Title: Placental DNA methylation signatures of maternal smoking during pregnancy and potential impacts on fetal growth

Authors

Todd M. Everson[1,*,¥]
Marta Vives-Usano[2,3,4,*]
Emie Seyve[5,*]
Andres Cardenas[6,7]
Marina Lacasaña[4,8,9]
Jeffrey M. Craig[10,11,12]
Corina Lesseur[13]
Emily R. Baker[14]
Nora Fernandez-Jimenez[15,16,17]
Barbara Heude[18]
Patrice Perron[19]
Beatriz Gónzalez-Alzaga[8,9]
Jane Halliday[20,11]
Maya A. Deyssenroth[13]
Margaret R. Karagas[21]
Carmen Íñiguez[22,23,4]
Luigi Bouchard[24]
Pedro Carmona-Sáez[25]
Yuk J. Loke[10]
Ke Hao[26]
Thalia Belmonte[27]
Marie A. Charles[18]
Jordi Martorell-Marugán[25,28]
Evelyne Muggli[20,11]
Jia Chen[13]
Mariana F. Fernández[4,9,29]
Jorg Tost[30]
Antonio Gómez-Martín[31]
Stephanie J. London[32]
Jordi Sunyer[3,4,33,34]
Carmen J. Marsit[1,35,**]
Johanna Lepeule[5,**]
Marie-France Hivert[6,36,**]
Mariona Bustamante[2,3,4,33,**, ¥]

[*] *These authors contributed equally to this work*
[**] *These authors contributed equally to this work*
[¥] *Corresponding authors*

Correspondence

Correspondence can be addressed to Todd M. Everson (Todd.M.Everson@Emory.edu) and Mariona Bustamante (mariona.bustamante@isglobal.org).

Competing interests

The authors declare no competing interests.

Author Affiliations

[1] Department of Environmental Health, Rollins School of Public Health at Emory University, Atlanta, GA, USA
[2] Center for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, Barcelona, Spain
[3] Universitat Pompeu Fabra, Barcelona, Spain
[4] CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
[5] University Grenoble Alpes, Inserm, CNRS, IAB, 38000 Grenoble, France
[6] Department of Population Medicine, Harvard Medical School, Harvard Pilgrim Health Care Institute, Boston, MA, USA
[7] Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, Berkeley, CA, USA
[8] Andalusian School of Public Health, Granada, Spain
[9] Instituto de Investigación Biosantaria (ibs.GRANADA), Granada, Spain
[10] Molecular Epidemiology, Murdoch Children's Research Institute, Parkville, Australia
[11] Department of Paediatrics, University of Melbourne, Parkville, Australia
[12] Centre for Molecular and Medical Research, Deakin University, Geelong, Australia
[13] Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[14] Department of Obstetrics & Gynecology, Geisel School of Medicine at Dartmouth College, Lebanon, NH
[15] University of the Basque Country (UPV/EHU), Leioa, Spain
[16] Biocruces-Bizkaia Health Research Institute, Barakaldo, Spain
[17] Public Health Division of Gipuzkoa, Basque Government, San Sebastian, Spain
[18] Inserm U1153, Research Center for Epidemiology and Statistics (CRESS), Paris Descartes University, Villejuif, France
[19] Department of Medicine, University of Sherbrooke, Sherbrooke, QC, Canada
[20] Reproductive Epidemiology, Murdoch Children's Research Institute, Parkville, Australia
[21] Department of Epidemiology, Geisel School of Medicine at Dartmouth College, Hanover NH, USA
[22] Department of Statistics and Computational Research, Universitat de València, València, Spain
[23] Epidemiology and Environmental Health Joint Research Unit, FISABIO-Universitat Jaume I-Universitat de València, València, Spain
[24] Department of Biochemistry, University of Sherbrooke, Sherbrooke, QC, Canada
[25] Bioinformatics Unit, GENYO, Centre for Genomics and Oncological Research, Pfizer, University of Granada, Andalusian Regional Government, Granada, Spain
[26] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[27] IUOPA University of Oviedo, Oviedo, Spain
[28] Artys Health, Barcelona, Spain
[29] Biomedical Research Centre (CIBM) and School of Medicine, University of Granada, Granada, Spain
[30] Laboratory for Epigenetics and Environment, Centre National de Recherche en Génomique Humaine, CEA – Institut de Biologie François Jacob, Evry, France
[31] Genomics Unit. GENYO. Centre for Genomics and Oncological Research, Pfizer, University of Granada, Andalusian Regional Government, Granada, Spain

[32] Division of Intramural Research, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Durham, North Carolina, USA
[33] ISGlobal, Barcelona Institute for Global Health, Barcelona, Spain
[34] Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain
[35] Department of Epidemiology, Rollins School of Public health at Emory University, Atlanta, GA, USA
[36] Diabetes Unit, Massachusetts General Hospital, Boston, MA, USA

Abstract

Maternal smoking during pregnancy (MSDP) contributes to poor birth outcomes, in part through disrupted placental functions, which may be reflected in the placental epigenome. We meta-analyzed the associations between MSDP and placental DNA methylation (DNAm) and between DNAm and birth outcomes within the Pregnancy And Childhood Epigenetics (PACE) consortium (7 studies, N=1700, 344 with any MSDP). We identified 1224 CpGs that were associated with MSDP, of which 341 associated with birth outcomes and 141 associated with gene expression. Only 6 of these CpGs were consistent with the findings from a prior meta-analysis of cord blood DNAm, demonstrating substantial tissue-specific responses to MSDP. The placental MSDP associated CpGs were enriched for growth-factor signaling, hormone activity, inflammation, and vascularization, which play important roles in placental function. We demonstrate links between placental DNAm, MSDP and poor birth outcomes, which may better inform the mechanisms through which MSDP impacts placental function and fetal growth.

Keywords: DNA methylation, epigenetics, EWAS, pregnancy, smoking, birth weight, gestational age

Almost 1 in 10 pregnancies are impacted by the effects of maternal smoking during pregnancy (MSDP), with state-specific prevalence ranging from as low as 1.8% to as high as 27.1% in the USA[1], while in Europe, the prevalence of MSDP is estimated to range between 4.2% and 18.9%[2]. Consequently, the numerous health effects of MSDP remain a significant public health concern. The impact of this exposure on fetal development has been the source of significant investigation, resulting in MSDP being recognized as a cause of multiple negative pregnancy and birth outcomes[3].

The mechanisms that underlie this reproductive and developmental toxicity are partially understood and include molecular and anatomical changes of the placenta[4,5]. Additionally, experimental mouse models have recently highlighted the critical roles of proper placental function in ensuring successful pregnancy outcomes[6]. Epigenetic responses to prenatal exposures have emerged as potential intermediate links between early life exposures and developmental health outcomes, and epidemiologic studies of DNAm, particularly multi-cohort collaborative efforts, are powerful approaches to investigate these types of research questions[7]. So far, most studies of MSDP and epigenetics have focused on DNA methylation (DNAm) variations in cord blood, though some studies of placenta, peripheral blood and lung tissues have also been performed[8]. The Pregnancy and Childhood Epigenetics (PACE) consortium[9] published a large meta-analysis identifying thousands of MSDP-associated variations in DNAm within cord blood and child peripheral blood[10]. However, the placental epigenome has not been as thoroughly studied, though the placenta is likely a critical target organ of MSDP-associated toxicity. A handful of prior studies have examined the relationships between MSDP and (DNAm) in human placenta[11–14], identifying MSDP-associated loci some of which have been suggested partially mediate the effects of MSDP on lower birth weight[15].

These studies have begun to characterize the impact that MSDP has on the human placental epigenome but have been limited by small sample sizes and have not adjusted for placental tissue heterogeneity. We aimed to address these gaps by performing a fixed effects meta-analysis examining the relationships between MSDP and variations in the placental methylome across seven independent studies that are members of the PACE consortium. We also aimed to gain insights into the potential biological and health-related impacts of these associations by performing additional analyses with nearby mRNA expression, as well as functional, regulatory, and phenotypic enrichment, and a secondary meta-analysis of the associations between placental DNAm and birth outcomes. This study represents the largest and most comprehensive examination of MSDP associations with placental DNAm in human populations and provides novel insights into the placental molecular responses to MSDP and how these relate to fetal development.

Results

Study population

Seven American, Australian, and European studies (N=1,700) contributed to the epigenome-wide association study (EWAS) linking MSDP to placental DNAm: including Asking Questions about Alcohol in pregnancy (AQUA)[16], Study on the pre- and early postnatal determinants of child health and development (EDEN)[17], Genetics of Glucose regulation in Gestation and Growth (Gen3G)[18], Genetics, Early Life Environmental Exposures and Infant Development in Andalusia (GENEIDA), Environment and Childhood Project (INMA)[19], New Hampshire Birth Cohort Study (NHBCS)[20] and Rhode Island Child Health Study (RICHS)[21]. For this meta-analysis, 344 (20.2%) mothers reported any MSDP, defined as any cigarette smoking during any trimester of pregnancy. Any MSDP tended to be less prevalent in the cohorts from Canada and the USA compared to those from Australia and Europe (Table 1). Three cohorts (N=795, EDEN, GENEIDA and INMA) contributed to the EWAS of sustained MSDP, defined as maternal smoking throughout pregnancy, among which 163 (20.5%) mothers reported sustained MSDP. Distributions of covariates by cohort are provided in the supplementary materials (Excel Table S1).

