Improved prediction of chronological age from DNA methylation limits it as a biomarker of ageing

Qian Zhang[1], Costanza L. Vallerga[1], Rosie M Walker[2], Tian Lin[1], Anjali K. Henders[1], Grant W. Montgomery[1], Ji He[3], Dongsheng Fan[3], Javed Fowdar[4], Martin Kennedy[5], Toni Pitcher[6,7], John Pearson[5], Glenda Halliday[8], John B. Kwok[8], Ian Hickie[8], Simon Lewis[8], Tim Anderson[6,7], Peter A. Silburn[9], George D. Mellick[4], Sarah E. Harris[2,10], Paul Redmond[10], Alison D. Murray[11], David J. Porteous[2,10], Christopher S. Haley[12], Kathryn L. Evans[2], Andrew M. McIntosh[2,10,13], Jian Yang[1,9], Jacob Gratten[1,9], Riccardo E. Marioni[2,10], Naomi R. Wray[1,9], Ian J. Deary[10,14], Allan F. McRae[1,15], Peter M. Visscher[1,9,15]

[1] Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia

[2] Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU

[3] Department of Neurology, Peking University, Third Hospital, No. 49, North Garden Road, Haidian District, 100191, Beijing, China

[4] Griffith Institute for Drug Discovery (GRIDD), Griffith University, Brisbane, Australia

[5] Department of Pathology and Biomedical Science, University of Otago, Christchurch, New Zealand

[6] New Zealand Brain Research Institute, Christchurch, New Zealand

[7] Department of Medicine, University of Otago, Christchurch, New Zealand

[8] Brain and Mind Centre, Sydney Medical School, The University of Sydney, Sydney, Australia

[9] Queensland Brain Institute, The University of Queensland, Brisbane, Queensland 4072, Australia

[10] Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, EH8 9JZ

[11] Aberdeen Biomedical Imaging Centre, Lilian Sutton Building, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, UK

[12] MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU

[13] Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, EH10 5HF, UK

[14] Department of Psychology, University of Edinburgh, Edinburgh, EH8 9JZ

[15] Jointly supervised the study

Abstract

DNA methylation is associated with age. The deviation of age predicted from DNA methylation from actual age has been proposed as a biomarker for ageing. However, a better prediction of chronological age implies less opportunity for biological age. Here we used 13,661 samples (from blood and saliva) in the age range of 2 to 104 years from 14 cohorts measured on Illumina HumanMethylation450/EPIC arrays to perform prediction analyses. We show that increasing the sample size achieves a smaller prediction error and higher correlations in test datasets. We demonstrate that smaller prediction errors provide a limit to how much variation in biological ageing can be captured by methylation and provide evidence that age predictors from small samples are prone to confounding by cell composition. Our predictor shows a similar or better performance in non-blood tissues including saliva, endometrium, breast, liver, adipose and muscle, compared with Horvath's across-tissue age predictor.

Key words: DNA methylation, age prediction, best linear unbiased prediction, elastic net

Introduction

Ageing as a complex biological phenomenon is related to diseases and mortality [1,2], and chronological age has been widely used as a marker of ageing due to ease and accuracy of measurement [1]. However, chronological age 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, biomarkers of ageing have become popular as they can indicate the presence or severity of some disease states [4,5]. In 2013, Hannum *et al.* and Horvath built age predictors based on DNA methylation and implemented them as biomarkers of ageing [6,7]. DNA methylation as a part of the epigenome plays an essential role in the regulation of gene

expression in the human body [8,9]. Unlike DNA which is (mostly) stable across the lifetime of an individual, DNA methylation is dynamic, and previous studies have discovered a number of CpG sites associated with chronological age [10-12]. The age predictor developed by Hannum et al. was based on 482 blood samples with methylation measured on the Illumina 450K methylation arrays, and they reported a correlation of 0.91 and a Root Mean Square Error (RMSE) of 4.9 years in their test set [6]. Horvath's age predictor was based on 8,000 samples from different tissues and cell types, and probes of these samples were from the Illumina 27K DNA methylation arrays. He reported a correlation of 0.96 and a Median Absolute Deviation (MAD) of 3.6 years in the test set. Age Acceleration Residuals (AAR) is defined as the residuals from regressing predicted age on chronological age. It has been reported to be associated with mortality, obesity and other complex traits [13-16].

According to its definition, AAR is the prediction error in a chronological age predictor. Although previous survival analysis showed a significant association between AAR and mortality [13], AAR was found not to be a mitotic clock [17]. Therefore, whether the significance of that association is inflated (e.g. by potential confounders in the error) needs to be investigated. To study the use of predicted age from DNA methylation as a biomarker of ageing, we calculated AAR based on age predictors with different prediction accuracy and investigated the relationship between prediction accuracy and the significance of AAR in survival analysis. We also investigated the effect of training sample size and statistical method on age prediction.

In the present study, we built DNA methylation-based age predictors by integrating 13,661 samples (13,402 from blood and 259 from saliva) measured on 450K DNA methylation

arrays and Illumina EPIC (850K) arrays. Two approaches were evaluated: Elastic Net [18] and Best Linear Unbiased Prediction (BLUP) [19]. We discussed the implications of our results for the scope and utility of DNA methylation based age predictor as a biomarker for biological ageing. We also explored the factors that explain prediction accuracy. Finally, the performance of predictors on samples from tissues other than blood was investigated.

Results

Availability of DNA methylation in age prediction

We downloaded eight datasets from the public domain and used six datasets from our own studies (**Table 1**). All data underwent identical quality control criteria before statistical analyses (**Material and Methods**).

Estimation of variation in age from using all probes

We used the unrelated individuals from our two largest datasets (GS, N = 2,586, SGPD, N = 1,299) to estimate the proportion of the observed variation in age that is explained when fitting all probes simultaneously, using a mixed linear model analogous to estimating heritability from SNP data [20]. The proportion of variance of age explained by DNA methylation was close to 1 in both cohorts (proportion explained = 1, SE = 0.0036, REML analysis using the software package OSCA [21] in GS, and 0.99 in SGPD, SE = 0.058), indicating 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 a violation of assumptions, 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).

