## RESEARCH

# Applicability of epigenetic age models to next-generation methylation arrays

Leonardo D. Garma[1] and Miguel Quintela-Fandino[1*]

## Abstract

**Background**  Epigenetic clocks are mathematical models used to estimate epigenetic age based on DNA methylation at specific CpG sites. As new methylation microarrays are developed and older models discontinued, existing epigenetic clocks might become obsolete. Here, we explored the effects of the changes introduced in the new EPICv2 DNA methylation array on existing epigenetic clocks.

**Methods**  We tested the performance of four epigenetic clocks on the probeset of the EPICv2 array using a dataset of 10,835 samples. We developed a new epigenetic age prediction model compatible across the 450 k, EPICv1, and EPICv2 microarrays and validated it on 2095 samples. We estimated technical noise and intra-subject variation using two datasets with repeated sampling. We used data from (i) cancer survivors who had undergone different therapies, (ii) breast cancer patients and controls, and (iii) an exercise-based interventional study, to test the ability of our model to detect alterations in epigenetic age acceleration in response to theoretically antiaging interventions.

**Results**  The results of the four epiclocks tested are significantly distorted by the EPICv2 probeset, causing an average difference of up to 25 years. Our new model produced highly accurate chronological age predictions, comparable to a state-of-the-art epiclock. The model reported the lowest epigenetic age acceleration in normal populations, as well as the lowest variation across technical replicates and repeated samples from the same subjects. Finally, our model reproduced previous results of increased epigenetic age acceleration in cancer patients and in survivors treated with radiation therapy, and no changes from exercise-based interventions.

**Conclusion**  Existing epigenetic clocks require updates for full EPICv2 compatibility. Our new model translates the capabilities of state-of-the-art epigenetic clocks to the EPICv2 platform and is cross-compatible with older microarrays. The characterization of epigenetic age prediction variation provides useful metrics to contextualize the relevance of epigenetic age alterations. The analysis of data from subjects influenced by radiation, cancer, and exercise-based interventions shows that despite being good predictors of chronological age, neither a pathological state like breast cancer, a hazardous environmental factor (radiation), nor exercise (a beneficial intervention) caused significant changes in the values of the "epigenetic age" determined by these first-generation models.

**Keywords**  Epigenetics, Epigenetic clock, DNA methylation

## Background

In the rapidly evolving field of epigenomics, the development of epigenetic clocks has revolutionized our ability to gauge biological aging through DNA methylation patterns. Changes in the methylation state of CpG sites have proven to be highly correlated with chronological age [1, 2]. Thus, DNA methylation patterns are being increasingly used to gain insights into aging and

*Correspondence:
Miguel Quintela-Fandino
mquintela@cnio.es
[1] Breast Cancer Clinical Research Unit, Centro Nacional de Investigaciones Oncológicas–CNIO, Melchor Fernández Almagro, 3, Madrid 28029, Spain

associated pathologies [2] and have led to the development of several epigenetic clocks, which are predictive models that estimate the age of subjects based on methylation markers [3]. In these epigenetic clocks, the model prediction is interpreted as the "biological" or "epigenetic" age. Subjects whose biological age is greater than their chronological age are considered to have epigenetic age acceleration [4] (EAA). EAAs have been linked to increased risk for various health issues and mortality, independent of traditional risk factors [5–8]. Epigenetic clocks have thus not only underscored the potential of DNA methylation as a biomarker of aging [9] but also represent a valuable tool to gain insight into the complexities of multiple pathologies [4]. In cancer, EAAs have been linked to an increased risk of developing breast and colon cancer [10, 11]. Moreover, breast tissue from breast cancer patients has been shown to exhibit EAA [12]; likewise, cancer survivors do present an accelerated epigenetic age compared to noncancer patients, with an acceleration rate dependent on the treatment intensity received [13].

From the initial model of Bockland et al. published in 2011 [1], the field quickly evolved toward more robust models, such as the Hannum [14] and the Horvath [15] clocks, and later toward "second generation" approaches, which included features beyond DNA methylation with the promise of providing a "biological age" prediction that could better reflect age-related health status [16–18]. The field continues to evolve, with the implementation of new predictive models [19–21] and the development of species- [22, 23] and tissue-specific [24–27] epigenetic clocks. Most recently, Bernabeu et al. conducted a large-scale study to refine the predictive ability of both first (methylation-based) and second (including other features) generation epigenetic clocks, which yielded a significant improvement over existing models in the ability to predict chronological age from DNA methylation data across multiple cohorts with subjects of all ages [28].

The evolution of epigenetic clocks has run in parallel with that of methylation microarrays, whose coverage of DNA methylation sites increased from 27,578 in 2008 [29, 30] to 866,836 in 2015 [31]. In 2023, Illumina released the newest iteration of its series of methylation arrays, EPICv2, which targets 920,000 methylation sites with 936,866 probes [32]. Although the vast majority of probes have been conserved across different microarray models over the years [32–34], some of them have been lost. The probes on the latest (and currently the only commercially available) methylation microarray version cover more than 80% of the probes used by the most popular epigenetic clocks (PhenoAge [16], Horvath [15] [15], Horvath (2018) [35], Hannum [14], DunedinPACE [17], etc.), but they do not offer complete coverage for any of

them [36]. Given that earlier microarray models are now discontinued while there is still a need to use epigenetic clocks, it is necessary to evaluate whether the existing models can offer reliable results when using EPICv2 data.

Recent studies have shown that DNA methylation is a dynamic molecular feature, exhibiting changes which can affect epigenetic age predictions even within a 24-h timeframe [37–39]. Apsley et al. demonstrated that methylation values of probes in the EPICv1 array do exhibit significant changes using repeated sampling on the same subjects on 4 different timepoints within a 5-h interval. The effect was observed on multiple probes used by common epigenetic clocks, although the changes in the epigenetic age predictions were not directly reported [38]. More recently, Koncevičius et al. reported oscillations in the epigenetic age predictions produced by 17 epigenetic age models on 48 samples from the same subject obtained during a 72-h period, which they attributed to circadian oscillations [37]. Therefore, it is necessary to quantify the range of spontaneous variation in epigenetic age predictions to be able to distinguish truly biologically relevant alterations.

In this study, we examined the performance of existing epigenetic clocks using the CpG probes available in the new EPICv2 methylation array in human blood samples. We observed that the biological age predictions of these models were distorted due to missing probes in the EPICv2 array. To overcome this hurdle, we trained a model that can be applied to data obtained with 450 k, EPICv1, or EPICv2 chips and whose performance is at least on par with that of state-of-the-art models. We then estimated the variation of EAA in normal, non-pathological populations for our model and for four existing epiclocks. We used two datasets with repeated samples from the same subjects to obtain estimates of the contributions of technical and biological, intra-subject noise to the variation observed in the EAA distributions. The extent of these variations can set a threshold to distinguish normal from biologically relevant EAA changes. Finally, our model reproduced previously reported EAA differences attributed to radiation therapy and to sporadic breast cancer, and the negative results from an interventional study using physical exercise.

## Methods

### Study cohorts

We used exclusively publicly available methylation data from previous studies, which we collected from the GEO database [40, 41]. We provide the GEO accession number, related publication doi, number of samples, fraction of female subjects, and median age of each of the datasets used in Additional file 1: Table S1. The subjects and methods used to generate each of the datasets

### Data collection and preprocessing

Methylation data from previous studies were collected from the In all cases, the data were preprocessed: we used beta values provided by the original authors when available; otherwise, we computed the beta value from the methylated and unmethylated signals using the following formula:

$$\beta = \frac{M}{M + U + 100},$$

where $M$ and $U$ are the intensities of the methylated and unmethylated signals respectively. The main dataset consisted of 11,825 subjects with reported methylation, age, and sex data. We discarded 147 with nonnumerical age values, one with a reported age of 891 years, and one with a negative epigenetic age according to the cAge model. To generate training and test sets with even age distributions, we kept only samples with an age value present at least twice in the dataset. This process left a total of 10,835 samples on the main dataset, which were split into training and test sets with 7259 and 3576 subjects, respectively.

On the main dataset, we computed the average beta value of each probe across the whole dataset (Additional file 2: Table S2) and used this average to impute missing values. Among the probes used to train our model (see below) or for running other epigenetic clocks, we found 1.58% missing values. The number of probes missing per dataset is provided in Additional file 3: Table S3.

The validation dataset included data from 2098 subjects. Excluding subjects with missing or invalid sex and/or age annotations reduced the data to 2095 subjects. A total of 0.23% of the beta values were missing for the set of probes used to run our epiclock and the cAge models. The missing values were imputed using the average values derived from the main dataset. Additional file 3: Table S3 reports the number of probes missing per dataset.

The data from studies with repeated samples (GSE247197, GSE227809, 1 and 31 subjects, respectively), the cancer survivor dataset (GSE197674, 2138 subjects), the data from the breast cancer study (GSE148663, 32 subjects), and the data from the interventional study (GSE213363, 56 subjects) were also retrieved from the GEO database [40, 41]. We used the average beta values derived from the main datasets to impute beta values for 25 missing probes in the cancer survivor dataset.

