

RESEARCH

Open Access



Consistent cord blood DNA methylation signatures of gestational age between South Asian and white European cohorts

Wei Q. Deng^{1,2,3*}, Marie Pigeyre^{3,4,5}, Sandi M. Azab^{3,6}, Samantha L. Wilson⁷, Natalie Campbell³, Nathan Cawte⁴, Katherine M. Morrison⁸, Stephanie A. Atkinson⁸, Padmaja Subbarao^{3,9,10}, Stuart E. Turvey¹¹, Theo J. Moraes^{9,10}, Piush Mandhane¹², Meghan B. Azad¹³, Elinor Simons¹⁴, Guillaume Pare^{4,5,6,15} and Sonia S. Anand^{3,4,6*}

Abstract

Background Epigenetic modifications, particularly DNA methylation (DNAm) in cord blood, are an important biological marker of how external exposures during gestation can influence the *in-utero* environment and subsequent offspring development. Despite the recognized importance of DNAm during gestation, comparative studies to determine the consistency of these epigenetic signals across different ethnic groups are largely absent. To address this gap, we first performed epigenome-wide association studies (EWAS) of gestational age (GA) using newborn cord blood DNAm comparatively in a white European ($n = 342$) and a South Asian ($n = 490$) birth cohort living in Canada. Then, we capitalized on established cord blood epigenetic GA clocks to examine the associations between maternal exposures, offspring characteristics and epigenetic GA, as well as GA acceleration, defined as the residual difference between epigenetic and chronological GA at birth.

Results Individual EWASs confirmed 1,211 and 1,543 differentially methylated CpGs previously reported to be associated with GA, in white European and South Asian cohorts, respectively, with a similar distribution of effects. We confirmed that Bohlin's cord blood GA clock was robustly correlated with GA in white Europeans ($r = 0.71$; $p = 6.0 \times 10^{-54}$) and South Asians ($r = 0.66$; $p = 6.9 \times 10^{-64}$). In both cohorts, Bohlin's clock was positively associated with newborn weight and length and negatively associated with parity, newborn female sex, and gestational diabetes. Exclusive to South Asians, the GA clock was positively associated with the newborn ponderal index, while pre-pregnancy weight and gestational weight gain were strongly predictive of increased epigenetic GA in white Europeans. Important predictors of GA acceleration included gestational diabetes mellitus, newborn sex, and parity in both cohorts.

Conclusions These results demonstrate the consistent DNAm signatures of GA and the utility of Bohlin's GA clock across the two populations. Although the overall pattern of DNAm is similar, its connections with the mother's environment and the baby's anthropometrics can differ between the two groups. Further research is needed to understand these unique relationships.

Keywords Cord blood DNA methylation, Gestational age at birth, Epigenetic gestational age, Accelerated gestational age

*Correspondence:

Wei Q. Deng
dengwq@mcmaster.ca
Sonia S. Anand
anands@mcmaster.ca

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Epigenome-wide DNA methylation (DNAm) patterns across a variety of tissues and cells have been shown to accurately capture the “biological clock” in both adult [1–3] and pediatric populations [4–6], providing insights into an individual’s biological age, which can differ from their chronological age. These epigenetic clocks and their deviations from chronological age, otherwise known as accelerated aging, have been linked to a range of health outcomes and age-related diseases, suggesting that DNAm changes serve not just as markers of age, but also play a role in the mechanisms of aging that overlap with many chronic diseases [7, 8].

Chronological gestational age (GA) at birth—the actual time elapsed since the last menstrual period of a pregnant woman—is a fundamental component of neonatal care. Babies born preterm (before 37 weeks) or post-term (after 42 weeks of gestation) may have increased risks for various health conditions later in life [9–11], such as cardiovascular diseases, respiratory issues, and metabolic syndrome. In turn, maternal exposures, such as the mother’s health and lifestyle choices, many of which are modifiable, can also influence GA [12, 13]. On the other hand, biological GA at birth has been conceptualized as a measurement of fetal development that can deviate from the chronological gestational age. It is typically measured using biomarkers, such as epigenetic [5] or metabolic markers [14], which allow for a more nuanced understanding of fetal development and the impact of maternal exposures. In particular, epigenetic GA, which can be determined by DNAm GA clocks, has emerged as a more precise approach to measuring biological maturation in response to *in-utero* environmental exposures with potential predictive value for children’s future development and health.

Considering that DNAm patterns are tissue-specific and can reflect gene expression changes, maternal exposures can act as potential triggers to induce tissue-dependent DNAm alterations. For example, nutrition, cigarette smoking, alcohol use, parity, education, and physical health during gestation have been linked to DNAm shifts measured in the placenta [15–18] and newborn cord blood [19–37]. These exposures can reshape the epigenetic landscape of the fetus and placenta, and influence health in later life [38–42]. Further, a pronounced acceleration in GA (GAA), measured by how much the predicted epigenetic age is higher than the chronological age at birth, has also been associated with maternal exposures and offspring characteristics, including older maternal age, a higher pre-pregnancy body mass index (BMI), maternal smoking, gestational diabetes mellitus (GDM), pre-eclampsia, mode of delivery, higher birth weight and length, as well as male sex

in newborns [28, 43–50]. Meanwhile, placental and cord blood DNAm can provide complementary but distinct information, as placental DNAm reflects the interface between mother and fetus and its unique adaptations, while cord blood DNAm is more representative of the systemic fetal epigenetic profile at birth. Consequently, epigenetic clocks based on placental DNAm data do correlate with those derived for cord blood, but differ in terms of the CpGs included, with very little overlap [18], possibly due to the transient nature of placenta tissue and how specific tissues respond to environmental exposures.

The majority of studies on epigenetic GA, GAA, and the characterization of their associations focused on white Europeans. In particular, epigenetic GA clocks were derived almost exclusively in white European populations [5, 6, 51–53] and knowledge about their performance in non-white populations is limited. The general lack of ethnic and racial diversity in omics research restricts the generalizability of these findings. Certain racial groups, such as Black and Asian, are at a disadvantage as they have on average a shorter gestation duration [54] and are at a higher risk for pregnancy complications [55, 56]. South Asian women, owing to their distinct genetic backgrounds [57], environmental exposures [58], and socio-cultural practices [59], are nearly two times more likely to develop GDM as compared to white European women [60, 61]. These genetic and environmental differences become especially pronounced in the length of gestation and its relationship with birth weight. Non-white populations, with their distinct set of genetic and environmental exposures, may present a different profile of epigenetic GA and GAA, which warrants a comparative study of the associated characteristics among diverse populations.

Here, we propose to comparatively examine EWAS of GA and association studies of epigenetic GA and GAA between white European and South Asian birth cohorts with all participants living in Canada. First, to understand whether differences exist between the two populations, we contrasted characteristics of DNAm signatures at individual CpG levels using EWASs, and then at an aggregated level using epigenetic GA and GAA between the two populations. Second, to gain insights on whether these epigenetic markers are differentially or similarly influenced by prenatal factors in the two populations, we comprehensively examined associations of epigenetic GA, and GAA with an extensive collection of maternal exposures and offspring outcomes.

Methods

Study designs

CHILD is a prospective longitudinal birth cohort [62] that enrolled >3,600 pregnant women who gave birth between 2009 and 2012, in Vancouver, Edmonton,

Winnipeg, and Toronto, Canada. The South Asian Birth cohort (START) study is a prospective longitudinal birth cohort [63] that focused exclusively on people who originated from the Indian subcontinents known as South Asians. START recruited South Asian women between 18–40 years of age who were pregnant with a single fetus from the Peel region in Ontario, Canada between 2011 and 2013. These cohorts were not enriched for any clinical conditions and only singleton mothers were recruited. Our analyses focused on a subset of white European-origin mother–offspring pairs from CHILD and South Asian mother–offspring pairs from START, all of whom provided cord blood samples for genome-wide DNA methylation profiling.

Cord blood DNA methylation data in CHILD and START

Data quality processing has been described previously [64]. Briefly, 997 cord blood samples were hybridized to the Illumina Human-Methylation450K BeadChip (HM450K) array, covering CpG or 5′-C—phosphate—G—3′ sites in the entire genome [65]. We followed standard quality control procedures designed for HM450K using the R “*sesame*” package [66] and generated the β -matrix for further processing. Suppl. Table 1 summarizes the sample and probe inclusion/exclusion criteria. Briefly, duplicated probes, probes with detection threshold p -value < 0.05, more than 10% missing, or known to be cross-reactive or overlap with single nucleotide polymorphisms were removed [67, 68]. Further, for CpG probes with a missing rate < 10%, mean imputation was used to fill in the missing values. At each CpG site, the β -value reflects the ratio of methylated signals relative to total signals and is a continuous measure between 0 and 1. The cleaned datasets contain 504 START samples with 361,234 CpGs and 352 CHILD samples with 358,113 CpGs, covering CpG sites on autosomes and X chromosomes. Finally, we estimated cell-type proportions (CD8T, CD4T, Natural Killer cells, B cells, monocytes, granulocytes, and nucleated red blood cells) following a reference-based approach developed for cord blood [69] using the R package “*FlowSorted.CordBloodCombined.450k*”.

Maternal variables

Participants in the START and CHILD cohorts completed self-reported, study-specific questionnaires, which were used to derive standardized variables on personal and family medical histories, as well as social and cultural practices. Maternal diet was assessed using a previously validated ethnic-specific food frequency questionnaire (FFQ; [70] in START and a semi-quantitative multiethnic FFQ (created by the US-based Fred Hutchinson Cancer Center) in CHILD [70]).

Three validated maternal diet patterns [71] were derived from the FFQs: 1) the plant-based pattern was characterized by vegetables, legumes, fermented dairy, whole grains, nonmeat dishes, and a lack of red meat; 2) the Western pattern had a high loading of sweets, red and processed meats, French fries, starchy vegetables, condiments, and sweet drinks; and 3) the health-conscious pattern was characterized by seafood, poultry, and red meats; eggs; cruciferous vegetables; leafy greens; fruit; refined grains; stir-fried dishes; and condiments. Maternal smoking history (current, quit during pregnancy, quit before pregnancy, never-smoker) and household smoking exposure (hours per week) were self-reported. Maternal education questions were harmonized between CHILD and START to produce a single variable indicating the number of years in school. Socioeconomic status was captured by the validated social disadvantage index, a composite measure of household income, marital status, and maternal employment [72], and was weakly correlated with maternal education ($r^2 < 0.05$). Clinical variables of interest included maternal age at delivery, pre-pregnancy BMI (kg/m^2), pre-pregnancy weight (kg), gestational weight gain (kg), GDM status, gestational hypertension, pre-eclampsia, and parity (defined as the number of births before current pregnancy). GDM was determined by self-report only in CHILD, while in START, it was determined by a combination of oral glucose tolerance test (OGTT), self-report, and reported diabetic treatments, including insulin, pills, and restricted diet. The OGTT threshold (≥ 5.2 at baseline or ≥ 7.2 at 2 h) for South Asian women proposed in Born in Bradford [73, 74] was used in START.

