medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Maternal smoking DNA methylation risk score associated with health outcomes in offspring of European and South Asian ancestry

Wei Q. Deng[1,2,3*], Nathan Cawte[4], Natalie Campbell[3], Sandi M. Azab[3,5], Russell J de Souza[3,5], Amel Lamri[3,4], Katherine M. Morrison[6], Stephanie A. Atkinson[6], Padmaja Subbarao[7], Stuart E. Turvey[8], Theo J. Moraes[7,9], Koon K. Teo[3,4,5], Piush Mandhane[10], Meghan B. Azad[11], Elinor Simons[12], Guillaume Pare[4,5,13,14], Sonia S. Anand [3,4,5*]

1. Peter Boris Centre for Addictions Research, St. Joseph's Healthcare Hamilton, Hamilton, Canada.
2. Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Canada.
3. Department of Medicine, Faculty of Health Sciences, McMaster University, Hamilton, Canada.
4. Population Health Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, Hamilton, Canada.
5. Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada.
6. Department of Pediatrics, McMaster University, Hamilton, Canada.
7. Department of Pediatrics, University of Toronto, Toronto, Canada.
8. Department of Pediatrics, BC Children's Hospital, The University of British Columbia, Vancouver, Canada

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

9. Program in Translational Medicine, SickKids Research Institute, Toronto, Canada
10. Department of Pediatrics, University of Alberta, Edmonton, Canada
11. Children's Hospital Research Institute of Manitoba, Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, Canada
12. Section of Allergy and Immunology, Department of Pediatrics and Child Health, University of Manitoba, Canada
13. Thrombosis and Atherosclerosis Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, Hamilton, Canada
14. Department of Pathology and Molecular Medicine, McMaster University, Michael G. DeGroote School of Medicine, Hamilton, Canada

*Correspondence: dengwq@mcmaster.ca; anands@mcmaster.ca

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Maternal smoking has been linked to adverse health outcomes in newborns but the extent to which it impacts newborn health has not been quantified through an aggregated cord blood DNA methylation (DNAm) score. Here we propose the first cord blood DNAm score to capture the epigenetic signature of maternal smoking. We first examined the association between individual CpGs and cigarette smoking during pregnancy in two White European birth cohorts (n = 763), and smoking exposure in the White European and the South Asian birth cohorts (n = 887). Several previously reported genes for maternal smoking were supported, with the strongest and most consistent signal from the *GFI1* gene ($ps < 5\times10^{-5}$). Leveraging previously reported CpGs for maternal smoking, we constructed a cord blood epigenetic score of maternal smoking that was internally validated in one of the European-origin cohorts. This score was then tested for association with smoking status, secondary smoking exposure during pregnancy, and health outcomes in offspring measured after birth. The epigenetic maternal smoking score was strongly associated with smoking status during pregnancy ($p=1.04\times10^{-23}$) and self-reported smoking exposure ($p=6.56\times10^{-8}$) in White Europeans, but not with self-reported exposure ($p > 0.05$) in South Asians. The same score was consistently associated with smaller birth size ($p =1.19\times10^{-6}$) and lower birth weight ($p =8.79\times10^{-7}$) in the combined South Asian and White European cohorts. This cord blood epigenetic score can help identify babies exposed to maternal smoking and assess its long-term impact on growth. Notably, even minimal smoking exposure in South Asian mothers who were not active smokers showed a DNAm signature of small body size and low birthweight in newborns.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

Maternal smoking has adverse effects on offspring health including pre-term delivery (1,2), stillbirth (3), and low birth weight (4), and is associated with pregnancy complications such as maternal higher blood pressure, and gestational diabetes (5). Consistent with the Developmental Origins of Health and Disease (DOHaD) hypothesis, maternal smoking exposes the developing fetus to harmful chemicals in tobacco that negatively impact the health of newborns, resulting in early-onset metabolic diseases, such as childhood obesity (6–9). Yet self-reported smoking status is subject to underreporting among pregnant women (10–12). This could subsequently impact the effectiveness of interventions aimed at reducing smoking during pregnancy and may skew data on the risks associated with maternal smoking.

Differential DNA methylation has been established as a reliable biochemical response to cigarette smoking and was shown to capture the long-lasting effects of persistent smoking in ex-smokers (13–15). Recent large epigenome-wide association studies (EWAS) have robustly identified differentially methylated cytosine–phosphate–guanine (CpG) sites associated with adult smoking (14,16,17) and maternal smoking (18,19). Our recent systematic review of 17 cord blood epigenome-wide association studies (EWAS) demonstrated that out of the 290 CpG sites reported, 19 sites were identified in more than one study; all of them associated with maternal smoking (20). Furthermore, these findings have led to a more thorough investigation of the epigenetic mechanisms underlying associations between well-established epidemiological exposures and outcomes, such as the relationship between maternal smoking and birth weight in Europeans (19,21–24) and the less studied African American populations (23) as well as between maternal diet and cardiovascular health (25).

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The majority of cohort studies have focused on participants of European ancestry, but few were designed to assess the influence of maternal exposures on DNA methylation changes in non-Europeans (23,26). It has been suggested that ancestral background could influence both systematic patterns of methylation (27), such as cell composition and smoking behaviours (28). These systematic differences also contribute to different smoking-related methylation signals at individual CpGs (29). Thus, a comparative study of maternal smoking exposure is a first step towards generalizing existing EWAS results to other populations and a necessary step towards addressing health disparities that exist between populations due to societal privilege, including race or ethnicity and socioeconomic factors.

A promising direction in epigenetic studies of adult smoking is the application of a methylation score (30); this strategy can also be applied to disseminate current knowledge on differential DNA methylation studies of maternal smoking. A methylation score is usually tissue-specific and combines information from multiple CpGs using statistical models (31). Reducing the number of predictors and measurement noise in the data can lead to better statistical power and more interpretable effect sizes. It is also of interest to determine whether methylation scores demonstrate the capacity to predict outcomes in diverse human populations, given the presence of systematic differences in methylation patterns due to ancestral backgrounds (27).

