Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Blood transcriptomic associations of epigenetic age in adolescents

Dennis Khodasevich [a], Anne K Bozack[a], Saher Daredia[b], Julianna Deardorff[c], Kim G Harley[c], Brenda Eskenazi[c], Weihong Guo[c], Nina Holland[c], and Andres Cardenas[a]

[a]Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA, USA; [b]Division of Epidemiology, Berkeley Public Health, University of California, Berkeley, CA, USA; [c]Center for Environmental Research and Community Health (CERCH), Berkeley Public Health, University of California, Berkeley, CA, USA

ABSTRACT

Epigenetic aging in early life remains poorly characterized, and patterns of gene expression can provide biologically meaningful insights. Blood DNA methylation was measured using the Illumina EPICv1.0 array and RNA sequencing was performed in blood in 174 adolescent participants (age range: 14–15 years) from the CHAMACOS cohort. Thirteen widely used epigenetic clocks were calculated, and their associations with transcriptome-wide RNA expression were tested using the *limma-voom* pipeline. We found evidence for substantial shared associations with RNA expression between different epigenetic clocks, including differential expression of *MYO6* and *ZBTB38* across five clocks. The epiTOC2, principal component (PC) PhenoAge, Hannum, PedBE and PC Hannum clocks were associated with differential expression of the highest number of RNAs, exhibiting associations with 22, 8, 5, 3, and 2 transcripts respectively. Generally, biological clocks were associated with differential expression of more genes than chronological clocks, and PC clocks were associated with differential expression of more genes relative to their CpG-trained counterparts. A total of 17 associations in our study were replicated in an independent adult sample (age range: 40–54 years). Our findings support the biological relevance of epigenetic clocks in adolescents and provide direction for selection of epigenetic ageing biomarkers in adolescent research.

Background

Epigenetics provides a vital biomarker linking exposures, variation in gene expression, and related health outcomes [1]. Specifically, age-related DNA methylation changes have been shown to predict chronological age, leading to the development of epigenetic clocks. Epigenetic clocks are predictors of chronological and biological age derived from DNA methylation, and have proven to be a valuable resource for studying the determinants and consequences of biological aging in human populations. Epigenetic age acceleration (EAA), broadly defined as the deviation between one's epigenetic age prediction and their chronological age, has been found to be robustly associated with a variety of outcomes, including mortality, cardiovascular disease, and cognitive decline [2]. Similarly, epigenetic clocks can be sensitive to environmental exposures, including smoking, ambient air pollution, and early life adversity [3]. Epigenetic aging research has largely focused on adult populations, leaving epigenetic aging in childhood or adolescent populations relatively less understood. However, limited research in pediatric and adolescent populations has highlighted associations between epigenetic clocks and measures of growth and development, suggesting potentially distinct associations in younger populations compared to adult populations [4,5].

A myriad of epigenetic clocks have been developed, each with distinct applications. These include several epigenetic clocks trained to predict chronological age, such as the Horvath panTissue clock, Skin&Blood clock, Hannum clock, and PedBE clock [6–9]. Epigenetic clocks have also been developed to focus on different aspects of

the biological aging process including GrimAge, PhenoAge, epiTOC2, and DunedinPACE [10–13]. Each of these biological clocks accomplishes this focus on biological aging using different approaches; for example, DunedinPACE uses DNA methylation from a single timepoint to predict longitudinal trends of 19 indicators of organ system integrity [13]. Similarly, several existing epigenetic clocks have also been retrained by first utilizing principal component analysis (PCA) on the CpG-level data then training predictors on the principal components, with reported noise reduction benefits compared to the traditional CpG-based training process [14].

Transcriptome-wide expression data can impart insightful information on the biological and functional consequences of epigenetic aging in humans by helping determine which genes are differentially expressed in relation to epigenetic aging. An in-depth analysis of associations between several common epigenetic clocks and RNA expression in isolated CD14+ monocytes and the dorsolateral prefrontal cortex derived from an adult population found evidence for strong overlap in transcriptional associations across different epigenetic clocks [15]. Previous studies in various tissue types including cortical, lung, and liver tissues have consistently identified associations between epigenetic age acceleration and transcriptome-wide changes in RNA expression [16–18]. Similarly, epigenetic age acceleration measures from several common epigenetic clocks including Horvath, Skin&Blood, GrimAge, and PhenoAge have been found to be associated with differential expression of microRNAs in blood [19]. However, research on the relationship between epigenetic aging and RNA expression in pediatric populations, a period marked by rapid growth and development, is lacking.

Characterizing the transcriptomic signatures of epigenetic aging in pediatric populations may help elucidate the biological mechanisms underlying epigenetic aging, while also providing a principled method for evaluating the performance of different epigenetic clocks in terms of their functional consequences during this sensitive period of development. Here, we sought to address the key gaps in our understanding of epigenetic aging in a population of 14-year-old participants of the Center for the Health Assessment of Mothers and Children of Salinas (CHAMACOS) study, a primarily Latino population growing up in a low-income agricultural region of California, by performing a transcriptome-wide association study (TWAS) with several commonly used epigenetic clocks.

Methods

Study population

The Center for the Health Assessment of Mothers and Children of Salinas (CHAMACOS) study began with the recruitment of 601 pregnant women from farmworker communities of the Salinas Valley in California, starting in October 1999. At enrollment, women were ≤20 weeks of gestation, English- or Spanish-speaking, Medicare eligible, planning to deliver at the county hospital, and attending prenatal care visits at one of six local community clinics or hospitals. Of the 601 initial enrollees of the cohort, 526 were followed to delivery of live, singleton newborns in 2000–2001. Maternal education level (≤6th grade, 7th-12th grade, ≥High School Graduate) was obtained during pregnancy interviews, and child sex was abstracted from medical records. Children were followed up at regular intervals throughout childhood and adolescence. At the 14-year visit, body mass was measured using a foot-to-foot bioimpedance scale (Tanita TBF 300A, Arlington Heights, Illinois, USA), and height was measured in triplicate using a stadiometer and averaged. BMI was calculated as weight in kilograms divided by height in metres squared and compared with sex-specific BMI-for-age percentile data issued by the CDC in 2000. A categorical BMI variable was defined as underweight/normal weight (<85 percentile), overweight (≥85 percentile, <95 percentile), or obese (≥95 percentile). When the children were 14 years of age, a phlebotomist collected child blood samples via venipuncture, which were then refrigerated and transported to the University of California, Berkeley biorepository and stored at −80 °C until analysis. The University of California, Berkeley Committee for the Protection of Human Subjects approved all study activities. Written, informed

consent was obtained from all participating mothers, and written assent was obtained from children at age 14 years.

RNA sequencing

RNA sequencing was conducted by SeqMatic (SeqMatic, Fremont, CA). Total RNA was extracted from human whole blood preserved in PAXgene RNA tubes using the GeneJET Stabilized and Fresh Whole Blood RNA Extraction Kit (Thermo Scientific) at the Children's Environmental Health Laboratory, University of California, Berkeley. Prior to extraction following the manufacturer's instructions, the PAXgene samples were thawed at 4 °C and then incubated at room temperature for 2 hours. The total RNA samples were quantified using Agilent RNA ScreenTape on the Agilent 4200 Tapestation and Invirtrogen Qubit Broad Range DNA at SeqMatic. The lower concentration samples were repeated on Agilent RNA High Sensitivity Screen tape to validate the tentative quality control (QC). The RNA integrity number (RIN) score and the percentage of fragments larger than 200 nucleotides (DV200) were calculated. The minimum requirement for samples to pass the RNA QC was pre-specified at 1 ng/μl with DV200 greater than 50% as the minimum requirement for the downstream library preparation type (Illumina RNA Ligation with Ribo Zero Plus depletion kit). RNA samples with DNA contamination, where DNA was higher than 0–1% underwent additional DNase treatment using DNase I-NEB, 2UL/UL, following 'A Typical DNase I Reaction Protocol' (M0303). The DNA-free RNA was purified with Cytiva Sera-mag select beads (1: 1.8 × of beads) and eluted with 60 μl Nuclease free (NF) water. DNase-treated RNA samples were quantified again for concentration and DV200.