Genome-wide DNAm meta-analyses

We produced four statistical models for each CpG site, modeling the associations between DNAm with both any and sustained MSDP, with and without adjustment for putative cellular heterogeneity which was estimated with a reference free deconvolution algorithm[22]. Models were adjusted for maternal age, parity, and maternal education. Genomic inflation factors from the meta-analyses ranged from λ=2.0 to 2.8 (Excel Table S2, Supplemental Figure S1). Any MSDP was associated with DNAm at 532 CpG sites after Bonferroni adjustment (Excel Table S3), while sustained MSDP

was associated with DNAm at 568 CpG sites (Excel Table S4). After adjusting for putative cellular heterogeneity, any MSDP was associated with 878 CpGs (Excel Table S5), while sustained MSDP was associated with 894 CpGs (Excel Table S6). Among these, 10.7% of models for any MSDP and 3.9% of models for sustained MSDP, not adjusted for cell type, yielded heterogeneity p-values < 0.01, while only 5.5% of the models for any MSDP and 3.5% for sustained MSDP, adjusted for putative cellular heterogeneity, produced heterogeneity p-values < 0.01. Thus, heterogeneity in the associations across cohorts was lowest in the analyses that were adjusted for putative cellular heterogeneity and the CpGs from these models that yielded Bonferroni-significant associations (1224 unique CpGs) were carried forward for secondary analyses.

Overall, the Bonferroni significant CpGs were distributed throughout the genome, and the majority (68%) of CpGs exhibited lower DNAm in association with any and sustained MSDP (Figure 1). Of the 548 CpGs that yielded Bonferroni significant associations with both any and sustained MSDP, the absolute values of the parameter estimates were greater for the models of sustained MSDP compared to the estimates for any MSDP (Excel Table S7). In a secondary analysis that restricted the any MSDP models to the same three cohorts that contributed to the sustained MSDP models (EDEN, GENEIDA, and INMA), we again found that almost all CpGs (544 out of 548) exhibited increased magnitudes of association with increased duration of exposure (Excel Table S7, Supplemental Figure S2).

The most notable association was observed at cg27402634, located upstream of the *LEKR1* gene and the non-coding RNA *LINC00886*, which showed the largest differential DNAm and smallest p-values in all meta-analyses. Placentas that were exposed to any MSDP had 22.95% lower DNAm (95% CI: 21.01-24.99% lower

DNAm; p-value = 5.35E-119) and those exposed to sustained MSDP had 25.78% lower DNAm (95% CI: 24.02-27.54% lower DNAm; p-value = 7.22E-181) when compared to mothers that did not smoke at all during pregnancy. Though all cohorts observed substantial hypomethylation with MSDP at this CpG, the actual estimates of the associations were highly variable between cohorts for models of any MSDP (heterogeneity p-value = 2.66E-15), but relatively consistent for models of sustained MSDP (heterogeneity p-value = 1.55E-01) (Figure 2A). Overall, we consistency in the associations across cohorts for the vast majority of the 1224 CpGs that yielded Bonferroni-significant associations via meta-analysis: 92% and 97% of these CpGs yielded heterogeneity p-values > 0.01, for their associations with any and sustained MSDP respectively. In addition to cg27402634, we highlight those relationships that yielded that largest magnitudes of association: ($|\beta_{\text{Any MSDP}}|$) > 0.05 for cg26843110 (*EDC3*), cg20340720 (*WBP1L*), and cg17823829 (*KDM5B*) (Figure 2B-2D). We identified numerous other noteworthy relationships but due to the large number of Bonferroni-significant associations from our meta-analyses, we highlight the relationship among the 20 most statistically significant CpGs from the primary meta-analysis of any MSDP going forward (Table 2).

Expression Quantitative trait methylation (eQTM) analyses

We tested whether the DNAm levels at MSDP-associated CpGs were associated with the expression of nearby (± 250kb of the CpG) mRNA from 194 placental samples in the RICHS cohort, and corrected for multiple testing with the less conservative false discovery rate (FDR) due to the smaller sample size. We identified 170 associations between DNAm and gene expression within a 5% FDR (Excel Table S8), including 141 unique CpGs and 140 unique mRNAs; 72.4% of the eQTMs exhibited inverse associations, while inverse associations were most common and statistical significance

was strongest for CpGs that were closer to the transcription start site (Supplemental Figure S3). Among the top 20 CpGs from our meta-analysis, eight were identified as eQTMs (5% FDR) (Table 3). Higher placental DNAm at cg27402634, was associated with lower expression (β = -2.32, and FDR < 5%) of *LEKR1*. Additionally, DNAm at cg17823829 (annotated to *KDM5B*) was most strongly associated lower expression of *PPFIA4* (FDR < 5%), and DNAm at cg26843110 (annotated to *EDC3*) was nominally associated with higher expression of *CSK*, while cg20340720 (annotated to *WBP1L*) was not associated with any nearby mRNA expression.

Functional and regulatory enrichment analyses

Enrichment analyses were performed to gain insights into which biological processes may be impacted by these MSDP-associated CpGs. We performed gene-set enrichment analyses, which included 565 genes annotated to CpGs associated with any MSDP and 581 genes for sustained MSDP. 143 and 25 pathways were significantly (q-value < 0.05) enriched among CpGs associated with any MSDP (Excel Table S9) and sustained MSDP (Excel Table S10), respectively. In both gene-sets, "signaling by nerve growth factor (NGF)" was the most significant pathway (q-value = 2E-04 for any MSDP and q-value = 9.6E-03 for sustained MSDP). Other significantly enriched pathways were related to growth factors (VEGF, EGF, PDGF, IGF1R), hormones (aldosterone, insulin, TSH, GnRH), interleukins (IL2, IL4, IL7), myometrial contraction, vascular smooth muscle contraction, thrombin and platelet activation, signaling and aggregation. We also tested whether the genes annotated to MSDP-associated CpGs were enriched for regulatory regions of specific transcription factors (TFs). Most notably, our MSDP-associated CpGs were enriched for genes regulated by GATA1 and GATA2, as well as by RUNX1, TP63, SMAD4, AR, TP53 or PPARG (Excel Tables S11-S12).

We then examined whether the MSDP-associated CpG sites were enriched for CpG island locations, allele-specific germline differentially methylated regions (gDMR)[23], regulatory features from the placenta specific 15-chromatin state annotation from ROADMAP[24], or placenta specific partially methylated domains (PMD)[25], which contain placenta-specific repressed genes (annotated to the results files in Supplementary Tables S3-S6). The MSDP-associated CpGs were depleted in PMDs (Supplemental Figure S4), highly enriched in placenta enhancers and depleted in transcription start sites and inactive states (Supplemental Figure S5). Additionally, MSDP-associated CpGs were depleted in CpG islands and shores, while enriched in CpG island shelves and open sea positions (Supplemental Figure S6). We identified 3 CpGs (cg05211790 and cg16360861 at *RAI14*, and cg05575921 at *AHRR* gene) that were within two candidate maternal gDMRs, but our overall set of MSDP-associated CpGs were neither depleted nor enriched for confirmed gDMRs (Supplemental Figure S7).

Phenotype enrichment analyses

The genes annotated to our MSDP-associated CpGs were tested for phenotype enrichment using data from the database of Genotypes and Phenotypes (dbGAP). Our MSDP-associated genes were enriched for numerous phenotypes in dbGAP, including several adiposity phenotypes (body mass index (BMI), waist-hip ratio, and obesity), blood pressure, cardiovascular diseases, type 2 diabetes, asthma and respiratory function (Excel Tables S13-S14).

Proximity to genetic variants for birth outcomes

We aimed to understand whether the MSDP-associated CpGs that we identified co-localize within the same genomic regions as genetic variants that have been associated with birth outcomes via genome-wide association studies (GWAS). Thus we

292 investigated whether MSDP-associated CpGs were within ± 0.5 Mb (1 Mb window) of
293 single nucleotide polymorphisms (SNPs) that have been associated with birth weight
294 (BW), birth length (BL), head circumference (HC) and gestational age (GA)[26–31], which
295 have been added as annotations to the meta-analysis results files (Excel Tables S3-S6).
296 Of the 324 birth outcome SNPs in autosomal chromosomes, 94 SNPs (83 loci) and 108
297 (97 loci) were within 0.5 Mb of CpGs that were associated with any or sustained
298 MSDP, respectively (Excel Tables S15). Overall ~16% of the 1224 MSDP-associated
299 CpG sites were within 0.5 Mb of birth outcome SNPs, including cg27402634 (*LEKR1*),
300 cg26843110 (*EDC3*) and cg20340720 (*WBP1L*). We also explored whether our MSDP-
301 associated CpGs may be biased by methylation quantitative trait loci (mQTLs), in
302 which SNPs influence the methylation levels at nearby CpGs. Two studies have
303 examined this question in human placenta, identifying 866[32] and 4,342[33] placental
304 mQTLs. Our findings did not appear to be biased by genetic variation as only 9 of the
305 1,224 MSDP-associated CpGs are previously characterized placental mQTLs.