Offspring characteristics

Chronological GA at birth, reported in weeks and days, was collected from participants’ birth charts. As an additional quality control step, we excluded 10 newborns from CHILD and 16 from START who either lacked GA data or were born before 36 weeks, such that the remaining newborns had gestation at birth between 36 and 42 weeks. Offspring anthropometrics are available at birth from medical charts or measured at < 2 days old, and subsequently measured at the 1, 2, 3, and 5-year follow-up visits. These included height (cm), weight (g), ponderal index (PI; kg/m^3), and BMI (kg/m^2), at all visits; and the sum of skinfolds (triceps and subscapular skinfolds in mm), waist circumference (cm), hip circumference (cm), and waist-hip ratio (WHR) at 3- and 5-year visits.

EWAS of GA

The final datasets with cleaned DNAm and chronological GA at birth consisted of 342 and 490

mother-newborn pairs from CHILD and START, respectively. We performed an EWAS separately in each cohort, testing the association between DNAm β -values at each CpG and GA using a linear regression model. Contrary to existing EWASs where the methylation values are typically treated as the outcome and the exposure as the predictor; we reversed the regression such that the methylation levels were the predictors and chronological GA at birth as the outcome. This reverse regression approach is statistically more robust as GA is roughly normally distributed and produces similar evidence of association between the two sets of variables as the original approach [75]. Further, this reverse regression is consistent with our subsequent analysis using DNAm GA clock, a linear combination of additive effects over multiple CpGs, as the predictor and chronological GA at birth as the outcome. Though neither the plate or the row number, the two main sources of batch effects [76], was associated with GA using an omnibus test, there was more variability in the pairwise differences of GA between plate numbers ($p=0.005$ – 0.94 in CHILD and $p=0.01$ – 0.94 in START). Among the subset of samples with genetic data, there was no association between the first 10 genetic principal components and chronological GA at birth in START (adjusted $R^2=-0.0057$, $n=480$, p -value=0.7) or CHILD (adjusted $R^2=0.0022$, $n=290$, p -value=0.4). Considering the reduced sample size and the absence of evidence indicating confounding, we concluded that further adjustment for genetic ancestry within each cohort was not necessary. Thus, the final model adjusted for the plate number to correct for potential batch effects, the estimated compositions of cord blood cell types, maternal age, maternal education, maternal smoking history, social disadvantage index, parity, and newborn sex (coded as 0 for male and 1 for female). For association testing of X chromosome CpGs, we adopted a robust 2 degrees of freedom test proposed for X chromosome genetic associations to mitigate the influence of unknown X-inactivation [77]. We then meta-analyzed association results from CHILD and START using an inverse variance-weighted fixed-effect model, and CpGs that showed heterogeneity in effects were retained for further clarification. CpG sites located within genes have been labeled using the UCSC reference gene names provided by the annotation file from Illumina ('HumanMethylation450_15017482_v.1.2.bpm'). For each EWAS or meta-analysis, the false discovery rate (FDR) adjustment was used to control for multiple testing and an FDR-adjusted p -value < 0.05 was considered statistically significant. We further contrasted our primary findings with CpGs that were previously linked to GA using cord blood

DNAm as reported in Bohlin [51]. The concordance between signals from Bohlin and each EWAS, and those between START and CHILD, was tested using a two-sample proportion test.

EWAS functional analysis

Differentially methylated regions (DMRs) were identified using DMRcate [78], which is a data-driven method that incorporates only spatial annotations. The algorithm combines robust estimates at individual CpG sites via a Gaussian kernel within a given window to produce a smoothed estimate. We followed the recommended default parameters (e.g., ≥ 2 CpGs, Gaussian kernel bandwidth of 1000) to call DMRs. A DMR was deemed significant if the observed smoothed estimate was more extreme than its expected value given the same spatial information at an FDR-adjusted Stouffer's $p < 0.05$. Using genes that overlap with significant DMRs called in CHILD, START, and additionally, genes mapped from significant CpGs in Bohlin et al., we examined for enrichment in biological pathways using g:Profiler [79] (<https://biit.cs.ut.ee/gprofiler/snpense>; accessed on March 28, 2024) based on several databases, including gene ontology (GO), Kyoto encyclopedia of genes and genomes (KEGG), Reactome, and WikiPathways. HGNC gene symbols that had multiple ensemble IDs match were removed from the query. We deemed a pathway to be relevant if the adjusted p -value < 0.05. Finally, special consideration was given to the intersection of genes associated with type 2 diabetes (T2D) and GDM, informed by our preliminary findings on the EWAS and epigenetic GA clock. Specifically, we investigated whether significant CpGs were enriched in T2D and GDM genes, by searching the GWAS catalog [80] for genetic variants associated with T2D or GDM at genome-wide significance ($p < 5 \times 10^{-8}$). We only selected genetic variants reported in studies with a discovery sample size > 100,000 and published in the last five years, using the keywords "type 2 diabetes mellitus" (EFO ID: MONDO_0005148) and "gestational diabetes" (EFO ID: EFO_0004593). This approach not only ensures that our search reflects the most recent scientific findings but also assumes that any previously established relevant genetic variants would be sufficiently represented and validated within these large, well-powered studies.

Epigenetic GA and GAA

Epigenetic GA was estimated using Bohlin's cord blood [51], Knight's cord blood [5], and Mayne's placenta [50] GA clocks developed for HM450K data, as well as Lee's placenta [6] and Haftorn's cord blood [52] clock for EPIC data. These algorithms are implemented in the R package "methylclock" [81]. GAA was calculated as the residual

from a linear regression using the estimated epigenetic GA as the outcome and the chronological GA as the predictor such that GAA would be uncorrelated with epigenetic GA. A positive GAA indicates acceleration whereby the newborn's epigenetic GA is older than the chronological GA, and a negative value implies deceleration. We empirically assessed whether the GA clocks and the corresponding GAA were transferrable to the START cohort by examining the distribution of these scores between START and CHILD using a *t*-test.

Statistical analysis

The performance of various GA clocks to predict chronological GA was quantified using Pearson's correlation coefficient in both cohorts. The GA clock that most strongly correlates with the chronological GA was used as the estimated epigenetic GA in all subsequent analyses. We then explored maternal characteristics that were predictive of GA, epigenetic GA, and GAA using a stepwise model selection method. The search would yield a final deterministic model with a subset of the variables as evaluated by the Akaike information criterion. For completeness, we further tested the univariate association between GA, epigenetic GA, and GAA and individual traits in the following categories: 1) newborn sex, and birth anthropometrics [4 variables]; 2) offspring anthropometrics at each follow-up visit [20 variables]; and 3) maternal exposures [14 variables], using simple linear regression. An FDR adjustment was applied to univariate results and *q*-values were reported whereby we claimed suggestive evidence of association when $q < 0.1$. For offspring anthropometrics, whenever appropriate, the model adjusted for children's age and sex at each visit. The reported effect sizes were based on the original scale of each variable for ease of interpretation. Missing data were imputed using a random forest imputation algorithm implemented in the R package "missforest" [82]. All data processing and analyses were conducted in R v.4.1.0 [83].

Results

Study sample characteristics

Demographic characteristics and relevant covariates of the epigenetic subsamples (Table 1) were not statistically different ($p > 0.001$) from their respective full cohort after controlling for multiple hypothesis (Suppl. Table 2). Maternal characteristics, including parity, maternal age, pre-pregnancy weight, GDM, smoking history, education, social disadvantage index, and diet patterns, were significantly different between South Asian and white European women. Specifically, there was a much higher rate of GDM (36% in START vs. 4% in CHILD), and pregnant South Asian women in START were exclusively

Table 1 Sample characteristics of CHILD and START cohorts

	CHILD (N = 342)	START (N = 490)	P-value
<i>Maternal</i>			
Mother's Age			
Mean (SD)	32.7 (± 4.4)	30.1 (± 3.9)	< 0.001
Missing	4 (1.2%)	0 (0%)	
Parity			
Mean (SD)	0.72 (± 0.89)	0.80 (± 0.81)	0.191
Missing	0 (0%)	13 (2.6%)	
Pre-pregnancy BMI			
Mean (SD)	24.7 (± 5.2)	23.7 (± 4.4)	0.0236
Missing	131 (38.3%)	2 (0.4%)	
Pre-pregnancy weight			
Mean (SD)	68.8 (± 15.2)	62.6 (± 11.8)	< 0.001
Missing	126 (36.8%)	0 (0%)	
Gestational weight gain			
Mean (SD)	15.4 (± 6.2)	14.3 (± 7.5)	0.0413
Missing	120 (35.1%)	9 (1.8%)	
Gestational diabetes mellitus			
No	328 (95.9%)	313 (63.9%)	< 0.001
Yes	14 (4.1%)	176 (35.9%)	
Missing	0 (0%)	1 (0.2%)	
Gestational hypertension			
No	326 (95.3%)	479 (97.8%)	0.0801
Yes	16 (4.7%)	11 (2.2%)	
Pre-eclampsia			
No	331 (96.8%)	486 (99.2%)	0.0217
Yes	11 (3.2%)	4 (0.8%)	
Smoking history			
Never smoked	241 (70.5%)	487 (99.4%)	< 0.001
Quit before this pregnancy	68 (19.9%)	1 (0.2%)	
Quit during this pregnancy	17 (5.0%)	1 (0.2%)	
Currently smoking	11 (3.2%)	0 (0%)	
Missing	5 (1.5%)	1 (0.2%)	
Smoking exposure (hr/week)			
Mean (SD)	1.0 (± 7.8)	0.33 (± 2.7)	0.13
Missing	12 (3.5%)	39 (8.0%)	
Years of education			
Mean (SD)	17.0 (± 3.1)	15.8 (± 2.4)	< 0.001
Missing	9 (2.6%)	0 (0%)	
Social disadvantage index			
Mean (SD)	0.41 (± 0.95)	1.8 (± 1.4)	< 0.001
Missing	37 (10.8%)	69 (14.1%)	
Plant-based diet			
Mean (SD)	-0.48 (± 0.46)	1.6 (± 1.1)	< 0.001
Missing	21 (6.1%)	16 (3.3%)	
Health-conscious diet			
Mean (SD)	0.21 (± 0.81)	-0.42 (± 0.79)	< 0.001
Missing	21 (6.1%)	16 (3.3%)	
Western diet			