Selected purified and extracted RNA samples were subjugated to DNase digestion treatment and quantified again before starting the selected library preparation. Minimum RNA QC parameters for library preparation included 10 ng of total RNA with DV200 greater than 50% and very low DNA content in RNA (0–1%). RNA Reports were prepared and generated using the Cloud LIMS system, then reviewed before library preparation started. Illumina Stranded Total RNA prep, Ligation with the Ribo Zero Plus kit was used for RNA Sequencing library preparation. A total RNA input of up to 100 ng per sample was used as the starting material, and NF water was added to make the initial reaction volume to a minimum of 11 μl. Illumina Total RNA ligation with Ribo Zero Plus kit was selected to facilitate rich transcriptome analysis by depleting the ribosomal RNAs and globin RNAs while targeting long RNA species for library preparation. Total RNA libraries were pooled by equal volume by using Echo Lab Cyte for sequencing a small-scale test on MiSeq sequencer. The equal volume pool had the adapter dimer removed by using Cytiva Sera-mag beads and then quantified for concentration using an Agilent D1000 Screen. Library pool size 420 bp with sufficient concentration was carried forward to downstream dilution. The pool was diluted with HT1 hybridization Buffer and quantified by qPCR before loading into MiSeq. Normalized pool ratios were checked again in another small-scale MiSeq run before the final sequencing on the NovaSeq run.

RNA sequencing QC of fastq files was performed using *FastQC* and *multiQC* [20,21]. Import and processing of fastq files was performed using the *Rsubread* R package [22]. Raw sequence data was aligned to the GRCh38/hg38 genome build using the *align* function, and fragments were assigned and quantified using the *featureCounts* function. Normalization factors for library size were calculated using the trimmed mean of M-values (TMM) method [23]. Genes with a mean expression level of < 1 log counts per million were removed, leaving 15,358 RNAs for analysis.

DNA Methylation

A sample of 187 adolescents from the CHAMACOS cohort had stored blood samples available for DNA methylation analysis at the age of 14 years. DNA was isolated from the banked blood samples using QIAamp DNA Blood Maxi Kits (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol, with minor modifications, as previously described [24]. DNA aliquots of 1 μg were bisulphite converted using Zymo Bisulfite Conversion Kits (Zymo Research,

Orange, CA). DNA was amplified, enzymatically fragmented, purified, and applied to the Illumina Infinium EPICv1 BeadChips (Illumina, San Diego, CA, USA) according to the Illumina methylation protocol to measure DNA methylation. The EPIC chips were analysed using an Illumina Hi-Scan system. QC steps included the use of replicates and randomization of samples across chips and plates [25]. Methylation data were imported into R statistical software for preprocessing using *minfi* [26]. We first performed QC at the sample level, excluding samples with overall low intensities (<10.5) and technical replicates. We computed detection *p* values relative to the control probes and excluded probes with non-significant detection (p > 0.01). We preprocessed our data using functional normalization [27], adjusted for probe-type bias using the regression on correlated probes method [28], and used ComBat from the *sva* package to adjust for sample plate as a technical batch [29]. We visualized the data using density distributions at all processing steps and performed PC analyses to examine the associations of methylation variability with technical, biological, and measured traits and global DNAm variation using PCA plots. The estimated proportions of CD8T, CD4T, B cells, monocytes, and granulocytes were calculated using *EpiDish* [30].

We calculated six epigenetic clocks using the *methylCIPHER* R package, including Horvath's panTissue clock, Skin&Blood, Hannum, PedBE, PhenoAge, and epiTOC2 [31]. The DunedinPACE measure was calculated using the corresponding R package [13], and GrimAge and its component measures were calculated using the Clock Foundation online calculator [6]. We additionally calculated the principal component versions of Horvath's panTissue clock, Skin&Blood, Hannum, PhenoAge, GrimAge, and GrimAge components [14]. A full list of all included epigenetic clock measures is presented in Supplemental Table S1.

Statistical Analysis

The fit of each epigenetic age prediction to chronological age was assessed using Pearson's correlation coefficients and median absolute errors (MAE). (Supplemental Figure S1) Correlations between epigenetic age measures from each clock are presented in Supplemental Figure S2, and the number of CpG sites used by each epigenetic clock and the number of CpG sites shared between epigenetic clocks is presented in Supplemental Figure S3. To determine the strength of associations between each of the covariates and RNA expression, exploratory bivariate models considering each of the estimated cell proportions, child sex, chronological age, maternal education, and BMI were run using *limma-voom* pipeline from the *EdgeR* package using TMM normalized counts [32,33]. Associations between epigenetic age measures and transcriptome-wide RNA expression were calculated using the *limma-voom* pipeline, using TMM normalized counts and two levels of covariate adjustment. Minimally-adjusted models were adjusted for maternal education level during pregnancy, chronological age, and child sex, while fully-adjusted models were further adjusted for five cell-type proportions (CD8T, CD4T, B cells, monocytes, and granulocytes) calculated using *EpiDish* and child BMI category. Associations exhibiting Holm-Bonferroni-adjusted p-values <0.05 considering the number of independent tests to be 199,654 (15,358 genes * 13 clocks) were judged to be statistically significant. Furthermore, associations exhibiting Holm-Bonferroni-adjusted p-values <0.05 considering the number of independent tests to be 15,358 are discussed as suggestive. Genomic inflation factors (lambdas) were calculated for each model series to examine the potential for unmeasured confounding.

Enrichment of KEGG pathways for clocks exhibiting associations with at least five suggestive or significant differentially expressed RNAs was determined using the *enrichKEGG* function in the *clusterProfiler* R package, and KEGG pathways with an FDR-adjusted p-value <0.05 were considered significant [34]. Finally, we implemented Hallmark gene set enrichment testing using the *msigdbr* R package for clocks exhibiting at least five suggestive or significantly differentially expressed RNAs to further characterize potential biological processes implicated in our findings [35,36]. As a sensitivity analysis, we further examined univariate associations between GrimAge plasma protein components and expression of their corresponding transcripts using univariate generalized linear regression.

Next, we examined the extent to which the observed associations between each epigenetic clock and RNA expression in adolescents replicated in an unrelated population of 136 healthy adults from the public GEO GSE224624 dataset [37]. DNA methylation data from the GSE224624 dataset was used to generate epigenetic age measures and DNAm-derived cell proportion estimates as described above. RNAseq gene expression counts were expressed in TMM normalized counts, and associations between epigenetic age measures and transcriptome-wide RNA expression were calculated using limma models with empirical Bayes adjustment, further adjusted for participant sex, chronological age, and five cell-type proportions (CD8T, CD4T, B cells, monocytes, and granulocytes). Only significant or suggestive associations identified in the fully-adjusted analysis in the CHAMACOS cohort were tested for replication. Of the 60 unique genes found to be suggestively or significantly associated with an epigenetic age measure in the fully-adjusted CHAMACOS models, 39 were detected in the validation sample, allowing for testing of 62 of the 93 clock-RNA associations in the validation sample. Associations exhibiting Holm-Bonferroni-adjusted p-values <0.05 are discussed as significant, and all associations with unadjusted p-values <0.05 are further discussed as suggestive.

Results

Study population and introductory analysis

Study participants included 174 adolescent participants of the CHAMACOS cohort ranging from 14.0 to 15.1 years of age (SD = 0.22) at the time of sample collection. (Table 1) 54.6% of the study participants were female, 56.9% of the study population were overweight or obese, and 43% of the participants' mothers had less than or equivalent to a 6th grade education at the time of study enrollment.

The highest correlation between epigenetic age and chronological age was observed with the Skin&Blood clock ($r = 0.156$), while the lowest median absolute error was observed with the Horvath clock (MAE = 1.98), however, the limited age range of study participants (14 to 15.1 years) limits the ability to effectively evaluate the fit of epigenetic clocks. (Supplemental Figure S1) Epigenetic age measures from most of the included epigenetic clocks were moderately to strongly correlated with each other, with correlations ranging from −0.16 to 0.89. (Supplemental Figure S2) Some of the strongest correlations between epigenetic age measures were observed between the PC Horvath and PC Skin&Blood ($r = 0.89$) clocks, as well as between the PC Horvath and PC Hannum ($r = 0.87$) clocks. All of the included CpG-trained clocks shared at least 1 CpG site with another clock, however, relatively few CpG sites were shared between multiple epigenetic clocks, with the Horvath and Skin&Blood clocks sharing the most CpG sites (60 shared sites). (Supplemental Figure S3) The exploratory bivariate TWAS on each of the covariates revealed that DNAm-derived estimates of cell proportions were strongly associated with differential expression, ranging from 291 differentially expressed RNAs in relation to Monocyte cell proportion to 5,066 RNAs in relation to Granulocyte cell proportion. BMI category was associated with differential expression of 6 genes, sex was associated with 116

Table 1. Study population demographics. Count (%) provided for categorical variables and mean (range) provided for continuous variables.

Trait	Distribution
Chronological age (years)	14.1 (14.0–15.1)
Sex	
Male	79 (45.4%)
Female	95 (54.6%)
BMI Category	
Underweight/Normal	75 (43.1%)
Overweight	34 (19.5%)
Obese	65 (37.4%)
Maternal Education	
≤6th grade	75 (43.1%)
7th-12th grade	60 (34.5%)
≥High School Graduate	39 (22.4%)

genes, and maternal education was not associated with gene expression. Interestingly, in bivariate models, chronological age was only associated with differential expression of two genes (*ATG13*: logFC = −0.32, 95% CI: −0.43, −0.20; *ACTA1*: logFC = −0.74, 95% CI: −1.02, −0.46), likely due to the limited chronological age range of our study sample.

Transcriptomic analysis

The minimally-adjusted models adjusted for maternal education level during pregnancy, chronological age, and child sex produced a large number of hits for most epigenetic clocks (0 to 2,826 genes; *Holm-Bonferroni p-value <0.05*), with a general trend of a greater number of associations identified with biological clocks (172 to 2,826 genes) compared to chronological clocks (0 to 1,818 genes), and a greater number of associations identified PC-trained clocks (202 to 2,826 genes) compared to their corresponding CpG-trained clocks (0 to 2,067 genes). (Table 2) There was a high degree of overlap in significant associations identified for different clocks, including differential expression of lymphoid enhancer binding factor 1 (*LEF1*) and SATB homeobox 1 (*SATB1*) across 10 epigenetic clocks. Because of evidence for inflation in the minimally-adjusted models

(Supplemental Table S2) and strong associations between estimated cell proportions and gene expression in the introductory analysis, we focus most interpretation on the fully-adjusted models, which were additionally adjusted for the five estimated cell type proportions and BMI.

Complete summaries of all significant associations observed for each clock from the fully-adjusted models are summarized in Table 3. In the fully-adjusted models, the epiTOC2, PC PhenoAge, Hannum, PedBE, and PC Hannum measures were associated with differential expression of the highest number of genes, with 22, 8, 5, 3, and 2 RNAs respectively. (Table 2) (Supplemental Figure S4) A total of 24 RNAs were found to be suggestively or significantly differentially expressed across multiple epigenetic clocks. (Figure 1) When considering suggestive and significant associations, the highest degree of overlap in differential expression of a single RNA across multiple clocks was observed with differential expression of *MYO6* (myosin VI) in response to epigenetic age measures from 5 clocks (Hannum, PedBE, epiTOC2, PC Skin&Blood, and PC Hannum), as well as differential expression of *ZBTB38* (zinc finger and BTB domain containing 38) in response to epigenetic age measures from 5 clocks (Hannum, PhenoAge, epiTOC2, PC Hannum, and PC PhenoAge). The highest degree

Table 2. Summary detailing the number of RNA associations for each epigenetic clock. The number of component CpGs is specified for each non-PC clock. Minimal models were adjusted for chronological age, maternal education, and child sex, while fully adjusted models were additionally adjusted for BMI, and estimated cell proportions. RNA associations were considered significant if exhibiting a stringent Holm-Bonferroni-adjusted p-value <0.05 (considering 199,654 independent tests), and considered suggestive if exhibiting a Holm-Bonferroni-adjusted p-value <0.05 (considering 15,358 independent tests).

Clock	CpGs in Clock	Minimally-Adjusted		Fully-Adjusted	
		RNA Hits (suggestive)	RNA Hits (significant)	RNA Hits (suggestive)	RNA Hits (significant)
Chronological Clocks					
Horvath	353	0	0	0	0
Skin&Blood	391	6	2	0	0
Hannum	71	467	231	10	5
PedBE	94	197	65	9	3
PC Horvath	—	512	202	0	0
PC Skin&Blood	—	1499	918	3	0
PC Hannum	—	2637	1818	11	2
Biological Clocks					
GrimAge	1030	429	172	0	0
PhenoAge	513	2907	2067	1	0
DunedinPACE	173	484	209	1	1
epiTOC2	163	501	276	33	22
PC GrimAge	—	3279	2496	3	0
PC PhenoAge	—	3656	2826	22	8

Table 3. Significant RNA associations from fully-adjusted models, adjusted for chronological age, maternal education, child sex, BMI, and estimated cell proportions. RNA hits were considered significant with a stringent Holm-Bonferroni-adjusted p-value <0.05. Effect estimates correspond to the expected log-fold change in expression for a 1-unit increase in epigenetic age.

Clock	logFC	95% CI	P-Value	Gene ID	Gene Name
Hannum	0.10	0.07, 0.13	1.22E−09	KLRA1P	killer cell lectin like receptor A1, pseudogene
Hannum	0.11	0.07, 0.14	3.96E−09	MYO6	myosin VI
Hannum	0.05	0.03, 0.07	1.45E−07	PPP2R2B	protein phosphatase 2 regulatory subunit Bbeta
Hannum	0.10	0.06, 0.13	1.71E−07	GZMH	granzyme H
Hannum	0.03	0.02, 0.05	2.02E−07	ZBTB38	zinc finger and BTB domain containing 38
PedBE	0.43	0.29, 0.57	3.77E−09	FCRL6	Fc receptor like 6
PedBE	0.58	0.38, 0.77	3.64E−08	ZNF365	zinc finger protein 365
PedBE	0.53	0.34, 0.71	5.20E−08	MYO6	myosin VI
DunedinPACE	9.83	6.68, 12.98	5.25E−09	RAP1GAP	RAP1 GTPase activating protein
epiTOC2	8.2E−04	6.0E−04, 1.0E−03	1.54E−11	FCRL6	Fc receptor like 6
epiTOC2	1.1E−03	8.0E−04, 1.4E−03	4.56E−11	ZNF365	zinc finger protein 365
epiTOC2	1.0E−03	7.4E−04, 1.3E−03	5.89E−11	MYO6	myosin VI
epiTOC2	1.0E−03	7.3E−04, 1.3E−03	3.37E−10	GZMH	granzyme H
epiTOC2	5.6E−04	3.9E−04, 7.3E−04	7.08E−10	KIF21A	kinesin family member 21A
epiTOC2	9.6E−04	6.6E−04, 1.3E−03	1.53E−09	CADM1	cell adhesion molecule 1
epiTOC2	8.2E−04	5.6E−04, 1.1E−03	2.18E−09	TPRG1	tumor protein p63 regulated 1
epiTOC2	1.3E−03	8.7E−04, 1.7E−03	2.81E−09	GLB1L2	galactosidase beta 1 like 2
epiTOC2	3.4E−04	2.3E−04, 4.5E−04	4.34E−09	ZBTB38	zinc finger and BTB domain containing 38
epiTOC2	5.0E−04	3.4E−04, 6.7E−04	5.79E−09	TGFBR3	transforming growth factor beta receptor 3
epiTOC2	8.3E−04	5.6E−04, 1.1E−03	6.52E−09	KLRA1P	killer cell lectin like receptor A1, pseudogene
epiTOC2	1.1E−03	7.1E−04, 1.4E−03	1.24E−08	ARHGEF28	Rho guanine nucleotide exchange factor 28
epiTOC2	1.0E−03	6.8E−04, 1.4E−03	3.00E−08	LINC00943	long intergenic non-protein coding RNA 943
epiTOC2	9.0E−04	5.9E−04, 1.2E−03	3.93E−08	KIAA1671	KIAA1671
epiTOC2	1.4E−03	9.0E−04, 1.9E−03	6.25E−08	EFNA5	ephrin A5
epiTOC2	4.1E−04	2.7E−04, 5.5E−04	7.09E−08	GFI1	growth factor independent 1 transcriptional repressor
epiTOC2	4.9E−04	3.2E−04, 6.7E−04	8.52E−08	PPP2R2B	protein phosphatase 2 regulatory subunit Bbeta
epiTOC2	4.1E−04	2.7E−04, 5.6E−04	8.97E−08	MAF	MAF bZIP transcription factor
epiTOC2	9.3E−04	6.0E−04, 1.3E−03	9.03E−08	KLRC3	killer cell lectin like receptor C3
epiTOC2	1.1E−03	7.1E−04, 1.5E−03	1.12E−07	KLRC2	killer cell lectin like receptor C2
epiTOC2	7.1E−04	4.5E−04, 9.7E−04	1.78E−07	SOX13	SRY-box transcription factor 13
epiTOC2	4.6E−04	2.9E−04, 6.3E−04	1.99E−07	EOMES	eomesodermin
PC Hannum	0.09	0.06, 0.12	1.10E−07	ACSM1	acyl-CoA synthetase medium chain family member 1
PC Hannum	0.09	0.06, 0.12	2.50E−07	CYSLTR2	cysteinyl leukotriene receptor 2
PC PhenoAge	−0.05	−0.06, −0.03	1.09E−08	PLXDC1	plexin domain containing 1
PC PhenoAge	0.03	0.02, 0.04	1.64E−08	ZBTB38	zinc finger and BTB domain containing 38
PC PhenoAge	0.04	0.02, 0.05	2.66E−08	INPP1	inositol polyphosphate-1-phosphatase
PC PhenoAge	0.09	0.06, 0.12	4.96E−08	WFDC21P	WAP four-disulfide core domain 21, pseudogene
PC PhenoAge	−0.07	−0.09, −0.04	8.16E−08	NOG	noggin
PC PhenoAge	0.11	0.07, 0.15	2.09E−07	SIGLEC8	sialic acid binding Ig like lectin 8
PC PhenoAge	0.07	0.05, 0.10	2.24E−07	FRRS1	ferric chelate reductase 1
PC PhenoAge	−0.05	−0.06, −0.03	2.28E−07	CCR7	C-C motif chemokine receptor 7

of overlap between two epigenetic clocks, shared suggestive and significant associations with 10 RNAs, was found between epiTOC2 and the Hannum clock. We also found substantial overlap in associations observed between PC-trained clocks and their CpG-trained counterparts in the minimally-adjusted models, ranging from 50–100% of associations identified with the CpG clocks also being found with the corresponding PC clocks, with relatively less overlap in the fully-adjusted models, ranging from 20–100% of associations identified with the CpG clocks also being found with the corresponding PC clocks for clocks exhibiting suggestive or significant associations

with both the CpG and PC-trained versions. (Supplemental Table S3)

In addition to the shared association across several clocks, 36 RNAs exhibited distinct associations with only a single clock. DunedinPACE was associated with differential expression of *RAP1GAP* (logFC = 9.83, 95% CI: 6.68, 12.98). Distinct associations in relation to epiTOC2 included differential expression of *KIF21A* (logFC = 5.6E−04, 95% CI: 3.9E−04, 7.3E−04) and *TGFBR3* (logFC = 5.0E−04, 95% CI: 3.4E−04, 6.7E−04). PC PhenoAge was associated with differential expression of several distinct RNAs including *NOG* (logFC = −0.07, 95% CI: −0.09,
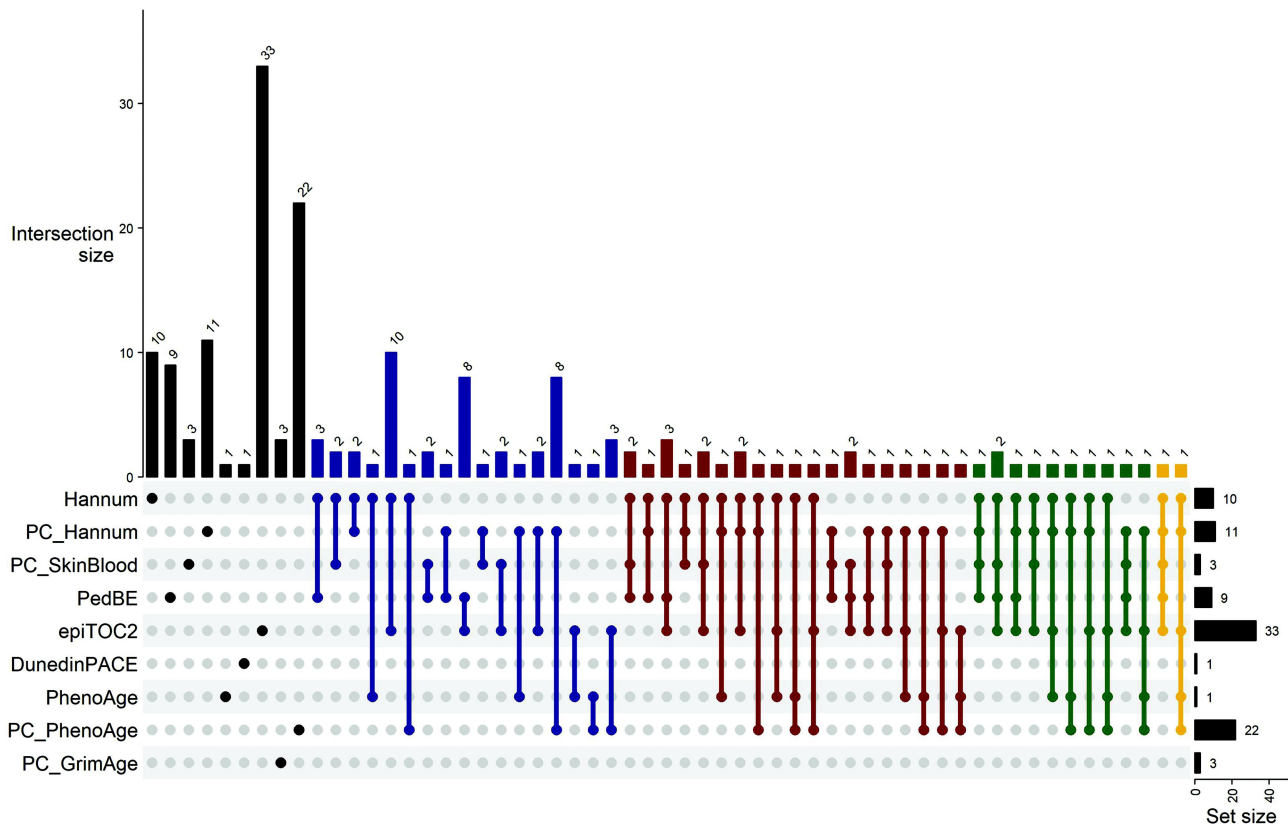
Figure 1. Number of suggestively or significantly differentially expressed RNAs associated with each epigenetic clock and number of common associations among clocks from the fully-adjusted models. Color corresponds to the degree of overlap (blue = 2 clocks, red = 3 clocks, green = 4 clocks, and yellow = 5 clocks).

−0.04) and *FRRS1* (logFC = 0.07, 95% CI: 0.05, 0.10). PC GrimAge was suggestively associated with the differential expression of three immunoglobulins: *IGKV2D–30* (logFC = 0.27, 95% CI: 0.16, 0.37), *IGLV3–1* (logFC = 0.26, 95% CI: 0.16, 0.36), and *IGKV4–1* (logFC = 0.22, 95% CI: 0.13, 0.30).

The GrimAge clock is a composite epigenetic clock consisting of seven epigenetic estimators of plasma protein levels (adrenomedullin (ADM), beta-2-microglobulin (B2M), cystatin C, growth differentiation factor 15 (GDF15), leptin, plasminogen activator inhibitor 1 (PAI1), and tissue inhibitor metalloproteinase 1 (TIMP1)) and smoking pack-years [10]. We further examined potential associations between GrimAge components and their corresponding transcripts. Excluding leptin, which was not detected at a sufficiently high frequency in the study population, 3/6 CpG-trained and 4/6 PC-trained GrimAge components were positively associated (unadjusted p-values <0.05) with expression of their corresponding transcripts,

indicating that GrimAge performs well in the adolescent study population despite being trained in an entirely adult sample. (Supplemental Table S4).

Each clock's epigenetic age estimates are derived from DNA methylation measurements from CpG sites throughout the genome, so we next set out to determine the extent to which observed suggestive and significant associations are directly related to the genes mapped to each clock's CpG sites. In the minimally-adjusted models, 77 PhenoAge associations, 17 GrimAge associations, 6 DunedinPACE associations, 1 epiTOC2 association, and 1 Hannum association were observed in genes mapped directly to component CpG sites. No associations were observed in genes mapped to component CpG sites in the fully-adjusted models, indicating that the observed associations are largely related to the overall epigenetic age signal, rather than being driven by associations related to changes in methylation at individual component CpG sites. A complete summary of each fully-adjusted

TWAS is provided in Supplemental Tables S5–17 [38].

Biological pathway enrichment

Using all genes found to be suggestively or significantly associated with an epigenetic clock as input for biological pathway enrichment, several KEGG pathways were found to be enriched in the minimally-adjusted models across multiple epigenetic clocks (*FDR* < 0.05). The Leishmaniasis (hsa05140) pathway was found to be significantly enriched across 10 different epigenetic clocks. In contrast, no KEGG pathways were significantly enriched for any of the epigenetic clocks in the fully-adjusted models. We then implemented hallmark gene set enrichment testing to further characterize the potential biological processes implicated in our findings. Several pathways were found to be significantly enriched across multiple clocks in the minimally-adjusted models, including Inflammatory Response and TNFa Signaling via NFKb, which were both found to be enriched across nine epigenetic clocks (*FDR* < 0.05). In the fully-adjusted models, the only significant pathway after multiple testing adjustment was the epithelial-mesenchymal transition pathway in relation to the PedBE clock.

External replication

We next examined the consistency of our results in an unrelated adult population by testing for replication of suggestive and significant findings in the GSE224624 external dataset consisting of 136 adults aged 40 to 54 years [37]. A total of 17 associations in the CHAMACOS sample were found to exhibit the same direction of association and an unadjusted p-value <0.05 in the GSE224624 dataset. (Table 4) This included nine associations related to epiTOC2, one association related to the PC Hannum clock, and seven associations related to PC PhenoAge. Among these associations, only one positive association between epiTOC2 and differential expression of *NUGGC* (logFC = 3.5E–04, 95% CI: 1.7E–04, 5.4E–04) remained significant after multiple testing adjustment.

Discussion

We performed a systematic cross-sectional analysis of transcriptomic associations with commonly used epigenetic clocks in an adolescent population. Within our primary study sample, the highest number of differentially expressed genes in fully-adjusted models were observed with the epiTOC2, PC PhenoAge, Hannum, PedBE, and PC Hannum

Table 4. Validated RNA associations from the GSE224624 dataset. Associations found to be suggestive or significant in the CHAMACOS sample and found to exhibit both the same direction of effect and an unadjusted p-value <0.05 in the GSE224624 dataset are displayed. ** indicates an association with a Holm-Bonferroni-adjusted p-value <0.05 in the GSE224624 dataset.

Clock	CHAMACOS Population			GSE224624 Population			
	logFC	95% CI	P Value	logFC	95% CI	P Value	Gene Name
epiTOC2**	6.8E–04	4.1E–04, 9.6E–04	2.55E–06	3.5E–04	1.7E–04, 5.4E–04	2.93E–04	*NUGGC*
epiTOC2	1.0E–03	7.4E–04, 1.3E–03	5.89E–11	3.1E–04	9.6E–05, 5.3E–04	5.18E–03	*MYO6*
epiTOC2	1.0E–03	7.3E–04, 1.3E–03	3.37E–10	2.9E–04	4.1E–05, 5.3E–04	2.23E–02	*GZMH*
epiTOC2	9.0E–04	5.9E–04, 1.2E–03	3.93E–08	2.9E–04	1.3E–05, 5.7E–04	4.01E–02	*KIAA1671*
epiTOC2	7.1E–04	4.5E–04, 9.7E–04	1.78E–07	1.9E–04	8.3E–06, 3.6E–04	4.04E–02	*SOX13*
epiTOC2	−4.6E–04	−6.4E–04, −2.9E–04	5.88E–07	−2.6E–04	−5.1E–04,-2.0E–05	3.44E–02	*ATP6V0E2-AS1*
epiTOC2	5.2E–04	3.2E–04, 7.3E–04	1.34E–06	2.0E–04	3.8E–05, 3.7E–04	1.66E–02	*MCOLN2*
epiTOC2	7.4E–04	4.5E–04, 1.0E–03	1.92E–06	2.4E–04	6.4E–05, 4.1E–04	7.43E–03	*RAB11FIP5*
epiTOC2	3.4E–04	2.0E–04, 4.7E–04	2.15E–06	1.0E–04	1.3E–05, 2.0E–04	2.62E–02	*SYT11*
PC Hannum	0.08	0.04, 0.11	2.40E–06	0.03	0.00, 0.06	2.40E–02	*PIK3R6*
PC PhenoAge	−0.05	−0.06, −0.03	1.09E–08	−0.03	−0.05, −0.01	6.37E–03	*PLXDC1*
PC PhenoAge	−0.07	−0.09, −0.04	8.16E–08	−0.05	−0.08, −0.02	4.22E–03	*NOG*
PC PhenoAge	−0.05	−0.06, −0.03	2.28E–07	−0.02	−0.03, −0.01	5.23E–03	*CCR7*
PC PhenoAge	0.09	0.05, 0.12	7.94E–07	0.04	0.00, 0.08	4.09E–02	*SLC29A1*
PC PhenoAge	−0.04	−0.06, −0.03	1.97E–06	−0.03	−0.05, −0.01	2.15E–03	*GPA33*
PC PhenoAge	−0.03	−0.04, −0.02	1.99E–06	−0.02	−0.03, −0.01	2.70E–03	*TRABD2A*
PC PhenoAge	0.06	0.04, 0.09	2.38E–06	0.03	0.00, 0.05	2.70E–02	*PIK3R6*

clocks. Our findings support the biological relevance of epigenetic clocks in adolescents, an age group that was not included in the training samples for several of the included epigenetic clocks, and may provide guidance for the selection of epigenetic aging biomarkers for research in adolescent populations. Furthermore, the CHAMACOS study sample represents a primarily Latino community, a population overlooked in the development of most epigenetic clocks. We found evidence of substantial overlap in differentially expressed RNAs when using epigenetic age measures derived from different clocks. Most epigenetic clocks tend to share relatively few CpG sites and epigenetic age measures from different clocks were only moderately correlated with each other, suggesting that shared associations are unlikely to be related to shared component CpG sites or direct correlation between epigenetic aging measures. For example, the Hannum and epiTOC2 clocks were found to share suggestive or significant associations with differential expression of 10 RNAs, but shared no CpG sites and their epigenetic age measures only had a relatively low correlation of 0.25.

A total of 24 RNAs were found to be suggestively or significantly differentially expressed across multiple epigenetic clocks in the fully-adjusted models. Myosin VI (MYO6), a motor protein involved in intracellular vesicle and organelle transport, and zinc finger and BTB domain containing 38 (ZBTB38), a transcriptional activator that binds methylated DNA, were both differentially expressed in response to epigenetic age estimates from five distinct clocks. Dysregulation of MYO6 has been implicated in the progression of several cancers including breast cancer [39], and MYO6 has been found to be upregulated in mouse hippocampus in response to traumatic stress exposure [40]. Psychosocial stressors have been found to be major contributors to epigenetic aging in a range of studies [41,42], which may be reflected in our findings of consistent positive associations between multiple epigenetic clocks and MYO6 expression. ZBTB38 has been implicated in a wide range of downstream processes including cell survival, homeostasis, and cancer progression [43–46]. ZBTB38 polymorphisms and expression have previously been implicated in birth size and growth rates [47], and epigenetic clocks have previously been associated with birth

size and pubertal timing in pediatric populations [4,5]. Therefore, our finding of consistent differential expression of ZBTB38 across multiple epigenetic clocks may be reflective of associations between epigenetic age estimates and altered growth patterns within the study population. Epigenetic age measures from four clocks were found to be associated with differential expression of Granzyme H (GZMH), a serine protease that is highly expressed in NK cells and is involved in NK-mediated apoptosis [48]. Epigenetic age measures from three clocks were found to be associated with differential expression of Fc receptor like 6 (FCRL6), a protein involved in cell surface receptor signaling. FCRL6 is an indicator of cytotoxic effector lymphocytes that is upregulated in conditions with chronic immune stimulation [49].

In addition to associations shared across several clocks, 36 significant or suggestive associations were found to be specific to a single clock in the fully-adjusted models. Each epigenetic clock was trained in unique samples, and, in the case of the biological clocks, trained to predict unique aspects of biological aging, which may help explain why different clocks exhibited unique associations. DunedinPACE was associated with differential expression of RAP1 GTPase activating protein (RAP1GAP). RAP1GAP plays a role in tumor suppression and the epithelial-mesenchymal transition, and lower expression has been implicated in poorer prognosis scores in several cancer types [50,51]. More research is required to help understand whether this finding of differential expression of RAP1GAP is related to the unique design of DunedinPACE, which captures the longitudinal changes in 19 indicators of multi-organ integrity. EpiTOC2, which is designed to capture the cumulative number of stem cell divisions in a tissue, was associated with differential expression of the highest number of genes in our primary study population, with nine associations further validated in an external adult population. Differential expression of unique genes in relation to epiTOC2 included higher expression of Kinesin Family Member 21A (KIF21A), a kinesin-like motor protein involved in neurodevelopment and cancer progression [52,53]. KIF21A has also been found to be overexpressed in patients with Down syndrome compared to controls [54]. EpiTOC2 was also associated with higher expression of LINC00943, a long non-

protein coding RNA previously found to be associated with gastric cancer and hepatocellular carcinoma progression [55,56]. The PC Hannum clock was suggestively associated with higher expression of Ankyrin Repeat and SOCS Box Containing 2 (ASB2), a protein involved in the regulation of NK cell migration [57], as well as Sphingomyelin Phosphodiesterase 3 (SMPD3), a protein involved in the regulation of skeletal development [58]. The PC PhenoAge clock was associated with differential expression of several distinct RNAs including noggin (NOG) and ferric chelate reductase 1 (FRRS1). Noggin is a key signaling molecule that plays a role in early patterning during embryonic development and has been suggested to be involved in processes such as adipogenesis and osteoblastogenesis [59,60]. PhenoAge is a DNAm-based predictor of phenotypic age that incorporates several biomarkers including serum glucose levels. Noggin has been previously found to influence serum glucose levels [61], which may explain the unique association between PC PhenoAge and differential expression of noggin. FRRS1 is involved in the reduction of ferric iron and overexpression has previously been observed in the blood of children with asthma, dermatitis, and rhinitis [62]. The PC GrimAge clock was suggestively associated with higher expression of several immunoglobulins involved in the immune response including IGKV2D–30, IGLV3–1, and IGKV4–1. GrimAge is a composite epigenetic clock consisting of several components related to immune function and inflammation, including B2M and GDF15, which may explain the unique associations between PC GrimAge and differential expression of these immunoglobulins.

A total of 17 associations identified in relation to epiTOC2, PC Hannum, and PC PhenoAge in the CHAMACOS cohort were replicated in an unrelated adult population, indicating some commonality in associations between adolescents and adults. EpiTOC2 was suggestively associated with higher expression of Mucolipin TRP Cation Channel 2 (MCOLN2), a cation channel protein that plays a role in chemokine release and innate immunity [63]. EpiTOC2 was also suggestively associated with higher expression of RAB11 Family Interacting Protein 5 (RAB11FIP5), a Rab11 effector protein involved in immune response and NK cellular function [64], as well as SRY-Box Transcription Factor 13 (SOX13), a critical transcription factor involved in the regulation of embryonic development and progression of several cancer types [65,66]. MYO6 and GZMH, which were consistently associated with a number of epigenetic clock measures in the CHAMACOS population, were also found to be suggestively associated with epiTOC2 in the adult population. Both the PC Hannum and PC PhenoAge clocks were suggestively associated with higher expression of Phosphoinositide-3-Kinase Regulatory Subunit 6 (PIK3R6), a lipid kinase implicated in clear cell renal cell carcinoma progression [67]. PC PhenoAge was also suggestively associated with lower expression of glycoprotein A33 (GPA33), a cell surface antigen involved in CD4+ T cell function [68], as well as lower expression of C-C Motif Chemokine Receptor 7 (CCR7), a GPCR expressed in several immune cells with relevance towards immune cell migration and autoimmune disease [69,70].

Several of the epigenetic clocks examined in our study, including each of the biological clocks, did not include pediatric participants in their training samples. Accordingly, several epigenetic clocks exhibited poor fit with chronological age in our primary study sample, although our ability to effectively evaluate the fit of the clocks was limited by the narrow age range of the study participants. However, our analysis aimed to examine the biological relevance of epigenetic age estimates, rather than directly evaluating the fit of epigenetic clocks with chronological age, and we found substantial evidence for differential expression of RNA in response to epigenetic age estimates from several epigenetic clocks. Similarly, previous studies have found associations between exposures, including childhood maltreatment and community violence, and epigenetic aging of biological clocks in adolescent populations [71,72], supporting the application of biological clocks in adolescents. Furthermore, most of the GrimAge component plasma protein predictors were found to be positively associated with expression of their corresponding transcripts, providing additional support that these predictors perform well despite

the original training sample not including paediatric samples. Taken together, these findings suggest that epigenetic clocks developed in entirely adult populations may still serve as useful biomarkers in adolescents.

Our study is subject to some limitations. First, our primary study sample consisted of adolescents from the CHAMACOS cohort ranging in age from 14 to 15 years, which may limit the generalizability of our findings to other populations and wider age groups. However, our work adds to the ongoing necessity of expanding epigenetics research to more diverse populations and addresses the pressing lack of research characterizing the impact of epigenetic aging in adolescents. Additionally, we were able to validate a portion of our findings in an unrelated adult population. Second, DNA methylation and RNA expression data were obtained from a mixed cell-type population, which limits cell-type specificity and generalizability to other tissue types. However, we adjusted for cell-type estimates, which effectively reduced the genomic inflation factors, suggesting proper control for cell-type heterogeneity within our analysis. Third, our study featured a relatively modest sample size. Fourth, our study utilized DNA methylation and RNA sequencing data measured at the same timepoint, which limited our analysis to examining cross-sectional associations. Lastly, we performed a large number of independent statistical tests in our analysis, increasing the probability of false positives. To account for this, we implemented a stringent multiple testing adjustment to identify high-confidence associations and further discussed suggestive associations that warrant further investigation.

We performed a transcriptome-wide association study of common epigenetic clocks in a population of adolescents from the CHAMACOS cohort. We identified both substantial shared associations between different clocks, suggesting the targeting of a common signal of biological ageing, as well as associations unique to a single clock. Several of the identified differentially expressed genes were involved in biological functions related to immune function and cancer progression. Most of the epigenetic clocks included in this study (with the exception of Horvath, Skin&Blood, PedBE, PC Horvath, PC Skin&Blood) did not include paediatric samples within their training datasets; thus, our findings support the biological relevance of epigenetic clocks in adolescent populations. In particular, we found that the epiTOC2, PC PhenoAge, Hannum, PC Hannum, and PedBE clocks were associated with differential expression of the highest number of genes.

Conclusions

We found that epigenetic aging is associated with differential expression of RNA in a population of adolescents from the CHAMACOS cohort, finding both substantial commonality in associations across different epigenetic clocks and associations unique to a single clock measure. Furthermore, several associations identified in adolescents were replicated in an independent adult sample. Our findings support the biological relevance of epigenetic clocks in adolescents and provide direction for the selection of epigenetic aging biomarkers for research in adolescent populations.

Acknowledgments

Author contributions

CRediT: **Dennis Khodasevich:** Conceptualization, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing; **Anne K Bozack:** Conceptualization, Methodology, Writing – review & editing; **Saher Daredia:** Conceptualization, Methodology, Writing – review & editing; **Julianna Deardorff:** Supervision, Writing – review & editing; **Kim G Harley:** Supervision, Writing – review & editing; **Brenda Eskenazi:** Supervision, Writing – review & editing; **Weihong Guo:** Supervision, Writing – review & editing; **Nina Holland:** Conceptualization, Methodology, Supervision, Writing – review & editing; **Andres Cardenas:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Funding

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

The authors did not receive study participant consent to publicly share CHAMACOS study data. CHAMACOS data analysed during the current study are available from the CHAMACOS study team upon submission of an Initial Data Use Application and institutional review board approval. Expanded information on working with CHAMACOS cohort data is available on the Center for Environmental Research and Community Health website (https://cerch.berkeley.edu/investigators) The GSE224624 dataset is publicly available from the Gene Expression Omnibus database.

Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki, and the University of California, Berkeley Committee for the Protection of Human Subjects approved all study activities. Written, informed consent was obtained for all participating mothers, and written assent was obtained from children at age 14 years.

ORCID

Dennis Khodasevich http://orcid.org/0000-0003-1412-8251

References

[1] Colwell ML, Townsel C, Petroff RL, et al. Epigenetics and the exposome: DNA methylation as a proxy for health impacts of prenatal environmental exposures. Exposome. 2023;3(1):osad001. doi: 10.1093/exposome/osad001

[2] Oblak L, van der Zaag J, Higgins-Chen AT, et al. A systematic review of biological, social and environmental factors associated with epigenetic clock acceleration. Ageing Res Rev. 2021;69:101348. doi: 10.1016/j.arr.2021.101348

[3] Dhingra R, Nwanaji-Enwerem JC, Samet M, et al. DNA methylation age—environmental influences, health impacts, and its role in environmental epidemiology. Curr Envir Health Rpt. 2018;5(3):317–327. doi: 10.1007/s40572-018-0203-2

[4] Simpkin AJ, Howe LD, Tilling K, et al. The epigenetic clock and physical development during childhood and adolescence: longitudinal analysis from a UK birth cohort. Int J Epidemiol. 2017;46(2):549–558. doi: 10.1093/ije/dyw307

[5] Binder AM, Corvalan C, Mericq V, et al. Faster ticking rate of the epigenetic clock is associated with faster pubertal development in girls. Epigenetics. 2018;13(1):85–94. doi: 10.1080/15592294.2017.1414127

[6] Horvath S. DNA methylation age of human tissues and cell types. Genome Biol. 2013;14(10):3156. doi: 10.1186/gb-2013-14-10-r115

[7] Horvath S, Oshima J, Martin GM, et al. Epigenetic clock for skin and blood cells applied to Hutchinson Gilford progeria syndrome and ex vivo studies. Aging (Albany NY). 2018;10(7):1758–1775. doi: 10.18632/aging.101508

[8] Hannum G, Guinney J, Zhao L, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Molecular Cell. 2013;49(2):359–367. doi: 10.1016/j.molcel.2012.10.016

[9] McEwen LM, O'Donnell KJ, McGill MG, et al. The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. Proc Natl Acad Sci USA. 2020;117(38):23329–23335. doi: 10.1073/pnas.1820843116

[10] Lu AT, Quach A, Wilson JG, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging (Albany NY). 2019;11(2):303–327. doi: 10.18632/aging.101684

[11] Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and healthspan. Aging (Albany NY). 2018;10(4):573–591. doi: 10.18632/aging.101414

[12] Teschendorff AE. A comparison of epigenetic mitotic-like clocks for cancer risk prediction. Genome Med. 2020;12(1):56. doi: 10.1186/s13073-020-00752-3

[13] Belsky DW, Caspi A, Corcoran DL, et al. DunedinPACE, a DNA methylation biomarker of the pace of aging. In: Deelen J, Tyler JK, Suderman M, Deelen J, editors. Elife. 2022;11:e73420. doi: 10.7554/eLife.73420

[14] Higgins-Chen AT, Thrush KL, Wang Y, et al. A computational solution for bolstering reliability of epigenetic clocks: implications for clinical trials and longitudinal tracking. Nat Aging. 2022;2(7):644–661. doi: 10.1038/s43587-022-00248-2

[15] Liu Z, Leung D, Thrush K, et al. Underlying features of epigenetic aging clocks in vivo and in vitro. Aging Cell. 2020;19(10):e13229. doi: 10.1111/acel.13229

[16] Horvath S, Erhart W, Brosch M, et al. Obesity accelerates epigenetic aging of human liver. Proc Natl Acad Sci USA. 2014;111(43):15538–15543. doi: 10.1073/pnas.1412759111

[17] Wolf EJ, Zhao X, Hawn SE, et al. Gene expression correlates of advanced epigenetic age and psychopathology in postmortem cortical tissue. Neurobiol Stress. 2021;15:100371. doi: 10.1016/j.ynstr.2021.100371

[18] Song MA, Mori KM, McElroy JP, et al. Accelerated epigenetic age, inflammation, and gene expression in lung: comparisons of smokers and vapers with non-smokers. Clin Epigenetics. 2023;15(1):160. doi: 10.1186/s13148-023-01577-8

[19] Jung AM, Furlong MA, Goodrich JM, et al. Associations between epigenetic age acceleration and microRNA expression among U.S. Firefighters. Epigenetic Insights. 2023;16:25168657231206301. doi: 10.1177/25168657231206301

[20] Babraham bioinformatics - FastQC a quality control tool for high throughput sequence data. [cited 2024 May 31]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

[21] Ewels P, Magnusson M, Lundin S, et al. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047–3048. doi: 10.1093/bioinformatics/btw354

[22] Liao Y, Smyth GK, Shi W. The R package rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Research. 2019;47(8):e47. doi: 10.1093/nar/gkz114

[23] Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11(3):R25. doi: 10.1186/gb-2010-11-3-r25

[24] Holland N, Furlong C, Bastaki M, et al. Paraoxonase polymorphisms, haplotypes, and enzyme activity in latino mothers and newborns. Environ Health Perspect. 2006;114(7):985–991. doi: 10.1289/ehp.8540

[25] Yousefi P, Huen K, Davé V, et al. Sex differences in DNA methylation assessed by 450 K BeadChip in newborns. BMC Genomics. 2015;16(1):911. doi: 10.1186/s12864-015-2034-y

[26] Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive bioconductor package for the analysis of infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363–1369. doi: 10.1093/bioinformatics/btu049

[27] Fortin JP, Labbe A, Lemire M, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol. 2014;15(11):503. doi: 10.1186/s13059-014-0503-2

[28] Niu L, Xu Z, Taylor JA. RCP: a novel probe design bias correction method for illumina methylation BeadChip. Bioinformatics. 2016;32(17):2659–2663. doi: 10.1093/bioinformatics/btw285

[29] Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118–127. doi: 10.1093/biostatistics/kxj037

[30] Teschendorff AE, Breeze CE, Zheng SC, et al. A comparison of reference-based algorithms for correcting cell-type heterogeneity in epigenome-wide association studies. BMC Bioinformatics. 2017;18(1):105. doi: 10.1186/s12859-017-1511-5

[31] Thrush KL, Higgins-Chen AT, Liu Z, et al. R methylCIPHER: a methylation clock investigational package for hypothesis-driven evaluation & research [internet]. bioRxiv. 2022 [cited 2024 May 31]. p. 2022.07.13.499978.

[32] Smyth GL. Linear models for microarray data. In: Gentleman R, Carey V, Huber W, Irizarry R Dudoit S, editors. Bioinformatics and computational biology solutions using R and Bioconductor [internet]. New York, NY: Springer; 2005 [cited 2024 May 31]. p. 397–420. doi: 10.1007/0-387-29362-0_23

[33] Law CW, Chen Y, Shi W, et al. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15(2):R29. doi: 10.1186/gb-2014-15-2-r29

[34] Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innov [Internet]. 2021;2(3):100141. doi: 10.1016/j.xinn.2021.100141

[35] MSigDB gene sets R package. [cited 2024 May 31]. Available from: https://igordot.github.io/msigdbr/

[36] Liberzon A, Birger C, Thorvaldsdóttir H, et al. The molecular signatures database hallmark gene set collection. Cell Syst. 2015;1(6):417–425. doi: 10.1016/j.cels.2015.12.004

[37] Sánchez-Cabo F, Fuster V, Silla-Castro JC, et al. Subclinical atherosclerosis and accelerated epigenetic age mediated by inflammation: a multi-omics study. Eur Heart J. 2023;44(29):2698–2709. doi: 10.1093/eurheartj/ehad361

[38] Khodasevich D, Bozack A, Daredia S, et al. Supplementary materials for "blood transcriptomic associations of epigenetic age in adolescents". 2025 [cited 2025 May 5]; Available from: https://purl.stanford.edu/kr571bm9410

[39] Zhan XJ, Wang R, Kuang XR, et al. Elevated expression of myosin VI contributes to breast cancer progression *via* MAPK/ERK signaling pathway. Cellular Signalling. 2023;106:110633. doi: 10.1016/j.cellsig.2023.110633

