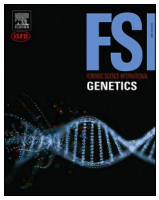
Forensic Science International: Genetics 74 (2025) 103142

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen



Exploring legal age estimation using DNA methylation

M. Boullón-Cassau [a], A. Ambroa-Conde [a], M.A. Casares de Cal [b], A. Gómez-Tato [b], A. Mosquera-Miguel [a], J. Ruiz-Ramírez [a], A. Cabrejas-Olalla [a], J. González-Bao [a], L. Casanova-Adán [a], M. de la Puente [a], A. Rodríguez [a], C. Phillips [a,c], MV. Lareu [a], A. Freire-Aradas [a,*]

[a] Forensic Genetics Unit, Institute of Forensic Sciences, Universidade de Santiago de Compostela, Spain
[b] CITMAga (Center for Mathematical Research and Technology of Galicia), University of Santiago de Compostela, Spain
[c] King's Forensics, Faculty of Life Sciences and Medicine, King's College, London, UK

ARTICLE INFO

Keywords:
DNA methylation
Legal age
Quantile regression
Horvath's skin & blood clock

ABSTRACT

Minors (subjects under the legal age, established at this study at 18 years) benefit from a series of legal rights created to protect them and guarantee their welfare. However, throughout the world there are many minors who have no way to prove they are underaged, leading to a great interest in predicting legal age with the highest possible accuracy. Current methods, mainly involving X-ray analysis, are highly invasive, so new methods to predict legal age are being studied, such as DNA methylation. To further such studies, we created two age prediction models based on five epigenetic markers: cg21572722 (*ELOVL2*), cg02228185 (*ASPA*), cg06639320 (*FHL2*), cg19283806 (*CCDC102B*) and cg07082267 (no associated gene), that were analysed in blood samples to determine possible limitations regarding DNA methylation as an effective tool for legal age estimation. A wide age range prediction model was created using a broad set of samples (14–94 years) yielding a mean absolute error (MAE) of ±4.32 years. A second model, the constrained age prediction model, was created using a reduced range of samples (14–25 years) yielding an MAE of ±1.54 years. Both models, in addition to Horvath's Skin & Blood epigenetic clock, were evaluated using a test set comprising 732 pairs of 18-year-old twins (N=426 monozygotic (MZ) and N=306 dizygotic (DZ) pairs), representing a relevant age of study. Through analysis of the two former age prediction models, we found that constraining the age of the samples forming the training set around the desired age of study significantly reduced the prediction error (from MAE: ±4.07 and ±4.27 years for MZ and DZ twins, respectively; to ±1.31 and ±1.3 years). However, despite low prediction errors, DNA methylation models are still prone to classify same-aged individuals in different categories (minors or adults), despite each sample belonging to the same twin pair. Additional evaluation of Horvath's Skin & Blood model (391 CpGs) led to similar results in terms of age prediction errors than if using only five epigenetic markers (MAE: ±1.87 and ±1.99 years for MZ and DZ twins, respectively).

1. Introduction

Europe as a continent, as well as other countries around the world face a challenge regarding unaccompanied minors who arrive at the borders. As an example, a mean of 30.710 undocumented minors arrived at the Spanish border last year (source: Eurostat, accessed on July 26th, 2024). One of the main problems regarding this issue is that many undocumented minors have no means of proving their legal age (18 years old at the present study). Proving whether they are minors or not is important for two main reasons: firstly, unaccompanied minors fall under the protection of the International Convention on the Rights of the Child [1]. This means they have a series of rights and guarantees for their welfare. Secondly, legal procedures in case of law violation differ depending on the age of the offender. For these reasons, unaccompanied minors who are unable to prove their legal age are submitted to non-medical and medical procedures to establish if they are minors or not [2]. The following are examples of such procedures: i) interviews analysing their conduct and psychological skills; ii) physical examination where anthropometric measures are taken, and sexual maturity is assessed; iii) radiological examination of the carpus from the left hand

* Corresponding author.
E-mail address: ana.freire@usc.es (A. Freire-Aradas).

https://doi.org/10.1016/j.fsigen.2024.103142
Received 27 July 2024; Received in revised form 3 September 2024; Accepted 4 September 2024
Available online 5 September 2024

using as reference the Greulich and Pyle Atlas [3]; iv) dental and radiological examination of the third molar applying the Demirjian system [4] and, v) radiological examination of the clavicle, among others approaches.

As can be seen, many of the medical procedures depend on radiological examination, but as there is no medical rationale for their use and they are highly intrusive methods, these analyses should be avoided. This is the main reason why new ways of predicting legal age, which do not pose a risk for the health of undocumented minors, are required. The European Economic and Social Committee (EESC) proposed a holistic evaluation as a less intrusive alternative, but this evaluation system is experiencing difficulties being implemented [2]. However, there is another procedure whose effectiveness for predicting legal age has been subject to debate in recent years - the use of DNA methylation.

