SUPPLEMENTARY DATA

Cohort-Specific Information

ALSPAC

Design and Study Population

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a prospective pregnancy cohort study, which enrolled 14,541 pregnant women residing in Avon, United Kingdom who had expected delivery dates between April 1st, 1991 and December 31st, 1992. As described previously (1, 2), detailed information has been collected on these participants and their offspring at regular intervals. Details of all data collected for ALSPAC are available through a fully searchable data dictionary, which is publicly available: http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/.

As part of the Accessible Resources for Integrated Epigenomic Studies (ARIES, http://www.ariesepigenomics.org.uk/) project, DNA methylation was generated for 1,018 mother-offspring pairs from the ALSPAC cohort, using the Infinium HumanMethylation450 BeadChip array (Illumina Inc., San Diego, United States). ARIES participants were selected based on availability of DNA samples at two time points for the mother (antenatal and at follow-up when the offspring were adolescents) and at three time points for the offspring (neonatal, childhood (age 7), and adolescence (age 17)). Cord blood DNA methylation data were included in the current meta-analyses.

Written informed consent was obtained for all ALSPAC participants, and ethical approval was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees.

Gestational Diabetes Classification

GDM classification for ALSPAC has been described previously (3). Briefly, at the time of recruitment, women were provided with a questionnaire to obtain information on existing maternal diabetes and past history of GDM. A standard protocol was used by research midwives to obtain information on GDM and glycosuria (recorded as none, trace, +, ++, +++, or more) for the index pregnancy from the mother's antenatal and postnatal medical records. At the time that ALSPAC participants were enrolled (1991-1992), the practice in the United Kingdom was for all women to be offered urine tests for glycosuria at each of their antenatal clinic visits. Universal screening of women with a random or fasting blood glucose level or with an oral glucose tolerance test (OGTT) was not performed, and diagnostic tests for GDM were only performed in women with established risk factors (e.g., family history of GDM, previous history of GDM or macrosomic birth, and South Asian ethnicity) or persistent glycosuria. Glycosuria was defined as a value of at least 13.9 mmoL/L or 250 mg/100 mL (according to the manufacturer (Bayer)) on a minimum of two occasions at any time during the pregnancy (4). Women were classified into one of four mutually exclusive categories: no evidence of glycosuria or diabetes, existing diabetes before the pregnancy, GDM, and glycosuria. Women classified as having GDM were included as GDM cases in the current meta-analysis, and women without evidence of glycosuria or diabetes were included as controls. For the current meta-analyses, a total of 889 mother-newborn pairs (22 with GDM) who had GDM information, cord blood DNA methylation, and complete covariate information were included.

SUPPLEMENTARY DATA

Methylation Measurements

Methods for methylation measurements in ALSPAC have been described previously (5). Briefly, cord blood was collected according to standard procedures. DNA methylation assays and data pre-processing was performed at the University of Bristol as part of the ARIES project. DNA was extracted using a standard protocol and was bisulfite-converted using the Zymo EZ DNA MethylationTM kit (Zymo, Irvine, CA). DNA methylation was then measured using the Infinium HM450 BeadChip assay (Illumina Inc, San Diego, CA), according to the standard protocol. Arrays were scanned using an Illumina iScan. An initial review of data quality was assessed using GenomeStudio (version 2011.1). A semi-random approach (sampling criteria were in place to ensure that all time points were represented on each array) was used to distribute ARIES samples across slides to minimize the possibility of potential confounding by batch. A wide range of batch variables were recorded in a purpose-built laboratory information management system (LIMS). The main batch variable was found to be the bisulfite conversion plate number. The LIMS also reported quality control metrics from the standard control probes on the 450k BeadChip for each sample. Samples with average probe p-values ≥ 0.1 were re-run. If a re-run was unsuccessful, the sample was excluded from further analyses. Additionally, genotype probes were compared with SNP-chip data from the same individual to identify and remove any sample mismatches. If an individual did not have genome-wide SNP data, the sample was flagged if there was a sex-mismatch based on X-chromosome methylation. Methylation data were pre-processed using R (version 3.0.1) with background correction. Subset quantile normalization was performed using a pipeline described by Touleimat and Tost (6). Probes with values > 3*interquartile range were excluded.

Covariates

Maternal age at delivery was derived from the mother's date of birth. Maternal body mass index (BMI) was calculated from self-reported height and pre-pregnancy weight, which were collected by questionnaire during the first trimester of pregnancy. Maternal education was determined by a questionnaire, completed during pregnancy, and was collapsed into the following two categories: less than A level or A level and above. Maternal smoking during pregnancy was defined as smoking at least one cigarette per day at any time during pregnancy, and was determined by questionnaire at the time of recruitment (median age of children: 2 months). Newborn sex was obtained from obstetric records. Participants with non-white European ancestry were excluded from all analyses. Proportions of seven cord blood cell types were estimated with the estimateCellCounts function in minfi (7), using the cord blood dataset (8). Models were adjusted for potential batch effects by including 10 surrogate variables, generated using the R package SVA (9), in regression models.

GOYA

Design and Study Population

The Genetics of Overweight Young Adults (GOYA) study has been described previously (10). Briefly, it includes a case-cohort sampled subset of 91,387 pregnant women recruited into the Danish National Birth Cohort between 1996 and 2002, which was approved by all the regional scientific ethics committees in Denmark, by the central scientific ethics committee for the whole of Denmark, and by the Danish Data Protection Board. Of 67,853 women who had given birth to a live infant, provided a blood sample during pregnancy, and had BMI information available, 3.6% with the largest residuals from the regression of BMI on age and parity (all entered as continuous variables) were selected as GOYA cases. The BMI for these 2,451 women ranged from 32.6 to 64.4. A similar number of women (n=2,450) was randomly sampled from the remaining cohort as controls. DNA methylation data were generated for the

offspring of 1,000 case and control mothers. For the current meta-analyses, we restricted to 432 GOYA mother-child pairs from the case (i.e., obese) group (N=28 with GDM), who had DNA methylation measures and all relevant covariates available.

Gestational Diabetes Classification

GDM was identified by the presence of ICD-10 code O24 during pregnancy in the National Patient Register, as described previously (11) Because some under-reporting of GDM was expected in the register, we also relied on self-reported information from telephone interviews (11). According to Danish guidelines during the study period, GDM was diagnosed by the use of a 1-step 75 g OGTT. Thresholds for GDM were based on international recommendations during this time period (12).

Methylation Measurements

Cord blood was collected according to standard procedures, spun and frozen at -80˚C. DNA methylation analysis and data pre-processing were performed at the University of Bristol. Following extraction, DNA was bisulfite converted using the Zymo EZ DNA MethylationTM kit (Zymo, Irvine, CA). Following conversion, methylation status was measured using the Illumina Infinium® HumanMethylation450k BeadChip assay according to standard protocol. The arrays were scanned using an Illumina iScan and initial quality review was assessed using GenomeStudio (version 2011.1). The level of methylation is expressed as a "Beta" value (β-value), ranging from 0 (no cytosine methylation) to 1 (complete cytosine methylation).

All DNA methylation results were cleaned and normalized in R (version 3.3.0) using the *meffil* package (13). Samples were excluded from downstream analysis if > 10% of probes had a p-value ≥ 0.01, and CpGs were excluded if their probe detection p-values were ≥ 0.01 across > 10% of samples. Sample sex was assessed by comparing HumanMethylation450k genotype probes to previous SNP-chip data, and comparing median chromosome Y and chromosome X probe intensities. Samples were excluded if they failed these checks. Data were normalized using the functional normalization approach in the *minfi* R package (7). Probes with values > 3*interquartile range were also excluded from analyses.

Covariates

Data on covariates were collected via a telephone interview at ~16 weeks gestation. Maternal age was derived from the mother's self-reported date of birth. Maternal pre-pregnancy BMI was calculated from self-reported height and weight. Socioeconomic status was defined using maternal education or occupation as follows: 1) manager or medium to long education, 2) work requiring a short training period or skilled manual labor, 3) unskilled or public service. Maternal smoking in pregnancy was defined as any/no reported smoking in pregnancy. Ten surrogate variables were generated and included in models to adjust for technical batch. Estimation of seven different cord white blood cell types (CD8+ T and CD4+ T lymphocytes, CD56+ natural killer cells, CD19+ B cells, CD14+ monocytes, granulocytes, nRBCs) were computed by the Houseman method (14) using the cord blood reference dataset and the default implementation of the estimateCellCounts function in the minfi package.

SUPPLEMENTARY DATA

Healthy Start

Design and Study Population

The Healthy Start study is an ongoing, prospective cohort study of mothers and children in Colorado, United States. Pregnant women were recruited from outpatient obstetrics clinics at the University of Colorado in 2009-2014. Eligible participants were aged 16 years or older, were pregnant with a single infant at fewer than 24 weeks gestation, with no history of diabetes, cancer, major psychiatric illness or asthma treated with steroids, and no prior stillbirths or extremely preterm births. A total of 1,410 pregnant women were recruited, representing approximately 50% of those eligible. Study procedures were approved by the Colorado Multiple Institutional Review Board and written informed consent was obtained from all participating mothers.

Gestational Diabetes Classification

Physician-diagnosed maternal GDM status was ascertained by medical record review. The standard clinical care is to offer all pregnant women screening for GDM with a 1-h, 50-g oral glucose challenge test. Women with a value ≥ 7.7 mmol/l during this screening are then asked to undergo a 3-h, 100-g diagnostic OGTT. GDM is diagnosed when two or more glucose values during the diagnostic OGTT meet or exceed the Carpenter-Coustan criteria for a positive test (15).

