Adapting Blood DNA Methylation Aging Clocks for Use in Saliva Samples With Cell-type Deconvolution

Fedor Galkin1*, Kirill Kochetov1, Polina Mamoshina1 and Alex Zhavoronkov1,2,3*

1Deep Longevity Limited, Hong Kong, China, 2Insilico Medicine Hong Kong Limited, Hong Kong Science and Technology Park, Hong Kong, China, 3The Buck Institute for Research on Aging, Novato, CA, United States

DeepMAge is a deep-learning DNA methylation aging clock that measures the organismal pace of aging with the information from human epigenetic profiles. In blood samples, DeepMAge can predict chronological age within a 2.8 years error margin, but in saliva samples, its performance is drastically reduced since aging clocks are restricted by the training set domain. However, saliva is an attractive fluid for genomic studies due to its availability, compared to other tissues, including blood. In this article, we display how cell type deconvolution and elastic net can be used to expand the domain of deep aging clocks to other tissues. Using our approach, DeepMAge’s error in saliva samples was reduced from 20.9 to 4.7 years with no retraining.

Keywords: aging, biogerontology, epigenetics, aging clock, deep learning, cell-type deconvolution, DNA methylation, domain adaptation

1 INTRODUCTION

Aging clocks are the centerpiece of the emerging longevity industry. They allow us to accurately quantify a hidden property of living organisms—the pace of aging.

Over the years the concept of aging clocks has been validated in numerous settings. An increased pace of aging has been shown to manifest in diseases, smoking, obesity, and other conditions that could be interpreted as reducing human longevity potential. At the same time, a slower pace of aging has been associated with lower mortality, and decreasing it is the aim of a new medical paradigm—longevity medicine (Bischof et al., 2021).

However, each aging clock comes with its own set of limitations, that stand in the way of its widespread adoption. Some of these limitations stem from the mathematical apparatus used to measure the pace of aging (Galkin et al., 2020a). Some of them are inherent to the information a certain data type can possibly convey about the state of the whole organism. Another set of limitations has its root in the training set domain.

Currently, all aging clocks are implemented within the supervised learning framework. A data set is collected and each sample is labeled with a target variable representing biological age. The most commonly used target variable is chronological age (CA), however more complex target variables have also been tried with great success (Levine et al., 2018; Lu et al., 2019).

A loss function is used to iteratively train a statistical model to correctly predict the age of a person until the best parameters for the model are established. When CA is used to train the model, the pace of aging is expressed as its prediction error: if an aging clock estimates someone to be five years older than their CA, their pace of aging is said to be increased.

All samples within a training set need to share a degree of similarity (Chicco, 2017). Otherwise, any modeling approach may not be able to identify the connection between the measured properties and aging.
2 MATERIALS AND METHODS

2.1 Aging Clock

The aging clock used in this study is DeepMAge originally published in Galkin et al. (2020b). DeepMAge is a deep-learning neural network that takes in a vector of \( \beta \)-values for 1,000 CpG sites present in Illumina BeadChip 27 K and Illumina BeadChip 450 K platforms.

DeepMAge has been trained on a collection of blood DNAm profiles and its behavior in saliva samples has not been described elsewhere.

2.2 Cell Type Deconvolution

To determine the cell type composition of the saliva samples, EpiDISH described in Zheng et al. (2018) was used. EpiDISH is available as an R package at https://github.com/sjczheng/ EpiDISH (v.2.6.0).

2.3 Data Collection

All data used in this study is publicly available at Gene Expression Omnibus (GEO). We selected the datasets according to the following criteria: 1) A data set had to contain epigenetic profiles obtained with an Illumina Infinium array; 2) A data set had to contain saliva and/or buccal swab samples; 3) A data set had to be annotated with age and sex information.