Table 1 (continued)

	CHILD (N = 342)	START (N = 490)	P-value
Mean (SD)	-0.16 (±0.64)	-0.51 (±0.65)	<0.001
Missing	21 (6.1%)	16 (3.3%)	
<i>Newborn</i>			
Gestational age (weeks)			
Mean (SD)	39.6 (±1.3)	39.3 (±1.1)	<0.001
Newborn sex			
Male	187 (54.7%)	234 (47.8%)	0.0581
Female	155 (45.3%)	256 (52.2%)	
Birth length (cm)			
Mean (SD)	51.7 (±2.5)	51.5 (±2.6)	0.381
Missing	65 (19.0%)	7 (1.4%)	
Birth weight (kg)			
Mean (SD)	3.5 (±0.48)	3.3 (±0.45)	<0.001
Missing	4 (1.2%)	1 (0.2%)	
Newborn BMI (kg/m ²)			
Mean (SD)	13.1 (±1.4)	12.3 (±1.4)	<0.001
Missing	66 (19.3%)	7 (1.4%)	
Newborn ponderal index (kg/m ³)			
Mean (SD)	25.4 (±3.1)	24.0 (±3.2)	<0.001
Missing	66 (19.3%)	7 (1.4%)	

non-smoking during pregnancy. There was also a noticeable difference in the pattern of phenotypic correlations between the two cohorts, both in size of correlation and directions (Suppl. Figure 1). In terms of newborn characteristics, the mean GA at birth in START (39.3 ± 1.1) was lower than in CHILD (39.6 ± 1.3) when restricted to gestation superior or equal to 36 weeks. As expected, South Asians had a lower birth weight (3.3 ± 0.45 kg vs. 3.5 ± 0.48 kg), but were of similar length as compared to white Europeans. Even after accounting for chronological GA at birth, newborn sex, social disadvantage index, and GDM, the difference in birth weight persisted ($p < 0.001$).

EWAS of GA

Figure 1 highlights the 1,652 and 2,136 differentially methylated CpGs associated with GA after FDR correction (above the red dashed line and in Suppl. Table 3) in CHILD and START, respectively. The number of significant CpGs overlapping between the two cohorts was 599 (Suppl. Figure 2). While the signal overlap between the START and CHILD was smaller than expected, the majority of signals separately from CHILD (73%) and START (72%) agreed with those reported in Bohlin's EWAS [51] in a much larger sample ($n > 1000$). Across these EWASs, there was considerable empirical inflation

at various significance thresholds as well as overall inflation at the medium (Suppl. Table 4), however, the level of inflation was comparable across the studies and had been similarly observed in the context of other -omics studies as a result of widespread true associations [84, 85]. The meta-analysis of all CpGs with heterogeneity p -value > 0.01 identified an additional 4664 CpGs (7492 CpGs total) that were significantly associated with GA. Only 108 CpGs that were significant in the meta-analysis showed marginal evidence for heterogeneity of effects (heterogeneity p -value < 0.01), however, the estimated effects of these associations were always in the same direction (Suppl. Figure 3). Approximately 78% of the CpGs identified as significant in Bohlin's EWAS (11,337 of the 14,501 reported in the EWAS catalog using ultrasound-estimated GA at birth) were also significant in both CHILD and START. To benchmark our results, we highlighted these CpGs in Fig. 1 in purple. In general, there was good agreement in signals identified between CHILD and START, as well as those between CHILD or START and Bolin's (Suppl. Figure 4). We identified 1479 and 9142 DMRs in CHILD and START, respectively, following the default FDR adjustment (Suppl. Tables 5 and 6). The largest DMR for CHILD contained 53 CpGs and was mapped to the *B3GALT4* and *WDR46* genes, while the largest DMR for START (5 CpGs) overlapped with the *PRR5L* gene. On average, there was equal representation of significant DMRs across the genome, however, we observed significant DMRs identified in START to contain more CpGs than in CHILD on average (Suppl. Figure 5). At an FDR-adjusted p -value threshold of < 0.05 , 451 out of 489 pathways identified for CHILD and 675 out of 906 in START were identical to those identified using the significant CpGs from Bohlin (Suppl. Table 7). The top pathways across the three EWASs captured biological processes related to fetal development and immune system function, such as proliferation, migration, differentiation of different types of cells, and anatomical structure development.

We retained 1509 unique gene regions from the GWAS catalog, of which 1508 were linked to T2D and 7 were associated with GDM. Of these, 130 were represented in genes mapped onto by significant CpGs identified in either CHILD, START, or both (Suppl. Table 8). In both CHILD and START, we found the number of genes containing at least one significant CpG to be overrepresented among known T2D genes (Suppl. Figure 6). Specifically, 71 out of 921 genes identified in CHILD (binomial test for proportion $p = 9.9 \times 10^{-5}$) and 97 out of 1158 genes identified in START (binomial test for proportion $p = 1.3 \times 10^{-7}$) were represented in genes reported in T2D or GDM GWASs (Suppl. Table 9).

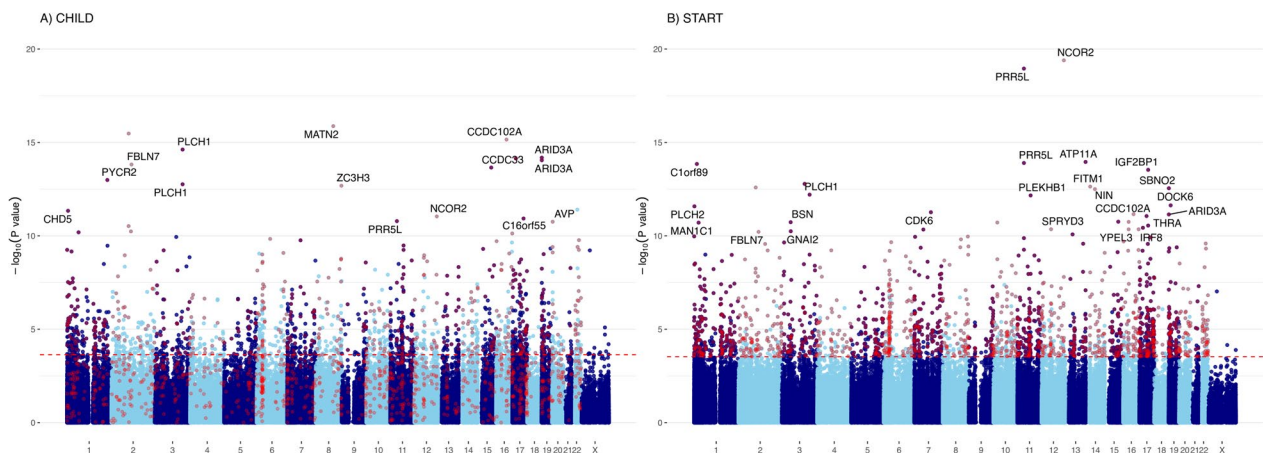


Fig. 1 Manhattan plots of the EWAS results of gestational age in CHILD (left) and START (right). Manhattan plots summarized the association *p*-values between cord blood DNA methylation levels and gestational age in CHILD (left) and START (right). The red line denotes the smallest $-\log_{10}(p\text{-value})$ that is below the FDR correction threshold of 0.05. The red dots represent established associations with gestational age based on ultrasound [51]

Epigenetic GA and GAA

Out of the six epigenetic GA clocks tested, the Knight and Bohlin clocks designed for the HM450 array had better overlap with CpGs in our data as compared to the EPIC clocks with high missingness (Table 2). In terms of performance, the cord blood-based Knight and Bohlin clocks had better correlation with the chronological GA, with the Bohlin clock being consistently the best in START ($r=0.66$; $p=6.9 \times 10^{-64}$) and CHILD ($r=0.71$; $p=6.0 \times 10^{-54}$), producing the highest correlation coefficient and the smallest median absolute difference (Table 2). Thus, all subsequent results were based on epigenetic GA estimated using Bohlin’s cord blood clock. The estimated epigenetic age using Bohlin’s clock and the chronological GA roughly followed a linear relationship,

while the residual approach ensured that GAA was uncorrelated with the chronological age (Fig. 2). Finally, there was no statistically significant difference between the mean estimated epigenetic GA in white European (40.6 ± 0.81) and South Asian (40.6 ± 0.74) newborns, respectively. Similarly, there was no difference in the GAA derived using the residual approach (Suppl. Table 10).