306 ***Association of DNAM at smoking associated loci and smoking related birth outcomes***

307 We also performed a second meta-analysis to examine the relationships between
308 DNAm with gestational age at birth, preterm birth, BW, BL, and HC z-scores, outcomes
309 that are known to be related to maternal smoking. Of the 1224 CpGs tested, 341
310 (27.9%) were related to at least one birth outcome after Bonferroni-adjustment
311 (0.05/1224). The majority of birth outcome associations were related to gestational age
312 at birth (298 CpGs) (Excel Table S16). Preterm delivery, for which only two cohorts
313 could contribute data (EDEN and NHBCS), produced similar associations found for
314 GA, though fewer CpGs were statistically significant (Excel Table S17). We also found
315 that numerous loci were associated birth size z-scores, with the majority of these being
316 associated with BW (43 CpGs) (Excel Table S18), followed by BL (20 CpGs) (Excel

Table S19) and HC (4 CpGs) (Excel Table S20). Some of the CpGs associated with GA were also associated with birth size measurements, even though they were standardized for GA, suggesting independent associations with both gestational duration and fetal growth (Supplemental Figure S8A), including 6 CpGs (annotated to *SYNJ2*, *PXN*, *PTPRE*, *IGF2BP2*, 4q21.1) shared between GA and BW, and 2 (annotated to *POLR3E* and *LOC441869*) shared between GA and BL. Among the CpGs that yielded at least one Bonferroni-significant association with birth outcomes, CpGs that tended to have positive associations with birth outcomes clustered together and were typically hypomethylated with MSDP, while CpGs that exhibited inverse associations with birth outcomes tended to be hypermethylated with exposure to MSDP (Supplemental Figure S8B).

Among our top 20 CpGs that were associated with any MSDP, 5 were associated with GA at birth and 4 were associated with BW z-scores after Bonferroni adjustment (Table 3). DNAm at cg27402634 (*LEKR1*) and cg20340720 (*WBP1L),* both located close to BW-SNPs and for which MSDP associated with lower DNAm, were associated with larger BW (p-value = 6.71E-07, and p-value = 2.42E-07). On the other hand, DNAm at cg26843110 (*EDC3;* hypomethylated in response to MDSP and also close to BW-SNPs) and at cg17823829 (*KDM5B;* hypermethylated) were associated with longer and shorter gestational ages at birth, respectively (p-value = 5.09E-12, and p-value = 9.12E-06). Forest plots of BW z-scores and GA for these four CpGs are shown in Figure 3.

We summarize the results for all secondary analyses with a circos plot, for those 548 CpGs that yielded Bonferroni significant associations with both any and sustained MSDP (Figure 4). Among these, 21 CpGs were associated both with mRNA expression

(FDR < 5%) and at least one birth outcome (Bonferroni-significant), which have been annotated with the gene symbols from their respective eQTM models.

Comparison with smoking-sensitive CpGs in cord blood

We assessed whether the DNAm signatures of MSDP in the placenta were consistent with MSDP associations in cord blood previously reported by the PACE consortium[10]. Only nine CpGs annotated to seven unique genes (*AHRR*, *CYP1A1*, *GNG12*, *PXN*, *RNF122*, *SLC23A2*, and *ZBTB4*) yielded Bonferroni-significant associations in both tissues, out of 1224 CpGs from our study and 568 CpGs from the cord blood study (Table 4). Of note, the CpGs within *CYP1A1* and *RNF122* showed opposite directions of association with MSDP in cord blood and placenta. We also compared the parameter estimates from our study that yielded associations with MSDP within 5% FDR to the parameter estimates of those CpGs described in cord blood also within a 5% FDR[10]. There was no overall correlation ($r^2<0.1$) of the regression coefficients across these two tissues (Supplemental Figure S9).

Discussion

We identified 1224 CpG sites with placental methylation levels that were associated with any or sustained MSDP. Differential DNAm was greater with increased duration of exposure at all loci that were associated with both any and sustained MSDP, and a large proportion of the MSDP-associated CpGs were related to birth outcomes. Those CpGs that were observed to have higher DNAm associated with MSDP, tended to be inversely associated with gestational age and birth size, while CpGs exhibiting lower DNAm with MSDP tended to be positively associated with gestational age and birth size.

The MSDP-associated loci with the most statistically significant association (cg27402634), also identified in prior smoking EWAS of placental tissues[15], yielded dramatically lower DNAm levels in association with MSDP exposure. This effect-size is much larger in magnitude (~25% difference for sustained MSDP) compared to what has generally been observed in most exposure-focused EWAS, though within the same range as a CpG site in *AHRR* (cg05575921; 18% difference between exposed and unexposed) from a prior EWAS of current smoking and blood DNAm[34]. Additionally, decreased placental DNAm at cg27402634 correlates with increased expression of *LEKR1*, and associates with smaller BW and BL. Thus MSDP-associated hypomethylation at this CpG would be consistent with the well-known effect of maternal smoking resulting in shorter gestation and smaller birth size.

The functional activities of cg27402634, or corresponding *LEKR1* gene, in human placental tissues are not known. However, GWAS findings provide evidence that genetic variants within this region might be involved in fetal growth and possibly metabolic programming. For instance, the SNP rs1482852 or its proxies (rs900400; rs13322435) have been associated with different parameters of fetal growth[35], adiposity

in newborns[36,37], maternal adiponectin levels, cord blood leptin[37], and insulin release after an oral glucose challenge[38]. These findings from genetic studies in combination with our current study, implicate that this locus on chromosome 3 (3q25.31) contains very active determinants of growth regulation and metabolic activity, and that placental DNAm at cg27402634 is highly responsive to maternal smoking. Future mechanistic work is necessary to investigate whether the placental epigenetic regulation at this locus specifically influences placental functions and/or overall growth and metabolic functions of the developing fetus.

We identified numerous other notable MSDP-associated loci in addition to those of cg27402634 (*LEKR1*), and highlight those CpGs yielding the strongest magnitudes of effect (cg20340720, cg26843110, and cg17823829). MSDP was associated with lower DNAm at cg20340720, located within *WBP1L* (also annotated as *C10orf26*), while lower DNAm at this CpG correlated with lower with BW and BL z-scores. Genetic variants nearby to this CpG have been related to BW[31] and blood pressure[39]. We also observed lower DNAm with MSDP at cg26843110, which is within the body of the *EDC3* gene, and is nearby to SNPs associated with BW (rs3784789[31]). Lower DNAm at cg26843110 associated with shorter GA at birth, and decreased expression of *CSK*, which is involved in trophoblast differentiation[40] as well as blood pressure and aldosterone regulation[41]. Finally, cg17823829 (annotated to *KDM5B*) was hypermethylated with MSDP. Higher DNAm at this CpG correlated with shorter GA at birth and with lower expression of *PPFIA4* gene, which can be induced in response to hypoxia[42].

Our enrichment analyses identified numerous pathways that are critical to placental growth and development, such as vascularization, hormone signaling and inflammatory cytokines. Multiple pathways involving vascular endothelial growth

406 factors (VEGF) and nerve growth factors (NGF) populated the top of our enrichment
407 lists. The VEGFs and their receptors are required for all steps of placental
408 vascularization[43], while nerve growth factor (NGF) modulates immune activity,
409 inflammation and angiogenesis in the placenta[44]. Thus, our findings may be related to
410 perturbed placental vascularization or angiogenic signaling. Altered placental
411 vasculature is the most common placental pathology identified in pregnancy
412 complications[43], and MSDP can result in placental vascular remodeling[5]. Our findings
413 may represent, in part, an epigenetic footprint of the placental vascular and angiogenic
414 response to MSDP. Our CpGs were also enriched for genes regulated by specific
415 transcription factors, most notably GATA1 and GATA2. Together with PPARG and
416 TP63, GATA factors are part of the core transcriptional regulatory circuit that guides
417 and maintains proper trophoblast differentiation[45,46]. Placentas lacking *PPARG* have
418 lethal defects in placental vascularisation[43], and angiogenic activity is reduced in
419 placentas lacking *GATA2*[47]. Additionally, our MSDP-associated CpGs were enriched in
420 placental enhancers, while depleted in transcription start sites, inactive states, PMDs[25]
421 and gDMRs[23], overall suggesting that the CpG sites that we identified are located in
422 active regulatory regions.

423 Genes that are annotated to the CpGs that we studied have been linked to human
424 health and disease traits via dbGAP, including a number of conditions that are part of
425 the metabolic syndrome (BMI, obesity, cholesterol, blood pressure), for which prior
426 links to MSDP have been reported[48,49]. This may indicate that the placental genes whose
427 regulation is impacted by MSDP, are involved in energy uptake and expenditure, lipid
428 and glucose metabolism, blood pressure regulation, and inflammation, which are some
429 of the key physiological processes that are disrupted in the pathogenesis of metabolic
430 syndrome[50]. The MSDP-associated CpGs were also enriched for genes linked to asthma

and impaired respiratory function, which are known to be caused by MSDP[51], but it is unclear if altered regulation of these genes in the placenta has consequences on the development of the respiratory system. Furthermore, 128 of the 324 SNPs that have previously been associated with birth size or gestational age at birth[26–31] were in similar genomic proximity (within 0.5 Mb) to our CpGs, suggesting the MSDP-associated differential methylation in the placenta occurs in regions of the genome that are heavily involved in growth and development. Many of these SNPs have been shown to be related to birth outcomes, glycemic traits, blood pressure, and height, while the largest GWAS of BW to date concludes that the link between lower BW and later cardio-metabolic traits are largely driven by shared genetic effects[31]. Future research is needed to characterize this convergence of genetic effects on growth and placental epigenetic responsiveness to MSDP within similar genomic proximities. It is possible that MSDP, genetic variation, and placental DNAm in these regions yield additive or interactive effects on birth outcomes; prior studies have addressed this question in blood[52,53]. Additionally, while our MSDP-associated CpGs are in similar genomic regions (1 Mb windows) with birth outcome SNPs, very few of our CpGs (9 of 1224) and very few of the birth outcome SNPs (3 of 324) are known placental mQTLs[32,33].

