Technical Notes

Prediction of lung function in adolescence using epigenetic aging: a machine learning approach

Md. Adnan Arefeen [1,†], Sumaiya Tabassum Nimi [1,†], M. Sohel Rahman [2], S. Hasan Arshad [3,4], John W. Holloway [5] and Faisal I. Rezwan [5,6,*]

[1] Department of Computer Science Electrical Engineering, University of Missouri-Kansas City, Missouri, USA; aa4cy@mail.umkc.edu (M.A.A), snvb8@mail.umkc.edu (S.T.N)
[2] Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh; msrahman@cse.buet.ac.bd (M.S.R.)
[3] Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton SO16 6YD, UK; S.H.Arshad@soton.ac.uk (S.H.A.)
[4] The David Hide Asthma and Allergy Research Centre, St Mary's Hospital, Isle of Wight PO30 5TG UK
[5] Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK; J.W.Holloway@soton.ac.uk (J.W.H.)
[6] School of Water, Energy and Environment, Cranfield University, Cranfield, UK; F.I.Rezwan@cranfield.ac.uk (F.I.R.)
* Correspondence: F.I.Rezwan@cranfield.ac.uk; Tel.: +44 (0) 1234 75 4627 (U.K.)
† Contributed equally and share first authorship

Abstract: Epigenetic aging has been found associated with a number of phenotypes and diseases. Few studies investigated its effect on lung function in relatively older people. However, this effect has not been explored in younger population. This study examines whether lung function at adolescent can be predicted with epigenetic age accelerations (AAs) using machine learning techniques. DNA methylation based AAs were estimated in 326 matched samples at two time points (at 10 years and 18 years) from the Isle of Wight Birth Cohort. Five machine learning regression models (linear, lasso, ridge, elastic net, and Bayesian ridge) were used to predict FEV_1 (Forced Expiratory Volume in one second) and FVC (Forced Vital Capacity) at 18 years from feature selected predictor variables (based on mutual information) and AA changes between the two time points. The best models were ridge regression ($R^2 = 75.21\% \pm 7.42\%$; RMSE $= 0.3768 \pm 0.0653$) and elastic net regression ($R^2 = 75.38\% \pm 6.98\%$; RMSE $= 0.445 \pm 0.069$) for FEV_1 and FVC, respectively. This study suggests that the application of machine learning in conjunction with tracking changes in AA over life span can be beneficial to assess the lung health in adolescence.

Keywords: Lung function, epigenetic aging, machine learning, feature selection, hyperparamter tuning

---

1. Introduction

In recent years, the concept of biological aging, as opposed to chronological aging, has gained considerable popularity in understanding aging process due to its stronger relation with phenotypes and diseases [1]. DNA methylation (DNAm), an epigenetic process, can provide biomarkers to estimate the biological aging, known as "epigenetic aging". There are several methods available to estimate epigenetic aging [2–6] and among them Horvath methods for epigenetic age estimation (DNAmAge) is used widely and has shown high accuracy, with an average correlation > 0.90 with chronological age [4]. Age acceleration (AA) is the difference between epigenetic age and chronological age, and both DNAmAge and AA are highly correlated with the chronological age. However, another epigenetic age acceleration measure calculated from the residuals of regression (AA_{res}), between epigenetic and chronological ages, is not correlated with chronological ages and is thought to represent true biological effects on age related phenotypes. In addition, another related measure is the intrinsic epigenetic age acceleration (IEAA), which is independent of age related changes of cellular composition of blood [7]. Several recent studies, using the Horvath methods, have found that age acceleration is associated with a number of diseases and phenotypes, such as obesity [8], Alzheimer's disease [9], Down's syndrome [10], Huntington disease [11], HIV [12], Parkinson's disease [13], earlier menopause [14], and overall mortality [15]. Studies have also shown that lung function can be influenced by epigenetic age accelerations as quantified in peripheral blood DNAm [16,17].

Lung development is a continuous process from childhood to adolescence [18]. Low adult lung function can be the result of poor growth in childhood that may cause excessive decline in adult life [19] and it has been found, in many studies, that children with poor lung function also experience reduced lung function at adulthood [20–24]. While lung function is dependent on age, gender, height, and ethnicity [18], it can be influenced by both genetics [25] and environmental exposures [26–28]. Studies have shown that DNAm, measured in peripheral blood, is associated with lung functions [29]. Changes in DNAm from childhood to adolescence have been found associated with lung function during adolescence in females [30]. Therefore, changes in DNAm aging from childhood to adolescence may have potential effect on lung function.