### Epigenetic age prediction using existing models

We applied the Horvath [15], Hannum [14], Pheno-Age [16], and cAge [28] epiclocks to the methylation data in the main dataset by using the methods and parameters reported by the authors for each case and validated using the implementations in the pyaging Python library [42]. In the case of the Horvath clock, this implies computing an initial value using a linear model and then applying additional transformations for the "young" ($\leq 20$) and "adult" ($> 20$) subjects [15]. Similarly, the cAge model uses one model to predict *age* and another to predict log(*age*); if the *age* prediction is $\leq 20$, then it is replaced by the exponential of the log(*age*) prediction [28].

### Training of the new epigenetic clock

We trained a general epigenetic age prediction model using the beta values of 10,000 probes and the squared beta values of 300 probes to perform a regularized linear regression of chronological age using elastic net [43]. The linear and quadratic features were chosen based on an EWAS study [28]: we chose the 10,000 probes with a significant linear association with age with the lowest association p-value and the 300 probes with a significant quadratic association with age with the lowest association p-value.

To train our general epigenetic age prediction model, we separated the main dataset into training and test sets using a 70/30 split stratified by age (i.e., all ages present in the dataset were sampled in both the training and test sets). Elastic net combines L1 and L2 regularizations, which in the scikit-learn implementation [44] are mixed in proportions given by the parameter L1_ratio, which takes values between 0 and 1. This parameter multiplies the L1 penalty term whereas the L2 term is multiplied by 1-L1_ratio. Based on its performance in previous studies, we set the L1_ratio to 0.5 [15, 45]. The value of the parameter alpha (i.e., 1/C), which multiplies both the L1 and L2 penalties, was optimized to 0.001 through fivefold cross-validation.

Age-specific models were trained using the same approach on 3 different batches of data with subjects of specific ages: age $\leq 46$ (young), $26 < \text{age} < 69$ (middle-aged), and age $\geq 59$ (old). The alpha value was set to 0.001 in all cases. The three age-specific models were then used to perform predictions on nonoverlapping *putative* age (age values assigned by the general model) groups: age $\leq 36$ (young), $36 < \text{age} < 59$ (middle-aged), and age $\geq 59$ (old).
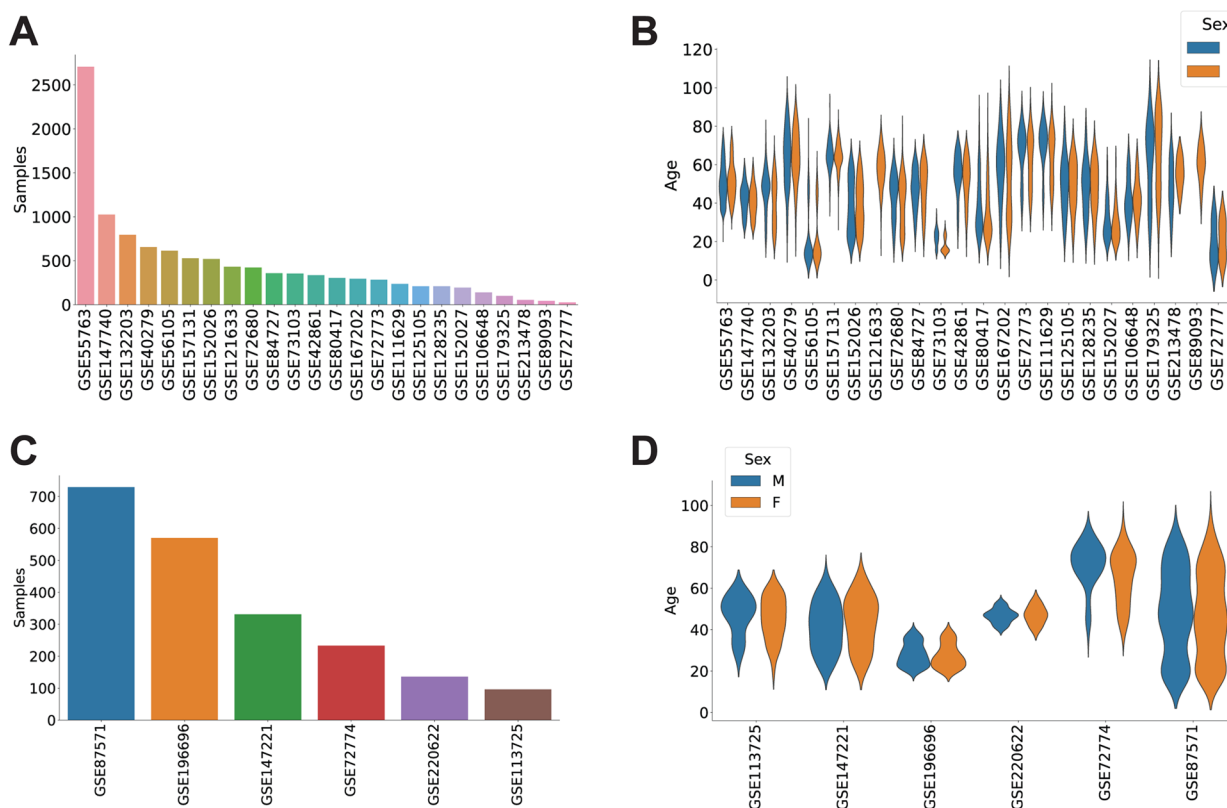
**Fig. 1** Public DNA methylation datasets compiled for the study. **A** Number of subjects present in each of the datasets included in the main set. **B** Distribution of ages and sex in the datasets included in the main set. **C** and **D** Distribution of the number of subjects and their age and sex in the datasets used for the validation set

## Results

### DNA methylation datasets

We explored public repositories to assemble a large and diverse DNA methylation dataset to examine the performance of existing epigenetic clocks and to train and test new models. We used public data to generate two separate datasets: (1) a main dataset for evaluating existing models and training new ones and (2) a validation dataset to test the generalizability of new models.

The main dataset included whole blood DNA methylation data of human subjects from 24 previous studies (Additional file 1: Table S1). We filtered the data to retain exclusively subjects labeled as controls, with reported age and sex values, resulting in a total of 11,825 individuals. To be able to split the data in a stratified fashion into training and test sets, we kept only the samples with an age value present at least twice in the dataset, leaving a total of 10,835 subjects (Fig. 1A). The subjects in this main dataset had ages between 8 and 96 years, with a mean of 47.68 years and an almost even distribution of sexes (5456 females and 5379 males) (Fig. 1B).

The validation dataset was composed of data from 6 additional studies (Additional file 1: Table S1). We selected only control subjects with reported age and sex values, which resulted in a total of 2095 individuals. The age range in this validation set was between 14 and 94 years, and the sex distribution was slightly biased, with 828 females and 1267 males (Fig. 1C–D).

### Existing epigenetic clocks generate distorted results from EPICv2 data

Although more than 80% of the probes used by existing epiclocks are conserved in the EPICv2 methylation array [36], applying these models to data limited to the probes present in the EPICv2 array results in differences in epigenetic age predictions. We applied four different models (Horvath [15], Hannum [14], PhenoAge [16], and the recent chronological age prediction model (cAge) from Bernabeu et al. [28]) to the main dataset of 10,835 individuals, using either all the required probes (*complete* models) or only those present in the new EPICv2 array (*truncated* models). The proportion of probes relevant for these models in the EPICv2 ranged from 84.5% (Hannum) to 97.1% (cAge).
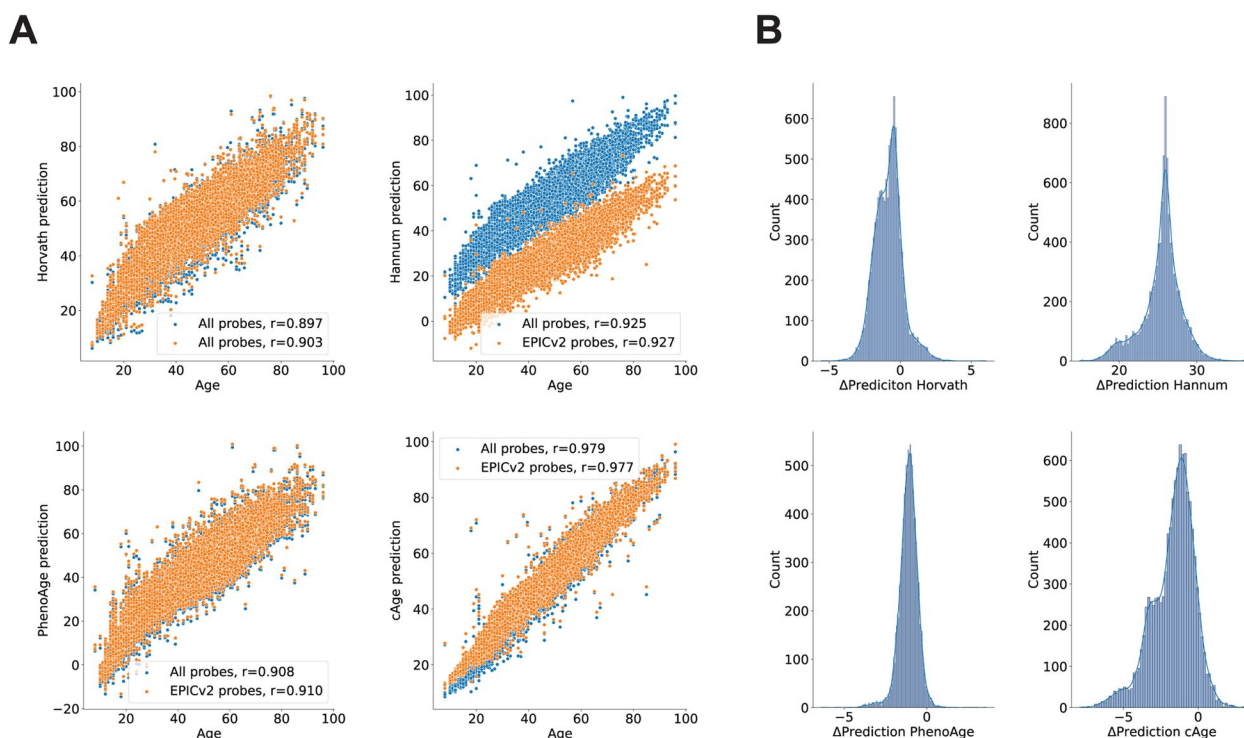
**Fig. 2** Epigenetic clock age prediction drifts caused by the loss of probes in the EPICv2 microarray. **A** Epigenetic age predictions by four existing epigenetic clocks in 10,835 subjects using either their complete sets of CpG probes (blue) or only those available in the EPICv2 array (orange). **B** Distribution of differences between the values predicted using the full sets of probes or only those present in the EPICv2 array (drift) for the same four epigenetic clocks

