Research Paper

NEOage clocks - epigenetic clocks to estimate post-menstrual and postnatal age in preterm infants

Stefan Graw[1], Marie Camerota[2], Brian S. Carter[3], Jennifer Helderman[4], Julie A. Hofheimer[5], Elisabeth C. McGowan[6], Charles R. Neal[7], Steven L. Pastyrnak[8], Lynne M. Smith[9], Sheri A. DellaGrotta[10], Lynne M. Dansereau[10], James F. Padbury[6], Michael O'Shea[5], Barry M. Lester[2,6,10,11], Carmen J. Marsit[1], Todd M. Everson[1]

[1]Gangarosa Department of Environmental Health, Emory University Rollins School of Public Health, Atlanta, GA 30322, USA
[2]Department of Psychiatry and Human Behavior, Brown University, Providence, RI 02906, USA
[3]Department of Pediatrics-Neonatology, Children's Mercy Hospital, Kansas City, MO 64108, USA
[4]Department of Pediatrics, Wake Forest School of Medicine, Winston-Salem, NC 27101, USA
[5]Department of Pediatrics, University of North Carolina School of Medicine, Chapel Hill, NC 27599, USA
[6]Department of Pediatrics, Brown Alpert Medical School and Women and Infants Hospital, Providence, RI 02912, USA
[7]Department of Pediatrics, University of Hawaii John A. Burns School of Medicine, Honolulu, HI 96813, USA
[8]Department of Pediatrics, Spectrum Health-Helen Devos Hospital, Grand Rapids, MI 49503, USA
[9]Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA 90502, USA
[10]Brown Center for the Study of Children at Risk, Brown Alpert Medical School and Women and Infants Hospital, Providence, RI 02912, USA
[11]Department of Psychiatry and Human Behavior, Brown Alpert Medical School, Providence, RI 02906, USA

Correspondence to: Todd M. Everson; **email:** Todd.M.Everson@Emory.edu
Keywords: DNA methylation, neonatal aging, epigenetic clock, preterm infants
Received: May 24, 2021 **Accepted:** September 28, 2021 **Published:** October 15, 2021

ABSTRACT

Epigenetic clocks based on DNA methylation (DNAm) can accurately predict chronological age and are thought to capture biological aging. A variety of epigenetic clocks have been developed for different tissue types and age ranges, but none have focused on postnatal age prediction for preterm infants. Epigenetic estimators of biological age might be especially informative in epidemiologic studies of neonates since DNAm is highly dynamic during the neonatal period and this is a key developmental window. Additionally, markers of biological aging could be particularly important for those born preterm since they are at heightened risk of developmental impairments. We aimed to fill this gap by developing epigenetic clocks for neonatal aging in preterm infants.

As part of the Neonatal Neurobehavior and Outcomes in Very Preterm Infants (NOVI) study, buccal cells were collected at NICU discharge to profile DNAm levels in 542 very preterm infants. We applied elastic net regression to identify four epigenetic clocks (NEOage Clocks) predictive of post-menstrual and postnatal age, compatible with the Illumina EPIC and 450K arrays. We observed high correlations between predicted and reported ages (0.93 – 0.94) with root mean squared errors (1.28 - 1.63 weeks).

Epigenetic estimators of neonatal aging in preterm infants can be useful tools to evaluate biological maturity and associations with neonatal and long-term morbidities.

INTRODUCTION

DNA methylation (DNAm) is one of the most studied epigenetic mechanisms and acts at the interface between the environment and human health. Changes in DNAm are also strongly correlated with aging [1] and are most dynamic during pediatric age [2]. Aging-related fluctuations in DNAm levels have been capitalized on by researchers to develop "epigenetic clocks", sets of CpG sites whose methylation extents have been shown to accurately predict chronological age and are thought to capture biological aging [2, 3]. These predicted ages are often referred to as epigenetic age or DNAm age. Greater DNAm age relative to chronological age, also known as age acceleration (AA), has been shown to be associated with age-related phenotypes in adults, such as frailty, chronic diseases and mortality [4].

A variety of epigenetic clocks have been developed to predict numerous age metrics in different tissue types and age ranges [5]. One of the most widely used pan-tissue clocks to estimate chronological age was created by Horvath and is based on over 8,000 samples from 51 healthy tissues (age range: 0-101 years) [6]. However, DNAm age estimates from Horvath's epigenetic clock become more precise as chronological age increases and are most variable in pediatric samples [7]. Hannum et al. developed a clock based on blood with an age range of 19-101 years [8] while other clocks are designed to capture physiological measures of biological age rather than chronological age. These include DNAm PhenoAge [9] and DNAm GrimAge [10] and are both blood-based. Many studies have successfully generated epigenetic clocks for various tissues, age ranges, and morbidities, leading to very promising predictors of chronological age in adults, and to potentially useful biomarkers for the diseases of aging. While some epigenetic clocks include children, most clocks are primarily focused on adults and extrapolating them to children results in inaccurate predictions [2, 11]. Additionally, AA metrics that are derived from these clocks may not be as relevant to the health conditions that are most important to children and adolescents. To address this issue McEwen et al. developed PedBE, an epigenetic clock that focuses on estimating chronological age of children ranging from 0 (birth) to 20 years old and is based on buccal epithelial cells [2]. However, the definition of chronological age becomes less meaningful proximal to birth and is especially skewed among infants born preterm. Infants born preterm might differ biologically from infants of the same chronologic or postnatal age that are born full-term. Epigenetic clocks, such as those developed by Knight et al. [12] or Bohlin et al. [13], have been created to capture gestational age (GA), i.e. the time from conception to birth. Both clocks are based on cord blood and therefore can only estimate GA, not postnatal age. To our knowledge, there exists no epigenetic clock that properly handles or is specialized for age prediction in preterm infants.

The WHO estimated 15 million infants, approximately 10% of live births, are born prematurely early every year (before completing 37 weeks of gestation) [14]. Preterm birth is not only associated with acute and long-term morbidities including chronic illnesses, brain injuries, and adverse neuromotor, cognitive, and behavioral outcomes [15], but it is also the leading cause of death worldwide among children under 5 years [14]. This leads to an immense emotional and financial burden for families and society. The Institute of Medicine reported in 2007 that the average medical

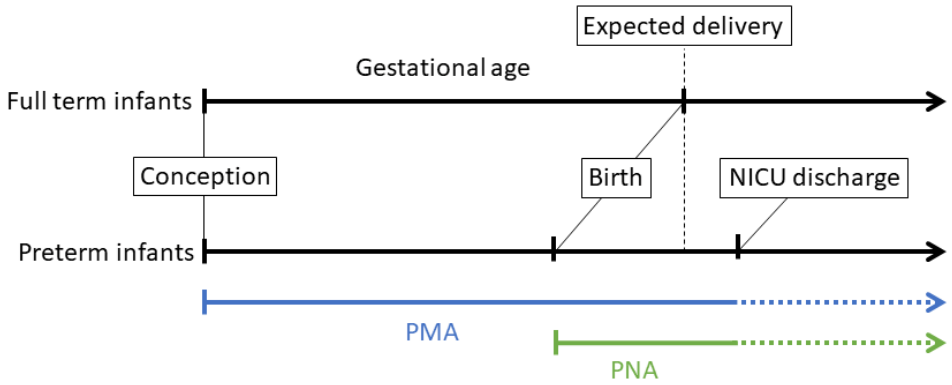

Figure 1. Illustration of different perinatal age metrics, measured in weeks and days, which we highlight for infants born preterm. Gestational age (GA) is defined as the time from conception to birth (expected delivery around 37-42 weeks typically refers to full-term birth, and <37 weeks refers to preterm birth). Post-menstrual age (PMA) refers to the time from conception onward, and postnatal age (PNA) is equivalent to chronological age and is the time elapsed after birth. In this study, buccal cell tissue was collected from infants at NICU discharge to profile DNA methylation.

costs of the first year were almost 10 times greater for preterm infants in the U.S., and results in a societal economic cost of $26.2 billion each year [16, 17].

Here, we present four NEOage (Neonatal Epigenetic Estimator of age) clocks, epigenetic clocks that are focused on age estimation of preterm infants based on their DNAm profile measured in an easily accessible tissue, buccal epithelial cells. Specifically, we investigated post-menstrual age (PMA), the time from conception to tissue collection at neonatal intensive care unit (NICU) discharge, and post-natal age (PNA, or chronological age), the time from birth to tissue collection (Figure 1). These epigenetic estimators of aging could be particularly important for preterm neonates because they may provide insight into early life aging, reflect health and development, and provide a measure of early life risk for neonatal morbidities or long-term neurodevelopmental impairments.

RESULTS

We applied elastic net regression to identify the sets of CpGs that are predictive of PMA and PNA in a unique population of 542 preterm neonates (see Table 1 for characteristics of the study sample). We compared the prediction performances of our NEOage clocks to two existing epigenetic clocks (Horvath's skin-blood clock and PedBE) by evaluating their performances in our Neonatal Neurobehavior and Outcomes in Very Preterm Infants (NOVI) data set (buccal cells) and an external saliva data set.