We compared our findings to those of a previous PACE meta-analysis of MSDP and cord blood DNAm[10]. Two of the CpGs that we identified within the *AHRR* gene (cg05575921 an eQTM for *AHRR* in placenta, and cg21161138), have been consistently observed to be sensors of MSDP exposure in cord blood. Only four other CpG sites, (annotated to *GNG12*, *PXN*, *ZBTB4*, and *SLC23A2*) were differentially methylated in both placenta and cord blood with the same direction of association in both tissues. Additionally, three CpG sites within *CYP1A1* and *RNF122* were identified in both meta-analyses but with different directions of association in cord blood versus placenta.

Interestingly, we observed *CYP1A1* to be hypomethylated in placenta with exposure to MSDP, which is consistent with studies of adipose, skin, and lung tissues[54], but this CpG was hypermethylated in the cord blood meta-analysis[10]. Additionally, the most statistically significant association with MSDP in placenta (cg27402634, *LEKR1*) was not associated with MSDP in the cord blood meta-analysis. These observations, and the lack of overall correlation in regression coefficients when comparing placental and cord blood responses to MSDP, suggest that there are unique tissue-specific molecular responses to this exposure.

The above findings should be interpreted within the context this study's limitations. MSDP was self-reported and subject to misclassification, though differential misclassification likely would have biased our findings towards the null. We modeled two different definitions for MSDP and found that the models with sustained MSDP produced larger magnitudes of association, as has been previously found in studies of blood DNAm[55]. Although these findings suggest that increased duration of MSDP is associated with greater differential DNAm, we did not assess dose-response patterns (ie. number of cigarettes or cotinine concentrations), which should be the focus of future investigations. Our study predominantly consisted of samples from mother-infant pairs of European ancestry, and thus additional studies involving diverse racial and ethnic backgrounds are needed in order to improve the generalizability of these findings. While it is unlikely that placental DNAm would influence maternal smoking, the observed associations between DNAm levels and reproductive outcomes could be due to reverse causation. Placenta is a heterogeneous tissue with multiple different cell types[56] that serve different functions and thus have different epigenetic states[57]. To correct for this, we estimated and adjusted for variability in placental DNAm that may be due to tissue heterogeneity. We utilized a data driven approach that was not based on

a methylome reference, as no references for placental cell-type methylomes are currently available. Adjustments for these estimates of putative cellular admixtures did reduce heterogeneity in the meta-analyses, thus improving the consistency of observed associations between MSDP and DNAm across these independent cohorts. However, it is possible that residual confounding may have influenced some of our results.

Despite these limitations, our study had numerous strengths, including a large overall sample-size and seven independent studies to identify these relationships. We used harmonized definitions of exposure variables and covariates, standardized protocols for quality control and pre-processing of DNAm data, and standardized methods for estimating/adjusting for tissue heterogeneity and for statistical analyses. We also performed secondary analyses involving mRNA expression, functional and phenotype enrichment, overlap with GWAS hits for reproductive outcomes, and meta-analyses of DNAm variation with birth outcomes to provide biological and health-related interpretations of our findings.

We identified a DNAm signature of MSDP in the placenta that shows substantial differences from that observed in cord blood, most notably the CpG in close proximity to *LEKR1*. Many of the identified MSDP-associated loci are involved in biological process that are known to play critical roles in placental development, including vascularization, angiogenesis, and inflammation. Additionally, a large proportion of the MSDP-associated CpGs were also associated with GA at birth, or birth size z-scores, suggesting these placental epigenetic variations may be intermediate molecular markers between MSDP and these outcomes. Further study is required to determine whether these epigenetic variations are causal mediators within these relationships or reflecting other processes.

506

Material and methods

Participating cohorts

Cohorts that are members of the PACE consortium were identified for participation in the current study if they had existing DNAm data quantified from placental tissue via the Illumina Infinium HumanMethylation450 BeadChip and if they had obtained information on self-reported smoking during pregnancy. The seven cohorts that contributed to the meta-analysis of any MSDP included AQUA[16], EDEN[17], Gen3G[18], GENEIDA, INMA[19], NHBCS[20] and RICHS[21]. EDEN, GENEIDA and INMA also contributed to the sustained MSDP stratified analyses. RICHS contributed RNAseq data for analyses with mRNA expression. All cohorts acquired ethics approval and informed consent from participants prior to data collection through local ethics committees. Exclusion criteria for this study were: non-singleton births, pre-eclampsia and DNAm not assessed in the fetal side of the placenta. All participants in the study were of European ancestry, except 1.85% of EDEN mothers. Detailed methods for each cohort are provided in the Supplementary Material (Supplemental Methods File).

Tobacco smoking definitions

Any MSDP was defined as "yes" if mothers reported smoking cigarettes at any time during pregnancy. Sustained MSDP was defined as "yes" when mothers reported smoking cigarettes at least in the 1st and 3rd trimester of pregnancy. For both exposure variables, the comparison group was defined as the mothers that reported no smoking during any of the pregnancy.

Placental genome-wide DNAm data acquisition, quality control and normalization

Placental DNAm from the fetal side was assessed with the Infinium Human-Methylation450 array (Illumina, San Diego, CA USA). See Supplementary Methods file for extra details on placenta collection, DNA extraction and DNAm acquisition in each

cohort. Quality control of DNAm was standardized across all cohorts. Low quality samples were filtered out and probes with detection p-values > 0.01 were excluded. Beta-values were normalized via functional normalization[58] and beta-mixture quantile normalization (BMIQ)[59] was applied to correct for the probe type bias. Cohorts applied ComBat to remove batch effects when applicable. Probes that hybridize to the X/Y chromosomes, cross-hybridizing probes and probes with SNPs at the CpG site, extension site, or within 10 bp of the extension site with an average minor allele frequency > 0.01 were filtered out[60]. Overall, 418,639 probes and 415,396 were available for modelling any MSDP and sustained MSDP, respectively. Finally, DNAm extreme outliers (<25th percentile – 3*IQR or >75th percentile + 3*IQR across all the samples) were trimmed.

Estimates of putative cellular heterogeneity

Placental putative cellular heterogeneity was estimated from DNAm data using a reference-free cell-mixture decomposition method[61]. The number of surrogate variables ranged from 2 to 5 depending on the cohort. Models for differential DNAm were corrected for the number of surrogate variables minus one to reduce multi-collinearity.

Genome-wide differential DNAM analyses

Within each cohort, robust linear regression from the "MASS" package[62] in R were used to account for potential heteroskedasticity while testing the associations between normalized DNAm beta values at each CpG with any MSDP and sustained MSDP. Models were adjusted for maternal age, parity, maternal education and cohort-specific variables first unadjusted for putative cellular heterogeneity then adjusted for cellular heterogeneity. We performed inverse variance-weighted fixed-effects meta-analyses using METAL[63]. The meta-analysis was performed independently by two groups to ensure consistent results. CpGs not retained in at least 2 cohorts were filtered

557 out. We used the Bonferroni adjustment to control for multiple testing. To examine
558 whether increased duration of exposure (sustained smoking versus any smoking)
559 yielded increased magnitudes of association, we calculated the percent change in the
560 coefficients between the two models $(|\beta_{Sustained}| - |\beta_{Any}|)/|\beta_{Any}| * 100$. Secondary analyses
561 were only performed on CpGs that yielded Bonferroni significant associations with any
562 or sustained MSDP in models that were adjusted for putative cellular heterogeneity.

563 ***Expression quantitative trait methylation (eQTM) loci***

564 We performed expression quantitative trait methylation (eQTM)[64] analyses in
565 the RICHS cohort. Transcription was measured via RNA-seq on 194 placentas. The
566 details of sample collection, assay, and QC for the RNA-seq data are presented in detail
567 elsewhere[65], and summarized in the Supplementary Material (Supplementary Methods
568 File). In this dataset, we identified 6523 unique transcripts annotated to an Ensembl ID
569 (GrCh37/hg19) and with a transcriptional start site (TSS) within 250 kb upstream or
570 downstream of 1184 out of the 1224 candidate CpGs. The association between DNAm
571 and expression levels was assessed via 10295 linear regression models using the MEAL
572 package[66] in R. We report the results for all models yielding nominally significant
573 associations (raw p-values <0.05), statistically significant eQTMs were determined at
574 5% FDR.

575 ***CpG site annotation***

576 We annotated CpGs to genes and CpG islands with notations from the Illumina
577 HumanMethylation 450K annotation file, and with several regulatory features using
578 publicly available data: placental 15-chromatin states[67] released from the ROADMAP
579 Epigenomics Mapping Consortium[24] (ChromHMM v1.10), placental partially
580 methylated domains (PMDs)[25] and placental germline differentially methylated regions
581 (gDRMs)[23].