DNA methylation is one of the most studied epigenetic changes that the genome undergoes. Epigenetic marks are molecular alterations in the genome that do not modify the underlying DNA sequence [5]. These changes are the result of the interaction between the genome and the environment. Lifestyle factors such as personal diet, intake of tobacco and/or alcohol and levels of physical activity, among others; can modify the epigenome over time. The nucleotides that may be methylated in the human genome are cytosines followed by a guanine in the 5'-3' direction of the DNA strand. For this reason, these positions are named CpG sites or CpG dinucleotides. Since the complementary strand is also a CpG in the 5'-3' direction, after cellular division the methylated site is present in both daughter strands, allowing its mitotic inheritance. These sites are normally found clustered together in certain parts of the genome called CpG islands [6]. In turn, these islands are usually found near transcription start sites and are associated with promoter regions. DNA methylation has been found to be an epigenetic way of regulating gene expression since methylated promotor regions are associated with silenced genes [7]. It is important to keep in mind that DNA methylation is tissue-dependent, which means that methylation patterns differ depending on what tissue they originate from [8]. Furthermore, in recent years, certain CpG sites have been identified that gradually change their methylation levels over time (hyper- or hypomethylation) [9–13]. Through the study of these age-correlated CpGs, age prediction models can be created to estimate the age of an individual through their DNA.

These age prediction models are termed epigenetic clocks, and the work that generated impetus in the field was Horvath's system [12], a model for age prediction in 51 tissues and cell types. Other models were built in the years following Horvath's early work, especially for the most studied forensic tissues of blood [14–16], saliva and/or buccal cells [15,17,18], semen [19,20] and bones [15,21]. Usually, a prediction error of ±3–4 years is considered an acceptable error for age prediction since it narrows down considerably the possible age of the DNA donor. However, for the prediction of legal age, such prediction error is too high and cannot be used. It would mean anyone between the ages of 14 and 22 could be equally classified as a minor or an adult and the result would be within the accepted error range. For this reason, in this study we built two age prediction models for blood with ages ranging from 14 to 94 years in the first model, and 14–25 in the second. These models were tested to see if they could predict legal age accurately with a total of 1464 18-year-old European twins (732 pairs). The result of this analysis was to determine whether DNA methylation is a suitable candidate to replace the current procedures to establish legal age or not.

2. Materials and methods

2.1. DNA methylation data for training and test sets

In order to create the training set, a total of 729 samples were obtained from dataset GSE87571 [10]. The DNA from these samples had been extracted from white blood cells of a Swedish cohort ranging from 14 to 94 years of age. The percentage of donors from different biological sexes was balanced between females (53.22 %) and males (46.78 %) throughout the whole age range of the sample set. Another cohort was assessed for the test set, using the data from dataset GSE105018 (DNA extracted from whole blood) [22]. From here, a total of 1464 18-year--old twin samples (732 pairs) from the United Kingdom were selected, of which 852 were monozygotic twins (MZ) and 612 were dizygotic (DZ). The percentage of donors from different biological sexes was balanced between females (48.77 %) and males (51.23 %). The DNA samples from both cohorts were bisulphite-converted and analysed using Illumina Infinium 450 K Human Methylation Beadchip to obtain their DNA methylation values.

2.2. CpG site selection

The CpG sites used were taken from the study of Freire-Aradas et al. [23], which created an age prediction model for blood comprising 7 DNA methylation markers. However, the technology used the Agena Bioscience EpiTYPER® system and only 5 of the 7 markers have a CpGID, which is necessary for working with data from Illumina. Since the samples from both the training and test set were studied using Illumina technology, only the 5 markers with a CpGID from this epigenetic clock were used in the present study. These markers are cg21572722 (*ELOVL2*), cg02228185 (*ASPA*), cg06639320 (*FHL2*), cg19283806 (*CCDC102B*) and cg07082267 (no associated gene).

2.3. Statistical analysis

All statistical analyses were performed using R software v. 4.3.1 [24]. The DNA methylation data from the two datasets previously mentioned, was obtained using the *minfi* R package [25]. Since the methylation data was analysed with Illumina Infinium 450 K Human Methylation Beadchip, the *IlluminaHumanMethylation450kmanifest* and *IlluminaHumanMethylation450kanno.ilmn12.hg19* R packages [26,27] were also needed. The age prediction models were built with a quantile regression model using the *quantreg* R package [28]. The quantiles of the quantile regression were established as 0.1 and 0.9. Validation of the prediction models was implemented with a k-fold cross-validation (k=10) using the *cvTools* R package [29]. The following parameters were studied in the k-fold validation: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), percentage of correctly classified individuals (chronological age inside the obtained age prediction intervals) and Spearman correlations between chronological and predicted age in both the training and test sets. Although when using quantiles, the MAE can be represented by the median instead of the mean, the mean was used in the present study for comparative purposes with additional models, where the MAE is usually based on the mean. The graphs representing the predicted age *versus* the chronological age for the models, as well as the graphs representing age predictions of the twin samples, were created using the *ggplot2* [30] and the *ggpubr* [31] R packages. The Horvath's Skin & Blood epigenetic clock [32] was also used to analyse the samples and, to do so, the R package *methylclock* [33] was used. In this case, the percentage of correctly classified individuals based on the age prediction intervals cannot be obtained because of the underlying statistical model. Finally, a leave-one-out cross-validation was performed to assess the correct classification rate of samples under and over 18 years old. In the leave-one-out cross-validation, only MAE, RMSE and percentage of correctly classified individuals were studied.