[40] Tamaki K, Kamakura M, Nakamichi N, et al. Upregulation of Myo6 expression after traumatic stress in mouse hippocampus. Neuroscience Letters. 2008;433(3):183–187. doi: 10.1016/j.neulet.2007.12.062

[41] Nwanaji-Enwerem JC, Cardenas A, Gao X, et al. Psychological stress and epigenetic aging in older men: the VA normative aging study. Transl Med Aging. 2023;7:66–74. doi: 10.1016/j.tma.2023.06.003

[42] Zannas AS, Arloth J, Carrillo-Roa T, et al. Lifetime stress accelerates epigenetic aging in an urban, African American cohort: relevance of glucocorticoid

signaling. Genome Biol. 2015;16(1):266. doi: 10.1186/s13059-015-0828-5

[43] de Dieuleveult M, Marchal C, Jouinot A, et al. Molecular and clinical relevance of ZBTB38 expression levels in prostate cancer. Cancers (Basel). 2020;12 (5):1106. doi: 10.3390/cancers12051106

[44] Marchal C, de Dieuleveult M, Saint-Ruf C, et al. Depletion of ZBTB38 potentiates the effects of DNA demethylating agents in cancer cells via CDKN1C mRNA up-regulation. Oncogenesis. 2018;7(10):82. doi: 10.1038/s41389-018-0092-0

[45] Chen J, Xing C, Yan L, et al. Transcriptome profiling reveals the role of ZBTB38 knock-down in human neuroblastoma. PeerJ. 2019;7:e6352. doi: 10.7717/peerj.6352

[46] Jing J, Liu J, Wang Y, et al. The role of ZBTB38 in promoting migration and invasive growth of bladder cancer cells. Oncol Rep. 2019;41(3):1980–1990. doi: 10.3892/or.2018.6937

[47] Parsons S, Stevens A, Whatmore A, et al. Role of ZBTB38 genotype and expression in growth and response to recombinant human growth hormone treatment. J Endocr Soc. 2022;6(3):bvac006. doi: 10.1210/jendso/bvac006

[48] Hou Q, Zhao T, Zhang H, et al. Granzyme H induces apoptosis of target tumor cells characterized by DNA fragmentation and Bid-dependent mitochondrial damage. Mol Immunol. 2008;45(4):1044–1055. doi: 10.1016/j.molimm.2007.07.032

[49] Schreeder DM, Pan J, Li FJ, et al. FCRL6 distinguishes mature cytotoxic lymphocytes and is upregulated in patients with B cell chronic lymphocytic leukemia. Eur J Immunol. 2008;38(11):3159–3166. doi: 10.1002/eji.200838516

[50] Zuo H, Gandhi M, Edreira MM, et al. Downregulation of rap1gap through epigenetic silencing and loss of heterozygosity promotes invasion and progression of thyroid tumors. Cancer Res. 2010;70(4):1389. doi: 10.1158/0008-5472.CAN-09-2812

[51] Yang Y, Zhang J, Yan Y, et al. Low expression of Rap1GAP is associated with epithelial-mesenchymal transition (EMT) and poor prognosis in gastric cancer. Oncotarget. 2016;8(5):8057–8068. doi: 10.18632/oncotarget.14074

[52] Puri D, Barry BJ, Engle EC. TUBB3 and KIF21A in neurodevelopment and disease. Front Neurosci. 2023;17:1226181. doi: 10.3389/fnins.2023.1226181

[53] Lucanus AJ, Thike AA, Tan XF, et al. KIF21A regulates breast cancer aggressiveness and is prognostic of patient survival and tumor recurrence. Breast Cancer Res Treat. 2022;191(1):63–75. doi: 10.1007/s10549-021-06426-x

[54] Salemi M, Barone C, Romano C, et al. KIF21A mRNA expression in patients with down syndrome. Neurol Sci. 2013;34(4):569–571. doi: 10.1007/s10072-012-1183-x

[55] Qi X, Liu L. The regulatory effect of lncRNA LINC00943 on the progression of hepatocellular carcinoma and its relationship with clinicopathological

features. Clin Res Hepatol Gastroenterol. 2024;48 (2):102273. doi: 10.1016/j.clinre.2023.102273

[56] Xu Y, Ji T, An N, et al. LINC00943 is correlated with gastric cancer and regulates cancer cell proliferation and chemosensitivity via hsa-miR-101-3p. Int J Clin Oncol. 2021;26(9):1650–1660. doi: 10.1007/s10147-021-01945-5

[57] Shin JH, Moreno-Nieves UY, Zhang LH, et al. AHR regulates NK cell migration via ASB2–mediated ubiquitination of filamin a. Front Immunol. 2021;12:624284. doi: 10.3389/fimmu.2021.624284

[58] Li J, Manickam G, Ray S, et al. Smpd3 expression in both chondrocytes and osteoblasts is required for normal endochondral bone development. Mol Cellular Biol. 2016;36 (17):2282–2299. doi: 10.1128/MCB.01077-15

[59] Sawant A, Chanda D, Isayeva T, et al. Noggin is novel inducer of mesenchymal stem cell adipogenesis. J Biol Chem. 2012;287(15):12241–12249. doi: 10.1074/jbc.M111.293613

[60] Gazzerro E, Gangji V, Canalis E. Bone morphogenetic proteins induce the expression of noggin, which limits their activity in cultured rat osteoblasts. J Clin Invest. 1998;102(12):2106–2114. doi: 10.1172/JCI3459

[61] Koga M, Engberding N, Dikalova AE, et al. The bone morphogenic protein inhibitor, noggin, reduces glycemia and vascular inflammation in db/db mice. Am J Physiol Heart Circ Physiol. 2013;305(5):H747–55. doi: 10.1152/ajpheart.00825.2012

[62] Lemonnier N, Melén E, Jiang Y, et al. A novel whole blood gene expression signature for asthma, dermatitis, and rhinitis multimorbidity in children and adolescents. Allergy. 2020;75(12):3248–3260. doi: 10.1111/all.14314

[63] Plesch E, Chen CC, Butz E, et al. Selective agonist of TRPML2 reveals direct role in chemokine release from innate immune cells. Elife. 2018;7:e39720. doi: 10.7554/eLife.39720

[64] Li D, Bradley T, Cain DW, et al. RAB11FIP5-deficient Mice exhibit Cytokine-related transcriptomic signatures. ImmunoHorizons. 2020;4(11):713–728. doi: 10.4049/immunohorizons.2000088

[65] Feng M, Fang F, Fang T, et al. Sox13 promotes hepatocellular carcinoma metastasis by transcriptionally activating Twist1. Lab Investig J Tech Met Pathol. 2020;100(11):1400–1410. doi: 10.1038/s41374-020-0445-0

[66] Gao T, Jiang B, Zhou Y, et al. SOX13 is a novel prognostic biomarker and associates with immune infiltration in breast cancer. Front Immunol. 2024;15:1369892. doi: 10.3389/fimmu.2024.1369892

[67] Yang J, Zhong X, Gao X, et al. Knockdown of PIK3R6 impedes the onset and advancement of clear cell renal cell carcinoma. Cell Adhes Migr. 2024;18(1):1–12. doi: 10.1080/19336918.2024.2353920

[68] Opstelten R, Suwandi JS, Slot MC, et al. GPA33 is expressed on multiple human blood cell types and

distinguishes CD4+ central memory T cells with and without effector function. Eur J Immunol. 2021;51 (6):1377–1389. doi: 10.1002/eji.202048744

[69] Han L, Zhang L. CCL21/CCR7 axis as a therapeutic target for autoimmune diseases. Int Immunopharmacol. 2023;121:110431. doi: 10.1016/j.intimp.2023.110431

[70] Hong W, Yang B, He Q, et al. New insights of CCR7 signaling in dendritic cell migration and inflammatory diseases. Front Pharmacol. 2022;13:841687. doi: 10. 3389/fphar.2022.841687

[71] Chang OD, Meier HCS, Maguire-Jack K, et al. Childhood maltreatment and longitudinal epigenetic aging: NIMHD social Epigenomics program. JAMA Netw Open. 2024;7(7):e2421877. doi: 10.1001/jamanet workopen.2024.21877

[72] dos Santos Oliveira NC, Katrinli S, de Assis SG, et al. Community and domestic violence are associated with DNA methylation GrimAge acceleration and heart rate variability in adolescents. Eur J Psychotraumatol. 2023;14 (2):2202054. doi: 10.1080/20008066.2023.2202054