NEOage clocks

We identified four epigenetic clocks predictive of either PMA or PNA that are compatible with the Infinium MethylationEPIC BeadChip (EPIC) array or Infinium HumanMethylation450 BeadChip (450k) array. The number of CpGs within each clock range from 303-522 CpGs with varying degrees of overlap between the clocks (see Figure 2). CpGs for each NEOage clock with the corresponding coefficients to calculate DNAm age are provided in the Supplementary Material (Supplementary Tables 1–4 and Supplementary Code 1).

To assess the prediction performances without reusing information we performed leave-one-out (LOO) cross-validation (additional information in 5.3 Development of the epigenetic clock) and evaluated prediction performances using correlations and root mean squared error ($RMSE = \sqrt{\sum_{i=1}^{N} \frac{(x_i - \hat{x}_i)^2}{N}}$ with x_i and $\hat{x}_i$ being the observed and estimated age, respectively). We observed very strong positive correlations between predicted and measured age metrics ($r > 0.9$ and p-values $< 10^{-16}$) with very similar correlation coefficients among our four NEOage clocks (Figure 3). The predictions for PMA achieved RMSEs of 1.28 for

Table 1. Characteristics of the study population (N=542).

Sample characteristics	N (%) / Median (IQR)
Infant sex	
Male	301 (55.5)
Female	241 (44.5)
Race and Ethnicity	
White	280 (52.2)
Black	123 (22.9)
Asian	41 (7.6)
Hawaiian / Pacific Islander	38 (7.1)
Other	54 (10.1)
Ethnicity	
Non-Hispanic	419 (78.2)
Hispanic	117 (21.8)
PMA (weeks)	38.57 (4.43)
PNA (weeks)	11.43 (6.39)
Gestational age (weeks)	27.29 (3.14)
Birthweight (grams)	919 (430)
Maternal age (years)	28.50 (9.25)
Serious infection	103 (19.11)
Bronchopulmonary dysplasia	277 (51.39)
Severe brain injury	69 (12.80)
Retinopathy	34 (6.31)

PMA, postmenstrual age; PNA, postnatal age.

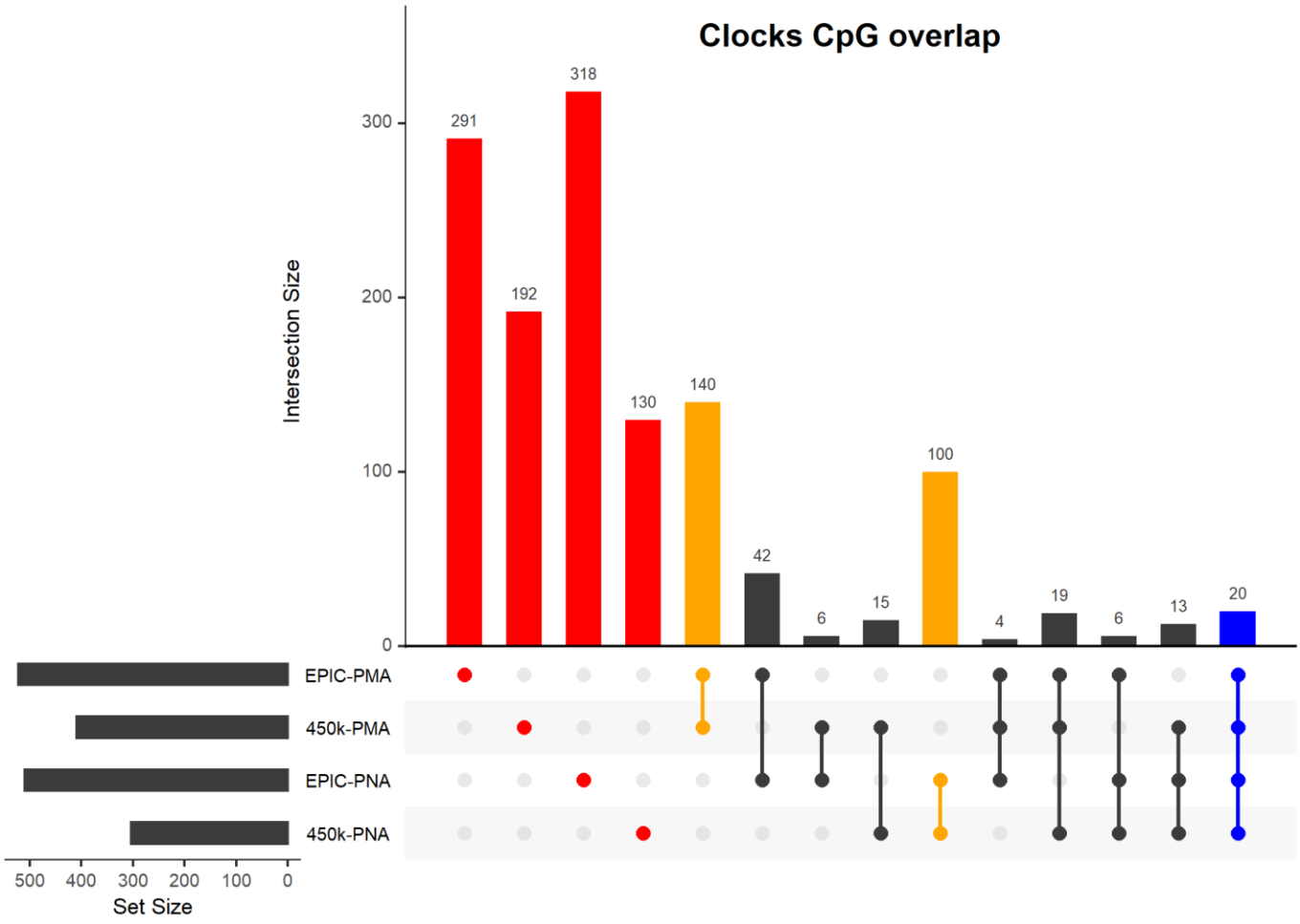

Figure 2. Upset plot of CpGs included in our four NEOage clocks. Highlighted in red are the number of CpGs that are unique to each individual clock. Highlighted in orange are the number of overlapping CpGs of clocks that are predictive of either PMA or PNA. Highlighted in blue are the number of CpGs that overlapped in all four clocks (additional information for the 20 common CpGs provided in Supplementary Table 13). Highlighted in black are the number of overlapping CpGs of clocks where at least one clock is predictive of PMA and at least one clock is predictive of PNA.

the 450k and EPIC clocks, while predictions of PNA resulted in a RMSEs of 1.63 and 1.55, for the 450k and EPIC clocks respectively. The scatterplots in Figure 3 in combination with the strong correlations and low RMSE indicate high accuracy of our NEOage clocks.

Next, we evaluated the prediction performance of our 450k clocks in an external independent data set that measured DNAm in saliva tissue using the 450k array. This external saliva data (GSE72120 [18]) includes preterm (n=34) and full-term infants (n=14) for which PMA (median = 40.15; IQR = 2.61 weeks) and PNA (median = 9.79; IQR = 7.64) were available. While Figure 4 visualizes both preterm and full-term infants, we first focused on only preterm infants in the prediction performance assessment of our NEOage clocks. Focusing on preterm infants of the saliva data allows for a more appropriate comparison of the two data sets. The prediction performances in the external saliva data set resulted in diminished but still strong correlations (PMA: r=0.61 and PNA: r=0.76), and lower RMSE for PMA (RMSE = 1.09) and similar RMSE for PNA (RMSE = 1.55), compared to the NOVI data set. However, it is important to note that the ranges of PMA and PNA in preterm infants of the saliva data are 38-42.6 and 6.9-17.6 weeks, respectively. These ranges are noticeably smaller than the ranges of PMA and PNA in the NOVI data set (PMA: 32.1-51.4 weeks; PNA: 2.7-25.3 weeks) and is likely one reason for lower correlation coefficients between predicted and reported ages in this dataset.