To date only two studies have explored the association of epigenetic aging and lung function. Marioni *et al.* [16] examined the association of various physical measures with epigenetic aging in over 1000 elderly adults (mean age of 69 ± 0.83 years) in the 1936 Mid-Lothian Birth Cohort, which followed up between 3 and 6 years. Lung function, considered as FEV_1 (forced expiratory volume in one second), showed a statistically weak (P-value = 0.05) association with DNAmAge with small effect size (<1 mL change in FEV_1 per additional year of epigenetic aging) and epigenetic aging explained only 0.33% of the variance in FEV_1 decline. In contrast, Rezwan *et al.* [17] explored the association of lung function in two cohorts, namely Swiss study of Air Pollution and Lung and heart Disease in Adults (SAPALDIA) and the European Community Respiratory Health Survey (ECRHS), at two-time points and found AA is cross-sectionally associated with lower FEV_1 (Forced Expiratory Volume in one second) and FVC (Forced Vital Capacity) in females at the follow-up time point only. The findings were both statistically significant and the effect sizes were larger, for FEV_1: between -5.00 mL to -3.02 mL and for FVC: between -8.06 mL to -4.61 mL, in comparison to the previous study. However, both studies dealt with the association of lung function in comparatively older adults focusing on lung function decline, and no such work has been undertaken to explore the effect of epigenetic age measures on the lung function development from childhood to adolescence.

Machine learning approaches are increasingly in use to address healthcare problems. However, to date, no study has been conducted to predict lung functions using machine learning approaches. Few studies incorporated machine learning in lung function tests [31] and diseases related to lung function, such as, Chronic obstructive pulmonary disease (COPD) and asthma [32,33]. Moreover, no work has been done yet to leverage the power of machine learning utilizing the effect of DNAmAge and AAs on lung function.

As part of the Isle of Wight Birth Cohort (IOWBC), DNAm in peripheral blood and lung function at age 10 and 18 years were obtained. Therefore, the aim of the study was to explore the efficacy of the use of machine learning regression models in predicting lung function for subjects at 18 year of age using their epidemiological and epigenetic aging data from both at 10 and 18 years of age.

2. Materials and Methods

2.1. Isle of Wight Birth Cohort

The IOWBC is a population birth cohort of 1536 newborns, recruited between 1989 and 1990 [34]. Informed consent for 1456 infants were obtained from the parents and were enrolled into the longitudinal study. Participants were followed up at 1 or 2, 4, 10, 18 and 26 years and peripheral blood samples were collected at birth (neonatal heel prick on Guthrie cards) and at 10, 18 and 26 years.

2.2. DNA extraction and microarray

DNA was extracted from peripheral blood samples for 326 matched 10 year and 18 year samples. DNAm levels were measured using the Infinium HumanMethylation450 and Methylation EPIC BeadChips from Illumina (Illumina, San Diego, CA, USA) for the 10 year and 18 year old sample, respectively. The CPACOR pipeline was used for quality control and pre-processing DNAm data (β values) [35] and batch effect correction was done using ComBat [36].

2.2. Measuring epigenetic aging

DNAmAge was calculated using the Horvath method, which uses 353 cytosine-phosphate-guanine sites (CpGs) from the Illumina Infinium HumanMethylation450 Beadchip arrays. Age acceleration residuals (AA_{res}) were obtained from a linear regression model by regression of DNAmAge on chronological age and further adjusted for blood cell counts to calculate Intrinsic Epigenetic Age Acceleration (IEAA). Age acceleration measures were estimated using an online calculator (available from: https://dnamage.genetics.ucla.edu/new).

2.4. Feature Selection

FEV_1 and FVC at age 18 were used as the outcome variables. Each subject's sex, weight, height, hay fever status, asthma status, eczema status and smoking status at age 18, and FEV_1 and FVC at age 10 with AA, AA_{res} and IEAA at age 18 were used as features. Mutual information between each feature and the target FEV_1 and FVC at age 18, respectively, was calculated, and features whose mutual information is > 0.1 were selected. Normalisation was undertaken on selected features.

2.5. Machine learning model

Five machine learning regression models: linear, lasso, ridge regression, elastic net, and Bayesian ridge regression were used to predict FEV_1 and FVC at age 18. The best subset of features from the feature selection was used, 10-fold cross-validation was performed along with fine-tuning the hyperparameters using grid search, where applicable. To select the best alpha (hyperparameter that controls the balances between minimizing residual sum of squares vs. minimizing sum of square of coefficients), the models were run for different ranges of alpha and the best alpha was empirically chosen to build the model. Further, age acceleration changes at 10 and 18 years were added by taking differences between epigenetic age acceleration between two age groups (henceforth denoted as: AA_{diff}, $AA_{resdiff}$, and $IEAA_{diff}$).

