

ARTICLE OPEN



A pan-tissue DNA-methylation epigenetic clock based on deep learning

Lucas Paulo de Lima Camillo¹✉, Louis R. Lapierre² and Ritambhara Singh^{1,3}✉

Several age predictors based on DNA methylation, dubbed epigenetic clocks, have been created in recent years, with the vast majority based on regularized linear regression. This study explores the improvement in the performance and interpretation of epigenetic clocks using deep learning. First, we gathered 142 publicly available data sets from several human tissues to develop AltumAge, a neural network framework that is a highly accurate and precise age predictor. Compared to ElasticNet, AltumAge performs better for within-data set and cross-data set age prediction, being particularly more generalizable in older ages and new tissue types. We then used deep learning interpretation methods to learn which methylation sites contributed to the final model predictions. We observe that while most important CpG sites are linearly related to age, some highly-interacting CpG sites can influence the relevance of such relationships. Using chromatin annotations, we show that the CpG sites with the highest contribution to the model predictions were related to gene regulatory regions in the genome, including proximity to CTCF binding sites. We also found age-related KEGG pathways for genes containing these CpG sites. Lastly, we performed downstream analyses of AltumAge to explore its applicability and compare its age acceleration with Horvath's 2013 model. We show that our neural network approach predicts higher age acceleration for tumors, for cells that exhibit age-related changes in vitro, such as immune and mitochondrial dysfunction, and for samples from patients with multiple sclerosis, type 2 diabetes, and HIV, among other conditions. Altogether, our neural network approach provides significant improvement and flexibility compared to current epigenetic clocks for both performance and model interpretability.

npj Aging (2022)8:4; <https://doi.org/10.1038/s41514-022-00085-y>

INTRODUCTION

One of the leading challenges in the field of aging research is measuring age accurately. Monitoring healthy individuals for decades to assess whether an intervention affects the aging process is prohibitive in terms of time and funding. The creation of epigenetic clocks, age predictors that use DNA methylation data, has given researchers tools to measure the aging process quantitatively. Moreover, recent works have demonstrated the effectiveness of precise epigenetic editing based on CRISPR with targeted DNA methylation or demethylation¹. Consequently, epigenetic clocks have the potential of not only measuring aging but also guiding epigenetic interventions.

Notably, two of the most well-known predictors are the ones developed by Hannum et al.² and Horvath³ in 2013. Hannum et al. developed a blood-based epigenetic clock with 71 CpG sites². Then Horvath showed epigenetic clocks could also accurately predict age across tissues, developing a predictor with 353 CpG sites³. Horvath's model has been widely used as it is seen as the state-of-the-art pan-tissue epigenetic clock for humans^{4–7}. Both of these works used simple regularized linear regression (ElasticNet) for feature selection and prediction⁸. More recent epigenetic clocks that predict mortality also use a linear combination of features^{9,10}. ElasticNet has been widely used to develop epigenetic clocks^{2,3,9–13}. Nevertheless, simple linear regression can display high bias and cannot capture non-linear feature-feature interactions in the data.

Interactions among variables can be taken into account by expanding the feature space with feature multiplication. However, incorporating pairwise CpG site interactions is unfeasible given

the high dimensionality of the DNA methylation data. Horvath's model³ selected 353 CpG sites out of total 21,368 sites. If the linear regression had taken into account all pairwise interactions, the feature space would grow to over 228 million. A large number of features is especially challenging due to the relatively low number of publicly available DNA methylation samples. Given the complexity of the epigenetic regulatory network, it is likely that important interactions among CpG sites are not captured in the current epigenetic clocks developed thus far.

Deep learning models have been successfully applied to transcriptomic and clinical blood biomarker data for age prediction^{14,15}. For DNA methylation data, Galkin et al. recently showed that a deep neural network model, DeepMAge¹⁶, gave slightly better prediction performance than Horvath's model in blood samples. However, the authors compared Horvath's pan-tissue predictor to a model trained only in blood DNA methylation data. Moreover, there was no in-depth exploration of why their deep learning model outperformed the ElasticNet model. Similarly, Levy et al.¹⁷ developed a deep learning framework to work with DNA methylation data that encodes the CpG sites into latent features for downstream analysis. They showed encouraging results for age prediction using a multi-layer perceptron; however, they investigated only one data set obtained from white blood cells. Therefore, currently, our understanding of the advantages of neural networks for this task in a pan-tissue setting is limited.

We introduce AltumAge, a deep neural network that uses beta values from 20,318 CpG sites common to the Illumina 27k, 450k and EPIC arrays for pan-tissue age prediction (summarized in Fig. 1a). We hypothesized that a neural network using all available

¹Department of Computer Science, Brown University, Providence, RI, USA. ²Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI, USA. ³Center for Computational Molecular Biology, Brown University, Providence, RI, USA. ✉email: lucas_camillo@alumni.brown.edu; ritambhara@brown.edu

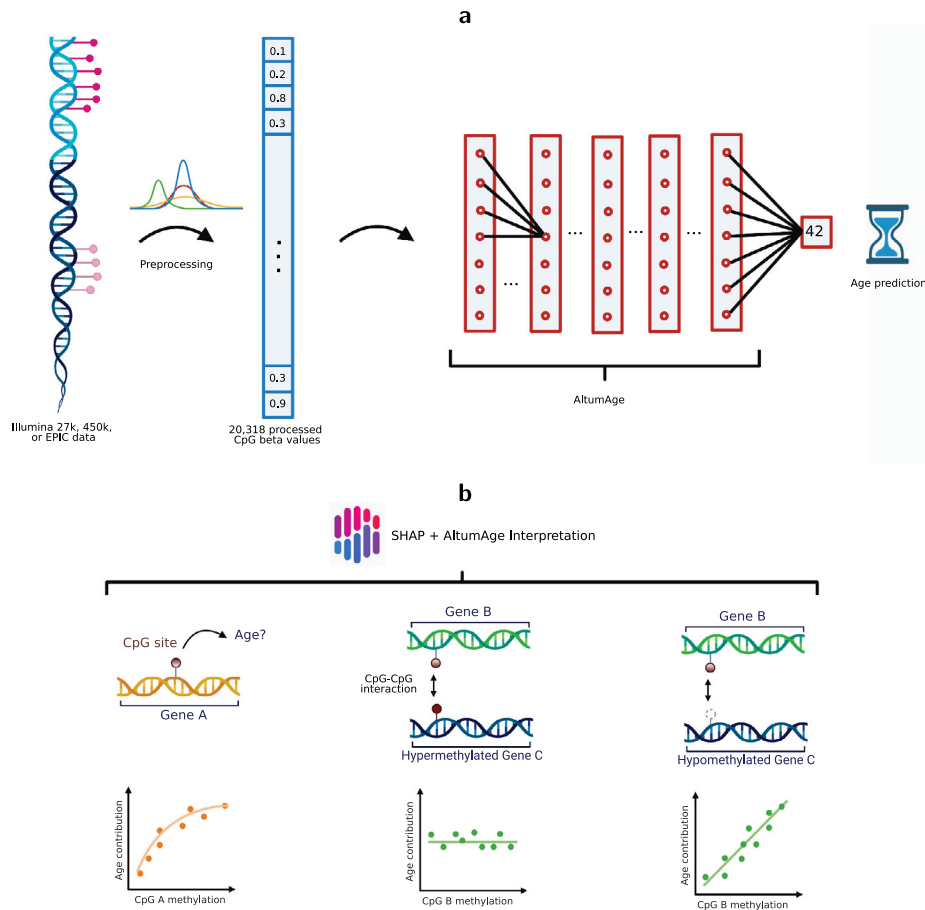


Fig. 1 | AltumAge model and interpretation. **a** DNA methylation data from Illumina 27k, 450k, or EPIC arrays are normalized with BMIQ and scaled. Then 20,318 CpG sites are selected as the input of the model. The information is processed through five hidden layer nodes with 32 nodes each, and the values of the last hidden layer nodes are combined into a single node as the age output in years. **b** For interpretation, a Shapley-values-based method, called SHAP¹⁸, is used to determine how the methylation status of a specific CpG site affects the age output of AltumAge. Relevant CpG sites generally present a primarily linear relationship (left) with the predicted age. However, interacting CpG sites can change such relationships. In some instances, we find that when a secondary CpG site is hypermethylated (middle), the methylation status of the first CpG is irrelevant for age prediction; when it is hypomethylated (right), then the methylation status becomes essential. Images created with Biorender.com.

CpG sites would be better suited to predict pan-tissue age using DNA methylation data due to their ability to (1) capture higher-order feature interactions and (2) leverage important information contained in the thousands of CpG sites not selected by ElasticNet models. AltumAge uses multi-layer perceptron layers (similar to^{16,17}) that account for non-linear interactions by combining multiple features into each node of the network. We trained AltumAge on samples from 142 different experiments, which, to our knowledge, is the largest compilation of DNA methylation data sets for human age prediction. The publicly available data were obtained from multiple studies that used Illumina 27k and Illumina 450k arrays. The code for the model can be found in our GitHub repository (<https://github.com/rsinghlab/AltumAge>).

We show that AltumAge has a significantly lower error for within-data set age prediction, is better able to generalize to new tissue types and older ages for cross-data set settings, and is more resistant to noise than ElasticNet. For inference, we apply the Shapley-value-based interpretation method, called SHAP¹⁸, on AltumAge to determine the contributions of different CpG sites towards age prediction (summarized in Fig. 1a). We confirm that the most important CpG sites have complex interactions resulting in non-linear relationships when predicting age. Such interactions may lead to mechanistic hypotheses on how the epigenetic network interacts to drive the aging phenotype. Additionally, we

find that the most important CpG sites are proximal to CTCF binding sites. However, CpG sites in known age-related pathways (SIRT, mTOR, and AMPK) do not seem relevant for age prediction. Finally, our downstream analysis reveals that AltumAge predicts higher age for tumors, cells with cellular immune dysfunction, cells with mitochondrial dysfunction, cells with high passage numbers, and samples with multiple sclerosis, type 2 diabetes, HIV, among others, than controls demonstrating its usefulness for age-related studies. Overall, we show that deep learning can improve both the performance and interpretation over the widely used ElasticNet-based models and present AltumAge as a useful tool for age prediction.

RESULTS

Performance Evaluation

Model selection. Neural networks can capture complex variable interactions when provided with a large number of high-dimensional data sets. We hypothesized that the same would be true for age prediction with DNA methylation data.

For each of 142 data sets, we split the total samples—60% for training and validation ($n = 8050$) and 40% for testing ($n = 5455$)—to avoid introducing any bias in the age, gender, and tissue type distributions. In the training and validation set, the data was

Table 1. Evaluation metrics of AltumAge and different linear models in the test set for within-data set age prediction.