In the end, 12 datasets were selected. Study accession numbers for the training set are: GSE78874—contains men and women of two ethnic backgrounds: Hispanics and Caucasians (\( N = 259 \) people), GSE80261—contains human buccal epithelial cells from children from the NeuroDevNet cohort (\( N = 216 \)), GSE94876—contains methylation profiles in buccal cells of long-term smokers and moist snuff consumers (\( N = 120 \)), GSE34035—contains saliva DNA from alcohol-dependent subjects (\( N = 112 \)), GSE50759—contains samples from buccal epithelium collected using exfoliative brushing (\( N = 96 \)), GSE82746—saliva samples from male identical twins discordant for sexual orientation (\( N = 84 \)), GSE42700 - buccal cell samples collected at birth and 18 months from 10 monozygotic and five dizygotic twin pairs from the Peri/postnatal Epigenetic Twins Study (PETS) cohort (\( N = 53 \)), GSE50586–buccal swabs (\( N = 20 \)).

Studies used in the verification set include: GSE92767—saliva from 54 males aged 18–73 years (\( N = 54 \)), GSE109042–human buccal epithelial cells from children with fetal alcohol spectrum disorder and control samples (\( N = 47 \)), GSE28217–contains primary oral squamous cell carcinoma, oral leukoplaikia, and normal oral mucosa (\( N = 12 \)). GSE48742–contains data from blood, saliva, buccal swab, and hair follicles, but only saliva and buccal swab samples were used (\( N = 5 \)). All samples were normalized using lumi, according to the protocol described in Galkin et al. (2020b), Du et al. (2008).

2.4 DeepMAge Prediction Adjustment

Several approaches were compared to find the optimal way of enabling accurate DeepMAge predictions for saliva samples:

- **No adjustment**: DeepMAge predictions were used as-is;
- **Simple adjustment**: No information from cell-type deconvolution was used. Actual age of the saliva samples was regressed as a function of its DeepMAge prediction alone;
- **Shift adjustment**: \( \Delta \text{Age} \) was regressed as a function of saliva samples’ cell types to obtain a correction term, where \( \Delta \text{Age} = \text{DeepMAge prediction} - \text{CA} \). Correction terms were added to DeepMAge output of the saliva samples to obtain the adjusted predictions;
• **Total age adjustment**: CA of saliva samples was regressed as a function of DeepMAge predictions and cell composition to obtain the adjusted predictions.

Any models used to derive adjustments were implemented with the ElasticNet fitter, as implemented in scikit-learn v.0.24.1 Python package. Optimal regression parameters were chosen based on the grid search defined as:

```
'alpha': (1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 0.0, 1.0, 10.0, 100.0)
```

with leave-one-study-out (LOSO) cross-validation (CV) folds. The final adjustment regressions were trained on all studies, save those in the verification set.

For simple and total age adjustments, transformed predicted age values were tried. The adjustment was described in detail in Horvath (2013). The age transformation may be expressed as:

\[
f(x) = \begin{cases} 
\frac{x + 1}{21} - 1 & \text{if } x > 20 \\
\ln\frac{x + 1}{21} & \text{if } x \leq 20 
\end{cases}
\]

\[
f^{-1}(x) = \begin{cases} 
21 \times \exp(x) - 1 & \text{if } x < 0 \\
21 \times x + 20 & \text{if } x \geq 0 
\end{cases}
\]

### 3 RESULTS

#### 3.1 Cell Composition in Saliva DNA Methylation Studies

The collection of publicly available studies we used was extremely diverse in its cell composition. The average fraction of immune cells in the samples varied within the 5.6–81.3% range. Most studies contained a negligible fraction of fibroblasts (< 5%), except for GSE28217 in which samples contained 23.3% fibroblast cells on average (**Figure 1A**).

Cell composition varied significantly not only between studies but also between the samples from the same study, illustrating the heterogeneity inherent to salivary samples processed according to the same study protocol (**Figure 1B**).

#### 3.2 Enabling Age Prediction in Saliva Samples

No adjustment for cell-type composition produced worse predictions than the baseline model – median age assignment (**Table 1**). The accuracy was increased with any of the proposed adjustments. Among the adjusters using non-transformed age, the correction based on delta age, immune and epithelial cell relative counts (“Delta age” in **Table 1**) produced the most accurate results – MAE = 6.34 years in the CV and MAE = 6.28 years in the verification sets.

The most accurate adjuster, however, used the age transformation described in Methods. The best adjuster among those tested is “Total age-transformed” with a MAE of 4.57 years in the CV set and 4.74 years in the verification sets.