3. Results

A total of 326 participants with matched data at 10 and 18 years were analysed. Descriptive statistics are given in Table S1.

2.1. Feature selection by mutual information regression

For FEV_1, four features were identified as the most important, namely height, sex, weight, at age 18 and FEV_1 at age 10 (Figure 1 and Table S2). AA, AA_{res} and IEAA exhibited lower mutual information scores (0.041, 0.028, and 0.003 respectively).

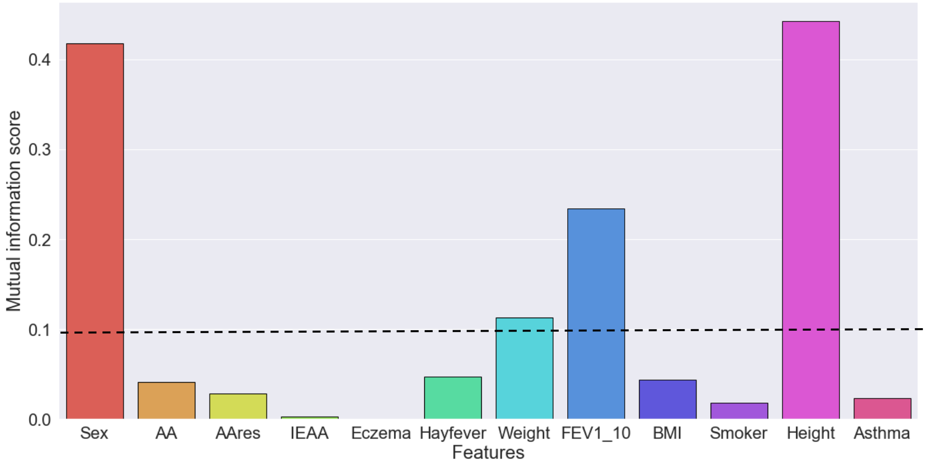


Figure 1. Mutual information score between each feature and the target FEV_1 at age 18. Mutual information score > 0.1 has been used as threshold for selecting the best features.

Similarly, for FVC, the same three features (height, sex, and weight) at age 18 and FVC at age 10 were identified as the most important (Figure S1 and Table S2). AA exhibited lower mutual information scores (0.029) and AA_{res} and IEAA have had mutual information of zero.

2.2. *Machine learning regression models for FEV_1*

With the best four features (height, sex, weight at age 18 and FEV_1 at age 10) for FEV_1, all the regression models performed almost similarly after tuning the hyperparameter. However, ridge regression model (with $\alpha = 0.4$) worked slightly better ($R^2 = 75.03\% \pm 7.37\%$ with RMSE = 0.378 ± 0.064) than other methods (Table 2). As expected based on the mutual information score, adding three age acceleration measures (AA, AA_{res}, and IEAA) respectively, with this four features, did not improve the predictions of FEV_1 (Table S3).

Table 2. Results of five regression models predicting FEV_1 using best features

Regression model	R^2	RMSE
Linear	74.98 ± 7.45	0.3781 ± 0.06380
Lasso ($\alpha = 0.0001$)	74.99 ± 7.45)	0.3781 ± 0.0639
Ridge ($\alpha = 0.4$)	75.03 ± 7.37	0.3780 ± 0.0639
Elastic Net ($\alpha = 0.001$)	75.00 ± 7.41	0.3781 ± 0.0640
Bayesian Ridge	75.01 ± 7.42	0.3780 ± 0.0639

The models were developed using best four features (height, sex, weight at age 18 and FEV_1 at age 10) as predictors of FEV_1. Here, R^2 = average goodness-of-fit measure for regression models represented as percentage and RMSE = average root mean square error

Changes of AA between the two time points (AA_{diff}, $AA_{resdiff}$, and $IEAA_{diff}$) were added with the four predictive features. Although none of the age acceleration differences were found significant during feature selection using mutual information regression, adding AA_{diff} with the other important features showed slight improvement in predicting FEV_1. The best performer was ridge regression model ($R^2 = 75.21\% \pm 7.42\%$ with RMSE = 0.3768 ± 0.0653) (Table 3 and Table S4).