Model	CpGs	MAE	MSE	R	Median Error
AltumAge	20318	2.153	29.486	0.980	-0.275
AltumAge with ElasticNet CpGs	903	2.302	30.455	0.979	-0.043
AltumAge with Horvath's CpGs	353	2.425	32.732	0.978	-0.271
ElasticNet	903	2.621	39.198	0.973	-0.006
Linear Regression with Horvath's CpGs	353	3.011	46.867	0.968	-0.003
Horvath's model	353	3.530	71.031	0.952	0.128

The median absolute error (MAE) and the median error are in units of year, while the mean squared error (MSE) is in units of year-squared. Also shown is the number of CpGs whose model importance is greater than zero. R stands for Pearson's correlation coefficient. Bold numbers indicate better performance.

further subdivided by data set, with 85 ($n = 4394$) for model selection and 57 ($n = 3656$) for validation. Supplementary Figure 1a shows a schematic of the division of the data, with the full list of data sets used in Supplementary Note 1. Several machine learning models were trained and validated (Supplementary Table 1, Supplementary Figure 2a) to pick the best performing models according to the mean squared error (MSE). We also included other evaluation metrics such as the median absolute error (MAE) and Pearson's correlation coefficient (R). Our primary baseline method is the ElasticNet model, a linear model with implementation following Horvath³. We also included two traditional machine learning methods that capture non-linear relationships in the data, random forest and support vector regression; however, they performed poorly on the validation set (MAE = 6.206 and 10.903, respectively). In addition, we tried the neural network TabNet, an attentive interpretable tabular learning method¹⁹. Although the MSE was marginally better than the ElasticNet baseline (MSE = 62.113 vs. 67.981), the MAE was slightly worse (MAE = 4.027 vs. 3.708). Lastly, we tuned the hyperparameters of the neural networks based on recent findings that highly regularized deep learning methods excel in tabular data prediction²⁰. The best neural network model based on both the MAE and MSE was dubbed AltumAge (from the Latin *altum*, meaning "deep"). For clarity, a regularized linear regression model with Horvath's log-linear age transformation, trained on our 142 data sets, and using the built-in hyperparameter tuning from the Python `glmnet` will be referred to as ElasticNet. On the other hand, the application of Horvath's original 2013 epigenetic clock, originally trained on 39 data sets in that paper³, will be referred to as Horvath's model.

AltumAge outperforms linear models for within-data set age prediction. Differences in performance among epigenetic clocks can generally be explained by three factors: the DNA methylation data, the model, and the input CpG sites (or the features). For within-data set age prediction, all of the training and validation set ($n = 8050$) was used to train all of the models while the model and input CpG sites were varied. Supplementary Figure 1b displays a simple schematic of the within-data set age prediction. Main results are presented in Table 1 (detailed in Supplementary Table 3), with a qualitative comparison in Supplementary Figure 2b.

We hypothesized that our large and diverse DNA methylation data might improve performance compared to other epigenetic clocks irrespective of model type, adding a confounding variable to any performance improvement seen with AltumAge. To understand the magnitude of such effect, we compared a replication of Horvath's model as seen in³ with a linear regression trained on our

143 data sets using the same set of 353 CpG sites. Indeed, the regression trained with our data has a lower error (MAE = 3.011 vs. 3.530; MSE = 46.867 vs. 71.031). ElasticNet, with its selected 903 CpG sites trained with our data, further improves the performance (MAE = 2.621, MSE = 39.198). This result shows that a larger training data set helps the age prediction performance.

Next, we aimed to determine whether the model type, i.e., a linear regression vs. a neural network, would significantly impact the performance. We, therefore, compared the aforementioned linear models with the neural network AltumAge using the same set of features. AltumAge outperformed the respective linear model with Horvath's 353 CpG sites (MAE = 2.425 vs. 3.011, MSE = 32.732 vs. 46.867) and ElasticNet-selected 903 CpG sites (MAE = 2.302 vs. 2.621, MSE = 30.455 vs. 39.198). This result shows that AltumAge outperforms linear models given the same training data and set of features.

Lastly, to compare the effect of the different sets of CpG sites, we trained AltumAge with all 20,318 CpG sites available and compared the results from the smaller sets of CpG sites obtained above. There is a gradual improvement in performance for AltumAge by expanding the feature set from Horvath's 353 sites (MAE = 2.425, MSE = 32.732) to 903 ElasticNet-selected CpG sites (MAE = 2.302, MSE = 30.455) to all 20,318 CpG sites (MAE = 2.153, MSE = 29.486). This result suggests that the expanded feature set helps improve the performance, likely because relevant information in the epigenome is not entirely captured by the CpG sites selected by an ElasticNet model.

Overall, these results indicate that even though more data samples lower the prediction error, AltumAge's performance improvement is greater than the increased data effect. Indeed, the lower error of AltumAge when compared to the ElasticNet is robust to other data splits (Alpaydin's Combined 5x2cv F test p -value = $9.71e-5$).

A direct comparison of AltumAge and Horvath's model reveals that AltumAge has fewer tissue types with a high MAE. In his 2013 paper, Horvath noticed poor calibration of his model in breast, uterine endometrium, dermal fibroblasts, skeletal muscle, and heart³. In our test data, a similarly poor predictive power was found for these tissue types for Horvath's model (breast MAE = 9.462; uterus MAE = 5.798; fibroblast MAE = 10.863; muscle MAE = 9.047; heart not included). AltumAge, on the other hand, had much lower errors for them (MAE = 3.681, 3.660, 4.089, 2.540 respectively). Furthermore, Horvath's model had an MAE > 10 years in 22 tissue types in the test data. AltumAge, on the other hand, had an MAE > 10 in only three tissue types.

Supplementary Figure 3 in particular shows how AltumAge, in contrast to Horvath's model, does not underestimate older ages (>60 years) to such an extent (median error = -2.808 vs. -4.677). Better performance in older age is fundamental in defining biomarkers of age-related diseases of which age is the biggest risk factor. Horvath's model tends to underestimate such population partly due to CpG saturation (beta value approaching 0 or 1 in certain genomic loci)²¹. Another reason might be the assumption that age-related CpG changes are linearly correlated with age after 20 years of age. AltumAge resolves these two problems by incorporating an expanded feature set and not using any pre-defined age transformation function that might inject bias in the data processing.

Of note, we were unable to compare AltumAge with DeepMAge¹⁶, another deep learning framework. Unfortunately, neither the code for DeepMAge nor a complete description of its architecture is available.

AltumAge is more generalizable than ElasticNet in older ages and in non-blood tissue types. Leave-one-data-set-out cross-validation (LOOCV) provides a way to understand the generalization

Table 2. Leave-one-data-set-out cross validation evaluation metrics for AltumAge and ElasticNet.

Model	MAE	MSE	R	Median Error
AltumAge	3.194	63.022	0.958	-0.179
ElasticNet	3.204	67.317	0.955	-0.090

The median absolute error (MAE) and the median error are in units of year, while the mean squared error (MSE) is in units of year-squared. R stands for Pearson's correlation coefficient. Bold numbers indicate better performance.

potential of a model to new unseen data sets. We performed this LOOCV analysis by leaving out the training samples of each data set (out of the 143) during model fitting. Therefore, the model training was performed using the training set of 142 data sets. Next, we evaluated the performance of this model on the test set of the left-out data set. Consequently, we trained 142 different models in total to evaluate the LOOCV performance for all 142 data sets (Table 2, Supplementary Figure 2c). Supplementary Figure 1c displays a simple schematic of data flow for the LOOCV age prediction.

Since AltumAge uses 20,318 CpG sites, we expected it to be more prone to noise and overfitting than a model with low variance such as ElasticNet, which effectively uses only a subset of CpG sites. Nevertheless, we see that AltumAge performs better than ElasticNet in MSE (63.022 vs. 67.317, Wilcoxon signed-rank test $p = 0.0018$) and has a slightly lower MAE (MAE = 3.194 vs. 3.204, Wilcoxon signed-rank test $p = 0.04$).

Next, we analyzed the results stratified by absolute error, age, data set, and tissue type. Fig. 2a shows a scatter plot of the LOOCV absolute error for each sample according to AltumAge and ElasticNet. Points above the black line favor ElasticNet while the opposite favors AltumAge. As shown by the 100-window rolling mean line, for samples with an absolute prediction error > 3.457 years, on average, AltumAge performs better. This observation is particularly apparent for large deviations. This result indicates, alongside the lower MSE, that AltumAge is more resistant to outliers than ElasticNet when generalizing to new samples.

Stratifying the results by age can give insights into particular strengths and weaknesses of each model. For example, while both models capture the age-related epigenetic drift given the correlation between absolute error and age (AltumAge Pearson's $R = 0.443$; ElasticNet Pearson's $R = 0.459$), AltumAge performs better on average for samples with age > 59 years (Fig. 2b). This result suggests that AltumAge better captures epigenetic changes during aging while ElasticNet better understands the developmental epigenome, since epigenetic changes during childhood and puberty are related to development but after they are mostly due to aging²².

Finally, we analyzed the performance of each model by data set and tissue type. As shown in Fig. 2c, AltumAge performs better than ElasticNet in the LOOCV MAE for data sets with a higher error in performance, i.e., that are more challenging to predict. In comparison, ElasticNet is superior for the ones with a lower error. We observe that most data sets with a low MAE are from newborns or blood samples, and the training set is skewed towards blood-based samples (see Supplementary Figure 4). Therefore, we hypothesized that ElasticNet may be simply performing better for overrepresented tissue types in the training set. To check this, we looked at the tissue types from data sets for which ElasticNet had a 50% worse LOOCV MAE than AltumAge (capturing a large deviation) and vice versa. As expected, ElasticNet does not generalize as well to a large variety of tissue types (Fig. 2d). At the same time, it performs better in blood-based samples (Fig. 2e). These observations imply that AltumAge can better generalize to more tissue types, likely capturing global age-

related epigenetic patterns, while ElasticNet could be focusing primarily on blood changes.

As each model has its benefits and drawbacks, we checked the performance of an ensemble of both methods. Interestingly, we observe a substantial decrease in both MAE (2.934) and MSE (58.986) by averaging the predictions of both models. These results indicate that combining deep learning and linear model predictions may further improve the age prediction performance.

AltumAge is more robust to noise than ElasticNet. Another desired property of epigenetic clocks is reliability. Noise derived from the experimental procedure, biological or technical replicates may negatively influence the model's reliability. AltumAge was trained with Gaussian noise and adversarial regularization to be more robust against random variation²³. Gaussian noise introduces normally distributed fluctuations in between hidden layers. Adversarial regularization includes artificial observations with subtle modifications in the loss function that attempt to fool the model into increasing the error. To assess the robustness of AltumAge and ElasticNet to noise, we gradually added artificial Gaussian noise in the beta value of each CpG site up to one standard deviation in the within-data test set and tracked MAE (Fig. 3a) and MSE (Fig. 3b). As expected, the error grows much faster in the ElasticNet model, particularly with the MSE, which is more swayed by outliers.

Furthermore, we examined an independent whole blood data set GSE55763 (not used in training or testing), which contains 2 technical replicates for each of its 36 samples. Ideally, the difference in prediction between the replicates would be zero. As shown in the histogram in Fig. 3c, the median absolute difference for AltumAge is 1.537 years, whereas for ElasticNet, 1.719 years, while the maximum absolute difference is 4.756 and 5.862 years, respectively. Despite no significant difference in distributions—likely due to the small sample size—the models differ in whether they capture an artifact effect between replicates. As anticipated, we do not observe a statistically significant effect from replicate one to two for AltumAge (linear mixed-effects $p = 0.720$). However, we see that ElasticNet predicted a higher age of 1.130 years for replicate two (linear mixed-effects $p = 0.002$). Overall, the results highlight that resistance to random noise may help in real-world scenarios, increasing model robustness and reliability.

Inference

Neural networks, particularly in the context of deep learning, used to be seen as “black-box” methods, as their interpretability was difficult. Regardless of the predictive power of ElasticNet models, they are easily understandable. Recently, various methods have been proposed to extract the contribution of features towards prediction in neural networks. They include interpretation based on model gradients^{24–26}, attention²⁷, among others. One such inference method is SHAP¹⁸, which uses a game-theoretic approach to aid in the explanation of machine learning methods. It can measure how one feature contributes to the output of deep neural networks. For our case, the SHAP value can be conceived as how much the value of one CpG site affects the age output of the model in years. Through the architecture of neural networks, it can also determine which CpG sites most highly interact with each other.