Maternal characteristics that are predictive of GA, epigenetic GA, and GAA

The final models for each outcome are summarized in Table 3. Between CHILD and START, the common set of predictors for GA included pre-eclampsia and GDM, but maternal age and dietary patterns were specific to START,

Table 2 Performance of cord blood DNA methylation gestational age clocks in CHILD and START

Dataset	Clock	Tissue	CpGs used (total)	Pearson’s r	P-value	MAD (days)
CHILD (n=342)	Knight	Cord blood	132 (148)	0.60	5.74E-35	9
	Bohlin	Cord blood	79 (96)	0.71	5.98E-54	4
	Mayne	Placenta	54 (62)	0.12	1.18E-02	6
	Lee.RPC	Placenta	950 (1125)	0.20	9.50E-05	6
	Lee.CPC	Placenta	950 (1125)	0.33	1.81E-10	7
	Haftorn-EPIC	Cord blood	71 (176)	0.13	7.73E-03	6
START (n=490)	Knight	Cord blood	132 (148)	0.44	4.09E-25	14
	Bohlin	Cord blood	79 (96)	0.66	6.90E-64	4
	Mayne	Placenta	54 (62)	0.13	0.0030	11
	Lee.RPC	Placenta	958 (1125)	0.024	0.59	10
	Lee.CPC	Placenta	958 (1125)	0.15	7.40E-04	9
	Haftorn-EPIC	Cord blood	74 (176)	0.028	0.53	8

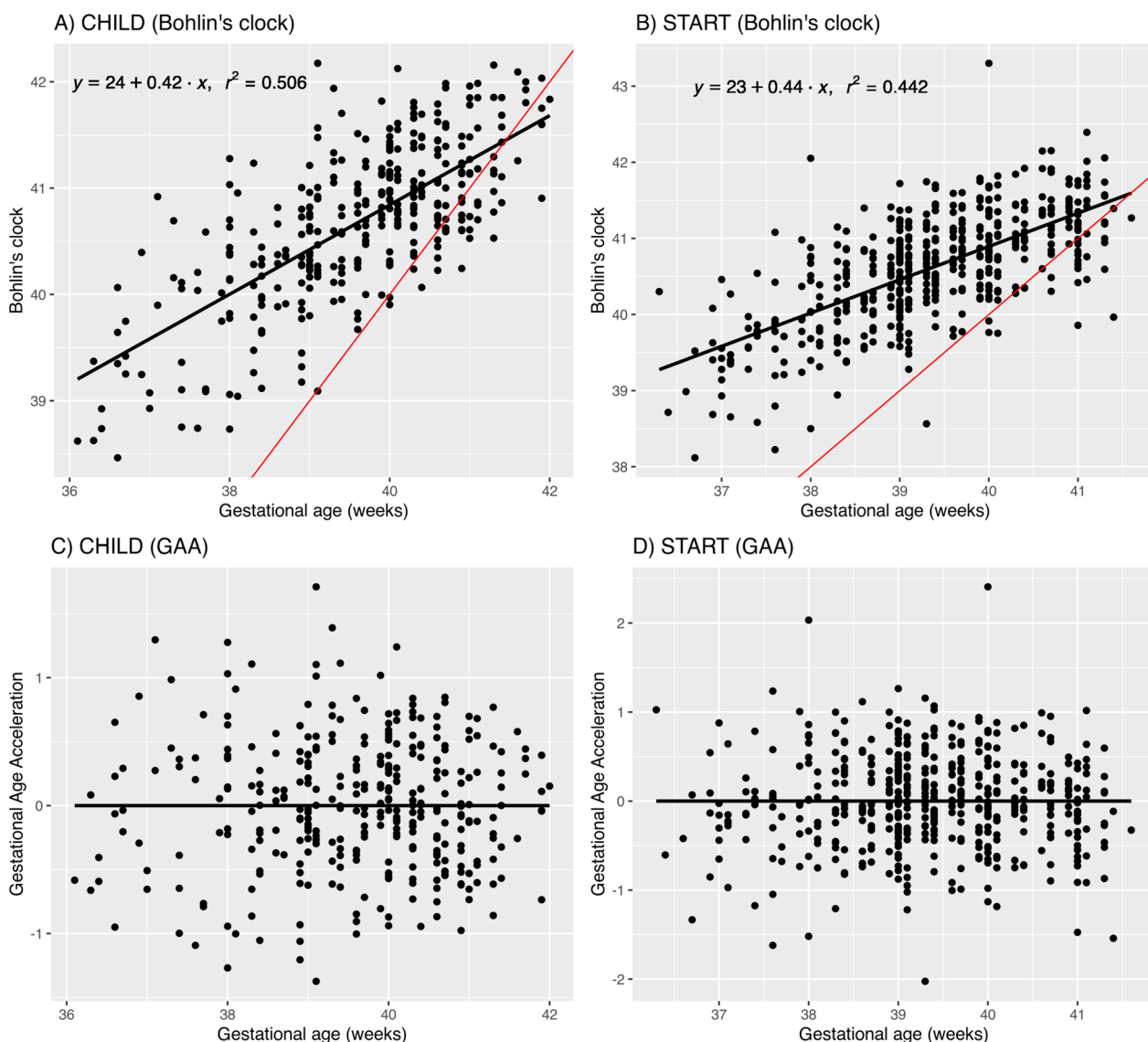


Fig. 2 Relationship between newborn epigenetic gestational age at birth based on cord blood DNA methylation data, chronological gestational age at birth, and gestational age acceleration. Panels A-B show the scatterplot of chronological gestational age at birth and Bohlin's gestational age clock in CHILD and START, respectively. Panels C-D show the scatterplot of chronological gestational age at birth and gestational age acceleration in CHILD and START, respectively. The black lines are the best-fitted line using ordinary least squares, with equations displayed in Panels A) and B). The red lines indicate the reference line of $y = x$

while pre-pregnancy weight, BMI, and gestational weight gain were unique predictors for GA in CHILD alone. These models explained 8.03% and 7.90% of the GA variance in CHILD and START, respectively. Meanwhile, parity, GDM, and newborn sex were important predictors in both CHILD and START for epigenetic GA, with gestational weight gain, pre-pregnancy weight and BMI being additionally implicated in CHILD. In particular, GDM was negatively associated with epigenetic GA, but more severely in

CHILD with a reduction of 0.66 ± 0.19 weeks as compared to the 0.16 ± 0.07 weeks in START. Together, these predictors explained 8.52% and 4.50% of the epigenetic GA variance, which was higher than that for GA in CHILD but lower than that in START. Both parity and newborn female sex were also negatively associated with GAA in CHILD and START, but GDM was only associated with GAA in CHILD while dietary patterns, gestational weight gain, and maternal smoking were uniquely associated with GAA in START (Table 3).

Table 3 Multivariable models using maternal characteristics to predict gestational age, epigenetic gestational age, acceleration of gestational age in CHILD and START

		CHILD								
		GA			DNAm GA			GAA		
		Estimate	Error	P-value	Estimate	Error	P-value	Estimate	Error	P-value
Predictors	Intercept	39.73	0.54	2.14E-209	41.33	0.34	5.10E-281	0.21	0.26	0.43
	Age (mother)							0.01	0.01	0.068
	Parity	-0.14	0.07	0.061	-0.14	0.04	0.0015	-0.11	0.03	0.0011
	BMI	-0.10	0.04	0.017	-0.04	0.02	0.085			
	Weight	0.04	0.01	0.0051	0.015	0.01	0.071			
	Weight gain	0.05	0.01	1.04E04	0.023	0.01	0.0025			
	Diabetes mellitus	-1.01	0.33	0.0023	-0.66	0.19	7.39E-04	-0.27	0.14	0.055
	Pre-eclampsy	-0.62	0.37	0.096						
	Smoking									
	Smoke									
	Disadvantage									
	Conscious diet									
	Western diet									
	Plant-based diet									
	Model Metric	Newborn sex				-0.15	0.08	0.058	-0.17	0.06
P-value		6.19E-06			2.77E-06			1.71E-04		
squared		8.03%			8.52%			5.29%		
		START								
		GA			DNAm GA			GAA		
		Estimate	Error	P-value	Estimate	Error	P-value	Estimate	Error	P-value
Predictors	Intercept	41.91	0.64	1.04E-243	41.90	0.44	3.72E-317	0.25	0.10	0.012
	Age (mother)	-0.04	0.01	3.1E04	-0.01	0.01	0.13			
	Parity				-0.11	0.04	0.0095	-0.09	0.03	0.0033
	BMI									
	Weight									
	Weight gain							-0.01	0.00	0.034
	Diabetes mellitus	-0.34	0.10	8.76E-04	-0.16	0.07	0.019			
	Pre-eclampsy	-0.76	0.53	0.15	-0.58	0.35	0.10			
	Smoking							0.38	0.24	0.11
	Smoke									
	Disadvantage									
	Conscious diet	-0.14	0.06	0.028						
	Western diet	0.16	0.07	0.034				-0.08	0.04	0.033
	Plant-based diet									
	Model Metric	Newborn sex				-0.11	0.07	0.11	-0.08	0.05
P-value		1.35E-08			2.51E-05			0.00056		
squared		7.90%			4.50%			3.40%		

Univariate association of maternal and offspring characteristics with epigenetic GA and GAA

Newborns from mothers with GDM on average had a lower epigenetic GA in both CHILD (0.6 weeks, $p=2.98 \times 10^{-3}$, $q=0.02$; Suppl. Table 11) and START (0.18 weeks, $p=6.58 \times 10^{-3}$, $q=0.039$; Suppl. Table 12),

but more severely in CHILD as compared to START. Parity was significantly associated with epigenetic GA in both CHILD ($p=0.0016$, $q=0.013$) and START ($p=1.0 \times 10^{-4}$, $q=0.0014$), indicating that the number of previous births reduces epigenetic GA. Additionally, in START, the age of the mother was negatively associated

with both GA and epigenetic GA (Suppl. Table 12). However, a subsequent subset analysis revealed that maternal age was not associated with epigenetic age when adjusted for parity ($p=0.49$) and that it was not associated with epigenetic GA when stratified based on the number of previous pregnancies (Suppl. Figure 7). Further, male sex was only associated with GAA in CHILD ($p=0.0044$; FDR-adjusted $p=0.036$; Suppl. Table 11) but not in START ($p=0.14$; Suppl. Table 12).

As expected, the associations between epigenetic GA and newborn weight and length were significant in both South Asian and white European cohorts ($q<0.0025$; Suppl. Tables 11 and 12). The most interesting contrast was the association with newborn PI, whereby a positive association with GA (0.063 weeks; $q=2.89\times 10^{-4}$) and epigenetic GA (0.034 weeks; $q=0.005$) was observed in START, but it was negatively associated with GAA in CHILD (0.032 weeks; $p=0.0009$; $q=0.019$), despite newborn weight and height being similarly positively associated with GA and epigenetic GA in both cohorts. Both epigenetic GA and GAA were positively associated with 1-year length in CHILD ($q<0.02$), and additionally GAA with 1-year PI and 2-year length ($q<0.05$; Suppl. Table 11). We did not observe any association between anthropometrics at subsequent follow-up visits and GA, epigenetic GA, or GAA in START (Suppl. Table 12).

Discussion

This is the first study to comparatively examine DNAm signatures of GA in two ethnic populations (living in the same country) using cord blood. We found consistent DNAm signatures of gestational age at individual CpG levels and confirmed that Bohlin's epigenetic GA clock correlated well with chronological GA in white Europeans and can be generalized to South Asians. In both populations, the GA clock was positively linked to newborn weight and length, and negatively to gestational diabetes, newborn female sex, and parity. Unique to South Asians, the GA clock was also associated with a higher newborn PI. We confirmed the associations of parity and newborn sex with GAA in both populations and discovered that white European newborn males exhibited twice as much accelerated GA as compared to South Asians.