Methylation Measurements

Umbilical cord blood was collected at delivery, and DNA was extracted from stored buffy coats using the QIAamp kit (Qiagen). Bisulfite conversion, labeling, and hybridization were performed in the University of Colorado Denver Genomics Core. Methylation analysis of cord blood samples (N=600) was conducted using the Infinium HumanMethylation450 BeadChip assay (Illumina Inc, San Diego, CA), and processed in the University of Colorado Genomics Core lab. Samples with predicted child sex inconsistent with reported sex were excluded (N=4). Probes with high detection p-value (> 0.05) or low beadcount (< 3 in at least 5% of samples) were excluded. The total number of CpGs analyzed was 484,261. The preprocessQuantile function in the R package Minfi (7) was used for normalization, and the ComBat function in the sva R package was used to adjust for potential batch effects (9). Probes with values > 3*interquartile range were excluded.

Covariates

Models were adjusted for the following covariates: infant sex (obtained from the medical record at delivery); maternal age; pre-pregnancy BMI (calculated from pre-pregnancy weight obtained from the medical record or, if unavailable, via self-report at enrollment, and height measured at the first study visit); self-reported race/ethnicity (non-Hispanic white, Hispanic, non-Hispanic African American, all others); education completed (categorized as less than high school, high school/GED, some college, completed college, and graduate degree), smoking during pregnancy (obtained from the mother via questionnaire and classified as any versus none); and estimated proportions of seven cell types, which were determined with the estimateCellCounts function in minfi (7), using the cord blood dataset (8).

SUPPLEMENTARY DATA

INMA

Design and Study Population

The INfancia y Medio Ambiente (INMA) – (Environment and Childhood) Project consists of a network of birth cohorts in Spain, which aim to study the role of environmental pollutants during pregnancy and early childhood in child growth and development (16). As described previously (16), women were included in the study if they were a resident in one of the study areas (Ribera d'Ebre, Menorca, Granada, Valencia, Sabadell, Asturias, or Gipuzkoa), were at least 16 years old, had a singleton pregnancy, had not used assisted reproduction, planned to deliver in the reference hospital, and had no communication problems. This study was approved by the Ethical Committee of the Municipal Institute of Medical Investigation and by the Ethical Committees of each participating center. Pregnant women received information about the study, both written and orally, and written informed consent was obtained from all participating parents. Data for the current study came from the INMA Sabadell cohort (children born between 2004 and 2007) (16). A total of 156 mother-newborn pairs (12 with GDM) with information on maternal GDM and offspring cord blood DNA methylation data and complete covariate information were included in the current analysis.

Gestational Diabetes Classification

Women who met the criteria for being low risk for GDM with no indications, defined as < 24 years old, of normal weight, with no family or personal history of diabetes, and with no prior pregnancy complications, were not tested for GDM. Women who were not considered at low risk underwent a glucose challenge test, where plasma glucose levels were measured 1 h after the woman was administered 50 g of glucose. If this value exceeded 130 mg/dl, women were administered a subsequent 3 h OGTT in a fasting state, where 100 g of glucose was provided, and plasma glucose measures were obtained at 0, 60, 120, and 180 minutes post-glucose administration. Women were classified as having impaired glucose tolerance if they had a plasma glucose concentration ≥ 130 mg/dl after the glucose challenge test and if one value from the OGTT exceeded one of the following Carpenter-Coustan cutoffs (17): 95 mg/dl at 0 minutes, 180 mg/dl at 60 minutes, 155 mg/dl at 120 minutes, or 140 mg/dl at 180. Women identified as having impaired glucose tolerance were excluded from the current analysis. Women were classified as having GDM if ≥ 2 OGTT values exceeded the same cutoffs listed above.

Methylation Measurements

Cord blood was collected in EDTA tubes, and DNA was extracted using the Chemagic DNA Blood Kits (Perkin Elmer) in a Chemagen Magnetic Separation Module 1 station at the Spanish National Genotyping Center (CEGEN, http://www.usc.es/cegen/). DNA concentration was determined by Nanodrop and picogreen. 500 ng of DNA was bisulfite-converted using the EZ 96-DNA methylation kit according to the manufacturer's standard protocol. Bisulfite conversion was verified using Sanger Sequencing. DNA methylation was measured using the Infinium HumanMethylation450 BeadChip array (Illumina Inc., San Diego, United States), with samples randomized to prevent batch effects. Standard male and female DNA samples were included in this step as controls. The resulting DNA methylation data was preprocessed using the minfi package in R (7), and the data was normalized using functional normalization and background correction (18). Probes with a detection p-value < 10×10^{-16} were excluded from the analyses (19). Probes with values > 3*interquartile range were also excluded.

SUPPLEMENTARY DATA

Covariates

Maternal education was determined by a questionnaire administered at enrollment (week 12 of pregnancy) and was categorized into three categories: less than high school, high school, and college or higher. Maternal smoking status (ever versus never) was determined by a questionnaire administered at 32 weeks of pregnancy.

For the current analysis, the age reported by the mother at the time of enrollment (week 12 of pregnancy) was used. Infant's sex was abstracted from the medical records. Maternal pre-pregnancy BMI was calculated from measured height and self-reported weight, collected by a questionnaire at enrollment (20). In INMA, self-reported pre-pregnancy weight is highly correlated with measured weight at 12 weeks of pregnancy ($r=0.96$; $P < 0.0001$). Ethnic ancestry of participants was determined using a questionnaire where the mothers were asked about their race and ethnic identity and the race and ethnic identity of the child's father. All participants included in the current analysis on GDM and newborn DNA methylation self-identified as "White" and were European. Ethnic subcategories included "both parents white and Spanish", "both parents white, but at least one not European", "both parents white, not Spanish, but European". All participants in these analyses were considered European. Estimation of seven different blood cell types (CD8+ T and CD4+ T lymphocytes, CD56+ natural killer cells, CD19+ B cells, CD14+ monocytes, granulocytes and nucleated red blood cells) was performed using the constrained projection/quadratic programming of Houseman (14) included in the estimateCellCounts function in the minfi package (7), using the cord blood reference (8) for projection.

PREDO

Design and Study Population

The Prediction and Prevention of Preeclampsia and Intrauterine Growth Restriction (PREDO) Study is a longitudinal multicenter pregnancy cohort study of Finnish women and their singleton children born alive between 2006-2010 (21). A total of 1,079 pregnant women were recruited, of whom 969 had one or more and 110 had none of the known risk factors for preeclampsia and intrauterine growth restriction. The recruitment took place in arrival order when these women attended the first ultrasound screening at 12+0 to 13+6 weeks+days of gestation in one of the ten hospital maternity clinics participating in the study. The cohort profile (21) contains details of the study design and inclusion criteria. The study protocol was approved by the Ethical Committees of the Helsinki and Uusimaa Hospital District and by the participating hospitals. Written informed consent was obtained from all women. The study is registered at ClinicalTrials.gov (identifier ISRCTN14030412).

Gestational Diabetes Classification

GDM status was determined from OGTT results, abstracted from the medical records. GDM, defined as diabetes which emerged or was first identified during pregnancy, was classified based on results from a one-step 75-g oral OGTT, using the International Association of the Diabetes and Pregnancy Study Groups (IADPSG) criteria (22): fasting, 1-h, or 2-h plasma glucose concentrations ≥ 5.1, 10.0 or 8.5 mmol/l.

SUPPLEMENTARY DATA

Methylation Measurements

DNA was isolated from cord blood and bisulfite converted using the EZ-96 DNA methylation kit (Zymo research Corporation, Irvine, USA). The Infinium HumanMethylation450 BeadChip array (Illumina Inc., San Diego, USA) was used to measure DNA methylation levels, reported as beta values ranging from 0 (no methylation) to 1 (complete methylation). To limit potential batch effects, samples were randomized across 96-well plates, based on gender and maternal risk factors for pre-eclampsia. A quality control pipeline was set up using the R package minfi (7). Three samples were excluded from analyses, because they were outliers based on median intensities. Furthermore, 20 samples showed discordance between phenotypic sex and estimated sex and were therefore excluded. Nine samples that were found to be contaminated by maternal DNA were also removed (23).

Methylation beta values were normalized using the funnorm function in the minfi package (7). We excluded any probes on the X or Y chromosome and also probes containing SNPs and cross-hybridizing probes according to Chen et al. (24) and Price et al (25). Furthermore, any CpGs with a detection p-value > 0.01 in at least 25% of the samples were excluded. Additionally, probes with values > 3*interquartile range were excluded. The final dataset contained 428,619 CpGs and 780 participants (N=180 with GDM). After normalization, two batches (slide and well) were significantly associated with DNA methylation levels and were therefore removed iteratively using the Combat method (26).

Covariates

Newborn's sex; maternal early pregnancy BMI, calculated from maternal weight and height measures verified by a nurse at the first visit to the antenatal clinic (M=8+4 weeks+days, SD=1+3 weeks+days of gestation); maternal smoking status; and maternal age were derived from medical records. Maternal smoking status in PREDO was recorded as never smoked during the pregnancy, quit smoking during the first trimester, or smoked throughout pregnancy. For the current analysis, these three smoking categories were recoded into two categories: never versus ever smoked during pregnancy. Maternal education was determined by a questionnaire and consisted of the following categories: lower secondary education or less, upper secondary education, lower tertiary education, or upper tertiary education.

Ancestry was assessed by including multi-dimensional scaling components from GWAS. Cord blood cell fractions were derived using the estimateCellCounts function in the R-package minfi (7), using the cord blood reference dataset (8).

RHEA/ENVIR*ON*AGE/Piccolipiù

Within the EXPOsOMICS collaborative European project, three population-based birth cohorts, ENVIRonmental influence ON AGEing in early life (ENVIR*ON*AGE), Rhea, and Piccolipiù, were combined to conduct DNA methylation analyses (27). Phenotypic variables were harmonized across the three cohorts, and their biospecimen were semi-randomized on the Infinium HumanMethylation450 BeadChip array (Illumina Inc, San Diego, CA), such that the latter would incorporate proportional representations of the three cohorts and ensure that batch effects do not completely confound with biological covariates of interest.