We present results for model inference using SHAP that assist with understanding AltumAge. To support the results obtained by SHAP, we also applied another method of determining feature importance called DeepPINK²⁸ (see Supplementary Information).

AltumAge captures relevant age-related CpG-CpG interactions. Epigenetic modifications can significantly influence gene expression. They can also impact genes that affect other epigenetic changes.

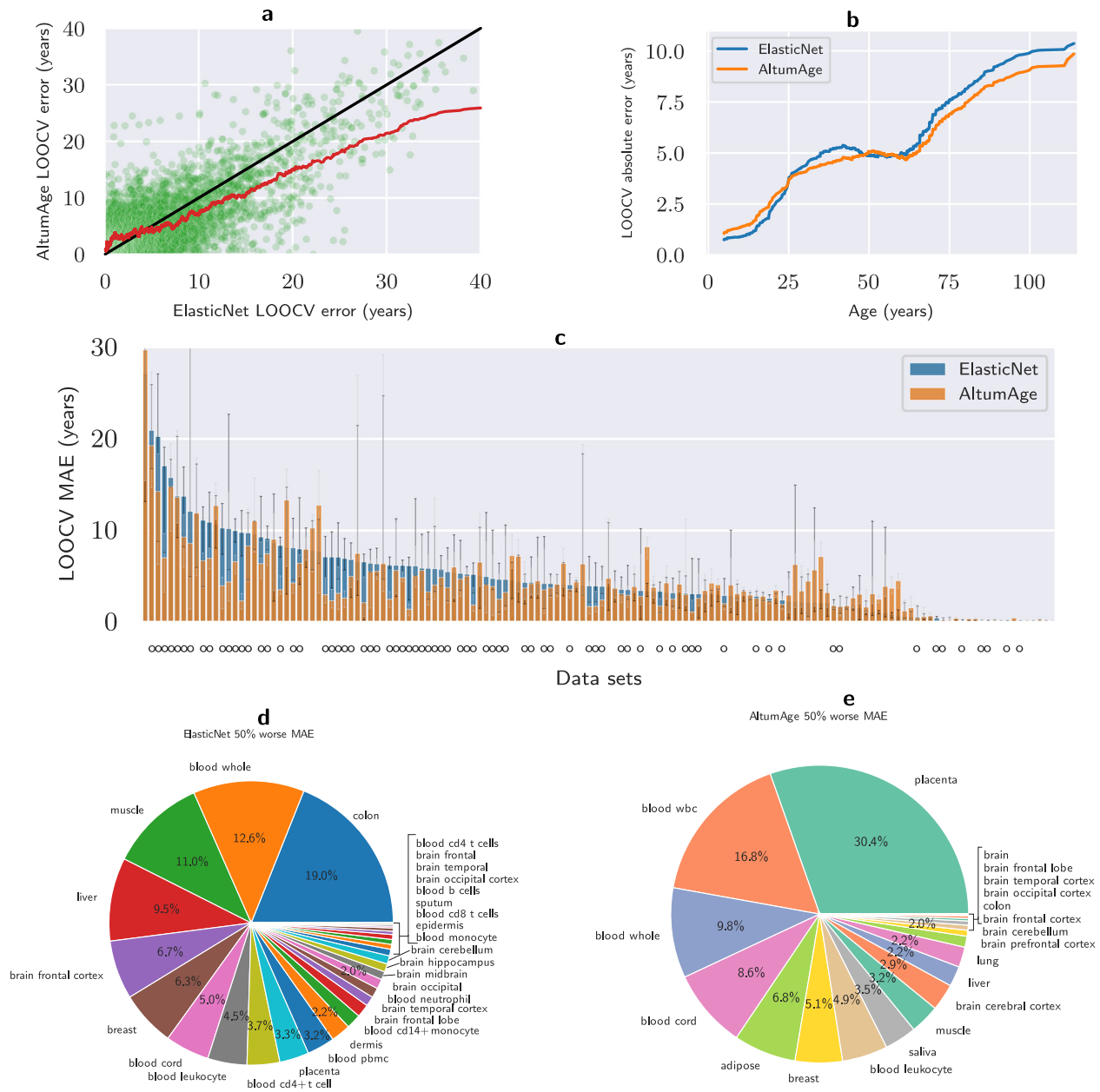


Fig. 2 Comparison of leave-one-data-set-out cross-validation (LOOCV) performance between AltumAge and ElasticNet. **a** Scatter plot contrasting the LOOCV absolute error of each model by sample. The black line separates the region in the graph in which AltumAge performs better (bottom right) versus where ElasticNet is superior (top left), and the red line is a 100-sample rolling mean. AltumAge outperforms ElasticNet, particularly in difficult-to-predict tissue types. **b** The 1000-sample rolling mean of the LOOCV absolute error of each model by age. AltumAge has a lower absolute error for age > 59 years on average. **c** Bar plot showing the LOOCV median absolute error (MAE) by data set for each model, with 95% confidence interval error bars calculated from 1000 bootstrap iterations. A circle below a bar represents data sets for which AltumAge had a lower LOOCV MAE than ElasticNet. **d** Pie plot showing which tissue types from data sets for which ElasticNet had at least a 50% worse MAE than AltumAge. (e) Pie plot with the converse. Overall, AltumAge can better generalize to more tissue types, whereas most of the improved ElasticNet performance comes from blood-based tissues.

Therefore, some CpG sites interact with others through the gene expression network and can work in tandem. Through SHAP, we show that AltumAge can measure how hyper- or hypomethylation of secondary CpG sites affects the relationship of a CpG of interest and age. Supplementary Figure 5 shows scatter plots of the nine most important CpG sites based on SHAP-based importance values assigned to the CpG sites of the samples in the test set. These nine CpG sites account for 1.83% of the total model importance (Supplementary Figure 6). The dependence plots show both the relationship of a CpG site with the predicted age and how that relationship can be affected by the value of a

secondary CpG site for a DNA methylation sample. This secondary CpG site has the highest interaction with the CpG of interest, as determined by SHAP values. One way to understand the effect of the secondary CpG site is to focus on the samples in the top and bottom deciles of its methylation value, looking for any differences that may arise due to hyper- or hypomethylation respectively. We categorized three different types of relationships between CpG site methylation value and age: (1) completely linear, which are independent of CpG-CpG interactions; (2) bivalently linear, whose slope is dependent on a secondary CpG site; and (3) non-linear, affected by a secondary CpG site.

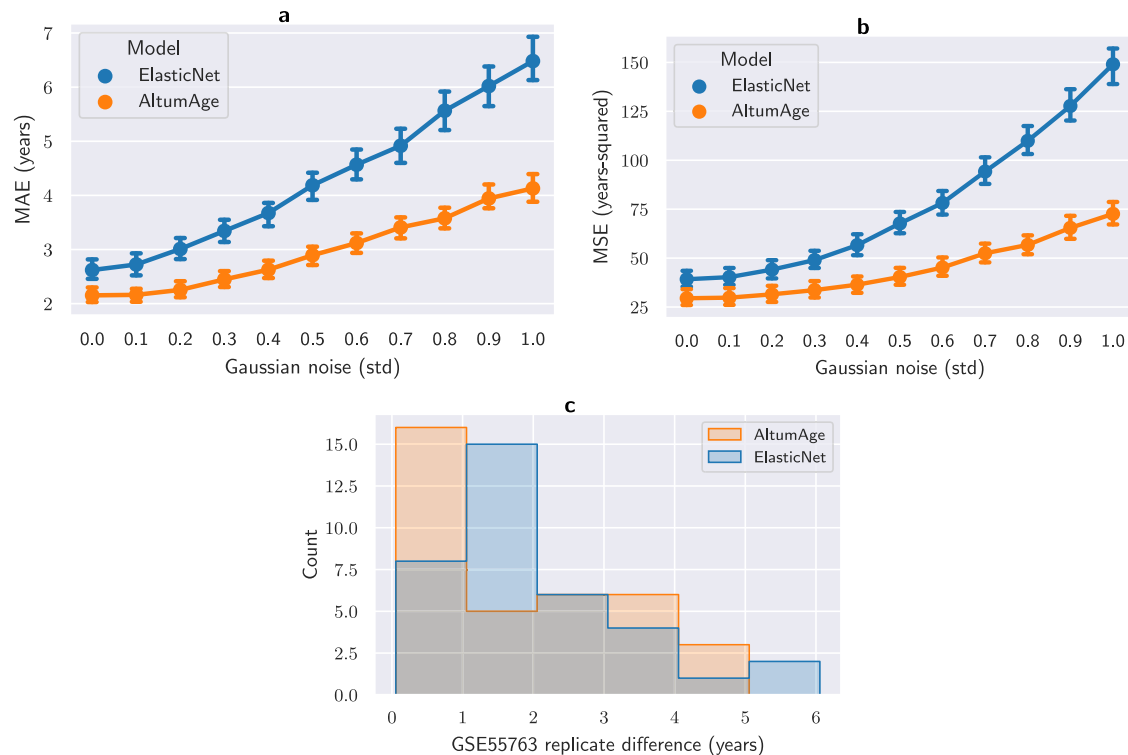


Fig. 3 Comparison of resistance to noise between AltumAge and ElasticNet. **a, b** Point plots show the increase in median absolute error (MAE) and mean squared error (MSE) per model when adding artificial Gaussian noise of up to one standard deviation for each feature. AltumAge is more resistant to noise in both metrics. Shown are the 99% confidence interval error bars calculated from 1000 bootstrap iterations. **c** Histogram of the difference in predicted age between two technical replicates in an independent whole blood data set (GSE55763). AltumAge has a lower median and maximum deviations than ElasticNet.

Out of the top nine CpG sites, only cg04084157 (Fig. 4a), the fourth most important, shows an almost completely linear relationship. We subdivided the samples in the test set into the top and bottom deciles for cg22736354, the most highly interacting CpG site. Both subsets display linear relationships despite its heteroscedasticity or unequal scatter (Fig. 4b), though with slightly different regression coefficients ($Z\text{-test}^1 p = 1.75e - 14$). The main interaction effect appears to be that when cg22736354 is hypermethylated, so is cg04084157, and when cg22736354 is hypomethylated, so is cg04084157. Of note, we repeated the aforementioned analysis for all top 1000 CpG sites but did not find a single CpG site without a statistically significant interaction, i.e., a difference in the linear regression coefficients for the top and bottom deciles with $p > 0.05$.

The effect of cg10523019 (Fig. 4c), the fifth most important CpG site, displays a bivalently linear relationship. The regression coefficient when cg19722847, the most highly interacting CpG site, is hypomethylated (coef = 0.653) is more than twice when it is hypermethylated (coef = 0.290, $Z\text{-test } p < 1.0e - 300$, Fig. 4d). This dual response may also shine a light on relevant age-related biological processes. cg19722847 is located in the gene IPO8, a gene that participates in nuclear transport, and cg10523019 lies in RHBDD1, a gene involved in proteolysis and apoptosis. IPO8 is known to interact with transcription factors NF- κ B and FOXO3 by allowing their nuclear transport and hence activation^{29,30}. Indeed, RHBDD1 is a known target of FOXO3³¹. The methylation status of cg10523019 becomes less relevant as transcription of RHBDD1 may become deficient due to lack of IPO8-and nuclear

FOXO3—regardless of RHBDD1 methylation. Laboratory experiments would have to be performed to more thoroughly characterize these relationships; however, it is possible to obtain data-driven hypotheses from these dependence plots.