Building multiple age predictors

To build our age predictors, we collected 14 cohorts and used a common set of 319,607 probes that passed quality control (**Material and Methods**) in all cohorts. We randomly combined 1 to 13 cohorts as a training set, and used the remaining cohorts as test sets. We repeated this step 65 times to generate different training sets with various sample sizes and age spectrum (**Material and Methods**). 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; (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. Two methods, namely Elastic Net [18] and BLUP [19] were used. Elastic Net was previously used by Horvath[7] and Hannum et al.[6] to build their age predictors and BLUP was used to predict age in Peters et al.[22]. These methods differ in how they select probes that are associated with age and how their effects are estimated. Results show that both methods have a decrease of RMSE (**Figure 1**) and an increase of correlation (**Supplementary Figure 1**) when the training sample size increased. The smallest RMSE based on Elastic Net was 2.04 years. 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 (**Supplementary Figure 2**).

Prediction accuracy and biological ageing

The difference between predicted age from the Hannum/Horvath predictors and chronological age (AAR) was found to be associated with all-cause mortality in later life[13]. To investigate the relationship between the significance of this association and the prediction accuracy of the predictor, we examined the association between AAR and mortality using the updated data in Marioni et al. [13]. These data were from two cohorts: LBC1921 (wave one, N = 436, N_{deaths} = 386) and LBC1936 (wave one, N = 906, N_{deaths} = 214) (**Materials and Methods**). Age predictors excluding LBC1921/LBC1936 as part of the training set (sample size ranges from 335 to 12,710) were used. We observed a decrease of the test statistics for the effect of AAR on mortality from the survival analysis (Cox regression) with increasing sample size in training data set (**Figure 2**). No significant ($P < 0.05$) associations between AAR and mortality was found based on the largest training sample size in either LBC1921 or LBC1936 using BLUP or Elastic Net (**Table 2 and Supplementary Table 1**). In contrast, results based on the age predictors of Hannum and Horvath were significant ($P < 0.05$, **Table 2**).

Enrichment analysis on AAR associated probes

Variation in cellular compositions is known to be associated with both DNA methylation [23] and mortality [24], which suggests it could be a confounder in the survival analysis. To investigate whether AAR is affected by cellular composition, we applied an epigenome-wide association study (EWAS) on AAR from different predictors. For each predictor, AAR associated ($P < 0.05/319,607$) probes were selected. We found that the selected CpG sites from age predictors of Hannum and Horvath were enriched in the probes that show heterogeneity in DNA methylation across cell types (cellular heterogeneity probes) [25] (**Table 3**), indicating AAR in these predictors was associated with variation in cellular composition. We also observed a decrease of the odds ratio of the enrichment test with the increase of

training sample size for both Elastic Net and BLUP based age predictors (**Figure 3**). No significant enrichment was found for the age predictors based on the largest training sample size (**Table 3**).

To examine the effect of variation in cellular content in the association between AAR and mortality, we re-ran the survival analysis based on AAR adjusting for white blood cell (WBC) counts (basophils, eosinophils, monocytes, lymphocytes, and neutrophils) (**Materials and Methods**). A decrease of the test statistics (from survival analysis) after correcting for the WBC counts was observed, especially when the training sample size is small (**Supplementary Figure 3**). The effect of AAR on survival is reduced the most (the changes of test statistics from survival analysis are largest) when adding WBC to the models using the Hannum and Horvath age predictors (**Supplementary Figure 3**). 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**). These results suggest that the significant associations between AAR and mortality is biased due to the existence of confounders like WBC counts, and that improved prediction accuracy of the age predictor reduces the effect of these confounders in the survival analysis.

Other factors related to chronological age prediction

Although the potential effect of confounders limit the "epigenetic clock" to be a biomarker of ageing, a chronological age predictor with good prediction accuracy would be a useful tool in forensics and/or other fields where chronological age is needed but not observed. To determine the factors that explain prediction accuracy, we examined the contribution of

age ranges (including absolute age difference between training and test set (Age_{diff}) and standard deviation of age (Age_{sd}) of the training set) to the RMSE/correlation of the prediction results in the test set by estimating the effect of Age_{diff}, Age_{sd} and sample size in the training set on the prediction accuracy jointly (**Material and Methods**). Results showed that RMSE was significantly associated ($P < 0.05$) with training sample size in 13 (out of 14) cohorts based on BLUP predictors, confirming that increasing the sample size leads to smaller prediction errors (**Figure 1**). In addition, eight out of 14 cohorts had a significant ($P < 0.05$) and positive Age_{diff} effect, indicating similar ages between training and test set results in better prediction accuracy (**Supplementary Table 2**). Five cohorts were found to have a statistically significant ($P < 0.05$) Age_{sd} effect on RMSE, suggesting the prediction accuracy benefits from a larger age range of the samples in the training set. Similar results were found based on Elastic Net (**Supplementary Table 3**). In addition, we did not observe any steady improvement using power-transformed Beta values (**Supplementary Figure 4**, **Materials and Methods**), the M values of DNA methylation or the arcsine square root transformed Beta values (**Supplementary Figure 5**, **Materials and Methods**).

There is a complex correlation structure in DNA methylation, and the effective number of independent methylation probes was previously reported to be around 200 [26], indicating a dense correlation structure. To compare the prediction performance between using the full probe set (319,607 probes) and a pruned probe set (128,405 probes) (**Material and methods**), we applied the same cross-validation steps to both probe sets using BLUP and Elastic Net. We identified a higher RMSE and a lower correlation for the pruned set (**Supplementary Figure 6**), indicating a loss of information when using fewer methylation probes for prediction. In addition, we found that probes in the age predictors of Hannum

and Horvath were not necessary for age prediction. Compared with these two predictors, better prediction accuracy can still be observed based on the probe set without the probes from these predictors (**Supplementary Figure 7, Material and methods**), consistent with widespread correlation among probes.

Age prediction in non-blood tissues

The majority of our samples are from 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 (**Figure 1**, **Supplementary Figure 1**). This increase is expected since samples from saliva were reported to exhibit more than 80% contamination by immune cells[27]. To quantify whether our predictor has a good performance in non-blood tissues, we downloaded 13 data sets (**Supplementary Table 4**) that contain samples from other tissues. We compared the performance of our predictor (samples are from blood and saliva, age predictor generated by Elastic Net) with Horvath's age predictor (samples from multiple tissues, using Elastic Net) in these cohorts and found that our predictor has better performance in samples from endometrium and saliva. On the other hand, Horvath's age predictor outperformed our predictor in samples from brain (**Figure 4**). Their performance in other tissues (breast, liver, adipose and muscle) were similar, even though training samples in our predictor are not from these tissues. These results demonstrate that our predictor can also be used to predict the chronological age of samples from non-blood tissues.