We observed a high degree of correlation between chronological age and the predicted biological age in all cases, with cAge obtaining the highest value ($r = 0.979$). After the models were truncated to the EPICv2 probe-set, they maintained this correlation, which was still strong ($r > 0.88$) in all cases (Fig. 2A).

However, limiting the number of probes distorted the epigenetic age predictions in all the models. In all four epiclocks, we observed a significant difference ($p < 0.001$, one sample $t$-test) in the biological age predicted by the complete and the truncated models (Fig. 2B). This effect was greater on the Hannum clock model, with an average difference of 25.45 years. The larger models PhenoAge (513 parameters) and cAge (4058 parameters) had average differences of $-1.11$ and $-1.67$ years, respectively, whereas the average difference on the Horvath clock was just $-0.82$ years.

To characterize the performance of these models in predicting chronological age, we computed the Pearson correlation coefficient between the predictions and chronological age ($r$), the mean squared error (MSE), and the median absolute error (MAE) for each of them. We computed the mean ($\mu_{EAA}$) and standard deviations ($\sigma_{EAA}$) of the EAA distributions associated to each epigenetic (Table 1).

**A new epigenetic clock compatible with the EPICv2 array**

The changes in the CpG probes present in the EPICv2 array produce significant alterations in the results of existing epigenetic clocks. To overcome this limitation and to produce robust biological age predictions using EPICv2 data, we trained a new epiclock model using CpG probes common to the 450 k, EPICv1, and EPICv2 methylation arrays. Based on the epigenome-wide association study (EWAS) from Bernabeu et al. [28], there are 42,728 probes with a significant ($p$-adjusted < 0.05) linear association with chronological age that are present in the three different methylation arrays. Similarly, according to the EWAS, there are 63,324 CpG probes with significant ($p$-adjusted < 0.05) quadratic associations with age which are present in the three microarray models (Fig. 3A).

Following the strategy employed by Bernabeu et al. [28], we ranked the probes by their p-value on the EWAS and preselected the first 10,000 with linear and the first 300 with quadratic associations to chronological age to train an epiclock using elastic net regression [43] (Additional file 4: Table S4).

We split the main dataset with 10,835 subjects into training and test sets using a 70/30 split, resulting in 7259 and 3576 samples, respectively, and trained a general

**Table 1** Metrics from all the complete (C) and truncated (T) models on the different datasets used in the study

| | | Hannum$_C$ | Hannum$_T$ | Horvath$_C$ | Horvath$_T$ | PhenoAge$_C$ | PhenoAge$_T$ | cAge$_C$ | cAge$_T$ | General model | Combined model |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Main dataset | $r$ | 0.92 | 0.93 | 0.90 | 0.90 | 0.91 | 0.91 | 0.98 | 0.98 | N/A | N/A |
| | MAE (years) | 7.58 | 19.08 | 6.37 | 6.65 | 7.05 | 6.44 | 2.55 | 3.57 | N/A | N/A |
| | MSE (years) | 81.58 | 410.80 | 65.85 | 69.57 | 78.43 | 66.92 | 12.82 | 20.88 | N/A | N/A |
| | $\mu_{EAA}$ (years) | 6.38 | −19.06 | 3.32 | 4.14 | −5.25 | −4.14 | 1.07 | 2.73 | N/A | N/A |
| | $\sigma_{EAA}$ (years) | 6.39 | 6.89 | 7.40 | 7.24 | 7.13 | 7.06 | 3.42 | 3.67 | N/A | N/A |
| Validation dataset | $r$ | 0.96 | 0.97 | 0.94 | 0.94 | 0.94 | 0.94 | 0.99 | 0.99 | 0.99 | 0.99 |
| | MAE (years) | 7.04 | 18.93 | 6.10 | 6.59 | 7.90 | 7.16 | 1.94 | 2.21 | 2.08 | 1.74 |
| | MSE (years) | 68.45 | 397.10 | 55.64 | 63.44 | 89.60 | 74.98 | 6.41 | 8.32 | 7.55 | 5.96 |
| | $\mu_{EAA}$ (years) | 6.21 | −18.92 | 3.08 | 4.24 | −6.85 | −5.86 | 0.01 | 1.37 | −0.54 | −0.21 |
| | $\sigma_{EAA}$ (years) | 5.47 | 6.27 | 6.80 | 6.74 | 6.53 | 6.38 | 2.53 | 2.54 | 2.70 | 2.43 |
| Koncevičius et al | $\sigma_{noise}$ (years) | 1.18 | 1.01 | 1.26 | 1.19 | 1.13 | 1.09 | 0.78 | 0.86 | 0.46 | 0.50 |
| | $\sigma_{subject}$ (years) | 2.20 | 1.74 | 1.99 | 1.85 | 2.37 | 2.31 | 0.98 | 1.04 | 0.89 | 0.83 |
| Apsley et al | $\sigma_{subject}$ (years) | 2.06 | 1.77 | 3.31 | 3.28 | 2.61 | 2.58 | 0.99 | 1.05 | 1.30 | 0.88 |
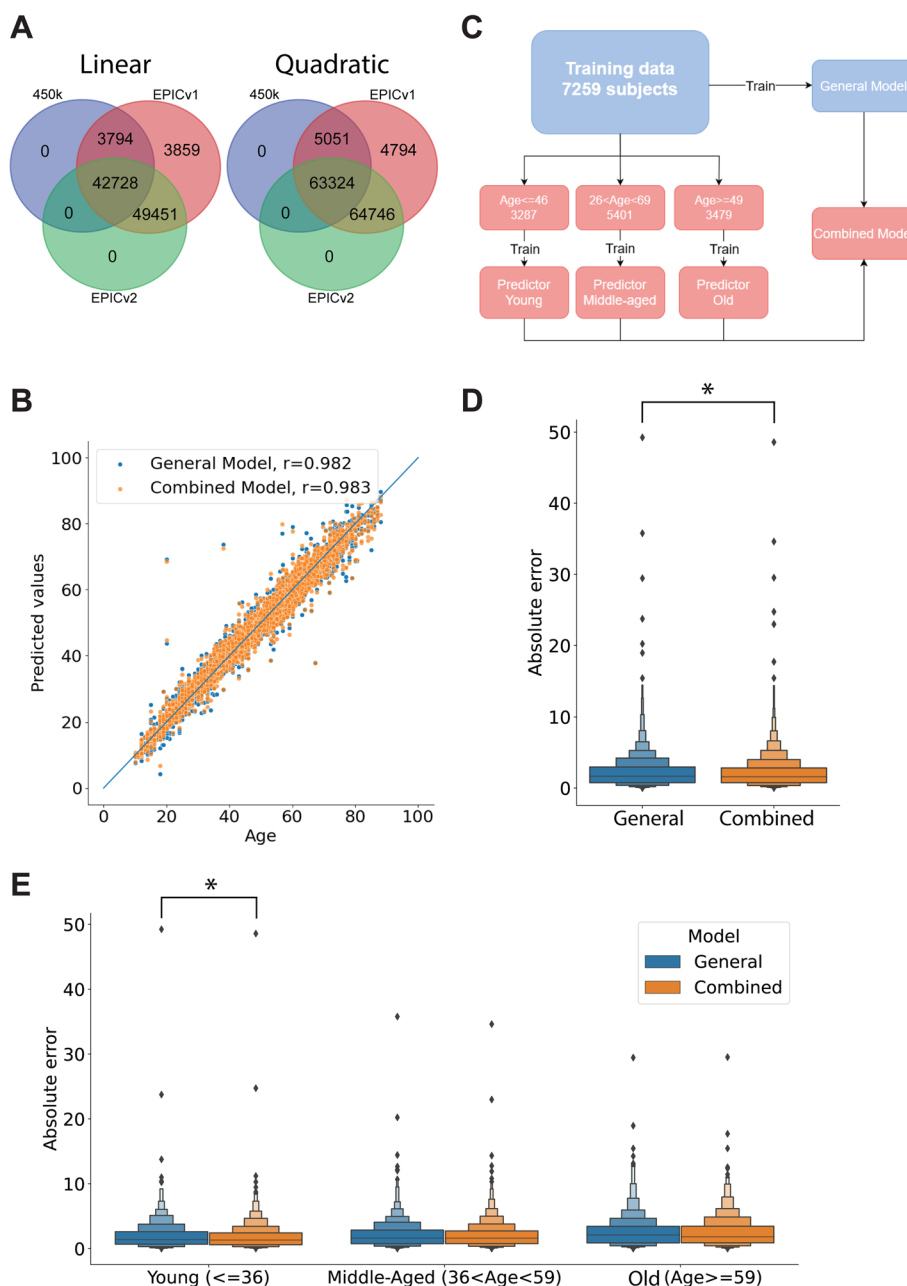
**Fig. 3** Results of the new epigenetic clocks in the test dataset. **A** Methylation probes with significant linear (left) and quadratic (right) associations with chronological age in the EWAS from Bernabeu et al. present in the 450 k, EPICv1, and EPICv2 arrays. **B** Prediction results on the test set by the General and Combined models. The blue line indicates the 1:1 correspondence. **C**. Schematic overview of the training of the different age prediction models. **D** Absolute error of the ages predicted in the test set by the general and combined models. **E** Absolute error from the predictions on the test set by each model, broken down by age group. *Indicates significant differences ($p < 0.05$, one-sided Wilcoxon rank-sum test) between groups