582 ***Enrichment analyses***

583 Functional enrichment analyses were performed at the gene level via
584 ConsensusPathDB[68] using KEGG, Reactome, Wikipathways, Biocarta as reference
585 gene-sets. ConsensusPathDB performs a hypergeometric test and corrects multiple-
586 testing with FDR. Enrichment for transcription factors and for phenotypes were
587 assessed at the gene level with EnrichR using ENCODE and ChEA consensus TFs from
588 ChIP-X database, and dbGaP database, respectively. EnrichR results were ranked using
589 the combined score (P-value computed using Fisher exact test combined with the z-
590 score of the deviation from the expected rank).[69] Enrichment for regulatory features was
591 assessed with the hypergeometric test, and P-values were Bonferroni-corrected for 15
592 (placental chromatin 15-states) and 6 (relation to CpG island) tests, respectively.

593 ***Overlap of MSDP-sensitive CpG sites and birth outcome SNPs***

594 Co-localization between MSDP-associated CpGs in placenta with previously
595 identified BW, BL, HC and GA SNPs from the largest genome-wide association studies
596 (GWAS) to date[26–31] was assessed using the GenomicRanges package in R[70]. We
597 identified which CpGs were located within 1 Mb windows (± 0.5 Mb) surrounding each
598 of the 324 autosomal SNPs, which correspond to 280 potential unique loci. Unique loci
599 were defined based on the criteria in Warrington et al. 2019[31], and linkage
600 disequilibrium in Europeans ($r^2 > 0.1$ in < 2Mb).

601 ***Association between DNAm and birth outcomes***

602 Within each cohort, robust linear regression models were utilized to test the
603 association between normalized DNAm beta values at each CpG as the independent
604 variable and gestational age at birth (inverse normal transformation of sex residuals),
605 BW z-scores, BL z-scores, and HC z-scores as the dependent variables. Logistic
606 regression was used to examine the relationships between DNAm and pre-term birth

607 (defined as <37 weeks of gestation). Birth size z-scores were calculated using
608 international references from the INTERGROWTH-21st Project[71] and standardized by
609 both gestational age and newborn sex. Models were adjusted for maternal age, parity,
610 maternal education, cohort-specific variables (see Supplemental Methods) and putative
611 cellular heterogeneity. Inverse variance-weighted fixed-effects meta-analyses[63] were
612 again used to estimate pooled associations. Multiple testing was controlled with the
613 Bonferroni adjustment ($\alpha = 0.05/1224$).

614 ***Comparison of MSDP-sensitive CpGs in placenta with cord blood***

615 We examined the consistency between MSDP-sensitive CpGs in placenta and in
616 cord blood[10]. First we compared the coefficients from the models for sustained MSDP
617 in cord blood, unadjusted for cellular heterogeneity, to results for both any and
618 sustained MSDP in placenta, adjusted for cellular heterogeneity, using Pearson
619 correlation coefficients.

620 All DNAm data processing and analyses were conducted in R, with the
621 exception of the meta-analyses which were performed with METAL.

622

623

References

1. Curtin, S. C. & Mathews, T. J. Smoking prevalence and cessation before and during pregnancy: data from the birth certificate, 2014. *Natl. Vital Stat. Reports* **65**, (2016).

2. Smedberg, J., Lupattelli, A., Mårdby, A.-C. & Nordeng, H. Characteristics of women who continue smoking during pregnancy□: a cross-sectional study of pregnant women and new mothers in 15 European countries. *BMC Pregnancy Childbirth* **14**, (2014).

3. United States Department of Health and Human Services. The Health Consequences of Smoking—50 Years of Progress A Report of the Surgeon General. *A Rep. Surg. Gen.* 1081 (2014). doi:NBK179276

4. Jauniaux, E. & Burton, G. J. Morphological and biological effects of maternal exposure to tobacco smoke on the feto-placental unit. *Early Hum. Dev.* **83**, 699–706 (2007).

5. Zdravkovic, T., Genbacev, O., Mcmaster, M. T. & Fisher, S. J. The Adverse Effects of Maternal Smoking on the Human Placenta: A Review. *Placenta* **26**, S81–S86 (2005).

6. Perez-Garcia, V. *et al.* Placentation defects are highly prevalent in embryonic lethal mouse mutants. *Nature* **555**, 463–468 (2018).

7. Felix, J. F. & Cecil, C. A. M. Population DNA methylation studies in the Developmental Origins of Health and Disease (DOHaD) framework. *J. Dev. Orig. Health Dis.* (2018). doi:10.1017/S2040174418000442

8. Richmond, R. C. & Joubert, B. R. Contrasting the effects of intra-uterine smoking and one-carbon micronutrient exposures on offspring DNA methylation. *Epigenomics* **9**, 351–367 (2017).

649 9. Felix, J. F. *et al.* Cohort profile: Pregnancy and childhood epigenetics (PACE)
650 consortium. *Int. J. Epidemiol.* **47**, 22-23u (2018).
651 10. Joubert, B. R. *et al.* DNA Methylation in Newborns and Maternal Smoking in
652 Pregnancy: Genome-wide Consortium Meta-analysis. *Am. J. Hum. Genet.* **98**,
653 680–696 (2016).
654 11. Wilhelm-Benartzi, C. S. *et al.* In Utero Exposures, Infant Growth, and DNA
655 Methylation of Repetitive Elements and Developmentally Related Genes in
656 Human Placenta. *Environ. Health Perspect.* **120**, 296–302 (2012).
657 12. Chhabra, D. *et al.* Fetal lung and placental methylation is associated with in utero
658 nicotine exposure. *Epigenetics* **9**, 1473–1484 (2014).
659 13. Maccani, J. Z. J., Koestler, D. C., Houseman, E. A., Marsit, C. J. & Kelsey, K. T.
660 Placental DNA methylation alterations associated with maternal tobacco smoking
661 at the RUNX3 gene are also associated with gestational age. *Epigenomics* **5**,
662 619–630 (2013).
663 14. Suter, M. *et al.* Maternal tobacco use modestly alters correlated epigenome-wide
664 placental DNA methylation and gene expression. *Epigenetics* **6**, 1284–1294
665 (2011).
666 15. Morales, E. *et al.* Genome-wide DNA methylation study in human placenta
667 identifies novel loci associated with maternal smoking during pregnancy. *Int. J.*
668 *Epidemiol.* **45**, 1644–1655 (2016).
669 16. Muggli, E. *et al.* Study protocol: Asking QUestions about Alcohol in pregnancy
670 (AQUA): A longitudinal cohort study of fetal effects of low to moderate alcohol
671 exposure. *BMC Pregnancy Childbirth* **14**, (2014).
672 17. Heude, B. *et al.* Cohort Profile: The EDEN mother-child cohort on the prenatal
673 and early postnatal determinants of child health and development. *Int. J.*

Epidemiol. **45**, 353–363 (2016).

18. Guillemette, L. *et al.* Genetics of Glucose regulation in Gestation and Growth (Gen3G): a prospective prebirth cohort of mother–child pairs in Sherbrooke, Canada. *BMJ Open* **6**, e010031 (2016).

19. Guxens, M. *et al.* Cohort Profile: The INMA—INfancia y Medio Ambiente—(Environment and Childhood) Project. *Int. J. Epidemiol.* **41**, 930–940 (2012).

20. Gilbert-diamond, D., Emond, J. A., Baker, E. R., Korrick, S. A. & Karagas, M. R. Relation between in Utero Arsenic Exposure and Birth Outcomes in a Cohort of Mothers and Their Newborns from New Hampshire. *Environ. Health Perspect.* **124**, 1299–1307 (2016).

21. Appleton, A. A. *et al.* Prenatal Programming of Infant Neurobehaviour in a Healthy Population. *Paediatr. Perinat. Epidemiol.* **30**, 367–375 (2016).

22. Houseman, E. A. *et al.* Reference-free deconvolution of DNA methylation data and mediation by cell composition effects. *BMC Bioinformatics* **17**, 259 (2016).

23. Hamada, H. *et al.* Allele-Specific Methylome and Transcriptome Analysis Reveals Widespread Imprinting in the Human Placenta. *Am. J. Hum. Genet.* **99**, 1045–1058 (2016).

24. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–329 (2015).

25. Schroeder, D. I. *et al.* The human placenta methylome. *Proc. Natl. Acad. Sci.* **110**, 6037–6042 (2013).

26. Horikoshi, M. *et al.* Genome-wide associations for birth weight and correlations with adult disease. *Nature* **538**, 248–252 (2016).

27. Beaumont, R. N. *et al.* Genome-wide association study of offspring birth weight in 86 577 women identifies five novel loci and highlights maternal genetic

effects that are independent of fetal genetics. *Hum. Mol. Genet.* **27**, 742–756 (2018).

28. van der Valk, R. J. P. *et al.* A novel common variant in DCST2 is associated with length in early life and height in adulthood. *Hum. Mol. Genet.* **24**, 1155–1168 (2015).

29. Taal, H. *et al.* Common variants at 12q15 and 12q24 are associated with infant head circumference. *Nat. Genet.* **44**, 532–538 (2012).