Discussion

We investigated the relationship between the prediction accuracy of an age predictor 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 13566). We ran survival analysis (based on age acceleration residuals AAR) using samples from LBC1921 and LBC1936, with AAR calculated using different age predictors. We observed a decrease in the significance of association between mortality and AAR 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.

We found potential effects of confounders in the association between AAR and mortality in the age predictors of Hannum and Horvath. 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 that AAR from these predictors is affected by variation in cellular composition. The further sensitivity analysis confirmed that, although the AAR from these two predictors are associated with mortality in LBC1921 and LBC1936, no significant ($P < 0.05/4$) association was observed after adjusting for white blood cell counts. This demonstrates that although the Hannum and Horvath age predictors appear to capture differences in biological ageing between people of the same age, these effects are largely driven by differences in cellular makeup of the samples, and limits their usage as a marker of biological ageing.

We also examined the factors that can affect the accuracy of chronological age prediction, including the effect of the training sample size, the age range of the training samples, the

number of probes used and the statistical methods utilised. We found a positive association between the training sample size and the prediction accuracy in test sets. Our predictors showed substantially improved prediction accuracy compared to using the estimated coefficients previously reported by Hannum [6] and Horvath [7] in blood samples. Most of this improvement appears to come from simply increasing the experimental sample size in the training set. We also found that increased similarity of ages between samples in the training and test data set can improve the prediction accuracy in the test sets (**Supplementary Tables 2 and 3**). We provide estimated effect sizes on chronological age from the largest training set of 13,566 individuals for both Elastic Net and BLUP in **Supplementary Table 5**.

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 (**Supplementary Figure 6**). 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 effects 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.

Although most of the samples in our age predictor are from 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) in samples from saliva and endometrium, but smaller correlations in samples from brain. These smaller correlations are expected since a large proportion (23.4%)

of training samples in Horvath's age predictor are from brain. Moreover, these two predictors have similar performance in other tissues, which implies that our age predictor is also useful in samples from non-blood tissues.

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 variation in chronological age in those samples could be effectively captured by all the DNA methylation probes on the arrays. For prediction, this implies that for a very large training set a near-perfect predictor of chronological age can be built. Our results showing that larger sample sizes lead to more accurate prediction is consistent with this implication. It is clear that DNA methylation measured in blood is associated with environmental exposures such as smoking, sex and BMI [28-30]. In addition, "age acceleration", the difference between actual age and that predicted from methylation, has been reported to be associated with a number of outcomes, including mortality [13,16]. However, there is currently no good DNA-methylation-based estimator of an individual's "epigenetic clock" that is free from confounders (e.g., white blood cell counts) and from prediction error caused by other factors (e.g., measurement error). The difference between actual and predicted age contains both a prediction error term based on unknown factors and possible effects of confounders. These confounders could bias the results when using "epigenetic clock" as a biomarker of ageing, for example the association between "age acceleration" and mortality is confounded by the variation in cellular composition.

Methods

Data

We collected 14 data cohorts with samples measured on the DNA methylation 450K chips and Illumina EPIC (850K) arrays (**Table 1**), eight of which were from the public domain and six datasets from the investigators. Details of the BSGS and LBC cohorts can be found in Powell et al. [31] and Deary et al. [32,33]. GS is a population and family based cohort recruited through the NHS Scotland general practitioner research network [34,35]. 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. [36]). For the purpose of this study, age at sample collection was the focus, disease status and ethnicity of individuals were not considered in any cohort. DNA methylation Beta value at each probe was used for analysis.

A total of 319,607 probes (No Pruned Set) passed our quality control and 128,405 probes (Pruned Set) were retained after pruning based upon the pairwise correlation of probes (see next section). To test the performance of age predictors in non-blood tissues, we downloaded 13 cohorts from GEO database with accession ID GSE61431 (brain)[37], GSE59685 (brain)[38], GSE80970 (brain), GSE101961 (breast)[39], GSE108213 (breast), GSE48325 (liver)[40], GSE61257 (adipose)[41], GSE61258 (liver)[41], GSE61259 (breast)[41], GSE88883 (breast)[42], GSE90060 (endometrium)[43], GSE92767 (saliva)[44], GSE99029 (saliva)[45].

Quality Control

All the samples were measured on either the Illumina HumanMethylation450 arrays or Illumina EPIC arrays. Probes with call rate less than 0.95 were removed, and probes found to contain SNPs or potentially cross-hybridizing to different locations were excluded from

further analysis [46]. After combining all the samples from different cohorts, a set of 319,607 probes remained (called No Pruned set). Pruning was performed by removing one of two probes on the same chromosome when their correlation (R^2) was higher than 0.2; this resulted in a set of 128,405 probes (called Pruned set). Both sets were used for further analysis. DNA methylation Beta value was standardized by removing the mean value and divided by the standard deviation for each sample.

Selection of DNA methylation cohorts

We collected 14 different cohorts in total, including a single cohort (GSE78874) measured in saliva rather than blood tissue. Since DNA methylation is sensitive to batch effects, cell type and tissue type[23], we applied a PCA analysis (using probes from the No Pruned Set) on the samples from these 14 cohorts to assess the presence of any "outlier" cohorts (i.e. cohorts with a low prediction accuracy from the age predictor based on the other cohorts). All the cohorts were closely matched with the exception of GSE78874 and GS (**Supplementary Figure 8)**. Samples in GSE78874 were from saliva instead of blood, and the samples in GS were measured using Illumina EPIC arrays instead of 450K DNA methylation arrays. To investigate if this difference could potentially adversely influence performance in age prediction for these two cohorts, we used a "leave-one-cohort-out" strategy to leave these two cohorts out as the test set separately and built the age predictor based on the remaining cohorts. We found both of them to have good prediction accuracy (GS: R = 0.98, RMSE = 3.52, GSE78874: R = 0.88, RMSE = 5.39), indicating a small difference between these two cohorts and other cohorts in age prediction. We used all cohorts for subsequent analyses.

Generation of training set

We generated training sets from the 14 cohorts. Each training set has a certain number of cohorts ranging between 1 to 13. For each number, we repeated random sampling five times. In total, 65 (13×5) training sets were generated.

