

UCSF

UC San Francisco Previously Published Works

Title

Impact of age-related changes in buccal epithelial cells on pediatric epigenetic biomarker research.

Permalink

<https://escholarship.org/uc/item/6f28k5pc>

Journal

Nature Communications, 16(1)

Authors

Merrill, Sarah
Konwar, Chaini
Fatima, Fizza
[et al.](#)

Publication Date

2025-01-12

DOI

10.1038/s41467-025-55909-8

Peer reviewed

Impact of age-related changes in buccal epithelial cells on pediatric epigenetic biomarker research

Received: 4 April 2024

Accepted: 2 January 2025

Published online: 12 January 2025

 Check for updates

Sarah M. Merrill^{1,2,22}, Chaini Konwar^{2,3,4,22}, Fizza Fatima^{2,3,4}, Kristy Dever^{2,3,4}, Julia L. MacIsaac^{2,3,4}, Nicole Letourneau^{5,6}, Gerald F. Giesbrecht^{6,7,8}, Deborah Dewey^{5,7,8}, Gillian England-Mason^{6,7}, Candace R. Lewis⁹, Dennis Wang^{10,11,12}, Ai Ling Teh^{10,11}, Michael J. Meaney^{10,13,14}, Andrea Gonzalez¹⁵, Jennie G. Noll¹⁶, Carolina De Weerth¹⁷, Nicole R. Bush¹⁸, Kieran J. O'Donnell¹⁹, S. Evelyn Stewart^{2,20} & Michael S. Kobor^{2,3,4,14,21} ✉

Cheek swabs, heterogeneous samples consisting primarily of buccal epithelial cells, are widely used in pediatric DNA methylation studies and biomarker creation. However, the decrease in buccal proportion with age in adults remains unexamined in childhood. We analyzed cheek swabs from 4626 typically developing children 2-months to 20-years-old. Estimated buccal proportion declined throughout childhood with both increasing chronological and predicted epigenetic age. However, buccal proportion did not associate with age throughout adolescence. Variability in buccal proportion increased with age through the entire developmental range. These trends held inversely true for neutrophil proportions. Correcting for buccal proportion attenuated the weak association with PedBE age acceleration to non-significance during initial estimation. Notably, correcting for buccal proportion attenuated the association of PedBE age acceleration with obsessive-compulsive disorder and strengthened the association with diurnal cortisol slope. Thus, the age-related change in children's oral cells is a crucial consideration for cell type-sensitive research.

There is increasing awareness of the importance of tissue heterogeneity and differences in cell-type proportions in the field of epigenetics. The flourishing field of human population epigenetics research in particular has almost exclusively utilized samples composed of multiple cell types. Each of these cell-type identities exhibits diverse phenotypic properties and associates with dynamic biological functions. DNA methylation, the most commonly investigated epigenetic mark in humans, plays a pivotal role in the establishment and maintenance of these distinct cell-type identities. With this biological feature of DNA methylation, cellular proportions can be estimated reliably by reference-based bioinformatics methods, as often cell counts cannot be directly measured due to feasibility and cost issues. Thus, accounting for heterogeneity in cell-type

proportions across individuals in mixed tissues has become critical in DNA methylation studies^{1,2}. This is of particular importance in current pediatric DNA methylation research, as this is predominated by heterogeneous peripheral tissue studies. For example, the most commonly used cheek swabs are primarily comprised of a single cell type, i.e., buccal epithelial cells (BECs)^{3,4}, but they also include a mixture of immune cells, mainly neutrophils and lymphocytes^{3,5}. As cheek swabs are obtained painlessly and noninvasively, they are an attractive source of tissue for safely collecting high-quality DNA from children and often employed to study pediatric DNA methylation-based biomarkers. However, there is little consistency in how and whether cell-type proportions are considered in modeling and interpreting these epigenetic tools.

A full list of affiliations appears at the end of the paper. ✉ e-mail: michael.kobor@ubc.ca

Cellular composition has been previously associated with aging, mortality, and DNA methylation tools¹. One study, using microscopic examination, found higher BEC proportions in samples from children compared to adults³. A follow-up study found a negative association between estimated BEC in a combination of cheek swabs and saliva samples with chronological age across the lifespan⁴. This decrease in estimated BEC proportion was especially pronounced in elderly populations. However, it remains unclear whether such correlations exist within pediatric populations, specifically from infancy to mid-childhood to adolescence—sensitive developmental periods for epigenetics research. The cellular proportions in pediatric cheek swab samples may be influenced by an array of complex factors over the course of normal pediatric maturation, which may be either absent or exacerbated in samples from older populations, including but not limited to primary tooth loss and permanent tooth eruption²; oral inflammatory conditions, such as gingivitis⁶ and periodontitis^{3,7}; and orthodontic appliances such as braces and retainers^{8,9}. Of concern in the interest of scientific inquiry, these oral events may also associate with researchers' variables of interest, such as development or socioeconomic status, simultaneously with changes in cell-type heterogeneity—potentially permitting false associations with children's environment and DNA methylation.

In extant DNA methylation literature, early life is associated with greater variability in DNA methylation than the period of adulthood, and epigenetic tools developed in adults are also less precise and have greater variability in pediatric populations than those developed uniquely for children^{10–12}. Thus, the increasing quantity of available pediatric cheek swabs, and subsequently the creation and application of tools for these samples such as the Pediatric-Buccal-Epigenetic (PedBE) clock¹³, necessitates exploration of the influence of BEC heterogeneity on epigenetic age estimation. Epigenetic clocks are tools trained in specific tissues and populations to estimate biological age based on DNA methylation¹⁴. In general, biological aging can be concordant (approximately equivalent), faster (biologically older), or slower (biologically younger) than chronological age¹⁵. These deviations between biological and chronological age, known as epigenetic age acceleration (EAA), can serve as biomarkers of both exposures and health outcomes throughout life, appear to be sensitive to intervention in children^{16,17}, and provide potentially powerful means of investigating molecular maturation^{18,19}. A number of epigenetic clocks have been designed for use in multiple tissues, including cheek swabs, such as the Horvath pan-tissue clock¹⁴ and the skin-blood clock²⁰; however, PedBE provides the most accurate estimation of chronological age in pediatric (0–20 years old) cheek swab and saliva samples¹⁰.

Thus, in the current study, we characterized the change of DNA methylation-estimated BEC proportion measured in cheek swabs from 4626 typically developing children over this developmental period.

Specifically, we investigated the association between estimated BEC proportion and chronological age, both cross-sectionally and longitudinally, in buccal samples from a large pediatric population collated from multiple independent cohorts throughout childhood and into adolescence. We also explored the association of predicted epigenetic age, namely PedBE, with pediatric BEC proportion, as cell-type heterogeneity is known to influence epigenetic age estimation in other tissues. Finally, we demonstrated the significance of accounting for estimated BEC proportion when employing pediatric epigenetic clocks with two exemplars relevant to child health. Overall, this study provides insight into the relations between estimated BEC proportion and children's age, as well as critical considerations for the application of bioinformatics tools to this issue that may improve precision in associations with childhood health phenotypes.

Results

Twelve cohorts of typically developing children were assembled with BEC samples from across pediatric development

Buccal DNA methylation profiles of 4626 (2154 M/2472 F) typically developing individuals ranging in age from 2 months to 20 years old were assembled in silico from 12 independent cohorts (Table 1). The majority of the cohorts (7/12) included in this study were described previously in the initial report of the PedBE clock¹³, and data were deposited in the publicly available Gene Expression Omnibus (GEO) SuperSeries data repository (accession: GSE137503). Data from another pediatric cohort were also downloaded from GEO (accession: GSE147058)²¹. The remaining four cohorts (APrON Neurotox, BEPAC, SEED, and Cohort 12²²) are described in the Supplementary Methods. The cohort of childhood-onset OCD patients, also described in the Supplementary Methods, was not included in the initial investigation of typically developing children and was only employed to examine EAA associations. Approaches implemented in the EpiDISH R package, specifically robust partial correlation (RPC)-based, were used to report estimated BEC proportion as it has been previously shown to correlate with actual cell counts and consistently predicted higher BEC proportions compared to CIBERSORT (CBS) and constrained projection (CP) based estimations across all cohorts (Supplementary Fig. 1)⁴.

Estimated BEC proportion declined with chronological age from infancy to the end of childhood but have no association with age during adolescence

To test whether inferred BEC proportion was associated with chronological age in pediatric samples, we examined the magnitude and direction of correlations between Box-Cox transformed RPC-estimated BEC proportion (as predicted with EpiDISH²³) and chronological age (reported in months) using Pearson's correlation coefficient (r). A negative linear relation was observed between RPC-

Table 1 | Descriptive cohort characteristics of the pediatric populations who provided cheek swab samples

Cohorts	<i>n</i>	% Female	Estimated buccal epithelial proportions range (mean)	Reported age range (months) (mean)	Illumina platform
APrON FetalPro	142	47.89%	0.84–0.99 (0.97)	0.92–3.71 (2.89)	450K
APrON Neurotox	278	46.76%	0.90–0.99 (0.98)	1.90–8.50 (3.267)	850K
BEPAC	662	46.22%	0.50–1 (0.94)	1.81–29.91 (12.50)	850K
GECKO	289	50.87%	0.50–0.97 (0.82)	83.76–155.59 (112.90)	450K
GSE147058	42	47.61%	0.51–0.97 (0.78)	91–116 (103.57)	850K
NDN	96	59.37%	0.64–0.98 (0.50)	60–216 (135.37)	450K
PAWS	191	53.4%	0.59–0.99 (0.88)	56.96–72.28 (64.00)	450K
SEED	316	53.16%	0.53–0.99 (0.88)	36.06–82.19 (55.01)	850K
WISC	174	55.75%	0.60–1 (0.90)	208.33–225.47 (216.90)	450K
GUSTO	1609	49.44%	0.50–0.99 (0.88)	2.77–78.00 (47.37)	850K
BIBO	438	46.80%	0.53–0.98 (0.86)	70.43–181.13 (121.09)	850K
Cohort 12	389	100%	0.50–0.98 (0.78)	138.02–220.41 (181.56)	850K

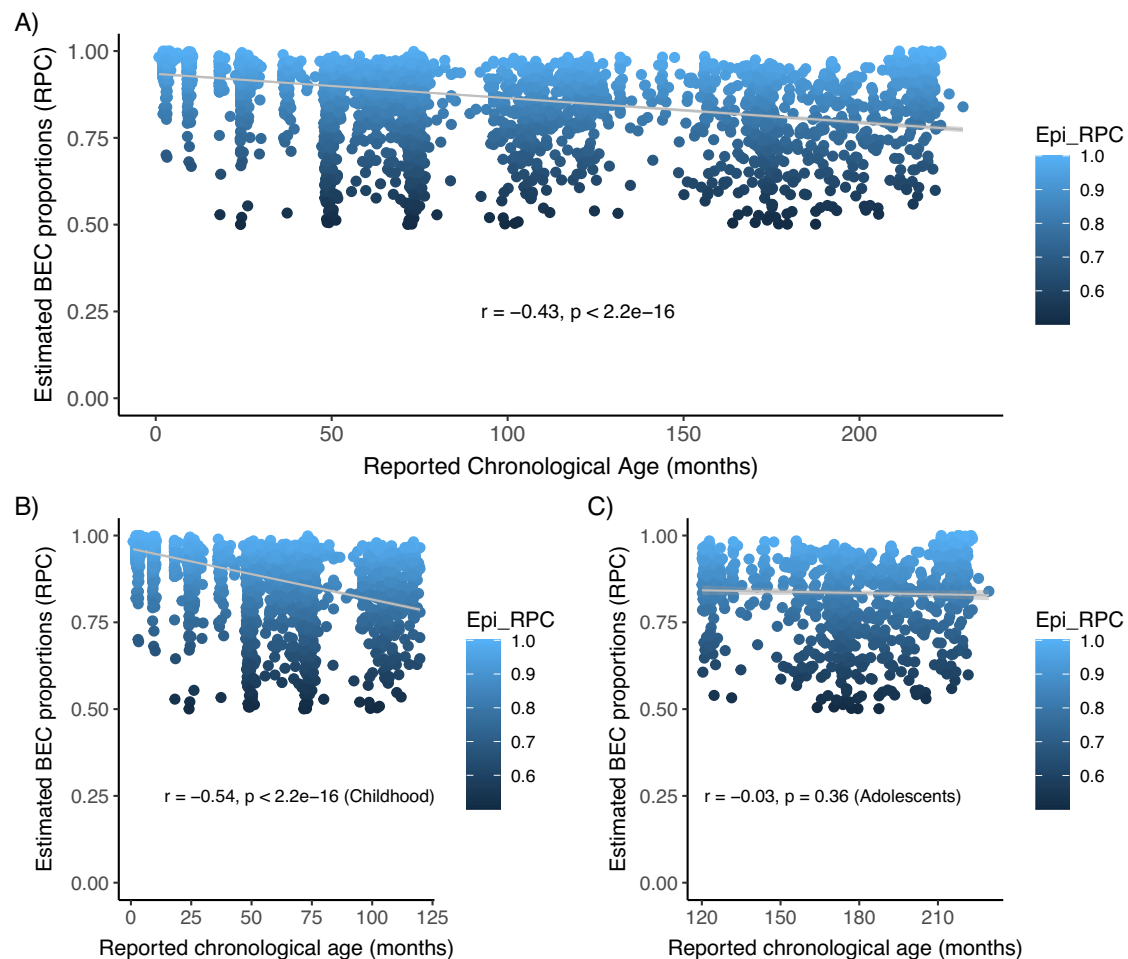


Fig. 1 | Age-related decline in estimated buccal epithelial cell proportions observed in childhood but not adolescence. Significant association of estimated buccal epithelial cell (BEC) proportion with reported chronological age in the pediatric window, using Pearson correlation on Box-Cox transformed estimated BEC proportion. **A** In the scatter plot, estimated BEC proportion estimated by the EpiDISH-RPC method was plotted on the y-axis against chronological age in months on the x-axis. A significant decrease in estimated BEC proportion with reported chronological age was observed in pediatric buccal swabs ($n = 4626$). **B** Pediatric