3. Results

3.1. A wide age range prediction model

A wide age range prediction model for blood was developed using a training set made up of DNA methylation data from 729 samples collected from GSE87571 [10] and the five age-correlated markers selected from the Freire-Aradas study [23]. The corresponding scatter

plots (DNA methylation values in front of chronological age) have been depicted in **Supplementary Fig. S1**. The samples from this training set ranged from 14 to 94 years of age and the developed age prediction model was based on quantile regression. **Supplementary Fig. S2** shows the results for the model representing the predicted age against chronological age. The black continuous line represents the 0.5 quantile while the dark red discontinuous lines represent the prediction intervals (0.1 and 0.9 quantiles). The grey line represents perfect correlation between predicted and chronological age. The training set gave an MAE of ±4.32 years, an RMSE of 5.51 and a percentage of correctly classified individuals of 79.84 %. To evaluate the accuracy of the model, a k-fold cross-validation divided the total samples randomly into k groups of similar size (in this case k=10). One by one, each of the groups is used as a test set while the other k-1 groups are used for the training set. The prediction error was similar to that obtained for the model (MAE: ±4.39 years and RMSE: 5.55) and the percentage of correctly classified individuals was 79.01 %. The correlation values for the training and test sets were 0.97 and 0.96, respectively.

An independent cohort was also used to validate the model. For this new test set, DNA methylation data from 18-year-old samples were used since it is the targeted age of this study. A total of 1464 18-year-old twin blood samples (732 pairs) were selected from GSE105018 [22], out of which 852 were MZ twins and 612 were DZ. The analyses for MZ and DZ twins were carried out separately in case there were differences between them. The predicted ages of each of the pairs of MZ and DZ twins is shown in Fig. 1. The test set comprising 852 MZ samples (426 pairs) gave an MAE of ±4.07 years, an RMSE of 5.17 and a percentage of correctly classified cases of 75 % (Fig. 1A). Out of the 852 total MZ samples, 257 (30.16 %) were predicted to be over 18 and 594 (69.72 %) were predicted to be under 18, while only one individual was predicted as 18 years old. When the test set was built with the 612 DZ samples (306 pairs), an MAE of ±4.27 years, an RMSE of 5.29 and a percentage of correctly classified cases of 74 % was obtained (Fig. 1B). Out of the 612 samples, 196 (32.03 %) were predicted over 18 and 416 samples (67.97 %) were predicted under 18.

There were some predictions where, in the same pair of twins, one was predicted as a minor and the other as an adult. From the total of 732 pairs of twins (if adding both MZ and DZ test sets), 219 of them (29.92 %) displayed this opposite result between individuals. These results are shown in **Supplementary Fig. S3**. When using this model, an MAE of ±3.61 years was obtained between twins belonging to the same pair.

3.2. A constrained age range prediction model

A constrained age range prediction model was developed using only samples ranging between 14 and 25 years for the training set, leaving the model with 163 samples for the training set. **Supplementary Fig. S4** shows the result of the model regarding predicted *versus* chronological age. The MAE for this model was ±1.54 years and the k-fold cross-validation (also with k=10), had a prediction error of MAE: ±1.62 years and an RMSE: 2.04 with a percentage of correctly classified individuals of 75.59 %. The correlation values for the training and test sets were 0.74 and 0.69, respectively.

The test sets of 852 MZ twins and 612 DZ twins (all 18-year-olds) were also tested on this model for validation purposes. The predicted ages of each of the pairs of MZ and DZ twins using the constrained age range model is shown in Fig. 2. The MZ test set resulted in an MAE of ±1.31 years, an RMSE of 1.66 and a percentage of correctly classified individuals of 82 % (Fig. 2A). Out of the 852 samples, 411 (48.24 %) were predicted over 18 years, 436 (51.17 %) were predicted under 18 and 5 (0.59 %) were predicted as exactly 18. Using the DZ test set, an MAE of ±1.3 years, an RMSE of 1.65 and a percentage of correctly classified individuals of 83 % was obtained (Fig. 2B). Out of the 612 total samples, 296 (48.37 %) were predicted over 18 and 316 (51.63 %) were predicted under 18.

This model also presented predictions where, in the same pair of twins, one was predicted as a minor and the other as an adult. Of 732 pairs of twins, this model predicted one twin as a minor and the other as an adult in 221 pairs of twins (30.19 %). These results are shown in **Supplementary Fig. S5**. When using this model, an MAE of ±1.21 years was obtained between twins belonging to the same pair.