While we observed strong predictive performance for our newly developed NEOage clocks, the existing Horvath skin-blood clock and PedBE clock did not predict PNA as accurately in preterm infants. As shown

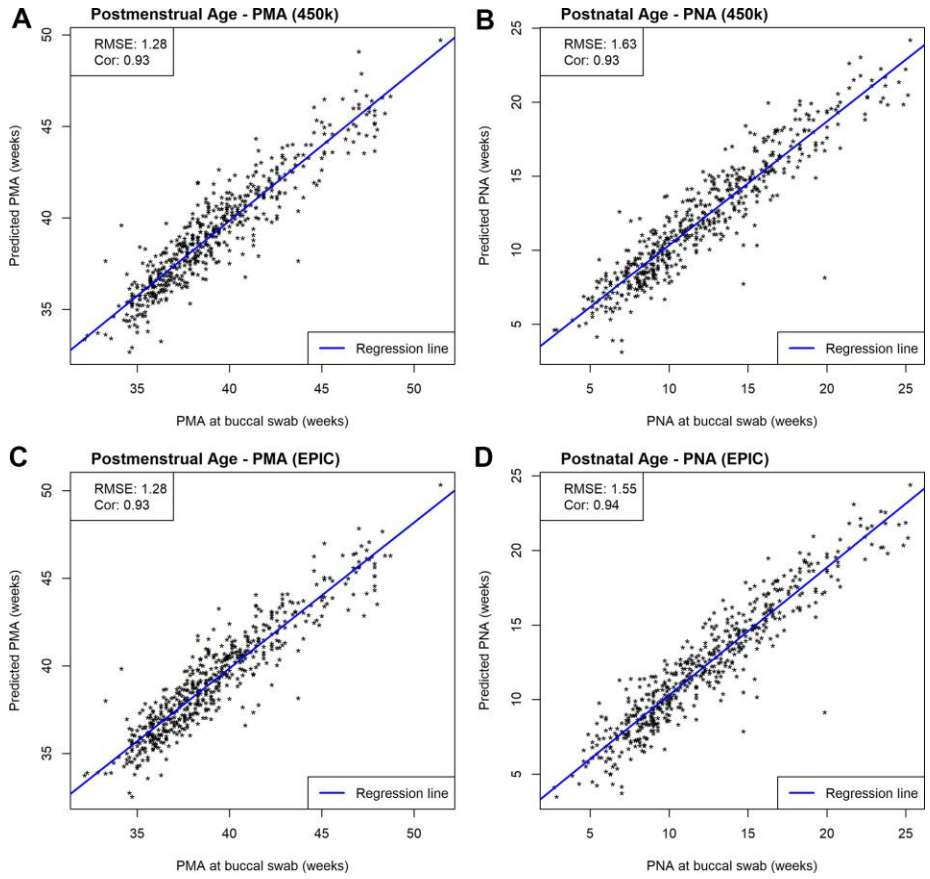

Figure 3. Scatterplots of estimated and measured age. Prediction performances are evaluated by RMSE and correlations between estimated and measured age metrics. (**A**) Scatterplots of estimated and measured PMA using our 450k NEOage clocks within NOVI. (**B**) Scatterplots of estimated and measured PNA using our 450k NEOage clocks within NOVI. (**C**) Scatterplots of estimated and measured PMA using our EPIC NEOage clocks within NOVI. (**D**) Scatterplots of estimated and measured PNA using our EPIC NEOage clocks within NOVI.

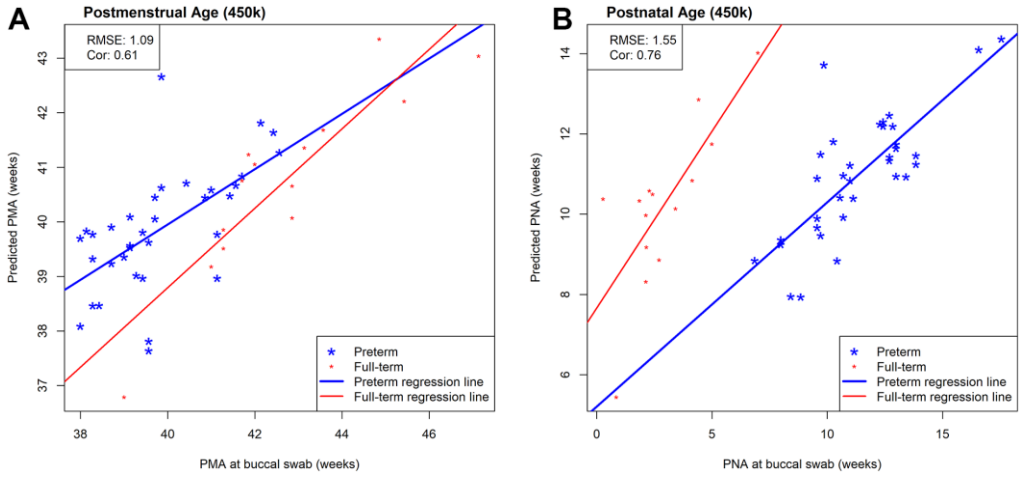

Figure 4. Scatterplots of estimated and measured age using our 450k NEOage clocks in an external saliva data set (GSE72120 [18]) that included full-term (red) and preterm (blue) infants. This saliva data set was measured by the 450k array. The reported prediction performances, RMSE and correlation coefficients between estimated and measured age metrics are based on preterm infants only, since our NOVI training data did not include any full-term infants. (**A**) Scatterplots of estimated and measured PMA. (**B**) Scatterplots of estimated and measured PNA.

in Figure 5, the correlations between estimated and measured PNA are moderate in the NOVI data set (Horvath: r = 0.44 and PedBE: r = 0.59). The RMSE are greater for both clocks, with a noticeably greater RMSE for Horvath's skin-blood clock (Horvath: *RMSE* = 49.68 and PedBE: *RMSE* = 8.68). Additionally, our NEOage clocks outperformed the existing clocks in the independent saliva data set. Analogously, Figure 6 displays both preterm and full-term infants. For preterm infants (highlighted in blue), Horvath skin-blood clock and PedBE exhibit weak correlations (Horvath: r = 0.31 and PedBE: r = 0.19) with RMSE of 38.49 and 12.93

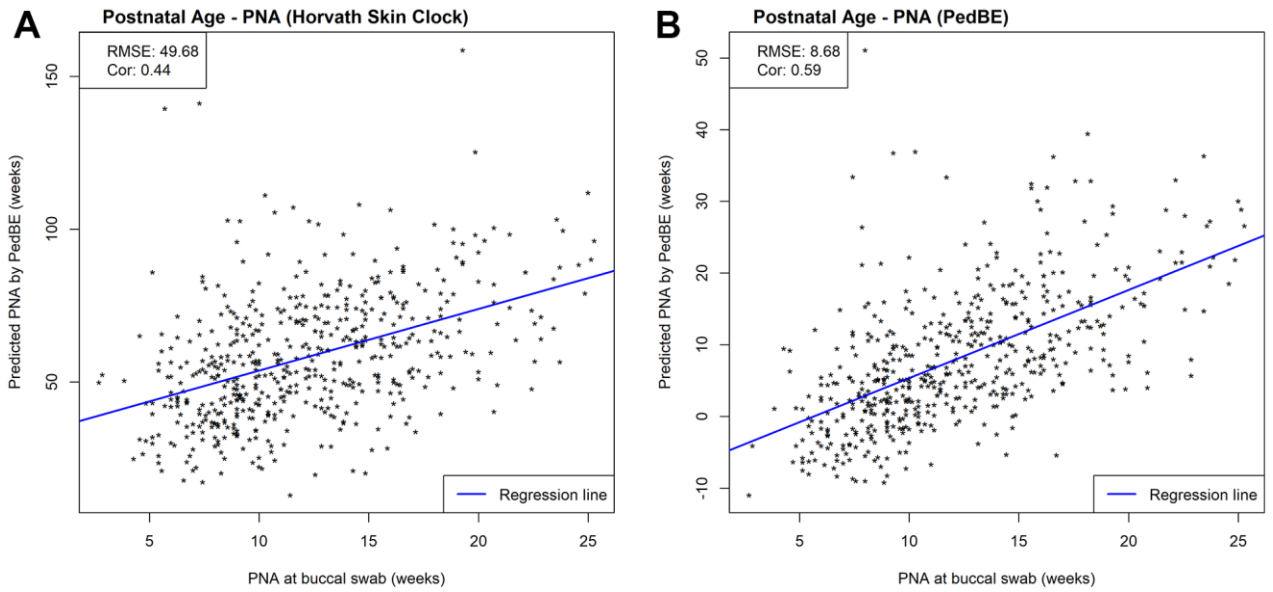

Figure 5. Scatterplots of PNA estimated by (**A**) Horvath's skin-blood clock and (**B**) PedBE and measured PNA within NOVI. Prediction performances are evaluated by RMSE and correlations between estimated and measured PNA.