In this paper, we investigated the epigenetic signature of maternal smoking on cord blood DNA methylation in newborns, as well as its health consequences for newborn and later life outcomes in one South Asian which refers to people who originate from the Indian subcontinent,

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

and two predominantly European-origin birth cohorts. Similar to the Born in Bradford study (32), we observed several differentiating epidemiological characteristics between South Asian and European-origin mothers. Notably, almost none of the South Asian mothers were current smokers and had low smoking rates pre-pregnancy as compared to European mothers, which is consistent with the broader trends of lower smoking rates in South Asian females (33). Another relevant observation is the small birth size and low birth weight in the South Asian newborns. These differences in newborn size and weight may be influenced by various factors, including maternal nutrition, genetics, and socioeconomic status. Keeping these differences in mind, we first conducted cohort-specific epigenetic association studies between available CpGs and maternal smoking in the predominantly European-origin cohorts, benchmarking with previously identified CpGs for maternal smoking and adult smoking. Second, we leveraged the reported summary statistics from existing large EWASs to construct a methylation risk score (MRS) for maternal smoking. The MRS was first internally validated in one of the European-origin cohorts and then tested in a second independent European-origin cohort. Third, we examined the association between maternal smoking MRS and newborn health outcomes, including length, weight, body mass index (BMI), ponderal index, and early-life anthropometrics in both European and South Asian cohorts.

Results

Cohort Sample Characteristics

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The analyses included 763 European mother-child pairs with cord blood DNAm data from the Canadian Healthy Infant Longitudinal Development (CHILD; *n* = 352) study and The Family Atherosclerosis Monitoring In earLY life (FAMILY; *n* = 411) study (34), and 880 South Asian mother-child pairs from The SouTh Asian biRth cohorT (START) study (35). We observed lower past smoking and missingness on smoking history among pregnant women in START as compared to CHILD or FAMILY using the epigenetic subsample (Table 1) and the overall sample (Supplementary Table 2). Pregnant women in START were significantly different from CHILD or FAMILY in that they were on average younger at delivery, had a lower BMI, and a higher rate of GDM, in line with other cohort studies in South Asian populations (36,37). As compared to START, newborn infants from CHILD and FAMILY had a longer gestational period, a higher birth weight, and a higher BMI at birth (Table 1; Supplementary Table 2). We observed no difference between cohorts in terms of parity or newborn sex in the epigenetic subsample (Table 1).

Within the European epigenetic subsample, of the 744 mother–newborn pairs with complete smoking history data, 40 (5.3%) newborns were exposed to current maternal smoking, which is on the lower end of the spectrum for the prevalence of smoking during pregnancy (9.2%-32.5%) among Canadians (38). In addition, mothers who smoked during pregnancy were on average younger, had fewer years of education, and had higher household exposure to smoking (Supplementary Table 4). However, there was no statistically significant difference between newborns exposed to current and none or previous smoking in terms of birth weight, birth length, gestational age, or cell compositions.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Epigenetic Association of Maternal Smoking in White Europeans

The two predominantly White European cohorts, FAMILY (n = 397) and CHILD (n = 347), contributed to the meta-analysis of maternal smoking for both the primary outcome of current smoking (Figure 1-A) and the secondary outcome of ever smoking (Supplementary Figure 1). The most significant associated CpGs with current maternal smoking were mapped to the growth factor independent 1 (*GFI1*) gene on chromosome 1, with cg12876356 as the lead (meta-analyzed $p = 2.6\times10^{-6}$; $q = 0.006$; Table 2). There were no CpGs associated with the ever-smoker status at an FDR of 0.05, though the top signal also coincided with the *GFI1* gene (Supplementary Figure 1). The meta-analysis of smoking exposure (hours per week) associations in the European-origin cohorts (Figure 1-B) identified only one CpG on chromosome 17, ccg01798813, that was also associated with maternal smoking and was consistent in the direction of association (Table 2). However, the meta-analysis of the combined samples including the South Asians from the START cohort did not yield any significant associations (Supplementary Figure 2).

The EWAS in CHILD did not yield significant CpGs after FDR correction (Supplementary Figure 3 and Supplementary Figure 4-A). The South Asian EWAS (n=504) identified 474 CpGs at FDR < 0.05, although some of these corresponded to previously identified CpGs for maternal smoking in Europeans (Supplementary Figure 4-B; highlighted in red).

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Methylation Risk Score (MRS) Captures Maternal Smoking and Smoking Exposure

The final MRSs, validated using CHILD European samples, included 11 and 114 CpG markers (Supplementary Tables 5-6) from the targeted array and the epigenome-wide HM450 array, respectively. Both produced methylation scores that were significantly associated with maternal smoking history (Supplementary Figure 5). There was no statistically significant difference between the two scores in all samples (p = 1.00) or among non-smokers (p = 0.24). Thus, we proceeded with the simpler MRS model constructed using the 11 CpGs in subsequent analyses for compatibility across cohorts.

The MRS was significantly associated with maternal smoking history in CHILD and FAMILY (Figure 2), but not in START – not surprisingly – due to the low number of ever-smokers (n = 5). A weak dose-dependent relationship between the MRS and the four categories of maternal smoking status in the severity of exposure ([0] = never smoked; [1] = quit before this pregnancy; [2] = quit during this pregnancy; [3] = currently smoking) was present in CHILD but was not replicated in FAMILY. The areas under the Receiver Operating Characteristic (ROC) curve (area under the curve or AUC) for detecting current smokers were 0.86 and 0.90 in CHILD and FAMILY, respectively, while the AUCs for detecting ever-smokers were 0.61 and 0.60, respectively. Meanwhile, the maternal smoking MRS was significantly associated with increased smoking exposure in the two White European cohorts ($p = 6.91 \times 10^{-4}$ in CHILD and $p = 1.35 \times 10^{-5}$ in FAMILY; Supplementary Table 7), but not in the South Asian birth cohort.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Among individuals who had never smoked, no statistically significant mean difference was observed in the distribution of the combined methylation score nor individual CpGs that contributed to the score between South Asian and European cohorts (Supplementary Figure 6). One CpG had a significant difference in variance (Supplementary Table 8), though it did not change the conclusion that there was no difference in mean, variance, and overall distribution of the combined methylation score. These results provided empirical support for the portability of a European-derived maternal smoking methylation score to South Asian populations.

Association between MRS and Offspring Anthropometrics

A higher maternal smoking MRS was significantly associated with lower birth weight and smaller birth size in both the European and South Asian cohorts (Table 3; Supplementary Table 7). The meta-analysis revealed no heterogeneity in the direction nor the effect size of associations between populations. The association between the MRS and several health metrics, including height or length, weight, and skinfolds, appeared to persist with similar estimated effects throughout early developmental years (Supplementary Tables 7 and 9), albeit the most significant effects were at birth, and the significance attenuated at later visits.