30. Zhang, G. *et al.* Genetic Associations with Gestational Duration and Spontaneous Preterm Birth. *N. Engl. J. Med.* **377**, 1156–1167 (2017).

31. Warrington, N. M. *et al.* Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat. Genet.* **51**, 804–814 (2019).

32. Do, C. *et al.* Mechanisms and Disease Associations of Haplotype-Dependent Allele-Specific DNA Methylation. *Am. J. Hum. Genet.* **98**, 934–955 (2016).

33. Delahaye, F. *et al.* Genetic variants influence on the placenta regulatory landscape. *PLOS Comput. Biol.* **14**, e1007785 (2018).

34. Joehanes, R. *et al.* Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet* **9**, 436–447 (2016).

35. Mook-Kanamori, D. O. *et al.* Variants near CCNL1/LEKR1 and in ADCY5 and fetal growth characteristics in different trimesters. *J. Clin. Endocrinol. Metab.* **96**, E810-5 (2011).

36. Urbanek, M. *et al.* The chromosome 3q25 genomic region is associated with measures of adiposity in newborns in a multi-ethnic genome-wide association study. *Hum. Mol. Genet.* **22**, 3583–96 (2013).

37. Hivert, M.-F. *et al.* Genetic determinants of adiponectin regulation revealed by pregnancy. *Obesity* **25**, 935–944 (2017).

724 38. Andersson, E. A. *et al.* The Birth Weight Lowering C-Allele of rs900400 Near
725 LEKR1 and CCNL1 Associates with Elevated Insulin Release following an Oral
726 Glucose Challenge. *PLoS One* **6**, e27096 (2011).
727 39. Li, C. *et al*. Genome-Wide Association Study Meta-Analysis of Long-Term
728 Average Blood Pressure in East Asians. *Circ. Cardiovasc. Genet.* **10**, e001527
729 (2017).
730 40. Daoud, G., Le bellego, F. & Lafond, J. PP2 regulates human trophoblast cells
731 differentiation by activating p38 and ERK1/2 and inhibiting FAK activation.
732 *Placenta* **29**, 862–70 (2008).
733 41. Kim, S.-M., Kang, J.-O., Lim, J. E., Hwang, S.-Y. & Oh, B. Csk Regulates Blood
734 Pressure by Controlling the Synthetic Pathways of Aldosterone. *Circ. J.* **82**, 168–
735 175 (2018).
736 42. Chan, M. C. *et al*. Tuning the transcriptional response to hypoxia by inhibiting
737 Hypoxia-inducible Factor (HIF) prolyl and asparaginyl hydroxylases. *J. Biol.*
738 *Chem.* **291**, 20661–20673 (2016).
739 43. Chen, D.-B. & Zheng, J. Regulation of Placental Angiogenesis. *Microcirculation*
740 **21**, 15–25 (2014).
741 44. Sahay, A. S., Sundrani, D. P. & Joshi, S. R. Neurotrophins: Role in Placental
742 Growth and Development. *Vitam. Horm.* **104**, 243–261 (2017).
743 45. Paul, S., Home, P., Bhattacharya, B. & Ray, S. GATA factors: Master regulators
744 of gene expression in trophoblast progenitors. *Placenta* **60**, S61–S66 (2017).
745 46. Bai, Q. *et al*. Dissecting the First Transcriptional Divergence During Human
746 Embryonic Development. *Stem Cell Rev.* **8**, 150–162 (2012).
747 47. Ma, G. T. *et al*. GATA-2 and GATA-3 regulate trophoblast-specific gene
748 expression in vivo. *Development* **124**, 907–14 (1997).

749 48. Dior, U. P. *et al.* Parental smoking during pregnancy and offspring cardio-
750 metabolic risk factors at ages 17 and 32. *Atherosclerosis* **235**, 430–7 (2014).
751 49. Behl, M. *et al.* Evaluation of the Association between Maternal Smoking,
752 Childhood Obesity, and Metabolic Disorders: A National Toxicology Program
753 Workshop Review. *Environ. Health Perspect.* **121**, 170–180 (2012).
754 50. Stančáková, A. & Laakso, M. Genetics of metabolic syndrome. *Rev. Endocr.*
755 *Metab. Disord.* **15**, 243–252 (2014).
756 51. Zacharasiewicz, A. Maternal smoking in pregnancy and its influence on
757 childhood asthma. *ERJ Open Res.* **2**, 00042–02016 (2016).
758 52. Dogan, M. V., Beach, S. R. H. & Philibert, R. A. Genetically contextual effects
759 of smoking on genome wide DNA methylation. *Am. J. Med. Genet. Part B*
760 *Neuropsychiatr. Genet.* (2017). doi:10.1002/ajmg.b.32565
761 53. Gao, X., Thomsen, H., Zhang, Y., Breitling, L. P. & Brenner, H. The impact of
762 methylation quantitative trait loci (mQTLs) on active smoking-related DNA
763 methylation changes. *Clin. Epigenetics* **9**, 1–13 (2017).
764 54. Tsai, P., Glastonbury, C. A., Eliot, M. N. & Bollepalli, S. Smoking induces
765 coordinated DNA methylation and gene expression changes in adipose tissue
766 with consequences for metabolic health. *bioRxiv Genomics* 1–21 (2018).
767 doi:10.1101/353581
768 55. Joubert, B. R. *et al.* Maternal smoking and DNA methylation in newborns: In
769 utero effect or epigenetic inheritance? *Cancer Epidemiol. Biomarkers Prev.* **23**,
770 1007–1017 (2014).
771 56. Wang, Y. & Zhao, S. Cell Types of the Placenta. (2010).
772 57. Fogarty, N. M. E., Burton, G. J. & Ferguson-Smith, A. C. Different epigenetic
773 states define syncytiotrophoblast and cytotrophoblast nuclei in the trophoblast of

774 the human placenta. *Placenta* **36**, 796–802 (2015).

775 58. Fortin, J.-P. *et al*. Functional normalization of 450k methylation array data
776 improves replication in large cancer studies. *Genome Biol.* **15**, 503 (2014).

777 59. Teschendorff, A. E. *et al.* A beta-mixture quantile normalization method for
778 correcting probe design bias in Illumina Infinium 450 k DNA methylation data.
779 *Bioinformatics* **29**, 189–196 (2013).

780 60. Chen, J. *et al*. CpGFilter: model-based CpG probe filtering with replicates for
781 epigenome-wide association studies. *Bioinformatics* **32**, 469–471 (2016).

782 61. Houseman, E. A., Molitor, J. & Marsit, C. J. Reference-free cell mixture
783 adjustments in analysis of DNA methylation data. *Bioinformatics* **30**, 1431–1439
784 (2014).

785 62. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S*. (Springer,
786 2002).

787 63. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis
788 of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

789 64. Teschendorff, A. E. & Relton, C. L. Statistical and integrative system-level
790 analysis of DNA methylation data. *Nat. Rev. Genet.* **Nov 13**, (2017).

791 65. Deyssenroth, M. A. *et al*. Whole-transcriptome analysis delineates the human
792 placenta gene network and its associations with fetal growth. *BMC Genomics* **18**,
793 (2017).

794 66. Ruiz-Arenas, C. & Gonzalez, J. MEAL: Perform methylation analysis. *R Packag.*
795 *version 1.10.1* (2018).

796 67. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and
797 characterization. *Nat. Methods* **9**, 215–6 (2012).

798 68. Kamburov, A. *et al*. ConsensusPathDB: Toward a more complete picture of cell

biology. *Nucleic Acids Res*. **39**, 712–717 (2011).

69. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, (2013).

70. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLOS Comput. Biol.* **9**, e1003118 (2013).

71. Villar, J. *et al.* International standards for newborn weight, length, and head circumference by gestational age and sex: The Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. *Lancet* **384**, 857–868 (2014)

Acknowledgements

We would like to thank all the families that participated in these studies for their generous contribution. Detailed acknowledgements and funding can be found in Supplementary Material.

Author contributions

TME, MVU, CJM, JL, MFH and MB conceived of and designed the study. Study-specific analyses were completed by JMC, (AQUA), ES (EDEN), AC (Gen3G), PCS and JMM (GENEIDA), MVU (INMA) and TME (NHBCS and RICHS). MV meta-analysed these results. TME, MV and MB performed the follow-up analyses. TME, MVU, JS, CJM, MFH, and MB interpreted the results. TME, MVU, and MB wrote the first draft of the manuscript. All authors (TME, MVU, ES, AC, ML, JMC, CL, ERB, NFJ, BH, PP, BGA, JH, MAD, MRK, CI, LB, PCS, YJL, KH, TB, MAC, JMM, EM, JC, MFF, JT, AGM, SJL, JS, CJM, JL, MFH, MB) read and critically revised drafts of the manuscript. Correspondence can be addressed to TME (Todd.M.Everson@Emory.edu) and MB (mariona.bustamante@isglobal.org).

824
825

Table 1: Distribution of any and sustained MSDP within participating cohorts.