An example of a non-linear CpG-age relationship comes from the seventh most important CpG cg12373771 (Fig. 4e). When cg07099407, the most highly interacting CpG site is hypermethylated, the relationship between cg12373771 and age output is still linear. However, when it is hypomethylated, it becomes slightly non-linear (Fig. 4f). While Pearson's correlation coefficient for a straight line is high (0.961), the residual plot (Supplementary Figure 7) shows underestimation at the boundaries with overestimation in the center. This pattern demonstrates non-linearity; a cubic regression corrects the bias of under and overestimation and increases the correlation coefficient to 0.980. Overall, our results using SHAP values demonstrate that AltumAge captures the non-linear interaction between CpG sites.

Note that despite their important effects on the predicted age, some of the CpG sites that interact with the most important CpG sites may themselves not be particularly relevant for the output. For example, the aforementioned cg07099407, the CpG with the highest interaction with the seventh most important CpG site, ranks 6315 and 3975 according to SHAP and DeepPINK, respectively, out of 20,318. These results suggest that ElasticNet may miss DNA loci that regulate other loci in aging, and this may partly explain AltumAge's performance improvement compared to ElasticNet.

Characterization of CpG sites by model interpretation. CCCTC-Binding factor (CTCF) is a transcription factor involved in the negative regulation of several cellular processes. It also contributes to long-range DNA interactions by affecting chromatin architecture.

¹ $z = (\text{coef}_1 - \text{coef}_2) / \sqrt{sd_1^2 + sd_2^2}$, here *coef* is the coefficient, *sd* is the standard deviation.

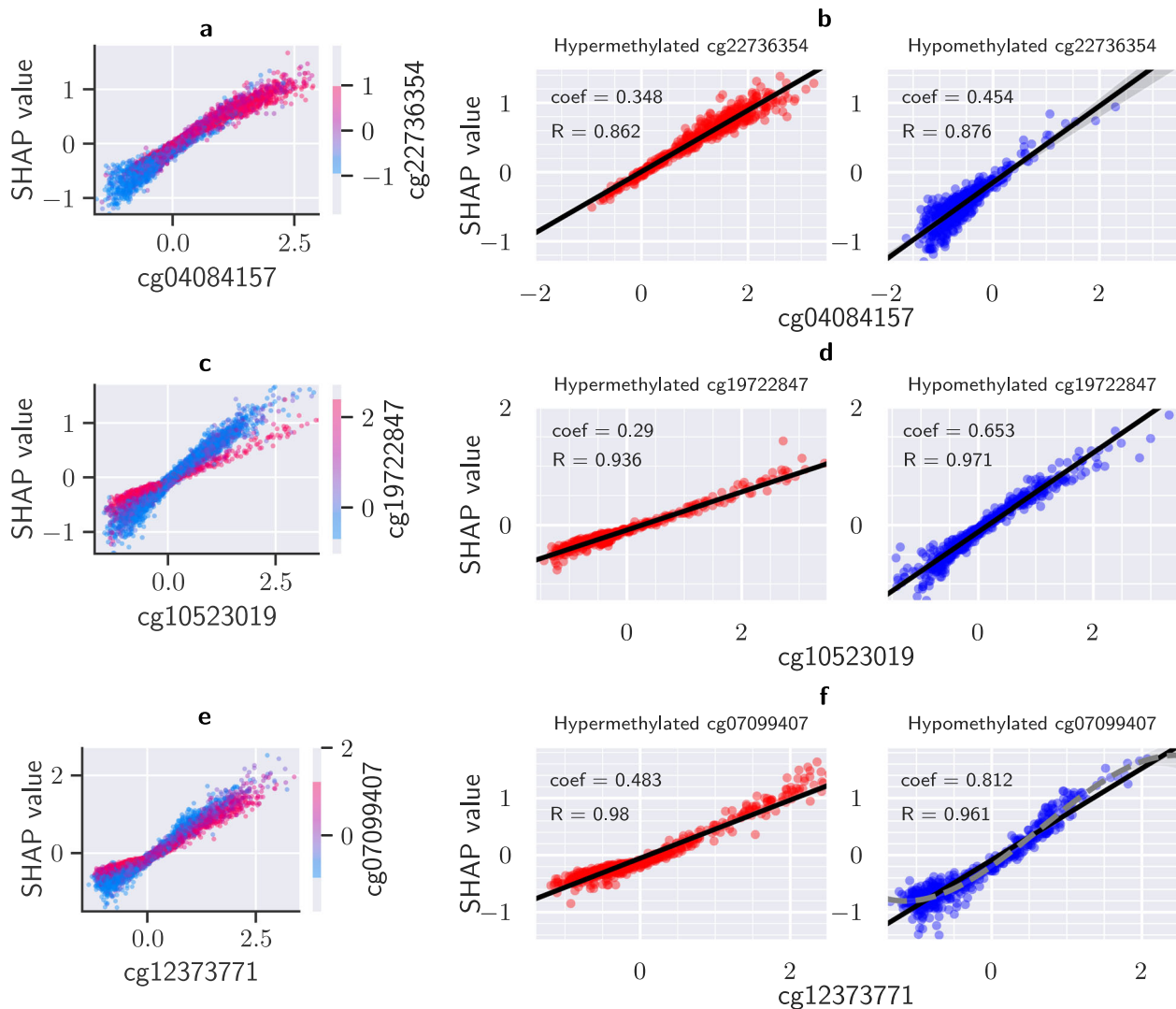


Fig. 4 Three main types of relationship between the scaled beta value of a CpG site and age according to SHAP value. It measures how one feature contributes to the output of deep neural networks. For our case, the SHAP value can be conceived as how much the value of one CpG site affects the age output of the model in years. The x-axis shows the scaled beta values for each specific CpG site; the y-axis, its SHAP value, and the coloring scheme, the scaled beta values for the CpG site with the highest interaction. The effect of a specific CpG site on the predicted age can vary based on a secondary CpG site. Dependence plots of the fourth (a), fifth (c), and seventh (e) most important CpG sites exemplify the three types of relationship. Samples into the top (red) and bottom (blue) deciles of the most highly interacting CpG site were divided, representing hyper- and hypomethylation respectively. The relationships are completely linear (b), bivalently linear (d), and non-linear (f). Regression lines are shown in (b), (d), and (f) with a 95% confidence interval calculated from 1000 bootstrap iterations. A cubic regression (dotted grey line) is also shown in (f) to demonstrate the better fit of the non-linear model.

We examined whether CpG sites with a higher SHAP importance were closer to CTCF binding sites. The 353 important CpG sites selected by Horvath's model were not closer to the CTCF binding sites when compared to the 21,368 control CpG sites from which the paper's ElasticNet model was trained (Mann-Whitney U-test p -value = 0.991). As for AltumAge, since it uses all of the 20,318 CpG sites as features, we compared the top 903 CpG sites to the control, as the ElasticNet model applied on the full data set selects 903 sites as important. These sites comprise 45.3% of SHAP importance. In line with previous studies^{32–34}, we find that the selected important CpG sites are overwhelmingly closer to CTCF binding sites (Mann-Whitney U-test p -value = 0.041, Supplementary Figure 8). This observation suggests that epigenetic alterations proximal to such loci that are involved in chromatin packing by affecting CTCF binding may be captured by AltumAge. This result is relevant because chromatin structure modifications have been associated with aging (see review³⁵).

Due to the close relationship between chromatin and aging, we hypothesized that different chromatin states would influence the importance of each CpG site. ChromHMM is a Hidden Markov Model used for the characterization of chromatin states³⁶. Annotations for several cell lines and tissue types are widely available online. Since AltumAge is a pan-tissue epigenetic clock, we used the mode of the 18-state annotation from 41 different tissues obtained from ENCODE for each CpG location³⁷ (Supplementary Figure 9, Supplementary Table 2). CpG SHAP importance values are indeed impacted by ChromHMM state (Kruskal–Wallis H-test p -value = $2.03e-78$). The chromatin states with the highest and second highest SHAP normalized median importance were ZNF genes and repeats (SHAP importance = $1.68e-03\%$, top 73th percentile of all CpG sites) and heterochromatin (SHAP importance = $1.68e-03\%$, top 58th percentile). Of note, the chromatin states with the highest DeepPINK normalized median importances were also heterochromatin (most important) and ZNF genes and

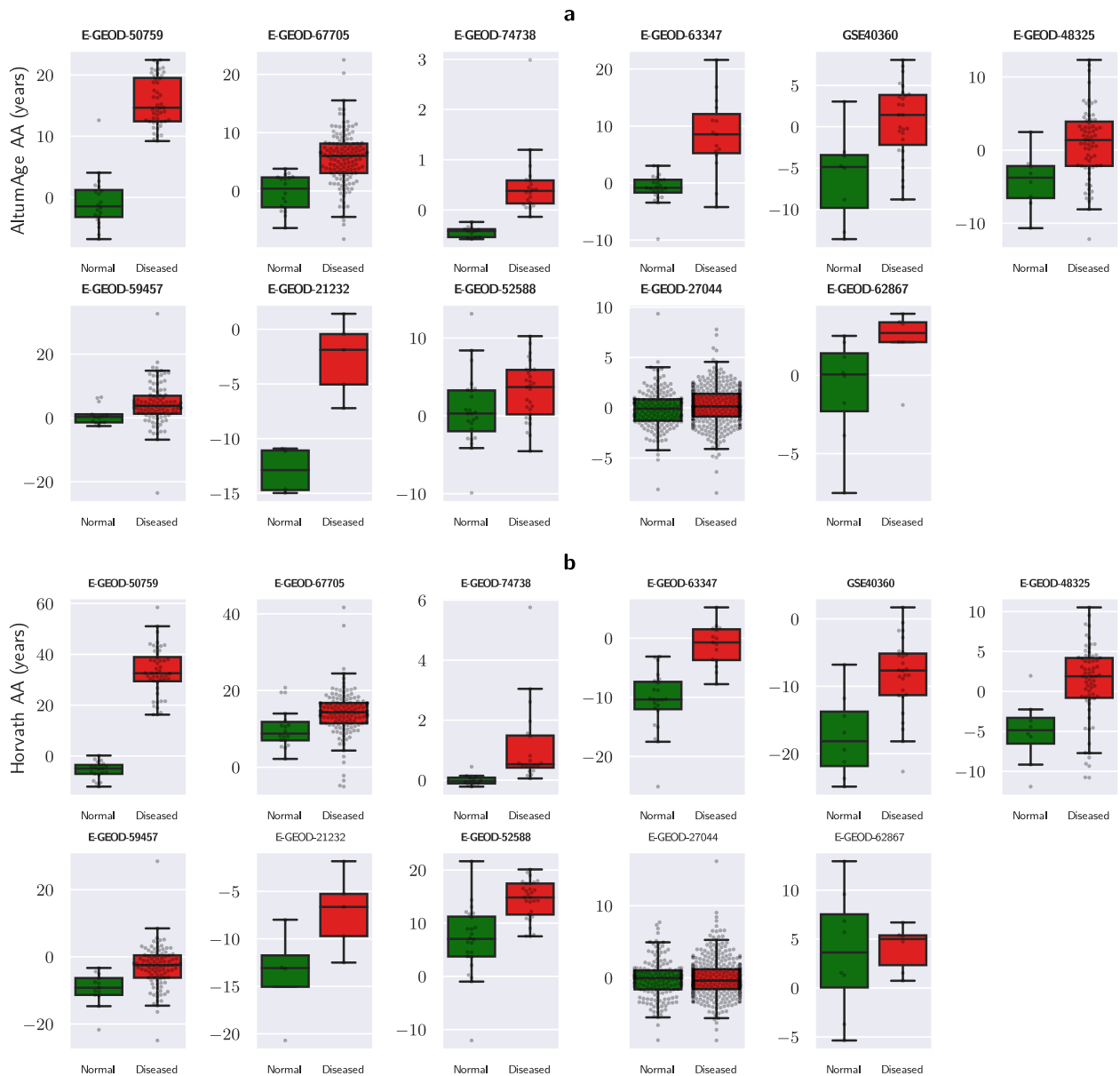


Fig. 5 Age acceleration of AltumAge and Horvath's model for various data sets with normal and diseased samples. Box plots show statistically significant (Mann-Whitney U-test p -value < 0.05) age acceleration (AA) with a title in bold according to AltumAge (**a**) and Horvath's model (**b**). Data sets include samples with autism (E-GEOD-50759, E-GEOD-67705, E-GEOD-59457), with HIV (E-GEOD-27044), from several pregnancy disorders (E-GEOD-74738), with multiple sclerosis (GSE40360), with non-alcoholic fatty liver disease (E-GEOD-48325), with type 2 diabetes (E-GEOD-21232), with Down syndrome (E-GEOD-52588), and with atherosclerosis (E-GEOD-62867). The box and whiskers show the quartiles of the data.

repeats (second most important). Heterochromatin is particularly interesting as several interventions that increase chromatin packing also extend lifespan³⁵. The full ChromHMM results can be found in Supplementary Note 2.