age-prediction model. The model used an L1 ratio of 0.5 [28], and the alpha value was optimized to 0.001 through fivefold cross-validation. This general model obtained a low prediction error (mean squared error (MSE) = 3.14 years) and a high correlation between the predicted and chronological ages ($r = 0.982$) in the test set (Fig. 3B).

Previous works have suggested that the relationship between methylation state and age is nonlinear [14, 28, 46–49]. Therefore, we decided to stratify our training data

into three age groups and train separate predictors for "young," "middle-aged," and "old" subjects. The predictors were intended to work on nonoverlapping age groups, but we did use overlapping age ranges during training to limit the inconsistencies between models assigned to contiguous groups. We produced a combined model in which the general model assigned a sample to one of the three age groups, and then the corresponding age-stratified model generated the age prediction (Fig. 3C). This combined model reduced the age prediction error significantly ($p < 0.05$, one-sided Wilcoxon rank-sum test) with respect to the general model (Fig. 3B, D), lowering the MSE to 3.06 years in the test set and increasing the correlation between the predicted and chronological ages to 0.983. Although the combined model reduced the error in all age groups, the difference was statistically significant ($p < 0.05$, one-sided Wilcoxon rank-sum test) only in the younger subjects (Fig. 3E). Weights for the general and age-specific models are provided in Additional file 5: Tables S5 to S8. The overlap of EPICv2 probes used by our general and combined models and by the rest of the epiclocks examined is presented schematically in Additional file 6: Fig. S1A and B. To underscore the numerical differences between models, we present the coefficients used for the overlapping probes in our models, cAge and the Horvath clock in Additional file 6: Fig. S1C.

### The performance of the new clock is *on par* with that of a state-of-the-art model

To examine the performance of our model, we compared it against cAge, a refined model that has shown significant improvement over previous epigenetic clocks [28]. We tested the complete cAge and our general and combined models on a validation dataset with 2095 samples from 6 studies. We included all the probes used for cAge and those required by our models and predicted the age of each sample based on the methylation data. In all cases, the predicted ages were strongly correlated with the chronological ages (Fig. 4A).

Our general model performed as well as cAge, with no significant difference ($p > 0.05$, one-sided Wilcoxon rank-sum test) in the distribution of errors between the two, whereas the error from the predictions of our combined (age-segregated) model was significantly lower ($p > 0.05$, one-sided Wilcoxon test) (Fig. 4B). By breaking down the data by age group, we could observe differences in the prediction error. In the young ($\leq 36$) and old ($\geq 59$) groups, the error from the combined model was significantly lower ($p > 0.05$, one-sided Wilcoxon test) than that of the cAge and the general models. In the group of middle-aged subjects, the error of the general model was significantly greater ($p > 0.05$, one-sided Wilcoxon test) than that of the other two models (Fig. 4C).

To be able to perform a more general comparison, we computed again the values of r, MSE, MAE, $\mu_{EAA}$, and $\sigma_{EAA}$ for our models and of the other four epigenetic clocks using the validation dataset (Table 1, Additional file 6: Fig. S2). These metrics were largely compatible with those obtained on the main dataset, with the exception of cAge, which had a large decrease in MAE, MSE, $\mu_{EAA}$, and $\sigma_{EAA}$.

### Technical and intra-subject EAA variations

It has recently been shown that epigenetic clocks are affected by spontaneous variation in DNA methylation values, leading to changes in epigenetic age predictions [37, 38]. Additionally, the experimental process to obtain methylation values might introduce another source of variation. In order to establish the magnitude of these variations, we applied our model and the four epiclocks discussed above to data obtained from repeated sampling of the same subjects.

First, we examined the data from the study of Koncevičius et al., which reported DNA methylation of blood samples of a 52-year-old male subject taken every 3 h over a period of 72 h [37] (GSE247197). This dataset included technical replicates, which we used to estimate the EAA deviations associated with technical noise, i.e., variations across multiple experimentally determined beta values on the same sample. All the (complete) models tested exhibited variation across technical replicates (Fig. 5A). To characterize this variation, we subtracted the mean value obtained on each set of replicates and then computed the standard deviation across all the mean-centered values. We interpreted the resulting measure as an estimate of the technical noise ($\sigma_{noise}$). The predictions of our general model had the lowest $\sigma_{noise}$ (0.78 years), whereas Horvath's model had the highest, reaching 1.26 years (Fig. 5B, Table 1). We then computed the standard deviation across all samples in the dataset as a first estimate for the intra-subject variability ($\sigma_{subject}$). The values of $\sigma_{subject}$ were in all cases larger than $\sigma_{noise}$, and in this case, the largest and lowest values corresponded to our combined model (0.83 years) and to PhenoAge (2.36 years), respectively (Fig. 5B, Table 1).

We then applied all the models to the data from Apsley et al., discarding the samples from subjects under the stress test. This data includes DNA methylation data from blood drawn from 31 subjects at 4 different time points over a period of 4 h and 45 min under stress or control conditions [38] (GSE227809). Using data from the 14 subjects in control condition, we computed the standard deviation across the mean-subtracted predictions of all subjects for all the models. We interpreted this measure as another estimate of $\sigma_{subject}$. The results were largely compatible with the estimate of $\sigma_{subject}$ from Koncevičius'
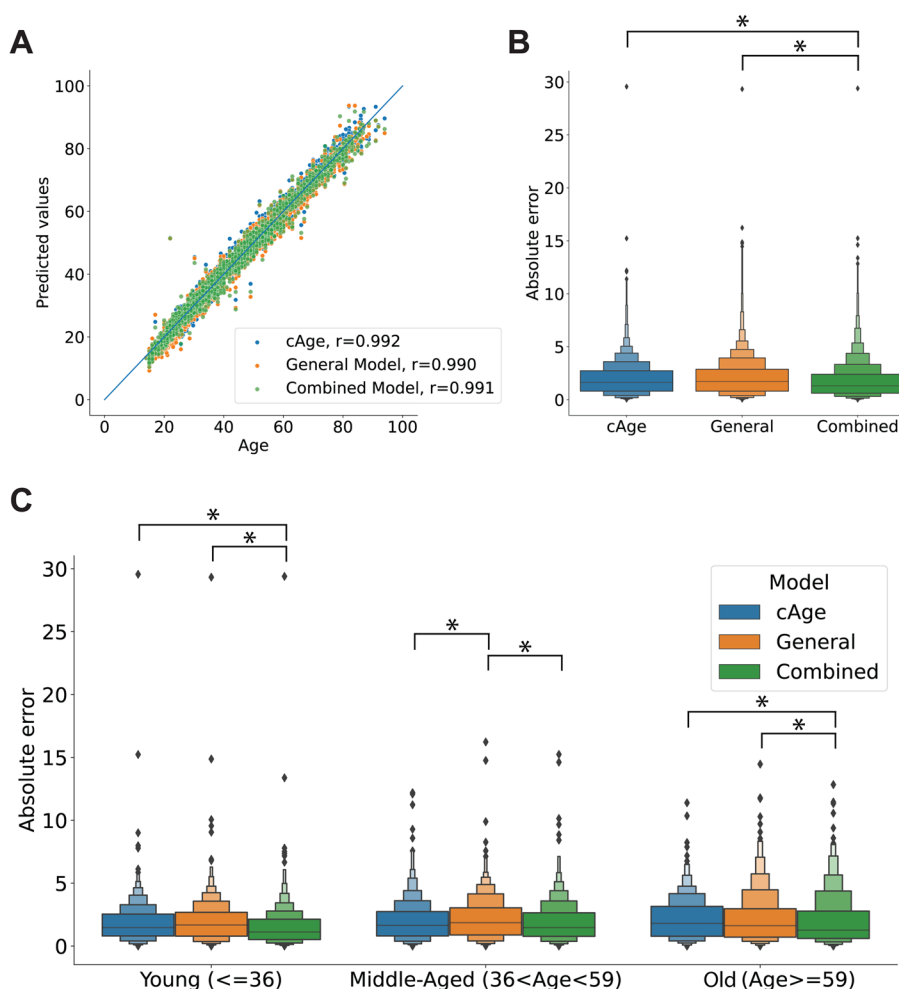
**Fig. 4** Age predictions on the validation set. **A** Chronological age and age values predicted by cAge and the general and combined models in the validation set. **B** Distribution of the absolute prediction error for each of the three models. **C** Distribution of the absolute error of each model in the validation set broken down by age group. *Indicates significant differences ($p < 0.05$, one-sided Wilcoxon rank-sum test) between groups