SUPPLEMENTARY DATA

Methylation Measurements for the Three Cohorts

Methods for DNA methylation measurements were the same for each cohort. Aliquots of cord blood samples (collected and frozen at birth at -80°C) were shipped on dry ice to the Epigenetics Group at the International Agency for Research on Cancer (IARC), Lyon, France, where DNA was extracted (QIAamp 96 DNA Blood Kit, Qiagen 51161), quantified (Quant-iT PicoGreen dsDNA Assay Kit, Molecular Probes P7589), and bisulfite converted (600 ng of DNA using EZ-96 DNA Methylation kit, Zymo Research D5004). DNA methylation was measured at 485,577 CpGs using the Infinium HumanMethylation450 BeadChip (Illumina Inc., San Diego, USA). The arrays were designed such that batch effects (e.g. sample position and intra- and inter-variability in arrays and chips) would not completely confound with biological covariates of interest. This design allows the retention of biological variation even after correction for technical variation. Raw intensity (.idat) files were handled in R using the minfi package (7) to calculate the methylation level at each CpG as the beta-value (β=intensity of the methylated allele (M)/(intensity of the unmethylated allele (U) + intensity of the methylated allele (M) + 100)), and the data were exported for quality control and processing. Cross-reactive probes (24) and low-quality probes (probes having bead counts < 3 in at least 5% of samples) were removed. Data quality was further assessed using box plots for the distribution of methylated and unmethylated signals, and multidimensional scaling plots and unsupervised clustering were used to check for sample outliers and potential gender mismatches, which were removed. Additionally, samples having > 1% of CpG sites with a detection p-value > 0.05 were removed. The remaining dataset was normalized using the funnorm normalization of the minfi package (7). Cohort and batch (sample plate, sentrix position) effects were then corrected for using surrogate variable analysis (9). Probes with values > 3*interquartile range were excluded from analyses.

RHEA

Design and Study Population

The mother-child Rhea study in Crete is a prospective cohort examining a population sample of pregnant women and their children, at the prefecture of Heraklion (N=1,500), who were recruited at the first comprehensive ultrasound exam (median 12 weeks gestation) between February 2007 and February 2008 (28). Pregnant women were included in Rhea if they lived in the study area, were ≥ 16 years of age, and had no communication handicap. Ethical approval was obtained by the Ethics Committee of the University Hospital at Heraklion and informed consent was given by all participants. The main study aims of the Rhea study are: 1. To characterize nutritional, environmental and psychosocial determinants of children's growth and development; 2. To focus on four primary outcome areas of research: (i) offspring growth and obesity, (ii) neuropsychological and behavioural development, (iii) allergies and asthma in childhood, and (iv) genotoxicity; 3. To evaluate mother's health during and after pregnancy; 4. To evaluate the interaction between environmental stressors and genetic variants in children's growth and health. A set of 100 newborns from the Rhea cohort is included in the EXPOsOMICS children studies, and cord blood DNA methylation data is available for these participants. A subsample of 88 mother-newborn pairs, eight with GDM, with complete covariate information for the investigation of GDM and newborn DNA methylation was included in the current meta-analyses.

SUPPLEMENTARY DATA

Gestational Diabetes Classification

GDM was classified in Rhea based on OGTT results, abstracted from medical records, as described previously (29, 30). Women were screened for GDM at 24–28 weeks gestation and were classified as having GDM at the index pregnancy if two or more of the four plasma glucose values obtained during the 100-g, 3-h OGTT were abnormal according to the criteria proposed by Carpenter and Coustan (15): fasting of 95 mg/dl or greater, 1 h of 180 mg/dl or greater, 2 h of 155 mg/dl or greater, and 3 h of 140 mg/dl or greater.

Covariates

Information on maternal age, pre-pregnancy BMI, smoking, education, and Greek origin was obtained from questionnaires administered at recruitment, which occurred between week 12 and 14 of pregnancy (at the time of the first routinely scheduled major ultrasound test). Maternal smoking status was categorized as smoking during pregnancy yes or no. Maternal education was defined as primary school, secondary school, and university degree or higher. Maternal race was defined as maternal Greek origin or not. As this cohort is quite homogeneous with respect to ethnicity, the ancestry of all newborns was set to European for the current study. Information on gender was collected by the midwives during the birth admission. Estimated proportions of seven cord blood cell types were determined with the estimateCellCounts function in minfi (7), using the cord blood dataset (8).

ENVIR*ON*AGE

Design and Study Population

The ongoing Belgian birth cohort ENVIR*ON*AGE is a longitudinal study, starting with recruitment at birth and follow-up at 4-6 years and in preadolescence (31). The birth cohort study was designed to understand the determinants of molecular ageing in early life and its role in the developmental origins of health and disease. Recruitment of ENVIR*ON*AGE mother-newborn pairs began in February 2010 and is ongoing. Women were enrolled when they arrived for delivery at the East-Limburg Hospital in Genk, Belgium. Mothers without a planned caesarean section who are able to fill out a questionnaire in Dutch are eligible for participation. Procedures are approved by the Ethical Committee of Hasselt University and the East-Limburg Hospital, and recruitment is carried out according to the Helsinki declaration. A set of 200 newborns from the ENVIR*ON*AGE cohort is included in the EXPOsOMICS children studies, for which data on cord blood DNA methylation is available. A subsample of 187 mother-newborn pairs, four with GDM, with complete covariate information was included in the current meta-analyses.

Gestational Diabetes Classification

GDM was classified in ENVIR*ON*AGE based on OGTT results, abstracted from medical records. All pregnant women were screened for GDM at 24-28 weeks gestation using a random 50 g 1-h glucose load test, followed by a diagnostic fasting 100 g 3-h OGTT if their screening test was positive. Women were classified as having GDM if two or more of the four plasma glucose values obtained during the OGTT were abnormal according to the criteria proposed by Carpenter and Coustan (15): fasting plasma glucose of 95 mg/dl or greater, 1 h plasma glucose of 180 mg/dl or greater, 2 h plasma glucose of 155 mg/dl or greater, and 3 h plasma glucose of 140 mg/dl or greater.

SUPPLEMENTARY DATA

Covariates

The variables newborn's sex and maternal age were obtained from medical records. Maternal BMI was calculated based on maternal weight and height values, measured at the first antenatal visit (week 7-9 of gestation); participants were asked to wear no shoes and to wear light clothing while these measures were obtained. Information on maternal smoking, maternal education, and newborn's ethnicity was obtained from questionnaires, which were administered post-delivery. Maternal smoking status was defined as smoking cigarettes during pregnancy yes or no. Newborns were classified as European when at least two grandparents were European or as non-European when at least three grandparents were of non-European origin. Maternal education was classified as primary school, secondary school, and university degree or higher. Estimated proportions of seven cord blood cell types were determined with the estimateCellCounts function in minfi (7), using the cord blood reference dataset (8).

Piccolipiù

Design and Study Population

Piccolipiù is a multicentric Italian birth cohort that recruited 3,338 newborns and their mothers in five centers: Turin, Trieste, Florence, Viareggio, and Rome between 2011 and 2015. Details about the study protocol have been published elsewhere (32). Briefly, singleton pregnant women were eligible to participate in Piccolipiù if they were ≥ 18 years of age, lived in the catchment area of the maternity centers, knew sufficient Italian to provide informed consent and to complete questionnaires, and could provide a telephone number for future communication. Families were contacted 6, 12, 24 and 48 months after delivery to collect follow-up information by questionnaire, and children underwent a medical examination at four years of age. Cord blood was collected and stored in a centralized biobank. Ethical approvals have been obtained from the Ethics Committees of the Local Health Unit Roma E (management center), of the Istituto Superiore di Sanità (National Institute of Public Health), and of each local center. Parents provided written informed consent. A sample of 99 newborns from the Turin Center, who resided in Turin, had growth data at birth and until at least two years of age, and provided a cord blood sample, was included in the EXPOsOMICS Children Studies. A subsample of 97 mother-newborn pairs, eight with GDM, with complete covariate information was included in the current meta-analysis of GDM and cord blood DNA methylation.

Gestational Diabetes Classification

Information on GDM was obtained from the baseline questionnaire, which was administered to the mothers at recruitment (in the Turin center participants are recruited at admittance to the hospital for delivery). Information on absence/presence of GDM was also gathered from medical records. Analyses for the current meta-analyses used the information obtained from the questionnaire. All GDM cases identified by questionnaires were confirmed by medical records except for one case, which was reported in questionnaires, but not in medical records. GDM had been ascertained from medical records, according to the IADPSG criteria (22). Women were classified as having GDM if fasting glucose was ≥ 92 mg/dl or if a 75 g OGTT with 1 hour plasma glucose ≥ 180 mg/dl (10.0 mmol/l) or a 2-hour plasma glucose ≥ 153 mg/dl (8.5 mmol/l). Pregnant women at high risk (one of the following risk factors: BMI ≥ 30; previous GDM; random plasma glucose pre-pregnancy or in the first trimester between 100 and 125 mg/dl) were screened for GDM at week 16-18 of gestation with a 75 g OGTT, while women at lower risk (one the following risk factors: age ≥ 35 years; BMI ≥ 25; family history, previous large for gestational age infant; mother born in countries with high risk for diabetes) were screened at week 24-28 of gestation.

SUPPLEMENTARY DATA

Covariates

Information on maternal age, pre-pregnancy BMI, smoking, education, and origin was obtained from a baseline questionnaire administered at recruitment (in the Turin center, participants are recruited at admittance to the hospital for delivery). Maternal smoking status was defined as smoking during pregnancy yes or no. Maternal education was categorized as primary school, secondary school, and university degree or higher. With respect to race, information on the country of birth of the mother and of the parents of the mother was available. As this cohort is quite homogeneous with respect to ethnicity, the ancestry of all newborns was set to European for the current study. Estimated proportions of seven cord blood cell types were determined with the estimateCellCounts function in minfi (7), using the cord blood reference dataset (8).

Project Viva

Design and Study Population

Project Viva is a population-based prospective pregnancy cohort of mothers and children in Eastern Massachusetts, United States, which has been described previously (33). Women were enrolled from 1999 to 2002, with a final enrollment of 2,128 live births. Exclusion criteria for the study included multiple gestation, an inability to answer questions in English, a gestational age ≥ 22 weeks at recruitment, and plans to move outside the study area prior to delivery. The study was approved by the Institutional Review Board of Harvard Pilgrim Health Care, and participating women provided written informed consent.