Table 3. Results of five regression models predicting FEV$_1$ using best features and AA$_{diff}$

Regression model	R^2	RMSE
Linear	75.16 ± 7.49	0.3770 ± 0.0652
Lasso (α = 0.0001)	75.16 ± 7.49	0.3770 ± 0.0652
Ridge (α = 0.4)	75.21 ± 7.42	0.3768 ± 0.0653
Elastic Net (α = 0.001)	75.16 ± 7.49	0.3770 ± 0.0653
Bayesian Ridge	75.19 ± 7.46	0.3768 ± 0.0652

The models were developed using best four features (height, sex, weight at age 18 and FEV$_1$ at age 10) with AA$_{diff}$ as predictors of FEV$_1$. Here, AA$_{diff}$ = AA at 18 – AA at 10, R^2 = average goodness-of-fit measure for regression models represented as percentage and RMSE = average root mean square error

2.2. Machine learning regression models for FVC

For FVC, using the best four features (height, sex, weight at age 18 and FVC at age 10), all the regression models performed with similar efficacy after tuning the hyperparameters. The elastic net regression model (with α = 0.0025) performed slightly better (R^2 = 75.35% ± 6.88% with RMSE = 0.445 ± 0.064) than other methods (Table 4). Adding three age acceleration measures (AA, AA$_{res}$, and IEAA) respectively, with this four features, did not improve the predictions of FVC (Table S5).

Table 4. Results of five regression models predicting FVC using best features

Regression model	R^2	RMSE
Linear	75.24 ± 7.10	0.4455 ± 0.0692
Lasso (α = 0.0001)	75.25 ± 7.08	0.4456 ± 0.0680
Ridge (α = 0.4)	75.24 ± 7.00	0.4458 ± 0.0673
Elastic Net (α = 0.0025)	75.35 ± 6.88	0.4450 ± 0.0673
Bayesian Ridge	75.25 ± 7.07	0.4456 ± 0.0678

The models were developed using best four features (height, sex, weight at age 18 and FVC at age 10) as predictors of FVC. Here, R^2 = average goodness-of-fit measure for regression models represented as percentage and RMSE = average root mean square error

While adding changes of AA (AA$_{diff}$, AA$_{resdiff}$, and IEAA$_{diff}$) with the four predictive features, showed almost similar prediction capacity for FVC (Table S6). The best performer was Elastic net regression model (R^2 = 75.38% ± 6.98% with RMSE = 0.445 ± 0.069) (Table 5)

Table 5. Results of five regression models predicting FVC using best features and AA$_{diff}$

Regression model	R^2	RMSE
Linear	75.26 ± 7.14	0.4456 ± 0.0693
Lasso (α = 0.0001)	75.27 ± 7.12	0.4456 ± 0.0692
Ridge (α = 0.4)	75.28 ± 7.12	0.4455 ± 0.0691
Elastic Net	75.38 ± 6.98	0.4448 ± 0.0690

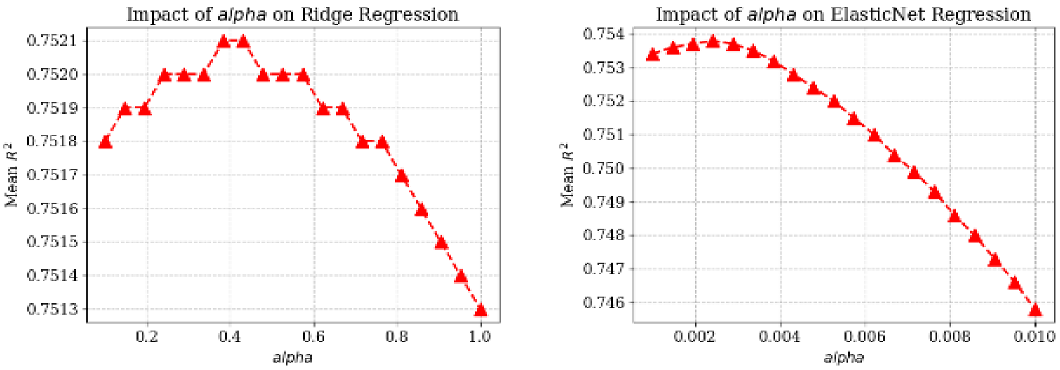
doi:10.20944/preprints202009.0168.v1

(α = 0.0025)		
Bayesian Ridge	75.28 ± 7.13	0.4455 ± 0.0692

The models were developed using best four features (height, sex, weight at age 18 and FVC at age 10) with AA_{diff} as predictors of FVC. Here, AA_{diff} = AA at 18 – AA at 10, R^2 = average goodness-of-fit measure for regression models represented as percentage and RMSE = average root mean square error

2.3. Effect of alpha on Ridge regression model

The choice of α affects the mean R^2 values for the regression models. Figure 2(a) shows how the choice of α affects the mean R^2 values for ridge regression for the FEV_1 prediction and the best R^2 value was achieved with $\alpha = 0.4$. Similar behavior is noticed for the elastic net regression for the FVC prediction (Figure 2b).