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 5,101 samples in the GS cohort, a subset of 2,586 unrelated individuals, with a genetic relationship coefficient below 0.05 and with no shared nuclear family environment were considered for the analysis. 1,299 unrelated (genetic relationship coefficient < 0.05) individuals with available age information in SGPD were selected. Variance of age was estimated by the REML method implemented in OSCA[21].

Prediction algorithm

We compared the age prediction performance of two methods, namely Elastic Net and BLUP. Both methods are based on a linear regression:

$$Y = \sum \beta_i X_i + e$$

where Y is the chronological age, X_i is the DNA methylation of probe i and e is the Gaussian noise.

Elastic Net is a regularized regression method[18], and its objective function is defined as:

$$L(\alpha, \beta) = ||Y - X\beta||^2 + \lambda(\frac{1-\alpha}{2}\left|\left|\beta\right|\right|_2^2 + \alpha\left|\left|\beta\right|\right|_1)$$

where α and lambda are regularisation parameters. $\left|\left|\beta\right|\right|_1$ is defined as $\sum_{i=1}^{n}|\beta_i|$ and $\left|\left|\beta\right|\right|_2^2$ equals $\sum_{i=1}^{n}{\beta_i}^2$, with n the number of probes. α is set to 0.5 and λ is chosen based on cross-validation. We used the implementation of Elastic Net from the Python package glmnet[47].

BLUP is special case of ridge regression with a fixed λ.

$$\hat{\beta} = (X'X + \lambda I)^{-1}X'Y \text{ with } \lambda = \frac{\sigma_e^2}{\sigma_u^2},$$

σ_u^2 the variance of the effect size of the probe set, and σ_e^2 the variance of the residuals. We used the R package rrBLUP[48] to build the age predictor, and σ_u^2 and σ_e^2 were estimated using the REML analysis implemented in this package.

Survival analysis

We followed the same analysis approach as previously described[13]. Briefly, Cox proportional hazards regression models were used to detect the association between the AAR and mortality with age at sample collection and sex as the covariates. AAR is defined as the difference between m_{age} and chronological age, where m_{age} is the predicted age correcting for plate, array, position on the array, and hybridisation date (all treated as fixed effect factors), all of which could be confounder in survival analysis. Additional adjustments of AAR were made for WBC counts measured on the same blood samples that were analysed for methylation. Hazard ratios for AAR were expressed per five years of methylation age acceleration (**Table 2**) and per standard deviation of methylation age acceleration (**Supplementary Table 1**), respectively. Cox models were performed utilizing the 'survival' library[49] in R. Samples from wave one of LBC1921 and LBC1936 were used in this analysis.

Transformation of DNA methylation

There are non-linear patterns in age-related DNA methylation[50]. To investigate if transformed data can remove the nonlinearity and hence improve the prediction accuracy, we selected eight DNA methylation cohorts with sample size larger than 600 to evaluate the impact of data transformation: LBC1921, LBC1936, GS, BSGS, SGPD, MND, GSE40279 and GSE42861. For each cohort, we randomly selected 70% of the samples as training set, and the remaining 30% were used as test set. Only 50,000 randomly selected probes were used for computational efficiency. Power parameter λ (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^{λ}. Only BLUP was used for age prediction because of its low bias. DNA methylation M value and arcsine square root transformed methylation Beta value were also used to compare to raw DNA methylation Beta value in prediction accuracy.

Age prediction without probes from age predictors of Horvath and Hannum

We compared the probes selected by Elastic Net (based on 13,566 training samples) with those in Horvath's and Hannum's age predictors. 11 out of the 514 probes in our analysis were identified in Horvath's age predictor and 30 in Hannum's age predictor. In addition, we estimated the squared correlation (R^2) of DNA methylation between probes selected by Elastic Net and probes from the age predictor of Hannum/Horvath. We found 11 (Elastic Net-Hannum) and 10 (Elastic Net-Horvath) pairs with an R^2 larger than 0.5 (**Supplementary Figure 9**), indicating that most of the probes selected by Elastic Net are not strongly correlated with those in the other two predictors. To quantify whether the probes in the Hannum and Horvarth predictors were necessary for age prediction, we re-built our age predictors by excluding these probes. No difference in prediction accuracy was found before

and after removing these probes for the BLUP based method (**Supplementary Figure 10**). The prediction accuracy decreased for the Elastic Net based method; however, its performance was still better than when using the Hannum and Horvath age predictors (**Supplementary Figure 7**).

Acknowledgements

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 GS:SFHS 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).

Author contributions

A.F.M and P.M.V conceived and designed the experiments. Q.Z performed all statistical analyses. Q.Z, A.F.M and P.M.V wrote the paper. R.E.M, I.J.D, J.Y and N.W.R advised on statistical methodology, C.L.V, R.M.W, T.L, A.K.H, G.W. M, J.H, D.F, J.F, M.K, T.P, J.P, G.H, J.B. K, I.H, S.L, T.A, P.A.S, G.D.M, S.E.H, P.R, A.D.M, D.J.P, C.S.H, K.L.E, A.M.M, J.G contributed data. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing financial interests.

Reference

1 Fries, J. F. Aging, natural death, and the compression of morbidity. *B World Health Organ* **80**, 245-250 (2002).

2 Gavrilov, L. A. & Gavrilova, N. S. The reliability theory of aging and longevity. *J Theor Biol* **213**, 527-545, doi:10.1006/jtbi.2001.2430 (2001).

3 Baker, G. T. & Sprott, R. L. Biomarkers of aging. *Experimental gerontology* **23**, 223-239 (1988).

4 McClearn, G. E. Biomarkers of age and aging. *Experimental gerontology* **32**, 87-94 (1997).

5 Johnson, T. E. Recent results: Biomarkers of aging. *Experimental Gerontology* **41**, 1243-1246, doi:10.1016/j.exger.2006.09.006 (2006).

6 Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular cell* **49**, 359-367 (2013).

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

8 Razin, A. & Riggs, A. D. DNA methylation and gene function. *Science* **210**, 604-610 (1980).

9 Bird, A. P. CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209-213 (1986).

10 Hernandez, D. G. *et al.* Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Human molecular genetics* **20**, 1164-1172 (2011).

11 Florath, I., Butterbach, K., Müller, H., Bewerunge-Hudler, M. & Brenner, H. Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites. *Human molecular genetics* **23**, 1186-1201 (2013).