Gestational Diabetes Classification

GDM was classified based on OGTT results, abstracted from medical records, using the Carpenter Coustan criteria (15). Clinicians screened all pregnant women at 26-28 weeks of gestation with a non-fasting oral glucose challenge test, in which venous blood was sampled 1 h after a 50 g oral glucose load. If the blood glucose exceeded 140 mg/dl, the clinician referred the woman for a fasting 3-h 100 g OGTT. OGTT results were considered abnormal if blood glucose levels exceeded 95 mg/dl at fasting, 180 mg/d 1 h post-glucose, 155 mg/dl 2 h post-glucose, or 140 mg/dl 3 h post-glucose. Women with two or more abnormal OGTT values were diagnosed with GDM. The remaining women served as controls for the current analysis.

Methylation Measurements

DNA was extracted from cord blood using the Qiagen Puregene Kit (Valencia, CA) and was bisulfite converted using the EZ-96 DNA Methylation kit (Zymo Research Corporation, Irvine, USA). DNA methylation levels were measured by Illumina FastTrack Microarray Services (San Diego, CA), using the Infinium HumanMethylation450 BeadChip kit (Illumina Inc., San Diego, USA). A stratified randomization was used to ensure balance of cohort characteristics across sample plates/batches. Arrays were rerun if detection p-values were not < 0.05 for > 99% of probes. Samples were excluded if there were any identity concerns (i.e., inconsistent genotyping and/or inferred sex). 485 unique samples remained, and a subset of 482 were included in the current meta-analyses (N = 23 GDM cases, N = 459 controls). Allosomal probes, non-CpG probes, and probes with detection p-values > 0.05 were excluded. Additionally, probes with values > 3*interquartile range were excluded. A total of 467,471 probes were included in the current analyses.

SUPPLEMENTARY DATA

Covariates

Information on maternal age, education, and race/ethnicity, was collected by interview or questionnaire at enrollment during the first trimester of pregnancy. Maternal age at delivery was included in statistical models as a continuous covariate. Maternal education was categorized into less than high school, high school/GED, some college, college degree, and graduate degree. Race/ethnicity of the mother was categorized as black, white, or other. Maternal smoking status during pregnancy was defined as ever or never smoking, based on current smoking reports in the early pregnancy questionnaire, smoking in the past three months based on the mid-pregnancy or delivery questionnaires, or indication of maternal smoking in maternal medical records. Cord blood cell fraction estimations were conducted with the estimateCellCounts function, using the "cordblood" designation for the compositeCellType input (8), in the minfi package in R (7). Maternal BMI for this study was calculated from mother's report of height and pre-pregnancy weight.

SUPPLEMENTARY DATA

Supplementary Acknowledgements

ALSPAC:

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. We would like to acknowledge Tom Gaunt, Oliver Lyttleton, Sue Ring, Nabila Kazmi, and Geoff Woodward for their earlier contributions to the generation of ARIES data (ALSPAC methylation data).

GOYA:

GOYA (Genomics of Obesity in Young Adults) was sampled as a case-cohort study within the Danish National Birth Cohort. The Danish National Birth Cohort was established with a significant grant from the Danish National Research Foundation. Additional support was obtained from the Danish Regional Committees, the Pharmacy Foundation, the Egmont Foundation, the March of Dimes Birth Defects Foundation, the Health Foundation and other minor grants. The DNBC Biobank has been supported by the Novo Nordisk Foundation and the Lundbeck Foundation.

INMA:

INMA researchers would like to thank all the participants for their generous collaboration. INMA researchers are grateful to Silvia Fochs, Nuria Pey, and Muriel Ferrer for their assistance in contacting the families and administering the questionnaires. A full roster of the INMA Project Investigators can be found at http://www.proyectoinma.org/presentacion-inma/listado-investigadores/en_listado-investigadores.html.

PREDO:

The PREDO study would not have been possible without the dedicated contribution of the PREDO study group members: H. Laivuori, P. Villa, E. Hämäläinen, A. Aitokallio-Tallberg, A-M. Henry, V. K. Hiilesmaa, T. Karipohja, R. Meri, S. Sainio, T. Saisto, S. Suomalainen-König, V-M. Ulander, T. Vaitilo (Department of Obstetrics and Gynaecology, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland), L. Keski-Nisula (Kuopio University Hospital, Kuopio Finland), E. Koistinen, T. Walle, R. Solja, P. Taipale (Northern Karelia Central Hospital, Joensuu, Finland), M. Kurkinen (Päijät-Häme Central Hospital, Lahti, Finland), P. Staven (Iisalmi Hospital, Iisalmi, Finland) and J. Uotila (Tampere University Hospital, Tampere, Finland). We also thank the PREDO cohort mothers, fathers, and children for their enthusiastic participation.

RHEA/ENVIR*ON*AGE/Piccolipiù:

We thank all of the participants of the Rhea study and the interviewers, statisticians and the hospital personnel for their cooperation and contribution to this study.

We are extremely grateful to the women and neonates participating in ENVIR*ON*AGE, as well as the staff of the maternity ward, the midwives, and the staff of the clinical laboratory of East-Limburg Hospital in Genk.

SUPPLEMENTARY DATA

Our thanks go to all the families who took part in the Piccolipiù study, to the midwives for their help in recruiting them, and to the whole Piccolipiù team, which includes doctors, nurses, research scientists, and computer/laboratory technicians.

The authors would also like to acknowledge Paolo Vineis for his input and coordination of the EXPOsOMICS project. We thank Mr. Cyrille Cuenin for his help and dedication in the methylome experimental workflow and the Genetic Cancer Susceptibility Group at IARC for their assistance in these experiments.

Project Viva:

We are indebted to the Project Viva mothers, children, and families

Funding Support

APS was supported by a grant from the National Institute of Environmental Health Sciences [grant number K99ES025817].

BC was supported by a postdoctoral fellowship from the Research Foundation - Flanders [grant number 12Q0517N].

CGH was supported by grants from the National Institute of Environmental Health Sciences [grant numbers K99ES030400 and T32ES013678]

CVB was supported by grants from the National Institute of Environmental Health Sciences [grant numbers NIH grants P30ES007048 and P50ES026086] and the Environmental Protection Agency [grant number 83615801-0]

ALSPAC:

JJ, DAL, CLR, and GCS are members of the MRC Integrative Epidemiology Unit, which receives funds from the University of Bristol and United Kingdom Medical Research Council [grant numbers MC_UU_00011/5, MC_UU_00011/6]. DAL's contribution to this work is supported by grants from the United States National Institutes of Health [grant number R01 DK1034] and the European Union's Seventh Framework Programme [grant number FP/2007-2013] / ERC Grant Agreement [grant number 66945; DevelopObese]. DAL is a National Institute of Health Research Senior Investigator [NF-SI-0611-10196]. The United Kingdom Medical Research Council and the Wellcome Trust [grant number 102215/2/13/2] and the University of Bristol provide core support for ALSPAC. The Accessible Resource for Integrated Epigenomics Studies (ARIES) which generated large scale methylation data was funded by the United Kingdom Biotechnology and Biological Sciences Research Council (grant numbers BB/I025751/1, BB/I025263/1]. Additional epigenetic profiling on the ALSPAC cohort was supported by the United Kingdom Medical Research Council Integrative Epidemiology Unit and the University of Bristol [grant numbers MC_UU_12013_1, MC_UU_12013_2, MC_UU_12013_5, MC_UU_12013_8], the Wellcome Trust [grant number WT088806], and the United States National Institute of Diabetes and Digestive and Kidney Diseases [grant number R01 DK10324]. The views expressed in this paper are those of the authors and not necessarily any funders. The funders had no influence on the content of the paper.

SUPPLEMENTARY DATA

GOYA:

Genotyping for the GOYA Study was funded by the Wellcome Trust [grant number 084762MA]. Generation of DNA methylation data was funded by the MRC Integrative Epidemiology Unit which is supported by the Medical Research Council [grant number MC_UU_12013/1-9] and the University of Bristol.

Healthy Start:

The Healthy Start study was supported by grants from the National Institute of Diabetes and Digestive and Kidney Diseases [grant number R01DK076648] and the National Institute of Environmental Health Sciences [grant number R01ES022934].

INMA:

INMA was funded by grants from the Instituto de Salud Carlos III Red INMA [grant numbers G03/176, CB06/02/0041], Spanish Ministry of Health [grant numbers FIS-PI041436, FIS-PI081151], Generalitat de Catalunya-CIRIT [grant number 1999SGR 00241], and the Fundació La marató de TV3 [grant number 090430].

PREDO:

The PREDO study has received funding from the Academy of Finland, EraNet Neuron, EVO (a special state subsidy for health science research), University of Helsinki Research Funds, the Signe and Ane Gyllenberg foundation, the Emil Aaltonen Foundation, the Finnish Medical Foundation, the Jane and Aatos Erkko Foundation, the Novo Nordisk Foundation, the Paivikki and Sakari Sohlberg Foundation, and the Sigrid Juselius Foundation.

RHEA/ENVIR*ON*AGE/Piccolipiù:

The Rhea project was financially supported by European Union (EU) grants for specific projects [grant numbers EU FP6-2003-Food-3-NewGeneris, EU FP6. STREP HiWATE, EU FP7 ENV.2007.1.2.2.2. Project no. 211250 ESCAPE, EU FP7-2008-ENV-1.2.1.4 Envirogenomarkers, EU FP7-HEALTH-2009-single stage CHICOS, EU FP7 ENV.2008.1.2.1.6. Proposal no. 226285 ENRIECO. EU-FP7-HEALTH-2012 Proposal no. 308333 HELIX], MeDALL FP7 European Union project [grant number 264357], and the Greek Ministry of Health (Program of Prevention of obesity and neurodevelopmental disorders in preschool children, in Heraklion district, Crete, Greece: 2011–2014; "Rhea Plus": Primary Prevention Program of Environmental Risk Factors for Reproductive Health, and Child Health: 2012–15).