Cohort	Country	Any MSDP			Sustained MSDP		
		Total N	N (%) Non-smokers	N (%) Smokers	Total N	N (%) Non-smokers	N (%) Smokers
AQUA	Australia	99	75 (75.76%)	24 (24.24%)	-	-	-
EDEN	France	647	446 (68.93%)	201 (31.07%)	570	446 (68.93%)	124 (21.75%)
Gen3G	Canada	151	138 (91.39%)	13 (8.61%)	-	-	-
GENEIDA	Spain	87	67 (77.01%)	20 (22.99%)	82	67 (77.01%)	15 (18.29%)
INMA	Spain	166	119 (71.69%)	47 (28.31%)	143	119 (71.69%)	24 (16.78%)
NHBCS	USA	310	290 (93.55%)	20 (6.45%)	-	-	-
RICHS	USA	240	221 (92.08%)	19 (7.92%)	-	-	-
TOTAL		1700	1356 (79.76%)	344 (20.24%)	795	632 (79.50%)	163 (20.50%)

MSDP = Maternal smoking during pregnancy.

Table 2. Meta-analysis results from models of any and sustained MSDP, for the 20 CpGs with the smallest p-values for the model of any MSDP, and the percent increase in effect size between sustained and any MSDP (% Change); all models adjusted for maternal age, parity, education and putative cellular heterogeneity; CpGs that were not annotated with a gene name in the Illumina 450K annotation file have been annotated with their genomic region (ie. 4q12).

CpG Annotations			Any MSDP			Sustained MSDP			% Change
CpG ID	Location	Annotated Gene (Region)	β_1	S.E.	P-value	β_1	S.E.	P-value	
cg17823829	chr1:202765754	*KDM5B* (Body)	0.06	0.006	1.79E-22	0.086	0.008	3.80E-27	42.98
cg26045080	chr1:36807363	*STK40* (Body)	0.024	0.003	6.66E-19	0.032	0.003	4.86E-21	31.12
cg00534380	chr2:101766586	*TBC1D8* (Body)	-0.044	0.004	2.53E-25	-0.061	0.006	1.03E-28	38.46
cg19246018	chr2:240031588	*HDAC4* (Body)	0.015	0.002	1.39E-20	0.016	0.002	6.52E-20	8.28
cg23752985	chr2:85803571	*VAMP8* (TSS1500)	-0.019	0.002	1.31E-19	-0.025	0.003	3.77E-16	32.98
cg00666842	chr2:88366145	*SMYD1* (TSS1500)	0.033	0.003	1.17E-22	0.049	0.004	2.23E-34	48.17
cg27402634	chr3:156536860	3q25.31	-0.23	0.01	5.40E-119	-0.258	0.009	7.20E-181	12.33
cg09491670	chr4:53529646	4q12	0.016	0.002	1.60E-19	0.023	0.002	3.96E-29	43.31
cg25585967	chr5:14452105	*TRIO* (Body)	0.04	0.005	7.53E-19	0.061	0.006	2.16E-23	51.74
cg12291408	chr7:100037572	7q22.1	-0.036	0.004	1.43E-18	-0.052	0.005	2.34E-22	45.15
cg14214914	chr9:131870304	*CRAT* (Body)	0.036	0.004	1.57E-20	0.05	0.005	1.20E-26	41.29
cg20340720	chr10:104512523	*WBP1L* (Body)	-0.054	0.004	5.38E-42	-0.076	0.005	2.19E-58	40.52
cg26648103	chr11:66791718	*SYT12* (5'UTR)	-0.033	0.004	8.98E-21	-0.043	0.005	4.30E-20	30.51
cg26115089	chr11:93846406	*HEPHL1* (3'UTR)	0.025	0.002	1.33E-28	0.032	0.003	2.00E-27	24.8
cg26843110	chr15:74935742	*EDC3* (Body)	-0.056	0.005	1.74E-29	-0.082	0.007	1.74E-36	47.84
cg26433445	chr16:81764289	16q23.3	0.024	0.003	5.85E-20	0.034	0.003	1.34E-23	40.25
cg24177452	chr17:27494295	*MYO18A* (5'UTR)	0.025	0.003	3.40E-18	0.031	0.004	2.84E-18	27.35
cg06716730	chr17:35851459	*DUSP14* (5'UTR)	-0.022	0.003	1.40E-18	-0.032	0.003	4.47E-22	46.79
cg03313447	chr19:41829042	*CCDC97* (3'UTR)	-0.02	0.002	1.26E-22	-0.029	0.002	1.93E-34	45.73
cg02341503	chr20:45947123	*ZMYND8* (Body)	-0.02	0.002	1.81E-18	-0.025	0.003	4.83E-18	25.74

β_1 = Coefficient of the association between DNAm and MSDP; MSDP = Maternal smoking during pregnancy; S.E. = Standard Error.

Table 3. Results from eQTM models, DNAm versus gestational age at birth, and DNAm versus BW Z-score models that yielded at least nominally significant associations, among the 20 CpGs that yielded the most statistically significant associations with any MSDP in the primary meta-analysis (sorted in the same order as Table 2).

CpG ID	Annotated Gene	eQTM				Gestational Duration			BW Z-score		
		eQTM Gene	β_1	S.E.	P-value	β_1	S.E.	P-value	β_1	S.E.	P-value
cg17823829	*KDM5B*	*PPFIA4*	-2.53	0.7	4.05E-04	-1.18	0.27	9.12E-06	-1.02	0.26	9.72E-05
cg26045080	*STK40*	*SH3D21*	-4.26	0.77	8.34E-08	-1.41	0.49	4.19E-03	-	-	-
cg00534380	*TBC1D8*	*TBC1D8*	-2.29	0.51	1.13E-05	3.53	0.39	9.26E-20	1.48	0.39	1.62E-04
cg19246018	*HDAC4*	-	-	-	-	-1.39	0.59	1.74E-02	-	-	-
cg23752985	*VAMP8*	*CAPG*	6.33	1.75	3.96E-04	5.47	0.69	1.40E-15	-	-	-
cg00666842	*SMYD1*	-	-	-	-	-	-	-	-	-	-
cg27402634	3q25.31	*LEKR1*	-2.32	0.67	7.14E-04	-	-	-	0.92	0.19	6.71E-07
cg09491670	4q12	*USP46*	-3.59	-	4.91E-04	-	-	-	-	-	-
cg25585967	*TRIO*	-	-	-	-	-0.97	0.37	8.64E-03	-1.35	0.36	2.06E-04
cg12291408	7q22.1	*ACTL6B*	-4.89	-	1.12E-02	-	-	-	1.31	0.39	6.82E-04
cg14214914	*CRAT*	*CRAT*	-4.76	0.59	6.03E-14	-0.96	0.39	1.35E-02	-	-	-
cg20340720	*WBP1L*	-	-	-	-	-	-	-	1.86	0.36	2.42E-07
cg26648103	*SYT12*	-	-	-	-	1.06	0.43	1.29E-02	1.66	0.42	7.56E-05
cg26115089	*HEPHL1*	-	-	-	-	-1.1	0.55	4.43E-02	-1.48	0.55	6.71E-03
cg26843110	*EDC3*	*CSK*	0.57	0.25	2.18E-02	2.3	0.33	5.09E-12	1.28	0.33	1.19E-04
cg26433445	16q23.3	-	-	-	-	-	-	-	-2.99	0.61	9.06E-07
cg24177452	*MYO18A*	*MYO18*	-1.37	0.43	1.79E-03	-	-	-	-	-	-
cg06716730	*DUSP14*	*DUSP14*	-3.07	0.96	1.59E-03	3.72	0.54	7.87E-12	-	-	-
cg03313447	*CCDC97*	*TGFB1*	-10.88	1.41	7.27E-13	-2.48	0.74	7.79E-04	2.29	0.73	1.58E-03
cg02341503	*ZMYND8*	*ZMYND8*	-1.72	0.64	7.37E-03	1.44	0.63	2.29E-02	3.19	0.62	2.70E-07

β_1 = Coefficient of the association between DNAm and the outcome variable (mRNA expression, Gestational duration, or BW Z-score); BW = Birth weight; S.E. = Standard Error.

Table 4: Differentially methylated CpGs with MSDP in placenta (from our meta-analysis) and cord blood (from a published PACE meta-analysis[10]) that yielded Bonferroni-significant associations in both tissues.