3.3. The Horvath's Skin & Blood epigenetic clock

The Skin & Blood age prediction model created by Horvath et al. [32], which contains 391 CpGs, was additionally used to analyse the 18-year-old twin samples. Since this model is already validated, k-fold cross validation was not performed. The predicted ages of each of the pairs of MZ and DZ twins using the Horvath's Skin & Blood model are

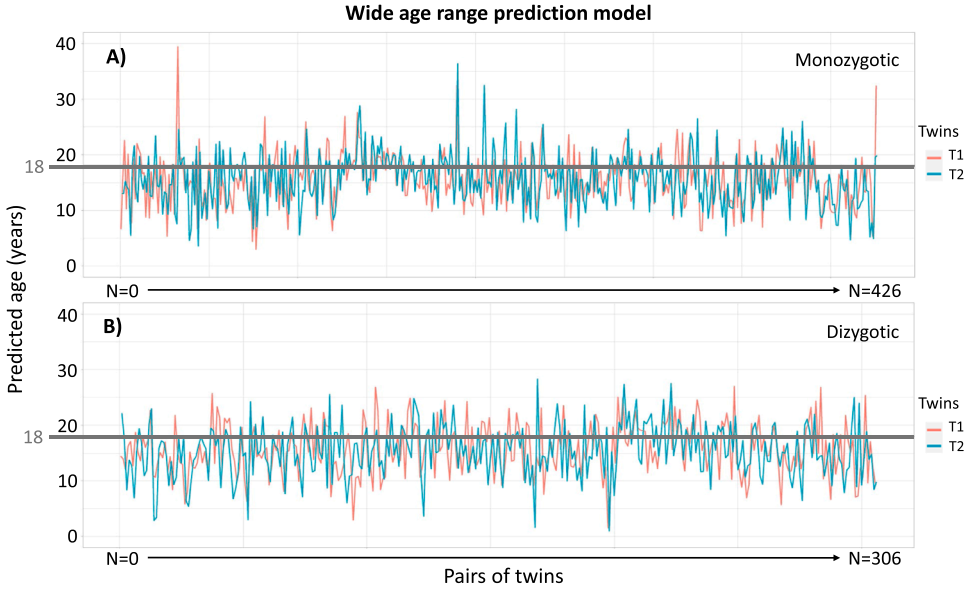


Fig. 1. Predicted age of each pair of 18-year-old twins using the wide age range prediction model (1 A: N=426 MZ pairs and 1B: N=306 DZ pairs). The twins from each pair are divided into twin T1 and T2, and represented by a red and a blue line, respectively. The grey horizontal line depicts the chronological age of these twins, 18 years old.

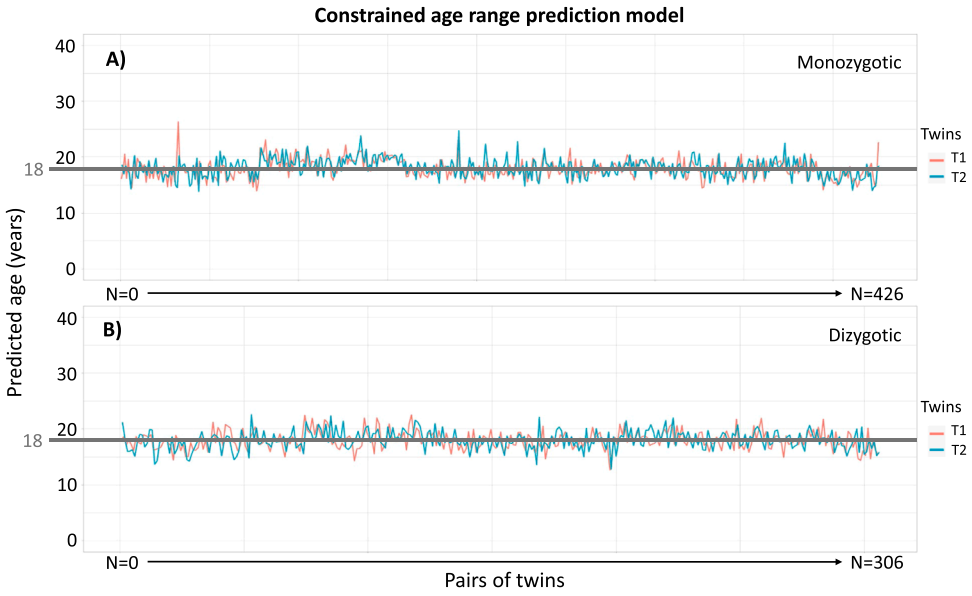


Fig. 2. Predicted age of each pair of 18-year-old twins using the constrained age range prediction model (2 A: N=426 MZ pairs and 2B: N=306 DZ pairs). The twins from each pair are divided into twin T1 and T2, and represented by a red and blue line, respectively. The grey horizontal line depicts the chronological age of these twins, 18 years old.

shown in Fig. 3**.** When analysing the MZ test set with this model, the prediction errors were MAE: ±1.87 years and RMSE: 2.4 (Fig. 3A). Out of the 852 samples, 538 (63.15 %) were predicted over 18 and 314 (36.85 %) were predicted under 18. When analysing the DZ test set, the prediction errors were MAE: ±1.99 years and RMSE: 2.54 (Fig. 3B). Out of the 612 total samples, 381 (62.25 %) were predicted over 18 and 231 (37.75 %) were predicted under 18.

As for the two previous models, Horvath's also gave predictions where, in the same pair of twins, one was predicted as a minor and the other an adult. Out of the total 732 pairs of twins, this model presents one twin as a minor and the other as an adult in 207 pairs of twins (28.28 %). These results are shown in **Supplementary Fig. S6**. When using this model, an MAE of ±1.74 years was obtained between twins belonging to the same pair.