samples from infancy to childhood, using a cut-off of <120 months (10 years), showed a stronger negative linear relationship between estimated BEC proportion and chronological age (reported in months, $n = 3668$). **C** Pediatric samples in adolescence (>120 months/10 years) exhibited no association between estimated BEC proportion and chronological age (reported in months, $n = 955$). Correlations are significant at multiple-test corrected p value < 0.05. “Source data are provided as a Source Data file”.

predicted BEC proportion and chronological age (Pearson’s on Box-Cox transformed values $r(4624) = -0.43$, $p < 2.2 \times 10^{-16}$) (Fig. 1A). We also observed the same negative relation between chronological age and BEC proportions estimated by the other two algorithms embedded in the EpiDISH R package (CBS and CP) (Supplementary Fig. 2).

However, upon visual inspection we observed the subset of the pediatric samples during adolescence did not conform well to the linear regression line. Thus, we conducted a spline regression analysis to assess if overall the model fit was improved by allowing for different slopes during childhood and adolescence for the regression between chronological age and estimated BEC proportion. The model fit was markedly improved by adding one knot at age 10, the developmentally relevant age of transition from childhood to adolescence (from adjusted $R^2 = 0.17$ with no knot to adjusted $R^2 = 0.25$ with one knot at age 10) (Supplementary Fig. 3)^{24,25}. Because the transition from childhood to adolescence is not at one age, per se, but defined as a developmental period that is generally understood in development to encompass the ages around 10 as well, we tested whether the model fit was equally or more improved by including a knot at age 9, 10, or 11 years^{24,25}. We found no difference in model fit with the exact age of the spline regression knot within this period of transition at age 9, 10, or 11

years (adjusted $R^2 = 0.25$), and as such we selected the midpoint age, 10 years (120 months) as the transition to separate our pediatric samples into two developmentally relevant time points: (i) infancy to childhood (0–9 years) and (ii) adolescence, including preadolescence through the end of adolescence (10–20 years). The resulting model indicated an even stronger negative linear relationship between chronological age and estimated BEC proportion throughout infancy to the end of childhood (Pearson’s on Box-Cox transformed values $r(3666) = -0.54$, $p < 2.2 \times 10^{-16}$) (Fig. 1B), a pattern which was consistent in the majority of the datasets (Supplementary Fig. 4). However, this plateaued into a lack of an association between age and estimated BEC proportion throughout adolescence (Pearson’s on Box-Cox transformed values $r(953) = -0.03$, $p = 0.36$) (Fig. 1C).

In contrast, the estimated neutrophil proportion linearly increased with chronological age in children, and therefore exhibited an inverse association with the observed correlation of estimated BEC proportion and chronological age (Pearson’s on Box-Cox transformed values $r(4624) = 0.23$, $p < 2.2 \times 10^{-16}$) (Supplementary Fig. 5). As with estimated BEC proportion, a better model fit was obtained by including a knot at age 10 (from adjusted $R^2 = 0.11$ with no knots to adjusted $R^2 = 0.18$ with one knot at age 10), with a strong positive association in

childhood and a lack of association during adolescence. Additionally, there were no associations between reported chronological age and estimated cell-type proportions for the other estimated cell types, including eosinophils, monocytes, CD4 T cells, CD8 T cells, B cells, natural killer cells, and fibroblasts (Supplementary Fig. 6).

Estimated BEC proportion was not influenced by DNA methylation platforms or biological factors: sex, breastfeeding status, and mode of delivery

Given that the EpiFiBIC reference (716 CpGs) embedded in the EpiDISH R package was created on data measured on the Illumina Infinium HumanMethylation450 (450K) BeadChip array, application of this reference to datasets processed on the Infinium MethylationEPIC (850K) array warrants investigation, as 36 CpGs from the EpiFiBIC reference panel are missing on the 850K. To explore this, we used the strength of a small, but matched subset of samples (24 matched pairs) from the APrON FetalPro and APrON Neurotox cohorts for which we had obtained DNA methylation measurements using both 450K and 850K arrays. Specifically, we estimated the buccal cell-type proportions, adopting the EpiFiBIC reference but with different sets of reference CpGs:

- Using the EpiFiBIC reference with 716 CpGs on 450K datasets and the same EpiFiBIC reference excluding the 36 missing CpGs on the 850K datasets
- Restricting the EpiFiBIC reference to the common/shared CpGs (680 CpGs) between the arrays and using this set of CpGs on both 450K and 850K datasets

Estimated proportions of BEC by EpiDISH-RPC did not differ based on differing number of CpGs employed in the deconvolution EpiFiBIC reference panel, for (i) ($t(23) = 1.1$, $p_{\text{adj}} > 0.05$) and for (ii) ($t(23) = -0.16$, $p_{\text{adj}} > 0.05$). Overall, EpiDISH-RPC method consistently predicted higher BEC proportions compared to the other methods (Supplementary Fig. 7). Further, CBS estimated BEC proportion showed no significant differences based on the array type, only when the cell-type estimates were calculated by limiting the EpiFiBIC reference to the common/shared CpGs between the arrays. In contrast, in the context of CP-based cell-type algorithm, employing either equality or inequality constraint, we noted significant differences between 450K and 850K calculated BEC proportions, with a large Cohens's d effect size (>1), in both the scenarios of using all the CpGs in the EpiFiBIC reference versus restricting the EpiFiBIC reference to the common/shared CpGs between the arrays. However, the magnitude of difference between the arrays based on CP estimated BEC proportion was relatively larger when differing numbers of CpGs on the EpiFiBIC reference were used to calculate the estimates.

As changes in cell-type proportions are often associated with sex in adult blood samples^{26,27} and potentially with oral-microbiome-relevant early-life variables, such as mode of delivery (C-section/Vaginal) and breastfeeding status (Yes/No)^{15–17}, we explored whether the estimated BEC proportion differed according to these variables. However, no significant associations with sex ($H(1) = 3.38$, $p = 0.06$), mode of delivery ($H(1) = 0.21$, $p = 0.64$), or breastfeeding status ($H(1) = 0.04$, $p = 0.83$) were found (Supplementary Fig. 8).

Estimated BEC proportion was associated with yearly dentist visits, but not daily tooth brushing in 48 months old

To explore the possible association between estimated BEC proportion from cheek swabs and oral health in children, we employed the available parent-reported dental information from the GUSTO cohort when the children were 48 months old. While there was no association with parent-reported daily tooth brushing ($F(2,296) = 0.16$, $p = 0.690$) (Supplementary Fig. 9), there was a small effect size association with yearly dental visits and estimated BEC proportion ($F(2,293) = 4.84$, $p = 0.024$, Cohen's $d = 0.357$). Children who visited the dentist yearly had a higher estimated BEC proportion than children who did not

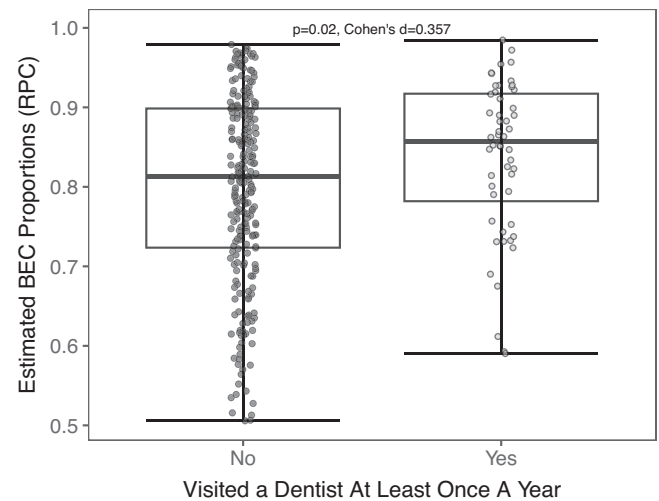


Fig. 2 | Regular dental visits associated with higher estimated buccal epithelial cell proportion in toddlers. Higher estimated buccal epithelial cell (BEC) proportion in children at 48 months ($n = 300$) whose parents reported visiting a dentist at least once in a year (0.84 ± 0.10 , minima = 0.59, maxima = 0.98, 1st Quantile = 0.78, 3rd Quantile = 0.91) compared to children who did not (0.80 ± 0.12 , minima = 0.50, maxima = 0.98, 1st Quantile = 0.72, 3rd Quantile = 0.90). Box plots of estimated BEC proportion by EpiDISH-RPC method were plotted on the y-axis against the categorical Yes/No response of yearly dental visits showed significantly higher BEC proportions, using Kruskal–Wallis test, in children who visited the dentist at least once a year. This difference was robust to differences in exact age at sample collection, sex, and maternally reported racial identity. “Source data are provided as a Source Data file”.

(Fig. 2). Those parents who reported fear of the dentist as their primary reason and lack of financial resources as a secondary reason for not maintaining yearly dental visits exhibited the lowest average BEC proportion, however we note that the sample sizes were small in these categories (Supplementary Fig. 10).

Estimated BEC proportion exhibited more variability with age in pediatric populations

In addition to the decrease in the predicted BEC proportion of pediatric cheek swabs, we found an increase in the range of estimated BEC proportion with increasing age among the youngest children in our sample. Therefore, we hypothesized that the variability of these BEC proportions increases over time in young children, and to test this, we estimated BEC proportion with age using data from three longitudinal pediatric cohorts with repeated sampling of the same children. As the time points were distinct between the cohorts (BEPAC: 2 and 24 months, $n = 244$; SEED: 48, 60, and 72 months, $n = 64$, GUSTO: 3, 9, and 48 months, $n = 114$, BIBO: 72 months, 120 months, and 168 months, $n = 121$), they were analyzed independently. For BEPAC, variance of BEC proportion was significantly higher at 24 months (0.0057) than at 2 months (0.0003) (med $\chi^2(1) = 132.99$, $p < 2.2 \times 10^{-16}$) (Fig. 3A). For SEED, variance was significantly higher at 72 months (0.0128) than 60 months (0.0051) (med $\chi^2(1) = 15.61$, $p = 7.77 \times 10^{-5}$) and 48 months (0.0055) (med $\chi^2(1) = 18.92$, $p = 1.35 \times 10^{-5}$) (Fig. 3B). For GUSTO, variance of estimated BEC proportion was significantly higher at 48 months (0.0119) than at 9 months (0.0018) (med $\chi^2(1) = 72.42$, $p < 2.2 \times 10^{-16}$) and 3 months (0.0006) (med $\chi^2(1) = 94.61$, $p < 2.2 \times 10^{-16}$) (Fig. 3C). For BIBO, variance of estimated BEC proportion were higher at 168 months (0.009) compared to 72 months (0.004) (med $\chi^2(1) = 9.92$, $p = 0.001$) and 120 months (0.005) (med $\chi^2(1) = 15.74$, $p = 7.264 \times 10^{-5}$) (Fig. 3D). Given the differing associations found for estimated BEC proportion and chronological age in childhood and adolescence, we also assessed if the variability of estimated BEC proportion differed between the two developmental stages. Indeed,