The ENVIR*ON*AGE birth cohort is supported by the European Research Council [grant number ERC-2012-StG.310898] and by funds from the Flemish Scientific Research council [grant number FWO, G.0.733.15.N].

Piccolipiù was approved and initially funded by the Italian National Centre for Disease Prevention and Control CCM grant [grant number 2010) and by the Italian Ministry of Health [grant number art 12, 12bis Dl.gs.vo 502/92).

SUPPLEMENTARY DATA

The methylation assays of the three cohorts were funded by the European Community's Seventh Framework Programme FP7/2007–2013 project EXPOsOMICS [grant number 308610]. ZH is Co-PI of this grant. ZH, AG, and the Epigenetics Group at IARC are supported by grants from the Institut National du Cancer (INCa, Plan Cancer-EVA-INSERM, France).

Project Viva:

The Project Viva cohort is funded by National Institutes of Health grants [grant numbers R01 HL111108, R01 NR013945, and R01 HD034568].

SUPPLEMENTARY DATA

Results from Meta-Analyses that were Unadjusted for Estimated Cord Blood Cell Proportions

Five differentially methylated CpGs (cg06455422, cg06905453, cg07569069, cg15114431, cg21380280) (uncorrected p-values = 2.1×10^{-7}, 2.0×10^{-7}, 1.5×10^{-7}, 6.5×10^{-8}, and 3.8×10^{-7}, respectively) were identified in meta-analyses that were adjusted only for the baseline set of covariates. After adjusting for estimated cord blood cell fractions, effect estimates for these CpGs were attenuated, and they were no longer identified as differentially methylated ($P_{FDR} \geq 0.10$). The genomic inflation factor was 1.25 prior to adjusting for estimated cord blood cell fractions and was reduced to 1.15 after adjusting for the estimated cord blood cell fractions

CpGs Identified as Differentially Methylated in Meta-Analyses Unadjusted for Estimated Cord Blood Cell Fractions*

CpG	Genomic Position	%Methylation Difference (95%CI)†	Direction by Cohort‡	Raw P-Value	FDR-Corrected P-Value	Bonferroni-Corrected P-Value	Heterogeneity P-Value§	I^2	Relation to CpG Island‖	Target Gene¶	Gene Region Feature Category#
cg06455422	chr13:98875360	-1.5% (-2.0%, -0.9%)	------	2.1×10^{-7}	0.02	0.08	0.05	53.8	Open Sea	*FARP1*	Body
cg06905453	chr19:14089284	-1.2% (-1.7%, -0.8%)	----+-	2.1×10^{-7}	0.02	0.08	4.4×10^{-3}	70.7	North Shore	*RFX1*	Body
cg07569069	chr19:11485349	0.2% (0.1%, 0.2%)	+-+-++	1.5×10^{-7}	0.02	0.06	1.9×10^{-3}	73.7	Island	*C19orf39*	TSS200
cg15114431	chr3:102775941	-0.3% (-0.4%, -0.2%)	--+---	6.5×10^{-8}	0.02	0.02	0.23	27.5	Open Sea	N/A	N/A
cg21380280	chr3:193311632	-0.3% (-0.4%, -0.2%)	++----	3.8×10^{-7}	0.03	0.15	0.15	38.8	South Shore	*OPA1*	Body

*Results are from inverse-weighted fixed effects meta-analysis, conducted using METAL. Each cohort independently ran robust linear regression models, adjusting for newborn's sex, maternal age (in years), maternal body mass index (early pregnancy or pre-pregnancy), maternal smoking status during pregnancy, maternal education, maternal genetic ancestry (if available) or maternal race/ethnicity, and estimated proportions of B cells, $CD8^+$ T cells, $CD4^+$ T cells, granulocytes, natural killer cells, monocytes, and nucleated red blood cells in cord blood.

†%Difference in newborn DNA methylation and 95% confidence interval, comparing the gestational diabetes case group to the control group

‡Direction of association between gestational diabetes and methylation at the locus of interest by cohort, ordered as follows: ALSPAC, GOYA, Healthy Start, INMA, PREDO, RHEA/ENVIR*ON*AGE/Piccolipiù, Project Viva

§The heterogeneity p-value and I^2 were calculated by METAL using Cochran's Q-test for heterogeneity.

‖Relationship to CpG islands from the UCSC database, annotated using the Illumina 450k manifest (34).

¶Target gene name(s) from the UCSC database, annotated using the Illumina 450k manifest (34).

#Gene region feature category describing the position of the CpG from UCSC, annotated using the Illumina 450k manifest (34). Gene body indicates that the CpG is located between the ATG and stop codon.

SUPPLEMENTARY DATA

References:

1. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. Int J Epidemiol. 2013;42(1):111-27.
2. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, et al. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. Int J Epidemiol. 2013;42(1):97-110.
3. Fraser A, Nelson SM, Macdonald-Wallis C, Lawlor DA. Associations of existing diabetes, gestational diabetes, and glycosuria with offspring IQ and educational attainment: the Avon Longitudinal Study of Parents and Children. Exp Diabetes Res. 2012;2012:963735.
4. Lawlor DA, Fraser A, Lindsay RS, Ness A, Dabelea D, Catalano P, et al. Association of existing diabetes, gestational diabetes and glycosuria in pregnancy with macrosomia and offspring body mass index, waist and fat mass in later childhood: findings from a prospective pregnancy cohort. Diabetologia. 2010;53(1):89-97.
5. Relton CL, Gaunt T, McArdle W, Ho K, Duggirala A, Shihab H, et al. Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES). Int J Epidemiol. 2015;44(4):1181-90.
6. Touleimat N, Tost J. Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics. 2012;4(3):325-41.
7. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363-9.
8. Bakulski KM, Feinberg JI, Andrews SV, Yang J, Brown S, S LM, et al. DNA methylation of cord blood cell types: Applications for mixed cell birth studies. Epigenetics. 2016;11(5):354-62.
9. Leek J, Johnson W, Fertig E, Storey J, Zhang Y, Torres L. SVA:Surrogate Variable Analysis. R Package Version 3280. 2018.
10. Paternoster L, Evans DM, Nohr EA, Holst C, Gaborieau V, Brennan P, et al. Genome-wide population-based association study of extremely overweight young adults–the GOYA study. PloS one. 2011;6(9):e24303.
11. Nohr EA, Timpson NJ, Andersen CS, Smith GD, Olsen J, Sørensen TI. Severe obesity in young women and reproductive health: the Danish National Birth Cohort. PloS one. 2009;4(12):e8444.
12. Lind T, Phillips PR. Influence of pregnancy on the 75-g OGTT: a prospective multicenter study. Diabetes. 1991;40(Supplement 2):8-13.
13. Min J, Hemani G, Smith GD, Relton CL, Suderman M. Meffil: efficient normalisation and analysis of very large DNA methylation samples. bioRxiv. 2017:125963.
14. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:86.
15. Carpenter MW, Coustan DR. Criteria for screening tests for gestational diabetes. American journal of obstetrics and gynecology. 1982;144(7):768-73.

16. Guxens M, Ballester F, Espada M, Fernandez MF, Grimalt JO, Ibarluzea J, et al. Cohort Profile: the INMA--INfancia y Medio Ambiente--(Environment and Childhood) Project. Int J Epidemiol. 2012;41(4):930-40.
17. Carpenter MW, Coustan DR. Criteria for screening tests for gestational diabetes. American Journal of Obstetrics & Gynecology. 1982;144(7):768-73.
18. Fortin J-P, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome biology. 2014;15(11):503.
19. Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan S-T, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. Genome biology. 2015;16(1):37.
20. Casas M, Chatzi L, Carsin AE, Amiano P, Guxens M, Kogevinas M, et al. Maternal pre-pregnancy overweight and obesity, and child neuropsychological development: two Southern European birth cohort studies. Int J Epidemiol. 2013;42(2):506-17.
21. Girchenko P, Lahti M, Tuovinen S, Savolainen K, Lahti J, Binder EB, et al. Cohort Profile: Prediction and prevention of preeclampsia and intrauterine growth restriction (PREDO) study. Int J Epidemiol. 2017;46(5):1380-1g.
22. Metzger BE, Gabbe SG, Persson B, Buchanan TA, Catalano PA, Damm P, et al. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care. 2010;33(3):676-82.
23. Morin AM, Gatev E, McEwen LM, MacIsaac JL, Lin DTS, Koen N, et al. Maternal blood contamination of collected cord blood can be identified using DNA methylation at three CpGs. Clin Epigenetics. 2017;9:75.
24. 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.
25. Price ME, Cotton AM, Lam LL, Farre P, Emberly E, Brown CJ, et al. Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. Epigenetics Chromatin. 2013;6(1):4.
26. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118-27.
27. Vineis P, Chadeau-Hyam M, Gmuender H, Gulliver J, Herceg Z, Kleinjans J, et al. The exposome in practice: Design of the EXPOsOMICS project. International journal of hygiene and environmental health. 2017;220(2 Pt A):142-51.
28. Chatzi L, Leventakou V, Vafeiadi M, Koutra K, Roumeliotaki T, Chalkiadaki G, et al. Cohort Profile: The Mother-Child Cohort in Crete, Greece (Rhea Study). Int J Epidemiol. 2017;46(5):1392-3k.
29. Vafeiadi M, Roumeliotaki T, Chalkiadaki G, Rantakokko P, Kiviranta H, Fthenou E, et al. Persistent organic pollutants in early pregnancy and risk of gestational diabetes mellitus. Environ Int. 2017;98:89-95.
30. Daraki V, Roumeliotaki T, Koutra K, Georgiou V, Kampouri M, Kyriklaki A, et al. Effect of parental obesity and gestational diabetes on child neuropsychological and behavioral development at 4 years of age: the Rhea mother-child cohort, Crete, Greece. Eur Child Adolesc Psychiatry. 2017;26(6):703-14.