Discussion

We examined the epigenetic signature of maternal smoking and smoking exposure using newborn cord blood samples from predominately European-origin and South Asian cohorts via

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

two strategies: an individual CpG-level EWAS approach, and a multivariate approach in the form of a methylation score. The EWAS results replicated the association between maternal smoking and CpGs in the *GFI1* gene that is well described in the literature with respect to smoking (14,16), maternal smoking (18,19,39,40), and birth weight (22). In the latter case, we observed a significant association with maternal smoking history and smoking exposure in European-origin newborns. Further, we noted a weak dose-dependent relationship between maternal smoking history and the methylation score in one European cohort (CHILD) but this was not replicated in the other (FAMILY). Since the timing and duration of maternal smoking during pregnancy were not directly available, these differences could play a role in the magnitude and specificity of DNA methylation changes in cord blood. Finally, the significant association of the MRS with the newborn health metrics in START, in the absence of mothers' active smoking, could be the result of underreporting of smoking, poor recall of the time of quitting, and/or due to air pollution exposure (41), leading to oxidative stress. This suggests that our cord blood DNAm signature of maternal smoking is perhaps not unique to cigarette smoking, but captures similar biochemical responses, for example, via the aryl hydrocarbon receptor (42,43). Our observation that a higher MRS was associated with lower birth weight and smaller birth length in both ethnic populations is thus consistent with the established link between oxidative stress and metabolic syndrome (44).

Contrary to DNA methylation studies of smoking in adults, where whole blood is often used as a proxy tissue, there are multiple relevant tissues for maternal smoking during pregnancy, including the placenta of the mother, newborn cord blood, and children's whole blood. However, methylation changes measured in whole blood or placenta of the mother, or cord blood of infants

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

showed substantially different patterns of association signals (45). Since methylation signals are known to be tissue-specific, it would be of interest for future research to combine differential methylation patterns from all relevant tissue to assess the immediate and long-term effects of maternal smoking. Another direction to further this line of research is to explore postnatal factors that mitigate prenatal exposures, for example, breastfeeding, which has been shown to have a protective effect against maternal tobacco smoking (46). Indeed, more research is necessary to understand the critical periods of exposure and the dose-response relationship between maternal smoking and cord blood DNA methylation changes. Ongoing efforts to monitor the offspring and collect data in the next decade are in progress to establish the long-term association between maternal smoking and cardio-metabolic health (34,35). As such, the constructed MRS can facilitate future research in child health and will be included as part of the generated data for others to access.

The strengths of this report include ethnic diversity, and fine phenotyping in a prospective and harmonized way with follow-up at multiple early childhood stages. This work is the first major multi-ancestry study that utilizes methylation scores to study maternal smoking and examines their portability from European-origin populations to South Asians. The use of MRS, as compared to individual CpGs, is a powerful tool to systematically investigate the influence of DNA methylation changes and whether it has lasting functional consequences on health outcomes. Our results converge with previous findings that epigenetic associations of maternal smoking are associated with newborn health, and add to the small body of evidence that these relationships extend to non-European populations and that different ancestral populations can experience the early developmental periods differently.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

A few limitations should be mentioned. In the context of existing epigenetic studies of maternal smoking, we were not able to replicate signals in other well-reported genes such as *AHRR*, *CYP1A1*, and *MYO1G*, however, the MRS was able to pick up signals from these genes (Supplementary Table 5). This could be due to several reasons. First, the customized array was designed in 2016 and many large EWASs on smoking and maternal smoking conducted more recently had not been included. However, we have shown that from a multivariate perspective, the MRS constructed using a targeted approach that was carefully designed can be equally powerful with the advantage of being cost-effective. Second, contrary to existing EWASs where the methylation values are typically treated as the outcome, and the exposure, such as smoking, as the predictor; we reversed the regression such that the methylation levels were the predictors and smoking exposure as the outcome. This reverse regression approach is robust and our choice to reverse the regression was motivated by the goal of constructing a smoking score that combines the additive effects at multiple CpGs, which would otherwise be unfeasible. Fourth, systematic ancestral differences in DNA methylation patterns had been shown to vary at individual CpGs in terms of their association with smoking (29). Converging with this conclusion, we also found the association with *GFI1* to be most consistent after adjusting for cell composition. Finally, maternal smoking is often associated with other confounding factors, such as socioeconomic status, other lifestyle behaviours, and environmental exposures. While we have done our best to control for well-known confounders that were available by study design, as in all observational studies, we could not account for unknown confounding effects.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

In conclusion, the epigenetic maternal smoking score we constructed was strongly associated with smoking status during pregnancy and self-reported smoking exposure in White Europeans, and with smaller birth size and lower birth weight in the combined South Asian and White European cohorts. The proposed cord blood epigenetic signature of maternal smoking has the potential to identify newborns who were exposed to maternal smoking *in utero* and to assess the long-term impact of smoking exposure on offspring health. In South Asian mothers with minimal smoking behaviour, the relationship between the methylation score and negative health outcomes in newborns is still apparent, indicating that DNA methylation response is sensitive to smoking exposure, even in the absence of active smoking.

Material and Methods

Study population

The NutriGen Alliance is a consortium consisting of four prospective, population-based birth cohorts that enrolled birthing mother and newborn pairs in Canada. Details of these cohorts have been described elsewhere (47). The current investigation focused on i). European-origin offspring from the population-based CHILD study who were selected for methylation analysis, ii). The Family Atherosclerosis Monitoring In early life (FAMILY) study that is predominately European-origin, and iii). The SouTh Asian biRth cohorT (START) study that is exclusively comprised of people who originated from the Indian subcontinent known as South Asians. The ethnicity of the parents was self-reported and recorded at baseline in all three cohorts. Biological

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

samples, clinical assessments, and questionnaires were used to derive health phenotypes and an array of genetic, epigenetic, and metabolomic data. The superordinate goal of the NutriGen study is to understand how nutrition, environmental exposures, and physical health of mothers impact the health and early development of their offspring using a multi-omics approach.