(**a**) (**b**)

Figure 2. Impact of hyperparameter (α) on (a) ridge regression and (b) elastic net regression

5. Discussion

Using the data at two time-points (10 and 18 years) from IOWBC, we have explored whether epigenetic aging can be utilised together with other features for predicting lung function in adolescence using machine learning regression models. Epigenetic age acceleration at 18 years did not contribute in improving the prediction of lung function at 18 years of age. However, using changes in age acceleration between 10 and 18 years improved the prediction of FEV_1 slightly, despite the fact that the mutual information scores thereof indicated otherwise. Similar improvement, although in an even smaller scale was observed for FVC.

This is a novel study that examines the effect of epigenetic age acceleration on lung function using supervised machine learning techniques. The previous two studies, examining the association between lung function and epigenetic aging, were performed in older population and were more focused on lung function decline rather than development. The participants from Mid-Lothian birth cohort study were 70 years at baseline and 76 years at follow-up, and participant from ALEC project were 37 to 61 years at baseline and 48 to 70 years at follow-up, whereas participants from IOWBC were matched samples at 10 and 18 years.

In this study, changes of epigenetic age acceleration, between 10 years and 18 years, was incorporated with the most informative features from the feature selection technique to develop the best regression models. Previous studies have found height, weight and sex to be important predictors of lung function [18,37,38]. This study confirms these previous observations and adds lung function at an earlier time point (10-years of age), which confirms the efficacy of machine learning in identifying predictors for lung function. Our study suggests that changes in epigenetic age acceleration between 10 and 18 years can improve the prediction of FEV_1 and FVC at 18 years of age. Based on the prediction performances of the five selected regression models, it can be postulated that any of these supervised machine learning techniques can be used for lung function prediction.

The fine-tuning of hyperparameters always plays a crucial role in the efficacy of a machine learning technique and we have shown that choice of the hyperparameter (α) changes the prediction result drastically. Therefore, a grid search was performed for identifying the most optimized parameters for the models to achieve the best prediction performance. This is evident from the higher average R^2 and lower RMSE values of each regression model. The best models can explain 75.16% and 75.38% of variance for FEV_1 and FVC, respectively through weight, height, sex at 18 years and lung function at 10 years in conjunction with changes of epigenetic age acceleration between 10 and 18 years. The RMSE values are also very low for each model (0.3768 ± 0.0653 and 0.4448 ± 0.0690, for FEV_1 and FVC model, respectively).

Our study has some limitations. Firstly, due to a relatively smaller sample size (n = 326), 10-fold cross-validation was used to generate average performance measures of the models rather than using a hold-out test set. However, the cross-validation method performs better to break the bias variance tradeoff in small datasets [39]. Secondly, epigenetic age derived from blood was used rather than lung tissue. However, successful use of epigenetic aging measured from blood is evident in a number of other non-blood related diseases and phenotypes, such as: developmental disorder [40], lung cancer [41], and metabolic syndrome [8]. Additionally, physiological changes, such as hormonal changes during adolescence, were not considered. Moreover, sex-stratified analysis, for lung function, has proven informative in other studies [17,30] and therefore, this could be implemented in this study as well. However, this would further lower the samples size (43.25% female) and may be impractical for this study.

In conclusion, while the full impact of epigenetic age acceleration is still unknown from DNA methylation measures, this study suggests that it can be utilised as one of the potential factors to predict adolescent lung function. It also suggests that application of machine learning in conjunction with tracking changes in epigenetic age acceleration over life span can be beneficial to assess lung health in adolescent and have the potential to be extended to adults.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: Mutual information score between each feature and the target which is FVC at age 18., Table S1: Summary of the variables IOWBC 10- and 18-year matched samples, Table S2: Mutual information regression scores for predicting FEV_1 and FVC at 18 years, Table S3: Results of five regression models predicting FEV_1 using best features and AAs., Table S4: Results of five regression models predicting FEV_1 using best features and $AA_{resdiff}$ and $IEAA_{diff}$, Table S5: Results of five regression models predicting FVC using best features and AAs, Table S6: Results of five regression models predicting FVC using best features and $AA_{resdiff}$ and $IEAA_{diff}$.