SUPPLEMENTARY DATA

31. Janssen BG, Madhloum N, Gyselaers W, Bijnens E, Clemente DB, Cox B, et al. Cohort Profile: The ENVIRonmental influence ON early AGEing (ENVIRONAGE): a birth cohort study. International journal of epidemiology. 2017;46(5):1386-7m.
32. Farchi S, Forastiere F, Vecchi Brumatti L, Alviti S, Arnofi A, Bernardini T, et al. Piccolipiù, a multicenter birth cohort in Italy: protocol of the study. BMC pediatrics. 2014;14:36.
33. Oken E, Baccarelli AA, Gold DR, Kleinman KP, Litonjua AA, De Meo D, et al. Cohort profile: project viva. Int J Epidemiol. 2015;44(1):37-48.
34. Hansen K. IlluminaHumanMethylation450kanno.ilmn12.hg19: Annotation for Illumina's 450k methylation arrays. R package version 060. 2016.

SUPPLEMENTARY DATA

Supplementary Table S1. Cohort-Specific Covariate Frequencies

	ALSPAC	GOYA	Healthy Start	INMA	PREDO	RHEA/ENVIR*ON*AGE/Piccolipiù	Viva
Ancestry or Race/Ethnicity (%)	889 (100.0) European	432 (100.0) European	139 (24.6) Hispanic, 305 (53.9) Non-Hispanic white, 83 (14.7) African American, 39 (6.9) All Others	156 (100.0) European	780 (100.0) European	353 (94.9) European, 19 (5.1) Non-European	57 (11.8) Black, 341 (70.7) White, 84 (17.5) Other
Maternal Age (Years), Mean (SD)	29.4 (4.6)	29.4 (4.0)	27.6 (6.2)	29.7 (3.6)	33.4 (5.7)	30.5 (4.8)	32.6 (5.4)
Maternal Body Mass Index (kg/m^2), Mean (SD)	23.5 (4.83)	37.0 (3.3)	26 (6.8)	23.5 (3.9)	26.9 (6.5)	23.8 (4.5)	24.7 (5.3)
Maternal Education*, (%)							
Level 1	458 (51.5)	70 (16.2)	86 (15.2)	36 (22.8)	37 (4.7)	41 (11.0)	11 (2.3)
Level 2	431 (48.5)	189 (43.8)	102 (18.0)	83 (53.2)	317 (40.6)	154 (41.4)	33 (6.8)
Level 3	--	173 (40.0)	124 (21.9)	37 (23.7)	181 (23.2)	177 (47.6)	118 (24.5)
Level 4	--	--	123 (21.7)	--	245 (31.4)	--	170 (35.3)
Level 5	--	--	131 (23.1)	--	--	--	150 (31.1)
Maternal Smoking Status During Pregnancy, N (%)							
Ever	122 (14)	98 (22.7)	50 (8.8)	53 (34.0)	32 (4.1)	119 (32.0)	54 (11.2)
Never	767 (86)	334 (77.3)	516 (91.2)	103 (66.0)	748 (95.9)	253 (68.0)	428 (88.8)
Fetal Sex, N (%)							
Male	438 (49.3)	223 (51.6)	292 (51.6)	87 (55.8)	411 (52.7)	197 (53.0)	252 (52.3)
Female	451 (50.7)	209 (48.4)	274 (48.4)	69 (44.2)	369 (47.3)	175 (47.0)	230 (47.7)

*Maternal education categories by cohort: ALSPAC (level 1 = less than A level, level 2 = A level or above), GOYA (level 1 = unskilled work or on public benefit, level 2 = work that requires shorter training or artisans, level 3 = managers or long-medium education), Healthy Start (level 1 = less than high school, levels 2 = high school/GED, level 3 = some college, level 4 = completed college, level 5 = graduate degree); INMA (level 1 = less than high school, level 2 = high school, level 3 = college or higher); PREDO (level 1 = primary, level 2 = secondary, level 3 = lower tertiary, level 4 = upper tertiary), RHEA/ENVIR*ON*AGE/Piccolipiù (level 1 = primary school, level 2 = secondary school, level 3 = university degree or higher), Project Viva (level 1 = less than high school, level 2 = high school/GED, level 3 = some college, level 4 = college degree, level 5 = graduate degree)

SUPPLEMENTARY DATA

Supplementary Table S4. Cohort and Model-Specific Lambdas and Probe Numbers

	Unadjusted for Estimated Cord Blood Cell Fractions*		Final Model (Additionally Adjusted for Estimated Cord Blood Cell Fractions)†	
Study	λ‡	Number of Probes§	λ‡	Number of Probes§
ALSPAC	2.7	373,668	1.2	373,668
GOYA	1.0	380,878	0.9	380,878
HEALTHY START	1.0	380,259	1.0	380,259
INMA	1.8	377,768	1.5	377,769
PREDO	0.9	371,173	1.0	371,173
RHEA/ENVIR*ON*AGE/Piccolipiù	1.2	380,878	1.1	380,878
Viva	2.0	378,476	1.6	378,476

*Robust linear regression models were used to evaluate associations between *in utero* exposure to gestational diabetes mellitus and methylation levels at each CpG on the Infinium HumanMethylation, adjusting for newborn's sex, maternal age, maternal body mass index (pre-pregnancy or early pregnancy), maternal smoking status during pregnancy, maternal education, and maternal genetic ancestry (if available) or maternal race/ethnicity. Inverse variance-weighted fixed effects meta-analyses were then run using METAL.

†Robust linear regression models were used to evaluate associations between *in utero* exposure to gestational diabetes mellitus and methylation levels at each CpG on the Infinium HumanMethylation, adjusting for newborn's sex, maternal age, maternal body mass index (pre-pregnancy or early pregnancy), maternal smoking status during pregnancy, maternal education, maternal genetic ancestry (if available) or maternal race/ethnicity, and estimated cord blood cell fractions. Inverse variance-weighted fixed effects meta-analyses were then run using METAL.

‡Lambda estimated using all probes provided by the cohort, after excluding cross-reactive probes, polymorphic probes, probes with a single nucleotide polymorphism at the single base pair extension, control probes, and probes on the X or Y chromosomes

§Number of probes included in the cohort-specific analysis, after excluding cross-reactive probes, polymorphic probes, probes with a single nucleotide polymorphism at the single base pair extension, control probes, and probes on the X or Y chromosomes

SUPPLEMENTARY DATA

Supplementary Table S5. Sample Sizes and λ Values for Leave-One-Out Meta-Analyses*

Cohort Excluded	N Cases	N Controls	λ
ALSPAC	295	2,495	1.14
GOYA	289	2,958	1.16
Healthy Start	285	2,828	1.17
INMA	305	3,216	1.12
PREDO	137	2,762	1.20
RHEA/ENVIR*ON*AGE/Piccolipiù	297	3,010	1.13
Project Viva	294	2,903	1.23

*Robust linear regression models evaluated *in utero* exposure to maternal gestational diabetes mellitus and cord blood methylation at each CpG site on the 450k array, adjusting for newborn's sex, maternal age, body mass index (pre-pregnancy or early pregnancy), maternal smoking status during pregnancy, maternal education, and maternal genetic ancestry (if available) or maternal race/ethnicity, and estimated cord blood cell fractions

SUPPLEMENTARY DATA

Supplementary Table S8. Differentially Methylated Regions Identified by Comb-p*

DMR	450k CpGs	Direction of Association	Sidák-Corrected Region P-Value	Nearby Genes	Regulatory Feature Group/Gene Group/Relation to Island
Chr1:248100345-248100614	cg00785941, cg03748376, cg04028570, cg08260406, cg08944170, cg20434529, cg20507276	-	1.8×10^{-9}	*OR2L13*	Promoter associated/1st Exon:5'UTR or TSS200/Island
Chr4:165878037-165878219	cg00393585, cg06481168, cg08992305, cg11630554, cg12861945, cg20697094	-	3.4×10^{-5}	*C4orf39*, *TRIM61*	Promoter associated cell type specific/TSS200:Body, Body:1st Exon, or Body:1st Exon:5'UTR/Island
Chr6:31148332-31148666	cg03078486, cg09179646, cg09357589, cg11805138, cg11811828, cg14036627, cg17931227, cg22291762, cg22701603, cg23252259, cg24427850, cg26668675, cg27547543	+	2.6×10^{-5}	NA	Unclassified/NA/Island or South Shore
Chr10:135342218-135342413	cg10862468, cg25330361	-	8.9×10^{-3}	*CYP2E1*	Unclassified/Body/Island
Chr15:74592566-74592786	cg04629595, cg05926586, cg21565421	+	8.0×10^{-5}	*CCDC33*	NA/Body/Open Sea

Abbreviations Used: TSS, transcription start site; UTR, untranslated region

*The comb-p program was run on results from robust linear regression models evaluating maternal gestational diabetes status as the exposure of interest and methylation levels at each CpG site as the outcome, adjusting for newborn's sex, estimated cord blood cell fractions, and maternal age, body mass index (pre-pregnancy or early pregnancy), maternal smoking status during pregnancy, maternal education, and maternal genetic ancestry (if available) or maternal race/ethnicity. Comb-p results were identical when a window of either 500 or 1,000 base pairs was used.

SUPPLEMENTARY DATA

Supplementary Table S9. Differentially Methylated Regions Identified by DMRcate*

DMR	450k CpGs	Minimum FDR†	Stouffer‡	Maximum Beta Fold Change§	Mean Beta Fold Change‖	Nearest Genes	Regulatory Feature Group/Gene Group/Relation to Island
500 base pair window							
chr1:248100407-248100614	cg00785941, cg03748376, cg04028570, cg08260406, cg08944170, cg20507276	1.3×10^{-8}	0.31	-0.03	-0.03	*OR2L13*, *CLK3P2*	Promoter Associated/1st Exon, 5'UTR, or TSS200/Island
1,000 base pair window							
chr10:135341870-135342620	cg00321709, cg10862468, cg19469447, cg23400446, cg24530264 cg25330361	4.9×10^{-11}	0.16	-0.04	-0.02	*CYP2E1*	Unclassified/Body/Island

Abbreviations Used: FDR, false discovery rate; UTR, untranslated region

*DMRcate was run on results from robust linear regression models evaluating maternal gestational diabetes status as the predictor and cord blood methylation at each CpG site as the outcome, adjusting for newborn's sex, estimated cord blood cell fractions, and maternal age, body mass index (pre-pregnancy or early pregnancy), maternal smoking status during pregnancy, maternal education, and maternal genetic ancestry (if available) or maternal race/ethnicity.