Methylation data processing and quality controls

Newborn cord blood samples were processed using two methylation array technologies. About half of the START samples and selected samples from CHILD were hybridized to the Illumina Human-Methylation450K BeadChip (HM450K) array, which covers CpGs in the entire genome (48) The raw methylation data were generated by the Illumina iScan software and separately pre-processed for START and CHILD using the "*sesame*" R package following pipelines designed for HM450K BeadChip (49). The remaining START and all FAMILY samples were profiled using a targeted array based on the Infinium Methylation EPIC designed by the Genetic and Molecular Epidemiology Laboratory (GMEL; Hamilton, Canada). The GMEL customized array includes ~3000 CpG sites that were previously reported to associate with complex traits or exposures and was designed to maximize discovery while keeping the costs of profiling epigenome-wide DNA methylation down. The targeted methylation data were pre-processed using a customized quality control pipeline and functions from the "*sesame*" R package recommended for EPIC.

Pre-processed data were then used to derive the β-value matrix, where each column gives the methylation level at a CpG site as a ratio of the probe intensity to the overall probe intensity.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Additional quality control filters were applied to the final beta-value matrices to remove samples with > 10% missing probes and CpG probes with >10% samples missing. Cross-reactive probes and SNP probes were removed as recommended for HM450 (50) and EPIC arrays (51,52). For CpG probes with a missing rate <10%, mean imputation was used to fill in the missing values. We further excluded samples that were either mismatches between reported sex and methylation-inferred sex or were duplicates. Finally, considering the low prevalence of smokers, we sought to reduce spurious associations by removing non-informative probes that were either all hypomethylated (β-value < 0.1) or hypermethylated (β-value > 0.9), which have been shown to have less optimal performance (53). A summary of the sample and probe inclusion/exclusion is shown in Supplementary Table 1. A detailed description of pre-processing and quality control steps is included in Supplementary Material.

Cell-type proportions (CD8T, CD4T, Natural Killer cells, B cells, monocytes, granulocytes, and nucleated red blood cells) were estimated following a reference-based approach developed for cord blood (54) and using R packages "*FlowSorted.CordBloodCombined.450k*" and "*FlowSorted.Blood.EPIC*". All data processing and subsequent analyses were conducted in R v.4.1.0 (55).

Phenotype data processing and quality controls

At the time of enrollment, all pregnant women completed a comprehensive questionnaire that collected information on prenatal diet, smoking, education, socioeconomic factors, physical activities and health as detailed previously (34,35). Maternal smoking history (0=never smoked,

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

1=quit before this pregnancy, 2=quit during this pregnancy, or 3=current smoker) was assessed during the second trimester (at baseline). Smoke exposure was measured as "number of hours exposed per week". Gestational diabetes mellitus (GDM) was determined based on a combination of oral glucose tolerance test (OGTT), self-report, and reported diabetic treatments (insulin, pills, or restricted diet). For South Asian mothers in START, the same OGTT threshold as Born in Bradford (26,32) was used, while the International Association of the Diabetes and Pregnancy Study Groups (IASDPSG) criteria (56) for OGTT were used in CHILD and FAMILY cohorts.

All newborn anthropometric measurements, including length, weight, waist, waist-hip-ratio (WHR), BMI, were collected at the time of birth. The newborns were then followed up at 1, 2, 3, and 5 years of age and provided basic anthropometric measurements, including height, weight, BMI, sum of the skinfolds (triceps skinfold and subscapular skinfold), WHR, as well as additional environmental exposures.

Phenotype and Methylation Data Consolidation

The current investigation examines the impact of maternal smoking or smoke exposure on DNA methylation derived from newborn cord blood in START and the two predominately European cohorts (CHILD and FAMILY). To maximize sample size in FAMILY and CHILD, we retained either self-identified or genetically confirmed Europeans (Supplementary Table 2). The cohorts consist of representative population samples without enrichment for any clinical conditions, though only singleton mothers were invited to participate.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The final analytical datasets, after combining the quality-controlled methylation data and phenotypic data, included 352, 411, and 890 mother-newborn pairs from CHILD, FAMILY, and START, respectively. Demographic characteristics and relevant covariates of the epigenetic subsample and the overall sample are summarized in Table 1 and Supplementary Table 3, respectively.

Epigenome-Wide Association of Maternal Smoking in European Cohorts

Since there were no current smokers in START (Table 1), we tested the association between maternal smoking and differential methylated sites in FAMILY (# of CpG = 2,544) and CHILD (# of CpG = 358,113). The primary outcome variable was "current smoker", defined by mothers self-identified as currently smoking during the pregnancy vs. those who never smoked or quit either before or during pregnancy. As a sensitive analysis, we included a secondary outcome variable "ever smoker", defined by mothers who are current smokers or have quit smoking vs. those who never smoked. A tertiary outcome was smoking exposure, measured by the number of hours a week reported by the expectant mothers, and was available in all cohorts. We summarized the type of analyses for different outcomes in Supplementary Table 3.

We first conducted a separate epigenetic association study in each cohort, testing the association between methylation β-values at individual CpGs and the smoking phenotype using either a logistic regression model for smoking status or a linear regression for smoking exposure as the outcome. The model adjusted for additional covariates including the estimated cell

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

compositions, maternal age, social disadvantage index, which is a continuous composite measure of social and economic exposures (57), mother's years of education, GDM, and parity. The smoking exposure variable was skewed, and a rank-based transformation was applied to mimic a standard normal distribution.

We then meta-analyzed association results for maternal smoking status in the European cohorts using an inverse variance-weighted fixed-effect model. The meta-analysis was conducted for 2,112 CpGs that were available in both CHILD (HM450K) and FAMILY (GMEL-EPIC). For the tertiary outcome, we conducted a meta-analysis including START using both a fixed-effect and a random-effect model to account for the potential heterogeneity. For each EWAS or meta-analysis, false discovery rate (FDR) adjustment was used to control multiple testing and we considered an FDR-adjusted p-value < 0.05 to be statistically significant.

Using DNA Methylation to Construct Predictive Models for Maternal Smoking

We sought to construct a predictive model in the form of a methylation risk score (MRS) using reported associations of maternal smoking. The proposed solution adapted the existing lassosum method (58) that was originally designed for polygenic risk scores, where the matrix of SNP genotypes (X) can be conveniently replaced by the β-value matrix. For more details, see the Supplementary Material. Briefly, an objective function under elastic-net constraint was minimized to obtain the elastic-net solution γ, where only summary statistics (b) and a scalar of the covariance between the β-values of the CpGs ($X'X$) are needed. The tuning parameters λ_1 and λ_2 were chosen by validating on the observed smoking history (as a continuous outcome) in

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

CHILD that produced the most significant model. The optimized λ_1 and λ_2 were then used to create a final model that entails a list of CpGs and their corresponding weights, which were then used to calculate a MRS for maternal smoking in the FAMILY and START samples.