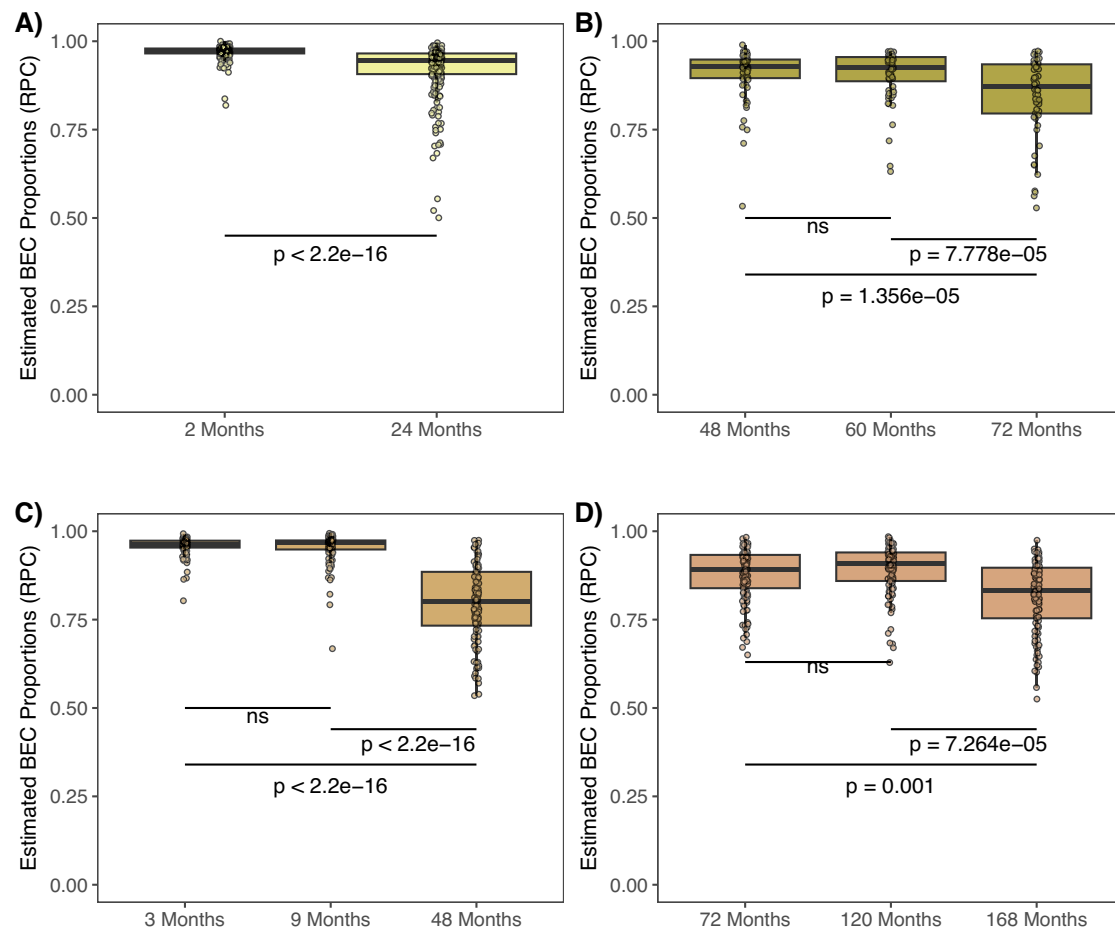


Fig. 3 | Variability of estimated buccal epithelial cell proportion increased across pediatric development in longitudinal cohorts. Increased variability of the estimated buccal epithelial cell (BEC) proportion over time across individuals in four longitudinal pediatric cohorts, as indicated by the different colors.

A In the boxplots, estimated BEC proportion by EpiDISH-RPC method were plotted on the y-axis against the BEPAC cohort samples ($n = 488$) corresponding to the two time points (2 months, 0.97 ± 0.02 , minima = 0.82, maxima = 1.00, 1st Quantile = 0.96, 3rd Quantile = 0.98; 24 months, 0.92 ± 0.08 , minima = 0.50, maxima = 0.99, 1st Quantile = 0.90, 3rd Quantile = 0.96). Variance of BEC proportion was significantly higher at 24 months compared to samples at 2 months. **B** In the boxplots, estimated BEC proportion by EpiDISH-RPC method were plotted on the y-axis against the SEED cohort samples ($n = 192$) corresponding to the three time points (48 months, 0.91 ± 0.07 , minima = 0.53, maxima = 0.99, 1st Quantile = 0.89, 3rd Quantile = 0.95; 60 months, 0.90 ± 0.07 , minima = 0.63, maxima = 0.97, 1st Quantile = 0.89, 3rd Quantile = 0.95; and 72 months, 0.85 ± 0.11 , minima = 0.53, maxima = 0.97, 1st Quantile = 0.79, 3rd Quantile = 0.93). Variance of estimated BEC proportion was significantly higher at 72 months compared to 60 months and at 48 months. **C** In

the boxplots, estimated BEC proportion by EpiDISH-RPC method were plotted on the y-axis against the GUSTO cohort samples ($n = 342$) corresponding to the three time points (3 months, 0.96 ± 0.03 , minima = 0.80, maxima = 0.99, 1st Quantile = 0.95, 3rd Quantile = 0.97; 9 months, 0.95 ± 0.04 , minima = 0.66, maxima = 0.99, 1st Quantile = 0.94, 3rd Quantile = 0.97; and 48 months, 0.79 ± 0.11 , minima = 0.53, maxima = 0.97, 1st Quantile = 0.73, 3rd Quantile = 0.88). Variance of estimated BEC proportion was significantly highest at 48 months compared to 3 months and 9 months. **D** In the boxplots, estimated BEC proportion by EpiDISH-RPC method were plotted on the y-axis against the BIBO cohort samples ($n = 363$) corresponding to the three time points (72 months, 0.88 ± 0.07 , minima = 0.65, maxima = 0.98, 1st Quantile = 0.83, 3rd Quantile = 0.93; 120 months, 0.89 ± 0.07 , minima = 0.63, maxima = 0.98, 1st Quantile = 0.86, 3rd Quantile = 0.94; and 168 months, 0.82 ± 0.10 , minima = 0.52, maxima = 0.97, 1st Quantile = 0.75, 3rd Quantile = 0.90). Variance of estimated BEC proportion was significantly highest at 168 months compared to 72 months and 120 months. Fligner–Killeen test of homogeneity of variance was used to compare variability. “Source data are provided as a Source Data file”.

mirroring the longitudinal analysis, children exhibited less variability (0.009) in their estimated BEC proportion overall than adolescents did (0.02) (med χ^2 (1) = 90.038, $p < 2.2e^{-16}$).

With this overall increase in estimated BEC variability across individuals with increasing age, we next explored whether the same individuals could be estimated reliably to have the highest and lowest BEC proportions over time utilizing data from the SEED, GUSTO, and BIBO cohorts available at three time points. Although within a relatively small temporal window, there was no consistent pattern of relative BEC proportion prediction within the sample. That is, the individuals with the lowest or highest estimated BEC proportion in SEED, GUSTO, and BIBO were not the same across the investigated time points. Further, the correlations in SEED quantified by Kendall’s τ_b (τ_b (62) = 0.09, $p = 0.29$), in GUSTO quantified by Kendall’s τ_b

(112) = 0.01, $p = 0.9$) and in BIBO quantified by Kendall’s τ_b (τ_b (119) = 0.09, $p = 0.14$) were the lowest when comparing the estimated BEC proportion for the same individual at the earliest and latest time points measured (i.e., in SEED 48 and 72 months, respectively (Fig. 4A), in GUSTO 3 and 48 months, respectively (Fig. 4B) and in BIBO 72 and 168 months, respectively (Fig. 4C)). Overall, these findings illustrate increasing variability in estimated BEC proportion with increasing age and the dynamism BEC composition during early development.

Estimated BEC proportion declined with increasing predicted epigenetic age in the period from infancy to childhood

Given that estimated BEC proportion declined with chronological age, specifically from infancy to the end of childhood, we hypothesized that this association may be reflected in epigenetic ages calculated using

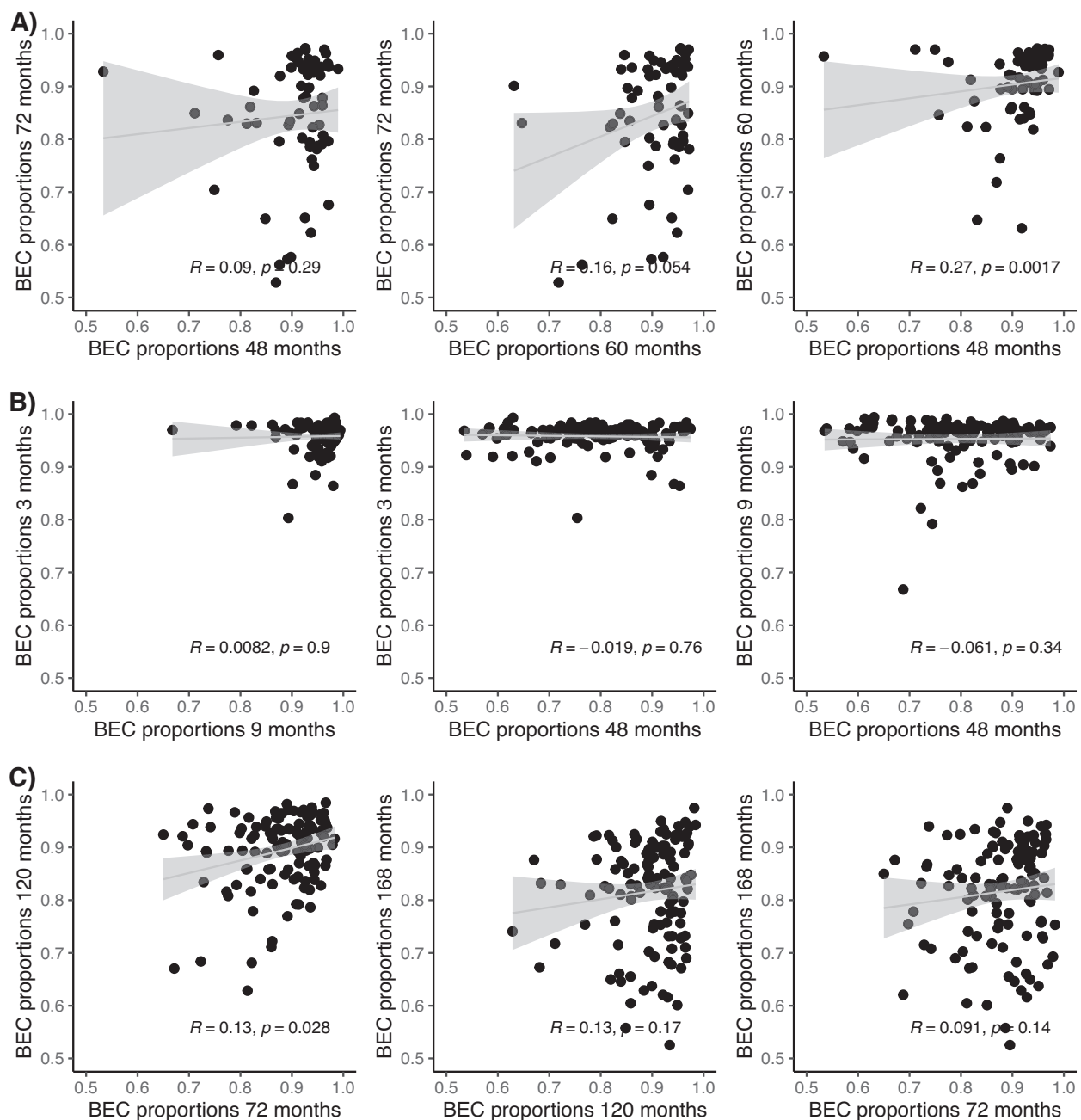


Fig. 4 | Estimated buccal epithelial cell proportions were inconsistent over time within the same children. Relative buccal epithelial cell (BEC) proportion lacked consistency over time within the same individuals compared across three time points (48 months, 0.90 ± 0.08 ; 60 months, 0.90 ± 0.07 ; and 72 months, 0.84 ± 0.11 ; $n = 64 \times 3$) in the SEED cohort (A), in GUSTO cohort (3 months, 0.96 ± 0.03 ; 9 months, 0.95 ± 0.04 ; and 48 months, 0.80 ± 0.10 ; $n = 114 \times 3$) (B) and in BIBO cohort (72 months, 0.88 ± 0.07 ; 120 months, 0.89 ± 0.06 ; and 168 months, 0.82 ± 0.10 ; $n = 121 \times 3$) (C). For example, a child with the highest predicted BEC proportion at 48 months was not more likely than another child to have the highest

predicted BEC proportion at 72 months. To observe the intra-individual variability in estimated BEC proportion over time, scatter plots of estimated BEC proportion by EpiDISH-RPC method for the same samples were graphed across time points. Correlations were determined using Kendall's tau, and this correlation, while not significant, also decreased over time, with the correlation between estimated BEC proportion for the same individuals being the least correlated between 48 and 72 months for SEED cohort, between 9 and 48 months for GUSTO cohort, and between 72 and 168 months for BIBO cohort. Estimated BEC proportions are presented as mean \pm SD. "Source data are provided as a Source Data file".