The total age adjustment can be expressed as the following formulas, with and without age transformation, respectively:

\[
f(DeepMAge_{Saliva}) = 0.93 \times f(DeepMAge_{Blood}) - 1.14 \times Epi \\
+ 1.05 \times IC - 1.18
\]  

(1)

**FIGURE 1** | Publicly available studies of DNAm in saliva samples have high inter- and intra-study cell composition variability. (A) Average cell composition of saliva samples in the studies used, estimated with EpiDISH DNAm deconvolution. N = Total number of samples in a study. (B) Immune cell distributions within the studies used. Boxes correspond to the IQR, whiskers protrude no further than 1.5xIQR. Boxes are colored according to their inclusion in the training or verification set.

**TABLE 1** | All cell-type variability adjustments significantly improve DeepMAge's accuracy in saliva samples. MAE = Mean Absolute Error, see MATERIALS AND METHODS for the description of the adjustments used.

| Adjustment               | Age transformed | MAE, years |
|--------------------------|-----------------|------------|
|                          | Train          | Test       |
| Baseline Non-transformed | 22.09          | 19.03      |
| None Non-transformed     | 26.36          | 20.86      |
| Simple Non-transformed   | 13.48          | 13.74      |
| Delta age Non-transformed| 12.54          | 13.87      |
| Total age Non-transformed| 7.76           | 6.83       |
| Total age Transformed    | 4.57           | 4.74       |

Galkin et al. DNAm Deconvolution for Aging Clocks
DeepMAgeSaliva = 0.85 × DeepMAgeBlood − 11.54 × Epi + 10.56 × IC − 15.76

(2)

where Epi and IC are the portions of epithelial and immune cells in the sample, respectively.

3.3 The Effect of Immune Cell Fraction on the Increase in Accuracy

Since DeepMAge was trained on a set of blood-derived DNAm profiles, we wondered if our adjuster-suggested shift in prediction was larger in samples with lower IC proportions. For this purpose, we measured the increase in accuracy between DeepMAgeBlood and DeepMAgeSaliva as the drop in distance to the chronological age after the adjustment.

We compared the increase in accuracy to the IC content of the samples to discover that IC content was strongly negatively correlated with the absolute adjustment magnitude ($R^2 = 0.72$, Pearson’s $r = −0.85$). Moreover, only 5% of all samples actually grew farther from the chronological age, among which most consist of > 50% IC (Figure 4).

4 DISCUSSION

In this article, we have demonstrated a method to expand an aging clock’s domain of applications to other tissues.

More specifically, the method we propose can be used to translate blood-trained aging clocks to saliva and buccal swab samples.

Saliva is one of the easiest to obtain tissues and is a good source of high-quality DNA. In comparison to blood, saliva requires no additional anticoagulant, the risk of disease transmission is lower, and there is no need for venipuncture (Langie et al., 2017). Although blood is the most widely used tissue for clinical tests, saliva in its own right has been shown to be a valuable biomarker for the early detection of such cancer types as oral cancer, head and neck squamous cell carcinoma, and breast cancer (Liyanage et al., 2019; Bryan et al., 2013; Nagler, 2009; Delmonico et al., 2015).

Moreover, saliva methylome was also shown to be positively correlated with methylation in blood for 88.5% (Smith et al., 2015). Thus, changes in salivary DNAm profiles might be indicative of organism-wide aging processes that can also be detected in blood DNAm levels. From a technical standpoint, however, saliva cannot be used as a substitute for DNAm aging clocks trained on blood samples only.

DeepMAge is a deep learning model, which was trained on a set of blood-derived DNAm profiles and shows poor performance in saliva samples. Salvaging its accuracy in other tissues can be formalized as a domain adaptation problem, for which a variety of theoretical solutions exist within the deep learning framework. Our solution, however, does not involve retraining the model or...
manipulating the target domain samples. It is based on introducing additional information (cell type composition) on the target domain samples to derive a linear offset for aging clock predictions. The reported adjustment has resulted in a saliva DNAm aging clock with a MAE of 4.7 years, which is comparable to the clock described in Bocklandt et al. (2011) which was trained in saliva and reached a MAE of 5.2 years.