Ageing-related pathways. One of the main interpretation advantages of AltumAge compared to ElasticNet is that the former effectively uses a much larger feature space. CpG sites in aging-related genes are often not selected within the couple hundred features of an ElasticNet model, thus making analyses of these CpG sites of interest impossible. AltumAge allows a closer look at the relationship of CpG sites in aging-related pathways even when these CpG sites are not particularly important for the final age prediction. It is worth analyzing the relative importance of CpG sites in well-known age-related pathways such as SIRT, mTOR, and AMPK^{38–40}.

Unexpectedly, most of the CpG sites in SIRT genes do not appear

relevant, at least directly, for age prediction using AltumAge. Located in SIRT2, cg27442349, accounting for 0.026% of the total SHAP importance and ranked 1119, has the highest SIRT SHAP importance value (Supplementary Figure 10a).

Out of the 67 proteins participating in the mTOR signaling pathway according to the PID Pathways data set⁴¹, cg11299964, located in MAPKAP1, has the highest SHAP importance of 0.066%, ranking 148. Surprisingly, mTOR was not particularly relevant, with its most important CpG site being cg04508649 (SHAP importance = 0.0062, rank 7705) (Supplementary Figure 10b).

In terms of the AMPK pathway, out of CpG sites in genes for the proteins that directly activate or inhibit AMPK from the KEGG database⁴², cg22461835, located in ADRA1A, has the highest SHAP importance of 0.061%, ranking 169 (Supplementary Figure 10c). Most, however, ranked below 1000.

We also performed KEGG pathway analysis on the genes related to

the top-ranking nine CpG sites using KEGGMapper⁴³. We found the following genes associated with four of them—NHLRC1, involved in proteolysis; KLF14, associated with type 2 diabetes; BCO1, involved in metabolic pathways, including biosynthesis of cofactors; and FZD9, involved in a range of age-related diseases, including cancer and neurodegeneration. Note that DNA methylation affects gene expression depending on its position. A methylated CpG site in an enhancer, promoter, or gene body may impact gene regulation differently. These findings show how methylation in specific loci in aging-related pathways can contribute to age prediction. This insight may not be possible to obtain using ElasticNet due to its focus on selecting only the most important CpG sites related to aging. For example, only cg1129964 (from MAPKAP1 mentioned above) was present among the 353 sites selected by Horvath's model.

Potential biological applications

The age acceleration, defined as the predicted age minus the real age, of epigenetic clocks have been shown to be related to several biologically relevant events and characteristics, such as obesity⁴⁴, menopause⁴⁵, diet⁴⁶, heart disease⁴⁷, anxiety⁴⁸, and even socioeconomic status⁴⁹, among others. Given the observed performance of AltumAge, we explore its applicability to such studies, for which Horvath's model has been a popular choice. Therefore, we assess AltumAge's age acceleration performance with respect to the Horvath's model in the following sections to understand the usefulness of its age prediction for downstream analyses.

AltumAge predicts higher age acceleration for certain pathologies. To determine whether some diseases accelerate aging, we

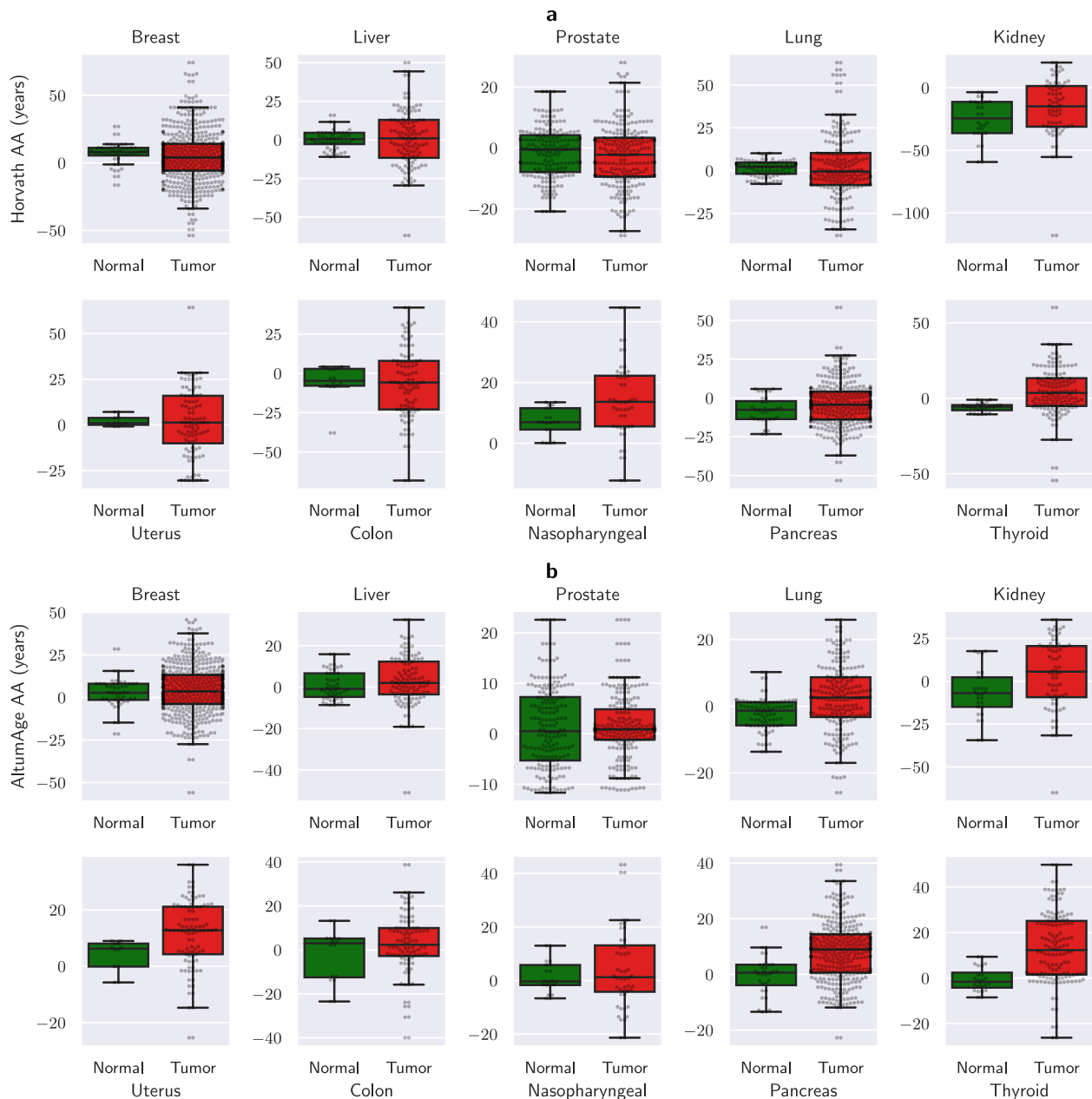


Fig. 6 Age acceleration of AltumAge and Horvath's model for various tissue types with normal and cancerous samples. Horvath's model **a** does not significantly predict higher AA for cancer, with some tissue-specific behavior. On the other hand, AltumAge **b** predicts higher median AA for tumors in all but one cancer tissue type. The box and whiskers show the quartiles of the data.

investigated the separated, diseased samples (not used in training, validation, or testing) from 41 data sets described in Supplementary Note 1. We then compared the age acceleration of these separated samples to the age acceleration of the test samples from the same data set in a pairwise fashion.

There were 12 statistically significant (Mann-Whitney U-test p -value < 0.05) comparisons according to AltumAge, 11 of which predicted higher age acceleration for the diseased group (Fig. 5a). These include ectodermal samples from patients with autism (E-GEOD-50759, p -value = 3.41×10^{-10}), blood samples from patients with HIV (E-GEOD-67705, p -value = 2.93×10^{-7}), fetal samples from several pregnancy disorders (E-GEOD-74738, p -value = 3.29×10^{-6}),

brain samples from patients with autism (E-GEOD-63347, p -value = 7.12×10^{-5}), brain samples from patients with multiple sclerosis (GSE40360, p -value = 0.00251), liver samples from patients with non-alcoholic fatty liver disease (E-GEOD-48325, p -value = 0.00431), brain samples from patients with HIV (E-GEOD-59457, p -value = 0.0131), pancreatic cells from patients with type 2 diabetes (E-GEOD-21232, p -value = 0.0158), blood samples from patients with Down syndrome (E-GEOD-52588, p -value = 0.0197), peripheral blood mononuclear cells from patients with autism (E-GEOD-27044, p -value = 0.0232), and blood vessel samples from patients with atherosclerosis (E-GEOD-62867, p -value = 0.0426). In Horvath's case, there were 10 statistically significant comparisons,