†Minimum false discovery rate adjusted p-value among the CpGs making up the differentially methylated region

‡Stouffer transformation of the group of false discovery rate-corrected p-values for individual CpG sites belonging to the differentially methylated region

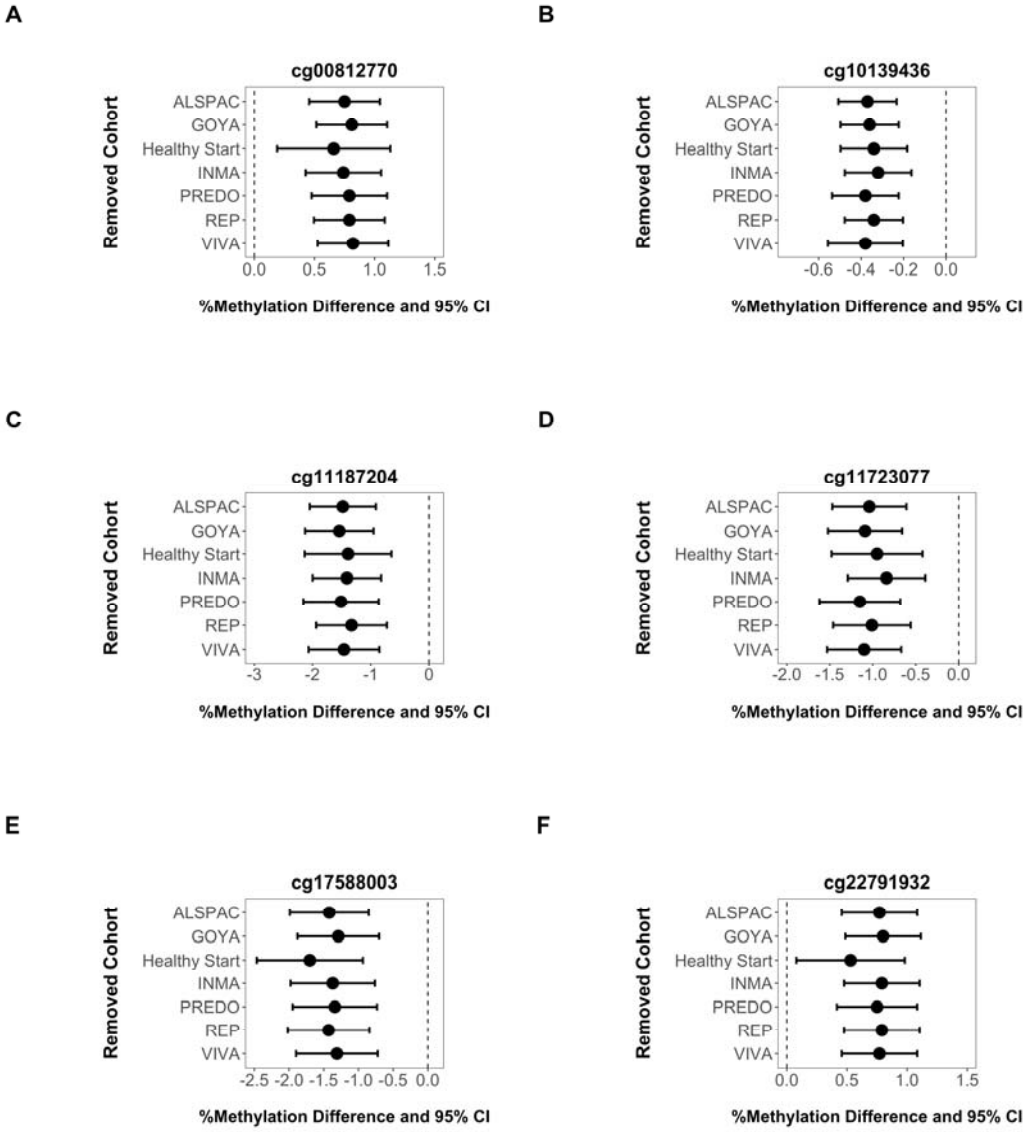
§Maximum absolute beta fold change within the region

‖Mean beta fold change within the region

SUPPLEMENTARY DATA

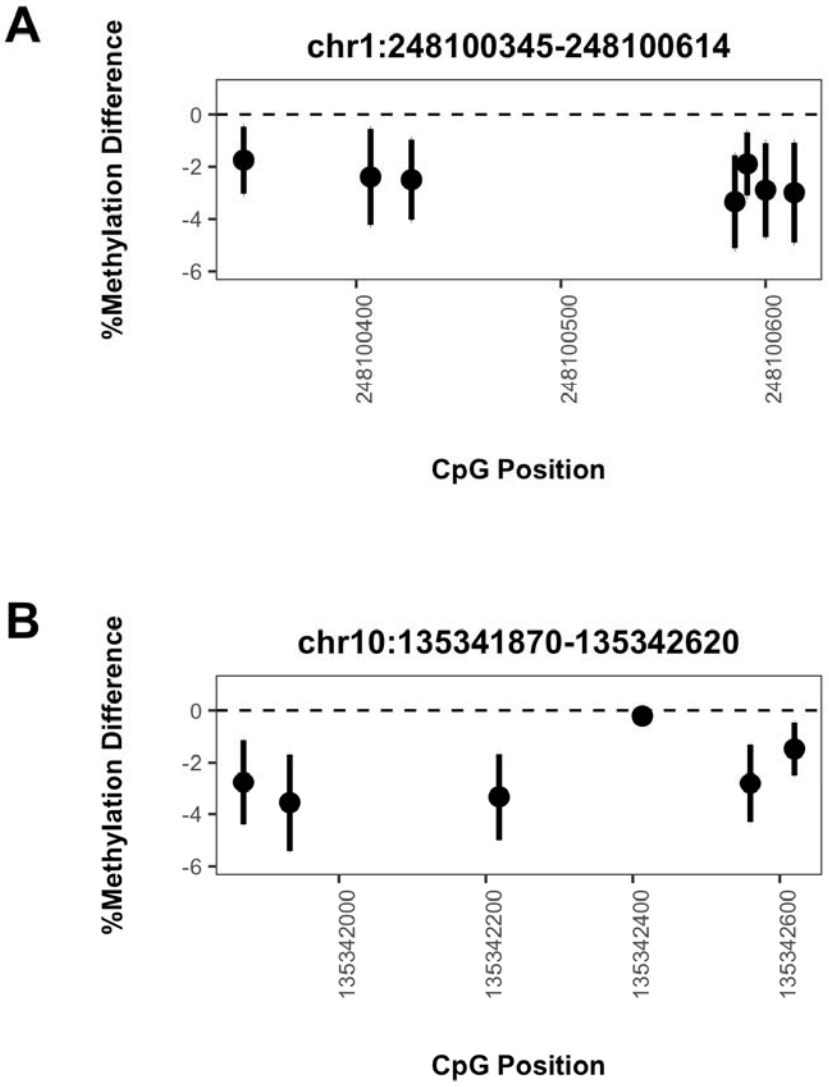
Supplementary Figure S1. Forest plots showing %methylation differences and 95% confidence intervals for leave-one out meta-analysis results for the six CpGs that were identified as differentially methylated by GDM status based on a $P_{FDR} < 0.10$ in the full meta-analysis: **(A)** cg00812770, **(B)** cg10139436, **(C)** cg11187204, **(D)** cg11723077, **(E)** cg17588003, and **(F)** cg22791932. The x-axis indicates the %methylation difference and 95% confidence interval for each CpG after removing the cohort listed on the y-axis and meta-analyzing results from the remaining six cohorts. Leave-one-out meta-analyses were run using results from robust linear regression models, which examined associations between *in utero* exposure to maternal gestational diabetes mellitus and cord blood DNA methylation levels, adjusting for newborn's sex, maternal age, maternal body mass index (early pregnancy or pre-pregnancy), maternal smoking status during pregnancy, maternal education, and maternal genetic ancestry (if available) or maternal race/ethnicity, and estimated cord blood cell fractions. REP stands for the pooled analysis of three cohorts (Rhea, ENVIR*ON*AGE, and Piccolipiù)