To estimate the cell content of salivary sample, we used EpiDISH—a cell type deconvolution algorithm that uses a reference of pure cell type DNAm profiles to evaluate the cell composition of bulk samples. This tool was verified using in silico mixtures of pure cell types. Its creators aimed to remove the effect of immune cell contamination in differential methylation studies in epithelial tissues.

Since DeepMAge was trained on blood data, the IC effect represents not the contamination, in this setting, but the signal the aging clock is prepared to recognize. Thus in some sense, EpiDISH may be said to adjust for the noise associated with epithelial DNAm profiles, bringing the samples closer to the original training domain.

However, this explanation for the efficiency of cell-type adjustment in the context of DeepMAge performance is an oversimplification. EpiDISH does not deconvolute DNAm bulk profiles into three distinct profiles of fibroblast, immune, and epithelial origin, the mixture of which would produce the observed profile. It only provides a vector of relative abundances for these three cell types in the sample. Thus, it is impossible to subtract the epithelial noise from the DNAm profiles. Another counterargument for the noise hypothesis is that EpiDISH significantly improves DeepMAge prediction quality even in data sets that have low immune cell abundance (GSE42700, GSE109042), and thus could be said to have an overwhelming noise to signal ratio (Figure 4).

Another, more probable explanation for the efficiency of EpiDISH is that the pace of epigenetic aging, as perceived by DeepMAge, is faster in epithelial samples. This follows from the total age correction formula: 1) DeepMAge overestimates the age of saliva samples (see <1 coefficient for DeepMAgeBlood, negative intercept), 2) a sample consisting completely of epithelial cells would be predicted 27.3 years older than it should be (see negative Epi coefficient, negative intercept).

Another major contributing factor, apparently, is that despite their different function and ontogenesis, buccal epithelium and immune blood cells share similar aging signatures. Since the correction can be interpreted to stretch the original DeepMAge prediction space, it may be argued that the changes in the organismal pace of aging may be more rapidly reflected in saliva than in blood. This quicker responsiveness and the difference in the aging rate between IC and epithelium may be
explained by a much higher turnover rate of buccal epithelium whose surface layer is replaced every 3 h (Sender and Milo, 2021; Dawes, 2003).

The tissue-specific pace of aging has been reported in some other publications (Hannum et al., 2013; Thompson et al., 2018), yet in them, the impact of variable cell composition is not explored. We believe that it is important to delineate the effects of the primary tissue and that of the contaminating cells.

We tried the approach similar to the one used in Hannum et al. (2013) ("Simple adjustment" in Table 1) (two-fold increase in accuracy). While it yielded better results than the non-adjusted DeepMAge, the best adjuster produced predictions almost three times as accurate (Table 1). We argue that simple tissue-specific linear offsets are insufficient to turn a single-tissue clock into a multi-tissue one. This approach ignores the marked variance of cell composition that is present even within the same specimen collection protocol. Samples from the same study can have no immune cell contaminants at all, or be composed of them completely (Figure 1). Nonetheless, both extremes and everything in-between will be identically labeled as “saliva”.

In the future, single-cell DNAm profiling may render the concept of a “tissue-specific” aging clock obsolete by designing biological age as a probabilistic function of individual cells (Trapp et al., 2021). But within the bulk sample context, DNAm deconvolution is an impressively efficient technique to solve the tissue domain problem. While working with public data repositories, it can also serve as a quality control tool to remove outliers, control a hidden source of technical variation, and fill in missing or misleading tissue labels.

From a practical standpoint, our results show that it is not necessary to retrain deep learning clocks or employ any other complex domain adaptation techniques to widen their range of use cases. Since the correction procedure is model agnostic, similar extensions may be developed for shallow learning aging clocks.

While blood is easily available in clinical settings, the necessity to schedule blood drawing stands in the way of consumer biogerontology applications. Using cell deconvolution to adapt the existing blood domain solutions to saliva samples will increase the adoption rate of the aging clock technology.