12 Bell, J. T. *et al.* Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS genetics* **8**, e1002629 (2012).

13 Marioni, R. E. *et al.* DNA methylation age of blood predicts all-cause mortality in later life. *Genome biology* **16**, 25 (2015).

14 Horvath, S. *et al.* Obesity accelerates epigenetic aging of human liver. *P Natl Acad Sci USA* **111**, 15538-15543, doi:10.1073/pnas.1412759111 (2014).

15 Wolf, E. J. *et al.* Accelerated DNA methylation age: Associations with PTSD and neural integrity. *Psychoneuroendocrino* **63**, 155-162, doi:10.1016/j.psyneuen.2015.09.020 (2016).

16 Chen, B. H. *et al.* DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging (Albany NY)* **8**, 1844-1865, doi:10.18632/aging.101020 (2016).

17 Yang, Z. *et al.* Correlation of an epigenetic mitotic clock with cancer risk. *Genome Biol* **17**, 205, doi:10.1186/s13059-016-1064-3 (2016).

18 Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**, 301-320 (2005).

19 Robinson, G. K. That BLUP is a good thing: the estimation of random effects. *Statistical science*, 15-32 (1991).

20 Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565-569, doi:10.1038/ng.608 (2010).

21 Zhang, F. *et al.* OSCA: a tool for omic-data-based complex trait analysis. *bioRxiv*, doi:10.1101/445163 (2018).

22 Peters, M. J. *et al.* The transcriptional landscape of age in human peripheral blood. *Nature communications* **6** (2015).

23 Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *Bmc Bioinformatics* **13**, doi: 10.1186/1471-2105-13-86 (2012).

24 Ruggiero, C. *et al.* White blood cell count and mortality in the Baltimore Longitudinal Study of Aging. *J Am Coll Cardiol* **49**, 1841-1850, doi:10.1016/j.jacc.2007.01.076 (2007).

25 Jaffe, A. E. & Irizarry, R. A. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biology* **15**, doi:10.1186/gb-2014-15-2-r31 (2014).

26 Shah, S. *et al.* Improving phenotypic prediction by combining genetic and epigenetic associations. *The American Journal of Human Genetics* **97**, 75-85 (2015).

27 Zheng, S. C. *et al.* A novel cell-type deconvolution algorithm reveals substantial contamination by immune cells in saliva, buccal and cervix. *Epigenomics* **10**, 925-940, doi:10.2217/epi-2018-0037 (2018).

28 Joehanes, R. *et al.* Epigenetic signatures of cigarette smoking. *Circulation: Cardiovascular Genetics* **9**, 436-447 (2016).

29 Mendelson, M. M. *et al.* Association of body mass index with DNA methylation and gene expression in blood cells and relations to cardiometabolic disease: a mendelian randomization approach. *PLoS medicine* **14**, e1002215 (2017).

30 Yousefi, P. *et al.* Sex differences in DNA methylation assessed by 450 K BeadChip in newborns. *BMC Genomics* **16**, 911, doi:10.1186/s12864-015-2034-y (2015).

31 Powell, J. E. *et al.* The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. *PLoS One* **7**, e35430, doi:10.1371/journal.pone.0035430 (2012).

32 Deary, I. J., Gow, A. J., Pattie, A. & Starr, J. M. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. *International journal of epidemiology* **41**, 1576-1584 (2011).

33 Taylor, A. M., Pattie, A. & Deary, I. J. Cohort Profile Update: The Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol*, doi:10.1093/ije/dyy022 (2018).

34 Smith, B. H. *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* **42**, 689-700, doi:10.1093/ije/dys084 (2013).

35 Navrady, L. B. *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* **47**, 13-14g, doi:10.1093/ije/dyx115 (2018).

36 Benyamin, B. *et al.* Cross-ethnic meta-analysis identifies association of the GPX3-TNIP1 locus with amyotrophic lateral sclerosis. *Nat Commun* **8**, 611, doi:10.1038/s41467-017-00471-1 (2017).

37 Pidsley, R. *et al.* Methylomic profiling of human brain tissue supports a neurodevelopmental origin for schizophrenia. *Genome Biol* **15**, 483, doi:10.1186/s13059-014-0483-2 (2014).

38 Lunnon, K. *et al.* Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat Neurosci* **17**, 1164-1170, doi:10.1038/nn.3782 (2014).

39 Song, M. A. *et al.* Landscape of genome-wide age-related DNA methylation in breast tissue. *Oncotarget* **8**, 114648-114662, doi:10.18632/oncotarget.22754 (2017).

40 Ahrens, M. *et al.* DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. *Cell Metab* **18**, 296-302, doi:10.1016/j.cmet.2013.07.004 (2013).

41 Horvath, S. *et al.* Obesity accelerates epigenetic aging of human liver. *Proc Natl Acad Sci U S A* **111**, 15538-15543, doi:10.1073/pnas.1412759111 (2014).

42 Johnson, K. C., Houseman, E. A., King, J. E. & Christensen, B. C. Normal breast tissue DNA methylation differences at regulatory elements are associated with the cancer risk factor age. *Breast Cancer Res* **19**, doi:10.1186/s13058-017-0873-y (2017).

43 Kukushkina, V. *et al.* DNA methylation changes in endometrium and correlation with gene expression during the transition from pre-receptive to receptive phase. *Sci Rep* **7**, 3916, doi:10.1038/s41598-017-03682-0 (2017).

44 Hong, S. R. *et al.* DNA methylation-based age prediction from saliva: High age predictability by combination of 7 CpG markers. *Forensic Sci Int Genet* **29**, 118-125, doi:10.1016/j.fsigen.2017.04.006 (2017).

45 Gopalan, S. *et al.* Trends in DNA Methylation with Age Replicate Across Diverse Human Populations. *Genetics* **206**, 1659-1674, doi:10.1534/genetics.116.195594 (2017).

46 Price, E. M. *et al.* Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenet Chromatin* **6**, doi: 10.1186/1756-8935-6-4 (2013).

47 Balakumar, B. J., Hastie, T., Friedman, J., Tibshirani, Simon, N. R. *Glmnet for Python*, <http://www.stanford.edu/~hastie/glmnet_python/> (2016).

48 Endelman, J. B. Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome* **4**, 250-255 (2011).

49 A Package for Survival Analysis in S v. 2.38 (2015).

50 Johnson, N. D. *et al.* Non-linear patterns in age-related DNA methylation may reflect CD4(+) T cell differentiation. *Epigenetics-Us* **12**, 492-503, doi:10.1080/15592294.2017.1314419 (2017).