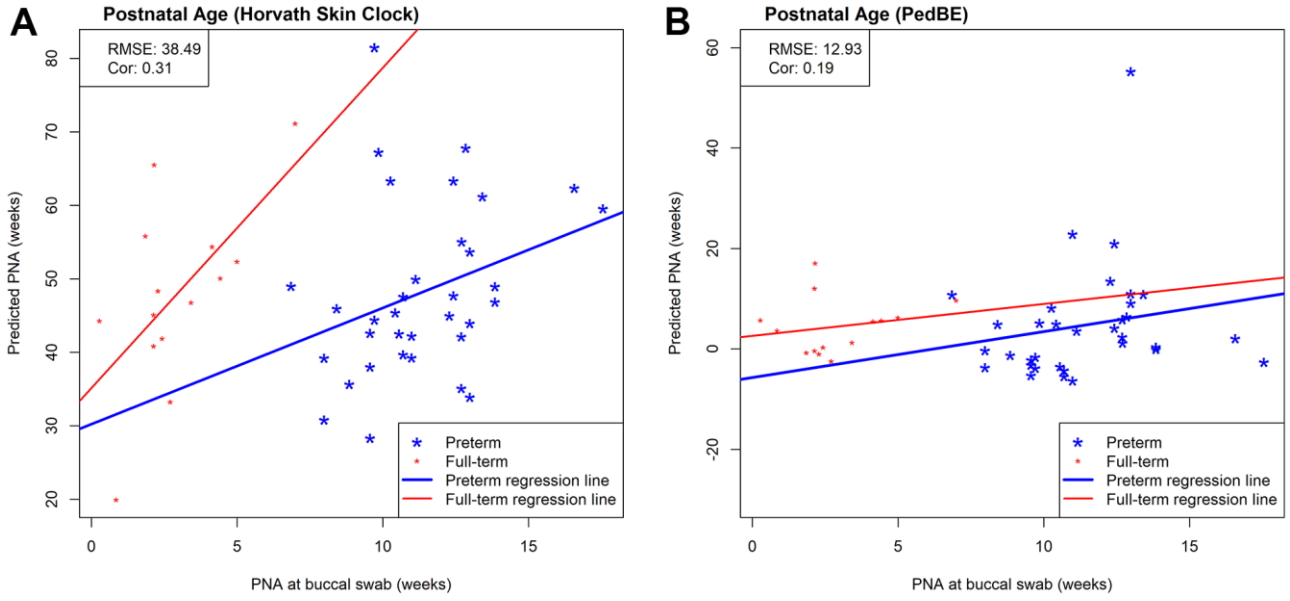

Figure 6. Scatterplots of measured PNA and PNA estimates by (**A**) Horvath's skin-blood clock and (**B**) PedBE in an external saliva data set (GSE72120 [18]). This saliva data set was measured by the 450k array and included full-term (red) and preterm (blue) infants. The reported prediction performances, RMSE and correlation coefficients, between estimated and measured age metrics are based on preterm infants only.

weeks, respectively. For full-term infants (highlighted in red), the Horvath skin-blood clock correlation is r = 0.60 and PedBE correlation is r = 0.20 with RMSE of 46.31 and 5.54 weeks, respectively. Interestingly, Horvath's clock yields a substantially better correlation between reported and predicted age for full-term infants compared to preterm infants, while the PedBE clock yielded weak correlations for both groups. Yet, while the correlations are stronger for Horvath's clock, the actual predicted ages were closer to the reported ages for the PedBE clocks. In contrast, PNA prediction of full-term infants using our NEOage 450k PNA clock has a stronger correlation (r = 0.76) than both existing clocks and a similar RMSE of 7.42 weeks compared to PedBE. The best prediction performance for the full-term infants resulted from our NEOage 450k PMA clock with a correlation of 0.90 and RMSE of 2.14 weeks.

Enrichment analysis

We performed enrichment analyses for the CpGs included in the four NEOage clocks that we characterized to evaluate potential pathway enrichments of genes associated with CpGs that we identified. No pathways or gene ontology (GO) terms were significantly enriched after False Discovery Rate (FDR) correction (FDR < 0.1), but the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways and GO terms that tended to have the smallest raw p-values included immune and inflammatory responses, endocrine activities, steroidogenesis, cellular proliferation, cellular differentiation and organization, and organ morphogenesis. Tables containing the 20 most significantly enriched pathways are provided in the Supplementary Material (Supplementary Tables 5–12).

DISCUSSION

While there has been some progress in addressing the lack of epigenetic clocks focusing on pediatric populations in recent years [2], to our knowledge, there currently exists no epigenetic clock that is specialized for preterm infants, nor for age prediction specific to the neonatal period. Preterm infants present a unique population due to the shift of their biological and chronological age progress relative to full-term infants. To fill this gap, we developed four NEOage clocks that are based on preterm infants from the NOVI study to estimate PMA and PNA (EPIC- and 450k-compatible) and include 303-522 CpGs. We demonstrate that our newly developed NEOage clocks outperform two established epigenetic clocks, Horvath's skin-blood clock and PedBE, both in our NOVI buccal data set and in an external saliva data set of infants that were born preterm.

A systematic deviation of full-term infants can be observed in Figures 4, 6. This shift appears to be more dominant in PNA predictions and might indicate that our PMA and PNA clocks capture a similar aging signature, but that our PNA clocks are more sensitive to the GA at birth. Pre- and full-term infants, as shown in Figure 4B, appear to have moderately similar regression slopes, but different intercepts, which is most likely a result of their different GA at birth. While extrapolation of our NEOage clocks outside of their training range is not recommended, it can be expected that prediction accuracy decreases with greater age differences (similar to extrapolating adult clocks to children, or pediatric clocks to the neonatal period). However, if extrapolation of age outside of our training age range but proximal to birth is necessary, our PMA clocks might be more appropriate.

We observed noticeable differences in RMSE when comparing reported ages to predicted ages from existing clocks [2, 6], predominantly in estimates from Horvath's skin-blood clock, but also PedBE. One possible explanation is that both clocks were not specifically developed for this age range. For these existing clocks, age is estimated in years, which was then transformed to weeks by multiplying by 52. Hence, any prediction errors might be amplified. In addition, PNA is greatly overestimated for all infants by Horvath's skin-blood clock, meaning that estimated PNA is greater than measured PNA for every infant.

While PMA seems to provide a more generalizable estimate of age, it comes with the limitation that the day of conception (reference point to calculate PMA) is not as precise of a measurement as day of birth (reference point to calculate PNA) and therefore is associated with a certain degree of uncertainty. Another limitation is the extension of these clocks to other tissue types, because our NEOage clocks are based on buccal cells collected via cheek swabs from preterm infants. Generalizing our NEOage clock to different tissue types will most likely compromise the prediction performance. Nevertheless, buccal swab is minimally invasive and thus is specifically important in pediatric and neonatal populations where more invasive sampling may deter study participation [19]. While blood samples provide large amounts of DNA with good quality, it requires an invasive and expensive procedure with technical difficulties, can be difficult or impossible to collect from preterm neonates, and causes discomfort and increased risk of infection [19]. In addition, buccal epithelial cells have been shown to be better proxy for the brain than peripheral blood [20]. The collection of buccal cells and saliva is less complicated, inexpensive

and non-invasive [19], with the added benefit of buccal cells being less heterogeneous [2, 20]. A possible contamination of prenatal fetal sample with maternal cells can be avoided by performing a short terminal repeats analysis [19].

With our newly developed NEOage clocks we aim to fill the gap of methylation clocks trained on pediatric samples [21] and based on buccal cells, an easily accessible tissue that requires no invasive procedures.

Our epigenetic estimators of neonatal aging in preterm infants might be particularly valuable in this population of neonates because it could allow us to gain insight into early life aging and reflect influences on subsequent health and development. Further, establishing precise estimators of PMA might help us to develop tools to more accurately determine the day of conception and measurements associated with it (e.g., PMA and GA).

CONCLUSIONS

We have introduced our four NEOage clocks that are specific to the assessment of epigenetic age in very preterm neonates. Our NEOage clocks are based on buccal cells, a tissue that is easily accessible and requires no invasive intervention. Postmenstrual age (PMA) and post-natal age (PNA) can be accurately estimated utilizing DNAm measured by either the Illumina 450k or EPIC array. We demonstrated that our NEOage clocks outperform two existing clocks by assessing their prediction performances in two preterm infant data sets. With our NEOage clocks, we have provided tools to examine neonatal aging, age acceleration and their association with neonatal health and development in a unique population of very preterm infants.

MATERIALS AND METHODS

Study participants

The Neonatal Neurobehavior and Outcomes in Very Preterm Infants (NOVI) Study was conducted at 9 university-affiliated NICUs in Providence, RI, Grand Rapids, MI, Kansas City, MO, Honolulu, HI, Winston-Salem, NC, and Torrance and Long Beach CA from April 2014 through June 2016. These NICUs were also Vermont Oxford Network (VON) participants. Eligibility was determined based on the following inclusion criteria: 1) birth at <30 weeks post menstrual age; 2) parental ability to read and speak English or Spanish and 3) residence within 3 hours of the NICU and follow-up clinic. Exclusion criteria included maternal age <18 years, maternal cognitive impairment, maternal death, infants with major congenital anomalies, including central nervous system, cardiovascular, gastrointestinal, genitourinary, chromosomal, and nonspecific anomalies, and NICU death. Parents of eligible infants were invited to participate in the study when survival to discharge was determined to be likely by the attending neonatologist. Overall, 704 eligible infants were enrolled. Researchers explained study procedures and obtained informed consent in accordance with each institution's review board. 542 children for which DNAm data was measured and passed QC were included in this analysis (characteristics presented in Table 1). The sample included 19% of infants with serious infection (sepsis or necrotizing enterocolitis), 51% with bronchopulmonary dysplasia, 13% with severe brain injury (parenchymal echodensity, periventricular leukomalacia, or ventricular dilatation), and 6% with severe retinopathy of prematurity. PMA in NOVI was calculated by adding PNA at buccal collection to the estimated GA at birth which was obtained via an established process [22, 23] and is described in detail by Everson et al. [15].

DNAm collection and pre-processing

Buccal cell tissue was collected from infants that were born very preterm (<30 weeks gestation), at NICU discharge (Figure 1), and DNAm levels were profiled using the EPIC array.