data; the largest difference was observed on Horvath's model, for which $\sigma_{subject}$ rose from 1.99 to 3.31 years. In all the other models, $\sigma_{subject}$ changed by less than 1 year. Also in this dataset, our combined model's predictions suffered the least variation, whereas those from Horvath's model changed the most (Fig. 5C, Table 1).

According to these results, the sum of technical and intra-subject variation of the EAA predictions represents ~50 to 60% of the population-wide $\sigma_{EAA}$ depending on the model. Taken together, the metrics in Table 1 provide a means to judge whether the magnitude of EAA findings puts them beyond the range of normal variations.

### The new epigenetic clock reflects the influence of radiation therapy and breast cancer

Several studies have employed different epigenetic clocks to assess the impact of pathologies and environmental

factors on epigenetic age acceleration (EAA) [50]. Using PhenoAge [16], Qin et al. demonstrated that exposure to different anticancer treatments had a significant influence on the EAA of childhood cancer survivors [51]. To validate the performance of our model, we tested it in public datasets that have already shown increased EAA in cancer patients. Using data made public in a subsequent study by Dong et al. [52] (GEO accession number GSE197674) which contained data from 2138 childhood cancer survivors, we studied the influence of radiation therapy (RT) on the EAA determined by cAge and our models. In this dataset, all the models produced epigenetic age predictions highly correlated with chronological age, with r values above 0.9 in all cases (Fig. 6A). When comparing subjects who had received RT in one or more body areas (chest, abdomen or pelvis, brain) to those who had not been exposed to RT, all models revealed that the latter group had a significantly lower
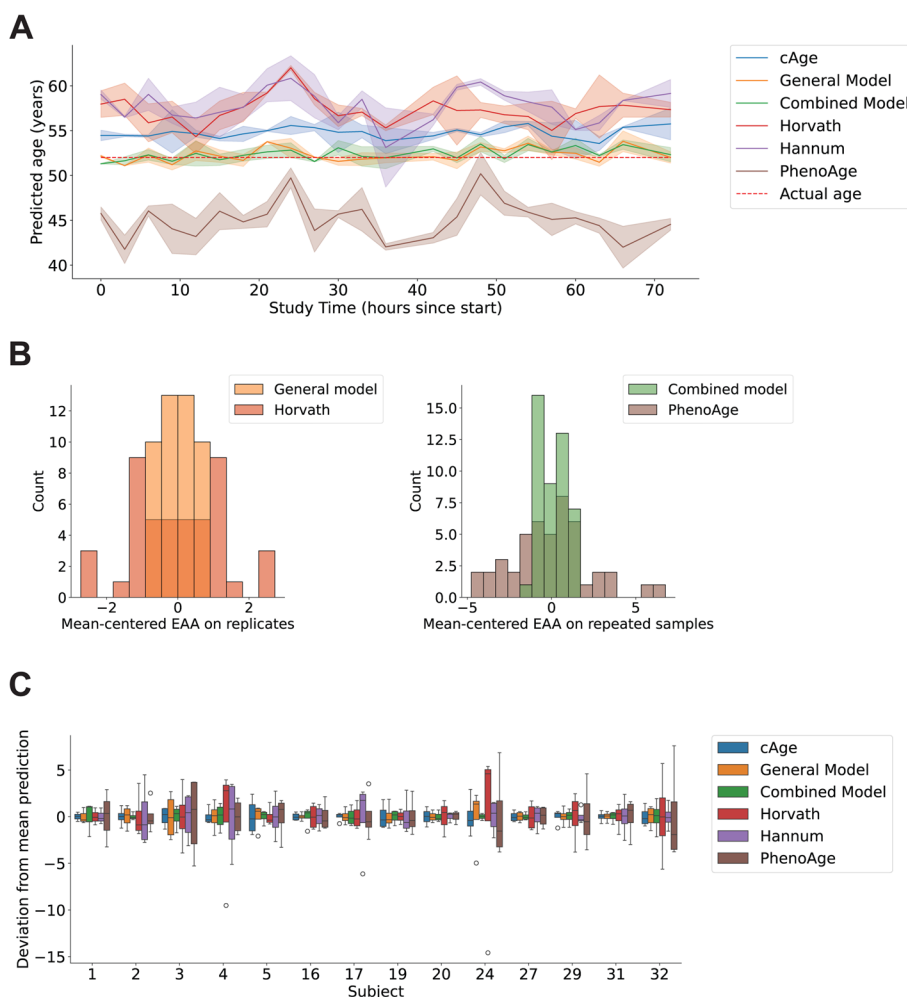
**Fig. 5** Age predictions on technical replicates and repeated samples. **A** Age values predicted by the different models on the dataset from Koncevičius et al., with multiple samples of the same subject (52-year-old male) obtained at different times and with replicates. Solid lines indicate the mean values, whereas the colored area marks the 95% confidence intervals. **B** Distribution of mean-centered EAA values on technical replicates (left) and repeated samples from the same subject (right) on the Koncevičius et al. data for the models with the highest and lowest standard deviations. **C** Distribution of mean-centered age predictions for the 14 subjects from the Apsley et al.'s dataset

($p < 0.05$, one-sided Wilcoxon test) EAA (Fig. 6A). This is in line with the results reported in the original study [51], and similar observations have been made in other studies in which radiation exposure was associated with an increased EAA [53, 54]. The rest of the complete and truncated models assign very large EAA values to virtually all samples and do not attribute significantly larger EAAs ($p > 0.05$ in all cases, one-sided Wilcoxon test) to the groups with RT treatments (Additional file 6: Fig. S3). In all cases, the magnitude of these EAA changes is smaller than the $\sigma_{EAA}$ in the general population (Table 2). Therefore, it is questionable whether the observed EAA differences (here as well as in previous studies) are a true, biologically relevant, reflection of the RT treatment during childhood.