The summary statistics of the discovery EWAS were obtained from EWAS catalog (http://www.ewascatalog.org/) reported under "PubMed ID 27040690" by Joubert and colleagues (18). The summary statistics were first screened to retain CpGs with sample size above 5000, and then restricted to analysis that examined "sustained maternal smoking in pregnancy effect on newborns adjusted for cell composition". Of the 2620 maternal smoking CpGs that passed the initial screening, 1,902 were available in CHILD but only 125 were common to CHILD, FAMILY, and START. To evaluate whether the targeted GMEL-EPIC array design has comparable performance as the epigenome-wide array to evaluate the epigenetic signature of maternal smoking, two MRSs were constructed, one using the 125 CpGs available in all cohorts – across the HM450K and targeted GMEL-EPIC arrays – and another using 1,902 CpGs that were only available in CHILD and a subset of START samples.

For the MRS constructed using the 125 common CpGs, we empirically assessed whether the resulting score was portable to the START cohort by examining the distribution of the final score as well as individual CpGs that contributed to the score among never smokers in terms of mean difference using an analysis of variance *F*-test, the variance difference using a Levene's test, and the overall distribution using a non-parametric Anderson-Darling test.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Finally, we tested the association between each maternal smoking MRS and smoking phenotypes in mothers, as well as offspring phenotypes, when applicable, adjusting for the child's age at each visit. The association results were meta-analyzed for phenotypes with homogeneous effects across the cohorts using a fixed-effect model.

Table 1. Characteristics of the epigenetic subsample (1,650 mother–newborn pairs) from the CHILD, FAMILY, START cohorts.

	CHILD-HM450K* (N=352)	FAMILY-GMEL** (N=411)	START-HM450* (N=504)	START-GMEL** (N=376)	P-value
Smoking History					
never smoked	247 (70.2%)	253 (61.6%)	501 (99.4%)	371 (98.7%)	<0.001
quit before this pregnancy	72 (20.5%)	58 (14.1%)	1 (0.2%)	2 (0.5%)	
quit during this pregnancy	17 (4.8%)	57 (13.9%)	1 (0.2%)	1 (0.3%)	
currently smoking	11 (3.1%)	29 (7.1%)	0 (0%)	0 (0%)	
Missing	5 (1.4%)	14 (3.4%)	1 (0.2%)	2 (0.5%)	
Smoking Exposure (hr/week)					
Mean (SD)	0.97 (± 7.6)	0.36 (± 1.8)	0.33 (± 2.7)	0.33 (± 3.1)	<0.001
Missing	12 (3.4%)	5 (1.2%)	42 (8.3%)	7 (1.9%)	
GDM					
YES	16 (4.5%)	345 (83.9%)	183 (36.3%)	142 (37.8%)	<0.001
NO	336 (95.5%)	66 (16.1%)	320 (63.5%)	231 (61.4%)	
Missing	0 (0%)	0 (0%)	1 (0.2%)	3 (0.8%)	
Years of Education					
Mean (SD)	16 (± 2.9)	17 (± 3.4)	16 (± 2.4)	16 (± 2.2)	<0.001
Missing	7 (2.0%)	3 (0.7%)	0 (0%)	0 (0%)	
Mother's Age					
Mean (SD)	33 (± 4.5)	32 (± 5.3)	30 (± 3.9)	30 (± 3.8)	<0.001
Missing	4 (1.1%)	0 (0%)	0 (0%)	0 (0%)	
Parity					
Mean (SD)	0.72 (± 0.88)	0.80 (± 1.0)	0.80 (± 0.81)	0.74 (± 0.83)	0.186
Missing	2 (0.6%)	0 (0%)	13 (2.6%)	7 (1.9%)	
Pre-pregnancy BMI					

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Mean (SD)	25 (± 5.4)	26 (± 6.4)	24 (± 4.4)	24 (± 4.6)	<0.001
Missing	132 (37.5%)	16 (3.9%)	2 (0.4%)	2 (0.5%)	
Newborn Sex					
Male	194 (55.1%)	211 (51.3%)	239 (47.4%)	188 (50.0%)	0.168
Female	158 (44.9%)	200 (48.7%)	265 (52.6%)	188 (50.0%)	
Plant Based Diet					
Mean (SD)	-0.48 (± 0.46)	0.19 (± 0.67)	1.6 (± 1.1)	1.4 (± 1.1)	<0.001
Missing	23 (6.5%)	36 (8.8%)	16 (3.2%)	13 (3.5%)	
Health Conscious Diet					
Mean (SD)	0.21 (± 0.81)	-0.73 (± 0.73)	-0.42 (± 0.79)	-0.47 (± 0.70)	<0.001
Missing	23 (6.5%)	36 (8.8%)	16 (3.2%)	13 (3.5%)	
Western Diet					
Mean (SD)	-0.15 (± 0.63)	1.1 (± 1.2)	-0.51 (± 0.65)	-0.51 (± 0.64)	<0.001
Missing	23 (6.5%)	36 (8.8%)	16 (3.2%)	13 (3.5%)	
Gestational Age (weeks)					
Mean (SD)	40 (± 1.4)	39 (± 1.5)	39 (± 1.3)	39 (± 1.3)	<0.001
Missing	4 (1.1%)	0 (0%)	0 (0%)	0 (0%)	
Birth Length (cm)					
Mean (SD)	52 (± 2.5)	50 (± 2.2)	51 (± 2.7)	51 (± 2.4)	<0.001
Missing	71 (20.2%)	10 (2.4%)	7 (1.4%)	8 (2.1%)	
Birth Weight (kg)					
Mean (SD)	3.5 (± 0.49)	3.5 (± 0.50)	3.3 (± 0.46)	3.2 (± 0.45)	<0.001
Missing	6 (1.7%)	0 (0%)	1 (0.2%)	1 (0.3%)	
Newborn BMI (kg/m2)					
Mean (SD)	13 (± 1.4)	14 (± 1.3)	12 (± 1.4)	12 (± 1.4)	<0.001
Missing	72 (20.5%)	10 (2.4%)	7 (1.4%)	8 (2.1%)	
Newborn Ponderal Index					
Mean (SD)	25 (± 3.1)	28 (± 2.5)	24 (± 3.2)	25 (± 3.2)	<0.001
Missing	2 (0.6%)	10 (2.4%)	7 (1.4%)	8 (2.1%)	
CD8T					
Mean (SD)	0.0094 (± 0.014)	0.040 (± 0.031)	0.019 (± 0.024)	0.056 (± 0.028)	<0.001
CD4T					
Mean (SD)	0.11 (± 0.056)	0.13 (± 0.060)	0.16 (± 0.069)	0.15 (± 0.054)	<0.001
NK					
Mean (SD)	0.021 (± 0.024)	0.031 (± 0.028)	0.019 (± 0.026)	0.028 (± 0.031)	<0.001
Bcell					
Mean (SD)	0.024 (± 0.021)	0.042 (± 0.030)	0.043 (± 0.033)	0.055 (± 0.028)	<0.001