tools such as the PedBE clock¹³, Horvath pan-tissue clock¹⁴, and skin-blood clock²⁰. For this analysis, we included only cohorts ($n = 3587$) that were not used in training the PedBE clock¹³ to avoid potentially biasing the results. As expected, we observed the same negative association of reference-estimated BEC proportion with PedBE-estimated age over time in children (Pearson's on Box-Cox transformed values $r(3585) = -0.53, p < 2.2 \times 10^{-16}$) (Fig. 5A); the association was stronger in the pediatric samples in the childhood period (<120

months /10 years) (Pearson's on Box-Cox transformed values $r(2942) = -0.61, p < 2.2 \times 10^{-16}$) compared to the samples in the adolescence period (>120 months/10 years), which showed no association between estimated BEC proportion and estimated pediatric age (Pearson's on Box-Cox transformed values $r(641) = -0.03, p = 0.41$) (Supplementary Fig. 1). Although PedBE is the only epigenetic clock designed specifically for pediatric cheek swabs, we also tested and observed the same negative association in two other relevant and

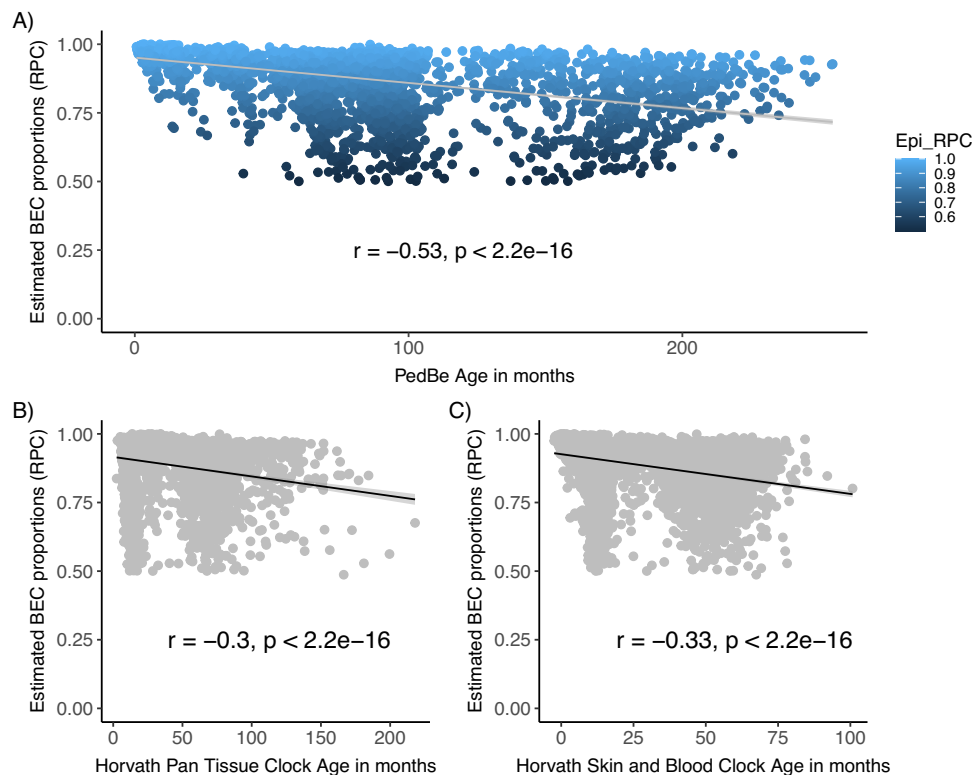


Fig. 5 | Estimated buccal epithelial cell proportions decreased with increased epigenetic age across pediatric samples. A strong significant decrease in estimated buccal epithelial cell (BEC) proportion with epigenetic age was observed in pediatric buccal samples ($n = 3587$), using Pearson correlation on Box-Cox transformed estimated BEC proportion. **A** Scatter plot of estimated BEC proportion (0.88 ± 0.10) estimated by EpiDISH-RPC method was plotted on the y-axis and showed a strong significant decline with PedBE clock predicted age (x-axis) in

pediatric buccal samples. **B** Scatter plot of estimated BEC proportion (0.88 ± 0.10) by EpiDISH-RPC method on the y-axis exhibited a significant decline with Horvath clock predicted age (x-axis) in pediatric buccal samples. **C** Scatter plot of estimated BEC proportion (0.88 ± 0.10) by EpiDISH-RPC method was represented on the y-axis and displayed a significant decrease with Horvath skin-blood predicted age (x-axis) in pediatric buccal samples. “Source data are provided as a Source Data file”.

tissue-appropriate epigenetic clocks, i.e., the Horvath pan-tissue (Pearson’s on Box-Cox transformed values $r(3585) = -0.30$, $p < 2.2 \times 10^{-16}$) (Fig. 5B) and skin-blood epigenetic clocks (Pearson’s on Box-Cox transformed values $r(3585) = -0.33$, $p < 2.2 \times 10^{-16}$) (Fig. 5C). In both childhood and adolescence samples, the epigenetic ages estimated from either Horvath’s pan-tissue (childhood: $r(2942) = -0.54$, $p < 2.2 \times 10^{-16}$; adolescence: $r(641) = -0.47$, $p < 2.2 \times 10^{-16}$) or skin-blood clock (childhood: $r(2942) = -0.52$, $p < 2.2 \times 10^{-16}$; adolescence: $r(641) = -0.22$, $p < 2.2 \times 10^{-16}$) revealed a negative association with estimated BEC proportion; however, the strength of this association was stronger in childhood.

Developmental period-specific statistical suppression effects of estimated BEC proportion were observed on the chronological age prediction of PedBE age

To investigate the contribution of estimated BEC proportion to PedBE age prediction by chronological age in childhood and adolescence, we examined models with and without the inclusion of estimated BEC proportion as a covariate. In childhood samples (<120 months/10 years), we fit the regression model of chronological age and estimated BEC proportion on PedBE age, which was statistically significant, $F(2,640) = 54,350$, $p < 2.2 \times 10^{-16}$, adjusted $R^2 = 0.97$. Next, removing the estimated BEC proportion from the regression model to leave chronological age as the sole predictor of PedBE age, the model was still statistically significant, $F(1,641) = 91,110$, $p < 2.2 \times 10^{-16}$, though with a slightly lower adjusted $R^2 = 0.96$. We also fit these two regression models on the adolescent samples (>120 months/10 years). While the regression model of chronological age and estimated BEC proportion predicting PedBE age was statistically significant, $F(2,640) = 813.4$,

$p < 2.2 \times 10^{-16}$, $R^2 = 0.71$, removing the estimated BEC proportion from the model and leaving chronological age as the sole predictor of PedBE age, while still statistically significant, $F(1,641) = 1097$, $p < 2.2 \times 10^{-16}$ yielded a substantially lower adjusted $R^2 = 0.63$.

When examining the model β coefficients, as expected, the coefficient for chronological age was -1, and this held true both in the sole predictor models and the models with estimated BEC proportion as a covariate, as well as in both developmental periods. This model coefficient indicates that for every unit of chronological age, there was a unit increase for PedBE age across all models. However, for samples in the infancy to childhood period, the model β coefficient of estimated BEC proportion as the sole predictor of PedBE age was -220.14 , significantly more negative than when in the model with chronological age (Table 2). Similarly, for the adolescence period samples, the model β coefficient when BEC proportion was the sole predictor of PedBE age was 4.58, much smaller than when in the model with chronological age (Table 2). These substantial differences in model β coefficients for estimated BEC proportion indicated the possibility of statistical suppressor effects in the more complex regression models^{28,29}. Statistical suppression occurs when the addition of a variable (estimated BEC proportion) to a regression model reveals a stronger relationship between the other variable (chronological age) and the outcome (PedBE age) than was evident without the suppressor, despite the suppressor having a negative or non-significant relationship with the outcome^{28,29}. Therefore, we examined the model β coefficients, structure coefficients, and commonality analysis results with the inclusion of estimated BEC proportion for the chronological age prediction of PedBE age in childhood and adolescence (Table 2).

Table 2 | Model indices of developmental period-specific prediction of PedBE age by chronological age and estimated BEC proportion

Predictor	Beta (β) coefficient	Structured coefficient	Commonality Analysis		
			Unique contribution	Shared contribution	Total contribution
Childhood (<120 months)					
Chronological age	1.1942	0.9974	0.6517	0.3170	0.9737
Estimated BEC proportion	-220.1476	-0.5750	0.0049		
Adolescence (>120 months)					
Chronological age	0.9247	0.9378	0.7173	-0.0860	0.7177
Estimated BEC proportion	4.5898	0.0231	0.0864		

Beta model regression coefficients, structural coefficients, representing the correlation between a predictor variable and the predictor-criterion values, and the unique, shared, and total contribution of each predictor from the commonality analysis are presented.

Estimated BEC proportion was not associated with DNA methylation age residuals corrected for chronological age and BEC proportion

We next examined whether the association of the residuals of DNA methylation age regressed onto chronological age correlated with estimated BEC proportion. These residuals are commonly used in applications of epigenetic clocks and are referred to as EAA, representing the predicted concordance between estimated epigenetic age and chronological age per individual^{15,30}. Therefore, residuals were obtained from two linear models:

- pediatric epigenetic age regressed only onto chronological age (beta coefficient for age, $b = 1.02$, $p < 2.2 \times 10^{-16}$)
- pediatric epigenetic age regressed onto both chronological age and estimated BEC proportion (beta coefficient for estimated BEC proportion, $b = -14.6$, $p < 2.2 \times 10^{-16}$)

As expected, residuals from the first model, epigenetic age regressed onto only chronological age, showed a moderate correlation with estimated BEC proportion (Pearson's on Box-Cox transformed values $r(3585) = -0.20$, $p < 2.2 \times 10^{-16}$) (Fig. 6A). However, residuals from the second model, epigenetic age regressed onto both chronological age and estimated BEC proportion, exhibited no association with the estimated BEC proportion (Pearson's on Box-Cox transformed values $r(3585) = -0.00$, $p = 1$) (Fig. 6B).

Further, we also conducted this analysis using age 10 years/120 months as the developmentally relevant age of transition from childhood to adolescence. Residuals were separately calculated for samples in each of the two developmental periods: childhood and adolescence. As expected, residuals extracted from the model when epigenetic age was regressed onto both chronological age and estimated BEC proportions showed no association with estimated BEC proportion, specifically at both developmental periods (childhood, $r(2942) = -0.00$, $p = 1$; adolescence, $r(641) = 0.00$, $p = 1$).

Including DNA methylation-estimated cell-type proportions in PedBE EAA calculation either strengthened or attenuated associations with childhood health-related variables

Accounting for estimated BEC proportion when calculating PedBE EAA strengthened the association with daily cortisol slope in monozygotic twins. As a proof-of-concept, we leveraged a unique publicly available cohort (GES147058), which although small with 22 monozygotic twin pairs ($n = 44$ children), had measured daily salivary cortisol concurrently with buccal DNA methylation samples, enabling us to examine the potential importance of accounting for estimated BEC proportion when predicting PedBE EAA in the association with a relevant biological variable. Again, residuals were obtained from the two linear models described above to calculate PedBE EAA; one with and the other without accounting for estimated BEC proportion (Table 3).

While both calculations of PedBE EAA were positively associated with daily cortisol slope steepness, an indicator of HPA axis function

and a potential, albeit nuanced, biomarker for physical and mental health conditions³¹⁻³³, the model with PedBE EAA accounting for estimated BEC proportion was more appropriate, as indicated by lower information criterion measures (Table 3). In addition, there was a stronger correlation, as evidenced by a larger R^2 , between PedBE EAA and daily cortisol slope when accounting for estimated BEC variability, even in monozygotic twins. There was a lower correlation between model residuals and estimated BEC proportion when PedBE was calculated with versus without accounting for estimated BEC proportion (Pearson's on Box-Cox transformed values $r(39) = 0.03$, $p = 0.82$ vs. $r(39) = 0.14$, $p = 0.37$, respectively) (Supplementary Fig. 12).