Next, we leveraged a dataset with DNA methylation data from peripheral blood leukocytes of sporadic breast cancer patients and control subjects [55] (GEO accession number GSE148663) to study the effect of the disease on the EAA determined by the different models. We computed the predicted epigenetic age for 22 sporadic breast cancer patients and 10 controls using our models and cAge. Although the correlation between the epigenetic age predictions and the chronological ages of the subjects remained high (0.914–0.951), our models predicted a significantly greater ($p < 0.05$, one-sided Wilcoxon test) EAA in the cancer group (Fig. 6B). As reported in the original study [55], the cancer patients had no previous cancer history and were sampled at the time of diagnosis, so the difference in EAA cannot be attributed to anticancer
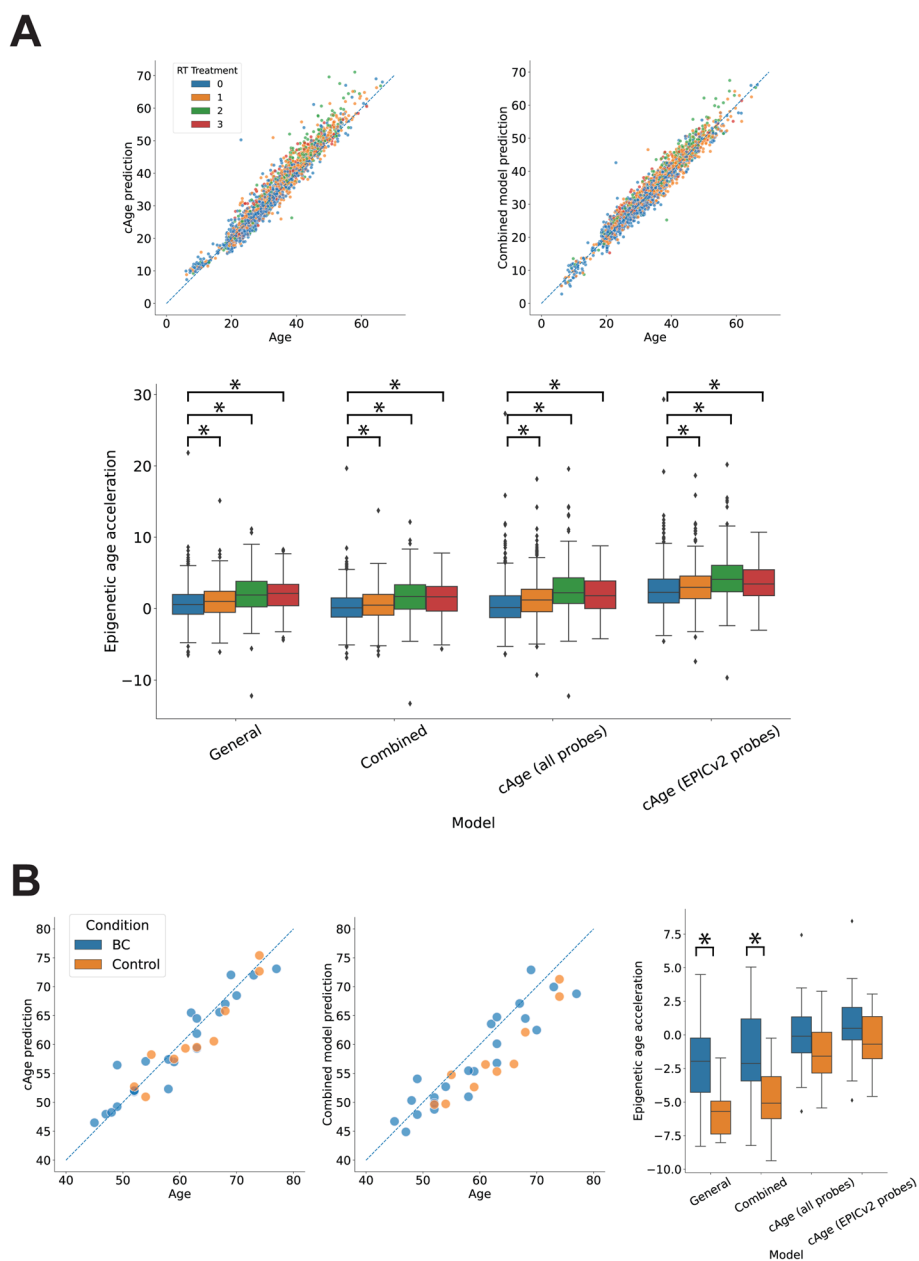
**Fig. 6** Effects of radiation therapy and spontaneous breast cancer in EAA. **A** Chronological age and predicted age for childhood cancer survivors, colored by the number of body areas (brain, chest, abdomen/pelvis) in which they received radiation therapy (top). Distribution of the EAA for each group of subjects determined by each of the models (bottom). **B** Chronological ages of spontaneous breast cancer patients (BC) and control subjects and the corresponding epigenetic ages determined by the combined model and cAge (left). Distribution of the EAA estimated for each group by the different models (right)

therapy but rather to the disease itself. The rest of the complete and truncated models do not assign a significantly greater EAA ($p > 0.05$ in all cases, one-sided Wilcoxon test) to the breast cancer group (Additional file 6: Fig. S4). In this case, the magnitude of the EAA differences between cancer patients and controls determined by our models are larger than the $\sigma_{EAA}$. Therefore, we can

consider that these changes do reflect differences larger than the normal variations across individuals in the general population (Table 3).

Finally, we explored data from an interventional study using resistance and aerobic training [56] (GEO accession number GSE213363). Besides its numerous benefits for overall health [57, 58] and aging [59, 60], physical activity

**Table 2** Differences between average EAA of the RT > 0 and RT = 0 groups observed in the different complete (C) and truncated (T) models

| | $Hannum_C$ | $Hannum_T$ | $Horvath_C$ | $Horvath_T$ | $PhenoAge_C$ | $PhenoAge_T$ | $cAge_C$ | $cAge_T$ | General model | Combined model |
|---|---|---|---|---|---|---|---|---|---|---|
| $EAA_{RT1-RT0}$ (years) | −0.82 | −1.23 | −1.65 | −1.65 | −1.70 | −1.70 | 0.96 | 0.54 | 0.39 | 0.39 |
| $EAA_{RT2-RT0}$ (years) | −0.19 | −1.07 | −1.68 | −1.68 | −1.87 | −1.87 | 2.37 | 1.73 | 1.34 | 1.48 |
| $EAA_{RT3-RT0}$ (years) | 0.35 | −0.15 | −0.65 | −0.65 | −0.35 | −0.35 | 2.05 | 1.46 | 1.43 | 1.44 |

has demonstrated an influence on DNA methylation [61, 62]. Yet, the study by Furtado et al. reported that after a 16-week exercise intervention, they did not detect significant changes in the epigenetic age calculated by the Horvath model [56]. We applied all the models discussed here to the data from subjects before and after the intervention and found the same result: none of the epiclocks showed significant reductions in epigenetic age after the intervention ($p > 0.05$ in all cases, one-sided Wilcoxon test), neither in the resistance nor in the aerobic training groups (Additional file 6: Fig. S5). This suggests that although the first-generation models examined here can predict chronological age with considerable accuracy, they are less responsive in terms of ability to reflect other biological factors or the effect of certain interventions.

## Discussion

Epigenetic clocks are valuable tools for research on aging and pathological states. In this work, we evaluated the applicability of existing epigenetic clocks to the data generated by EPICv2, the newest methylation microarray model from Illumina. The EPICv2 will phase out previous microarray models (namely, 450 k and EPIC), which were used for the development of some widely used epigenetic clocks (e.g., Hannum, Horvath). As the EPICv2 discontinued the use of some of the probes used by existing epiclocks [36], we felt urged to test whether this would affect the performance of the models.

We generated a large DNA methylation dataset by compiling data from whole blood samples obtained in 24 different studies. This dataset allowed us to quantify the effect of running 4 different epiclocks (Hannum [14],

Horvath [15], PhenoAge [16], and the cAge model from Bernabeu et al. [28]) restricted to the set of methylation probes present in the EPICv2 methylation array. Our observations indicate that the results of these 4 models are significantly altered by probes absent from the EPICv2 array.

Epigenetic age models are routinely applied in epigenetic research and are also exploited commercially [63–65]. In the face of commercial discontinuation of the 450 k and EPICv1 arrays, our findings suggest that both researchers and commercial vendors alike will need to update their epigenetic clock models to make them compatible with EPICv2 data. Future studies and commercial solutions using this new microarray will require readjusted or new epigenetic age models, which are currently lacking.

As our results show, *none* of the epiclocks tested work as intended on data generated by the EPICv2 microarrays, showing significant differences between the epigenetic ages predicted using data from different chip models. Therefore, we sought to produce a model compatible across the 450 k, EPICv1, and EPICv2 microarray platforms, aiming to obtain results in accordance with those of existing methods rather than outperforming them. The approach that we used closely followed the one applied by Bernabeu et al. in the development of their cAge model [28], so that we would rely on established methodologies. Our results on the training and test datasets indicate that our epiclock offered very high performance in the prediction of ages from DNA methylation values. Similar to Bernabeu's model, these results highlight the benefit of using feature preselection, nonlinear terms, and

**Table 3** Average EAA differences between cancer patients and healthy controls observed in the different complete (C) and truncated (T) models

| | $Hannum_C$ | $Hannum_T$ | $Horvath_C$ | $Horvath_T$ | $PhenoAge_C$ | $PhenoAge_T$ | $cAge_C$ | $cAge_T$ | General model | Combined model |
|---|---|---|---|---|---|---|---|---|---|---|
| $EAA_{BC}-EAA_{Control}$ (years) | 2.41 | 2.46 | −0.03 | −0.03 | −0.76 | −0.76 | 1.30 | 1.27 | 3.37 | 3.12 |

age-based stratification. We followed a simple strategy regarding these aspects, as we relied on a prior EWAS study [28] to preselect features, we considered only linear and quadratic features, and we stratified our subjects into three arbitrary age groups. We are confident that follow-up work can improve the results presented here by, for example, conducting more thorough feature selection, deriving more features from the original beta values, and/or exploring different data stratification strategies.

Despite its simplicity, our implementation outperformed cAge in the validation dataset, where it obtained significantly lower prediction errors. These results demonstrate that our model can be generalizable (as it is applicable to new data) and that its performance is on par with that of state-of-the-art models.

We tested the variability of epigenetic age predictions on two datasets with repeated samples from the same subjects obtained hours apart from each other. The results indicated that the predictions of cAge and our combined model had the lowest variations across replicates and across repeated samples. We claim that these spontaneous variations of epigenetic age predictions should be considered when examining EAA-related findings: differences smaller than those seen on repeated samples of the same subject obtained on the same day could hardly be considered biologically relevant. Consequently, the significance of previous reports of altered EAAs based on the models d here might need to be re-evaluated.

To demonstrate the ability of our model to reproduce the results of previous models in the detection of EAA alterations, we first applied it to methylation data from cancer survivors [51]. The results of our epiclock indicated an increased EAA induced by radiation therapy, in agreement with previous studies [53, 54]. However, the magnitude of this increase (0.39 to 1.44 years) was smaller than the variation we observed in the general population ($\sigma_{EAA}$). Thus, we cannot claim that the EAA changes detected are a reflection of the RT effect on the subjects.