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Mono					
Mean (SD)	0.013 (± 0.019)	0.045 (± 0.029)	0.027 (± 0.031)	0.074 (± 0.035)	<0.001
Gran					
Mean (SD)	0.80 (± 0.096)	0.60 (± 0.13)	0.72 (± 0.14)	0.49 (± 0.13)	<0.001
nRBC					
Mean (SD)	0.078 (± 0.084)	0.12 (± 0.11)	0.069 (± 0.11)	0.14 (± 0.11)	<0.001
MNLR					
Mean (SD)	6.6 (± 6.0)	3.3 (± 3.1)	4.0 (± 3.1)	1.9 (± 0.98)	<0.001
Missing	6 (1.7%)	0 (0%)	3 (0.6%)	0 (0%)	

* HM450 array

** EPIC customized array from GMEL

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 2. Meta-analysis results of maternal smoking and smoking exposure that were significant after false discovery rate correction in European cohorts.

	CHR	Probe_ID	Position	Estimate	Std..Error	Pvalue	HetPval	CHILD	FAMILY	UCSC Ref Gene	FDR	Reported Association EWAS catalog
Maternal **Smoking**												
	1	cg12876356	92481269	-1.11	0.22	7.33E-07	0.51	0.02	9.45E-06	***GFI1***	0.0019	MS; S; AC; BW
	1	cg09935388	92482032	-1.15	0.24	2.26E-06	0.52	0.02	2.71E-05	***GFI1***	0.0029	MS; GA; S; AC; BMI; BW
	1	cg14179389	92482405	-1.48	0.32	5.03E-06	0.73	0.01	1.12E-04	***GFI1***	0.0035	MS; S
	1	cg18146737	92481144	-0.92	0.20	5.58E-06	0.50	0.04	3.95E-05	***GFI1***	0.0035	MS; S; AC; BW
	1	cg09662411	92480576	-0.94	0.22	1.64E-05	0.29	0.10	3.85E-05	***GFI1***	0.0083	MS; S
	1	cg18316974	92481479	-0.74	0.18	3.58E-05	0.33	0.13	7.34E-05	***GFI1***	0.0152	MS; S; AC; BW
	17	cg01798813	2494783943	-0.83	0.21	1.09E-04	0.34	0.02	1.59E-03	*NA*	0.0395	A; GA; BMI
Smoking **Exposure**												
	1	cg09935388	92482032	-0.18	0.04	1.39E-05	0.23	0.15	2.45E-05	***GFI1***	0.04	MS; GA; S; AC MI; BW
	17	cg01798813	2494783943	-0.18	0.04	3.30E-05	0.13	0.00035	0.013	*NA*	0.04	A; GA; BMI

MS: maternal smoking

GA: gestational age

AC: alcohol consumption

BMI: body mass index

T2D: type 2 diabetes

A: age

BW: birth weight

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 3. Significant associations between maternal smoking methylation risk score and phenotypes in CHILD, FAMILY and START.

	Meta-analyzed Effect	Std. Error	Pvalue	HetPval	q-value	START HM450*	START GMEL**	CHILD HM450*	FAMILY GMEL**
Smoking Exposure (hr/week)	0.12	0.02	5.68E-07	0.03	3.69E-06	0.37	1.93E-01	6.91E-04	1.35E-05
Birth Length (cm)	-0.31	0.06	4.46E-07	0.31	3.69E-06	4.25E-03	1.83E-04	0.050	0.105
Birth Weight (kg)	-0.06	0.01	5.53E-07	0.80	3.69E-06	0.029	0.023	0.006	3.98E-03
1-year Length (cm)	-0.44	0.08	3.53E-08	0.45	9.18E-07	0.017	1.48E-04	0.021	7.76E-03
1-year Weight (kg)	-0.10	0.03	2.16E-03	0.92	9.38E-03	0.31	0.14	0.057	0.083
2-year Length (cm)	-0.48	0.11	2.02E-05	0.33	4.75E-05	0.010	0.00094	0.32	0.054
3-year Height (cm)	-0.27	0.11	0.014	0.92	0.047	0.43	0.32	0.20	0.074
3-year Skinfold	0.45	0.12	2.25E-04	0.57	9.74E-04	0.37	0.41	0.012	0.0057

* HM450 array, ** EPIC customized array from GMEL

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure Legends:

Figure 1. Manhattan plots of the meta-analyzed association between cord blood DNAm and maternal smoking in Europeans.

Manhattan plots summarized the meta-analyzed association *p*-values between cord blood DNA methylation levels and current maternal smoking (A) or smoking exposure (B) at a common set of 2,114 CpG sites. The red line denotes the smallest -log10(*p*-value) that is below the FDR correction threshold of 0.05. The red dots represent established associations with maternal smoking reported in Joubert and colleagues (18).

Figure 2. Relationships between maternal smoking MRS and maternal smoking history categories for each of the studies.

Maternal smoking methylation score (y-axis) was shown as a function of maternal smoking history (x-axis) in levels of severity for prenatal exposure for each study. Each severity level was compared to the never-smoking group and the corresponding two sample *t*-test *p*-value was reported. An omnibus test *p*-value to test whether a mean difference in methylation score was present among all smoking history categories.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Supporting Information captions

Supplementary Methods.

Suppl. Table 1. Quality controls for the inclusion/exclusion of samples and methylation probes.

Suppl. Table 2. Characteristics of the overall sample include 5176 mother–newborn pairs from the CHILD, FAMILY, and START cohorts.