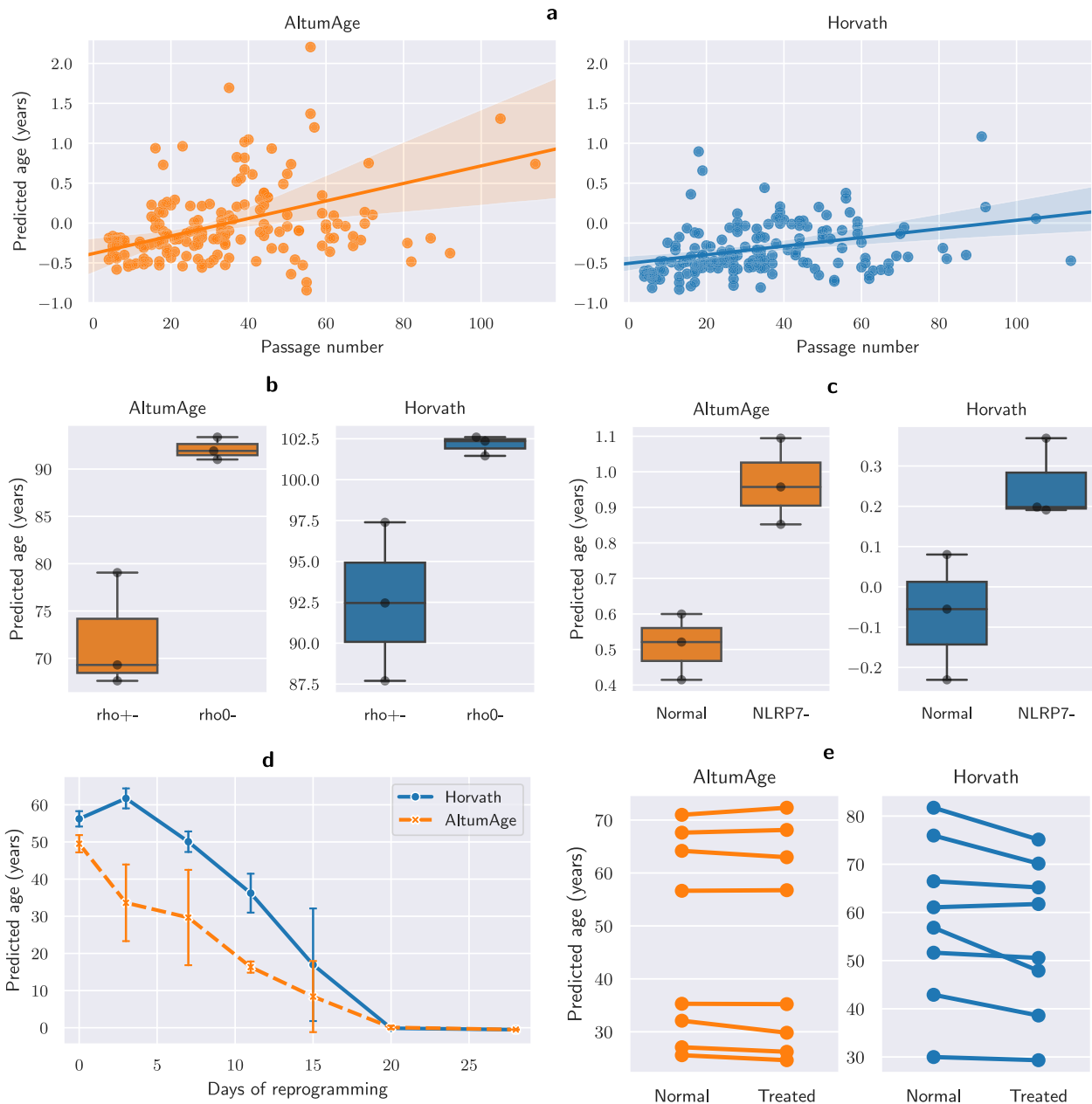


Fig. 7 Analysis of the effect of cell passage number, mitochondrial depletion, NLRP7 knockdown, continuous cellular reprogramming, and transient cellular reprogramming on predicted age comparing AltumAge and Horvath's model. **a** Scatter plot of predicted age of iPSCs and ESCs by passage number with best fit line with 95% confidence interval calculated from 1000 bootstrap iterations. **b** Box plots showing predicted age in 143B cells with mitochondrial depletion (rho0-) or control (rho+-). **c** Box plots showing predicted age of H9 ESCs with NLRP7 knockdown (NLRP7-) or control (Normal). **d** Time course analysis of OSKM reprogramming with standard deviation error bars. **e** Point plot showing age prediction in human fibroblasts and endothelial cells before and after transient reprogramming. The box and whiskers (**b**, **c**) show the quartiles of the data.

eight of which predicted higher age acceleration for the diseased group (Fig. 5b).

Given the high number of pairwise comparisons, it is likely that a few will be statistically significant by chance, as the expected number of false null hypothesis rejections for 41 pairwise comparisons with a p -value threshold of 0.05 is 2.05. Indeed, AltumAge predicted a lower age acceleration for blood samples of patients with schizophrenia (E-GEOD-41169, p -value = 0.0161) and Horvath's model predicted a lower age acceleration for blood samples of patients with osteoporosis (GSE99624, p -value = 0.00531) and in sun-exposed skin (E-GEOD-51954, p -value = 0.0181); these may be artifacts.

Of note, there was no statistically significant age acceleration for both AltumAge and Horvath's model in several data sets that include patients with obesity, Crohn's disease, schizophrenia, asthma, chronic obstructive pulmonary disease, among others. Overall, however, these observations indicate that AltumAge predicts higher age acceleration for certain pathologies and may indicate which ones are epigenetically age-related.

AltumAge predicts higher age acceleration for cancer. Cancer cells display several genetic and epigenetic aberrations which have been related to aging and mortality by epigenetic clocks^{50–52}. Liu et al.⁵³ have reported that some age predictors consistently estimate higher age acceleration for tumors, whereas others show tissue-specific behavior. Therefore, we examined the age acceleration of cancer samples from 14 data sets comprising 10 tissue types in total for AltumAge, using Horvath's model as a benchmark (Fig. 6). Overall, Horvath's model was not able to differentiate between normal and tumor samples (Mann–Whitney U-test p -value = 0.156, Fig. 6a). Its median age acceleration for cancer was marginally higher in seven tissue types and lower in another three (breast, prostate, colon). AltumAge, in contrast, predicts overall higher age acceleration for cancer when compared to normal tissue by 4.542 years (Mann–Whitney U-test p -value = 8.28e–12, Fig. 6b). The median age acceleration was higher for tumors in every single tissue type examined except colon, and the mean age acceleration was higher for all tumors. These results indicate that AltumAge can generally differentiate between normal and cancerous tissue by predicting a higher age acceleration and could be useful to studies focusing on the relationship between cancer and aging. Note that the age accelerations of both models had a much higher overall variance in cancer versus normal tissue (AltumAge Levene's test p -value = 7.32e–10; Horvath's model Levene's test p -value = 3.63e–12). This observation might be the consequence of a more chaotic epigenome in tumors.

AltumAge differentiates cells with age-related hallmarks. To assess whether AltumAge can capture biologically relevant changes in vitro, we examined independent data sets (not used for training, validation, or testing) to understand the effect of cell passage number, cellular senescence, mitochondrial depletion, NLRP7 knockdown, transient cellular reprogramming, and continuous cellular reprogramming. We compared AltumAge with Horvath's model as a benchmark.

We looked into a study (with accession ID GSE30653⁵⁴) which contains information of induced pluripotent stem cells (iPSCs) and embryonic stem cells (ESCs) by passage number. We observe that AltumAge detects a correlation (Pearson's R = 0.351, p -value = 2.85e–06) between the predicted age and the passage number. As shown by Fig. 7a, cells begin with a slightly negative age that increases as they are passaged. Horvath's model also detects a significant correlation (Pearson's R = 0.273, p -value = 3.245e–04), albeit it is weaker when compared to AltumAge. The increase in age with passage number is also more subtle. The effect of passage number on predicted age indicates that AltumAge might be sensitive to cellular exhaustion due to passaging cell.

Given that cellular senescence is a well-known hallmark of aging and can be caused by excessive replication in vitro, we analyzed a study (with accession ID GSE91069⁵⁵) which provides methylation data from fibroblasts at different stages of senescence (Supplementary Figure 11). Early passage cells did not have a statistically different predicted age with AltumAge (regression coefficient t -test p -value = 0.078), and only slightly with Horvath's model (regression coefficient t -test p -value = 0.046). None of the other comparisons were statistically significant.

Mitochondrial dysfunction is another important hallmark of aging. A study (with accession ID GSE100249⁵⁶) contains data on 143B cells chronically depleted of mitochondrial DNA (rho0-). AltumAge predicts a higher age for cells with mitochondrial dysfunction (regression coefficient t -test p -value = 0.036, Fig. 7b). Horvath's model also predicts a suggestive increase (regression coefficient t -test p -value = 0.057).

The NLRP gene family of receptors, primarily expressed in immune cells, is involved in the normal response to inflammation. Mutations in some of these genes are involved in immune system malfunction, excessive inflammation, and disease^{57–59}, suggesting it may also have ramifications in aging. Moreover, cells with NLRP7 knockdown display aberrant CpG methylation patterns⁶⁰. We, therefore, analyzed DNA methylation data from H9 ESCs (accession ID GSE45727⁶⁰) with or without NLRP7 knockdown. AltumAge and Horvath's model predict a higher age for knock-down cells (regression coefficient t -test p -value = 0.009 and 0.019, respectively, Fig. 7c). When H9 cells are exposed to BMP4 differentiating medium, both models are still able to capture the increase in age (regression coefficient t -test p -value = 0.024 and 0.013 respectively, Supplementary Figure 12).

Cellular reprogramming is currently the most effective way to reduce the biological age of a cell. Thus, we analyzed the methylation from a study (with accession ID GSE54848⁶¹) to check whether reprogramming with the four Yamanaka factors (Oct-3/4, Sox2, Klf4 and c-Myc) would lead to rejuvenation. Indeed, there is a steady decrease in predicted age for AltumAge until day 20 (Fig. 7d). Horvath's model, intriguingly, displays an initial increase in predicted age at day 3. This likely artifact is not present in AltumAge's reprogramming time course.

Lastly, we investigated whether AltumAge could capture a rejuvenation event caused by transient expression of several reprogramming factors in aged human fibroblasts and endothelial cells (GSE142439)⁶². Sarkar et al.⁶² demonstrated that several biomarkers such as H3K9me3, SIRT1, HP1y, and β -galactosidase were restored to a youthful state after treatment. Horvath's model was able to capture a decrease in epigenetic age (linear mixed-effects model p -value = 0.004, Fig. 7e). Interestingly, there was no difference before and after the intervention according to AltumAge (linear mixed-effects model p -value = 0.966, Fig. 7b). While the researchers tracked the cells until six days after treatment, it is possible that the apparent restoration of youthful biomarkers would not endure. Indeed, studies have shown that transient reprogramming causes only temporary rejuvenation^{63–65}. Altogether, AltumAge performs as well as Horvath's model for most cases and, more importantly, it captures an expanded global methylation landscape. It can also robustly recognize age-related epigenetic patterns while potentially avoiding overestimating the impact of temporary interventions and detecting interventional artifacts.

DISCUSSION

The creation of new quantitative aging measurements has been rapidly expanding with the burgeoning field of the biology of aging. Epigenetic clocks are a tool that can aid researchers to understand better and to measure the aging process. In 2013, Horvath showed it was possible to use just a couple of CpG sites to predict a person's age based on DNA methylation accurately. It

was a giant leap in the field. However, his 2013 ElasticNet model or other versions of linear models are still widespread despite recent advances in machine learning. The accuracy of such linear models was so good that it was difficult to imagine a model significantly outperforming it⁶⁶. Other deep learning methods, which slightly outperform ElasticNet, have focused thus far only in a single tissue type¹⁶¹⁷.

We show that our neural network-based model, AltumAge, is an overall better age predictor than ElasticNet. While our more comprehensive and larger data does improve the performance, the capability of neural networks to detect complex CpG-CpG interactions and the expanded feature set with 20,318 CpG sites also contribute to its lower error. For within-data set prediction - which is the case for several studies which create a new epigenetic clock - AltumAge performs drastically better than state-of-the-art methods. Even for LOOCV analysis, while the improved performance of AltumAge over ElasticNet was not as substantial, it performed better in older ages and new tissue types. Arguably, a more generalizable model like AltumAge can better capture pan-tissue age-related changes.