Next, we applied our model to data obtained from breast cancer patients. Our epiclock revealed a significant increase in the EAA of cancer patients compared to that of control subjects, suggesting that the epigenetic age predicted by the model could be sensitive to this particular pathology. In this case, the average EAA difference between the cancer patients and the control group (3.13–3.37 years) was larger than the $\sigma_{EAA}$ from the non-pathological subjects in the validation dataset (2.43–2.7 years). Therefore, these changes could indeed be a reflection of the pathological state. Notably, the cAge model did not detect a significant difference in the EAA between the control and the cancer groups.

Finally, we analyzed data from an interventional study aimed at improving the health of patients suffering from polycystic ovary syndrome. As in the original study, none of the models we tested reported significant differences in the epigenetic age of the subjects before and after the intervention. Considering the large and extensively documented benefits of physical exercise on health, these results support the idea that despite being good predictors of chronological age, first-generation epigenetic clocks do not necessarily reflect biological factors such as pathological states, environmental exposures, or interventions.

As a relevant limitation of our model, we would like to highlight that it is limited to methylation data obtained from whole blood samples, a constraint that is also shared by multiple other models [28, 66, 67]. Its applicability to data obtained from other tissues has not been assessed. Likewise, we would like to emphasize that its ability to reflect pathological states beyond the sporadic breast cancer cases we have shown here remains to be explored.

Taken together, our results demonstrate that our epigenetic clock is compatible with data generated using the 450 k, EPICv1, and EPICv2 microarray platforms. Its epigenetic age predictions are highly correlated with chronological age in control subjects of all ages. Our model exhibited consistency across technical replicates and across repeated samples of the same subjects, with lower variation than the rest of the models tested. As a first-generation epigenetic clock, our model can predict chronological age from DNA methylation data with high accuracy, but its ability to reflect the effects of environmental exposures, pathological states, or beneficial interventions is limited. Our work solves a technical barrier derived from technological development that has not yet been addressed and has important implications both for aging research and for biotechnology companies offering services in this field.

## Conclusions

Our results reveal that existing epigenetic clocks experience significant distortions due to the transition to the new EPICv2 microarray, underscoring the need for their adaptation to ensure full compatibility with its data. We developed a new epigenetic clock model that is compatible across 450 k, EPICv1, and EPICv2 platforms, achieving superior accuracy in chronological age prediction for the new microarray and even outperforming a state-of-the-art model in validation tests.

To enable a more nuanced interpretation of epigenetic age acceleration (EAA), we established benchmarks by quantifying normal variation, technical noise, and intra-subject variability. These metrics provide a critical context for assessing the biological relevance of observed

EAA changes and suggest that some previously reported alterations may require re-evaluation given these variability thresholds.

Our findings highlight both the potential and limitations of first-generation epigenetic clocks. While these clocks excel in predicting chronological age, their ability to reflect biological factors—such as pathological states, environmental exposures, or lifestyle interventions—appears constrained. Given the higher accuracy of our new model in predicting chronological age and its minimal response to theoretical anti-aging interventions, an important question arises: Is the concept of biological aging actually measurable or modifiable, and is it distinct from chronological age, at least with the current technology? Regardless of the answer, this question necessitates further, more comprehensive studies, which must be conducted using tools as precise as our new model. Future research should also focus on extending EPICv2 compatibility to second-generation models to further refine our understanding of biological aging.

## Abbreviations

EAA    Epigenetic age acceleration
cAge    Chronological age prediction model
EWAS    Epigenome-wide association study
MSE    Mean squared error
RT    Radiation therapy

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-024-01387-4.

---

Additional file 1: Table S1. GEO accession numbers, number of samples, fraction of samples from female subject, median age and related publication doi for each of the datasets in the main and validation sets.

Additional file 2: Table S2. Mean methylation beta value for each probe in the Main set.

Additional file 3: Table S3. Number of probes missing on each dataset on the Main and the Validation sets.

Additional file 4: Table S4. Chromosome, position, associated gene and order for each of the probes selected for training the models.

Additional file 5: Tables S5 to S8. Weights of the general and the age-specific models.

Additional file 6. Supplementary figures S1 to S5.

---

## Availability of data and materials

The datasets analyzed during the current study are available in the GEO repository:
• Main set: GSE55763 [68], GSE157131 [69], GSE147740 [70], GSE132203 [71], GSE40279 [14], GSE56105 [72], GSE152026 [72], GSE121633 [73], GSE72680 [74], GSE84727 [75], GSE73103 [76], GSE42861 [77], GSE72773 [78], GSE80417 [75], GSE167202 [79], GSE111629 [80], GSE125105 [81], GSE128235 [74], GSE152027 [82], GSE106648 [83], GSE179325 [84], GSE213478 [85], GSE72777 [78], GSE89093 [86].
• Validation set: GSE87571 [87], GSE196696 [88], GSE147221 [82], GSE72774 [89], GSE220622 [90], GSE113725 [91].
• Datasets with repeated samples: GSE247197 [37], GSE227809 [38].
• Case studies (cancer survivors, breast cancer, exercise-based intervention): GSE197674 [52], GSE148663 [55], GSE213363 [56].
Further details on the main validation sets are provided in Additional file 1: Table S1.

## Data availability

The manuscript present an analysis of publicly available data. All the data sources are listed and are accessible.

## Declarations

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare no competing interests.

## References

1.  Bocklandt S, et al. Epigenetic predictor of age PloS one. 2011;6: e14821.
2.  Jones MJ, Goodman SJ, Kobor MS. DNA methylation and healthy human aging. Aging Cell. 2015;14:924–32.
3.  Ryan CP. "Epigenetic clocks": Theory and applications in human biology. Am J Hum Biol. 2021;33: e23488.
4.  Oblak L, van der Zaag J, Higgins-Chen AT, Levine ME, Boks MP. A systematic review of biological, social and environmental factors associated with epigenetic clock acceleration. Ageing Res Rev. 2021;69: 101348.
5.  Bozack AK, et al. DNA methylation age at birth and childhood: performance of epigenetic clocks and characteristics associated with epigenetic age acceleration in the Project Viva cohort. Clin Epigenetics. 2023;15:62.
6.  Christiansen L, et al. DNA methylation age is associated with mortality in a longitudinal Danish twin study. Aging Cell. 2016;15:149–54.
7.  Marioni RE, et al. DNA methylation age of blood predicts all-cause mortality in later life. Genome Biol. 2015;16:1–12.
8.  Perna L, et al. Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. Clin Epigenetics. 2016;8:1–7.
9.  Marioni RE, et al. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. Int J Epidemiol. 2015;44:1388–96.
10. Valencia CI, Saunders D, Daw J, Vasquez A. DNA methylation accelerated age as captured by epigenetic clocks influences breast cancer risk. Front Oncol. 2023;13:1029.
11. Berstein FM, et al. Assessing the causal role of epigenetic clocks in the development of multiple cancers: a Mendelian randomization study. Elife. 2022;11: e75374.
12. Rozenblit M, et al. Evidence of accelerated epigenetic aging of breast tissues in patients with breast cancer is driven by CpGs associated with polycomb-related genes. Clin Epigenetics. 2022;14:30.

13. Gehle SC, et al. Accelerated epigenetic aging and myopenia in young adult cancer survivors. Cancer Med. 2023;12:12149–60.

14. Hannum G, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell. 2013;49:359–67.

15. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14:1–20.

16. Levine ME, et al. An epigenetic biomarker of aging for lifespan and healthspan. Aging (Albany NY). 2018;10:573.

17. Belsky DW, et al. DunedinPACE, a DNA methylation biomarker of the pace of aging. Elife. 2022;11: e73420.

18. Lu AT, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging (Albany NY)   2019;11:303.

19. Freire-Aradas A, et al. A common epigenetic clock from childhood to old age. Forensic Sci Int Genet. 2022;60: 102743.

20. Prosz A, et al. Biologically informed deep learning for explainable epigenetic clocks. Sci Rep. 2024;14:1306.

21. Tomusiak, A. et al. Development of a novel epigenetic clock resistant to changes in immune cell composition. 2023. bioRxiv 2023–03.

22. Caulton A, et al. Development of epigenetic clocks for key ruminant species. Genes. 2021;13:96.

23. Zoller JA, et al. DNA methylation clocks for clawed frogs reveal evolutionary conservation of epigenetic aging. GeroScience. 2024;46:945–60.

24. Sala C, et al. Where are we in the implementation of tissue-specific epigenetic clocks? Front Bioinform. 2024;4:1306244.

25. Pośpiech E, Bar A, Pisarek-Pacek A, et al. Epigenetic clock in the aorta and age-related endothelial dysfunction in mice. GeroScience. 2024;46:3993–4002. https://doi.org/10.1007/s11357-024-01086-3.

26. Voisin S, et al. An epigenetic clock for human skeletal muscle. J Cachexia Sarcopenia Muscle. 2020;11:887–98.

27. Coninx E, et al. Hippocampal and cortical tissue-specific epigenetic clocks indicate an increased epigenetic age in a mouse model for Alzheimer's disease. Aging (Albany NY). 2020;12:20817.