3.4. Differences in age prediction of minors and adults

To study if samples close to the age of 18 years old behave differently in terms of age prediction, minors from 14 to 17 years old (N=65), as well as adults between 19 and 25 years old (N=79) were studied. To do this, the fitted values (0.5 quantile) for the individuals from the training set (GSE87571) corresponding to this age range were evaluated using leave-one-out cross-validation with the wide and constrained age range prediction models. This cross-validation step was not possible for Horvath's Skin & Blood model since the information related to the training set corresponding to this model is unavailable. The leave-one-out cross-validation consists of using all the samples except for one to form the training set with the left-out sample used as the test sample. This procedure is repeated for all samples, to obtain age predictions for each. The

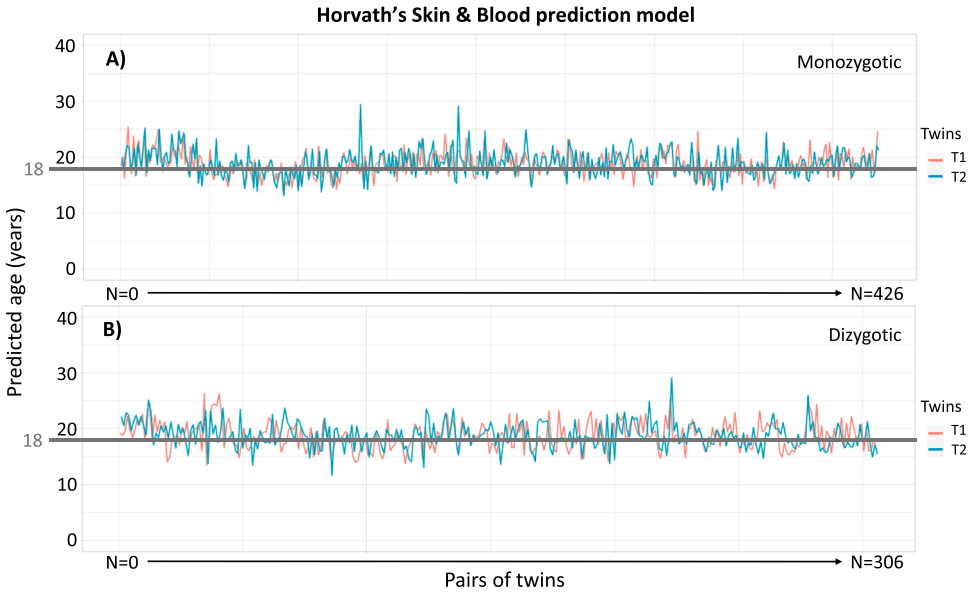


Fig. 3. Predicted age of each pair of 18-year-old twins using Horvath's Skin & Blood prediction model (3 A: N=426 MZ pairs and 3B: N=306 DZ pairs). The twins from each pair are divided into twin T1 and T2, and represented by a red and a blue line, respectively. The grey horizontal line depicts the chronological age of these twins, 18 years old.

resulting predictions were used to test for differences between predicting the age of minors and adults.

Fig. 4A shows the results for the wide age range prediction model, where the minors yielded a mean error of −1.42 years while the adults yielded a mean error of +1.77 years. Since the chronological age was subtracted from the predicted age, the positive results indicate overestimation, and the negative results underestimation. The same procedure was applied for the constrained age range prediction model depicted in Fig. 4B, where the minors yielded a mean error of +1.35 years, and the adults yielded a mean error of −1.2 years.

4. Discussion

Being able to predict legal age with the highest possible accuracy is a matter of interest, since many minors have no other way to prove they are underaged. Current methods, mostly involving X-ray analyses, are highly invasive, so new options such as DNA methylation should be considered.

In this study, two age prediction models based on blood were created. These two models (the wide and the constrained model), as well as the established Horvath's Skin & Blood clock, were used to predict the age of 732 pairs of 18-year-old twins. As seen in the results, the MAE for the constrained model was lower than the one obtained for Horvath's Skin & Blood clock and both of their MAEs were lower than the one obtained for the wide model. To further compare these three models, the tendency of each of them to over- or underestimate an individual's age was studied.

Regarding the percentages of all three models when classifying both test sets (MZ and DZ) into minors and adults, the wide model tended to classify them as minors while the constrained and the Horvath's Skin & Blood model tended to classify them as adults. To confirm these tendencies, we subtracted the chronological age of each twin to the age predicted by the model. In this case, the positive results indicate overestimation, and the negative results indicate underestimation. The results obtained can be seen in **Supplementary Fig. S7**. When predicting the MZ and DZ twins, it was observed that the wide model did tend to underestimate age in both groups, with mean errors of −2.38 and −2.43 years, respectively. The constrained model had little difference between under and overprediction both in MZ and DZ with mean errors of +0.01 and +0.01, respectively. In contrast, the Horvath's Skin & Blood clock tended to overestimate age in both groups, with a mean error of +0.83 for MZ and +0.79 for DZ.

Since we are aiming to predict legal age, the prediction model should always favour the minor, meaning the prediction model used should tend to underestimate the age of the analysed individual, if they are a minor. For this reason, the age prediction model not only needs the lowest MAE possible but must also not overestimate age of the young. When comparing the wide and constrained model, the latter clearly has a lower MAE (±4.32 and ±1.54 years respectively). However, when comparing the tendency to under and overestimate the age of minors and adults, the wide model tends to underestimate the age of minors and overestimate those of adults, whilst the constrained model has the opposite tendency. Overestimating minors and underestimating adults is contrary to the desired objective, so the constrained model is less appropriate for predicting legal age, as it might have seemed at first.