Genomic DNA was extracted from buccal swab samples, collected near term-equivalent age, using the Isohelix Buccal Swab system (Boca Scientific), quantified using the Quibit Fluorometer (Thermo Fisher, Waltham, MA, USA) and aliquoted into a standardized concentration for subsequent analyses. DNA samples were plated randomly across 96-well plates and provided to the Emory University Integrated Genomics Core for bisulfite modification using the EZ DNA Methylation Kit (Zymo Research, Irvine, CA, USA), and subsequent assessment of genome-wide DNAm using the Illumina MethylationEPIC Beadarray (Illumina, San Diego, CA, USA) following standardized methods based on the manufacturer's protocol. The pre-processing of the data followed a modified workflow described by Everson et al. [15]. Array data were normalized via Noob normalization [24, 25] and samples with more than 5% of probes yielding detection p-values > 1.0E-5 or mismatch between reported and predicted sex were excluded. In addition, one of two duplicated samples was omitted (we retained the duplicate sample with smallest detection p-values). Probes with median detection p-values < 0.05, probes measured on the X or Y chromosome, probes that had single nucleotide polymorphisms (SNP) within the binding region or that could cross-hybridize to other regions of the genome were excluded [26]. Then, array

data were standardized across Type-I and Type-II probe designs with beta-mixture quantile normalization [27, 28]. After exclusions, 706,323 probes were available from 542 samples for this study. These data are accessible through NCBI Gene Expression Omnibus (GEO) via accession series GSE128821.

Development of the epigenetic clocks

Since data from the EPIC and 450k arrays are widely used in ongoing research projects, we considered two sets of data for all analyses: (1) a complete data set (706,323 probes) with logit transformed beta-values (m-values) that is compatible with EPIC arrays (hereafter referred to as the EPIC data set) and (2) a subset of the logit-transformed data (364,410 probes) that is compatible with both EPIC and 450k arrays (hereafter referred to as the 450k data set). Penalized regression models ("glmnet" function in glmnet R package [29]) were fit to both data sets to identify sets of CpGs (NEOage clocks) predictive of PMA and PNA (4 total clocks: PMA-EPIC, PNA- EPIC, PMA-450k and PNA-450k). The alpha parameter of glmnet was set to 0.5 (elastic net regression) and lambda (PMA-EPIC: 0.049, PNA- EPIC: 0.0677, PMA-450k: 0.097 and PNA-450k: 0.2038) was chosen such that the mean cross-validated error is minimized with 10-fold cross validation ("lambda.min" result from "cv.glmnet" function in glmnet R package [29]).

We fit a series of penalized regression models to both data sets (EPIC and 450k) applying LOO cross-validation. This procedure allowed us to assess prediction performances but also limit overfitting and selection bias. In LOO cross-validation, a model is trained on all but one sample to make a prediction for that held-out sample. This step is repeated until each sample is held out and predicted once and results in N potentially unique sets of CpGs for a given outcome, where N is the sample size. Because our sample contained multiple births (e.g., twins), we additionally removed all siblings from the training set of all non-singleton children. The performance of predicted age outcomes was evaluated by examining their correlation with the measured outcome and RMSE.

In addition, prediction performances of models trained using the complete (not LOO approach) 450k data set (450k NEOage clocks) were evaluated in an independent publicly available data set (GSE72120 [18]) that contained DNAm from the 450k array for 34 preterm and 14 full-term infants with information on PMA and PNA. This data set was chosen because to our knowledge it is the closest comparable data, but it is important to point out the difference between both data sets, as one measured DNAm of buccal swabs via the EPIC array and the other profiled DNAm in saliva using the 450k array. We evaluated the performance of our PMA-450k and PNA-450k NEOage clocks in the test sample by examining the correlation between predicted and measured outcomes. We also report the RMSE.

Application of existing epigenetic clocks

To compare our newly-developed NEOage clocks to existing clocks, we applied Horvath's skin-blood clock [30] and the PedBE clock [2] to estimate PNA in our data and in the independent external data set. Both existing clocks were trained on pediatric epithelial samples, and thus could be applicable to our data. However, the skin-blood clock was also trained on blood samples and thus can estimate age from DNA derived from multiple tissue types, while PedBE is specific to buccal epithelium. Additionally, while the skin-blood clock is a life-course clock that was trained on samples from infants, children, and adults, the PedBE clock is a pediatric-specific clock. The coefficients and codes for estimating age via these existing clocks are available in the original publications via the Supplementary Materials [30] and the author's webpage [2]. Horvath's skin-blood clock includes 391 CpGs and was developed with DNA from human fibroblasts, keratinocytes, buccal cells, endothelial cells, blood, and saliva (age range: 0-92 years). Out of the 391 CpGs, 345 CpGs were available in the NOVI and saliva data set. For the NOVI data set, 42 out of the 46 missing CpGs were substituted with closest CpGs within 5,000bp. The remaining 4 missing CpGs were omitted; 3 CpGs did not have CpGs available in our data that were within 5,000bp and 1 CpG was located on chromosome X (excluded during data preprocessing). Analogously for the saliva data set, 40 of the 46 missing CpGs were substituted with closest CpGs within 5,000bp. The remaining 6 missing CpGs were omitted; 5 CpGs did not have CpGs available in the saliva data set that were within 5,000bp and 1 CpG was located on chromosome X. The PedBE clock (age range: 0-20 years), developed with pediatric buccal epithelial cells, consists of 94 CpGs. There were 5 CpGs not available in the NOVI and saliva data set, which were substituted by the closest CpGs within 5,000bp. No CpGs were omitted. Performance of predicted PNA was evaluated by their correlation with the measured PNA and RMSE.

Enrichment analysis

To gain insights into the biological functions of the genes associated with the identified CpGs included in the four NEOage clocks, we performed an enrichment analysis. We utilized the "gometh" function in missMethyl Bioconductor package [31], that performs a hypergeometric test, while taking the number of CpG

sites per gene into account. For the enrichment analysis involving the CpGs of our 450k NEOage clocks, we specified the array type to be "450k" and provided a list of CpGs that were considered (364,410 probes) for the "all.cpg" argument of "gometh". Analogously, we specified the array type to be "EPIC" for the enrichment analysis involving the CpGs of our EPIC NEOage clocks and provided a list of CpGs that were considered (706,323 probes). We evaluated both options for databases provided by "gometh": GO and KEGG.

Data availability statement

The DNA methylation data generated in the current study are available in the NCBI GEO via accession series GSE128821. R codes used for the analyses presented in the paper are available upon request to the corresponding author.

Abbreviations

450k: Infinium HumanMethylation450 BeadChip; AA: Age Acceleration; CpG: Cytosine-phosphate-guanine; EPIC: Infinium MethylationEPIC BeadChip; FDR: False Discovery Rate; GA: Gestational Age; GEO: Gene Expression Omnibus; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; LOO: Leave-One-Out; NEOage: Neonatal Epigenetic Estimator of age; NICU: Neonatal Intensive Care Unit; NOVI: Neonatal Neurobehavior and Outcomes in Very Preterm Infants; PMA: Post-Menstrual Age; PNA: Post-Natal Age; RMSE: Root Mean Squared Error; SNP: single nucleotide polymorphisms.

AUTHOR CONTRIBUTIONS

Dr. Graw designed the study, analyzed and interpreted data, drafted the article and revised critically for important intellectual content, and approved the final version as submitted. Dr. Camerota reviewed and revised critically for important intellectual content and approved the final version as submitted. Dr. Carter reviewed and revised critically for important intellectual content and approved the final version as submitted. Dr. Helderman reviewed and revised critically for important intellectual content and approved the final version as submitted. Dr. Hofheimer reviewed and revised critically for important intellectual content and approved the final version as submitted. Dr. McGowan reviewed and revised critically for important intellectual content and approved the final version as submitted. Dr. Neal reviewed and revised critically for important intellectual content and approved the final version as submitted. Dr. Pastyrnak reviewed and revised critically for important intellectual content and approved the final version as submitted. Dr. Smith reviewed and revised critically for important intellectual content and approved the final version as submitted. Dr. DellaGrotta reviewed and revised critically for important intellectual content and approved the final version as submitted. Dr. Dansereau reviewed and revised critically for important intellectual content and approved the final version as submitted. Dr. Padbury reviewed and revised critically for important intellectual content and approved the final version as submitted. Dr. O'Shea reviewed and revised critically for important intellectual content and approved the final version as submitted. Dr. Lester reviewed and revised critically for important intellectual content and approved the final version as submitted. Dr. Marsit reviewed and revised critically for important intellectual content and approved the final version as submitted. Dr. Everson conceptualized and designed the study, interpreted data, drafted the article and revised critically for important intellectual content, and approved the final version as submitted.

ACKNOWLEDGMENTS

We would like to thank the Emory and WIH lab teams, NOVI Study Coordinators, and the NOVI families who made this work possible.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

FUNDING

This work was supported by NIH Grants NICHD R01HD072267 (Lester and O'Shea), R01HD084515 (Lester and Everson), UH3OD023347 (Lester, Marsit, and O'Shea), and the HERCULES Center (P30 ES019776).

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. Johnson AA, Akman K, Calimport SR, Wuttke D, Stolzing A, de Magalhães JP. The role of DNA methylation in aging, rejuvenation, and age-related disease. Rejuvenation Res. 2012; 15:483–94. https://doi.org/10.1089/rej.2012.1324 PMID:23098078

2. McEwen LM, O'Donnell KJ, McGill MG, Edgar RD, Jones MJ, MacIsaac JL, Lin DT, Ramadori K, Morin A, Gladish N, Garg E, Unternaehrer E, Pokhvisneva I, et al. The PedBE clock accurately estimates DNA methylation age

in pediatric buccal cells. Proc Natl Acad Sci USA. 2020; 117:23329–35.
https://doi.org/10.1073/pnas.1820843116
PMID:31611402

3. Bell CG, Lowe R, Adams PD, Baccarelli AA, Beck S, Bell JT, Christensen BC, Gladyshev VN, Heijmans BT, Horvath S, Ideker T, Issa JJ, Kelsey KT, et al. DNA methylation aging clocks: challenges and recommendations. Genome Biol. 2019; 20:249.
https://doi.org/10.1186/s13059-019-1824-y
PMID:31767039

4. Fransquet PD, Wrigglesworth J, Woods RL, Ernst ME, Ryan J. The epigenetic clock as a predictor of disease and mortality risk: a systematic review and meta-analysis. Clin Epigenetics. 2019; 11:62.
https://doi.org/10.1186/s13148-019-0656-7
PMID:30975202

5. Bergsma T, Rogaeva E. DNA Methylation Clocks and Their Predictive Capacity for Aging Phenotypes and Healthspan. Neurosci Insights. 2020; 15:2633105520942221.
https://doi.org/10.1177/2633105520942221
PMID:32743556

6. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013; 14:R115.
https://doi.org/10.1186/gb-2013-14-10-r115
PMID:24138928

7. Simpkin AJ, Hemani G, Suderman M, Gaunt TR, Lyttleton O, Mcardle WL, Ring SM, Sharp GC, Tilling K, Horvath S, Kunze S, Peters A, Waldenberger M, et al. Prenatal and early life influences on epigenetic age in children: a study of mother-offspring pairs from two cohort studies. Hum Mol Genet. 2016; 25:191–201.
https://doi.org/10.1093/hmg/ddv456
PMID:26546615

8. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan JB, Gao Y, Deconde R, Chen M, Rajapakse I, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell. 2013; 49:359–67.
https://doi.org/10.1016/j.molcel.2012.10.016
PMID:23177740

9. Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, Hou L, Baccarelli AA, Stewart JD, Li Y, Whitsel EA, Wilson JG, Reiner AP, et al. An epigenetic biomarker of aging for lifespan and healthspan. Aging (Albany NY). 2018; 10:573–91.
https://doi.org/10.18632/aging.101414
PMID:29676998

10. Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, Hou L, Baccarelli AA, Li Y, Stewart JD, Whitsel EA, Assimes TL, Ferrucci L, Horvath S. DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging (Albany NY). 2019; 11:303–27.
https://doi.org/10.18632/aging.101684
PMID:30669119

11. Goldman J, Becker ML, Jones B, Clements M, Leeder JS. Development of biomarkers to optimize pediatric patient management: what makes children different? Biomark Med. 2011; 5:781–94.
https://doi.org/10.2217/bmm.11.96
PMID:22103612

12. Knight AK, Craig JM, Theda C, Bækvad-Hansen M, Bybjerg-Grauholm J, Hansen CS, Hollegaard MV, Hougaard DM, Mortensen PB, Weinsheimer SM, Werge TM, Brennan PA, Cubells JF, et al. An epigenetic clock for gestational age at birth based on blood methylation data. Genome Biol. 2016; 17:206.
https://doi.org/10.1186/s13059-016-1068-z
PMID:27717399

13. Bohlin J, Håberg SE, Magnus P, Reese SE, Gjessing HK, Magnus MC, Parr CL, Page CM, London SJ, Nystad W. Prediction of gestational age based on genome-wide differentially methylated regions. Genome Biol. 2016; 17:207.
https://doi.org/10.1186/s13059-016-1063-4
PMID:27717397

14. WHO. Preterm birth. 2018.
https://www.who.int/news-room/fact-sheets/detail/preterm-birth

15. Everson TM, O'Shea TM, Burt A, Hermetz K, Carter BS, Helderman J, Hofheimer JA, McGowan EC, Neal CR, Pastyrnak SL, Smith LM, Soliman A, DellaGrotta SA, et al. Serious neonatal morbidities are associated with differences in DNA methylation among very preterm infants. Clin Epigenetics. 2020; 12:151.
https://doi.org/10.1186/s13148-020-00942-1
PMID:33076993

16. Institute of Medicine. Preterm Birth: Causes, Consequences, and Prevention. Washington (DC): National Academies Press 2007.
https://doi.org/10.17226/11622
PMID:20669423

17. Wilhelm-Benartzi CS, Koestler DC, Karagas MR, Flanagan JM, Christensen BC, Kelsey KT, Marsit CJ, Houseman EA, Brown R. Review of processing and analysis methods for DNA methylation array data. Br J Cancer. 2013; 109:1394–402.
https://doi.org/10.1038/bjc.2013.496
PMID:23982603

18. Sparrow S, Manning JR, Cartier J, Anblagan D, Bastin ME, Piyasena C, Pataky R, Moore EJ, Semple SI, Wilkinson AG, Evans M, Drake AJ, Boardman JP.

Epigenomic profiling of preterm infants reveals DNA methylation differences at sites associated with neural function. Transl Psychiatry. 2016; 6:e716. https://doi.org/10.1038/tp.2015.210 PMID:26784970

19. Said M, Cappiello C, Devaney JM, Podini D, Beres AL, Vukmanovic S, Rais-Bahrami K, Luban NC, Sandler AD, Tatari-Calderone Z. Genomics in premature infants: a non-invasive strategy to obtain high-quality DNA. Sci Rep. 2014; 4:4286. https://doi.org/10.1038/srep04286 PMID:24598548

20. Theda C, Hwang SH, Czajko A, Loke YJ, Leong P, Craig JM. Quantitation of the cellular content of saliva and buccal swab samples. Sci Rep. 2018; 8:6944. https://doi.org/10.1038/s41598-018-25311-0 PMID:29720614

21. Kling T, Wenger A, Carén H. DNA methylation-based age estimation in pediatric healthy tissues and brain tumors. Aging (Albany NY). 2020; 12:21037–56. https://doi.org/10.18632/aging.202145 PMID:33168783

22. O'Shea TM, Allred EN, Dammann O, Hirtz D, Kuban KC, Paneth N, Leviton A, and ELGAN study Investigators. The ELGAN study of the brain and related disorders in extremely low gestational age newborns. Early Hum Dev. 2009; 85:719–25. https://doi.org/10.1016/j.earlhumdev.2009.08.060 PMID:19765918

23. McElrath TF, Hecht JL, Dammann O, Boggess K, Onderdonk A, Markenson G, Harper M, Delpapa E, Allred EN, Leviton A, and ELGAN Study Investigators. Pregnancy disorders that lead to delivery before the 28th week of gestation: an epidemiologic approach to classification. Am J Epidemiol. 2008; 168:980–89. https://doi.org/10.1093/aje/kwn202 PMID:18756014

24. Liu J, Siegmund KD. An evaluation of processing methods for HumanMethylation450 BeadChip data. BMC Genomics. 2016; 17:469. https://doi.org/10.1186/s12864-016-2819-7 PMID:27334613

25. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014; 30:1363–69. https://doi.org/10.1093/bioinformatics/btu049 PMID:24478339

26. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, Van Djik S, Muhlhausler B, Stirzaker C, Clark SJ. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol. 2016; 17:208. https://doi.org/10.1186/s13059-016-1066-1 PMID:27717381

27. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013; 29:189–96. https://doi.org/10.1093/bioinformatics/bts680 PMID:23175756

28. Pidsley R, Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics. 2013; 14:293. https://doi.org/10.1186/1471-2164-14-293 PMID:23631413

29. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010; 33:1–22. https://doi.org/10.18637/jss.v033.i01 PMID:20808728

30. Horvath S, Oshima J, Martin GM, Lu AT, Quach A, Cohen H, Felton S, Matsuyama M, Lowe D, Kabacik S, Wilson JG, Reiner AP, Maierhofer A, et al. Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and *ex vivo* studies. Aging (Albany NY). 2018; 10:1758–75. https://doi.org/10.18632/aging.101508 PMID:30048243

31. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics. 2016; 32:286–88. https://doi.org/10.1093/bioinformatics/btv560 PMID:26424855

SUPPLEMENTARY MATERIALS

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1–4.

Supplementary Table 1. CpGs and corresponding coefficients of the 450k-PMA NEOage clock.

Supplementary Table 2. CpGs and corresponding coefficients of the 450k-PNA NEOage clock.

Supplementary Table 3. CpGs and corresponding coefficients of the EPIC-PMA NEOage clock.

Supplementary Table 4. CpGs and corresponding coefficients of the EPIC-PNA NEOage clock.

Supplementary Table 5. Top 20 pathways from GO pathway analysis for CpGs included in the 450k-PMA NEOage clock.

	ONTOLOGY	TERM	N	DE	P.DE	FDR
GO:0051240	BP	positive regulation of multicellular organismal process	1717	55.5	0.000101530604212178	1
GO:0007343	BP	egg activation	5	2	0.000488637355841379	1
GO:0071230	BP	cellular response to amino acid stimulus	63	7	0.000539308264807122	1
GO:0120162	BP	positive regulation of cold-induced thermogenesis	96	8	0.000645109263846585	1
GO:0005086	MF	ARF guanyl-nucleotide exchange factor activity	18	4	0.000719253481406075	1
GO:0032011	BP	ARF protein signal transduction	18	4	0.00124599435874616	1
GO:0032012	BP	regulation of ARF protein signal transduction	18	4	0.00124599435874616	1
GO:0043200	BP	response to amino acid	106	8	0.00126159591846291	1
GO:0031669	BP	cellular response to nutrient levels	223	11	0.00174701422017418	1
GO:0098772	MF	molecular function regulator	1668	45	0.00185581614546516	1
GO:0051239	BP	regulation of multicellular organismal process	3091	80	0.00185939492722176	1
GO:0002687	BP	positive regulation of leukocyte migration	120	7	0.00199912401286309	1
GO:0031668	BP	cellular response to extracellular stimulus	254	12	0.00200618606306946	1
GO:0071496	BP	cellular response to external stimulus	322	14	0.00209524302848006	1
GO:0051954	BP	positive regulation of amine transport	34	4	0.00234424561108088	1
GO:0045623	BP	negative regulation of T-helper cell differentiation	15	3	0.00247101470290018	1
GO:0106106	BP	cold-induced thermogenesis	140	9	0.00249643374976729	1
GO:0120161	BP	regulation of cold-induced thermogenesis	140	9	0.00249643374976729	1
GO:0042636	BP	negative regulation of hair cycle	5	2	0.00301157639678554	1
GO:0009966	BP	regulation of signal transduction	2976	75.33333333	0.00329481415897157	1

Supplementary Table 6. Top 20 pathways from KEGG pathway analysis for CpGs included in the 450k-PMA NEOage clock.

	Description	N	DE	P.DE	FDR
path:hsa04974	Protein digestion and absorption	92	7	0.0025432527716253	0.826463250084872
path:hsa04144	Endocytosis	242	12.5	0.00483311842154896	0.826463250084872
path:hsa04512	ECM-receptor interaction	86	6	0.0131031855183759	1
path:hsa04151	PI3K-Akt signaling pathway	331	13.5	0.0218884783349282	1
path:hsa00450	Selenocompound metabolism	17	2	0.0265841978233973	1
path:hsa04972	Pancreatic secretion	93	5	0.0269679372038782	1
path:hsa04971	Gastric acid secretion	75	5	0.0273248262064296	1
path:hsa04150	mTOR signaling pathway	151	7	0.0374334371410264	1
path:hsa04640	Hematopoietic cell lineage	89	4	0.0462122030513505	1
path:hsa04727	GABAergic synapse	84	5	0.0469447204065581	1
path:hsa03410	Base excision repair	32	2.5	0.0642826135929407	1
path:hsa04064	NF-kappa B signaling pathway	95	4	0.0737709112366848	1
path:hsa00500	Starch and sucrose metabolism	30	2	0.0746391342946885	1
path:hsa04510	Focal adhesion	193	8	0.0828799780172142	1
path:hsa04261	Adrenergic signaling in cardiomyocytes	144	6	0.0930959869894971	1
path:hsa05412	Arrhythmogenic right ventricular cardiomyopathy	74	4	0.106935318655637	1
path:hsa05200	Pathways in cancer	506	15	0.109313257881781	1
path:hsa00920	Sulfur metabolism	10	1	0.111928784033015	1
path:hsa04657	IL-17 signaling pathway	88	3	0.120766341082755	1
path:hsa03022	Basal transcription factors	41	2	0.122128583254498	1

Supplementary Table 7. Top 20 pathways from GO pathway analysis for CpGs included in the 450k-PNA NEOage clock.

	ONTOLOGY	TERM	N	DE	P.DE	FDR
GO:0042127	BP	regulation of cell population proliferation	1591	40	0.0000534190886299528	1
GO:0008283	BP	cell population proliferation	1897	44	0.000139796711649994	1
GO:0008285	BP	negative regulation of cell population proliferation	715	23	0.000144783294410145	1
GO:0090191	BP	negative regulation of branching involved in ureteric bud morphogenesis	2	2	0.000224072825185314	1
GO:2001252	BP	positive regulation of chromosome organization	164	9.5	0.000246284844795411	1
GO:0031616	CC	spindle pole centrosome	14	3	0.000330509425042432	1
GO:0033044	BP	regulation of chromosome organization	323	12.5	0.000664900082049856	1
GO:0045843	BP	negative regulation of striated muscle tissue development	55	5	0.000970314025106637	1
GO:1901187	BP	regulation of ephrin receptor signaling pathway	2	2	0.000971984410108729	1
GO:0048635	BP	negative regulation of muscle organ development	56	5	0.00108076577962778	1
GO:0097028	BP	dendritic cell differentiation	38	4	0.00125990537999378	1
GO:0060021	BP	roof of mouth development	88	7	0.00141641819931624	1
GO:1901862	BP	negative regulation of muscle tissue development	58	5	0.00153736679951646	1
GO:0030513	BP	positive regulation of BMP signaling pathway	32	4	0.00161955302866369	1
GO:0016202	BP	regulation of striated muscle tissue development	143	8	0.00163674469616389	1
GO:0048385	BP	regulation of retinoic acid receptor signaling pathway	16	3	0.00165543877096897	1
GO:0045082	BP	positive regulation of interleukin-10 biosynthetic process	4	2	0.00168336883391311	1
GO:0070534	BP	protein K63-linked ubiquitination	49	4	0.00175553877998935	1
GO:0048634	BP	regulation of muscle organ development	147	8	0.0019218112555276	1
GO:1901861	BP	regulation of muscle tissue development	146	8	0.00194065212296168	1

Supplementary Table 8. Top 20 pathways from KEGG pathway analysis for CpGs included in the 450k-PNA NEOage clock.

	Description	N	DE	P.DE	FDR
path:hsa04913	Ovarian steroidogenesis	50	3	0.0305045691205889	1
path:hsa04911	Insulin secretion	83	4	0.0580354487974986	1
path:hsa00920	Sulfur metabolism	10	1	0.0635494328977979	1
path:hsa00330	Arginine and proline metabolism	46	2	0.0660623210263602	1
path:hsa04664	Fc epsilon RI signaling pathway	66	3	0.0689829305658202	1
path:hsa00250	Alanine, aspartate and glutamate metabolism	36	2	0.0691860293452126	1
path:hsa04120	Ubiquitin mediated proteolysis	132	4	0.0765639489819126	1
path:hsa04261	Adrenergic signaling in cardiomyocytes	144	5	0.0797431390863962	1
path:hsa04020	Calcium signaling pathway	228	7	0.0802610307505027	1
path:hsa01100	Metabolic pathways	1400	21.5	0.0827951900967627	1
path:hsa00520	Amino sugar and nucleotide sugar metabolism	47	2	0.0837183667094073	1
path:hsa05221	Acute myeloid leukemia	64	3	0.0853382746769124	1
path:hsa04925	Aldosterone synthesis and secretion	95	4	0.0941511750167515	1
path:hsa04962	Vasopressin-regulated water reabsorption	43	2	0.0958227630574678	1
path:hsa00590	Arachidonic acid metabolism	61	2	0.0990080964986279	1
path:hsa04927	Cortisol synthesis and secretion	63	3	0.103475903215217	1
path:hsa05110	Vibrio cholerae infection	49	2	0.139903479319178	1
path:hsa04610	Complement and coagulation cascades	81	2	0.144076070674707	1
path:hsa00910	Nitrogen metabolism	15	1	0.144600617616157	1
path:hsa04725	Cholinergic synapse	111	4	0.144777378086761	1

Supplementary Table 9. Top 20 pathways from GO pathway analysis for CpGs included in the EPIC-PMA NEOage clock.