Previous EWASs of GA have found overlapping signals with other outcomes, such as birth weight, while EWASs of birthweight were enriched for CpGs previously associated with prenatal smoking, folic acid intake, and maternal hypertension or pre-eclampsia [26, 53, 86]. Since the strongest association we found with the epigenetic GA clock was GDM, we were interested in whether GDM or T2D genetics could be implicated in the epigenetics of GA. We found two CpGs significantly associated with GA in START that were present in the

TCF7L2 gene, whose risk alleles increase the risk of T2D by reducing insulin secretion and are also associated with a lower BMI [87, 88]. This lends support to the hypothesis that GDM in South Asians is predominantly due to insulin deficiency, which is consistent with their lower birth weight [89], whereby comparatively white Europeans may have a stronger insulin resistance component. Meanwhile, *GLIS3* is another recognized gene for diabetes that is associated with the development of beta cells [90] and neonatal diabetes [91]. Further, *ADA* gene, which has been linked to T2D [92] and several serum metabolites [93], was also mapped to CpGs significantly associated in START but was just below statistical significance in CHILD. Overall, there was an overrepresentation of genes that were known to associate with T2D in both START and CHILD, but with a more pronounced enrichment in START. The shared genes could be part of the mechanisms through which metabolic risks are transmitted from mothers to offspring, contributing to a transgenerational cycle of metabolic disorders [94–96].

Despite widespread differences in DNAm patterns between South Asians and white Europeans in an adult population [97], at an aggregate level, the Bohlin clock consisting of 132 CpGs produced a robust measure of biological GA with comparable performance in both cohorts. While the epigenetic GA clocks can adequately capture biological GA in both populations, they displayed subtle yet differential associations with maternal exposures and effects on offspring. The multitude of maternal exposures, ranging from nutrition, substance use, and metabolic factors to chronic health conditions, presented a complex tapestry of influences on GA, epigenetic GA and GAA. In general, South Asian women had low to no smoking and a diet that was mostly plant-based, and their newborns often exhibited a lower birthweight and a lower GA as compared to white Europeans. In this particular population, we found evidence of Western dietary pattern to influence GA and GAA. But this should be interpreted in the context of differences in the correlation between two populations: a plant-based diet score was positively correlated with Western and health-conscious diets in South Asian women ($r=0.35$ and 0.32), but a choice of plant-based diet implied a reduced Western diet ($r=-0.21$) and a strong preference for health-conscious diet ($r=0.64$) in white European women (Suppl. Figure 1).

In both cohorts, parity and newborn female sex were found to be negatively associated with epigenetic GA and GAA. While both maternal age and parity have been shown to associate with GAA [18, 49], our univariate and multivariable analyses pointed to parity being independently associated with epigenetic GA and GAA, but not maternal age. Consistent with this claim, a previous

study showed a negative association between parity and maternal telomere length, which is an indicator of cellular aging. Advanced maternal age (e.g., > 35) is typically considered a risk factor for pregnancy-related outcomes [98, 99]. Here we showed that maternal age was negatively associated with GA and epigenetic GA in the univariate analysis, but only in the South Asian cohort, suggesting it could be capturing the effect of parity. This was supported by the observation that parity was more strongly correlated with maternal age in START (Pearson's $r=0.45$) than CHILD ($r=0.27$), reflecting a higher proportion of older mothers being multiparous among South Asians. This could also explain why maternal age has the most significant impact on negative outcomes in South Asian women [100]. On the other hand, our finding that newborn males have accelerated GA in both CHILD and START is consistent with established literature that male sex was associated with accelerated aging in newborns [43] and in adult populations [101].

In maternal health outcomes, the observed prevalence of GDM was consistent with those reported for the two populations in general [102, 103]. We observed a negative relationship between GDM and epigenetic GA in both cohorts, as well as GDM and GAA in CHILD. The direction of associations was in line with a previous report [48]. However, there was no evidence to support the role of GDM on GAA in the South Asian cohort despite much higher prevalence and sample size. We were able to replicate previous findings of pre-pregnancy weight and BMI having an impact on epigenetic GA in CHILD [43], but there was a coherent indication that gestational weight gain might be the more relevant exposure to GA and epigenetic GA in both cohorts.

The findings of this study should be interpreted in the light of several limitations. Cord blood DNAm in newborns reflects the regulation of gene activity during embryonic development, cellular differentiation, and response to environmental factors, and is thus a valuable source of information for assessing maternal risk factors and predicting future health outcomes. However, depending on the tissue types, resolution of the epigenetic data (e.g., single cell vs. tissue), technology (HM450K vs. EPIC), and method of GA clock construction, the performance of existing GA clocks varied substantially. The main observation is that the barriers to trans-ethnic EWAS analysis lie in various choices of technical and study design. While tissue-specific clocks generated roughly similar results between the South Asian and the white European cohorts (Table 3), we also observed that the robust Lee clock designed for pregnancies without complications produced very different results, possibly due to the large difference in GDM cases (36% in START vs. 4% in CHILD). For example, different

tissue types might reveal tissue-specific responses, a placental clock could produce an elevated association with maternal exposures vs. offspring outcomes [6], which is complementary to existing studies of GA using cord blood. We have shown that DNAm clock, while universal in its foundational concept, can exhibit specific patterns and variations when analyzed across different ethnic groups. Factors contributing to these disparities might range from genetic backgrounds to varied environmental and socio-cultural practices that might not be readily available, such as alcohol consumption [34, 104] and air pollution [105, 106], both are known to alter epigenetics. These discrepancies, while enriching our understanding of epigenetic variability between ethnic groups, also serve as a clarion call for tailored healthcare and research approaches across populations.

In summary, this paper comparatively examined the transferability and health implications of epigenetic GA and GAA in white European and South Asian birth cohorts living in the same country. Two key findings are the consistent associations of parity, GDM, and newborn sex with epigenetic GA in both South Asian and white European cohorts, but with weaker effects observed in the South Asian cohort; and the positive association between epigenetic GA and newborn ponderal index that is exclusive to South Asian newborns. The epigenetic GA clock presents a promising avenue to bridge our understanding of maternal exposures and offspring health outcomes, and these differentiating characteristics of associations across the cohorts, enrich this narrative. Further research in the South Asian population, both adult and pediatric, is needed to strengthen these findings.

Abbreviations

BMI	Body mass index
CpGs	5'—C—phosphate—G—3'
DNAm	DNA methylation
EWAS	Epigenome-wide association study
GA	Gestational age
GAA	Gestational age acceleration
GDM	Gestational diabetes mellitus
GWAS	Genome-wide association study
OGTT	Oral glucose tolerance test
PI	Ponderal index
T2D	Type 2 diabetes
WHR	Waist-to-hip ratio

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13148-024-01684-0>.

Additional file 1 Table S1. Quality controls for the inclusion/exclusion of samples and methylation probes. Table S2. Characteristics of CHILD and START epigenetic subsample vs. the overall cohort. Table S3. CpGs associated with gestational age in CHILD or START with FDR adjusted p -value < 0.05. Table S4. Measures of type I error inflation for EWAS association p -values in CHILD and START. Table S5. Significantly associated DMR

identified in CHILD. Table S6. Significantly associated DMR identified in START. Table S7. Significantly enriched biological pathways for DMRs associated with chronological GA at birth in CHILD and START. Table S8. GWAS genes associated with T2D or GDM mapped to CpGs significantly associated with GA in START or CHILD. Table S9. Enrichment of GWAS genes associated with T2D or GDM using a two-sample binomial test. Table S10. Differences in gestational age, epigenetic gestational age and acceleration between studies or sex. Table S11. Associations between gestational age, epigenetic gestational age, acceleration of gestational age, and maternal and offspring characteristics in CHILD. Table S12. Associations between gestational age, epigenetic gestational age, acceleration of gestational age, and maternal and offspring characteristics in START

Additional file2 Figure S1. Heatmap of maternal and offspring characteristics in CHILD and START. Heatmap in Panels A-B) display the correlation matrix between the maternal and offspring variables in CHILD and START, respectively. Each cell represents the correlation coefficient between the variables on the corresponding x and y axes, ranging from -1 (strong negative correlation, shown in blue) to +1 (strong positive correlation, shown in red). Cells with a correlation close to 0 are colored in neutral (white), indicating no correlation. The color intensity increases with the strength of the relationship

Additional file3 Figure S2. A Venn Diagram for significant CpGs identified in CHILD, START and Bohlin

Additional file4 Figure S3. Relationship between CpG association effect size and heterogeneity of effect. Panel A) shows the scatterplot of estimated association effect in CHILD (x-axis) and START (y-axis) for CpGs that were significant in the meta-analysis (FDR adjusted $p < 0.05$) that also were heterogenous in their effects (Heterogeneous $p < 0.01$) (#CpG = 108); for the same set of CpGs, Panel B) shows the relationship between the absolute difference in beta-values (taken as the mean difference of CHILD and START) and the $-\log_{10}$ Wilcoxon test p-values, where Wilcoxon test p-value > 0.05 corresponded to no evidence for difference in distribution of CpG between CHILD and START, and those between 0.05 and 1×10^{-5} , 1×10^{-5} and 1×10^{-10} , and $< 1 \times 10^{-10}$ as small, medium and large difference. Each CpG that had been mapped to a gene using the "sesame" annotation package was labeled with the corresponding gene. The solid gray line is the best fitted line for the linear relationship between the effect sizes and the dashed gray line represents the reference of $y=x$

Additional file5 Figure S4. Distributions and scatterplots of CpGs association effect sizes identified in CHILD, START, and Bohlin. Panel A) shows the distribution of estimated association effects for CpGs that were significant in either CHILD or START (#CpG = 3164); Panel B) is the scatterplot of estimated effects in CHILD (x-axis) vs. START (y-axis) for all CpGs present in CHILD, START, and Bohlin (# CpGs = 11337); Panels C) and D) show the scatterplot of estimated effects in CHILD (Panel C) or START (Panel D) with respect to those identified in Bohlin. The solid gray line is the best fitted line for the linear relationship between the effect sizes and the dashed gray line represents the reference of $y=x$

Additional file6 Figure S5. Distributions of the number of CpGs and DMRs. Panel A) shows the distributions of the number of CpGs that were in the significant DMRs identified in CHILD and START; Panel B) is a scatterplot of the number of DMRs identified in CHILD (x-axis) vs. START (y-axis) across the chromosomes. The dashed line represents the reference of $y=x$