51 Horvath, S. *et al.* An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biology* **17**, doi:10.1186/s13059-016-1030-0 (2016).

52 Horvath, S. *et al.* Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol* **13**, R97, doi:10.1186/gb-2012-13-10-r97 (2012).

53 Liu, Y. *et al.* Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat Biotechnol* **31**, 142-147, doi:10.1038/nbt.2487 (2013).

54 Li, Y. *et al.* An Epigenetic Signature in Peripheral Blood Associated with the Haplotype on 17q21.31, a Risk Factor for Neurodegenerative Tauopathy. *Plos Genet* **10**, doi: 10.1371/journal.pgen.1004211 (2014).

1 Table 1: Description of DNA methylation cohorts

COHORT[1]	SAMPLE SIZE[2]	NUMBER OF SAMPLES WITH VALID AGE	MEAN AGE(SD)	AGE RANGE	SOURCE	DISEASE
LBC1921[32,33]	692	692	82.3 (4.3)	[77.8, 90.6]	blood	Not Available
LBC1936[32,33]	2326	2326	72.4 (2.8)	[67.7, 77.7]	blood	Not Available
BSGS[31]	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[36]	695	600	45.2 (15.0)	[17.0,76.0]	blood	Motor Neuron Disease (MND): 497, Control: 198
GS[34,35]	5101	5100	48.5(14.0)	[18.0,94.5]	blood	Not Available
GSE72775[51]	335	335	70.2 (10.3)	[36.5,90.5]	blood	Not Available
GSE78874[51]	259	259	68.8(9.7)	[36.0,88.0]	saliva	Not Available
GSE72773[51]	310	310	65.6 (13.9)	[35.1,91.9]	blood	Not Available
GSE72777[51]	46	46	14.7 (10.4)	[2.2,35.0]	blood	Not Available
GSE41169[52]	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[53]	689	689	51.9 (11.8)	[18.0,70.0]	blood	Rheumatoid Arthritis:354, Control:335
GSE53740[54]	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

2
3 [1] LBC = Lothian Birth Cohort; BSGS = Brisbane Systems Genomics Study; SGPD = Systems Genomic of Parkinson's Disease
4 consortium; MND = Motor Neuron Disease cohort; GS = Generation Scotland. Cohorts with prefix GSE are from the GEO
5 database.
6 [2] The number of samples in each cohort. Some samples in LBC were measured from the same individual but at different
7 chronological age.

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 counts correction)

	LBC1921 wave one	**LBC1936 wave one**
N	436	906
N_{deaths}	386	214
Chronological Age: mean (SD)[1]	79.1 (0.6)	69.5 (0.8)
Before cell counts correction		
Hannum: mean (SD)	80.3 (6.2)	71.3 (5.7)
Hannum: Hazard Ratio (P-value, 95% CI)[2]	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 (P-value, 95% CI)	1.09 (0.038, 1.00-1.20)	1.19 (0.0022, 1.06-1.32)
Elastic Net: mean (SD)[3]	77.4 (3.6)	72.5 (3.2)
Elastic Net: Hazard Ratio (P-value, 95% CI)	1.08 (0.38, 0.91-1.27)	1.00 (0.96, 0.79-1.28)
BLUP: mean (SD)[3]	77.3 (3.3)	72.5 (2.8)
BLUP: Hazard Ratio (P-value, 95% CI)	1.20 (0.066, 0.99-1.46)	1.25 (0.12, 0.95-1.64)
After cell counts correction		
Hannum: Hazard Ratio (P-value, 95% CI)	1.10 (0.057, 1.00-1.21)	1.11 (0.15, 0.96-1.29)
Horvath: Hazard Ratio (P-value, 95% CI)	1.07 (0.13, 0.98-1.17)	1.14 (0.032, 1.01-1.28)
Elastic Net: Hazard Ratio (P-value, 95% CI)[3]	1.07 (0.39, 0.91-1.27)	1.03 (0.79, 0.82-1.31)
BLUP: Hazard Ratio (P-value, 95% CI)[3]	1.21 (0.05, 1.00-1.48)	1.21 (0.17, 0.92-1.60)

[1] Mean (predicted) age and its standard deviation.
[2] 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.
[3] Both results of Elastic Net and BLUP were based on the age predictor with largest training sample size (sample size = 10,411 for LBC1936 and sample size = 12,710 for LBC1921).

Table 3: Enrichment test on the AAR associated CpG sites from different methods

	Number of significant associations (P < 0.05/319,607)	λ_{median}	Number of CpG sites showing significant cellular heterogeneity	Odds ratio (P-value)[1]
Hannum	12,015	3.6	4,958	3.85 ($P < 2.2\times10^{-16}$)
Horvath	18,847	5.4	5,955	2.53 ($P < 2.2\times10^{-16}$)
Elastic Net[2]	159	2.1	21	0.78 (P = 0.33)
BLUP[2]	793	2.6	130	1.00 (P = 1.0)

[1] The odd ratio for the enrichment of EWAS significant CpG sites in the probe set showing significant cellular heterogeneity.
[2] Both results of Elastic Net and BLUP were based on the age predictor with largest training sample size (sample size = 10,411 for LBC1936 and sample size = 12,710 for LBC1921).