5 CONCLUSION

Cell-type deconvolution can be used to expand the applicability of aging clocks to the tissues outside of their original training domain and significantly increase their accuracy for initially unintended use-cases.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in the Gene Expression Omnibus online (GEO) repository. The accession numbers can be found below: GSE78874, GSE80261, GSE94876, GSE34035, GSE50759, GSE28746, GSE42700, GSE50586, GSE92767, GSE109042, GSE8217, GSE48472.

AUTHOR CONTRIBUTIONS

FG—Formal analysis, Software, Visualization, Writing (original draft), Writing (review) KK—Software PM—Project administration, Methodology, Supervision, Writing (review) AZ—Conceptualization, Resources, Supervision.

ACKNOWLEDGMENTS

The authors thank K. Anoshkin for valuable comments and suggestions that let us improve the quality of the manuscript.

REFERENCES

Aps, J. K. M., Van den Maagdenberg, K., Delanghe, J. R., and Martens, L. C. (2002). Flow Cytometry as a New Method to Quantify the Cellular Content of Human Saliva and its Relation to Gingivitis. Clinica Chim. Acta Int. J. Clin. Chem. 321, 35–41. doi:10.1016/s0003-9969(02)00062-1

Bischof, E., Schelby-Knudsen, M., Siow, R., and Moskalev, A. (2021). Longevity Medicine: Upskilling the Physicians of Tomorrow. The Lancet Healthy Longevity 2, e187–e188. doi:10.1016/s2666-7568(21)00024-6

Bocklandt, S., Lin, W., Sehl, M. E., Sanchez, F. J., Sinsheimer, J. S., Horvath, S., et al. (2011). Epigenetic Predictor of Age. PLoS ONE 6, e14821. doi:10.1371/journal.pone.0014821

Bryan, A. D., Magnan, R. E., Hooper, A. E. C., Harlaar, N., and Hutchison, K. E. (2013). Physical Activity and Differential Methylation of Breast Cancer Genes Assayed from Saliva: a Preliminary Investigation. Ann. Behav. Med. A Publ. Soc. Behav. Med. 45, 89–98. doi:10.1007/s12160-012-9411-4

Chicco, D. (2017). Ten Quick Tips for Machine Learning in Computational Biology. BioData Mining 10, 35. doi:10.1186/s13040-017-0155-3

Dawes, C. (2003). Estimates, from Salivary Analyses, of the Turnover Time of the Oral Mucosal Epithelium in Humans and the Number of Bacteria in an Edentulous Mouth. Arch. Oral Biol. 48, 329–336. doi:10.1006/ arbo.2003.9969(03)0014-1

Delmonico, L., Moreira, A. D. S., Franco, M. F., Esteves, E. B., Scherrer, L., Gallo, C. V. d. M., et al. (2015). CDKN2A (p14ARF/p16INK4a) and ATM Promoter Methylation in Patients with Impalpable Breast Lesions. Hum. Pathol. 46, 1540–1547. doi:10.1016/j.humpath.2015.06.016

Du, P., Kibbe, W. A., and Lin, S. M. (2008). Lumi: a Pipeline for Processing Illumina Microarray Bioinformatics (Oxford, England) 24, 1547–1548. doi:10.1093/ bioinformatics/btn224

Fitzgerald, K. N., Hodges, R., Hanes, D., Stack, E., Chiehivili, D., Szyf, M., et al. (2021). Potential Reversal of Epigenetic Age Using a Diet and Lifestyle Intervention: a Pilot Randomized Clinical Trial. Aging (Albany NY) 13, 9419–9432. doi:10.18632/aging.202913

Galkin, F., Mamoshina, P., Aliper, A., de Magalhães, J. P., Gladyshev, V. N., and Zhavoronkov, A. (2020a). Biohorology and Biomarkers of Aging: Current State-Of-The-Art, Challenges and Opportunities. Ageing Res. Rev. 60, 101050. doi:10.1016/j.arr.2020.101050

Galkin, F., Mamoshina, P., Kochetov, K., Sidorenko, D., and Zhavoronkov, A. (2020b). DeepMAge: A Methylation Aging Clock Developed with Deep Learning. Aging Dis. 11, 946–957. doi:10.14336/AD.2020.1202