Additional file7 Figure S6. Overlap of Significant CpG-Mapped Genes with Known Type 2 Diabetes and Gestational Diabetes Genes Identified in GWAS Studies. This karyogram illustrates the chromosomal distribution of previously identified T2D or GDM genes from GWAS catalog in the human genome. Each chromosome is represented by a distinct bar, with the p-arm on the left and the q-arm on the right. Highlighted bands on each chromosome indicate the cytogenetic bands that

have been stained and the centromeres are highlighted in red. Genes of interest are marked with black vertical lines along the chromosomes, with only the names of genes that were mapped to CpGs identified in either START (red) or CHILD (green) or both (black) annotated to their respective locations

Additional file8 Figure S7. Scatterplots of maternal age and DNA methylation predicted gestational age. Panel A) shows the relationship stratified by parity with p-value for each level displayed in the left margin, Panel B) shows the overall relationship and the estimated linear model

Acknowledgements

We express our sincere gratitude to all the participating families and the START and CHILD study teams, including interviewers, nurses, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, and receptionists. We thank the CHILD Cohort Study (CHILD) participant families for their dedication and commitment to advancing health research. CHILD was initially funded by CIHR and AllerGen NCE. Visit CHILD at childstudy.ca. We would like to acknowledge the Genetic and Molecular Epidemiology Laboratory (GMEL), an associate of Hamilton Health Sciences and McMaster University, for their indispensable contributions to this work. The technical staff of GMEL conducted all epigenetic profiling, including sample processing and other technical operations. We thank the members of the Nutrigen Alliance for providing the data: Sonia S. Anand; Stephanie A. Atkinson; Meghan Azad; Allan B. Becker; Jeffrey Brook; Judah A. Denburg; Dipika Desai; Russell J. de Souza; Milan K. Gupta; Michael Kobor; Diana L. Lefebvre; Wendy Lou; Piushkumar J. Mandhane; Sarah McDonald; Andrew Mente; David Meyre; Theo J. Moraes; Katherine M. Morrison; Guillaume Paré; Malcolm R. Sears; Padmaja Subbarao; Koon K. Teo; Stuart E. Turvey; Julie Wilson; Salim Yusuf; Gita Wahi; Michael A. Zylinski.

Author contributions

WQD conceptualized the study, designed the methodology, performed data analyses and visualization, and prepared the first draft and subsequent revisions. MP, SMA, SLW, KMM, GP and SAA contributed to the interpretation of data and results and substantively revised the manuscript. NCawte contributed to methylation data curation, pre-processing, and quality control methodology. NC was responsible for phenotypic data curation and pre-processing. PS, SET, TJM, PM, MBA, and ES contributed to data curation, the interpretation of data, and revision. SMA, NC, and SSA contributed to project administration. GP and SSA contributed to study conceptualization and data curation. SSA was responsible for funding acquisition and resources. All authors edited and critically reviewed the manuscript.

Funding

This study was funded by the Canadian Institutes of Health Research Metabolics Team Grant: MWG-146332. The Canadian Institutes of Health Research (CIHR), Debbie and Don Morrison and the Allergy, Genes, and Environment Network of Centres of Excellence (AllerGen NCE), Women's and Children's Health Research Institute provided core support for establishing the CHILD Study. Dr. Anand is supported by a Tier 1 Canada Research Chair in Ethnicity and CVD and Heart, Stroke Foundation Chair in Population Health, a grant from the Canadian Partnership Against Cancer, Heart and Stroke Foundation of Canada and Canadian Institutes of Health Research. Dr. Azad is supported by a Tier 2 Canada Research Chair in the Developmental Origins of Chronic Disease. Dr. Subbarao is supported by a Tier 1 Canada Research Chair in Pediatric Asthma and Lung Health. Dr. Turvey is supported by a Tier 1 Canada Research Chair in Pediatric Precision Health.

Availability of data and materials

The EWAS summary statistics are available from the EWAS catalog at https://www.ewascatalog.org/?study=27717397_gestational_age_ultrasoundestimated_gestational_age. Summary statistics generated in the current study, including a total of 2 primary association studies and the meta-analyzed results are available upon request. All scripts to reproduce the results and figures can be found at <https://github.com/WeiAkaneDeng/EpigeneticResearch/tree/main/EpigeneticGA>.

Declarations

Ethics approval and consent to participate

Ethical approval was obtained independently from the Hamilton Integrated Research Ethics Board: CHILD (REB 07–2929) and START (REB 10–640). CHILD was additionally approved by the respective Human Research Ethics Boards at McMaster University, the Universities of Manitoba, Alberta, and British Columbia, and the Hospital for Sick Children. Legal guardians of each participant provided written informed consent. Written informed consent was obtained from the parent/guardian (participating mother) for each study separately. We also have now obtained additional ethics board approval from HIREB (REB 16592) for using the data from the two cohorts together without additional consent from the participants.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Peter Boris Centre for Addictions Research, St. Joseph's Healthcare Hamilton, Hamilton, Canada. ²Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Canada. ³Department of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, Canada. ⁴Population Health Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, Hamilton, Canada. ⁵Thrombosis and Atherosclerosis Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, Hamilton, ON, Canada. ⁶Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada. ⁷Department of Obstetrics and Gynecology, McMaster University, Hamilton, Canada. ⁸Department of Pediatrics, McMaster University, Hamilton, Canada. ⁹Hospital for Sick Children, Department of Pediatrics, University of Toronto, Toronto, Canada. ¹⁰Program in Translational Medicine, SickKids Research Institute, Toronto, Canada. ¹¹Department of Pediatrics, BC Children's Hospital, The University of British Columbia, Vancouver, Canada. ¹²Department of Pediatrics, University of Alberta, Edmonton, Canada. ¹³Department of Pediatrics and Child Health, Children's Hospital Research Institute of Manitoba, University of Manitoba, Winnipeg, Canada. ¹⁴Section of Allergy and Immunology, Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, Canada. ¹⁵Department of Pathology and Molecular Medicine, Michael G. DeGroote School of Medicine, McMaster University, Hamilton, Canada.

Received: 28 February 2024 Accepted: 23 May 2024

Published online: 06 June 2024

References

- Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol.* 2013;14(10):115.
- Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda SV, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell.* 2013;49(2):359–67.
- Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinsheimer JS, Horvath S, et al. Epigenetic predictor of age. *PLoS ONE.* 2011;6(6):e1482.
- McEwen LM, O'Donnell KJ, McGill MG, Edgar RD, Jones MJ, MacIsaac JL, et al. The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. *Proc Natl Acad Sci U S A.* 2020;117(38):23329–35.
- Knight AK, Craig JM, Theda C, Bækvad-Hansen M, Bybjerg-Grauholm J, Hansen CS, et al. An epigenetic clock for gestational age at birth based on blood methylation data. *Genome Biol.* 2016. <https://doi.org/10.1186/s13059-016-1068-z>.
- Lee Y, Choufani S, Weksberg R, Wilson SL, Yuan V, Burt A, et al. Placental epigenetic clocks: Estimating gestational age using placental DNA methylation levels. *Aging.* 2019;11(12):4238–53.
- Kennedy BK, Berger SL, Brunet A, Campisi J, Cuervo AM, Epel ES, et al. Geroscience: linking aging to chronic disease. *Cell.* 2014;159(4):709–13.
- Franceschi C, Garagnani P, Morsiani C, Conte M, Santoro A, Grignolio A, et al. The continuum of aging and age-related diseases: Common mechanisms but different rates. *Front Med.* 2018. <https://doi.org/10.3389/fmed.2018.00061>.
- McMillen IC, Robinson JS. Developmental origins of the metabolic syndrome: Prediction, plasticity, and programming. *Physiol Rev.* 2005;85(2):571–633.
- Barker DJP. The origins of the developmental origins theory. *J Intern Med.* 2007;261(5):412–7.
- Barker DJP, Eriksson JG, Forsén T, Osmond C. Fetal origins of adult disease: Strength of effects and biological basis. *Int J Epidemiol.* 2002;31(6):1235–9.
- Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. *The Lancet.* 2008;371(9606):75–84.
- Behrman RE, Butler AS. Preterm Birth: Causes, Consequences, and Prevention. *Preterm Birth: Causes, Consequences, and Prevention.* 2007;1–772.
- Liang L, Rasmussen MLH, Piening B, Shen X, Chen S, Röst H, et al. Metabolic dynamics and prediction of gestational age and time to delivery in pregnant women. *Cell.* 2020;181(7):1680–1692.e15.
- Cardenas A, Lutz SM, Everson TM, Perron P, Bouchard L, Hivert MF. Mediation by placental DNA methylation of the association of prenatal maternal smoking and birth weight. *Am J Epidemiol.* 2019;188(11):1878–86.
- Everson TM, Vives-Usano M, Seyve E, Cardenas A, Lacasaña M, Craig JM, et al. Placental DNA methylation signatures of maternal smoking during pregnancy and potential impacts on fetal growth. *Nat Commun.* 2021. <https://doi.org/10.1038/s41467-021-24558-y>.
- van Otterdijk SD, Binder AM, Michels KB. Locus-specific DNA methylation in the placenta is associated with levels of pro-inflammatory proteins in cord blood and they are both independently affected by maternal smoking during pregnancy. *Epigenetics.* 2017;12(10):875–85.
- Dieckmann L, Lahti-Pulkkinen M, Kvist T, Lahti J, DeWitt PE, Cruceanu C, et al. Characteristics of epigenetic aging across gestational and perinatal tissues. *Clin Epigenetics.* 2021. <https://doi.org/10.1186/s13148-021-01080-y>.
- Nakamura A, François O, Lepeule J. Epigenetic alterations of maternal tobacco smoking during pregnancy: a narrative review. *Int J Environ Res Public Health.* 2021;18(10):5083.
- Witt SH, Frank J, Gilles M, Lang M, Treutlein J, Streit F, et al. Impact on birth weight of maternal smoking throughout pregnancy mediated by DNA methylation. *BMC Genomics.* 2018. <https://doi.org/10.1186/s12864-018-4652-7>.
- Markunas CA, Xu Z, Harlid S, Wade PA, Lie RT, Taylor JA, et al. Identification of DNA Methylation Changes In Newborns Related To Maternal Smoking During Pregnancy. *Environ Health Perspect.* 2014;122(10):1147–53.
- Zhang B, Hong X, Ji H, Tang WY, Kimmel M, Ji Y, et al. Maternal smoking during pregnancy and cord blood DNA methylation: new insight on sex differences and effect modification by maternal folate levels. *Epigenetics.* 2018;13(5):505–18.
- Küpers LK, Xu X, Jankipersadsing SA, Vaez A, La Bastide-van GS, Scholtens S, et al. DNA methylation mediates the effect of maternal smoking during pregnancy on birthweight of the offspring. *Int J Epidemiol.* 2015;44(4):1222–37.
- Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, et al. DNA methylation in newborns and maternal smoking in pregnancy: genome-wide consortium meta-analysis. *Am J Hum Genet.* 2016;98(4):680–96.
- Xu R, Hong X, Zhang B, Huang W, Hou W, Wang G, et al. DNA methylation mediates the effect of maternal smoking on offspring birthweight: a birth cohort study of multi-ethnic US mother–newborn pairs. *Clin Epigenetics.* 2021. <https://doi.org/10.1186/s13148-021-01032-6>.
- Antoun E, Titcombe P, Dalrymple K, Kitaba NT, Barton SJ, Flynn A, et al. DNA methylation signatures in cord blood associated with birthweight are enriched for dmCpGs previously associated with maternal