Author Contributions: Conceptualization, F.I.R., J.W.H. and M.S.R.; formal analysis, M.A.A. and S.T.N.; Cohort assessment and data generation S.H.A., J.W.H.; writing—original draft preparation, M.A.A., S.T.N. and F.I.R.; writing—review and editing, A.A., S.T., F.I.R., J.W.H., M.S.R. and S.H.A.; supervision, F.I.R., J.W.H. and M.S.R.; project administration, M.A.A., S.T.N. and F.I.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Institutes of Health USA (Grant No. R01 HL082925, R01 AI091905, R01 AI121226, and R01 HL132321) and Asthma UK (Grant No. 364), which supported the assessment and methylation analysis of the Isle of Wight Birth Cohort.

Acknowledgments: We would like to acknowledge the help of all the staff at The David Hide Asthma and Allergy Research Centre in undertaking the assessments of the Isle of Wight birth cohort and Nikki Graham for technical support. Our sincere thanks to the participants and their families who helped us with this project over the last three decades.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Horvath, S.; Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet*. **2018**, *19*, 371–384, doi:10.1038/s41576-018-0004-3.
2. Bocklandt, S.; Lin, W.; Sehl, M.E.; Sánchez, F.J.; Sinsheimer, J.S.; Horvath, S.; Vilain, E. Epigenetic Predictor of Age. *PLoS ONE* **2011**, *6*, e14821, doi:10.1371/journal.pone.0014821.
3. Hannum, G.; Guinney, J.; Zhao, L.; Zhang, L.; Hughes, G.; Sadda, S.; Klotzle, B.; Bibikova, M.; Fan, J.-B.; Gao, Y.; et al. Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates. *Mol. Cell* **2013**, *49*, 359–367, doi:10.1016/j.molcel.2012.10.016.
4. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **2013**, *14*, R115, doi:10.1186/gb-2013-14-10-r115.
5. Jones, M.J.; Goodman, S.J.; Kobor, M.S. DNA methylation and healthy human aging. *Aging Cell* **2015**, *14*, 924–932, doi:10.1111/acel.12349.
6. Weidner, C.; Lin, Q.; Koch, C.; Eisele, L.; Beier, F.; Ziegler, P.; Bauerschlag, D.; Jöckel, K.-H.; Erbel, R.; Mühleisen, T.; et al. Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol.* **2014**, *15*, R24, doi:10.1186/gb-2014-15-2-r24.
7. Chen, B.H.; Marioni, R.E.; Colicino, E.; Peters, M.J.; Ward-Caviness, C.K.; Tsai, P.-C.; Roetker, N.S.; Just, A.C.; Demerath, E.W.; Guan, W.; et al. DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging* **2016**, *8*, 1844–1865, doi:10.18632/aging.101020.
8. Quach, A.; Levine, M.E.; Tanaka, T.; Lu, A.T.; Chen, B.H.; Ferrucci, L.; Ritz, B.; Bandinelli, S.; Neuhouser, M.L.; Beasley, J.M.; et al. Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging* **2017**, *9*, 419–446, doi:10.18632/aging.101168.
9. Levine, M.E.; Lu, A.T.; Bennett, D.A.; Horvath, S. Epigenetic age of the pre-frontal cortex is associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive functioning. *Aging* **2015**, *7*, 1198–1211, doi:10.18632/aging.100864.
10. Horvath, S.; Garagnani, P.; Bacalini, M.G.; Pirazzini, C.; Salvioli, S.; Gentilini, D.; Di Blasio, A.M.; Giuliani, C.; Tung, S.; Vinters, H.V.; et al. Accelerated epigenetic aging in Down syndrome. *Aging Cell* **2015**, *14*, 491–495, doi:10.1111/acel.12325.
11. Horvath, S.; Langfelder, P.; Kwak, S.; Aaronson, J.; Rosinski, J.; Vogt, T.F.; Eszes, M.; Faull, R.L.M.; Curtis, M.A.; Waldvogel, H.J.; et al. Huntington's disease accelerates epigenetic aging of human brain and disrupts DNA methylation levels. *Aging* **2016**, *8*, 1485–1512, doi:10.18632/aging.101005.
12. Horvath, S.; Levine, A.J. HIV-1 Infection Accelerates Age According to the Epigenetic Clock. *J. Infect. Dis.* **2015**, *212*, 1563–1573, doi:10.1093/infdis/jiv277.
13. Horvath, S.; Ritz, B.R. Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging* **2015**, *7*, 1130–1142, doi:10.18632/aging.100859.
14. Levine, M.E.; Lu, A.T.; Chen, B.H.; Hernandez, D.G.; Singleton, A.B.; Ferrucci, L.; Bandinelli, S.; Salfati, E.; Manson, J.E.; Quach, A.; et al. Menopause accelerates biological aging. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 9327–9332, doi:10.1073/pnas.1604558113.
15. Marioni, R.E.; Shah, S.; McRae, A.F.; Chen, B.H.; Colicino, E.; Harris, S.E.; Gibson, J.; Henders, A.K.; Redmond, P.; Cox, S.R.; et al. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol.* **2015**, *16*, 25, doi:10.1186/s13059-015-0584-6.

