors' contributions

AFM and PMV conceived and designed the experiments. QZ performed all statistical analyses. QZ, AFM, and PMV wrote the paper. REM, IJD, JY, and NWR advised on the statistical methodology. CLV, RMW, TL, AKH, GW M, JH, DF, JF, MK, TP, JP, GH, JB K, IH, SL, TA, PAS, GDM, SEH, PR, ADM, DJP, CSH, KLE, AMM, and JG contributed the data. All authors read and approved the final manuscript.

### Funding

This research was supported by the Australian Research Council (DP160102400), the Australian National Health and Medical Research Council (1078037, 1078901, 1103418, 1107258, 1127440 and 1113400), and the Sylvia & Charles Viertel Charitable Foundation. Riccardo Marioni was supported by Alzheimer's Research UK Major Project Grant [ARUK-PG2017B-10]. Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping and DNA methylation profiling of the GSS/FHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland, and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award "Stratifying Resilience and Depression Longitudinally" (STRADL) Reference 104036/Z/14/Z).

### Availability of data and materials

Our age predictor from the largest training set of 13,566 individuals is available online: <https://github.com/qzhang314/DNAm-based-age-predictor> [39]. GSE40279 [6], GSE72775 [26], GSE78874 [26], GSE72773 [26], GSE72777 [26], GSE41169 [27], GSE42861 [28] and GSE53740 [29] are from GEO database.

### Ethics approval and consent to participate

Following written 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), and the Scotland A Research Ethics Committee (waves 2 and 3 07/MRE00/58). Research was carried out in compliance with the Declaration of Helsinki. All donated blood and information in SGPD were collected in accordance with the National Health and Medical Research Council's National Statement on Ethical Conduct in Human Research, with written consent obtained from each participant. Ethical approval for the SGPD was obtained from the Human Research Ethics Committee (HREC) at Griffith University (Protocol No: ESK/04/11/HREC). All components of GS received ethical approval from the NHS Tayside Committee on Medical Research Ethics (REC Reference Number: 05/S1401/89). GS has also been granted Research Tissue Bank status by the Tayside Committee on Medical Research Ethics (REC Reference Number: 10/S1402/20), providing generic ethical approval for a wide range of uses within medical research.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia. <sup>2</sup>Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK. <sup>3</sup>Department of Neurology, Peking University, Third Hospital, No. 49, North Garden Road, Haidian District, Beijing 100191, China. <sup>4</sup>Griffith Institute for Drug Discovery (GRIDD), Griffith University, Brisbane, Australia. <sup>5</sup>Department of Pathology and Biomedical Science, University of Otago, Christchurch, New Zealand. <sup>6</sup>New Zealand Brain Research Institute, Christchurch, New Zealand. <sup>7</sup>Department of Medicine, University of Otago, Christchurch, New Zealand. <sup>8</sup>Brain and Mind Centre, Sydney Medical School, The University of Sydney, Sydney, Australia. <sup>9</sup>Queensland Brain Institute, The University of Queensland, Brisbane, Queensland 4072, Australia. <sup>10</sup>Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH8 9JZ, UK. <sup>11</sup>Aberdeen Biomedical Imaging Centre, University of Aberdeen, Lillian Sutton Building, Foresterhill, Aberdeen AB25 2ZD, UK. <sup>12</sup>MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK. <sup>13</sup>Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh EH10 5HF, UK. <sup>14</sup>Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK.

Received: 9 May 2019 Accepted: 16 August 2019

Published online: 23 August 2019

### References

- Fries JF. Aging, natural death, and the compression of morbidity. *Bull World Health Organ.* 2002;80:245–50.
- Gavrilov LA, Gavrilova NS. The reliability theory of aging and longevity. *J Theor Biol.* 2001;213:527–45.
- Baker GT, Sprott RL. Biomarkers of aging. *Exp Gerontol.* 1988;23:223–39.
- McClearn GE. Biomarkers of age and aging. *Exp Gerontol.* 1997;32:87–94.
- Johnson TE. Recent results: biomarkers of aging. *Exp Gerontol.* 2006;41:1243–6.
- Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada S, Klotzle B, Bibikova M, Fan J-B, Gao Y. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell.* 2013;49:359–67 GEO: GSE40279.
- Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013;14:3156.
- Yang Z, Wong A, Kuh D, Paul DS, Rakyant VK, Leslie RD, Zheng SC, Widschwendter M, Beck S, Teschendorff AE. Correlation of an epigenetic mitotic clock with cancer risk. *Genome Biol.* 2016;17:205.
- Stubbs TM, Bonder MJ, Stark AK, Krueger F, Team BIAC, von Meyenn F, Stegle O, Reik W. Multi-tissue DNA methylation age predictor in mouse. *Genome Biol.* 2017;18:68.
- Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, Hou L, Baccarelli AA, Stewart JD, Li Y, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY).* 2018;10:573–91.
- Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, Hou L, Baccarelli AA, Li Y, Stewart JD, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY).* 2019;11:303–27.
- Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol.* 2015;16:25.
- Horvath S, Erhart W, Brosch M, Ammerpohl O, von Schonfels W, Ahrens M, Heits N, Bell JT, Tsai PC, Spector TD, et al. Obesity accelerates epigenetic aging of human liver. *Proc Natl Acad Sci U S A.* 2014;111:15538–43.
- Wolf EJ, Logue MW, Hayes JP, Sadeh N, Schichman SA, Stone A, Salat DH, Milberg W, McGlinchey R, Miller MW. Accelerated DNA methylation age: associations with PTSD and neural integrity. *Psychoneuroendocrinology.* 2016;63:155–62.
- Chen BH, Marioni RE, Colicino E, Peters MJ, Ward-Caviness CK, Tsai PC, Roetker NS, Just AC, Demerath EW, Guan W, et al. DNA methylation-based