SUPPLEMENTARY DATA



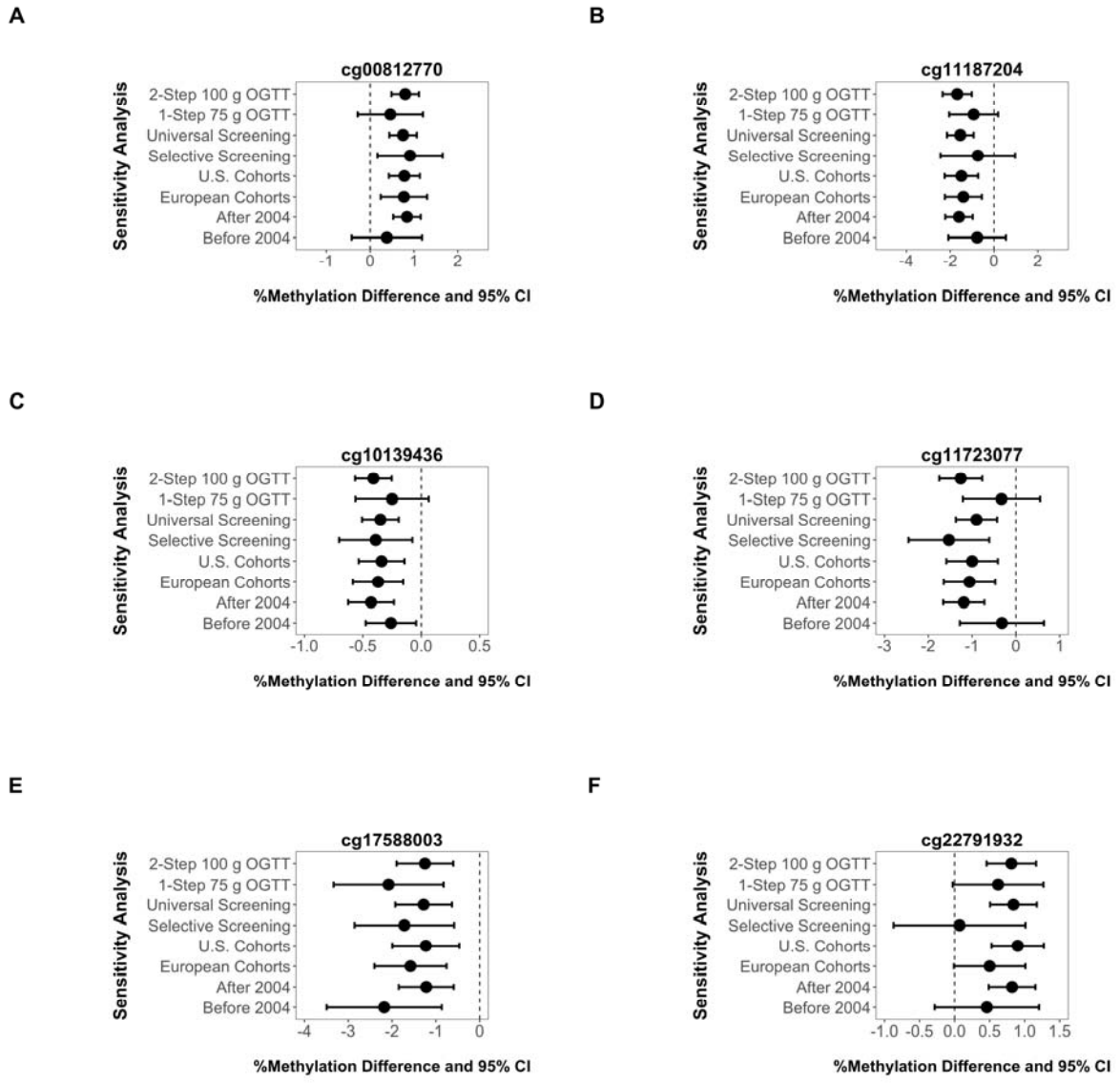
SUPPLEMENTARY DATA

Supplementary Figure S2. %Methylation differences for individual CpGs contained within the two regions identified as differentially methylated in cord blood by both comb-p and DMRcate in relation to *in utero* exposure to maternal gestational diabetes mellitus. **(A)** Seven CpGs are contained within the differentially methylated region identified on chromosome 1 and **(B)** six CpGs are contained within the differentially methylated region identified on chromosome 10. The %methylation difference comparing newborns exposed, versus unexposed, to maternal gestational diabetes mellitus *in utero* is shown on the y-axis with the corresponding 95% confidence interval. The genomic position of the CpG is shown on the x-axis**.**



SUPPLEMENTARY DATA

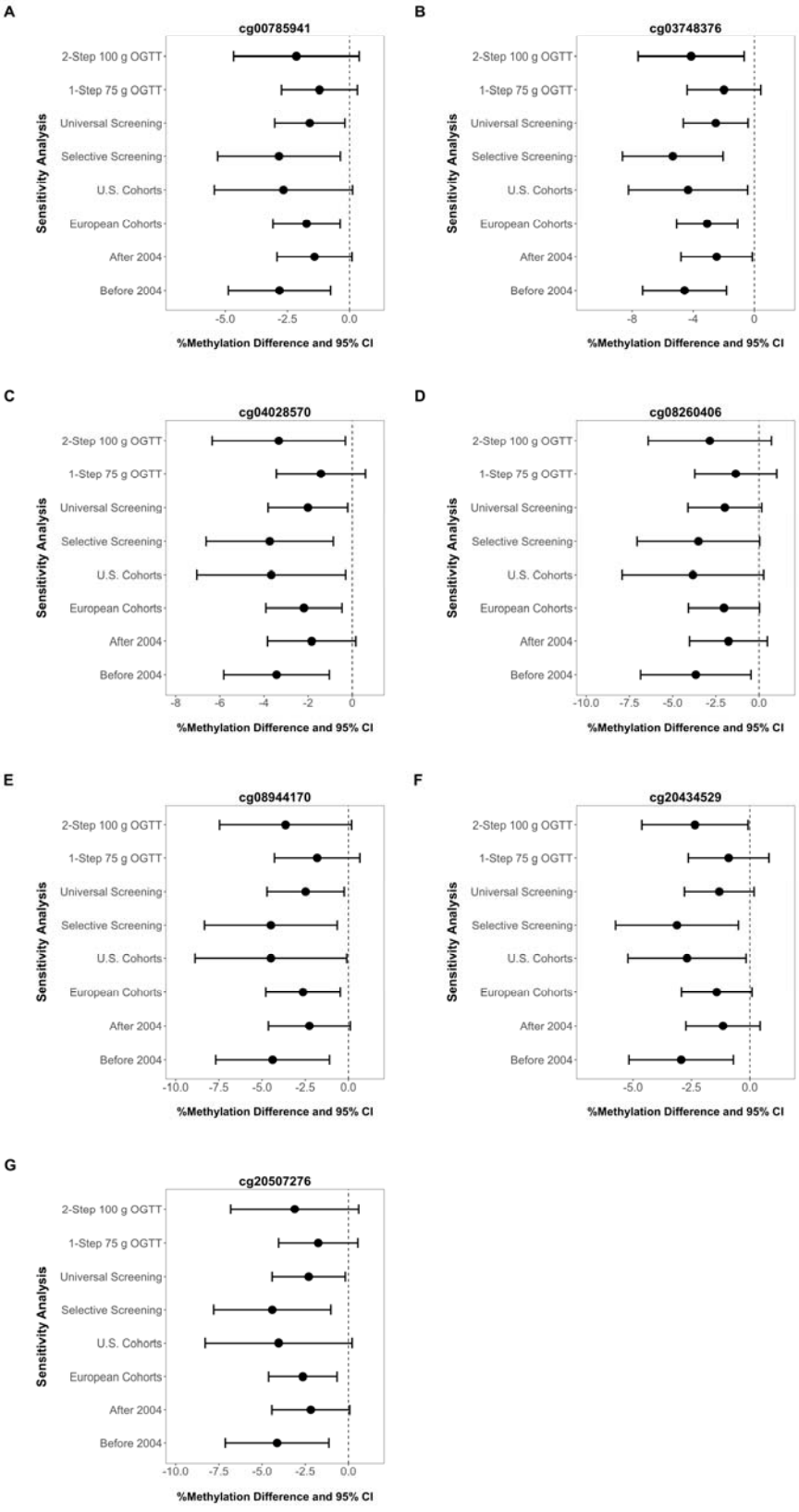
Supplementary Figure S3. Sensitivity Analyses for Individual CpG Results. A series of sensitivity meta-analyses were run, which compared results after restricting to cohorts with gestational diabetes mellitus (GDM) cases that had been classified by a 2-step 100 g oral glucose tolerance test (OGTT) versus a 1-step 75 g OGTT; with GDM cases that had been identified by universal versus selective screening; European versus U.S. cohorts; and cohorts that recruited participants prior to versus after 2004. The results are presented for the six CpGs that were found to be differentially methylated (based on a $P_{FDR} < 0.10$) in the individual CpG meta-analyses: **A)** cg00812770, **B)** cg11187204, **C)** cg10139436, **D)** cg11723077, **E)** cg17588003, and **F)** cg22791932. The x-axis indicates the %methylation difference and corresponding 95% confidence interval, comparing newborns who were exposed to maternal gestational diabetes *in utero* to unexposed newborns. The y-axis specifies the subgroup analysis.



SUPPLEMENTARY DATA

Supplementary Figure S4. Sensitivity Analyses for CpGs within the Differentially Methylated Region Identified on Chromosome 1. A series of sensitivity meta-analyses were run, which compared results after restricting to cohorts with gestational diabetes mellitus (GDM) cases that had been classified by a 2-step 100 g oral glucose tolerance test (OGTT) versus a 1-step 75 g OGTT; with GDM cases that had been identified by universal versus selective screening; European versus U.S. cohorts; and cohorts that recruited participants prior to versus after 2004. The results are presented for the seven CpGs contained within the differentially methylated region identified on chromosome 1 by both comb-p and DMRcate: **A)** cg00785941, **B)** cg03748376, **C)** cg04028570, **D)** cg08260406, **E)** cg08944170, **F)** cg20434529, **G)** cg20507276. The x-axis indicates the %methylation difference and corresponding 95% confidence interval, comparing newborns who were exposed to maternal gestational diabetes *in utero* to unexposed newborns. The y-axis specifies the subgroup analysis.

SUPPLEMENTARY DATA



SUPPLEMENTARY DATA

Supplementary Figure S5. Sensitivity Analyses for CpGs within the Differentially Methylated Region Identified on Chromosome 10. A series of sensitivity meta-analyses were run, which compared results after restricting to cohorts with gestational diabetes mellitus (GDM) cases that had been classified by a 2-step 100 g oral glucose tolerance test (OGTT) versus a 1-step 75 g OGTT; with GDM cases that had been identified by universal versus selective screening; European versus U.S. cohorts; and cohorts that recruited participants prior to versus after 2004. The results are presented for the six CpGs contained within the differentially methylated region identified on chromosome 10 by both comb-p and DMRcate: **A)** cg00321709, **B)** cg10862468, **C)** cg19469447, **D)** cg23400446, **E)** cg24530264, and **F)** cg25330361. The x-axis indicates the %methylation difference and corresponding 95% confidence interval, comparing newborns who were exposed to maternal gestational diabetes *in utero* to unexposed newborns. The y-axis specifies the subgroup analysis.