28. Bernabeu E, et al. Refining epigenetic prediction of chronological and biological age. Genome Medicine. 2023;15:12.

29. Bibikova M, et al. Genome-wide DNA methylation profiling using Infinium® assay. Epigenomics. 2009;1:177–200.

30. Weisenberger DJ, Van Den Berg D, Pan F, Berman B, Laird P. Comprehensive DNA methylation analysis on the Illumina Infinium assay platform. San Diego: Illumina; 2008.

31. Pidsley R, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol. 2016;17:1–17.

32. Noguera-Castells A, García-Prieto CA, Álvarez-Errico D, Esteller M. Validation of the new EPIC DNA methylation microarray (900K EPIC v2) for high-throughput profiling of the human DNA methylome. Epigenetics. 2023;18:2185742.

33. Sandoval J, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics. 2011;6:692–702.

34. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics. 2016;8:389–99.

35. Horvath S, et al. Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. Aging (Albany NY). 2018;10:1758.

36. Kaur D, et al. Comprehensive evaluation of the Infinium human MethylationEPIC v2 BeadChip. Epigenetics Communications. 2023;3:6.

37. Koncevičius K, et al. Epigenetic age oscillates during the day. Aging Cell. 2024;23:e14170.

38. Apsley AT, et al. Biological stability of DNA methylation measurements over varying intervals of time and in the presence of acute stress. Epigenetics. 2023;18:2230686.

39. Oh G, et al. Circadian oscillations of cytosine modification in humans contribute to epigenetic variability, aging, and complex disease. Genome Biol. 2019;20:1–14.

40. Barrett T, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2012;41:D991–5.

41. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30:207–10.

42. de Lima Camillo LP. pyaging: a Python-based compendium of GPU-optimized aging clocks. Bioinformatics. 2024;40:btae200.

43. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B Stat Methodol. 2005;67:301–20.

44. Pedregosa F, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.

45. Zhang Q, et al. Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. Genome medicine. 2019;11:1–11.

46. Johnson ND, et al. Non-linear patterns in age-related DNA methylation may reflect CD4+ T cell differentiation. Epigenetics. 2017;12:492–503.

47. Okada D, Cheng JH, Zheng C, Kumaki T, Yamada R. Data-driven identification and classification of nonlinear aging patterns reveals the landscape of associations between DNA methylation and aging. Hum Genomics. 2023;17:8.

48. Carlsen L, Holländer O, Danzer MF, Vennemann M, Augustin C. DNA methylation-based age estimation for adults and minors: considering sex-specific differences and non-linear correlations. Int J Legal Med. 2023;137:635–43.

49. Bell CG, et al. DNA methylation aging clocks: challenges and recommendations. Genome Biol. 2019;20:1–24.

50. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat Rev Genet. 2018;19:371–84.

51. Qin N, et al. Epigenetic age acceleration and chronic health conditions among adult survivors of childhood cancer. J Natl Cancer Inst. 2021;113:597–605.

52. Dong Q, et al. Genome-wide association studies identify novel genetic loci for epigenetic age acceleration among survivors of childhood cancer. Genome Medicine. 2022;14:32.

53. Sehl ME, Carroll JE, Horvath S, Bower JE. The acute effects of adjuvant radiation and chemotherapy on peripheral blood epigenetic age in early stage breast cancer patients. NPJ Breast Cancer. 2020;6:23.

54. Xiao C, et al. Epigenetic age acceleration, fatigue, and inflammation in patients undergoing radiation therapy for head and neck cancer: a longitudinal study. Cancer. 2021;127:3361–71.

55. Cappetta M, et al. Discovery of novel DNA methylation biomarkers for non-invasive sporadic breast cancer detection in the Latino population. Mol Oncol. 2021;15:473–86.

56. Miranda Furtado CL, et al. Resistance and aerobic training increases genome-wide DNA methylation in women with polycystic ovary syndrome. Epigenetics. 2024;19:2305082.

57. Ruegsegger GN, Booth FW. Health benefits of exercise. Cold Spring Harb Perspect Med. 2018;8: a029694.

58. Mandolesi L, et al. Effects of physical exercise on cognitive functioning and wellbeing: biological and psychological benefits. Front Psychol. 2018;9:509.

59. DiPietro L. Physical activity in aging: changes in patterns and their relationship to health and function. J Gerontol A Biol Sci Med Sci. 2001;56:13–22.

60. Paterson DH, Jones GR, Rice CL. Ageing and physical activity: evidence to develop exercise recommendations for older adults. Appl Physiol Nutr Metab. 2007;32:S69–108.

61. Grazioli E, et al. Physical activity in the prevention of human diseases: role of epigenetic modifications. BMC Genomics. 2017;18:111–23.

62. Ferioli M, et al. Role of physical exercise in the regulation of epigenetic mechanisms in inflammation, cancer, neurodegenerative diseases, and aging process. J Cell Physiol. 2019;234:14852–64.

63. Biological Age Test | Horvath's Clock | myDNAge. https://mydnage.com/.

64. TruDiagnostic.com. https://www.trudiagnostic.com/.

65. Elysium Health - Healthy Aging Supplements. https://www.elysiumhealth.com/.

66. Lee Y, et al. Blood-based epigenetic estimators of chronological age in human adults using DNA methylation data from the Illumina MethylationEPIC array. BMC Genomics. 2020;21:1–13.

67. Knight AK, et al. An epigenetic clock for gestational age at birth based on blood methylation data. Genome Biol. 2016;17:1–11.

68. Wahl S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature. 2017;541:81–6.

69. Kho M, et al. Epigenetic loci for blood pressure are associated with hypertensive target organ damage in older African Americans from the genetic epidemiology network of Arteriopathy (GENOA) study. BMC Med Genomics. 2020;13:1–10.

70. Robinson O, et al. Determinants of accelerated metabolomic and epigenetic aging in a UK cohort. Aging Cell. 2020;19: e13149.
71. Kilaru V, et al. Critical evaluation of copy number variant calling methods using DNA methylation. Genet Epidemiol. 2020;44:148–58.
72. McRae AF, et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. Genome Biol. 2014;15:1–10.
73. Kurushima Y, Tsai P, Castillo-Fernandez J, et al. Epigenetic findings in periodontitis in UK twins: a cross-sectional study. Clin Epigenetics. 2019;11(1):27.
74. Zannas AS, et al. Epigenetic upregulation of FKBP5 by aging and stress contributes to NF-κB–driven inflammation and cardiovascular risk. Proc Natl Acad Sci. 2019;116:11370–9.
75. Hannon E, et al. An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. Genome Biol. 2016;17:1–16.
76. Voisin S, et al. Many obesity-associated SNPs strongly associate with DNA methylation changes at proximal promoters and enhancers. Genome medicine. 2015;7:1–16.
77. Liu Y, et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. Nat Biotechnol. 2013;31:142–7.
78. Horvath S, et al. An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. Genome Biol. 2016;17:1–23.
79. Konigsberg IR, et al. Host methylation predicts SARS-CoV-2 infection and clinical outcome. Communications medicine. 2021;1:42.
80. Chuang YH, et al. Parkinson's disease is associated with DNA methylation levels in human blood and saliva. Genome medicine. 2017;9:1–12.
81. Arloth J, et al. DeepWAS: Multivariate genotype-phenotype associations by directly integrating regulatory information using deep learning. PLoS Comput Biol. 2020;16: e1007616.
82. Hannon E, et al. DNA methylation meta-analysis reveals cellular alterations in psychosis and markers of treatment-resistant schizophrenia. Elife. 2021;10: e58430.
83. Kular L, et al. DNA methylation as a mediator of HLA-DRB1* 15: 01 and a protective variant in multiple sclerosis. Nat Commun. 2018;9:2397.
84. Barturen G, et al. Whole blood DNA methylation analysis reveals respiratory environmental traits involved in COVID-19 severity following SARS-CoV-2 infection. Nat Commun. 2022;13:4597.
85. Oliva M, et al. DNA methylation QTL mapping across diverse human tissues provides molecular links between genetic variation and complex traits. Nat Genet. 2023;55:112–22.
86. Roos L, et al. Integrative DNA methylome analysis of pan-cancer biomarkers in cancer discordant monozygotic twin-pairs. Clin Epigenetics. 2016;8:1–16.
87. Johansson Å, Enroth S, Gyllensten U. Continuous aging of the human DNA methylome throughout the human lifespan. PLoS ONE. 2013;8: e67378.
88. Webster AP, et al. Donor whole blood DNA methylation is not a strong predictor of acute graft versus host disease in unrelated donor allogeneic haematopoietic cell transplantation. Front Genet. 2024;15:1242636.
89. Horvath S, Ritz BR. Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. Aging. 2015;7:1130.
90. Sánchez-Cabo F, et al. Subclinical atherosclerosis and accelerated epigenetic age mediated by inflammation: a multi-omics study. Eur Heart J. 2023;44:2698–709.
91. Crawford B, et al. DNA methylation and inflammation marker profiles associated with a history of depression. Hum Mol Genet. 2018;27:2840–50.

## Publisher's Note