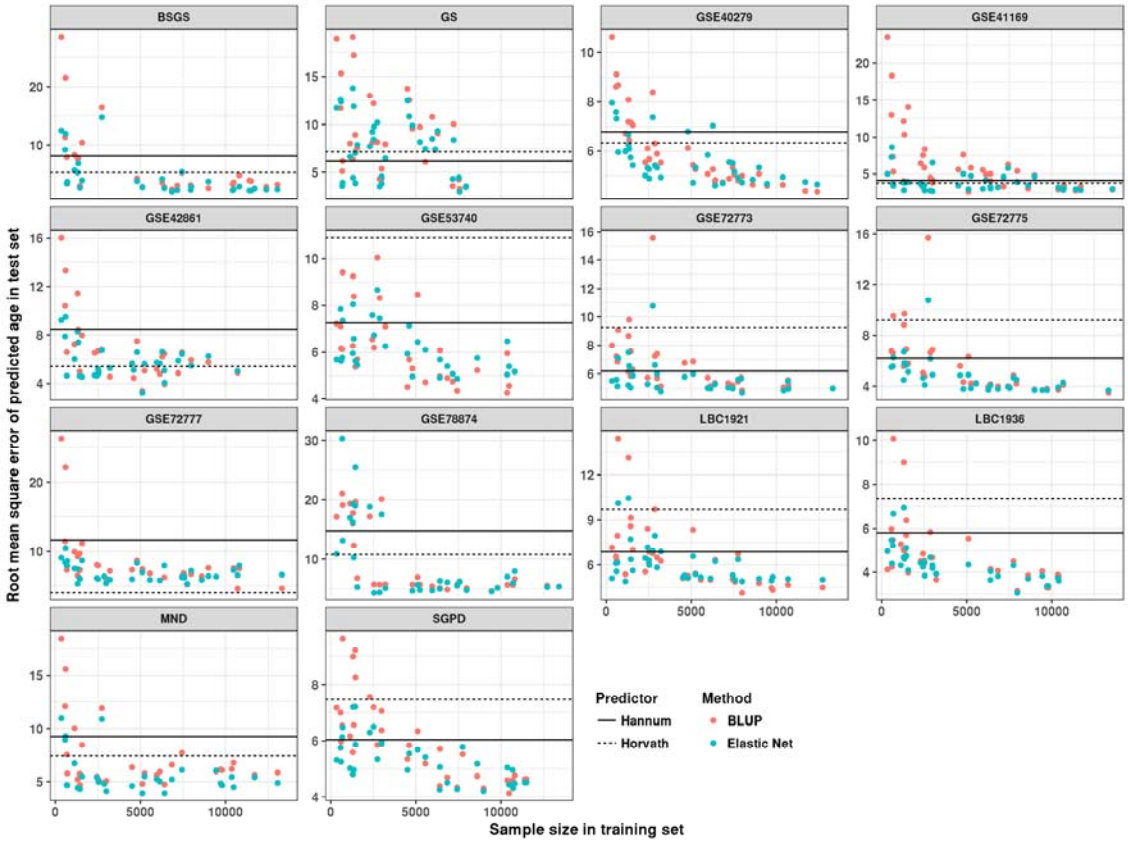

Figure 1: The relationship between training sample size and predictor error measured at the square root of the mean squared error (RMSE) in test data sets. Each point represents the RMSE of the test result based on predictors with different sample size and methods. Prediction results from Horvath are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

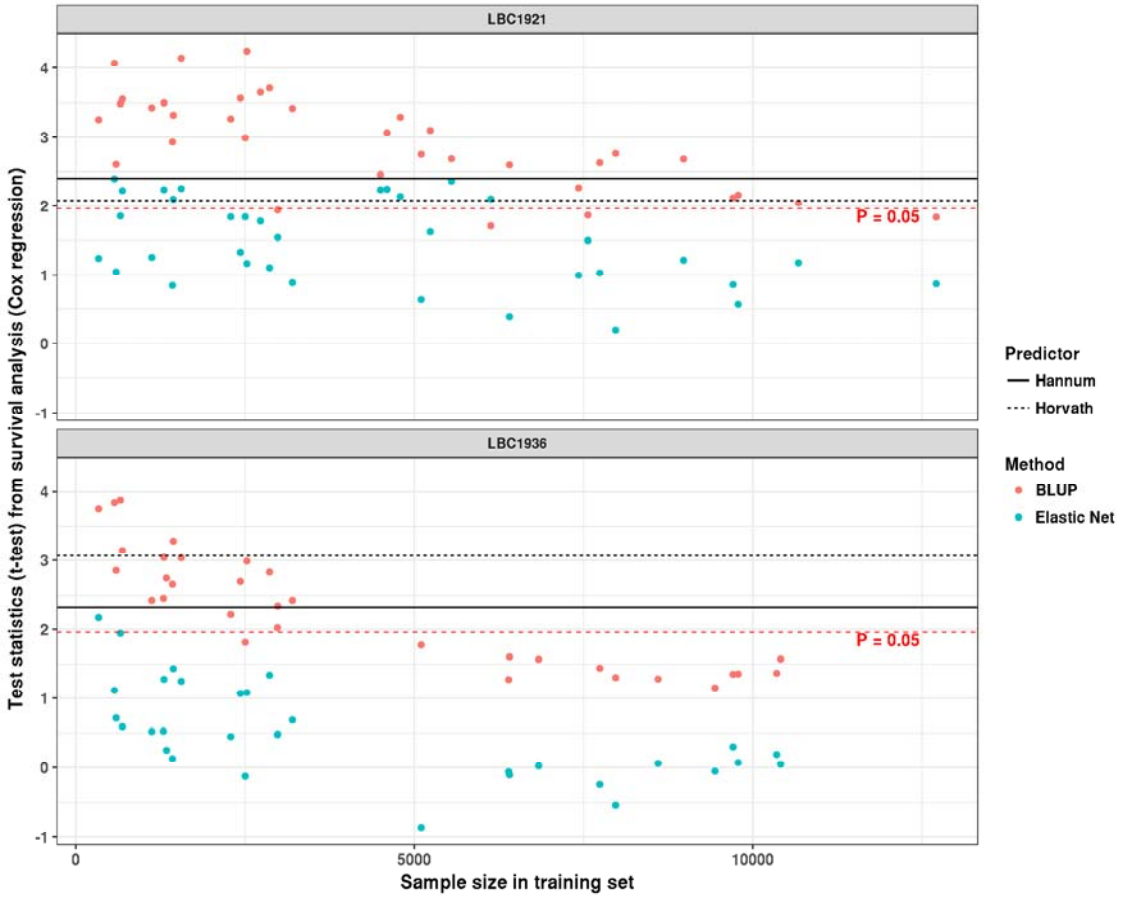

Figure 2: Relationship between the training sample size and the test statistics (t-test) from the association between age acceleration residual (AAR) and mortality. Each point represents the test statistic from the survival analysis based on the predicted ages from predictors with different training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

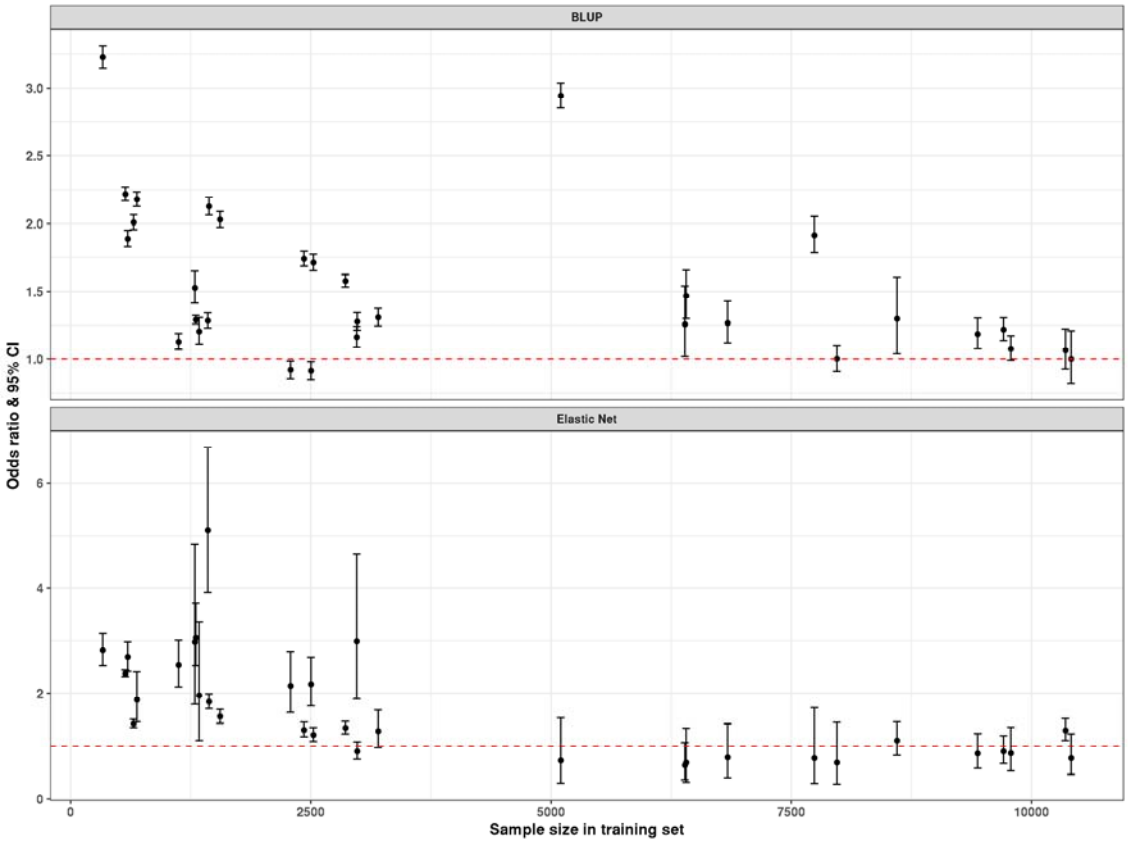

Figure 3: The change of odds ratio from the enrichment test with the increase of training sample size. The enrichment test examines whether AAR associated CpG sites are enriched in probes with cellular heterogeneity.

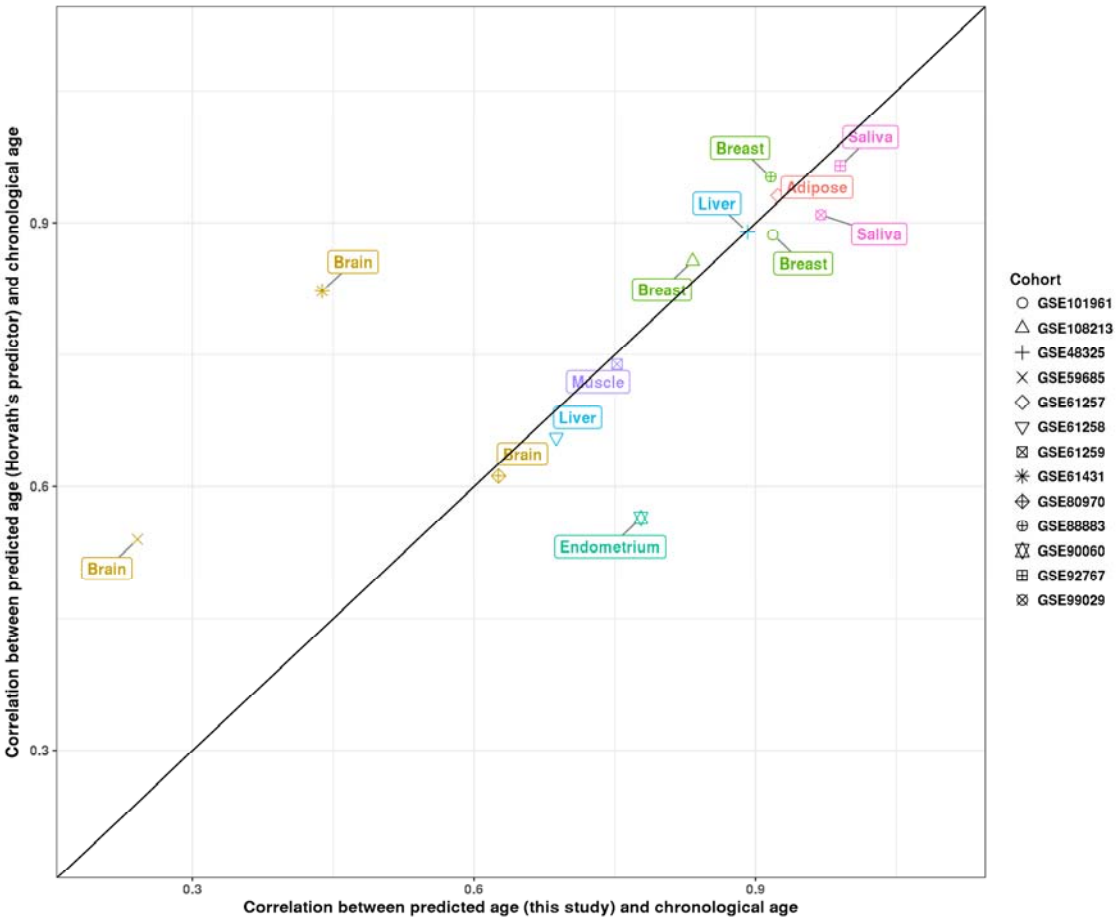

Figure 4: Comparison of prediction performance (correlation) between the predictor from this study (based on Elastic Net) and Horvath's age predictor in non-blood samples.