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 in the age range of 2 to 104 years from 14 cohorts measured on Illumina HumanMethylation450/EPIC arrays to perform prediction analyses using Elastic Net and Best Linear Unbiased Prediction. We show that increasing the sample size achieves a smaller prediction error and higher correlations in test datasets. Our predictors achieved prediction errors of about 4.5 years across cohorts, in contrast to >7 years for the widely-used Horvath and Hannum predictors. 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.

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, defined as the residuals from regressing predicted age on chronological age, have been reported to be associated with mortality, obesity and other complex traits [13-16].

Although the reported prediction accuracy in both Hannum *et al.* and Horvath's age predictors were high, there are three possible ways to improve the prediction of chronological age: (1) increase the size of the training sample; (2) increase the coverage of DNA methylation in the genome, and (3) use better statistical methods. Hannum's age predictor was based upon 482 samples only, but with denser coverage (the Illumina 450K methylation chip) than the Horvath predictor, which was from a larger sample size, at the expense of a lower coverage (the Illumina 27K Methylation microarray). Both of these landmark studies used Elastic Net methodology to build the age predictor.

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. We estimated the effect of sample size and age

range on the prediction accuracy in test data sets. Two approaches were evaluated in the prediction accuracy of age: Elastic Net [17] and Best Linear Unbiased Prediction (BLUP) [18]. The performance of predictors on brain samples was also investigated. Finally, we discuss the implications of our results for the scope and utility of DNA methylation based age predictor as a biomarker for biological ageing.

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 two largest datasets (GS, N = 5,101, SGPD, N = 1,556) 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 [19]. There were 2,586 and 1,299 unrelated individuals in GS and SGPD, respectively (**Material and Methods**). The proportion of variance of age explained by DNA methylation was close to 1 (proportion explained = 1, SE = 0.0036, REML analysis using the software package OSCA (http://cnsgenomics.com/software/osca)) 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).

Effect of training sample size on age prediction

To explore the factors contributing to prediction accuracy in the test cohorts, we performed a cross-validation analysis of the 14 cohorts using a common set of 319,607 probes that survived quality control (**Material and Methods**) in all cohorts. We randomly selected 1 to 13 cohorts as a test set, and used the remaining cohorts as a training set. We repeated this step 65 times to generate different training sets with various sample size and age spectrum. 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 [17] and BLUP [18] were compared. 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.[20]. 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 when the sample size increased (**Supplementary Figure 2**).

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. 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 can contribute to the better prediction accuracy. 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 Tables 1 and 2**).

Consistency of coefficients of probes between different methods

To investigate the consistency of coefficients estimated in the two statistical methods, we performed a pairwise analysis where we estimated the correlation between the coefficient of all probes for two methods. We found that this correlation was significant ($P < 2.2\times10^{-16}$) regardless of the training sample size, and that the larger the training sample size, the higher the correlation (R increases from 0.44 to 0.67, **Supplementary Figure 3**). This suggests the performance of age predictors based on these two methods will become similar with the increase of training sample size. Furthermore, we identified that cg16867657 (*ELOVL2*) was the probe with the largest positive coefficient (0.52 years per

standard deviation of DNA methylation) and cg01620164 (*FIGN*) had the largest negative coefficient (-0.64 years per standard deviation of DNA methylation) for BLUP based age predictors. cg16867657 was also found to be the probe with the second largest coefficient in the Elastic Net based age predictor, but not for cg01620164 (**Supplementary Table 3**). All the above results were from the age predictor based on 13,566 training samples, the coefficient of each probe is listed in **Supplementary Table 4** and the corresponding Manhattan plot is presented in **Supplementary Figure 4.**

Effect of probe sets on age prediction

There is a complex correlation structure in DNA methylation, and the effective number of independent methylation probes was previously reported to be around 200 [21], 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 5**), indicating a loss of information when using fewer methylation probes for prediction. This difference is more stable when the training sample size is large. Hence, despite the small effective number of independent probes, using all available information is better than pruned a set with a lower correlation structure.