- hypertension or pre-eclampsia, smoking and folic acid intake. *Epigenetics*. 2022;17(4):405–21.
27. Canouil M, Khamis A, Keikkala E, Hummel S, Lobbens S, Bonnefond A, et al. Epigenome-wide association study reveals methylation loci associated with offspring gestational diabetes mellitus exposure and maternal methylome. *Diabetes Care*. 2021;44(9):1992–9.
 28. Howe CG, Cox B, Fore R, Jungius J, Kvist T, Lent S, et al. Maternal gestational diabetes mellitus and newborn DNA methylation: findings from the pregnancy and childhood epigenetics consortium. *Diabetes Care*. 2020;43(1):98–105.
 29. Shiao S, Wang L, Liu H, Zheng Y, Drong A, Joyce BT, et al. Prenatal gestational diabetes mellitus exposure and accelerated offspring DNA methylation age in early childhood. *Epigenetics*. 2020;16(2):186–95.
 30. Küppers LK, Fernández-Barrés S, Mancano G, Johnson L, Ott R, Vioque J, et al. Maternal dietary glycemic index and glycemic load in pregnancy and offspring cord blood DNA methylation. *Diabetes Care*. 2022;45(8):1822–32.
 31. Monasso GS, Voortman T, Felix JF. Maternal plasma fatty acid patterns in mid-pregnancy and offspring epigenetic gestational age at birth. *Epigenetics*. 2022;17(11):1562–72.
 32. Lecorguille M, McAuliffe FM, Twomey PJ, Viljoen K, Mehegan J, Kelleher CC, et al. Maternal glycaemic and insulinemic status and newborn DNA methylation: findings in women with overweight and obesity. *J Clin Endocrinol Metabol*. 2023;108(1):85–98.
 33. Nakamura AL, Pryor L, Ballon M, Lioret S, Heude B, Charles MA, et al. Maternal education and offspring birth weight for gestational age: the mediating effect of smoking during pregnancy. *Eur J Public Health*. 2020;30(5):1001–6.
 34. Sharp GC, Arathimos R, Reese SE, Page CM, Felix J, Küppers LK, et al. Maternal alcohol consumption and offspring DNA methylation: findings from six general population-based birth cohorts. *Epigenomics*. 2018;10(1):27–42.
 35. Küppers LK, Fernández-Barrés S, Nounu A, Friedman C, Fore R, Mancano G, et al. Maternal Mediterranean diet in pregnancy and newborn DNA methylation: a meta-analysis in the PACE Consortium. *Epigenetics*. 2022;17(11):1419–31.
 36. Sharp GC, Salas LA, Monnereau C, Allard C, Yousefi P, Everson TM, et al. Maternal BMI at the start of pregnancy and offspring epigenome-wide DNA methylation: Findings from the pregnancy and childhood epigenetics (PACE) consortium. *Hum Mol Genet*. 2017;26(20):4067–85.
 37. Sbihi H, Jones MJ, MacIsaac JL, Brauer M, Allen RW, Sears MR, et al. Prenatal exposure to traffic-related air pollution, the gestational epigenetic clock, and risk of early-life allergic sensitization. *J Allerg Clin Immunol*. 2019;144(6):1729–31.
 38. Monasso GS, Jaddoe VVW, Küppers LK, Felix JF. Epigenetic age acceleration and cardiovascular outcomes in school-age children: the generation R study. *Clin Epigenetics*. 2021. <https://doi.org/10.1186/s13148-021-01193-4>.
 39. Kanney N, Patki A, Chandler-Laney P, Garvey WT, Hidalgo BA. Epigenetic age acceleration in mothers and offspring 4–10 years after a pregnancy complicated by gestational diabetes and obesity. *Metabolites*. 2022;12(12):1226.
 40. Murray R, Kitaba N, Antoun E, Titcombe P, Barton S, Cooper C, et al. Influence of maternal lifestyle and diet on perinatal DNA methylation signatures associated with childhood arterial stiffness at 8 to 9 years. *Hypertension*. 2021;78(3):787–800.
 41. Rzehak P, Saffery R, Reischl E, Covic M, Wahl S, Grote V, et al. Maternal smoking during pregnancy and DNA-methylation in children at age 5.5 years: Epigenome-wide-analysis in the European Childhood Obesity Project (CHOP)-study. *PLoS ONE*. 2016;11(5):e015554.
 42. Richmond RC, Simpkin AJ, Woodward G, Gaunt TR, Lyttleton O, McArdle WL, et al. Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the avon longitudinal study of parents and children (ALSPAC). *Hum Mol Genet*. 2015;24(8):2201–17.
 43. Khouja JN, Simpkin AJ, O’Keeffe LM, Wade KH, Houtepen LC, Relton CL, et al. Epigenetic gestational age acceleration: a prospective cohort study investigating associations with familial, sociodemographic and birth characteristics. *Clin Epigenetics*. 2018;10(1):1–9.
 44. Euclydes V, Gomes C, Gouveia G, Gastaldi VD, Feltrin AS, Camilo C, et al. Gestational age acceleration is associated with epigenetic biomarkers of prenatal physiologic stress exposure. *Clin Epigenetics*. 2022;14(1):152.
 45. Marques I, Santos S, Monasso GS, Fossati S, Vrijheid M, Nieuwenhuijsen M, et al. Associations of green and blue space exposure in pregnancy with epigenetic gestational age acceleration. *Epigenetics*. 2023;18(1):216531.
 46. Cardenas A, Ecker S, Fadadu RP, Huen K, Orozco A, McEwen LM, et al. Epigenome-wide association study and epigenetic age acceleration associated with cigarette smoking among Costa Rican adults. *Sci Rep*. 2022. <https://doi.org/10.1038/s41598-022-08160-w>.
 47. Bozack AK, Rifas-Shiman SL, Gold DR, Laubach ZM, Perng W, Hivert MF, et al. DNA methylation age at birth and childhood: performance of epigenetic clocks and characteristics associated with epigenetic age acceleration in the Project Viva cohort. *Clin Epigenetics*. 2023. <https://doi.org/10.1186/s13148-023-01480-2>.
 48. Ladd-Acosta C, Vang E, Barrett ES, Bulka CM, Bush NR, Cardenas A, et al. Analysis of pregnancy complications and epigenetic gestational age of newborns. *JAMA Netw Open*. 2023;6(2):e230672.
 49. Girchenko P, Lahti J, Czamara D, Knight AK, Jones MJ, Suarez A, et al. Associations between maternal risk factors of adverse pregnancy and birth outcomes and the offspring epigenetic clock of gestational age at birth. *Clin Epigenetics*. 2017. <https://doi.org/10.1186/s13148-017-0349-z>.
 50. Mayne BT, Leemaqz SY, Smith AK, Breen J, Roberts CT, Bianco-Miotto T. Accelerated placental aging in early onset preeclampsia pregnancies identified by DNA methylation. *Epigenomics*. 2017;9(3):279–89.
 51. Bohlin J, Håberg SE, Magnus P, Reese SE, Gjessing HK, Magnus MC, et al. Prediction of gestational age based on genome-wide differentially methylated regions. *Genome Biol*. 2016. <https://doi.org/10.1186/s13059-016-1063-4>.
 52. Haftorn KL, Lee Y, Denault WRP, Page CM, Nustad HE, Lyle R, et al. An EPIC predictor of gestational age and its application to newborns conceived by assisted reproductive technologies. *Clin Epigenetics*. 2021. <https://doi.org/10.1186/s13148-021-01055-z>.
 53. Hannon E, Schendel D, Ladd-Acosta C, Grove J, Hansen CS, Hougaard DM, et al. Variable DNA methylation in neonates mediates the association between prenatal smoking and birth weight. *Philos Trans R Soc B Biol Sci*. 2019;374(1770):20180120.
 54. Patel RR, Steer P, Doyle P, Little MP, Elliott P. Does gestation vary by ethnic group? A London-based study of over 122 000 pregnancies with spontaneous onset of labour. *Int J Epidemiol*. 2004;33(1):107–13.
 55. Hinkle SN, Schisterman EF, Liu D, Pollack AZ, Yeung EH, Mumford SL, et al. Pregnancy complications and long-term mortality in a diverse cohort. *Circulation*. 2023;147(13):1014–25.
 56. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, et al. Heart disease and stroke statistics—2021 update. *Circulation*. 2021. <https://doi.org/10.1161/CIR.0000000000000950>.
 57. Lamri A, Limbachia J, Schulze KM, Desai D, Kelly B, de Souza RJ, et al. The genetic risk of gestational diabetes in South Asian women. *Elife*. 2022. <https://doi.org/10.7554/eLife.81498>.
 58. Rana A, de Souza RJ, Kandasamy S, Lear SA, Anand SS. Cardiovascular risk among South Asians living in Canada: a systematic review and meta-analysis. *CMAJ Open*. 2014;2(3):E183–91.
 59. Greenhalgh T, Clinch M, Afsar N, Choudhury Y, Sudra R, Campbell-Richards D, et al. Socio-cultural influences on the behaviour of South Asian women with diabetes in pregnancy: Qualitative study using a multi-level theoretical approach. *BMC Med*. 2015. <https://doi.org/10.1186/s12916-015-0360-1>.
 60. Anand SS, Gupta MK, Schulze KM, Desai D, Abdalla N, Wahi G, et al. What accounts for ethnic differences in newborn skinfold thickness comparing South Asians and White Caucasians? Findings from the START and FAMILY Birth Cohorts. *Int J Obes*. 2016;40(2):239–44.
 61. Farrar D, Fairley L, Santorelli G, Tuffnell D, Sheldon TA, Wright J, et al. Association between hyperglycaemia and adverse perinatal outcomes in south Asian and white British women: analysis of data from the Born in Bradford cohort. *Lancet Diabetes Endocrinol*. 2015;3(10):795.
 62. Subbarao P, Anand SS, Becker AB, Befus AD, Brauer M, Brook JR, et al. The Canadian Healthy Infant Longitudinal Development (CHILD)