Other models for age prediction in young donor samples already exist [34–36]. McEwen et al.'s [35] PedBE clock for buccal samples has an MAE of ±0.35 years using 94 CpGs and 1032 samples. However, the samples used to create this clock range from 0 to 20 years of age. If our objective is to predict legal age, a wider range of age should be used for adults. Aanes et al.'s prediction model covers this range [36], based on 973 blood samples ranging from 12 to 25 years of age. Aanes' model gave a mean absolute deviation below 0.7 years using 267 CpGs. Both studies created their corresponding prediction models using an elastic net regression, in contrast to the quantile regression used in the current study. Quantile regression was chosen because DNA methylation presents heteroscedastic variability as epigenetic changes accumulate over the years.

One of the main differences between the two forementioned studies and the present one, are the number of samples used in each study. A higher number of samples used in the training set produces a better performing age prediction model. The other main difference between studies is the number of CpGs used in each model. Both McEwen's and Aanes' models use a much higher number of CpGs (94 and 267 respectively) than our study (5 CpGs). To see if a higher number of epigenetic markers yields a better prediction regarding legal age, we compared the constrained model (5 CpGs) with Horvath's Skin & Blood model (391 CpGs) and found no significant differences between the two, so a high number of CpGs is not needed to obtain a highly accurate prediction model.

Regardless of having prediction models with low error, it is important to take into consideration that two same-aged individuals can have

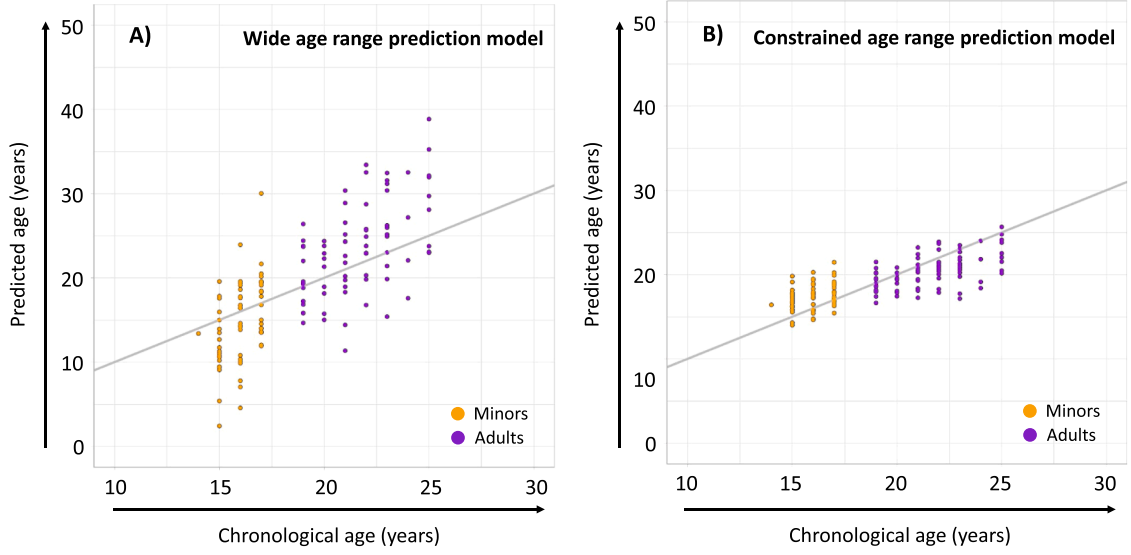


Fig. 4. Prediction tendencies for the wide and constrained age range prediction models (4A and 4B, respectively) regarding both categories (minors: N=65 and adults: N=79). The predictions for the minors are depicted in orange while the predictions for the adults are depicted in purple. The light grey line is the perfect correlation between predicted and chronological age.

very different predictions. This was observed when comparing the predictions between same-pair twins. In some cases, one twin was classified as a minor while the other was classified as an adult. In all three prediction models studied here, approximately 30 % of pairs of twins behaved in this fashion, despite the differences in their MAEs (±4.32 years the highest and ±1.54 years the lowest). Since twin samples are supposed to have minimal differences between them when comparing results, it can be expected that this percentage would be higher between same-aged individuals who are not related. Additionally, differences in lifestyles of the individuals included in the test cohort could potentially explain the outliers observed in the graphical representations.

One consideration is the fact that different populations might yield different results. It has been noted that methylation values in CpGs vary from one population to another [37]. In this study, the training set was comprised of a Swedish cohort and the test set from a UK cohort. These two cohorts can be considered to have the same European biogeographic origin. Further studies should be made to establish if different biogeographic origins would influence legal age prediction.

The main aim of this study was not to create a legal age prediction model, since Aanes' age predictor for adolescents and young adults already has promising results. Here we strive to determine whether DNA methylation is a suitable candidate to replace the current procedures to determine legal age and detect possible limitations to take into consideration when using methylation-based prediction models to determine said age. We found that using a higher number of samples to create the training set as well as limiting their age to a constrained range around the desired age, reduced considerably the prediction error, whereas using a higher number of epigenetic markers did not suppose a difference. Also, this error should not be the only factor taken into consideration when selecting a model. Prioritizing the underprediction of minors and the overprediction of adults and not the other way around, should also be an important aspect of the prediction model. Finally, low prediction error models are still prone to classify same-aged individuals into the different minors or adults' categories.