Goh, W. W. B., Wang, W., and Wong, L. (2017). Why Batch Effects Matter in Omics Data, and How to Avoid Them. Trends Biotechnol. 35, 498–507. doi:10.1016/j.tibtech.2017.02.012

Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., et al. (2013). Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. Mol. Cell 49, 359–367. doi:10.1016/j.molcel.2012.10.016
Horvath, S. (2013). DNA Methylation Age of Human Tissues and Cell Types. *Genome Biol.* 14, R115. doi:10.1186/gb-2013-14-10-r115

Houseman, E. A., Zelterman, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., et al. (2012). DNA Methylation Arrays as Surrogate Measures of Cell Mixture Distribution. *BMC Bioinformatics* 13, 86. doi:10.1186/1471-2105-13-86

Langie, S. A. S., Moisse, M., Declerck, K., Koppen, G., Godderis, L., Vanden Berghe, W., et al. (2017). Salivary DNA Methylation Profiling: Aspects to Consider for Biomarker Identification. *Basic Clin. Pharmacol. Toxicol.* 121, 93–101. doi:10.1111/bcpt.12721

Levine, M. E., Lu, A. T., Quach, A., Chen, B. H., Assimes, T. L., Bandinelli, S., et al. (2018). An Epigenetic Biomarker of Aging for Lifespan and Healthspan. *Aging* 10, 573–591. doi:10.18632/aging.101414

Liyanage, C., Wathupola, A., Muraleetharan, S., Perera, K., Punyadeera, C., and Udagama, P. (2019). Promoter Hypermethylation of Tumor-Suppressor Genes p16INK4a, RASSF1A, TIMP3, and PCQAP/MED15 in Salivary DNA as a Quadruple Biomarker Panel for Early Detection of Oral and Oropharyngeal Cancers. *Biomolecules* 9, 148. doi:10.3390/biom9040148

Lu, A. T., Quach, A., Wilson, J. G., Reiner, A. P., Aviv, A., Raj, K., et al. (2019). DNA Methylation GrimAge Strongly Predicts Lifespan and Healthspan. *Aging* 11, 303–327. doi:10.18632/aging.101684

Nagler, R. M. (2009). Saliva as a Tool for Oral Cancer Diagnosis and Prognosis. *Oral Oncol.* 45, 1006–1010. doi:10.1016/j.oraloncology.2009.07.005

Sender, R., and Milo, R. (2021). The Distribution of Cellular Turnover in the Human Body. *Nat. Med.* 27, 45–48. doi:10.1038/s41591-020-01182-9

Smith, A. K., Kilaru, V., Klengel, T., Mercer, K. B., Bradley, B., Conneely, K. N., et al. (2015). DNA Extracted from Saliva for Methylation Studies of Psychiatric Traits: Evidence Tissue Specificity and Relatedness to Brain. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 168B, 36–44. doi:10.1002/ajmg.b.32278

Teschendorff, A. E., Breeze, C. E., Zheng, S. C., and Beck, S. (2017). A Comparison of Reference-Based Algorithms for Correcting Cell-type Heterogeneity in Epigenome-wide Association Studies. *BMC Bioinformatics* 18, 105. doi:10.1186/s12859-017-1511-5

Theda, C., Hwang, S. H., Czajko, A., Loke, Y. J., Leong, P., and Craig, J. M. (2018). Quantitation of the Cellular Content of Saliva and Buccal Swab Samples. *Scientific Rep.* 8, 9444. doi:10.1038/s41598-018-25311-0

Thompson, M. J., Chwilokowska, K., Rubbi, L., Luis, A. J., Davis, R. C., Srivastava, A., et al. (2018). A Multi-Tissue Full Lifespan Epigenetic Clock for Mice. *Aging* 10, 2832–2854. doi:10.18632/aging.101590

Teschendorff, A. E., Breeze, C. E., Zheng, S. C., and Beck, S. (2017). A Comparison of Reference-Based Algorithms for Correcting Cell-type Heterogeneity in Epigenome-wide Association Studies. *BMC Bioinformatics* 18, 105. doi:10.1186/s12859-017-1511-5

Copyright © 2021 Galkin, Kochetov, Mamoshina and Zhavoronkov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.