Suppl. Table 3. A summary of available analyses and outcome variables in each cohort.

Suppl. Table 4. Characteristics of the epigenetic subsample from CHILD and FAMILY cohorts stratified by smoking status.

Suppl. Table 5. A summary of 11 CpGs available on the targeted array that contribute to the final methylation risk score.

Suppl. Table 6. A summary of 114 CpGs available on the HM450 array that contribute to the epigenome-wide methylation risk score.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Suppl. Table 7. Association between maternal smoking methylation risk score and phenotypes in CHILD, FAMILY and START.

Suppl. Table 8. Association between maternal smoking methylation risk score and phenotypes in CHILD and FAMILY.

Suppl. Table 9. A summary of statistical tests to evaluate the difference in distribution of methylation score and relevant CpGs between South Asian and European cohorts.

Supplementary Figure 1. Manhattan plots of the meta-analyzed association between cord blood DNA methylation and ever maternal smoking in Europeans.

The meta-analyzed association *p*-values for ever maternal smoking and methylation levels at 2,114 CpG sites were summarized in the Manhattan plot. Ever maternal smoking was defined to compare those who were currently smoking or quitted before or during this pregnancy vs. those never smoked. The red line denotes the smallest -log10(*p*-value) that is below the FDR correction threshold of 0.05. The red dots represent established associations with maternal smoking reported in Joubert and colleagues (18).

Supplementary Figure 2. Manhattan plots of the meta-analyzed association between cord blood DNA methylation and smoking exposure in the combined European and South Asian cohorts.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The meta-analyzed association *p*-values for smoking exposure and methylation levels at 2,114 CpG sites were summarized in the Manhattan plot. The red dots represent established associations with maternal smoking reported in Joubert and colleagues (18).

Supplementary Figure 3. Manhattan plots of the Epigenome-wide associations between cord blood DNAm and maternal smoking in CHILD.

Manhattan plots summarized the association *p*-values between cord blood DNA methylation levels and current maternal smoking (A) or ever maternal smoking (B) at 200,050 CpG sites. The red line denotes the smallest -log10(*p*-value) that is below the FDR correction threshold of 0.05. The red dots represent established associations with maternal smoking reported in Joubert and colleagues (18).

Supplementary Figure 4. Manhattan plots of the Epigenome-wide associations between cord blood DNAm and smoking exposure in CHILD and START.

Manhattan plots summarized the association *p*-values between cord blood DNA methylation levels and smoking exposure in CHILD (A) or START (B) at 200,050 and 218,982 CpG sites, respectively. The red line denotes the smallest -log10(*p*-value) that is below the FDR correction threshold of 0.05. The red dots represent established associations with maternal smoking reported in Joubert and colleagues (18).

Supplementary Figure 5. A comparison of results for maternal smoking MRS constructed using the targeted array and an epigenome-wide arrays in CHILD.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Maternal smoking methylation score (y-axis) was shown as a function of maternal smoking history (x-axis) in levels of severity ([0] = never smoked; [1] = quit before this pregnancy; [2] = quit during this pregnancy; [3] = currently smoking) for prenatal exposure for each study. Each severity level was compared to the never smoking group and the corresponding two sample *t*-test *p*-value was reported. An omnibus test *p*-value to test whether a mean difference in methylation score was present among all smoking history categories.

Supplementary Figure 6. Boxplots of methylation score and contributing CpGs stratified by study.

The boxplots captured the standardized maternal smoking methylation score and CpGs values (y-axis) stratified by study. The top panels included CpGs that were positively correlated with maternal smoking, or more methylated when exposed to prenatal smoking, while the bottom panels included CpGs and the score that were negatively associated with maternal smoking.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Acknowledgements

We express our sincere gratitude to all the participating families and the START, FAMILY, and CHILD study teams, including interviewers, nurses, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, and receptionists.

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. Zulyniak.

This study was funded by the Canadian Institutes of Health Research Metabolomics Team Grant: MWG-146332. Dr. Anand is supported by a Tier 1 Canada Research Chair in Ethnicity and CVD

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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.

Data availability statement

The summary statistics used to construct methylation risk scores are available from EWAS catalog at http://www.ewascatalog.org/?trait=maternal%20smoking%20in%20pregnancy with additional filters of PubMID 27040690 and analysis on "Sustained maternal smoking in pregnancy effect on newborns adjusted for cell composition".

Summary statistics generated in the current study, including a total of 8 primary association studies (three smoking phenotypes in three cohorts) and 3 sets of meta-analyzed results in Europeans are available upon request. All scripts to reproduce and validate the predictive model can be found at https://github.com/WeiAkaneDeng/EpigeneticResearch/tree/main/MaternalSmoking.

Conflicts of interest

No conflict of interest.

Ethics Statement

Ethical approval was obtained independently from the Hamilton Integrated Research Ethics Board: CHILD (REB 07–2929), FAMILY (REB 02–060), and START (REB 10–640). CHILD was additionally approved by the respective Human Research Ethics Boards at McMaster

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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 three cohorts together without additional consent from the participants.

Reference

1. Stock SJ, Bauld L. Maternal smoking and preterm birth: An unresolved health challenge. Vol. 17, PLoS Medicine. 2020.
2. Liu B, Xu G, Sun Y, Qiu X, Ryckman KK, Yu Y, et al. Maternal cigarette smoking before and during pregnancy and the risk of preterm birth: A dose–response analysis of 25 million mother–infant pairs. PLoS Med. 2020;17(8).
3. Marufu TC, Ahankari A, Coleman T, Lewis S. Maternal smoking and the risk of still birth: systematic review and meta-analysis. BMC Public Health. 2015;15(1).
4. Ventura SJ, Hamilton BE, Mathews TJ, Chandra A. Trends and variations in smoking during pregnancy and low birth weight: Evidence from the birth certificate, 1990-2000. Pediatrics. 2003;111(5 II).
5. U.S. Department of Health and Human Services. The Health Consequences of Smoking—50 Years of Progress A Report of the Surgeon General. A Report of the Surgeon General. 2014.
6. Montgomery SM, Ekbom A. Smoking during pregnancy and diabetes mellitus in a British longitudinal birth cohort. Br Med J. 2002;324(7328).
7. Toschke A, Koletzko B, Slikker W, Hermann M, von Kries R. Childhood obesity is associated with maternal smoking in pregnancy. Eur J Pediatr. 2002;161(8).
8. Oken E, Levitan EB, Gillman MW. Maternal smoking during pregnancy and child overweight: Systematic review and meta-analysis. Vol. 32, International Journal of Obesity. 2008.
9. Philips EM, Santos S, Trasande L, Aurrekoetxea JJ, Barros H, von Berg A, et al. Changes in parental smoking during pregnancy and risks of adverse birth outcomes and childhood overweight in Europe and North America: An individual participant data meta-analysis of 229,000 singleton births. PLoS Med. 2020;17(8).
10. England LJ, Grauman A, Qian C, Wilkins DG, Schisterman EF, Yu KF, et al. Misclassification of maternal smoking status and its effects on an epidemiologic study of pregnancy outcomes. Nicotine and Tobacco Research. 2007;9(10).
11. Shipton D, Tappin DM, Vadiveloo T, Crossley JA, Aitken DA, Chalmers J. Reliability of self reported smoking status by pregnant women for estimating smoking prevalence: A retrospective, cross sectional study. BMJ (Online). 2009;339(7732).