We also 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 6**), 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 (**Figure 2**). 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**). This suggests that the probes used for age prediction are not limiting, and BLUP has a better tolerance than Elastic Net in age prediction when some age-associated probes are missing.

Tissue specificity in age prediction

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**). The same trend was observed for the correlation results in brain tissue samples of the GSE59685 (GEO accession ID) cohort (**Materials and Methods**) using BLUP estimators derived from blood and saliva samples (**Figure 3A**), but not for RMSE (**Figure 3B**). RMSE was not improved by increasing the training sample size. It was highly variable for small training samples and then became steady when the training sample size was over 5,000. The large and variable RMSE in the brain samples is

not caused by batch effects since we can observe a small RMSE for the blood samples in the same cohort.

To quantify whether the large RMSE is consistent in the brain samples from other cohorts, we downloaded two additional data sets (GSE61431 and GSE80970) that contain brain samples and compared predicted ages with chronological age in these cohorts. The age predictor based on 13,566 training samples using the BLUP method was applied. We observed a high RMSE and correlation in all cohorts (**Table 2**), indicating the large RMSE is caused by the bias of the age predictors due to tissue-specific DNA methylation. Compared to the results based on the age predictor of Horvath (training samples are from multiple tissues), higher RMSE and comparable correlations were observed in our analysis, suggesting our age predictor could have a similar performance with the age predictor of Horvath in samples from brain tissues when only correlations are considered. The predicted age and chronological age in the same brain regions from three separate cohorts was found to have similar regression slopes (**Supplementary Figure 8**). We also found this similarity between samples from frontal cortex and superior temporal gyrus, indicating the relationship between DNA methylation and chronological age is close between these two brain regions. Similar results were found from the Elastic Net based age predictor (**Supplementary Table 5**, **Supplementary Figure 9**).

Transformed DNA methylation in age prediction

We found a significant correlation between prediction residuals (predicted age - chronological age) and predicted age from BLUP ($R = -0.18$, $P < 2.2\times10^{-16}$, 95% CI [-0.19, -0.16], **Supplementary Figure 10**) and Elastic Net ($R = -0.11$, $P < 2.2\times10^{-16}$, 95% CI [-0.12, -

0.089]) -based age predictors respectively (13,566 training samples). Two sources of bias could explain this significant correlation: (1) violation of the assumption of linearity; (2) bias induced by the regularization step in the prediction methods. To investigate the possible effect of nonlinearity, we applied several data transformation methods to the raw DNA methylation beta value before training and used the transformed data in age prediction within each cohort (**Materials and Methods**). We found that RMSE was reduced by 7.6% in LBC1921 and 6.3% in LBC1936 compared to the result based on raw data (**Supplementary Figure 11**); however, the decrease was substantially lower in other cohorts (around 2%). Additionally, the smallest RMSE was obtained with variable power transformation in each cohort (from 0.2 to 0.7), so that there is not a single power transformation to improve age prediction in all cohorts. Furthermore, we did not observe any improvement using the M values of DNA methylation or using the arcsine square root transformation (**Supplementary Figure 12**).

Prediction accuracy and biological ageing

The difference (Δ_{age}) between predicted age from the Hannum/Horvath predictors and chronological age 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 Δ_{age} 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**). We used age predictors excluding LBC1921/LBC1936 as part of the training set (sample size ranges from 335 to 12,710). We observed a decrease of the test statistics from the survival analysis (t-test) with increasing sample size in training data set (**Figure 4**). No

significant associations between Δ_{age} and mortality was found based on the largest training sample size in either LBC1921 or LBC1936 using BLUP or Elastic Net (**Table 3 and Supplementary Table 6**). Furthermore, we applied this survival analysis based on Δ_{age} adjusting for white blood cell (WBC) counts (basophils, eosinophils, monocytes, lymphocytes, and neutrophils) since WBC counts were found to be associated with DNA methylation [22] and mortality [23] (**Materials and Methods**). A large and positive change of the test statistics after correcting for the WBC counts was observed, especially when the training sample size is small (**Supplementary Figure 13**). These results suggest that the significant associations between Δ_{age} and mortality could be biased due to the existence of confounders like WBC counts, and that improving the prediction accuracy of the age predictor could reduce the effect of these confounders in the survival analysis.

Discussion

We investigated the factors that can affect the accuracy of chronological age prediction by combining around 14,000 samples from 14 cohorts, 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. Prediction performance using estimated coefficients from either Elastic Net or BLUP were similar when the training sample size was over 8,000. Our predictors showed substantially improved prediction accuracy compared to using the estimated coefficients previously reported by Hannum [6] and Horvath [7]. 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 1 and 2**). In addition, the slopes from regressing chronological age on predicted age based on our predictors were closer to 1 than the other two public age predictors in most cohorts (11 out of 14), indicating that our predictors are more unbiased (**Supplementary Figure 14**). 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 4**.

Predicted ages in samples from brain tissues showed high RMSE but strong correlation (**Table 2**). The matched predicted age between three different cohorts in the same brain tissue implies this large RMSE can be induced by the bias of our model since there are only samples from blood and saliva in the training dataset and DNA methylation is tissue specific [24]. Despite this, the high correlation in the prediction results reflects that the changing rate of DNA methylation is correlated across tissues, as reported previously[7]. In addition, although RMSE is much larger, the correlation in our analysis is close to that from Horvath's age predictor, which suggests that our predictor can be used in brain tissues when only relative age is needed. Moreover, we found similar slopes for the regression of predicted age on chronological age for samples from the prefrontal cortex and superior temporal gyrus, indicating a similar association between DNA methylation and age across these two brain regions.

Notwithstanding the highly correlated pattern of DNA methylation across the genome, we observed a decline of prediction accuracy when using the pruned probe set, so that including more probes in the training model is beneficial, especially when the training sample size is small (**Supplementary Figure 7**). 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.

We defined the prediction residual as the difference between the predicted age and the chronological age in the test set. Residuals from the prediction model in the test set can be decomposed into errors caused by bias and stochastic variation [25]. We identified a significant correlation between prediction residuals and predicted age, and attributed this correlation to the bias of the prediction model. We observed a considerable difference of RMSE between the training (RMSE $\approx$ 0) and test set for the BLUP based method, which means BLUP is more prone to "overfitting" than Elastic Net, as expected from the assumption that all probes have an effect **(Supplementary Figure 15)**. Although the prediction error was smaller using Elastic Net, the bias in Elastic Net was higher, and the performance between BLUP and Elastic Net became similar when the training sample size was larger than 8,000.

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. Therefore, the higher the prediction accuracy and smaller the prediction

error (the difference between predictor and actual value), the less opportunity there is for DNA methylation to capture biological ageing. It is clear that DNA methylation measured in blood is associated with environmental exposures such as smoking, sex and BMI [26-28]. 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.

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. [29] and Deary et al. [30,31]. GS is a population and family based cohort recruited through the NHS Scotland general practitioner research network [32,33]. 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. [34]). 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). We also downloaded three cohorts from GEO database with accession ID GSE61431[35], GSE59685[36] and GSE80970 to test the performance of age predictors in brain tissues. GSE59685 contains 80 samples from whole blood and 451 samples from brain samples. These brain samples are from four types of brain regions including 112 samples from the cerebellum, 108 samples from the entorhinal cortex, 114 samples from prefrontal cortex and 117 samples from superior temporal gyrus. GSE61431 contains 44 samples from the cerebellum and 43 samples from the prefrontal cortex. There are 142 samples from the prefrontal cortex and 144 samples from the superior temporal gyrus in GSE80970.

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 [37]. 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[22], 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 16)**. 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.

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 (http://cnsgenomics.com/software/osca).

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[17], 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[38].

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[39] to build the age predictor, and σ_u^2 and σ_e^2 were estimated using the REML analysis implemented in this package.

Transformation of DNA methylation

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 (from 0.1 to 2) transformation was applied to the original Beta value of DNA methylation, and 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.

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 Δ_{age} and mortality with age at sample collection and sex as the covariates. Δ_{age} 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 Δ_{age} were made for WBC counts measured on the same blood samples that were analysed for methylation. Hazard ratios for Δ_{age} were expressed per five years of methylation age acceleration (**Table 3**) and per standard deviation of methylation age acceleration (**Supplementary Table 5**), respectively. Cox models were performed utilizing the 'survival' library[40] in R. Samples from wave one of LBC1921 and LBC1936 were used in this analysis.

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 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).

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

19 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).

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

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

22 Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *Bmc Bioinformatics* **13**, doi:Artn 8610.1186/1471-2105-13-86 (2012).

23 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).

24 Davies, M. N. *et al.* Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biology* **13**, doi:ARTN R4310.1186/gb-2012-13-6-r43 (2012).

25 Geman, S., Bienenstock, E. & Doursat, R. Neural Networks and the Bias Variance Dilemma. *Neural Comput* **4**, 1-58, doi:DOI 10.1162/neco.1992.4.1.1 (1992).

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

27 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).

28 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).

29 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).

30 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).

31 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).

32 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).

33 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).

34 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).

35 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).

36 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).

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

38 Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).

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

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

41 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).

42 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).

43 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).

44 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).

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[30,31]	692	692	82.3 (4.3)	[77.8, 90.6]	blood	Not Available
LBC1936[30,31]	2326	2326	72.4 (2.8)	[67.7, 77.7]	blood	Not Available
BSGS[29]	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[34]	695	600	45.2 (15.0)	[17.0,76.0]	blood	Motor Neuron Disease (MND): 497, Control: 198
GS[32,33]	5101	5100	48.5(14.0)	[18.0,94.5]	blood	Not Available
GSE72775[41]	335	335	70.2 (10.3)	[36.5,90.5]	blood	Not Available
GSE78874[41]	259	259	68.8(9.7)	[36.0,88.0]	saliva	Not Available
GSE72773[41]	310	310	65.6 (13.9)	[35.1,91.9]	blood	Not Available
GSE72777[41]	46	46	14.7 (10.4)	[2.2,35.0]	blood	Not Available
GSE41169[42]	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[43]	689	689	51.9 (11.8)	[18.0,70.0]	blood	Rheumatoid Arthritis:354, Control:335
GSE53740[44]	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

[1] 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.

[2] The number of samples in each cohort. Some samples in LBC were measured from the same individual but at different chronological age.

Table 2: Prediction results on the samples from four different brain regions in three different cohorts. Age predictor is trained based on 13,566 samples and using the BLUP method.

	GSE59685			GSE61431			GSE80970		
Brain Region	**RMSE**[1]	**R (SE)**[2]	**Horvath RMSE/R**[3]	**RMSE**	**R (SE)**	**Horvath RMSE/R**	**RMSE**	**R (SE)**	**Horvath RMSE/R**
entorhinal cortex	28.0	0.62 (0.08)	8.4/0.66	NA[4]	NA	NA	NA	NA	NA
prefrontal cortex	31.2	0.74 (0.06)	10.4/0.74	23.7	0.91 (0.07)	13.0/0.91	35.0	0.60 (0.07)	16.3/0.63
superior temporal gyrus	31.6	0.76 (0.06)	9.3/0.81	NA	NA	NA	35.2	0.68 (0.06)	17.2/0.60
cerebellum	33.7	0.62 (0.07)	24.0/0.77	22.1	0.78 (0.1)	19.4/0.78	NA	NA	NA

[1] Root Mean Square Error of the predicted age in the test set.
[2] Correlation between predicted age and chronological age in the test set.
[3] Root Mean Square Error of the predicted age and the correlation between predicted age and chronological age based on Horvath's Age predictor in the test set.
[4] NA means there is no DNA methylation in this cohort for the brain region.

Table 3: Summary details of two LBC cohorts and the relationship between all-cause mortality and predicted age from different methods

	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)
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)

[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).

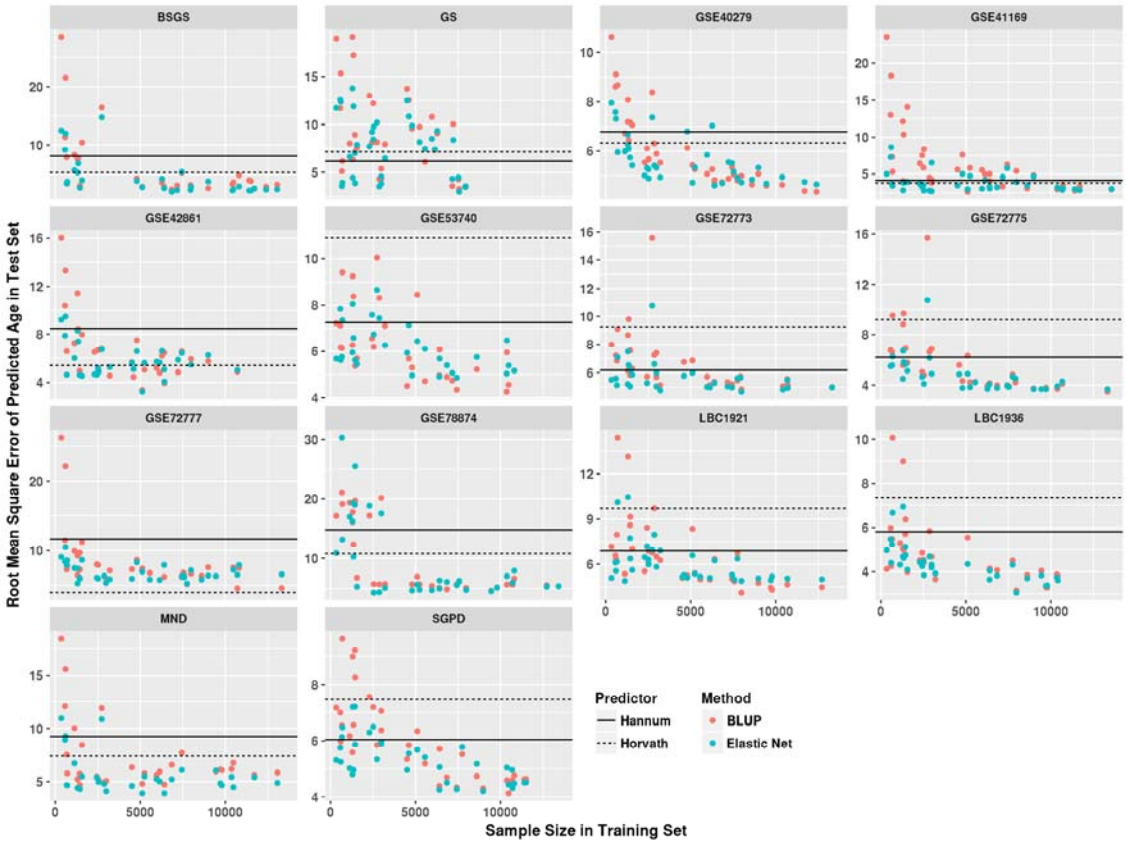

Figure 1: RMSE in test data sets. Each point represents the RMSE of the test result based on predictors with different sample size and methods. Prediction results from Horvath are marked as black dash line, and black solid line represents prediction result from Hannum's age predictor.

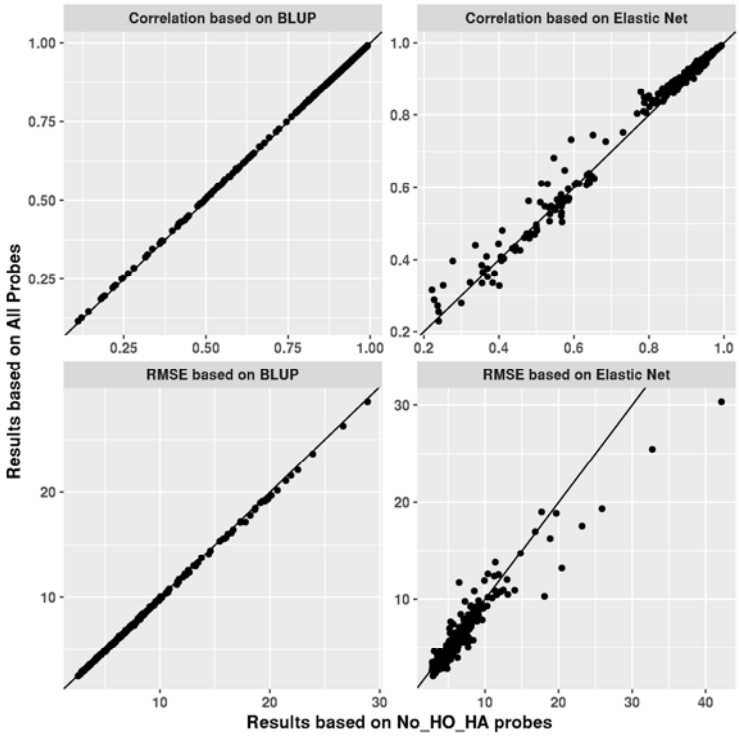

Figure 2: Comparison of prediction performance between the full probe set and probe set excluding probes from Hannum's and Horvath's Age predictors, using BLUP and Elastic Net. Each point represents prediction accuracy of each cohort based on predictors built from different training data sets.

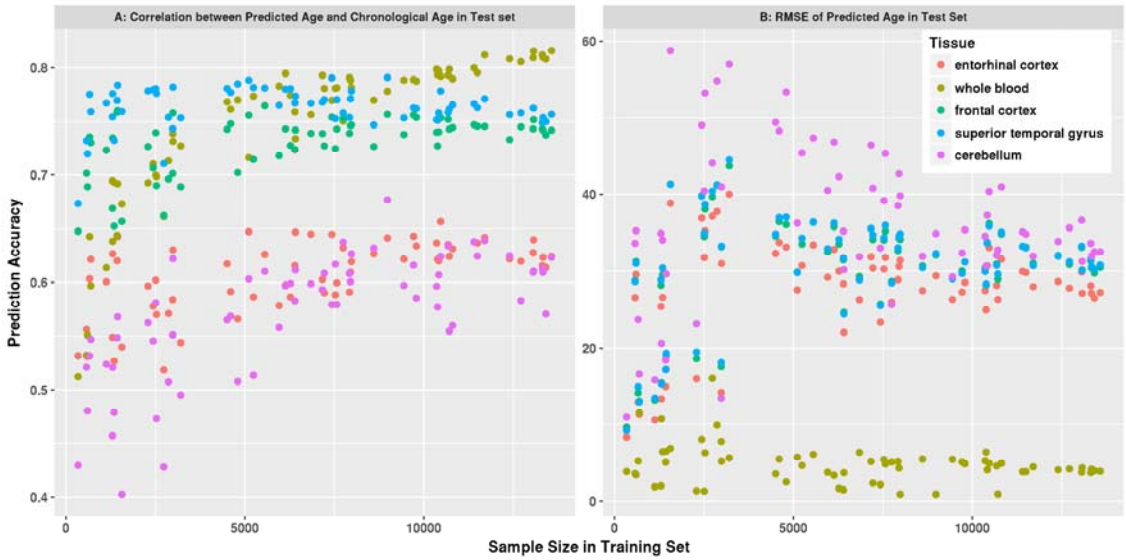

Figure 3: A: Correlation between predicted age and chronological age for the samples in cohort GSE59685, which contains brain and blood samples from the same individuals. Each point represents the correlation between the predicted age and chronological age based on age predictors with different training samples. Each colour represents one cell type. B: RMSE of predicted age for the samples in cohort GSE59685 that contains brain and blood samples from the same individuals.

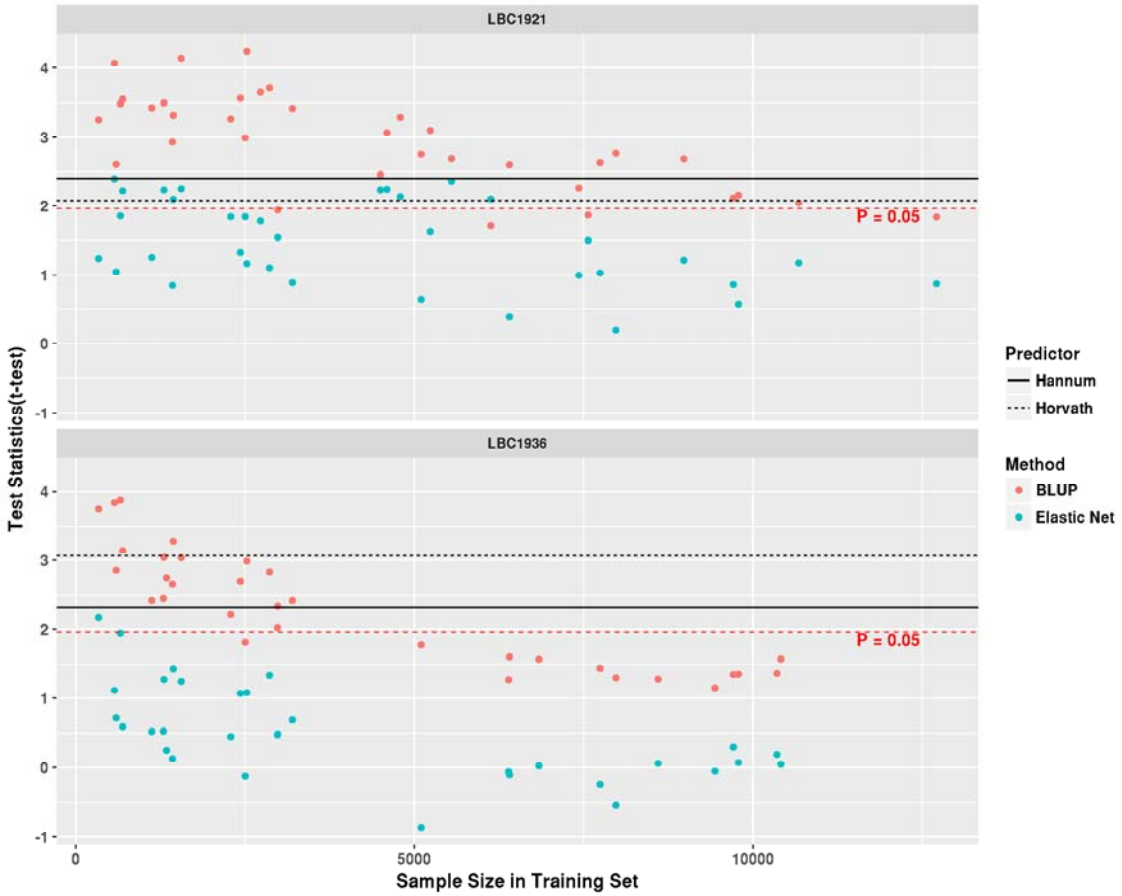

Figure 4: Relationship between the training sample size and the test statistics from the association between Δ_{age} and mortality. Each point represents the test statistic from the survival analysis based on the predicted ages from predictors with different training sample sizes.