16. Marioni, R.E.; Shah, S.; McRae, A.F.; Ritchie, S.J.; Muniz-Terrera, G.; Harris, S.E.; Gibson, J.; Redmond, P.; Cox, S.R.; Pattie, A.; et al. The epigenetic clock is correlated with physical and cognitive fitness in the Lothian Birth Cohort 1936. *Int. J. Epidemiol.* **2015**, *44*, 1388–1396, doi:10.1093/ije/dyu277.
17. Rezwan, F.I.; Imboden, M.; Amaral, A.F.; Wielscher, M.; Schaffner, E.; Jarvis, D.; Probst-Hensch, N.; Holloway, J.W. Association between accelerated biological aging and adult lung function. *Eur. Respir. J.* **2017**, *50*, doi:10.1183/1393003.congress-2017.OA2945.
18. Miller, M.R.; Hankinson, J.; Brusasco, V.; Burgos, F.; Casaburi, R.; Coates, A.; Crapo, R.; Enright, P.; van der Grinten, C.P.M.; Gustafsson, P.; et al. Standardisation of spirometry. *Eur. Respir. J.* **2005**, *26*, 319–338, doi:10.1183/09031936.05.00034805.
19. Dyer, C. The interaction of ageing and lung disease. *Chron. Respir. Dis.* **2012**, *9*, 63–67, doi:10.1177/1479972311433766.
20. Stern, D.A.; Morgan, W.J.; Wright, A.L.; Guerra, S.; Martinez, F.D. Poor airway function in early infancy and lung function by age 22 years: a non-selective longitudinal cohort study. *Lancet Lond. Engl.* **2007**, *370*, 758–764, doi:10.1016/S0140-6736(07)61379-8.
21. Belgrave, D.C.M.; Granell, R.; Turner, S.W.; Curtin, J.A.; Buchan, I.E.; Le Souëf, P.N.; Simpson, A.; Henderson, A.J.; Custovic, A. Lung function trajectories from pre-school age to adulthood and their associations with early life factors: a retrospective analysis of three population-based birth cohort studies. *Lancet Respir. Med.* **2018**, *6*, 526–534, doi:10.1016/S2213-2600(18)30099-7.
22. Bui, D.S.; Lodge, C.J.; Burgess, J.A.; Lowe, A.J.; Perret, J.; Bui, M.Q.; Bowatte, G.; Gurrin, L.; Johns, D.P.; Thompson, B.R.; et al. Childhood predictors of lung function trajectories and future COPD risk: a prospective cohort study from the first to the sixth decade of life. *Lancet Respir. Med.* **2018**, *6*, 535–544, doi:10.1016/S2213-2600(18)30100-0.
23. Tai, A.; Tran, H.; Roberts, M.; Clarke, N.; Wilson, J.; Robertson, C.F. The association between childhood asthma and adult chronic obstructive pulmonary disease. *Thorax* **2014**, *69*, 805–810, doi:10.1136/thoraxjnl-2013-204815.
24. Sears, M.R.; Greene, J.M.; Willan, A.R.; Wiecek, E.M.; Taylor, D.R.; Flannery, E.M.; Cowan, J.O.; Herbison, G.P.; Silva, P.A.; Poulton, R. A longitudinal, population-based, cohort study of childhood asthma followed to adulthood. *N. Engl. J. Med.* **2003**, *349*, 1414–1422, doi:10.1056/NEJMoa022363.
25. Tarnoki, D.L.; Tarnoki, A.D.; Lazar, Z.; Medda, E.; Littvay, L.; Cotichini, R.; Fagnani, C.; Stazi, M.A.; Nisticó, L.; Lucatelli, P.; et al. Genetic and environmental factors on the relation of lung function and arterial stiffness. *Respir. Med.* **2013**, *107*, 927–935, doi:10.1016/j.rmed.2013.02.002.
26. Adam, M.; Schikowski, T.; Carsin, A.E.; Cai, Y.; Jacquemin, B.; Sanchez, M.; Vierkötter, A.; Marcon, A.; Keidel, D.; Sugiri, D.; et al. Adult lung function and long-term air pollution exposure. ESCAPE: a multicentre cohort study and meta-analysis. *Eur. Respir. J.* **2015**, *45*, 38–50, doi:10.1183/09031936.00130014.
27. Burchfiel, C.M.; Marcus, E.B.; Curb, J.D.; Maclean, C.J.; Vollmer, W.M.; Johnson, L.R.; Fong, K.O.; Rodriguez, B.L.; Masaki, K.H.; Buist, A.S. Effects of smoking and smoking cessation on longitudinal decline in pulmonary function. *Am. J. Respir. Crit. Care Med.* **1995**, *151*, 1778–1785, doi:10.1164/ajrccm.151.6.7767520.
28. Sunyer, J. Lung function effects of chronic exposure to air pollution. *Thorax* **2009**, *64*, 645–646, doi:10.1136/thx.2009.115071.
29. Imboden, M.; Wielscher, M.; Rezwan, F.I.; Amaral, A.F.S.; Schaffner, E.; Jeong, A.; Beckmeyer-Borowko, A.; Harris, S.E.; Starr, J.M.; Deary, I.J.; et al. Epigenome-wide association study of lung function level and its change. *Eur. Respir. J.* **2019**, doi:10.1183/13993003.00457-2019.