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

12. Salmasi G, Grady R, Jones J, McDonald SD. Environmental tobacco smoke exposure and perinatal outcomes: a systematic review and meta-analyses. Vol. 89, Acta obstetricia et gynecologica Scandinavica. 2010.
13. Shenker NS, Ueland PM, Polidoro S, Van Veldhoven K, Ricceri F, Brown R, et al. DNA methylation as a long-term biomarker of exposure to tobacco smoke. Epidemiology. 2013;24(5).
14. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic Signatures of Cigarette Smoking. Circ Cardiovasc Genet. 2016;9(5).
15. Guida F, Sandanger TM, Castagné R, Campanella G, Polidoro S, Palli D, et al. Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. Hum Mol Genet. 2015;24(8).
16. Sikdar S, Joehanes R, Joubert BR, Xu CJ, Vives-Usano M, Rezwan FI, et al. Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking. Epigenomics. 2019;11(13).
17. Zeilinger S, Kühnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation. PLoS One. 2013 May 17;8(5).
18. 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).
19. 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. Philosophical Transactions of the Royal Society B: Biological Sciences. 2019;374(1770).
20. Akhabir L, Stringer R, Desai D, Mandhane PJ, Azad MB, Moraes TJ, et al. DNA methylation changes in cord blood and the developmental origins of health and disease – a systematic review and replication study. BMC Genomics. 2022;23(1).
21. 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;19(1).
22. Küpers LK, Xu X, Jankipersadsing SA, Vaez A, La Bastide-van Gemert S, 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).
23. 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;13(1).
24. 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).
25. 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).
26. 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;8.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

27. 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 [Internet]. 2022 Dec 1 [cited 2023 Mar 9];14(1). Available from: https://pubmed.ncbi.nlm.nih.gov/36243740/
28. Choquet H, Yin J, Jorgenson E. Cigarette smoking behaviors and the importance of ethnicity and genetic ancestry. Transl Psychiatry. 2021;11(1).
29. Elliott HR, Tillin T, McArdle WL, Ho K, Duggirala A, Frayling TM, et al. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. Clin Epigenetics. 2014;6(1).
30. Bollepalli S, Korhonen T, Kaprio J, Anders S, Ollikainen M. EpiSmokEr: A robust classifier to determine smoking status from DNA methylation data. Epigenomics. 2019;11(13).
31. Yousefi PD, Suderman M, Langdon R, Whitehurst O, Davey Smith G, Relton CL. DNA methylation-based predictors of health: applications and statistical considerations. Vol. 23, Nature Reviews Genetics. 2022.
32. 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).
33. 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).
34. Morrison KM, Atkinson SA, Yusuf S, Bourgeois J, McDonald S, McQueen MJ, et al. The Family Atherosclerosis Monitoring In earLY life (FAMILY) study. Rationale, design, and baseline data of a study examining the early determinants of atherosclerosis. Am Heart J. 2009;158(4).
35. 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;13(1).
36. Brydon P, Smith T, Proffitt M, Gee H, Holder R, Dunne F. Pregnancy outcome in women with type 2 diabetes mellitus needs to be addressed. Int J Clin Pract. 2000;54(7).
37. 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 [Internet]. 2015 Oct 1 [cited 2023 Jun 8];3(10):795. Available from: /pmc/articles/PMC4673084/
38. Lange S, Probst C, Rehm J, Popova S. National, regional, and global prevalence of smoking during pregnancy in the general population: a systematic review and meta-analysis. Lancet Glob Health. 2018;6(7).
39. 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).
40. 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).
41. Rider CF, Carlsten C. Air pollution and DNA methylation: Effects of exposure in humans. Vol. 11, Clinical Epigenetics. 2019.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

42. Vogel CFA, Van Winkle LS, Esser C, Haarmann-Stemmann T. The aryl hydrocarbon receptor as a target of environmental stressors – Implications for pollution mediated stress and inflammatory responses. Vol. 34, Redox Biology. 2020.
43. Reynolds LM, Wan M, Ding J, Taylor JR, Lohman K, Su D, et al. DNA Methylation of the Aryl Hydrocarbon Receptor Repressor Associations with Cigarette Smoking and Subclinical Atherosclerosis. Circ Cardiovasc Genet. 2015;8(5).
44. Roberts CK, Sindhu KK. Oxidative stress and metabolic syndrome. Vol. 84, Life Sciences. 2009.
45. 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;12(1).
46. Moshammer H, Hutter HP. Breast-feeding protects children from adverse effects of environmental Tobacco smoke. Int J Environ Res Public Health. 2019;16(3).
47. 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. Journal of Nutrition. 2016;146(11).
48. 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).
49. 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;46(20).
50. 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).
51. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. Nucleic Acids Res. 2017;45(4).
52. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol. 2016;17(1).
53. Hillary RF, McCartney DL, McRae AF, Campbell A, Walker RM, Hayward C, et al. Identification of influential probe types in epigenetic predictions of human traits: implications for microarray design. Clin Epigenetics. 2022;14(1).
54. 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;11(1).
55. 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.
56. Metzger BE. International Association of Diabetes and Pregnancy Study Groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Vol. 33, Diabetes Care. 2010.
57. 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).
58. Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized regression on summary statistics. Genet Epidemiol. 2017;(February):469–80.

medRxiv preprint doi: https://doi.org/10.1101/2023.09.24.23295907; this version posted October 12, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .