The Future of Families and Child Wellbeing Study Biomarker Appendage

9- and 15-year Follow-Up Waves

March 2023

Bendheim-Thoman Center for Research on Child Wellbeing
Wallace Hall
Princeton University
Princeton, NJ 08544
http://crcw.princeton.edu

Columbia Population Research Center
1255 Amsterdam Avenue, Room 715
Columbia University
New York, NY10027
http://cupop.columbia.edu

Department of Molecular Biology
Notterman Lab
Lewis Thomas Lab
Princeton University
Princeton, NJ 08544

Population, Neurodevelopment, and Genetics Program
Institute for Social Research
University of Michigan
Ann Arbor, MI 48106

Funding for the FFCWS Biomarker data was provided through multiple awards from the National Institutes of Health. In addition to other FFCWS funding acknowledgements, when using the methylation data products please acknowledge R01 HD076592, R01 MH103761, R01HL149869, and R01 MD011716. When using only the telomere length data products please acknowledge R01 HD076592.

Prepared by the staff at the Bendheim-Thoman Center for Research on Child Wellbeing (CRCW) and the Department of Molecular Biology, Princeton University and the Population, Neurodevelopment and Genetics Program, University of Michigan. For more information about the FFCWS, please visit our web site at https://ffcws.princeton.edu/ or email ffdata@princeton.edu.

Contents

1. DATA APPENDAGE OVERVIEW 3
2. FILE LAYOUT 3
3. VARIABLE NAMING CONVENTION 3
4. DATA COLLECTION AND PROCESSING PROCEDURES 4
   4.1 SAMPLE COLLECTION 4
      4.1.1 9-Year Follow-up 4
      4.1.2 15-Year Follow-up 4
   4.2 TRANSFER OF SAMPLES TO THE BIOREPOSITORY/ LABORATORY 5
   4.3 DNA PURIFICATION AND STORAGE 5
5. BIOMARKER MEASUREMENT 6
   5.1 Telomere Length (m5_tl, k6_tl, and k5_adjtl) 6
   5.2 Additional Telomere Length Variables (*_tlmeasured, *_tlpassedqc) 7
   5.3 DNA Methylation Variables 7
      5.3.1 DNA Methylation Data Acquisition and Processing 7
      5.3.2 Cell Proportion Estimates 7
   5.4 Epigenetic biomarkers 8
      5.4.1 Note about missing probes 8
      5.4.2 Individual DNA methylation biomarkers 9
      5.4.3 Additional methylation biomarker variables 11
6. MISSING FLAG 12
7. DATA DICTONARY 12
8. REFERENCES 14

1. DATA APPENDAGE OVERVIEW

The Future of Families & Child Wellbeing Study (FFCWS) Biomarker Data Appendage contains telomere length data and DNA methylation age (DNAmAge) data for focal children from the FFCWS and their biological mothers. To obtain and process genetic information, saliva samples were provided by focal children and their biological mothers during in-home visit assessments at the 9-year follow-up wave. Saliva samples were collected from the focal children again during the 15-year follow-up wave. Several studies have shown associations between adversity and both telomere length (TL) and DNAmAge (e.g., [1]). The goal of collecting biologic information was to allow researchers to use these biomarkers to test hypotheses about the relationships between exposure to adverse environments and child development and health. Two aspects of the FFCWS design make it especially suitable for studying such interactions. First, because non-marital births were oversampled, children in the study are disproportionately exposed to the kinds of family and community stresses that have been shown to be associated with these biomarkers. Second, because of the study's longitudinal design and its focus on family relationships and community contexts, the study provides substantial data on children's cumulative exposure to stressful family and community environments.

In addition to telomere length, this data appendage includes variables generated from DNA methylation data from Illumina EPIC and 450K methylation arrays. These include DNAmAges generated from 13 epigenetic clocks, cell proportions derived from the DNA methylation data and chronological age. Additional variables indicate if a DNA sample was available, if it passed quality control (QC), if TL was measured and if the TL passed quality control.

2. FILE LAYOUT

This file contains 4898 observations and 165 variables (including idnum) and is sorted by *idnum*.

3. VARIABLE NAMING CONVENTION

Variable names are 5-21 characters long. The first two characters of all variables except the idnum indicate the focal person (k for focal child, m for biological mother) and wave (5 for the 9-year follow-up and 6 for the 15-year follow-up). For methylation clock variables, the third character is m for methylation and the fourth character indicates the array platform (k for 450K and e for EPIC). The characters after the underscore ("_") are specific for the biomarker. The table below provides further examples and variable naming for TL and flag variables.

Variable name structure

Variable name			Description
Prefix	Wave	Leaf	
k	5	mk_*	Illumina Infinium Human Methylation450K (450K) array variables
k	5	me_*	Illumina Infinium MethylationEPIC (EPIC) array variables

m/k	5/6	_dnaavailable	Whether a participant had the DNA data
m/k	5/6	_tl*/_adjtl	Telomere length variables

4. DATA COLLECTION AND PROCESSING PROCEDURES

4.1 SAMPLE COLLECTION

4.1.1 9-Year Follow-up

As part of the Year 9 follow-up wave, we attempted to collect saliva samples for genetic analysis from all focal children and biological mothers completing the in-home visit activities. In cases where a biological father or non-parental figure was the primary caregiver, or the biological mother was not present for the in-home visit, a saliva sample was collected from only the child. Families completing the home visit activities received a $65 payment to the primary caregiver and a $30 payment to the child. No additional remuneration was provided specifically for the contribution of saliva samples. Ultimately, 3,403 in-home visits were conducted; 2,884 unique child samples and 2,670 unique mother samples were collected.

Our survey subcontractor, Westat Inc., arranged sample collection. Westat interviewers used the Oragene® DNA Self-Collection Kit to collect saliva samples from focal children and biological mothers during the Home Visit. The Self-Collection Kit is a repository for the collection, preservation, and transportation of saliva. The respondents were instructed to spit into the container until the liquid portion reached a line on the interior of the container (the ideal volume of saliva to be collected was 2 ml). The container was then capped. In the process of screwing the cap onto the container, a liquid preservative was released. The container was then put into a small plastic biohazard bag that contained absorbent material if the container were to leak. The plastic bag was then put into a mailer.

In cases where the child had developmental or physical limitations prohibiting the interviewer from collecting the full sample by having the child spit into the collection kit, the child accessory kit was used. The child accessory kit contained a set of five saliva sponges used with the Oragene® self-collection kits. The saliva sponges were inserted into the child's mouth and moved around the upper and lower cheek pouches on both sides of the mouth to collect saliva. The sponges were stored inside the containers and then sealed as described above. Respondents were instructed to rinse their mouth out 5 minutes prior to the saliva sample collection. They were also provided with a packet of sugar and instructed to use ¼ tsp. if they were having difficulty stimulating saliva. After completing a Home Visit, interviewers mailed the specimen containers (placed in the bubble wrap mailers) back to Westat.

4.1.2 15-Year Follow-up

During the Year 15 follow-up wave, saliva was collected from the focal children (now teenagers) using Oragene DNA Self-Collection Kits (OGR-500) as described for the year 9 follow-up with the following modifications. For those who did not complete a home visit, saliva collection kits were sent to participants via mail and after collection participants returned the kits to Westat via FedEx. Participants were discouraged from eating or drinking anything within 30 minutes prior to sample collection. Upon completion of the saliva collection, all participants received $20.

4.2 TRANSFER OF SAMPLES TO THE BIOREPOSITORY/ LABORATORY

As Westat received specimen containers from the field, they were inspected to make sure that samples did not contain personal identifiers and placed in a shipment box with other received collection kits. Until they were mailed, these boxes were secured in the locked field room and maintained at room temperature. On an approximately monthly basis during the field period, Westat shipped boxes of specimen containers at room temperature to the laboratory of Dr. Daniel Notterman, Co-Principal Investigator of the FFCWS, in the Department of Molecular Biology at Princeton University. A transmittal form containing the IDs of the enclosed containers was emailed to lab staff. The lab confirmed receipt of the boxes with Westat.

Saliva collection kits were shipped monthly from October 2007 through May 2010 and from April 2014 through March 2017, respectively, for the 9- and 15- year follow-ups to the Notterman laboratory at Princeton by FedEx from Westat. Upon receipt of the shipments, lab technicians used a barcode reader to inventory the individual samples. These data were imported into a Microsoft Access database where a full inventory of receipted samples is kept.

4.3 DNA PURIFICATION AND STORAGE

Extraction was completed 1 to 2 weeks after receipt of samples from Westat. DNA was extracted from the entire sample using the Oragene® prepIT•L2P Laboratory Protocol for Manual Purification of DNA (DNA Genotek). Briefly, when samples were ready to be processed, they were incubated at 50°C in a water incubator for a minimum of 1 hour. The mixed Oragene®-DNA/saliva sample was transferred to a 15 ml centrifuge tube. A 1/25 ul volume portion of Oragene®- prepIT•L2P solution was added to the microcentrifuge tube and mixed by vortexing for a few seconds. The sample was incubated on ice for 10 minutes, then centrifuged at room temperature for 10 minutes at 3,500 rpm. The clear supernatant was carefully transferred with a pipet into a fresh centrifuge tube, avoiding the precipitate at the bottom of the tube. A volume of room temperature 100% ethanol equal to the volume of the supernatant was added to the supernatant and gently mixed by inversion 10 times. The sample was allowed to stand for 10 minutes at room air to allow the DNA to fully precipitate. The tube was then centrifuged for 10 minutes at room temperature at 3,500 rpm. The supernatant was decanted and discarded, taking care to avoid disturbing the DNA pellet. An ethanol wash consisting of 1 ml of 70% ethanol was added to the tube without disturbing the pellet. After standing at room temperature for 1 minute, the tube was gently swirled to completely remove the ethanol, taking care not to disturb the pellet. The pellet was air dried after which the DNA was rehydrated by adding 0.5 to 1.0 ml of TE solution (10 mM Tris-HCl, 1 mM EDTA, pH 7.5), vortexing the sample for 30 seconds, incubating it at room temperature, and transferring the rehydrated DNA to 3 x 1.7 ml microcentrifuge tubes for storage (2 tubes were stored at -80°C and one in the lab refrigerator at 4°C). DNA concentration was determined using the Quant-iT PicoGreen dsDNA Assay Kit (ThermoFisher).

5. BIOMARKER MEASUREMENT

5.1 Telomere Length (m5_tl, k6_tl, and k5_adjtl)

Telomere length (TL) was determined using a quantitative real-time polymerase chain reaction (qPCR) assay that incorporates a double-stranded oligomer standard to permit the measurement of absolute TL (in kilobases (kb) per telomere as previously described [1–3]). More specifically, an 84-mer double stranded oligonucleotide containing the sequence TTAGGG was used to create a standard curve for telomere quantity and a 79-mer double stranded oligonucleotide containing sequence from the *36B4* gene was used to create a standard curve for the reference gene. The PCR efficiency of each plate was between 90-110% and the R^2 of the standard curve was greater than 0.997. TL was calculated by dividing the telomere quantity by the reference gene quantity. This was then divided by 92 to determine TL/telomere. For each primer pair (telomere or 36B4), samples were measured in triplicate, and the results were averaged. Distribution of samples in the 96-well plates was randomized, and each plate contained repeats from previous runs to detect and limit potential batch effects. To mitigate batch effects, reference DNA from a cell line with a relatively short telomere (3C167b) and a fibroblast cell line after stable integration of the hTERT gene (cell line NHFpreT) were included in each run (both cell lines were a gift from Dr Yuanjun Zhao of Pennsylvania State University; [4,5]). In our laboratory, 3C167b has a mean TL of 3.1 kb, whereas NHFpreT has a mean TL of 16.8 kb. Reference DNA was harvested at a single time, aliquoted, and frozen. The reference cell line telomere and 36B4 quantities were used to normalize variation between runs. The geometric mean of the two cell line telomere quantities from each run was divided by the geometric mean of the two cell line telomere quantities from all the runs to create a normalization factor for each run. Each sample telomere quantity was divided by its run's normalization factor. This procedure was repeated for the 36B4 quantities. The normalized Tel quantities were divided by the normalized 36B4 quantities to generate telomere length. This was then divided by 92.

A replicate sample (DNA from volunteers) was included in triplicate in all plates, and the results of this measurement were used to compute an inter-run coefficient of variation, which was <11% across all runs. We estimated inter-assay intraclass correlation coefficients (ICCs) to estimate the consistency and correlation of TL following the suggestion of recent work on TL reliability [6–8]. We estimated a two-way, single measurement, absolute agreement, random effects model, known as ICC(A,1) and for average measurements ICC(A,k) in McGraw and Wong's terminology [9]. In total, there were 242 batches run for Year 9 and 126 batches run for Year 15. We calculated the ICC at both Year 9 (ICC (A,1) = 0.91 (95 percent CI[1] 0.82–1.0); ICC(A,k) = 0.99 (95 percent CI 0.99–1.0)) and Year 15 (ICC(A,1) = 0.95 (95 percent CI 0.84–1.0); ICC(A,k) = 0.99 (95 percent CI 0.99–1.0)).

The Year 9 and Year 15 samples were collected approximately 6 years apart. To minimize batch effects and ensure comparability, we re-ran 228 samples from the Year 9 collection using the same reagents and conditions as the Year 15 collection. The difference in average TL between the original and repeat samples was not statistically significant. We estimated a linear

[1] CI = Confidence Interval

relationship between the repeat and original measurements using this sample and adjusted the overall original Year 9 TL data by that correction factor.

It is recommended that users perform a natural log transformation of the data prior to statistical analysis to deal with outliers and correct for the positive skew of the data.

5.2 Additional Telomere Length Variables (*_tlmeasured, *_tlpassedqc)

In addition to telomere length, there are binary variables that indicate whether TL was measured (1=yes, 0=no), and if the measurement passed QC or if it was trimmed as indicated above (1=passed QC, 0=did not pass QC/trimmed).

5.3 DNA Methylation Variables

5.3.1 DNA Methylation Data Acquisition and Processing

Approximately 500 ng of genomic DNA (quantified using the Quant-iT Picogreen ds DNA Assay Kit as described above) was subjected to bisulfite conversion using the EZ-96 DNA Methylation Kit (Zymo Research) and analyzed with the Illumina Infinium Human Methylation450K (450K) or Illumina Infinium MethylationEPIC (EPIC) array according to the manufacturer's protocol. Bisulfite conversion and array processing was performed by the Pennsylvania State College of Medicine Genome Sciences Core facility[2]. Age 9 and 15 samples were run at the same time to minimize technical variation. Otherwise, samples were randomized.

The red and green image pairs were read into R [10]. QC of the methylation data was initially performed with EWAStools [11]. Probes were removed if the detection value was greater than 0.01 or 0.05 for the 450K or EPIC arrays, respectively. Probes were also removed if the number of methylated or unmethylated bead count was fewer than four. Probes were also removed if they were identified by the ENmix function QCinfo using the default parameters [12]. Samples were removed if they had outlier methylation or bisulfite conversion values, as identified by the ENmix QC function or if the sex predicted from the methylation data differed from the recorded sex. If the sequential samples from the same individual exhibited genetic discordance between visits the sample was flagged (see DNA_flag above). The ENmix preprocessENmix and rcp functions were used to normalize dye bias, apply background correction and adjust for probe-type bias [12,13].

5.3.2 Cell Proportion Estimates

Four different methods/reference panels were used to estimate cell proportions from the methylation data.

*EpiDISH Estimates (*_epi, *_fib, *_ic)*

[2] The Genome Sciences Core (RRID:SCR_021123) services and instruments used in this project were funded, in part, by the Pennsylvania State University College of Medicine via the Office of the Vice Dean of Research and Graduate Students and the Pennsylvania Department of Health using Tobacco Settlement Funds (CURE). The content is solely the responsibility of the authors and does not necessarily represent the official views of the University or College of Medicine. The Pennsylvania Department of Health specifically disclaims responsibility for any analyses, interpretations or conclusions.

EpiDISH infers the proportions of a priori known cell-types present in a sample representing such cell-types [14]. The variables ending in epi, fib, and ic represent the estimated proportions of epithelial cells, fibroblasts, and immune cells, respectively.

*Child Saliva Panel Estimates (*_epithelial, *_immune)*

Derived using the Middleton reference panel estimates for saliva in the EWAStools package [11,15]. The panel was created by collecting saliva from 22 children, ranging in age from seven to 16 years of age, sorting each sample into two fractions: immune and epithelial cells. DNA was then isolated, and DNA methylation analyzed on EPIC arrays. Variables represent the proportion of epithelial (*_epithelial) and immune (*_immune) cells.

*Blood Immune Cell Panel (*_gr, *_nk, *_b, *_cd4, *_cd8, *_mo)*

Derived using the Salas reference panel estimates for blood immune cell proportions in EWAStools [11,16]. Variables represent the proportion of granulocytes (*_gr), natural killer cells (*_nk), B cells (*_b), CD4 cells (*_cd4), CD8 cells (*_cd8) and monocyles (*_mo).

Plasma blasts, $CD8^+$ $CD28^-$ $CD45RA^-$ T cells, naïve CD8 T cells (Horvath method [17])

Variables represent estimated abundance measures of plasma blasts (*_plasmablast), $CD8^+$ $CD28^-$ $CD45RA^-$ T cells (*_cd8pcd28ncd45ran) and naïve CD8 T cells, which are $CD45RA^+CCR7^+$ T cells (*_cd8_naive). These should not be interpreted as counts or percentages but rather as ordinal abundance measures. Do not convert to proportions. Negative values indicate very low values. The author (S. Horvath) of the method would not set a negative value to zero but would not object if one does.

5.4 Epigenetic biomarkers

Before documenting each DNA methylation biomarker below, it is important to note that slight variations in the algorithm code, imputation of missing values, and even sample processing used by different packages or websites can result in highly correlated (>0.97), but slightly different measures of the same epigenetic biomarker. For example, some packages may impute missing CpGs while others omit them completely, and some packages include the intercept while others do not. Thus, although the overall mean might be slightly different, the associations with other variables should be similar. Data users should be cautious about interpreting mean differences between versions of the same epigenetic biomarker for these reasons.

5.4.1 Note about missing probes

Subsets of probes used in the various epigenetic biomarker algorithms were missing due to QC or array platform. Some of the clocks represented in this data release impute these missing probes using a "Gold Standard." The number of probes used in each clock (Probe Number), proportion of missing probes for each array platform, and the "Gold Standard" used for imputation (if performed) were:

	Clock*	450K	EPIC	Probe Number	Imputation
1	Horvath	0.023	0.102	353	Horvath27K
2	Hannum	0.014	0.169	71	GrimCpG
3	PedBE	0.043	0.074	94	None
4	SkinBlood	0.043	0.054	391	None
5	PhenoAge	0.023	0.045	513	None
6	Grim	0.09	0.161	30084	GrimCpG
7	PoAm38	0.109	0.109	46	Dunedin
8	PoAm45	0.046	0.069	173	Dunedin
9	PC	0.012	0.076	78464	Hannum2013

*Please see sections below for details regarding each epigenetic biomarker.

5.4.2 Individual DNA methylation biomarkers

Horvath Pan-Tissue Estimator (*horvath)

This clock, developed in 2013 by Steve Horvath, was trained and validated via penalized regression using 82 Illumina DNA methylation array datasets (Infinium27K and Infinium450K) which included 8,000 samples from of more than 30 different tissue and cell types collected from cord blood, children, and adults. Through the use of a transformed version of chronological age as an outcome measure, it defined a clock based on 353 CpGs to predict epigenetic age across a broad spectrum of human cell types and tissues. [17].

Hannum (*hannum)

This clock was developed by Gregory Hannum, Trey Ideker, Kang Zhang and colleagues [18] using Illumina 450K data which analyzed DNA methylation of blood samples from 656 human adults ranging in age from 19 to 101. The aging model included gender, BMI, diabetes status, ethnicity, and batch as covariates and chronological age as the outcome measure.

The Skin and Blood Clock (*skinblood)

Realizing that existing epigenetic clocks did not perform well with *ex vivo* samples, Steve Horvath and colleagues developed the skin and blood epigenetic clock in 2018 to improve epigenetic age estimates of human fibroblasts, keratinocytes, buccal cells, endothelial cells, skin, and blood samples [19]. Illumina 450K and EPIC array data from buccal, epithelial, fibroblast, skin, endothelial, keratinocyte, lymphoblast, dermis and epidermis cells, as well as whole blood, cord blood and saliva were used to create this estimator based upon 391 CpGs. Despite sharing 45 and 60 CpGs with the Hannum [18] and original Horvath [17] clocks, the skin and blood epigenetic clock estimates only moderately correlate with those of these earlier

clocks. The clock was reported to outperform existing DNA methylation-based biomarkers to estimate the chronological ages of human doners of microvascular endothelial cells, skin cells, fibroblasts, keratinocytes, coronary artery endothelial cells, lymphoblastoid cells, blood, and saliva samples. This clock also shows strong age correlations in sorted neurons, glia, brain, liver and bone samples [19].

The Pediatric-Buccal-Epigenetic (PedBE) Clock (*pedbe)

This epigenetic clock was developed by Lisa McEwen, Steve Horvath, and Michael Kobor to improve the estimated epigenetic age derived from pediatric buccal sample data [20]. The clock used a training set consisting of datasets from 1032 Illumina 450K or EPIC arrays of buccal cell genomic DNA from individuals ranging from 0.17 to 19.47 years of age. The PedBE clock is comprised of weighted DNAm values at 94 CpG sites. Estimated DNAmAge values using this clock are reported to be highly correlative with those values generated from Horvath's pan tissue clock, but had a lower test error, reported as the median absolute difference between chronological age and pediatric DNAm age [20]. This clock was also predictive of age in data generated from saliva samples, but with increased variability, possibly due to the saliva dataset age being reported in years instead of days.

DNAmPhenoAge (*phenoage)

Developed by Morgan Levine, Steve Horvath and colleagues [21], this epigenetic clock was trained using a novel two-step method to generate a lifespan predictor. The first step used a Cox penalized regression model in which the hazard of mortality was regressed on clinical markers (albumin, creatinine, serum glucose, C-reactive protein, lymphocyte percent, mean cell volume, red blood cell distribution width, alkaline phosphatase, and white blood cell count), and chronological age to predict phenotypic age. Elastic net regression where the phenotypic age was predicted by blood DNA methylation data was then used to identify the 513 CpGs comprising the DNAm PhenoAge measure.

DNAm GrimAge (*grim)

The DNAm GrimAge epigenetic clock was developed by Ake Lu, Steve Horvath and colleagues. The authors used Framingham Heart Study data, including DNA methylation data from the HumanMethylation450K BeadChip array, from 2356 individuals composed of 888 pedigrees to construct a mortality risk estimator from DNA methylation data [22,23]. First, estimators for twelve plasma proteins and smoking pack years based on blood methylation data were developed. These DNAm estimators, together with chronological age and sex were then regressed on time-to-death (due to all-cause mortality) using an elastic net Cox regression model which selected the following covariates: seven DNAm-based surrogate plasma protein markers (adrenomedullin (*_adm), beta-2-microglobulim (*_B2M), cystatin C (*_Cystatin_C), growth/differentiation factor 15 (GDF-15; *_GDF_15), leptin (*_leptin), plasminogen activator inhibitor type 1 (PAI-1; *_pai_1), tissue inhibitor metalloproteinases 1 (TIMP-1; *_TIMP_1)), DNAm pack-years (*_PACKYRS), chronological age, and sex. The resulting value was transformed to be in the unit of years to generate DNAm GrimAge [23]. AgeAccelGrim is the raw residual resulting from regressing observed GrimAge on chronological age [23]. The authors also examined the inclusion of imputed blood cell composition in their multivariate Cox regression

models and demonstrated that AgeAccelGrim remained highly predictive of lifespan and time-to-coronary heart disease. DNAm biomarkers remained predictive of lifespan and time-to-CHD after adjusting for blood cell counts. With the exception of leptin, where inclusion of blood cell counts increased significance, the adjustment generally reduced significance.

Dunedin Pace of Aging Methylation (DunedinPoAm; *poam38)

The DunedinPoAm biomarker differs from previous DNA methylation biomarkers in that it uses DNA methylation at a single time point to predict aging-related decline [24]. Individuals age at different rates. The development of this biomarker by Daniel Belsky and colleagues builds upon their previous work that defined Pace of Aging, which incorporated change in participants in the Dunedin Study over a 12 year period for 16 biomarkers (glycated hemoglobin, forced expiratory volume in one second (FEV_1), forced vital capacity ratio (FEV_1/FVC), blood pressure (mean arterial pressure), total cholesterol, leukocyte telomere length, blood urea nitrogen (BUN), cardiorespiratory fitness (VO_2Max), waist-hip ratio, body mass index, lipoprotein(a), triglycerides, high density lipoprotein, apolipoprotein B100/A1 ratio, white blood cell count and periodontal disease). Additionally, the measure incorporated change in high sensitivity C-reactive protein and creatine clearance over a six-year period [25]. This Pace of Aging measure was used in elastic-net regression to create the DunedinPoAm biomarker (*_PoAm38).

Dunedin Pace of Aging Calculated from the Epigenome (DunedinPACE; *poam45)

This biomarker is an updated version of DunedinPoAm which takes into account all previous measures as well as an additional time point (to extend the time period to 20 years), an additional measure (estimated glomerular filtration rate (eGFR)), and improved probe QC [26].

Principal Components of Epigenetic Biomarkers (*_pc*)

These variables have been adjusted for technical variations as described in Higgens-Chen et al. [27]. Please see below for correlation between the original variable and PC adjusted variable.

Original Variable	PC adjusted variable
*_horvath	*_pchorvath1
*_skinblood	*_pchorvath2
*_hannum	*_pchannum
*_phenoage	*_pcphenoage
*_grim	*_pcgrim

5.4.3 Additional methylation biomarker variables

*Bio_4HAStatic (*_bio4hastatic)*

This corresponds to the BioAge4HAStatic measure that extends the predicted age measures based on the 71 CpGs used in the Hannum clock [18] using four epigenetic input variables as described by Dr. Horvath [17,28].

*Chronological age (*_age)*

Chronological age derived from sample receipt age. Please use this instead of other age variables in the public dataset when working with the biomarker data.

*Batch (*_batch)*

Indicates the batch in which the sample was processed (bisulfite conversion and array).

6. MISSING FLAG

**_dnaavailable*: The flag variable indicates the whether a participant provided a sample and if DNA was successfully extracted from that sample, as well as the missing values in the DNA data. More specifically, the following codes are used.

1 Yes

A participant provided a sample and DNA was successfully extracted from that sample.

0 No

A participant did not provide a sample or DNA was not successfully extracted from that sample.

-9 Not in wave

Family did not participant in this assessment.

-3 Missing

Saliva sample failed quality control.

7. DATA DICTONARY

Note: Variable names listed in the data dictionary exclude the first two characters indicating the respondent and wave. "mk/me" as the third and fourth characters are explained in the *Variable Naming Convention* section.

Variable	**Description**
*_tl	Telomere length data
*_adjtl	Adjusted telomere length data
*_tlmeasured	The telomere length was measured.
*_tlpassedqc	TL measurement passed QC or trimmed to pass QC.
_dnaavailable	Whether a participant had the DNA data
*mk_age*me_age	Chronological age derived from sample receipt age
*mk_batch*me_batch	The batch in which the sample was processed (bisulfite conversion and array)
*mk_horvath*me_horvath	Horvath pan-tissue estimator
*mk_skinblood*me_skinblood	The skin and blood clock
*mk_pedbe*me_pedbe	The pediatric-buccal-epigenetic clock
*mk_phenoage*me_phenoage	DNAm PhenoAge
*mk_poam38*me_poam38	Dunedin Pace of Aging Methylation
*mk_poam45*me_poam45	Dunedin Pace of Aging Calculated from the Epigenome

*mk_plasmablast*me_plasmablast	Estimated abundance measures of plasma blasts
*mk_cd8pcd28ncd45ran*me_cd8pcd28ncd45ran	Estimated abundance measures of $CD8^{+}$ $CD28^{-}$ $CD45RA^{-}$ T cells
*mk_cd8_naive*me_cd8_naive	Estimated abundance measures of naïve CD8 T cells, which are $CD45RA^{+}CCR7^{+}$ T cells
*mk_hannum*me_hannum	Hannum
*mk_grim*me_grim	DNAm GrimAge
*mk_gdf_15*me_gdf_15	Growth/differentiation factor 15
*mk_b2m*me_b2m	Beta-2-microglobulim
*mk_cystatin_c*me_cystatin_c	Cystatin C
*mk_timp_1*me_timp_1	Tissue inhibitor metalloproteinases 1
*mk_adm*me_adm	Adrenomedullin
*mk_pai_1*me_pai_1	Plasminogen activator inhibitor type 1
*mk_leptin*me_leptin	Leptin
*mk_packyrs*me_packyrs	DNAm pack-years
*mk_bio4hastatic*me_bio4hastatic	Weighted Hannum clock
*mk_pchorvath1*me_pchorvath1	Principal components adjusted horvath pan-tissue estimator
*mk_pchorvath2*me_pchorvath2	Principal components adjusted skin and blood clock
*mk_pchannum*me_pchannum	Principal components adjusted hannum
*mk_pcpheno*me_pcpheno	Principal components adjusted DNAm PhenoAge
*mk_pcgrim*me_pcgrim	Principal components adjusted DNAm GrimAge
*mk_immune*me_immune	The proportion of immune cells on child saliva panel estimates
*mk_epithelial*me_epithelial	The proportion of epithelial cells on child saliva panel estimates
*mk_epi*me_epi	The proportion of epithelial cells on epiDISH estimates
*mk_fib*me_fib	The proportion of fibroblasts on epiDISH estimates
*mk_ic*me_ic	The proportion of immune cells on epiDISH estimates
*mk_gr*me_gr	The proportion of granulocytes on blood immune cell panel
*mk_nk*me_nk	The proportion of natural killer cells on blood immune cell panel
*mk_b*me_b	The proportion of B cells on blood immune cell panel
*mk_cd4*me_cd4	The proportion of CD4 cells on blood immune cell panel

*mk_cd8*me_cd8	The proportion of CD8 cells on blood immune cell panel
*mk_mo*me_mo	The proportion of monocyles on blood immune cell panel

8. REFERENCES

[1] Mitchell C, Hobcraft J, McLanahan SS, Siegel SR, Berg A, Brooks-Gunn J, et al. Social disadvantage, genetic sensitivity, and children's telomere length. Proc Natl Acad Sci U S A 2014;111:5944–9. https://doi.org/10.1073/pnas.1404293111.

[2] O'Callaghan NJ, Fenech M. A quantitative PCR method for measuring absolute telomere length. Biol Proced Online 2011;13:3. https://doi.org/10.1186/1480-9222-13-3.

[3] Mitchell C, McLanahan S, Schneper L, Garfinkel I, Brooks-Gunn J, Notterman D. Father loss and child telomere length. Pediatrics 2017;140. https://doi.org/10.1542/peds.2016-3245.

[4] Cheng, Zhao Y, Wang S, Zhang F, Russo M, McMahon SB, et al. Repression of telomerase gene promoter requires human-specific genomic context and is mediated by multiple HDAC1-containing corepressor complexes. FASEB J 2017;31:1165–78. https://doi.org/10.1096/fj.201601111R.

[5] Wang S, Zhu J. Evidence for a relief of repression mechanism for activation of the human telomerase reverse transcriptase promoter. J Biol Chem 2003;278:18842–50. https://doi.org/10.1074/jbc.M209544200.

[6] Eisenberg DT. Telomere length measurement validity: the coefficient of variation is invalid and cannot be used to compare quantitative polymerase chain reaction and Southern blot telomere length measurement techniques. Int J Epidemiol 2016;45:1295–8. https://doi.org/10.1093/ije/dyw191.

[7] Lindrose AR, McLester-Davis LWY, Tristano RI, Kataria L, Gadalla SM, Eisenberg DTA, et al. Method comparison studies of telomere length measurement using qPCR approaches: A critical appraisal of the literature. PLoS One 2021;16:e0245582. https://doi.org/10.1371/journal.pone.0245582.

[8] Verhulst S, Aviv A, Benetos A, Berenson GS, Kark JD. Do leukocyte telomere length dynamics depend on baseline telomere length? An analysis that corrects for "regression to the mean." Eur J Epidemiol 2013;28:859–66. https://doi.org/10.1007/s10654-013-9845-4.

[9] McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychological Methods 1996;1:30–46. https://doi.org/10.1037/1082-989X.1.1.30.

[10] R Core Team. R: A language and environment for statistical computing 2021.

[11] Heiss JA, Just AC. Improved filtering of DNA methylation microarray data by detection p values and its impact on downstream analyses. Clin Epigenetics 2019;11:15. https://doi.org/10.1186/s13148-019-0615-3.

[12] Xu Z, Niu L, Li L, Taylor JA. ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip. Nucleic Acids Res 2016;44:e20. https://doi.org/10.1093/nar/gkv907.

[13] Niu L, Xu Z, Taylor JA. RCP: a novel probe design bias correction method for Illumina Methylation BeadChip. Bioinformatics 2016;32:2659–63. https://doi.org/10.1093/bioinformatics/btw285.
[14] Teschendorff AE, Breeze CE, Zheng SC, Beck S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. BMC Bioinformatics 2017;18:105. https://doi.org/10.1186/s12859-017-1511-5.
[15] Middleton LYM, Dou J, Fisher J, Heiss JA, Nguyen VK, Just AC, et al. Saliva cell type DNA methylation reference panel for epidemiological studies in children. Epigenetics 2022;17:161–77. https://doi.org/10.1080/15592294.2021.1890874.
[16] Salas LA, Koestler DC, Butler RA, Hansen HM, Wiencke JK, Kelsey KT, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. Genome Biol 2018;19:64. https://doi.org/10.1186/s13059-018-1448-7.
[17] 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.
[18] Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, 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.
[19] Horvath S, Oshima J, Martin GM, Lu AT, Quach A, Cohen H, 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.
[20] McEwen LM, O'Donnell KJ, McGill MG, Edgar RD, Jones MJ, MacIsaac JL, et al. The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. Proc Natl Acad Sci U S A 2020;117:23329–35. https://doi.org/10.1073/pnas.1820843116.
[21] Levine ME, Lu AT, Quach A, Chen BH, Assimes TL, Bandinelli S, 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.
[22] Dawber TR, Meadors GF, Moore FE. Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health 1951;41:279–81. https://doi.org/10.2105/ajph.41.3.279.
[23] Lu AT, Quach A, Wilson JG, Reiner AP, Aviv A, Raj K, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging (Albany NY) 2019;11:303–27. https://doi.org/10.18632/aging.101684.
[24] Belsky DW, Caspi A, Arseneault L, Baccarelli A, Corcoran DL, Gao X, et al. Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. Elife 2020;9:e54870. https://doi.org/10.7554/eLife.54870.
[25] Belsky DW, Caspi A, Houts R, Cohen HJ, Corcoran DL, Danese A, et al. Quantification of biological aging in young adults. Proc Natl Acad Sci U S A 2015;112:E4104-4110. https://doi.org/10.1073/pnas.1506264112.
[26] Belsky DW, Caspi A, Corcoran DL, Sugden K, Poulton R, Arseneault L, et al. DunedinPACE, a DNA methylation biomarker of the pace of aging. Elife 2022;11:e73420. https://doi.org/10.7554/eLife.73420.
[27] Higgins-Chen AT, Thrush KL, Wang Y, Minteer CJ, Kuo P-L, Wang M, et al. A computational solution for bolstering reliability of epigenetic clocks: Implications for clinical trials and

longitudinal tracking. Nat Aging 2022;2:644–61. https://doi.org/10.1038/s43587-022-00248-2.
[28] Horvath S. TUTORIAL for the Online Age Calculator: Estimate DNA methylation age 2013.