Accounting for estimated BEC proportion when calculating PedBE EAA attenuated the association with OCD diagnosis. We performed a second proof-of-concept analysis utilizing a cohort including individuals diagnosed with childhood-onset OCD ($n = 28$) and controls ($n = 31$) to test if estimated BEC proportion influences the association of PedBE EAA with a child health condition. As described above, we calculated PedBE EAA from residuals extracted from two linear models and, as expected, epigenetic age regressed onto both chronological age and estimated BEC proportion showed no correlation with the estimated BEC proportion (Pearson's on Box-Cox transformed values $r(57) = 0$, $p = 0.99$) (Supplementary Fig. 13). We also confirmed no significant association ($r(57) = 0.02$, $p = 0.54$) of chronological age on the residuals of the model fitted on PedBE EAA regressed only onto chronological age. We compared and fitted two ANCOVA models on PedBE EAA from residuals accounting for chronological age and for both chronological age and estimated BEC proportion. Although the OCD group showed increased PedBE EAA compared to the controls in both models (Fig. 7), the model fitted on PedBE EAA accounting for both chronological age and estimated BEC proportion showed an attenuated association, as demonstrated by the higher although still significant p value ($F(1,37) = 12.59$, $p = 0.001$ vs. $F(1,37) = 6.29$, $p = 0.016$, respectively). In addition, the model employing PedBE EAA regressed onto age and BEC proportion was slightly more appropriate (AIC = 241.01) than the model with PedBE EAA regressed only onto age (AIC = 245.91).

Discussion

A comprehensive understanding of changes in estimated BEC proportion and its variability over pediatric development is critical for the accurate utilization of DNA methylation and DNA methylation-based tools in pediatric populations. Taking advantage of a large cohort of samples from across the pediatric age range, we observed significant changes with age in BEC and neutrophil proportions, estimated using robust Epi-RPC estimation, specifically within the developmental period of infancy through the end of childhood, which is operationally defined as 10 years old in this cohort^{24,25,34}, but statistically our associations held true at 9 and 11 years as well. However, estimated BEC and neutrophil proportions changed in opposite directions—that is as BEC

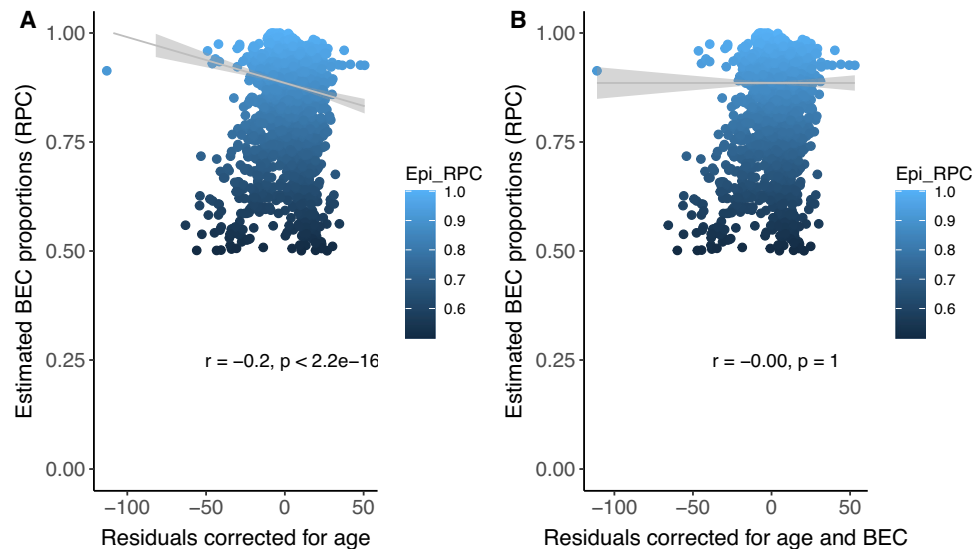


Fig. 6 | Correlation of estimated buccal epithelial cell proportion with epigenetic age acceleration was absent only after correcting residuals for cell type. **A** A weak correlation was reported between estimated buccal epithelial cell (BEC) proportion and DNA methylation residuals corrected for chronological age ($n = 3587$), using Pearson correlation on Box-Cox transformed estimated BEC proportion. In the scatter plot, estimated BEC proportion by EpiDISH-RPC method was plotted on the y-axis and DNA methylation residuals (0.00 ± 10.99) regressed only onto chronological age was on the x-axis. **B** No correlation of estimated

BEC proportion with DNA methylation residuals (-0.00 ± 10.90) corrected for both chronological age and estimated BEC proportion was observed. In the scatter plot, estimated BEC proportion by EpiDISH-RPC method was plotted on the y-axis and DNA methylation residuals regressed onto chronological age and BEC proportions were represented on the x-axis. Estimated BEC proportion and DNA methylation residuals are presented as mean values \pm SD. “Source data are provided as a Source Data file”.

Table 3 | Comparison of two linear mixed effect models which tested for the association of daily cortisol slope steepness with PedBE Epigenetic Age Acceleration (EAA): one without accounting for estimated BEC proportion (left column) and one with estimated BEC proportion correction (right column)

Predictors	PedBE EAA without BEC correction			PedBE EAA with BEC correction		
	Estimates	CI	P value	Estimates	CI	P value
(Intercept)	0.49	-0.20-1.18	0.16	0.52	-0.17-1.20	0.134
Mean cortisol slope	1.48	0.19-2.47	0.024*	1.4	0.28-2.53	0.016
Random effects						
a^2	0.41					
Too	0.24 _{twinpairs}			0.23 _{twinpairs}		
ICC	0.37					
Observations	41			41		
Marginal R^2 / Conditional R^2	0.12/0.44			0.14/0.44		
AIC	104.4			103.2		
BIC	111.2			110		

Associations remain significant after multiple-test correction.

proportion decreased, neutrophils increased. Additionally, both cell types increased in variability with age—both across samples and within longitudinal repeated samples. In contrast, in the adolescent period, from pre- to late adolescence, we observed no association between estimated BEC proportion and age. However, estimated BEC proportion variability was much greater in adolescence compared to the childhood period. BEC proportion decline with chronological age was recently reported across the lifespan and into older ages, yet we discovered a difference in this association within the pediatric age range, pointing to the importance of examining cell-type associations with age within developmental periods, as well as across them⁴. As expected, this association also influenced the interpretation of epigenetic age estimation tools, such as PedBE, in childhood where estimated BEC proportion in children’s cheek swabs declined strongly with predicted pediatric epigenetic age, though, again, not within adolescence.

However, this association was no longer observed if differences in estimated BEC proportion were controlled for in calculating the DNA methylation age residuals, also known as EAA. In adolescence, accounting for estimated BEC proportion improved model fit for EAA likely by eliminating noise during this more variable developmental period. Together, these findings indicated the importance of accounting for variation in BEC proportion in samples collected throughout childhood and adolescence.

From infancy through childhood, there was a significant decrease in estimated BEC proportion in cheek swabs. This may be explained by the oral epithelial morphological changes reported in studies in rodents, which showed increases in size along with reduced density, and decreased proliferation of oral epithelial cells with age^{35,36}, as well as the development of these cells during early-life craniofacial development as the head and mouth grow³⁷. In fact, it has been

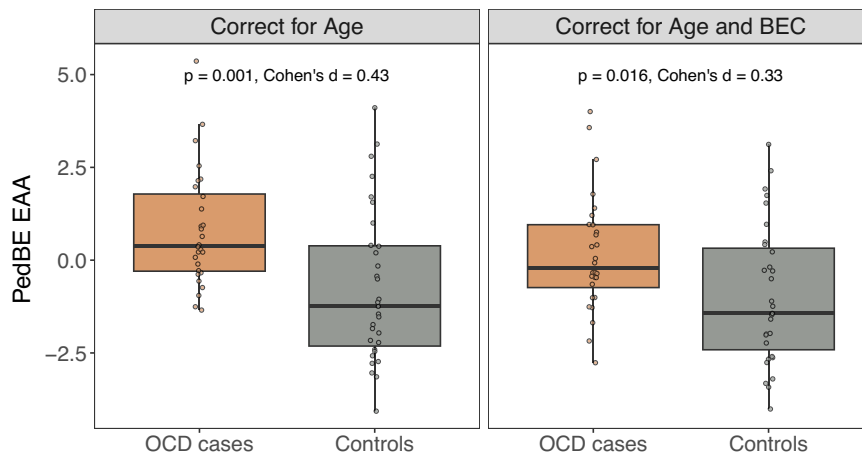


Fig. 7 | Association of PedBE epigenetic age acceleration and childhood-onset OCD diagnosis attenuated after accounting for estimated buccal epithelial cell proportions. A significant but attenuated association of PedBE Epigenetic Age Acceleration (EAA) with childhood-onset obsessive-compulsive disorder (OCD) diagnosis was observed when EAA calculation accounted for both chronological age and estimated BEC proportion in the cohort (ANCOVA). Box plots of PedBE EAA on the y-axis against the OCD-cases ($n = 28$, 0.82 ± 1.57 , minima = -1.34 , maxima = 5.36 , 1st Quantile = -0.30 , 3rd Quantile = 1.78) and controls ($n = 31$, -0.74 ± 2.04 ,

minima = -4.06 , maxima = 4.10 , 1st Quantile = -2.31 , 3rd Quantile = 0.39) showed a significant but attenuated association, when PedBE EAA accounted for both chronological age and estimated BEC proportion, whereas when PedBE EAA corrected only for age a lower p value was obtained when OCD-cases (0.16 ± 1.57 , minima = -2.76 , maxima = 4.00 , 1st Quantile = -0.73 , 3rd Quantile = 0.95) were compared to the controls (-0.95 ± 1.86 , minima = -4.01 , maxima = 3.11 , 1st Quantile = -2.41 , 3rd Quantile = 0.32). “Source data are provided as a Source Data file”.

hypothesized that epidermal growth factor receptor function regulates normal craniofacial development, especially mandibular morphogenesis, partially through inducing activity in oral epithelial cells^{38,39}. However, the decrease in BEC proportion was accompanied by a reciprocal increase in estimated neutrophil proportion, with the estimated proportion of no other immune cell type showing an association with chronological age. The directly proportional nature of this association speaks to the dual functions of the oral epithelium as both a skin- and mucosa-like barrier with immunological properties to exclude microorganisms and fight infections^{40,41}. Neutrophils are white blood cells found primarily in the gingiva, or outer gums, after birth⁴¹, and are the main innate immune cells recruited to protect the oral epithelium in response to bacterial invasion and inflammation. Therefore, the oral neutrophil proportion and activity are closely associated with oral health and the oral microbiome⁴¹. Our data are in support of the hypothesis that this juxtaposition of BEC to neutrophil proportion estimated within a pediatric cheek swab may be informative of the immunological activity within a child’s mouth. For example, 48-month-old children in the GUSTO cohort whose parents brought them to the dentist at least yearly had a higher BEC proportion than those children who did not. This may be an indicator of an association of oral health events or perhaps oral health prevention and BEC proportions in young children. As those children whose parent’s reported fear of the dentist as their primary reason for not visiting had the lowest BEC proportion compared to all other groups reporting not visiting, this may indicate BEC is lowest in the group avoiding in-office dental care regardless of need, rather than a difficulty of access or absence of concerns in this cohort. However, daily oral hygiene at 48 months old was not associated with BEC proportions, so this may indicate an association with either deep cleaning or the necessity of a dental visit as precipitating these oral cell-type differences. Many factors can alter the developing oral microbiome. Though factors known to alter the microbiome, such as breastfeeding status and mode of delivery, were not associated with inter-individual variability in BEC proportion within the 3-month-old cohort, we note that the variability was particularly low in the youngest cohorts. These factors, and other untested exposures, may contribute to variability as these children age and should be a consideration for further research.