- study: Examining developmental origins of allergy and asthma. *Thorax*. 2015;70(10):998–1000.
63. Anand SS, Vasudevan A, Gupta M, Morrison K, Kurpad A, Teo KK, et al. Rationale and design of South Asian Birth Cohort (START): a Canada-India collaborative study. *BMC Public Health*. 2013. <https://doi.org/10.1186/1471-2458-13-79>.
 64. Deng WQ, Cavte N, Campbell N, Azab SM, Souza RJ de, Lamri A, et al. Maternal smoking DNA methylation risk score associated with health outcomes in offspring of European and South Asian ancestry. *Elife*. 2024;13.
 65. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. *Genomics*. 2011;98(4):288–95.
 66. Zhou W, Triche TJ, Laird PW, Shen H. SeSAMe: Reducing artifactual detection of DNA methylation by Infinium BeadChips in genomic deletions. *Nucleic Acids Res*. 2018. <https://doi.org/10.1093/nar/gky691>.
 67. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 2013;8(2):203–9.
 68. Okamura K, Kawai T, Hata K, Nakabayashi K. Lists of HumanMethylation450 BeadChip probes with nucleotide-variant information obtained from the Phase 3 data of the 1000 Genomes Project. *Genom Data*. 2016;7:67–9.
 69. Gervin K, Salas LA, Bakulski KM, Van Zelm MC, Koestler DC, Wiencke JK, et al. Systematic evaluation and validation of reference and library selection methods for deconvolution of cord blood DNA methylation data. *Clin Epigenetics*. 2019. <https://doi.org/10.1186/s13148-019-0717-y>.
 70. Kelemen LE, Anand SS, Vuksan V, Yi Q, Teo KK, Devanesan S, et al. Development and evaluation of cultural food frequency questionnaires for South Asians, Chinese, and Europeans in North America. *J Am Diet Assoc*. 2003;103(9):1178–84.
 71. de Souza RJ, Zulyniak MA, Desai D, Shaikh MR, Campbell NC, Lefebvre DL, et al. Harmonization of food-frequency questionnaires and dietary pattern analysis in 4 ethnically diverse birth cohorts. *J Nutr*. 2016;146(11):2343–50.
 72. Anand SS, Razak F, Davis AD, Jacobs R, Vuksan V, Teo K, et al. Social disadvantage and cardiovascular disease: development of an index and analysis of age, sex, and ethnicity effects. *Int J Epidemiol*. 2006;35(5):1239–45.
 73. Raynor P, Duley L, Small N, Tuffnell D, Wild C, Wright J, et al. Born in Bradford, a cohort study of babies born in Bradford, and their parents: protocol for the recruitment phase. *BMC Public Health*. 2008. <https://doi.org/10.1186/1471-2458-8-327>.
 74. Wright J, Small N, Raynor P, Tuffnell D, Bhopal R, Cameron N, et al. Cohort profile: the born in Bradford multi-ethnic family cohort study. *Int J Epidemiol*. 2013;42(4):978–91.
 75. Galton F. Regression towards mediocrity in hereditary stature. *J Anthropol Inst Great Brit Ireland*. 1886;15:246.
 76. Price EM, Robinson WP. Adjusting for batch effects in DNA methylation microarray data, a lesson learned. *Front Genet*. 2018. <https://doi.org/10.3389/fgene.2018.00083>.
 77. Chen B, Craiu RV, Strug LJ, Sun L. The X factor: a robust and powerful approach to X-chromosome-inclusive whole-genome association studies. *Genet Epidemiol*. 2021;45(7):694–709.
 78. Peters TJ, Buckley MJ, Statham AL, Pidsley R, Samaras K, Lord VR, et al. De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin*. 2015. <https://doi.org/10.1186/1756-8935-8-6>.
 79. Kolberg L, Raudvere U, Kuzmin I, Adler P, Vilo J, Peterson H. G:Profiler-interop-erable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res*. 2023;51(W1):W207–12.
 80. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res*. 2023;51(1D):D977–85.
 81. Pelegi-Siso D, De Prado P, Ronkainen J, Bustamante M, Gonzalez JR. Methylclock: a Bioconductor package to estimate DNA methylation age methylclock. *Bioinformatics*. 2021;37(12):1759–60.
 82. Stekhoven DJ, Bühlmann P. Missforest-Non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012;28(1):112–8.
 83. R Core Team. R core team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>. 2021.
 84. Bulik-Sullivan B, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291–5.
 85. Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, et al. Genomic inflation factors under polygenic inheritance. *Eur J Human Genet*. 2011;19(7):807–12.
 86. Küpers LK, Monnereau C, Sharp GC, Yousefi P, Salas LA, Ghantous A, et al. Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight. *Nat Commun*. 2019;10(1):1893.
 87. Vujkovic M, Keaton JM, Lynch JA, Miller DR, Zhou J, Tcheandjieu C, et al. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 14 million participants in a multi-ancestry meta-analysis. *Nat Genet*. 2020;52(7):680–91.
 88. Pervjakova N, Moen GH, Borges MC, Ferreira T, Cook JP, Allard C, et al. Multi-ancestry genome-wide association study of gestational diabetes mellitus highlights genetic links with type 2 diabetes. *Hum Mol Genet*. 2022;31(19):3377–91.
 89. Hattersley AT, Tooke JE. The fetal insulin hypothesis: an alternative explanation of the association of low birthweight with diabetes and vascular disease. *Lancet*. 1999;353(9166):1789–92.
 90. Kang HS, Kim YS, ZeRuth G, Beak JY, Gerrish K, Kilic G, et al. Transcription factor glis3, a novel critical player in the regulation of pancreatic β -cell development and insulin gene expression. *Mol Cell Biol*. 2009;29(24):6366–79.
 91. Senée V, Chelala C, Duchatelet S, Feng D, Blanc H, Cossec JC, et al. Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. *Nat Genet*. 2006;38(6):682–7.
 92. Sakae S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet*. 2021;53(10):1415–24.
 93. Feofanova EV, Chen H, Dai Y, Jia P, Grove ML, Morrison AC, et al. A genome-wide association study discovers 46 loci of the human metabolome in the hispanic community health study/study of Latinos. *Am J Hum Genet*. 2020;107(5):849–63.
 94. Boney CM, Verma A, Tucker R, Vohr BR. Metabolic syndrome in childhood: association with birth weight, maternal obesity, and gestational diabetes mellitus. *Pediatrics*. 2005;115(3):e290–6.
 95. Rkzhay-Jaf J, O'Dowd JF, Stocker CJ. Maternal Obesity and the Fetal Origins of the Metabolic Syndrome. *Curr Cardiovasc Risk Rep*. 2012;6(5):487–95.
 96. Armitage J, Poston L, Taylor P. Developmental origins of obesity and the metabolic syndrome: the role of maternal obesity. *Front Hormone Res*. 2007;36:73–84.
 97. Elliott HR, Burrows K, Min JL, Tillin T, Mason D, Wright J, et al. Characterisation of ethnic differences in DNA methylation between UK-resident South Asians and Europeans. *Clin Epigenetics*. 2022. <https://doi.org/10.1186/s13148-022-01351-2>.
 98. Frederiksen LE, Ernst A, Brix N, Lauridsen LLB, Roos L, Ramlau-Hansen CH, et al. Risk of adverse pregnancy outcomes at advanced maternal age. *Obstet Gynecol*. 2018;131(3):457–63.
 99. Londero AP, Rossetti E, Pittini C, Cagnacci A, Driul L. Maternal age and the risk of adverse pregnancy outcomes: a retrospective cohort study. *BMC Pregnancy Childbirth*. 2019. <https://doi.org/10.1186/s12884-019-2400-x>.
 100. Flenady V, Koopmans L, Middleton P, Frøen JF, Smith GC, Gibbons K, et al. Major risk factors for stillbirth in high-income countries: A systematic review and meta-analysis. *The Lancet*. 2011;377(9774):1331–40.
 101. Horvath S, Gurven M, Levine ME, Trumble BC, Kaplan H, Allayee H, et al. An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol*. 2016. <https://doi.org/10.1186/s13059-016-1030-0>.
 102. Reitsma MB, Flor LS, Mullany EC, Gupta V, Hay SI, Gakidou E. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and initiation among young people in 204 countries and territories, 1990–2019. *Lancet Public Health*. 2021;6(7):e472–81.
 103. McIntyre HD, Catalano P, Zhang C, Desoye G, Mathiesen ER, Damm P. Gestational diabetes mellitus. *Nat Rev Dis Prim*. 2019. <https://doi.org/10.1038/s41572-019-0098-8>.
 104. Lohoff FW, Clarke TK, Kaminsky ZA, Walker RM, Birmingham ML, Jung J, et al. Epigenome-wide association study of alcohol consumption in N = 8161 individuals and relevance to alcohol use disorder pathophysiology:

identification of the cystine/glutamate transporter SLC7A11 as a top target. *Mol Psychiatry*. 2022;27(3):1754–64.

105. Song AY, Feinberg JI, Bakulski KM, Croen LA, Fallin MD, Newschaffer CJ, et al. Prenatal exposure to ambient air pollution and epigenetic aging at birth in newborns. *Front Genet*. 2022. <https://doi.org/10.3389/fgene.2022.929416>.
106. Gruziova O, Xu CJ, Breton CV, Annesi-Maesano I, Antó JM, Auffray C, et al. Epigenome-wide meta-analysis of methylation in children related to prenatal NO₂ air pollution exposure. *Environ Health Perspect*. 2017;125(1):104–10.

Publisher's Note

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