- measures of biological age: meta-analysis predicting time to death. *Aging* (Albany NY). 2016;8:1844–65.
16. Zheng SC, Widschwendter M, Teschendorff AE. Epigenetic drift, epigenetic clocks and cancer risk. *Epigenomics*. 2016;8:705–19.
  17. Field AE, Robertson NA, Wang T, Havas A, Ideker T, Adams PD. DNA methylation clocks in aging: categories, causes, and consequences. *Mol Cell*. 2018;71:882–95.
  18. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc*. 2005;67:301–20.
  19. Robinson GK. That BLUP is a good thing: the estimation of random effects. *Stat Sci*. 1991;6:15–32.
  20. Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol*. 2011;41:1576–84.
  21. Taylor AM, Pattie A, Deary IJ. Cohort profile update: the Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol*. 2018;47:1042–1042r.
  22. Powell JE, Henders AK, McRae AF, Caracella A, Smith S, Wright MJ, Whitfield JB, Dermitzakis ET, Martin NG, Visscher PM, Montgomery GW. The Brisbane systems genetics study: genetical genomics meets complex trait genetics. *PLoS One*. 2012;7:e35430.
  23. Benyamin B, He J, Zhao Q, Gratten J, Garton F, Leo PJ, Liu Z, Mangelsdorf M, Al-Chalabi A, Anderson L, et al. Cross-ethnic meta-analysis identifies association of the GPX3-TNIP1 locus with amyotrophic lateral sclerosis. *Nat Commun*. 2017;8:611.
  24. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, Deary IJ, Macintyre DJ, Campbell H, McGilchrist M, et al. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol*. 2013;42:689–700.
  25. Navrady LB, Wolters MK, MacIntyre DJ, Clarke TK, Campbell AI, Murray AD, Evans KL, Seckl J, Haley C, Milburn K, et al. Cohort profile: Stratifying Resilience and Depression Longitudinally (STRADL): a questionnaire follow-up of Generation Scotland: Scottish Family Health Study (GS:SFHS). *Int J Epidemiol*. 2018;47:13–14g.
  26. Horvath S, Gurven M, Levine ME, Trumble BC, Kaplan H, Allayee H, Ritz BR, Chen B, Lu AT, Rickabaugh TM, et al. An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol*. 2016;17:171 GEO: GSE72775, GSE78874, GSE72773 and GSE72777.
  27. Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, van Eijk K, van den Berg LH, Ophoff RA. Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol*. 2012;13:R97 GEO: GSE41169.
  28. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinus L, Acevedo N, Taub M, Ronninger M, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol*. 2013;31:142–7 GEO: GSE42861.
  29. Li Y, Chen JA, Sears RL, Gao FY, Klein ED, Karydas A, Geschwind MD, Rosen HJ, Boxer AL, Guo WL, et al. An epigenetic signature in peripheral blood associated with the haplotype on 17q21.31, a risk factor for neurodegenerative tauopathy. *PLoS Genet*. 2014;10:e1004211 GEO: GSE53740.
  30. Shah S, Bonder MJ, Marioni RE, Zhu Z, McRae AF, Zernakova A, Harris SE, Liewald D, Henders AK, Mendelson MM. Improving phenotypic prediction by combining genetic and epigenetic associations. *Am J Hum Genet*. 2015;97:75–85.
  31. Zhang FT, Chen WH, Zhu ZH, Zhang Q, Nabais MF, Qi T, Deary IJ, Wray NR, Visscher PM, McRae AF, Yang J. OSCA: a tool for omic-data-based complex trait analysis. *Genome Biol*. 2019;20:107.
  32. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*. 2010;42:565–9.
  33. Johnson ND, Wiener HW, Smith AK, Nishitani S, Absher DM, Amett DK, Aslibekyan S, Conneely KN. Non-linear patterns in age-related DNA methylation may reflect CD4(+) T cell differentiation. *Epigenetics*. 2017;12:492–503.
  34. Therneau TM. A package for survival analysis in S. 2.38 edition; 2015.
  35. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *Bmc Bioinformatics*. 2012;13:86.
  36. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol*. 2014;15:R31.
  37. Kim S, Eliot M, Koestler DC, Wu WC, Kelsey KT. Association of neutrophil-to-lymphocyte ratio with mortality and cardiovascular disease in the Jackson Heart Study and Modification by the Duffy antigen variant. *JAMA Cardiol*. 2018;3:455–62.
  38. Zheng SC, Webster AP, Dong D, Feber A, Graham DG, Sullivan R, Jevons S, Lovat LB, Beck S, Widschwendter M, Teschendorff AE. A novel cell-type deconvolution algorithm reveals substantial contamination by immune cells in saliva, buccal and cervix. *Epigenomics*. 2018;10:925–40.
  39. Zhang Q. DNA methylation based chronological age predictor. Github. 2019. <https://github.com/qzhang314/DNAm-based-age-predictor/tree/v1.0.0/>. <https://doi.org/10.5281/zenodo.3369456>.
  40. Zhang Q, Marioni RE, Robinson MR, Higham J, Sproul D, Wray NR, Deary IJ, McRae AF, Visscher PM. Genotype effects contribute to variation in longitudinal methylome patterns in older people. *Genome Med*. 2018;10:75.
  41. Lin Q, Weidner CI, Costa IG, Marioni RE, Ferreira MRP, Deary IJ, Wagner W. DNA methylation levels at individual age-associated CpG sites can be indicative for life expectancy. *Aging-US*. 2016;8:394–401.

## Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