Not only did the average estimated BEC proportion decrease and neutrophil proportion increase during childhood (<120 months/10 years), but both cell types also increased in variability with age, including adolescence (>120 months/10 years). Therefore, the youngest infants in these cohorts had the highest BEC proportions, and they were also the least variable. In contrast, adolescents were the most variable. An individual’s estimated BEC proportion in early life did not accurately predict estimated BEC proportion at later time points, with decreasing correlations between larger time steps within the first few years of life. Taken together, these results indicated low intra-individual consistency coupled with robust inter-individual trends toward variability with age. Therefore, it is tempting to speculate most children experience an increase in immunological activity in the oral epithelium with age, but this level of activity, indicated by an increase in estimated neutrophil proportion and subsequent decrease in estimated BEC proportion, varies both within and between individuals with age in early life. While we may not see an overall increase in oral immune cells, specifically neutrophils, throughout adolescence, we do see a greater amount of variability in oral neutrophils (and inversely BECs) within this developmental timeframe. For example, one study of adolescent cheek swabs found a decrease in BEC proportion with a subsequent increase in oral cells containing the bacterium *Actinobacillus actinomycetemcomitans* with the introduction of orthodontic appliances, which may speak to one avenue of variability observed in adolescence⁹. There are many other potential explanations for this variability, including tooth development, tooth loss and eruption, oral hygiene and habits, use of orthodontic or dental apparatuses, diet and environmental exposures, and other oral health concerns^{8,9,42–44}. These events may also differ across developmental stages, such as the importance of tooth loss and eruption in childhood versus oral hygiene later in the pediatric period, but also may point to any event that would increase blood in the mouth transiently. One compelling possibility is the incidence of pediatric gingivitis^{2,5,6,45–47}, an inflammatory condition of the gums reflecting bacterial challenge, which is especially common during pediatric development and can appear within the first few months of life⁴⁸. This inflammation is often transient and superficial, but indicates the increasing presence of gingival *Bacteroides melaninogenicus*, which is mostly absent in young children but then develops to be almost universally present in the mouths

of adolescents, who have the highest prevalence of gingivitis of any age group^{49–51}. The presence and severity of both gingivitis and similar oral health events have been associated with increased oral neutrophil proportions and/or decreased BEC proportions^{3,41,52}. Future research is needed to characterize the oral environment and other features associated with BEC proportion in pediatric populations.

As estimated BEC proportion was strongly associated with chronological age in childhood (<120 months/10 years), it was also associated with predicted epigenetic age based on pediatric cheek swab samples. The PedBE clock¹³, specifically trained in pediatric cheek swab samples, showed an even stronger association between calculated epigenetic age and estimated BEC proportion than chronological age in childhood. This indicates epigenetic tools may be particularly sensitive to age-related changes in cell type. However, there were also significant associations between estimated BEC proportion and predicted epigenetic age using both the Horvath pan-tissue¹⁴ and Horvath skin-blood²⁰ clocks during early life, which were largely trained outside the pediatric age range. This association between cell-type and epigenetic age prediction in all cheek swab-relevant clocks may be due to the computational approaches underlying epigenetic clock creation, which would likely prioritize age-related cell-type differences in selection of informative DNA methylation sites^{13,14}. Similarly, multiple epigenetic clocks, trained exclusively in adults, were shown to be associated with age-related immunological cell-type changes in blood⁵³. However, none of the DNA methylation sites chosen for the PedBE clock are used for cell type prediction in buccal swabs^{23,54}, suggesting they may be influenced by cell type but are not the sites most indicative of cell identity. Further evidence for this were the large model beta coefficients of estimated BEC proportion when predicting PedBE age, even accounting for chronological age in both childhood and adolescence.

The increases in β coefficients when estimated BEC proportion is in the model with chronological age to predict PedBE, as well as examination of the model fit (R^2), structural coefficients, and commonality analysis, suggested statistical suppression effects with estimated BEC proportion across the pediatric age range. Specifically, we observed a classical suppression effect in childhood and a negative suppression effect in adolescence. In childhood, the high common variance measured by the commonality analysis implies that chronological age and BEC proportion explained overlapping components of PedBE age variance. This indicates a classical suppression effect, where estimated BEC proportion enhanced the predictive power of chronological age by accounting for variance that chronological age cannot explain alone. This results in only a small improvement in model fit, as chronological age alone is already predictive of PedBE age and estimated BEC proportion is correlated with chronological age. However, in adolescence, the low common variance coupled with a more substantial model fit improvement, suggests a negative suppression effect, which is notably statistically rare^{28,29,55}. In this context, we observed estimated BEC proportion was slightly suppressing irrelevant variance in chronological age, despite their lack of association in adolescence, which subsequently allowed the model to better explain the variance in PedBE age. Together, these indices suggested that including estimated BEC proportion helps clarify the prediction of PedBE age made by chronological age by eliminating noise. As such, though for differing reasons, we recommend researchers to account for estimated BEC proportion when examining EAA in children and adolescents. This association of epigenetic age with cell type in early life may contribute to the relative accuracy of the PedBE clock in pediatric cheek swabs in comparison to other epigenetic clocks¹⁰. Therefore, the power of epigenetic age as a biomarker is likely related to some extent to the associations between cell-type proportions and the trait of interest⁵⁶. The expected changes in cell-type proportion heterogeneity with age may not be synonymous to, but informative of biological age prediction.

Due to the stronger association between estimated BEC proportion and predicted epigenetic age than chronological age, it was not surprising to find a moderate correlation between EAA and BEC proportion during childhood. However, the addition of differences in estimated BEC proportion as a covariate when calculating EAA led to the disappearance of the correlation between EAA and estimated BEC proportion. These data thus strongly support and expand our previous recommendation when creating the PedBE clock¹ to correct for estimated BEC proportion when calculating PedBE EAA. Given the primary use of PedBE age, or any epigenetic age estimation, to compare these residuals (PedBE EAA) from the regression of chronological age on PedBE age, incorporation of estimated BEC proportion, in calculating EAA can lead to robust and accurate epigenetic age estimates. This, in turn, will likely enhance the validity and replicability of the resulting statistical analyses. Given the primary use of PedBE age, or any epigenetic age estimation, to compare these residuals (PedBE EAA) from the regression of chronological age on PedBE age^{14,56}, incorporation of estimated BEC proportion, in calculating EAA can lead to robust and accurate epigenetic age estimates. This, in turn, will likely enhance the validity and replicability of the resulting statistical analyses. If this is not possible, or not preferred, then we recommend evaluating for associations and heteroskedasticity with estimated BEC proportion within pediatric samples and correct models accordingly. Due to statistical concerns with confounding and heteroskedasticity across the pediatric age range, specifically due to the increased variability observed with increased age, we would also extend this sentiment to all epigenetic age investigations employing pediatric cheek swabs. We also caution researchers in future studies to not assume BEC proportion stability in pediatric cohorts over time, especially when conducting longitudinal research or when combining independent, age-heterogeneous samples.

With the associations of estimated BEC proportion and EAA calculated by all cheek swab-relevant epigenetic clocks, it was important to consider the differences between models when calculating EAA accounting for this cell type. For example, controlling for differences in estimated cell-type proportions in blood was found to attenuate the associations of all-cause mortality with both the Hannum and Horvath pan-tissue clocks¹. Therefore, two proof-of-concept models were considered: a categorical ANCOVA comparison of childhood-onset OCD diagnosis and a linear mixed model comparison of diurnal cortisol slope steepness in monozygotic twin pairs. These comparisons are indicative of two areas of research in which PedBE EAA has been applied, i.e., developmental perspectives on mental health and stress responses^{13,19,57–59}. In both examples, accounting for BEC proportion during PedBE EAA calculation resulted in a more appropriate model, based on both the information criterion and magnitude of the association of model residuals with BEC proportion. However, these model differences may be exaggerated over larger timescales or with increasing variability. Although robustly associated with PedBE EAA, these regressions could be attenuated or amplified when considering estimated BEC proportion depending on variables of interest, cohort characteristics, and statistical assumptions of modeling. These findings demonstrate the importance of accounting for estimated BEC proportion when exploring potential associations with pediatric epigenetic aging mirroring previous findings in other age groups and tissue types^{3,4,53}.

While these analyses had many strengths, especially the large sample size and range of ages from 2 months through 20 years of age, our study also had some limitations. The present study analyses were appropriately powered and the pediatric cheek swab sample sizes were larger than have been analyzed in this developmental window in previous studies; however, a more granular spread across a wider range of ages may reveal a clearer picture of the association between estimated BEC and age. Similarly, these data did not have accompanying genetic information, and given the association between DNA

methylation and the underlying genetic architecture for the epigenetic clocks^{60,61}, analyses would benefit from examining the potential of genotype to affect these cell-type predictions and associations with age. While the cohorts were processed together, and buccal samples from the majority of the cohorts were run by the same facility, the combining of multiple cohorts (450K and 850K) rather than one, large, uniform collection likely introduced some technical variation at this level. However, in the context of DNA methylation-based calculated cell types, we demonstrate that EpiDISH-RPC-estimated BEC proportions are congruent between arrays, irrespective of whether all the CpGs in the EpiFibIC reference are used for BEC estimations or limiting the references to the shared CpGs between the arrays. This finding highlights EpiDISH-RPC as the robust algorithm for BEC estimations, specifically when collating proportions across the two platforms. Although four large longitudinal cohorts were available, the inclusion of more repeated samples with additional time points and using identical methods of collection and storage would have enhanced the robustness of the cell-type proportion variability findings across the pediatric age range, especially in relation to intra-individual variability. Similarly, while oral-microbiome-relevant information was available for the infant cohorts, these analyses would have benefited from direct and more encompassing assessments of the children's oral health and oral microbiome. Finally, although multiple cell-type estimation references were investigated and compared with previous cytology findings in the literature, our analyses did not include true cell counts from cytology or histology although previous research has documented generally high concordance. However, this approach allowed for the examination of the precise subset of probes used for BEC proportion estimation by estimation method and array platform, which revealed array-specific differences with CP method, either using equality or inequality constraint, that should be considered if using this method in the future.

The strong associations of chronological age with both BEC and neutrophil average proportion in childhood, but not in adolescence, and increasing variability with age across the pediatric period must be taken into consideration when using pediatric cheek swabs for any analyses in which cell type is relevant, such as determination of epigenetic age. The model indices suggested accounting for estimated BEC proportion could meaningfully affect PedBE age prediction, and subsequently EAA, in both childhood and adolescence. This would be particularly important if oral differences were confounded with both estimated cell-type proportions—either through correlation or heteroskedasticity—and variables of interest, such as in adolescent populations where parental income can be associated with both the variable of interest and whether the child has orthodontic appliances, which reduce BEC proportions in pediatric cheek swabs^{9,62,63}. Cell-type-sensitive molecular studies based on cheek swab samples should consider the relations of the oral landscape to differences in both amount and variability of estimated cell-type proportions and investigate how the oral environment is related to exposures and phenotypic outcomes. Taken together, these findings provide robust, replicable, and interpretable results from one of the very few tissues that can be obtained safely and noninvasively in pediatric populations.

Methods

Ethics approval and consent to participate

The majority of the datasets were obtained from the publicly available GEO repository and details regarding ethics approval of these cohorts were presented in a previous publication¹. In addition, research related to five cohorts that are not deposited in GEO (SEED, BEPAC, APron Neurotox, OCD, and Cohort 12) presented in this manuscript was performed in compliance with local, state, and national regulations for the ethical treatment of human subjects. Ethics approval information is provided in the Supplementary Methods.

Developmental period definition

Here, we define developmental periods across the span of youth based on the existing American Psychological Association age guidelines and American Academy of Pediatrics, though these are corroborated in other disciplines, such as pediatric nursing, as well. Infancy is defined as 0–2 years, childhood from 2 to 10 years, preadolescence from 10 to 12 years, and adolescence from 12 to 20 years^{24,25,34}. Though the exact age span varies across individuals and is based primarily on physiological maturity milestones, we use these age ranges as our development guides.

DNA methylation data preprocessing

Data preprocessing and subsequent analyses were performed in the R statistical environment (version 4.0.3). Multiple sample quality checks were performed^{64,65}, and samples were excluded if they failed technical control metrics, such as extension, hybridization, specificity, array staining, target removal, and bisulfite conversion. Samples with poor detection p values in >1% of the probes and in which <3 beads contributed to the signal for >1% of the probes were also eliminated. Sex was predicted from the DNA methylation intensities on the sex chromosomes.

Cell-type estimation

Cell types in cheek swabs were estimated using the Smith (2015) method⁵⁴ and HEpiDISH, an iterative hierarchical version of the EpiDISH R package²³. BEC proportions in saliva were predicted based on a buccal dataset (accession: GSE46573) and a FACS-sorted leukocyte dataset (accession: GSE35069) as described by Smith et al.⁵⁴ using the CP approach⁶⁰. The HEpiDISH algorithm embedded in the EpiDISH R package computes sample-specific cellular proportions using three methods: CP (using the default inequality constraint unless otherwise specified)⁶⁰; CIBERSORT (CBS)⁶¹; and RPC⁶⁶. As estimated cell-type proportions are not normally distributed, we used the nonparametric Kendall's tau correlation⁶⁷ to determine if the reference-estimated BEC proportion was correlated between the methods implemented in the EpiDISH R package and the Smith (2015) approach.