CRediT authorship contribution statement

Miguel Boullón-Cassau: Writing – original draft, Validation, Formal analysis, Data curation. **Adrián Ambroa-Conde:** Writing – review & editing. **María de los Ángeles Casares de Cal:** Methodology. **Antonio Gómez-Tato:** Methodology. **Ana Mosquera-Miguel:** Writing – review & editing. **Jorge Ruiz-Ramírez:** Writing – review & editing. **Amaia Cabrejas-Olalla:** Writing – review & editing. **Javier González-Bao:** Writing – review & editing. **Lucía Casanova-Adán:** Writing – review & editing. **María de la Puente:** Writing – review & editing. **Amelia Rodríguez:** Writing – review & editing. **Christopher Phillips:** Writing – review & editing. **Victoria Lareu:** Funding acquisition. **Ana Freire-Aradas:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization.

Declaration of Competing Interest

The authors have declared no conflict of interest.

Acknowledgements

MBC is supported by the "Programa de axudas á etapa predoutoral" funded by the Consellería de Cultura, Educación e Formación Profesional e Universitaria from Xunta de Galicia, Spain (ED481A-2024-163). JRR is supported by the "Programa de axudas á etapa predoutoral" funded by the Consellería de Cultura, Educación e Ordenación Universitaria e da Consellería de Economía, Emprego e Industria from Xunta de Galicia, Spain (ED481A-2020/039). JGB is supported by predoctoral grant "Ayudas para la Formación del Profesorado Universitario" (reference FPU22/02683) funded by Ministerio de Universidades from Gobierno de España. LCA is supported by a predoctoral grant funded by "Fundación Caixa Rural Galega Tomás Notario Vacas" in collaboration agreement with the Universidade de Santiago de Compostela. MdP is supported by grant IJC2020–042638-I funded by the Gobierno de España MCIN/AEI/10.13039/501100011033 and the European Union "NextGenerationEU/PRTR". MVL is supported by grant PID2019–107876RB-I00 funded by the Ministerio de Educación, Cultura y Ciencia, Spain (MCIN/AEI/ 10.13039/501100011033) and grant PID2022–141224OB-I00 funded by MCIN/AEI/ 10.13039/501100011033 and "ERDF A way of making Europe".

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fsigen.2024.103142.