30. Sunny, S.K.; Zhang, H.; Rezwan, F.I.; Relton, C.L.; Henderson, A.J.; Merid, S.K.; Melén, E.; Hallberg, J.; Arshad, S.H.; Ewart, S.; et al. Changes of DNA methylation are associated with changes in lung function during adolescence. *Respir. Res.* **2020**, *21*, 80, doi:10.1186/s12931-020-01342-y.
31. Topalovic, M.; Das, N.; Burgel, P.-R.; Daenen, M.; Derom, E.; Haenebalcke, C.; Janssen, R.; Kerstjens, H.A.M.; Liistro, G.; Louis, R.; et al. Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. *Eur. Respir. J.* **2019**, *53*, doi:10.1183/13993003.01660-2018.
32. Finkelstein, J.; Jeong, I.C. Machine learning approaches to personalize early prediction of asthma exacerbations. *Ann. N. Y. Acad. Sci.* **2017**, *1387*, 153–165, doi:10.1111/nyas.13218.
33. Tomita, K.; Nagao, R.; Touge, H.; Ikeuchi, T.; Sano, H.; Yamasaki, A.; Tohda, Y. Deep learning facilitates the diagnosis of adult asthma. *Allergol. Int.* **2019**, *68*, 456–461, doi:10.1016/j.alit.2019.04.010.
34. Arshad, S.H.; Holloway, J.W.; Karmaus, W.; Zhang, H.; Ewart, S.; Mansfield, L.; Matthews, S.; Hodgekiss, C.; Roberts, G.; Kurukulaaratchy, R. Cohort Profile: The Isle Of Wight Whole Population Birth Cohort (IOWBC). *Int. J. Epidemiol.* **2018**, doi:10.1093/ije/dyy023.
35. Lehne, B.; Drong, A.W.; Loh, M.; Zhang, W.; Scott, W.R.; Tan, S.-T.; Afzal, U.; Scott, J.; Jarvelin, M.-R.; Elliott, P.; et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol.* **2015**, *16*, doi:10.1186/s13059-015-0600-x.
36. Johnson, W.E.; Li, C.; Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **2006**, *8*, 118–127, doi:10.1093/biostatistics/kxj037.
37. Mottram, C. *Ruppel's manual of pulmonary function testing*; 2018; ISBN 978-0-323-44560-3.
38. LoMauro, A.; Aliverti, A. Sex differences in respiratory function. *Breathe Sheff. Engl.* **2018**, *14*, 131–140, doi:10.1183/20734735.000318.
39. Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A.J. Machine learning algorithm validation with a limited sample size. *PLOS ONE* **2019**, *14*, e0224365, doi:10.1371/journal.pone.0224365.
40. Walker, R.F.; Liu, J.S.; Peters, B.A.; Ritz, B.R.; Wu, T.; Ophoff, R.A.; Horvath, S. Epigenetic age analysis of children who seem to evade aging. *Aging* **2015**, *7*, 334–339, doi:10.18632/aging.100744.
41. Levine, M.E.; Hosgood, H.D.; Chen, B.; Absher, D.; Assimes, T.; Horvath, S. DNA methylation age of blood predicts future onset of lung cancer in the women's health initiative. *Aging* **2015**, *7*, 690–700, doi:10.18632/aging.100809.