Deep learning models have shown promise in several biological tasks, given their good performance on unstructured data. They have been for many years seen as “black-box” models, but new interpretation methods have made it possible to get interesting insights for them. Our interpretation of AltumAge provides a detailed relationship between each of the 20,318 CpG sites and age, showing that while most CpG sites are mostly linearly related with age, some important ones are not. Given recent advances in epigenetic editing¹, finding these DNA methylation sites to delay or reverse aging may be necessary for future interventions to tackle the disease. AltumAge allied with other deep learning inference methods can provide information on highly interacting CpG sites. Sometimes the primary locus of an epigenetic editing intervention, given its place in the genome, may be difficult to target because of the chromatin structure. Consequently, knowing secondary CpG sites that affect how the CpG of interest interacts with age could guide such interventions. We show that one can obtain biological hypotheses for the same from the data using AltumAge. For example, we observe that cg19722847 located inside the gene IPO8 could regulate cg10523019, which lies in RHBDD1. Analysis of ChromHMM annotations shows that the top-ranking CpG sites are associated with gene regulatory regions and CTCF binding sites. Finally, we highlight the age-related KEGG pathways obtained for genes with these CpG sites, indicating that the model is learning valuable biological information from the data.

We also explore how age acceleration as determined by AltumAge has potentially meaningful biological applications. AltumAge predicts higher age acceleration for cells with immune and mitochondrial, similarly to Horvath’s model. Moreover, AltumAge displays a higher correlation between cell passage number and predicted age. More importantly, AltumAge, in contrast to Horvath’s model, predicts higher age acceleration for cancer. It seems that epigenetic clocks highly predictive of mortality show this behavior consistently⁵³. Age acceleration in tumors can be thought of as a further deviation from the original state in Waddington’s landscape. Intriguingly as well, AltumAge does not uncover a rejuvenation event from transient reprogramming⁶². This observation is possibly due to the temporary nature of the rejuvenation event, which may not globally change age-related DNA methylation patterns. Since AltumAge considers a much larger portion of the epigenome, it may be more resistant to detecting transient events. Another contrast with Horvath’s model is that during a reprogramming time course study, it does not show an initial increase in predicted age, likely an artifact.

In future work, it would be interesting to create a deep learning model with Illumina’s EPIC array with the roughly 850 thousand

CpG sites to understand more deeply how genomic location can affect influence in aging. In addition, by having several CpG sites in a single gene, it is also possible to better understand how methylation in different positions may affect the contribution of a particular gene to the aging process. Currently, however, there are only a few EPIC array data sets publicly available. Another interesting application for deep learning in the aging field is the creation of epigenetic clocks that directly predict mortality. Currently, the state-of-the-art mortality predictor is GrimAge⁶⁷, which was created based on a linear Cox proportional hazards model. We anticipate that using neural networks to include non-linear relationships and CpG-CpG interactions would result in a better lifespan predictor.

Overall, we have shown that deep learning represents an improvement in performance over current approaches for epigenetic clocks while at the same time providing new, relevant biological insights about the aging process.

METHODS

DNA methylation data sets

For model training and testing, we gathered normal tissue samples from 142 publicly available DNA methylation data sets from the Gene Expression Omnibus, Array Express, and The Cancer Genome Atlas, comprising of the platforms Illumina Infinium HumanMethylation27 and the Illumina Infinium HumanMethylation450. All selected data sets had both processed beta values and age available for all samples. Missing values per data set were imputed with a KNN imputer from `scikit-learn`. Next, the data was normalized using the beta mixture quantile normalization (BMIQ) with the optimized code from Horvath, called BMIQCalibration^{3,68}. 13 data sets contained tumor samples which were separated for later analysis. Samples that failed BMIQ normalization were removed. Then, each data set was split 60% for training ($n = 8050$) and 40% for testing ($n = 5455$). The within-data set split ensures the distribution of age, gender, and tissue type between training and testing sets are unbiased (Supplementary Figure 4). In the training set, the data was further randomly subdivided by data set, with 85 ($n = 4394$) for model selection and 57 ($n = 3656$) for validation. After the best model was selected, it was trained in the full training set ($n = 8050$) and analyzed in the test set. Supplementary Figure 1a shows a schematic of the division of the data. The full list of data sets used is available in Supplementary Note 1.

For twelve data sets in which gestational week was available, the encoding for age is the following:

$$y = 7 * \frac{w - 40}{365} \quad (1)$$

where w is the gestational week, and y is the age in years. A gestational week below 40 would have negative age; for instance, 30 weeks would be encoded as $7*(30 - 40)/365 = -0.192$. In the US in 2013, the average birth occurred at an estimated 38.5 weeks⁶⁹. This number has changed slightly over time, and since preterm deliveries skew the average more than later-term births, we considered gestational week 40 as age 0 in such data sets.

For the cancer age acceleration analysis, we compared the test set of 13 data sets with the aforementioned separated cancer samples. These data sets were GSE32393, GSE37988, GSE26126, GSE63384, GSE59157, GSE32867, GSE30759, GSE31979, GSE77955, GSE52068, GSE49149, GSE39004. We further added GSE53051, which contains normal and tumor samples from five tissue types, for the analysis. It is worth noting that analyzing GSE53051 separately did not change the outcome of higher tumor age acceleration predicted with AltumAge vs. no difference with Horvath’s model. In total, we compared 434 normal and 1856 cancer observations across 10 different tissue types (Figure 6).

Lastly, a brief description of all data sets used, either for training, validation, testing, or any other analysis, is in Supplementary Note 1. Any sample that was excluded is explicitly mentioned there alongside brief notes.

CpG site annotation

For the annotation of CpG sites, GENCODE and Zhou et al’s annotations were used^{70,71}. 41 data sets from ENCODE with the 18-state ChromHMM information were gathered³⁷: ENCF717HFZ, ENCF718AGZ, ENCF371WNR, ENCF318XQO, ENCF340OUL, ENCF893CAJ, ENCF151PZS, ENCF098CED, ENCF273PJW, ENCF377YFI, ENCF773VYR, ENCF928QES, ENCF786HDE,

ENCF827FZN, ENCF364PIY, ENCF802QCI, ENCF021NNN, ENCF510ZEI, ENCF175NGE, ENCF670DBL, ENCF825ZCZ, ENCF912ILE, ENCF725WBV, ENCF829SZB, ENCF483NRC, ENCF717RYX, ENCF249ZBG, ENCF2050TD, ENCF765OKG, ENCF820YPO, ENCF685BMF, ENCF545ZMG, ENCF294UQS, ENCF104ZSA, ENCF370EGY, ENCF860FWW, ENCF177TTP, ENCF151ZGD, ENCF743GHZ, ENCF990YHL, and ENCF036WIO. Since AltumAge is a pan-tissue clock, the mode of each state was chosen for each CpG site.

Model selection

Multiple machine learning models were tested in the validation set. The evaluation metrics were median absolute error (MAE), mean squared error (MSE), Pearson's correlation coefficient (R), and median error.

To validate and train the models, the beta value of each CpG site was scaled with a robust scaler which removes the median and scales according to the interquartile range. A robust scaler was chosen as opposed to a standard scaler (mean = 0, var = 1) to better resist distortions caused by outliers. The robust scaler was fitted with the training data of each model simply as a step before model training. In addition, only 20,318 CpG sites common to all three platforms Illumina Infinium HumanMethylation27, Illumina Infinium HumanMethylation450, Infinium Methylation EPIC were chosen as features.

The non-neural network models trained with `scikit-learn` were support vector regression (with an RBF kernel) and random forest with the standard hyperparameters. ElasticNet, trained with `glmnet`, used the built-in λ optimization with parameters $\alpha = 0.5$ and $n_splits = 10$. Moreover, for ElasticNet, Horvath's age transformation was used³.

To select the best performing neural network with `tensorflow`, we conducted a grid search with the following hyperparameters: number of hidden layers (2, 5, or 8), number of neurons per dense layer (32 or 64), activity and kernel L2 regularization coefficients (0, 0.0034, or 0.0132), dropout (0 or 0.1), Gaussian dropout (0 or 0.1), batch normalization (yes or no), activation function (ReLU or SeLU), and learning rate (0.0002, 0.0005, or 0.001). The following parameters were held constant: optimizer (Adam), batch size (256), number of epochs (300), loss function (MSE), and learning rate decay by a factor of 0.2 after a 30-epoch plateau in the training loss. The weights with the lowest training loss were chosen. After selecting the best hyperparameters, we trained the neural network with adversarial regularization with `neural_structured_learning` with multiplier = 0.05, `adv_step_size` = 0.005. The main idea of adversarial regularization is to train a model with adversarially-perturbed data (called adversarial examples) in addition to the original training data. These adversarial examples are constructed by intentionally adding noise and cause the model to mispredict. By training with such examples, the model learns to be robust against adversarial perturbations (or noise) when making predictions.

We dubbed the best performing deep neural network as AltumAge. It consists of 5 hidden layers, 32 neurons per layer, activity and kernel regularization coefficients of 0.0034, no dropout, gaussian dropout of 0.1, with batch normalization, SeLU activation, and learning rate of 0.001. AltumAge was also tested in the validation set with a smaller set of features using the CpG sites selected from the ElasticNet.

To compare our deep learning approach with other state-of-the-art neural networks, we tried TabNet, an attentive interpretable tabular learning method¹⁹. Similarly, TabNet was trained for 300 epochs.

The results for the validation set with the full list of models, including the replication of Horvath's model³, is in Supplementary Table 1. Support vector regression was by far the worst performer (MAE = 14.229, MSE = 458.956), followed by random forest (MAE = 6.833, MSE = 165.354). All other models displayed $R > 0.9$, with AltumAge having the lowest MAE (3.563) and MSE (57.071).

SHAP and DeepPINK

To obtain the SHAP values for AltumAge, the function `GradientExplainer` from `shap` was used on the test set. For the DeepPINK importance values and feature selection, the standard architecture and number of epochs was used²⁸. To create the knockoff features for DeepPINK, the function `knockoff.filter` from the R 4.0.2 package `knockoff` version 0.3.3 was used with the importance statistic based on the square-root lasso.

Both SHAP and DeepPINK importance values were normalized so that their sum would equal to 100. Each importance value thus represents a percent contribution of a certain feature.

Statistical analysis

All statistical tests were conducted with packages `scikit-learn`, `statsmodel`, or `scipy`. Tests that were one-sided were explicitly mentioned in the main text; all others were two-sided.

To assess the performance of the models, we used median absolute error (MAE), mean squared error (MSE), Pearson's correlation coefficient (R), and median error. To compare the performance of ElasticNet and AltumAge in the within-data set setup, we also used Alpaydin's Combined 5x2cv F test⁷², which combines several metrics to compare the statistical significance in error between two models across different cross-validation splits. This ensures that any improved performance is not simply due to a favourable random split.

DATA AVAILABILITY

The list of all the data sets used, a summary of the results per data set, and detailed instructions to run AltumAge can be found in the paper's GitHub repository (<https://github.com/rsinghlab/AltumAge>). The GitHub also links to a Google Drive where our gathered DNA methylation data is publicly available.

CODE AVAILABILITY

Detailed instructions on how to run AltumAge can be found in the paper's GitHub repository (<https://github.com/rsinghlab/AltumAge>). All code was run with Python (3.9.1) and packages `scikit-learn` (0.24.2), `pandas` (1.3.0), `numpy` (1.19.5), `glmnet` (1.1), `tensorflow` (2.5.0), `neural_structured_learning` (1.3.1), `statsmodel` (0.10.2), `scipy` (1.7.0), `shap` (0.39.0).