	ONTOLOGY	TERM	N	DE	P.DE	FDR
GO:0060090	MF	molecular adaptor activity	244	19.5	0.0000261984905881029	0.299312820997077
GO:0030674	MF	protein-macromolecule adaptor activity	204	17.5	0.0000264948943079647	0.299312820997077
GO:0009653	BP	anatomical structure morphogenesis	2628	93.3333333333333	0.00036075400639525	1
GO:0061061	BP	muscle structure development	649	30	0.000817243519837221	1
GO:0071149	CC	TEAD-2-YAP complex	2	2	0.000829613936436568	1
GO:0060187	CC	cell pole	2	2	0.00124016610690871	1
GO:0043005	CC	neuron projection	1245	51	0.00135112812526085	1
GO:0001725	CC	stress fiber	66	8	0.00149094428870597	1
GO:0097517	CC	contractile actin filament bundle	66	8	0.00149094428870597	1
GO:2000096	BP	positive regulation of Wnt signaling pathway, planar cell polarity pathway	8	3	0.00164092892480337	1
GO:0034330	BP	cell junction organization	629	32	0.00190548862751066	1
GO:0019215	MF	intermediate filament binding	14	3	0.00192713729059663	1
GO:0032432	CC	actin filament bundle	72	8	0.00213251244179354	1
GO:0032289	BP	central nervous system myelin formation	3	2	0.00247281453328891	1
GO:0090258	BP	negative regulation of mitochondrial fission	3	2	0.0026011582719471	1
GO:0003012	BP	muscle system process	457	21.8333333333333	0.00276093280594124	1
GO:0044297	CC	cell body	537	25.5	0.0030219067814788	1
GO:1904636	BP	response to ionomycin	4	2	0.00306956109657764	1
GO:1904637	BP	cellular response to ionomycin	4	2	0.00306956109657764	1
GO:0071936	MF	coreceptor activity involved in Wnt signaling pathway	8	3	0.00306994667197237	1

Supplementary Table 10. Top 20 pathways from KEGG pathway analysis for CpGs included in the EPIC-PMA NEOage clock.

	Description	N	DE	P.DE	FDR
path:hsa04310	Wnt signaling pathway	160	11	0.00563804820140634	1
path:hsa03022	Basal transcription factors	41	3	0.0382585766688834	1
path:hsa04150	mTOR signaling pathway	151	8	0.0409548512453372	1
path:hsa04514	Cell adhesion molecules	133	7	0.055839619373039	1
path:hsa04916	Melanogenesis	101	6	0.0575374173755753	1
path:hsa04750	Inflammatory mediator regulation of TRP channels	97	6.5	0.0678487404242832	1
path:hsa04971	Gastric acid secretion	75	5	0.0701572968381165	1
path:hsa00780	Biotin metabolism	3	1	0.0745147803124464	1
path:hsa04744	Phototransduction	27	2	0.080805733488801	1
path:hsa00630	Glyoxylate and dicarboxylate metabolism	30	2	0.0970358200187176	1
path:hsa04625	C-type lectin receptor signaling pathway	103	5	0.0984908711417847	1
path:hsa05146	Amoebiasis	98	5	0.0985952263276953	1
path:hsa05205	Proteoglycans in cancer	199	9	0.0989759531598264	1
path:hsa00515	Mannose type O-glycan biosynthesis	23	2	0.0996538362015267	1
path:hsa03410	Base excision repair	32	2.5	0.101507997479545	1
path:hsa04144	Endocytosis	244	10	0.102761294195724	1
path:hsa05031	Amphetamine addiction	66	4	0.107648802928742	1
path:hsa04152	AMPK signaling pathway	117	6	0.109755427511973	1
path:hsa04080	Neuroactive ligand-receptor interaction	320	9.16666666666667	0.125111384608538	1
path:hsa04070	Phosphatidylinositol signaling system	92	5	0.12605022083487	1

Supplementary Table 11. Top 20 pathways from GO pathway analysis for CpGs included in the EPIC-PNA NEOage clock.

ONTOLOGY		TERM	N	DE	P.DE	FDR
GO:0110111	BP	negative regulation of animal organ morphogenesis	33	7	0.0000150081855686758	0.339094944738661
GO:1905331	BP	negative regulation of morphogenesis of an epithelium	16	5	0.0000864716932197235	0.951641205696613
GO:0060686	BP	negative regulation of prostatic bud formation	4	3	0.000172823666721358	0.951641205696613
GO:0030510	BP	regulation of BMP signaling pathway	88	9	0.000218810440561404	0.951641205696613
GO:0090191	BP	negative regulation of branching involved in ureteric bud morphogenesis	2	2	0.000250881472848229	0.951641205696613
GO:0048645	BP	animal organ formation	63	8.5	0.000320898779391021	0.951641205696613
GO:0060685	BP	regulation of prostatic bud formation	5	3	0.000364936681518837	0.951641205696613
GO:0060688	BP	regulation of morphogenesis of a branching structure	53	7	0.000371355157976279	0.951641205696613
GO:0090192	BP	regulation of glomerulus development	14	4	0.000382316795422211	0.951641205696613
GO:0072283	BP	metanephric renal vesicle morphogenesis	14	4	0.000504258477605173	0.951641205696613
GO:0030509	BP	BMP signaling pathway	146	11	0.000555551138375544	0.951641205696613
GO:0032675	BP	regulation of interleukin-6 production	143	9	0.000582041777636664	0.951641205696613
GO:0005402	MF	carbohydrate:cation symporter activity	19	4	0.000606490301815559	0.951641205696613
GO:0090185	BP	negative regulation of kidney development	16	4	0.000644477997689534	0.951641205696613
GO:0047045	MF	testosterone 17-beta-dehydrogenase (NADP+) activity	4	2	0.000704248067798324	0.951641205696613
GO:2000343	BP	positive regulation of chemokine (C-X-C motif) ligand 2 production	9	3	0.000725385020377788	0.951641205696613
GO:0032755	BP	positive regulation of interleukin-6 production	91	7	0.000751360618487751	0.951641205696613
GO:0030432	BP	peristalsis	10	3	0.000839410257077136	0.951641205696613
GO:0072006	BP	nephron development	137	11	0.000849343536609647	0.951641205696613
GO:0030513	BP	positive regulation of BMP signaling pathway	32	5	0.000906591713866725	0.951641205696613

Supplementary Table 12. Top 20 pathways from KEGG pathway analysis for CpGs included in the EPIC-PNA NEOage clock.

	Description	N	DE	P.DE	FDR
path:hsa05110	Vibrio cholerae infection	49	5	0.0044303057296943	1
path:hsa04961	Endocrine and other factor-regulated calcium reabsorption	51	5	0.0130902253578704	1
path:hsa04925	Aldosterone synthesis and secretion	95	7	0.0177978978310381	1
path:hsa04920	Adipocytokine signaling pathway	66	5	0.022338795028672	1
path:hsa04650	Natural killer cell mediated cytotoxicity	118	6.5	0.0315357173452436	1
path:hsa04217	Necroptosis	147	6	0.0396709182209861	1
path:hsa04261	Adrenergic signaling in cardiomyocytes	144	8	0.0399713503271485	1
path:hsa00980	Metabolism of xenobiotics by cytochrome P450	74	3	0.0439232520019814	1
path:hsa04970	Salivary secretion	86	5	0.0473923210414602	1
path:hsa00511	Other glycan degradation	18	2	0.0566603395732447	1
path:hsa05217	Basal cell carcinoma	63	4	0.062286340491488	1
path:hsa05412	Arrhythmogenic right ventricular cardiomyopathy	74	5	0.0666165663705422	1
path:hsa04020	Calcium signaling pathway	228	10.5	0.0733984670436083	1
path:hsa04060	Cytokine-cytokine receptor interaction	278	7	0.0823487304128945	1
path:hsa04978	Mineral absorption	54	3	0.0861725768552945	1
path:hsa04727	GABAergic synapse	84	5	0.0868375944591549	1
path:hsa04061	Viral protein interaction with cytokine and cytokine receptor	97	3.5	0.0971045985176263	1
path:hsa04935	Growth hormone synthesis, secretion and action	116	6	0.103216705086696	1
path:hsa05414	Dilated cardiomyopathy	92	5	0.10503579562344	1
path:hsa04066	HIF-1 signaling pathway	105	5	0.111209287977195	1

Supplementary Table 13. Annotation of the 20 common CpGs of the NEOage clocks.

Name	chr	pos	UCSC_RefGene_Name
cg05394010	chr16	2546596	TBC1D24
cg04777726	chr19	49340489	PLEKHA4;PLEKHA4;HSD17B14
cg24541835	chr1	12651540	DHRS3
cg05265234	chr22	38884016	DDX17;DDX17;DDX17;DDX17;DDX17;DDX17
cg21219851	chr17	78898189	RPTOR;RPTOR
cg00049440	chr9	73026643	KLF9
cg01454951	chr3	71730677	EIF4E3;EIF4E3;EIF4E3;EIF4E3
cg21664351	chr11	19841423	NAV2;NAV2;NAV2
cg12266861	chr1	35449720	
cg07318287	chr1	154377429	IL6R;IL6R
cg04862002	chr17	9074365	NTN1
cg00465247	chr13	50703477	
cg01138164	chr7	96648447	
cg13942103	chr1	111177829	
cg21135560	chr8	144946659	EPPK1
cg01916724	chr3	51975220	PARP3;PARP3;RRP9
cg05624226	chr5	180325954	BTNL8;BTNL8;BTNL8;BTNL8;BTNL8
cg13624964	chr14	86088696	FLRT2
cg17995197	chr2	26408167	FAM59B
cg06002476	chr3	8617065	

Supplementary Code

Please browse Full Text version to see the data of Supplementary Code 1.

Supplementary Code 1. R code example to calculate DNAm age using NEOage clocks.