To ensure that our cell-type estimations are as comparable and robust across models as possible, we compared BEC proportions estimated by different algorithms using both nonconstrained (EpiDISH-RPC, EpiDISH-CBS) and constrained methods (EpiDISH-CP, inequality constraint), Smith (2015)²³. One CpG (cg00769161) used to infer buccal cell types in the EpiFibIC reference (from EpiDISH) was different from those used in the Smith (2015) method. Eight CpGs (cg01012879, cg05344747, cg25757820, cg26538782, cg02780988, cg10624395, cg1160673, cg16429499) from the HEpiDISH blood reference (188 CpGs) were shared with the Smith (2015) reference (500 CpGs). None of the 94 DNA methylation sites used to calculate PedBE age overlapped with CpGs used for cell-type deconvolution in the EpiDISH and Smith (2015) methods¹³. Although the estimated BEC proportion was comparable across all approaches (Kendall's τ_b (4624) = 0.64 to 0.91, $p < 2.2 \times 10^{-16}$), both EpiDISH-RPC and EpiDISH-CBS consistently predicted higher BEC proportions across all cohorts, were strongly correlated (Kendall's τ_b (4624) = 0.91, $p < 2.2 \times 10^{-16}$), and showed the same correlations when the cohorts were investigated independently (Supplementary Fig. 1). Therefore, all results were presented with predicted cell-type proportions derived from EpiDISH-RPC estimation.

Epigenetic age calculation

Pediatric epigenetic age was calculated using the publicly available PedBE tool (available from <https://github.com/kobor-lab/PublicScripts/blob/master/PedBE.Md>)¹³, which multiplies the DNA methylation beta values of 94 CpG sites by predetermined weights and adds an intercept to estimate epigenetic age. DNA methylation age residuals were extracted from a linear model of pediatric epigenetic age regressed onto chronological age or chronological age and BEC

proportion. As recommended for the PedBE tool, data were background color corrected using the noob R package¹³.

Horvath pan-tissue and Horvath skin-blood epigenetic clock ages were calculated using the online DNA Methylation Age Calculator developed by the clocks' creator (<https://dnamage.genetics.ucla.edu/new>)^{14,20}.

Estimated cell-type proportion regression analysis

Linear regression in stats R package with Box-Cox transformation was used to evaluate the associations between chronological age and RPC-estimated cell-type proportions in pediatric cheek swab samples. Box-Cox transformation using the boxcox function in the mass R package was used to transform skewed estimated cell-type proportions into an approximately normal distribution⁶⁸. Two estimated proportions were tested across all datasets: BECs (right-skewed) and neutrophils (left-skewed). In addition, linear regressions of Box-Cox transformed RPC estimates of BEC proportion and both epigenetic age and EAA residuals were performed. All model assumptions for these analyses were tested and met.

Further, to allow for different slopes during childhood and adolescence for the regression between chronological age and estimated BEC and neutrophil proportions, we performed a spline regression analysis using the ss function in npreg R package by including a knot at the developmentally relevant transition period.

Technical, demographic, and oral health measures with estimated cell-type proportion

To examine whether estimated cell-type proportions differed by the technology of the two most commonly used DNA methylation arrays, we used a Pearson's correlation test on a subset of samples from the APron FetalPro and APron Neurotox cohorts (24 matched pairs) for which we had obtained DNA methylation measurements using both the 450K and 850K. Next, we explored whether the estimated BEC proportion differed according to the sex and oral-microbiome-relevant early-life variables including mode of delivery and breastfeeding status using both 450K and 850K platforms, both represented in these data, using Kruskal–Wallis test. Sex, inferred from DNA methylation intensities of the XX and XY chromosomes, was explored across all the samples but excluding samples that were represented at more than one timepoint in a dataset, while mode of delivery and breastfeeding status were examined in only the APron cohort for which these data were available ($n = 124$).

Additionally, we employed the GUSTO cohort to investigate the potential association of estimated cell-type proportion and oral health. The GUSTO cohort timepoint at 48 months old asked parents if their children brushed their teeth daily and if they took their child to the dentist yearly with yes or no response options. These responses were compared to estimated BEC type proportion in children with DNA methylation data at 48 months ($n = 298$) using an ANCOVA model controlling for age at DNA methylation sample, sex, and maternal reported ethnicity. Of parents who reported they did not take their child to the dentist yearly, they were asked to provide the primary reason why they had not with the available options of no money, no time, transportation difficulties, fear of drills/injections and dentists, the child's teeth did not bother them, and other reasons.

Analysis of estimated cell-type proportion variance in longitudinal cohorts

Four cohorts with repeated longitudinal BEC sampling (BEPAC, SEED, GUSTO, and BIBO) were independently utilized to examine variability of RPC-estimated BEC proportion over time. The BEPAC cohort consisted of 244 child pairs for which DNA methylation data were available at 2 and 24 months. Data at three time points (48, 60, and 72 months) for 64 children were quantified for the SEED cohort, whereas data at 3, 9, and 48 months for 114 children in GUSTO cohort and data at 72, 120,

and 168 months for 121 children in BIBO cohort was used in this analysis. The Fligner–Killeen test of homogeneity of variance⁶⁹ was performed to compare variances for the non-normally distributed RPC-estimated BEC proportion.

Assessment of estimated BEC proportion contribution to PedBE age prediction

To further understand the association of estimated BEC proportion to the prediction of PedBE age by chronological age, we compared the differences in model β coefficients for estimated BEC proportion across the two developmental periods. Upon discovering large discrepancies in estimated BEC proportion model β coefficients when BEC proportion was a covariate and when BEC proportion was a sole predictor, we investigated the possibility of statistical suppression^{28,29,55}. This exploration included the effects of estimated BEC proportion as a covariate to chronological age in predicting PedBE age, specifically: (a) employing model indices including model fit (adjusted R^2) and (b) structure coefficients extracted from the regression models, which were fitted using the stats R package⁷⁰. Structure coefficients reflect the correlation between a predictor variable and the outcome-predicted values and reflect the unique strength of the predictor as specified by the model²⁸. Additionally, we conducted a commonality analysis using the yhat R package⁷¹ to dissect the unique contribution of the independent predictors and the shared contribution of the predictors in explaining the overall variance of the model²⁹.

Estimated cell-type proportion correction when calculating PedBE EAA in an OCD cohort

Previously biobanked cheek swabs from a cohort of children with pediatric-onset OCD were drawn from BC Children's Hospital BioBank. After sample preprocessing, 28 OCD-affected cases and 31 age-matched controls from the Wisconsin Study of Families and Work cohort ranging in age from -7 to 18 years remained. PedBE EAA was calculated using residuals extracted from linear models by regressing chronological age alone onto biological age (PedBE EAA without BEC) and by regressing both chronological age and estimated BEC proportion onto biological age (PedBE EAA with BEC). Subsequently, we fitted two ANCOVA models, on PedBE EAA with BEC and PedBE EAA without BEC, accounting for covariates, including age, predicted sex, and batch variables.

Estimated cell-type proportion correction when calculating PedBE EAA in cortisol diurnal slope in monozygotic twins

In one cohort (accession: GSE147058) of monozygotic twins recruited prospectively for a longitudinal investigation of child health⁷², daily cortisol slope was measured from saliva samples concurrently with BEC DNA methylation samples. The children ranged in age from -7.5 to 9 years (mean = 8.5 years). The methods for saliva collection and daily cortisol slope calculation are described elsewhere²⁷. Slope coefficients were negative, as cortisol declines over the course of the day²⁷. PedBE EAA was calculated as described above for the OCD cohort. Individually, a mixed effect linear model was run with the lmer function in the lme4 R package with a random effect of twin pair, where PedBE EAA with and without BEC predicted daily cortisol slope.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data used in this study are available in the GEO SuperSeries under accession numbers GSE137503 and GSE147058 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE147058>]. However, data from five cohorts (SEED, BEPAC, APron Neurotox, OCD, and Cohort 12) are not yet publicly available because they include sensitive information

related to children's epigenomics, mental health, trauma exposure, and therapy. To request access to this data, kindly contact the corresponding author and expect a response within 2 weeks. Source data are provided with this paper.

Code availability

Code for this manuscript is available https://github.com/kobor-lab/PedBE_BEC/blob/main/Script.

References

- Zhang, Q. et al. Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. *Genome Med.* **11**, 54 (2019).
- Bimstein, E. & Matsson, L. Growth and development considerations in the diagnosis of gingivitis and periodontitis in children. *Pediatr. Dent.* **21**, 186–191 (1999).
- Theda, C. et al. Quantitation of the cellular content of saliva and buccal swab samples. *Sci. Rep.* **8**, 1–8 (2018).
- Wong, Y. T. et al. A comparison of epithelial cell content of oral samples estimated using cytology and DNA methylation. *Epigenetics* **17**, 327–334 (2022).
- Williams, D. W. et al. Human oral mucosa cell atlas reveals a stromal-neutrophil axis regulating tissue immunity. *Cell* **184**, 4090–4104.e15 (2021).
- Khan, S., Khan, A. U. & Hasan, S. Genotoxic assessment of chlorhexidine mouthwash on exfoliated buccal epithelial cells in chronic gingivitis patients. *J. Indian Soc. Periodontol.* **20**, 584–591 (2016).
- Tadin, A., Gavic, L., Roguljic, M., Jerkovic, D. & Zeljezic, D. Nuclear morphological changes in gingival epithelial cells of patients with periodontitis. *Clin. Oral. Investig.* **23**, 3749–3757 (2019).
- Gursoy, U. K. et al. The role of nickel accumulation and epithelial cell proliferation in orthodontic treatment-induced gingival overgrowth. *Eur. J. Orthod.* **29**, 555–558 (2007).
- Leung, N. M., Chen, R. & Rudney, J. D. Oral bacteria in plaque and invading buccal cells of young orthodontic patients. *Am. J. Orthod. Dentofac. Orthop.* **130**, 698.e11–698.e18 (2006).
- Kling, T., Wenger, A. & Carén, H. DNA methylation-based age estimation in pediatric healthy tissues and brain tumors. *Aging* **12**, 21037–21056 (2020).
- Pérez, R. F. et al. Longitudinal genome-wide DNA methylation analysis uncovers persistent early-life DNA methylation changes. *J. Transl. Med.* **17**, 1–16 (2019).
- Alisch, R. S. et al. Age-associated DNA methylation in pediatric populations. *Genome Res.* **22**, 623–632 (2012).
- McEwen, L. M. et al. The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. *Proc. Natl. Acad. Sci. USA* **117**, 23329–23335 (2020).
- Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, 1–20 (2013).
- Jones, M. J., Goodman, S. J. & Kobor, M. S. DNA methylation and healthy human aging. *Aging Cell* **14**, 924–932 (2015).
- Sullivan, A. D. W. et al. Intervening after trauma: child–parent psychotherapy treatment is associated with lower pediatric epigenetic age acceleration. *Psychol. Sci.* **35**, 1062–1073 (2024).
- Sullivan, A. D. W. et al. Parenting practices may buffer the impact of adversity on epigenetic age acceleration among young children with developmental delays. *Psychol. Sci.* **34**, 1173–1185 (2023).
- Oblak, L., van der Zaag, J., Higgins-Chen, A. T., Levine, M. E. & Boks, M. P. A systematic review of biological, social and environmental factors associated with epigenetic clock acceleration. *Ageing Res. Rev.* **69**, 101348 (2021).
- Palma-Gudiel, H., Fañanás, L., Horvath, S. & Zannas, A. S. Chapter Five—Psychosocial stress and epigenetic aging. in *International Review of Neurobiology* Vol. 150 (eds. Clow, A. & Smyth, N.) 107–128 (Academic Press, 2020).
- Teschendorff, A. E. Epigenetic clocks galore: a new improved clock predicts age-acceleration in Hutchinson Gilford Progeria Syndrome patients. *Aging* **10**, 1799–1800 (2018).
- Lewis, C. R. et al. Harsh parenting predicts novel HPA receptor gene methylation and NR3C1 methylation predicts cortisol daily slope in middle childhood. *Cell Mol. Neurobiol.* **41**, 783–793 (2021).
- Noll, J. G. et al. An observational study of Internet behaviours for adolescent females following sexual abuse. *Nat. Hum. Behav.* **6**, 74–87 (2022).
- Zheng, S. C. et al. EpiDISH web server: epigenetic dissection of intra-sample-heterogeneity with online GUI. *Bioinformatics* **36**, 1950–1951 (2019).
- American Psychological Association. Childhood. In *APA Dictionary of Psychology*. <https://dictionary.apa.org/childhood> (2015).
- Hatfield, N. T. *Broadribb's Introductory Pediatric Nursing* (Lippincott Williams & Wilkins, 2007).
- Bongen, E. et al. Sex differences in the blood transcriptome identify robust changes in immune cell proportions with aging and influenza infection. *Cell Rep.* **29**, 1961–1973.e4 (2019).
- Engelbrecht, H.-R. et al. Sex differences in epigenetic age in Mediterranean high longevity regions. *Front. Aging* **3**, 1007098 (2022).
- Henard, D. H. Suppressor variable effects: toward understanding an elusive data dynamic. <https://eric.ed.gov/?id=ED416215> (1998).
- Kraha, A., Turner, H., Nimon, K., Zientek, L. & Henson, R. Tools to support interpreting multiple regression in the face of multicollinearity. *Front. Psychol.* **3**, 44 (2012).
- Ryan, C. P. “Epigenetic clocks”: theory and applications in human biology. *Am. J. Hum. Biol.* **33**, e23488 (2021).
- Valentino, K., Hibbel, L. C., Speidel, R., Fondren, K. & Ugarte, E. Longitudinal effects of maltreatment, intimate partner violence, and Reminiscing and Emotion Training on children's diurnal cortisol regulation. *Dev. Psychopathol.* **33**, 868–884 (2021).
- Kumari, M. et al. Cortisol secretion and fatigue: Associations in a community based cohort. *Psychoneuroendocrinology* **34**, 1476–1485 (2009).
- Hsiao, F.-H. et al. The self-perceived symptom distress and health-related conditions associated with morning to evening diurnal cortisol patterns in outpatients with major depressive disorder. *Psychoneuroendocrinology* **35**, 503–515 (2010).
- Hardin, A. P. et al. Age limit of pediatrics. *Pediatrics* **140**, e20172151 (2017).
- Elias, W. Y. Age-dependent differential expression of apoptotic markers in rat oral mucosa. *Asian Pac. J. Cancer Prev.* **19**, 3245–3250 (2018).
- Hill, M. W. Influence of age on the morphology and transit time of murine stratified squamous epithelia. *Arch. Oral. Biol.* **33**, 221–229 (1988).
- Chai, Y. & Maxson, R. E. Jr Recent advances in craniofacial morphogenesis. *Dev. Dyn.* **235**, 2353–2375 (2006).
- Miettinen, P. J. et al. Epidermal growth factor receptor function is necessary for normal craniofacial development and palate closure. *Nat. Genet.* **22**, 69–73 (1999).
- van der Zee, E., Jansen, I., Hoeben, K., Beertsen, W. & Everts, V. EGF and IL-1 alpha modulate the release of collagenase, gelatinase and TIMP-1 as well as the release of calcium by rabbit calvarial bone explants. *J. Periodontol. Res.* **33**, 65–72 (1998).
- Abu Eid, R., Sawair, F., Landini, G. & Saku, T. Age and the architecture of oral mucosa. *AGE* **34**, 651–658 (2012).
- Uriarte, S. M., Edmisson, J. S. & Jimenez-Flores, E. Human neutrophils and oral microbiota: a constant tug-of-war between a harmonious and a discordant coexistence. *Immunol. Rev.* **273**, 282–298 (2016).

42. Childs, W. C. III & Gibbons, R. J. Selective modulation of bacterial attachment to oral epithelial cells by enzyme activities associated with poor oral hygiene. *J. Periodontal Res.* **25**, 172–178 (1990).
43. Li, H. et al. The impacts of delivery mode on infant's oral microflora. *Sci. Rep.* **8**, 11938 (2018).
44. Tsang, A. K. L. Teething, teething pain and teething remedies. *Int. Dent. South Afr.* **12**, 48–61 (2010).
45. Colombo, A. V., Da Silva, C. M., Haffajee, A. & Colombo, A. P. V. Identification of intracellular oral species within human crevicular epithelial cells from subjects with chronic periodontitis by fluorescence in situ hybridization. *J. Periodontal Res.* **42**, 236–243 (2007).
46. Davis, G. & Gibbons, R. J. Accessible sialic acid content of oral epithelial cells from healthy and gingivitis subjects. *J. Periodontal Res.* **25**, 250–253 (1990).
47. Vaahtoniemi, L. H., Räisänen, S. & Stenfors, L.-E. Effect of chlorhexidine and toothbrushing on the presence of bacteria on gingival and buccal epithelial cells. *Oral. Microbiol. Immunol.* **9**, 315–317 (1994).
48. Modéer, T. & Wondimu, B. Periodontal disease in children and adolescents. *Dent. Clin. North Am.* **44**, 633–658 (2000).
49. Bailit, H. L., Baldwin, D. C. & Hunt, E. E. The increasing prevalence of gingival *Bacteroides melaninogenicus* with age in children. *Arch. Oral. Biol.* **9**, 435–438 (1964).
50. Kumar, A., Dhanraj, M., Muralidharan, N. P. & Jain, A. Composition of subgingival microbial flora in children. *Drug Invent. Today* **10**, 1272–1275 (2018).
51. Yang, E. Y. et al. Periodontal pathogen detection in gingiva/tooth and tongue flora samples from 18- to 48-month-old children and periodontal status of their mothers. *Oral. Microbiol. Immunol.* **17**, 55–59 (2002).
52. Cortés-Vieyra, R., Rosales, C. & Uribe-Querol, E. Neutrophil functions in periodontal homeostasis. *J. Immunol. Res.* **2016**, e1396106 (2016).
53. Jonkman, T. H. et al. Functional genomics analysis identifies T and NK cell activation as a driver of epigenetic clock progression. *Genome Biol.* **23**, 24 (2022).
54. Smith, A. K. et al. DNA extracted from saliva for methylation studies of psychiatric traits: evidence tissue specificity and relatedness to brain. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **168**, 36–44 (2015).
55. Martínez Gutierrez, N. & Cribbie, R. Incidence and interpretation of statistical suppression in psychological research. *Can. J. Behavioural Sci. Rev. can. des. Sci. du Comport.* **53**, 480–488 (2021).
56. Bell, C. G. et al. DNA methylation aging clocks: challenges and recommendations. *Genome Biol.* **20**, 1–24 (2019).
57. Nishitani, S. et al. Altered epigenetic clock in children exposed to maltreatment. *Psychiatry Clin. Neurosci.* **75**, 110–112 (2021).
58. Clausen, E. S., Binder, A. M. & Non, A. L. Epigenetic age associates with psychosocial stress and resilience in children of Latinx immigrants. *Epigenomics* **13**, 1677–1699 (2021).
59. Dammering, F. et al. The pediatric buccal epigenetic clock identifies significant ageing acceleration in children with internalizing disorder and maltreatment exposure. *Neurobiol. Stress* **15**, 100394 (2021).
60. Houseman, E. A. et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinform.* **13**, 86 (2012).
61. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
62. McDade, T. W. & Koning, S. M. Early origins of socioeconomic inequalities in chronic inflammation: evaluating the contributions of low birth weight and short breastfeeding. *Soc. Sci. Med.* **269**, 113592 (2021).
63. Mejia, G. C. et al. Socioeconomic status, oral health and dental disease in Australia, Canada, New Zealand and the United States. *BMC Oral. Health* **18**, 176 (2018).
64. Konwar, C. et al. Risk-focused differences in molecular processes implicated in SARS-CoV-2 infection: corollaries in DNA methylation and gene expression. *Epigenetics Chromatin* **14**, 1–18 (2021).
65. Merrill, S. M. et al. Associations of peripheral blood DNA methylation and estimated monocyte proportion differences during infancy with toddler attachment style. 1–30. <https://doi.org/10.1080/14616734.2021.1938872> (2021).
66. Teschendorff, A. E., Breeze, C. E., Zheng, S. C. & Beck, S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. *BMC Bioinform.* **18**, 105 (2017).
67. Sen, P. K. Estimates of the regression coefficient based on Kendall's tau. *J. Am. Stat. Assoc.* **63**, 1379–1389 (1968).
68. Atkinson, A. C., Riani, M. & Corbellini, A. The Box–Cox transformation: review and extensions. *Stat. Sci.* **36**, 239–255 (2021).
69. Conover, W. J., Johnson, M. E. & Johnson, M. M. A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics* **23**, 351–361 (1981).
70. stats package—RDocumentation <https://www.rdocumentation.org/packages/stats/versions/3.6.2> (2019).
71. Nimon, K., Oswald, F. & Roberts, J. K. Package 'yhat'. R package version 2.0-4. <https://cran.r-project.org/package=yhat> (2009).
72. Lemery-Chalfant, K. et al. Arizona Twin Project: specificity in risk and resilience for developmental psychopathology and health. *Twin Res. Hum. Genet.* **22**, 681–685 (2019).

Acknowledgements

This work would not have been possible without the extensive work of Drs. James N. Reynolds, Marilyn J. Essex, and W. Thomas Boyce in collecting, curating, managing, and publically depositing the Neuro-DevNet, Wisconsin Study of Families and Work, and GECKO cohorts, respectively. Additionally, we are grateful for the efforts of Drs. Leah D. Doane and Kathryn Lemery-Chalfant in their collection, custody, and public data deposition of the Arizona Twin Project. We also thank Tanya Erb for her administrative assistance in facilitating data transfer between cohorts. We thank the many families without whose participation this work would not be possible. This work was made possible through the funding of multiple agencies. Dr. Michael S. Kobor is the Canada Research Chair in Social Epigenetics and the Edwin S.H. Leong Chair in Healthy Aging - a UBC President's Excellence Chair. Funding information of the GEO datasets utilized in the current study has been described in a previous publication¹³. For the five cohorts that have not been deposited in GEO (SEED, BEPAC, APrON Neurotox, OCD cohorts, and Cohort 12), details on funding agencies are provided in the Supplementary Methods.

Author contributions

Conceptualization: M.S.K., S.M.M., and C.K.; biological data generation: J.L.M., K.D., and M.S.K.; biological data curation and processing: C.K., S.M.M., and F.F.; phenotypic data curation: S.M.M., C.K.; data analysis: S.M.M. and C.K.; interpretation of results: S.M.M. and C.K.; figures/ visualization: C.K. and S.M.M.; manuscript review: S.M.M., C.K., M.S.K., F.F., K.D., J.L.M., N.L., G.F.G., D.D., G.E.M., M.J.M., D.W., A.L.T., N.R.B., C.R.L., J.G.N., K. O'D., C.D.W., A.G., and E.S.S.; cohort owners/custodians: N.L., G.F.G., D.D., M.J.M., D.W., A.L.T., G.E.M., N.R.B., C.R.L., A.G., S.E.S., J.G.N., K. O'D., C.D.W., M.S.K.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-55909-8>.

Correspondence and requests for materials should be addressed to Michael S. Kobor.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

¹Department of Psychology, University of Massachusetts Lowell, Lowell, MA, USA. ²British Columbia Children's Hospital Research Institute, University of British Columbia, Vancouver, BC, Canada. ³Department of Medical Genetics, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada. ⁴Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC, Canada. ⁵Faculty of Nursing and Cumming School of Medicine, Departments of Pediatrics, Community Health Sciences and Psychiatry, University of Calgary, Calgary, AB, Canada. ⁶Owerko Centre, Alberta Children's Hospital Research Institute, University of Calgary, Calgary, AB, Canada. ⁷Department of Paediatrics and Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada. ⁸Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada. ⁹School of Life Sciences, Arizona State University, Tempe, AZ, USA. ¹⁰Institute for Human Development and Potential (IHDP), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore. ¹¹Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore. ¹²National Heart and Lung Institute, Imperial College London, London, UK. ¹³Douglas Mental Health University Institute, McGill University, Montreal, QC, Canada. ¹⁴Program in Child and Brain Development, CIFAR, Toronto, ON, Canada. ¹⁵Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, ON, Canada. ¹⁶Mount Hope Family Center and Department of Psychology, University of Rochester, Rochester, NY, USA. ¹⁷Donders Institute for Brain, Cognition and Behaviour and Radboud University, Nijmegen, The Netherlands. ¹⁸Department of Psychiatry and Behavioral Sciences, Department of Pediatrics, University of California, Berkeley, CA, USA. ¹⁹Yale Child Study Center and Department of Obstetrics Gynaecology and Reproductive Sciences, Yale School of Medicine, New Haven, CT, USA. ²⁰British Columbia Mental Health and Substance Use Services Research Institute and Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada. ²¹Edwin S.H. Leong Centre for Healthy Aging, University of British Columbia, Vancouver, BC, Canada. ²²These authors contributed equally: Sarah M. Merrill, Chaini Konwar. ✉ e-mail: michael.kobor@ubc.ca