Received: 17 September 2021; Accepted: 24 March 2022;

Published online: 19 April 2022

REFERENCES

- Nuñez, J. K. et al. Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell* **184**, 2503–2519.e17 (2021).
- Hannum, G. et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **49**, 359–367 (2013).
- Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, 3156 (2013).
- Fahy, G. M. et al. Reversal of epigenetic aging and immunosenescent trends in humans. *Aging Cell* **18**, e13028 (2019).
- Chen, L. et al. Effects of Vitamin D3 supplementation on epigenetic aging in overweight and obese African Americans suboptimal Vitamin D status: A randomized clinical trial. *J. Gerontol. Ser. A: Biol. Sci. Med. Sci.* **74**, 91–98 (2019).
- Olova, N., Simpson, D. J., Marioni, R. E. & Chandra, T. Partial reprogramming induces a steady decline in epigenetic age before loss of somatic identity. *Aging Cell* **18**, e12877 <https://doi.org/10.1111/accel.12877> (2019).
- Fitzgerald, K. N. et al. Potential reversal of epigenetic age using a diet and lifestyle intervention: A pilot randomized clinical trial. *Aging* **13**, 9419–9432 (2021).
- Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc.: Ser. B (Stat. Methodol.)* **67**, 301–320 (2005).
- Levine, M. E. et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging* **10**, 573–591 (2018).
- Lu, A. T. et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* **11**, 303–327 (2019).
- Thompson, M. J. et al. A multi-tissue full lifespan epigenetic clock for mice. *Aging* **10**, 2832–2854 (2018).
- Horvath, S. et al. Reversing age: Dual species measurement of epigenetic age with a single clock. en. In: bioRxiv p. 2020.05.07.082917. <https://doi.org/10.1101/2020.05.07.082917> (2020).
- Sugrue, V. J. et al. Castration delays epigenetic aging and feminizes dna methylation at androgen-regulated loci. *Elife* **10**, e64932 (2021).
- Holzschek, N. et al. Modeling transcriptomic age using knowledge-primed artificial neural networks. *NPJ Aging Mech. Dis.* **7**, 1–13 (2021).
- Putin, E. et al. Deep biomarkers of human aging: Application of deep neural networks to biomarker development. *Aging* **8**, 1021 (2016).
- Galkin, F., Mamoshina, P., Kochetov, K., Sidorenko, D. & Zhavoronkov, A. DeepMAge: A Methylation Aging Clock Developed with Deep Learning. *Aging Dis.* **12**, 1252–1262 (2021).
- Levy, J. J. et al. MethylNet: An automated and modular deep learning approach for DNA methylation analysis. *BMC Bioinform* **21**, 1–15 (2020).

18. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **30**, 4765–4774 (2017).
19. Arik, S. & Pfister, T. Tabnet: Attentive interpretable tabular learning. *Proceedings of the AAAI Conference on Artificial Intelligence* **35**, 6679–6687 (2021).
20. Kadra, A., Lindauer, M., Hutter, F. & Grabocka, J. Well-tuned simple nets excel on tabular datasets. *Adv. Neural Inf. Process. Syst.* **34**, 23928–23941 (2021).
21. El Khoury, L. Y. et al. Systematic underestimation of the epigenetic clock and age acceleration in older subjects. *Genome Biol.* **20**, 283 (2019).
22. Booth, L. N. & Brunet, A. The aging epigenome. *Mol. Cell* **62**, 728–744 (2016).
23. Goodfellow, I. J., Shlens, J. & Szegedy, C. Explaining and harnessing adversarial examples. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings* (eds. Bengio, Y. & LeCun, Y.) <http://arxiv.org/abs/1412.6572> (2015).
24. Selvaraju, R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision*, 618–626 (2017).
25. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic attribution for deep networks. In *Proceedings of the 34th International Conference on Machine Learning, PMLR*, **70**, 3319–3328 (2017).
26. Smilkov, D., Thorat, N., Kim, B., Viégas, F. B. & Wattenberg, M. SmoothGrad: removing noise by adding noise. In *CoRR* abs/1706.03825. <http://arxiv.org/abs/1706.03825> (2017).
27. Bahdanau, D., Cho, K. & Bengio, Y. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 2015, Conference Track Proceedings*. (eds. Bengio, Y. & LeCun, Y.) <http://arxiv.org/abs/1409.0473> (2015).
28. Lu, Y., Fan, Y., Lv, J. & Stafford Noble, W. DeepPINK: Reproducible feature selection in deep neural networks. *Adv. Neural Inf. Process. Syst.* **31**, 8676–8686 (2018).
29. Du, Y. et al. Importin 8 is involved in human periodontitis by the nf-kb pathway. *Int. J. Clin. Exp. Pathol.* **12**, 711 (2019).
30. Putker, M. et al. Evolutionary acquisition of cysteines determines foxo paralogue-specific redox signaling. *Antioxid. Redox Signal.* **22**, 15–28 (2015).
31. Webb, A. E., Kundaje, A. & Brunet, A. Characterization of the direct targets of foxo transcription factors throughout evolution. *Aging cell* **15**, 673–685 (2016).
32. Han, Y. et al. New targeted approaches for epigenetic age predictions. *BMC Biol.* **18**, 1–15 (2020).
33. Wang, Y. et al. Epigenetic influences on aging: A longitudinal genome-wide methylation study in old Swedish twins. *Epigenetics* **13**, 975–987 (2018).
34. Yuan, T. et al. An integrative multi-scale analysis of the dynamic dna methylation landscape in aging. *PLoS Genet.* **11**, e1004996 (2015).
35. de Lima Camillo, L. P. & Quinlan, R. B. A. A ride through the epigenetic landscape: Aging reversal by reprogramming. *Geroscience* **43**, 463–485 (2021).
36. Ernst, J. & Kellis, M. ChromHMM: Automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
37. Davis, C. A. et al. The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).
38. Zhao, L. et al. Sirtuins and their biological relevance in aging and age-related diseases. *Aging Dis.* **11**, 927–945 (2020).
39. Weichhart, T. mTOR as regulator of lifespan, aging and cellular senescence. *Gerontology* **64**, 127–134 (2018).
40. Burkewitz, K., Zhang, Y. & Mair, W. B. AMPK at the nexus of energetics and aging. *Cell metabolism* **20**, 10–25 (2014).
41. Rouillard, A. D. et al. The harmonizome: A collection of processed data sets gathered to serve and mine knowledge about genes and proteins. *Database* **2016**, baw100 (2016).
42. Kanehisa, M. & Sato, Y. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci.: Publ. Protein Soc.* **29**, 28–35 (2020).
43. Kanehisa, M. & Sato, Y. Kegg mapper for inferring cellular functions from protein sequences. *Protein Sci.* **29**, 28–35 (2020).
44. Horvath, S. et al. Obesity accelerates epigenetic aging of human liver. *Proc. Natl Acad. Sci.* **111**, 15538–15543 (2014).
45. Levine, M. E. et al. Menopause accelerates biological aging. *Proc. Natl Acad. Sci.* **113**, 9327–9332 (2016).
46. Sae-Lee, C. et al. Dietary intervention modifies dna methylation age assessed by the epigenetic clock. *Mol. Nutr. Food Res.* **62**, 1800092 (2018).
47. Horvath, S. et al. An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol.* **17**, 1–23 (2016).
48. Marečková, K. et al. Epigenetic clock as a correlate of anxiety. *NeuroImage: Clin.* **28**, 102458 (2020).
49. Fiorito, G. et al. Socioeconomic position, lifestyle habits and biomarkers of epigenetic aging: a multi-cohort analysis. *Aging* **11**, 2045 (2019).
50. Zheng, Y. et al. Blood epigenetic age may predict cancer incidence and mortality. *EBioMedicine* **5**, 68–73 (2016).
51. Ambatipudi, S. et al. Dna methylome analysis identifies accelerated epigenetic aging associated with postmenopausal breast cancer susceptibility. *Eur. J. Cancer* **75**, 299–307 (2017).
52. Levine, M. E. et al. Dna methylation age of blood predicts future onset of lung cancer in the women’s health initiative. *Aging* **7**, 690 (2015).
53. Liu, Z. et al. Underlying features of epigenetic aging clocks in vivo and in vitro. *Aging Cell* **19**, e13229 (2020).
54. Nazor, K. L. et al. Recurrent variations in dna methylation in human pluripotent stem cells and their differentiated derivatives. *Cell Stem Cell* **10**, 620–634 (2012).
55. Xie, W. et al. Dna methylation patterns separate senescence from transformation potential and indicate cancer risk. *Cancer Cell* **33**, 309–321 (2018).
56. Lozoya, O. A. et al. Mitochondrial nicotinamide adenine dinucleotide reduced (nadh) oxidation links the tricarboxylic acid (tca) cycle with methionine metabolism and nuclear dna methylation. *PLoS Biol.* **16**, e2005707 (2018).
57. de Torre-Minguela, C., Mesa del Castillo, P. & Pelegrin, P. The nlrp3 and pyrin inflammasomes: Implications in the pathophysiology of autoinflammatory diseases. *Front. Immunol.* **8**, 43 (2017).
58. Meng, G., Zhang, F., Fuss, I., Kitani, A. & Strober, W. A mutation in the nlrp3 gene causing inflammasome hyperactivation potentiates th17 cell-dominant immune responses. *Immunity* **30**, 860–874 (2009).
59. Carriere, J., Dorfleutner, A. & Stehlik, C. Nlrp7: From inflammasome regulation to human disease. *Immunology* **163**, 363–376 (2021).
60. Mahadevan, S. et al. Nlrp7 affects trophoblast lineage differentiation, binds to overexpressed yy1 and alters cpg methylation. *Human Mol. Genet.* **23**, 706–716 (2014).
61. Ohnuki, M. et al. Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. *Proc. Natl Acad. Sci.* **111**, 12426–12431 (2014).
62. Sarkar, T. J. et al. Transient non-integrative expression of nuclear reprogramming factors promotes multifaceted amelioration of aging in human cells. *Nat. Commun.* **11**, 1–12 (2020).
63. Ocampo, A. et al. In vivo amelioration of age-associated hallmarks by partial reprogramming. *Cell* **167**, 1719–1733 (2016).
64. Manukyan, M. & Singh, P. B. Epigenome rejuvenation: Hp1β mobility as a measure of pluripotent and senescent chromatin ground states. *Sci. Rep.* **4**, 1–8 (2014).
65. Guo, L. et al. Generation of induced progenitor-like cells from mature epithelial cells using interrupted reprogramming. *Stem Cell Rep.* **9**, 1780–1795 (2017).
66. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* **19**, 371–384 (2018).
67. Lu, A. T. et al. Dna methylation grimage strongly predicts lifespan and healthspan. *Aging* **11**, 303 (2019).
68. Teschendorff, A. E. et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29**, 189–196 (2013).
69. Martin, J., Osterman, M. & Kirmeyer, S. Measuring Gestational Age in Vital Statistics Data: Transitioning to the Obstetric Estimate. *Natl. Vital Stat. Rep.* **64**, 1–20 (2015).
70. Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. *Nucl. Acids Res.* **47**, D766–D773 (2019).
71. Zhou, W., Laird, P. W. & Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res.* **45**, e22 (2017).
72. Alpaydm, E. Combined 5 × 2 cv f test for comparing supervised classification learning algorithms. *Neural Comput.* **11**, 1885–1892 (1999).

AUTHOR CONTRIBUTIONS

L.P.L.C. conceived of the presented idea. R.S and L.P.L.C. designed the methodology and the experiments. L.P.L.C. conducted all the experiments. L.R.L. assisted with the analysis and biological interpretation of the results. All authors discussed the results and contributed to the final manuscript.

FUNDING

L.R.L. is funded by a R01 (AG051810) and R21(AG068922) from the National Institute on Aging.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41514-022-00085-y>.

Correspondence and requests for materials should be addressed to Lucas Paulo de Lima Camillo or Ritambhara Singh.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022