References

[1] United Nations, Convention on the rights of the child, Treaty Ser. Vol. 1577 (1989) 3.
[2] J. Folguera Crespo, C. Puerta Ruiz De Azúa, S. Moya García, El procedimiento de determinación de la edad de los menores extranjeros no acompañados, Actual. Jur. ídica Ur. ía Men. éndez Vol. 58 (2022) 221–236.
[3] W. Greulich, I. Pyle, Radiographic atlas of skeletal development of the hand and wrist, Stanford University Press, London, 1959.
[4] Demirjian A., Goldstein H., Tanner J.M. A new system of dental age assessment. Vol. 45. 1973. p. 211–227.
[5] R. Holliday, Epigenetics: a historical overview, Epigenetics Vol. 1 (2006) 76–80.
[6] M. Gardiner-Garden, M. Frommer, CpG islands in vertebrate genomes, J. Mol. Biol. Vol. 196 (1987) 261–282.
[7] W. Doerfler, Review article. DNA methylation-A regulatory signal in eukaryotic gene expression, J. Gen. Virol. Vol. 57 (1981) 1–20.
[8] J. Ohgane, S. Yagi, K. Shiota, Epigenetics: the DNA methylation profile of tissue-dependent and differentially methylated regions in cells, Placenta Vol. 29 (2008) 29–35.
[9] S. Bocklandt, W. Lin, M.E. Sehl, F.J. Sánchez, J.S. Sinsheimer, S. Horvath, et al., Epigenetic predictor of age, PLoS ONE Vol. 6 (2011) 14821.
[10] Å. Johansson, S. Enroth, U. Gyllensten, Continuous aging of the human DNA methylome throughout the human lifespan, PLoS ONE Vol. 8 (2013) 67378.
[11] G. Hannum, J. Guinney, L. Zhao, L. Zhang, G. Hughes, S.V. Sadda, et al., Genome-wide methylation profiles reveal quantitative views of human aging rates, Mol. Cell Vol. 49 (2013) 359–367.
[12] S. Horvath, DNA methylation age of human tissues and cell types, Genome Biol. Vol. 14 (2013) 115.
[13] I. Florath, K. Butterbach, H. Müller, M. Bewerunge-hudler, H. Brenner, Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60 novel age-associated CpG sites, Hum. Mol. Genet. Vol. 23 (2014) 1186–1201.
[14] A. Freire-Aradas, L. Girón-Santamaría, A. Mosquera-Miguel, A. Ambroa-Conde, C. Phillips, M. Casares de Cal, et al., A common epigenetic clock from childhood to old age, Forensic Sci. Int.: Genet. Vol. 60 (2022) 102743.
[15] A. Woźniak, A. Heidegger, D. Piniewska-Róg, E. Pośpiech, C. Xavier, A. Pisarek, et al., Development of the VISAGE enhanced tool and statistical models for epigenetic age estimation in blood, buccal cells and bones, Aging (Albany NY) Vol. 13 (2021) 6459–6484.
[16] B. Bekaert, A. Kamalandua, S.C. Zapico, W. Van de Voorde, R. Decorte, Improved age determination of blood and teeth samples using a selected set of DNA methylation markers, Epigenetics Vol. 10 (2015) 922–930.
[17] A. Ambroa-Conde, L. Girón-Santamaría, A. Mosquera-Miguel, C. Phillips, M. A. Casares de Cal, A. Gómez-Tato, et al., Epigenetic age estimation in saliva and in buccal cells, Forensic Sci. Int.: Genet. Vol. 61 (2022) 102770.
[18] M.H. Lee, J.H. Hwang, K.M. Seong, J.J. Ahn, S.J. Kim, S.Y. Hwang, et al., Application of droplet digital PCR method for DNA methylation-based age prediction from saliva, Leg. Med. Vol. 54 (2022) 101992.
[19] A. Pisarek, E. Pośpiech, A. Heidegger, C. Xavier, A. Papież, D. Piniewska-Róg, et al., Epigenetic age prediction in semen-marker selection and model development, Aging Vol. 13 (2021) 19145–19164.
[20] H.Y. Lee, S.E. Jung, Y.N. Oh, A. Choi, W.I. Yang, K.J. Shin, Epigenetic age signatures in the forensically relevant body fluid of semen: A preliminary study, Forensic Sci. Int.: Genet. Vol. 19 (2015) 28–34.
[21] H.Y. Lee, S.R. Hong, J.E. Lee, I.K. Hwang, N.Y. Kim, J.M. Lee, et al., Epigenetic age signatures in bones, Forensic Sci. Int.: Genet. Vol. 46 (2020) 102261.
[22] E-Risk Study [Internet]. [cited 2024 Jul 3]. Available from: ⟨https://eriskstudy.com/about-the-e-risk-study/⟩.
[23] A. Freire-Aradas, C. Phillips, A. Mosquera-Miguel, L. Girón-Santamaría, A. Gómez-Tato, M. Casares De Cal, et al., Development of a methylation marker set for forensic age estimation using analysis of public methylation data and the Agena Bioscience EpiTYPER system, Forensic Sci. Int.: Genet. Vol. 24 (2016) 65–74.
[24] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2023.
[25] M.J. Aryee, A.E. Jaffe, H. Corrada-Bravo, C. Ladd-Acosta, A.P. Feinberg, K. D. Hansen, et al., Minfi: A flexible and comprehensive Bioconductor package for

the analysis of Infinium DNA methylation microarrays, Bioinformatics Vol. 30 (2014) 1363–1369.
[26] Hansen K., Aryee M. IlluminaHumanMethylation450kmanifest: Annotation for Illumina's 450k methylation arrays. R package version 0.4.0. 2012.
[27] Hansen K. IlluminaHumanMethylation450kanno.ilmn12.hg19: Annotation for Illumina's 450k methylation arrays. R package version 0.6.1. 2021.
[28] Koenker R. (2023). quantreg: Quantile Regression R package version 5.95.
[29] Alfons A. (2012). cvTools: Cross-validation tools for regression models. R package version 0.3.2.
[30] H. Wickham, ggplot2: Elegant graphics for data analysis (Springer-Verlag New York), Springer-Verlag New York, 2016.
[31] Kassambara A. (2023). ggpubr: "ggplot2" Based Publication Ready Plots. R package version 0.6.0.
[32] S. Horvath, J. Oshima, G.M. Martin, A.T. Lu, A. Quach, H. Cohen, et al., Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies, Aging Vol. 10 (2018) 1758–1775.
[33] D. Pelegí-Sisó, P. De Prado, J. Ronkainen, M. Bustamante, J.R. González, Methylclock: a bioconductor package to estimate DNA methylation age methylclock: a bioconductor package to estimate DNA methylation age, Bioinformatics Vol. 37 (2021) 1759–1760.
[34] C. Li, W. Gao, Y. Gao, C. Yu, J. Lv, R. Lv, et al., Age prediction of children and adolescents aged 6-17 years: an epigenome-wide analysis of DNA methylation, Aging Vol. 10 (2018) 1015–1026.
[35] L.M. McEwen, K.J. O'Donnell, M.G. McGill, R.D. Edgar, M.J. Jones, J.L. MacIsaac, et al., The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells, Proc. Natl. Acad. Sci. USA 117 (38) (2020) 23329–23335.
[36] Aanes H., Bleka Ø., Dahlberg P.S., Carm K.T., Lehtimäki T., Raitakari O., et al. (2023). A new blood based epigenetic age predictor for adolescents and young adults. Vol. 13, Scientific Reports. Nature Research.
[37] J. Fleckhaus, P. Bugert, N.A.M. Al-Rashedi, M.A. Rothschild, Investigation of the impact of biogeographic ancestry on DNA methylation based age predictions comparing a Middle East and a Central European population, in: Forensic Science International: Genetics, Vol. 67, Elsevier Ireland Ltd, 2023 102923.