CpG Annotation		Cord blood* (Sustained MSDP)			Placenta (Sustained MSDP)			Placenta (Any MSDP)		
CpG ID	Gene	β_1	S.E.	P-value	β_1	S.E.	P-value	β_1	S.E.	P-value
cg25189904	*GNG12* (TSS1500)	-0.024	0.002	1.38E-32	-0.020	0.003	2.54E-13	-0.014	0.002	1.99E-10
cg05575921	*AHRR* (Body)	-0.064	0.002	1.64E-193	-0.030	0.005	5.01E-09	-0.015	0.004	1.83E-04
cg21161138	*AHRR* (Body)	-0.022	0.001	1.78E-54	-0.026	0.004	5.25E-11	-0.018	0.003	2.55E-10
cg08327744	*RNF122* (Body)	-0.010	0.002	1.37E-08	0.020	0.003	3.12E-09	0.016	0.003	2.21E-09
cg15893360	*PXN* (Body)	-0.011	0.002	1.66E-08	-0.024	0.004	3.61E-08	-0.015	0.003	3.43E-06
cg12101586	*CYP1A1* (TSS1500)	0.045	0.003	3.67E-50	-0.045	0.008	7.15E-08	-0.026	0.006	9.12E-06
cg23680900	*CYP1A1* (TSS200)	0.003	0.001	8.73E-08	-0.018	0.002	6.26E-21	-0.006	0.001	7.35E-06
cg07565956	*ZBTB4* (5'UTR)	-0.006	0.001	2.50E-10	-0.023	0.003	4.02E-18	-0.016	0.002	2.26E-16
cg16547579	*SLC23A2* (5'UTR)	-0.013	0.002	2.00E-10	-0.048	0.009	1.47E-08	-0.031	0.007	1.42E-06

β_1 = Coefficient of the association between DNAm and MSDP; MSDP = Maternal smoking during pregnancy; S.E. = Standard Error

* Main model in Joubert et al: sustained MSDP adjusted for cell type heterogeneity

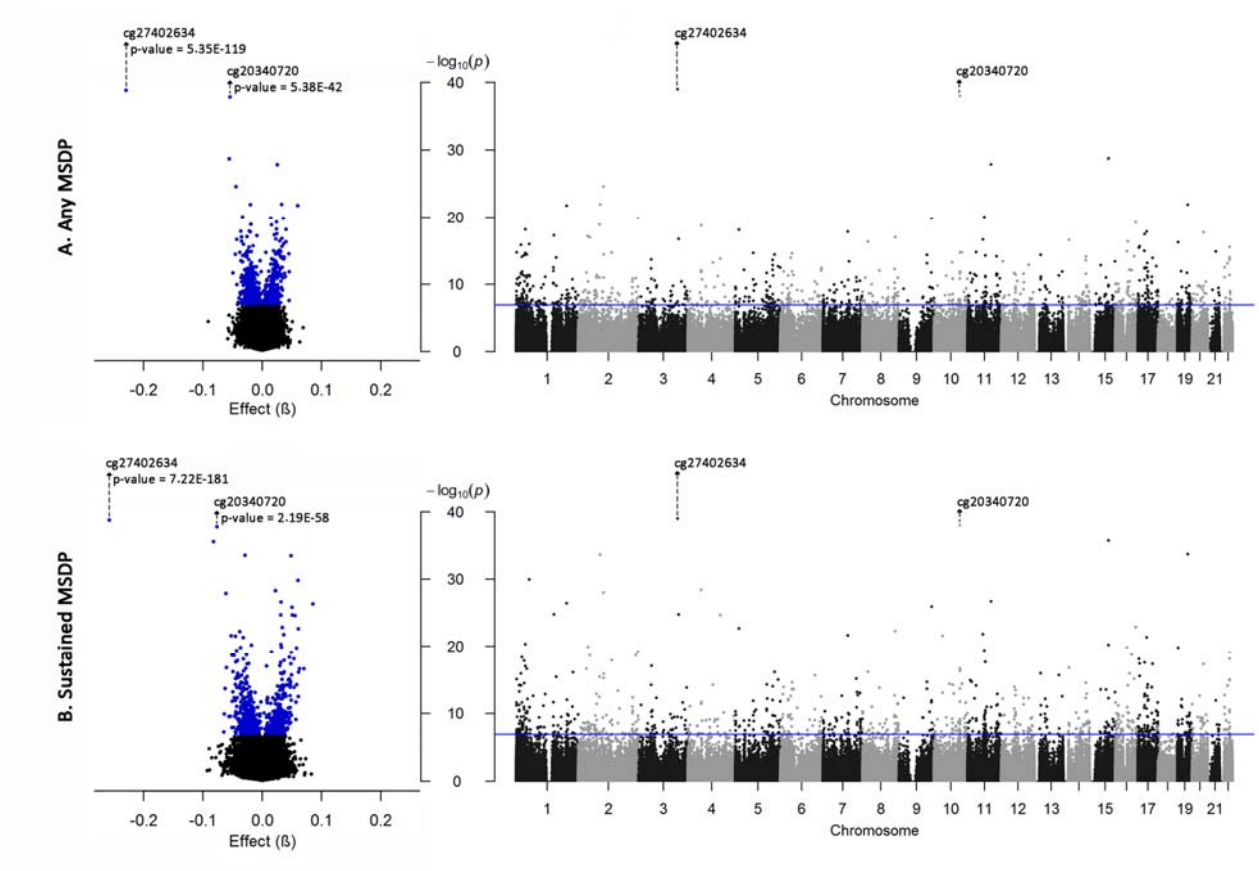

Figure 1. Volcano and Manhattan plots of the association between any MSDP (total N=1700, exposed = 344) (A) and sustained MSDP (total N=795, exposed = 163) (B) with placental DNAm adjusted for maternal age, parity, maternal education and putative cellular heterogeneity. In the volcano plots, the x-axis shows the difference in DNAm (effect) with a possible range between 0 and 1, while the x-axis in the Manhattan plot represents genomic location; both plots share the same y-axis with $-\log_{10}(P)$. Bonferroni thresholds for statistical significance are shown as blue dots and a blue horizontal line, for volcano and Manhattan plots, respectively. The y-axes were truncated to a minimum p-value of $1*10^{-40}$ (or maximum $-\log_{10}(P)$ of 40), to allow for better visualization of the majority of our results. The CpGs that were impacted by y-axis truncation are indicated with arrows and annotated with their actual p-values.

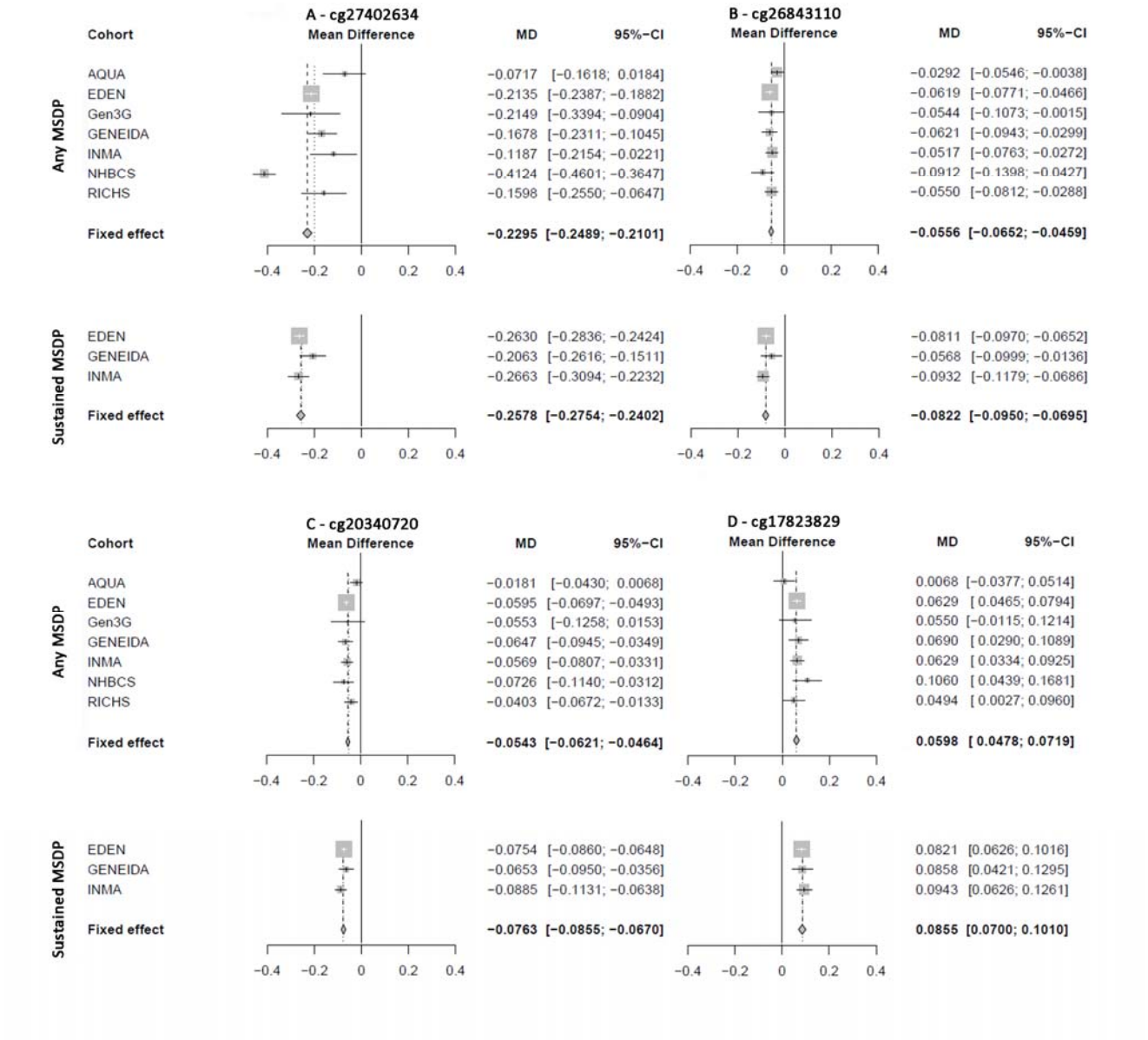

Figure 2. Forest plots of the cohort specific estimates of association and fixed-effect meta-analysis estimates of association between placental DNAm levels at (A) cg27402634, (B) cg26843110, (C) cg20340720, and (D) cg17823829 with any MSDP and sustained MSDP; models adjusted for maternal age, parity, maternal education, and putative cellular heterogeneity.

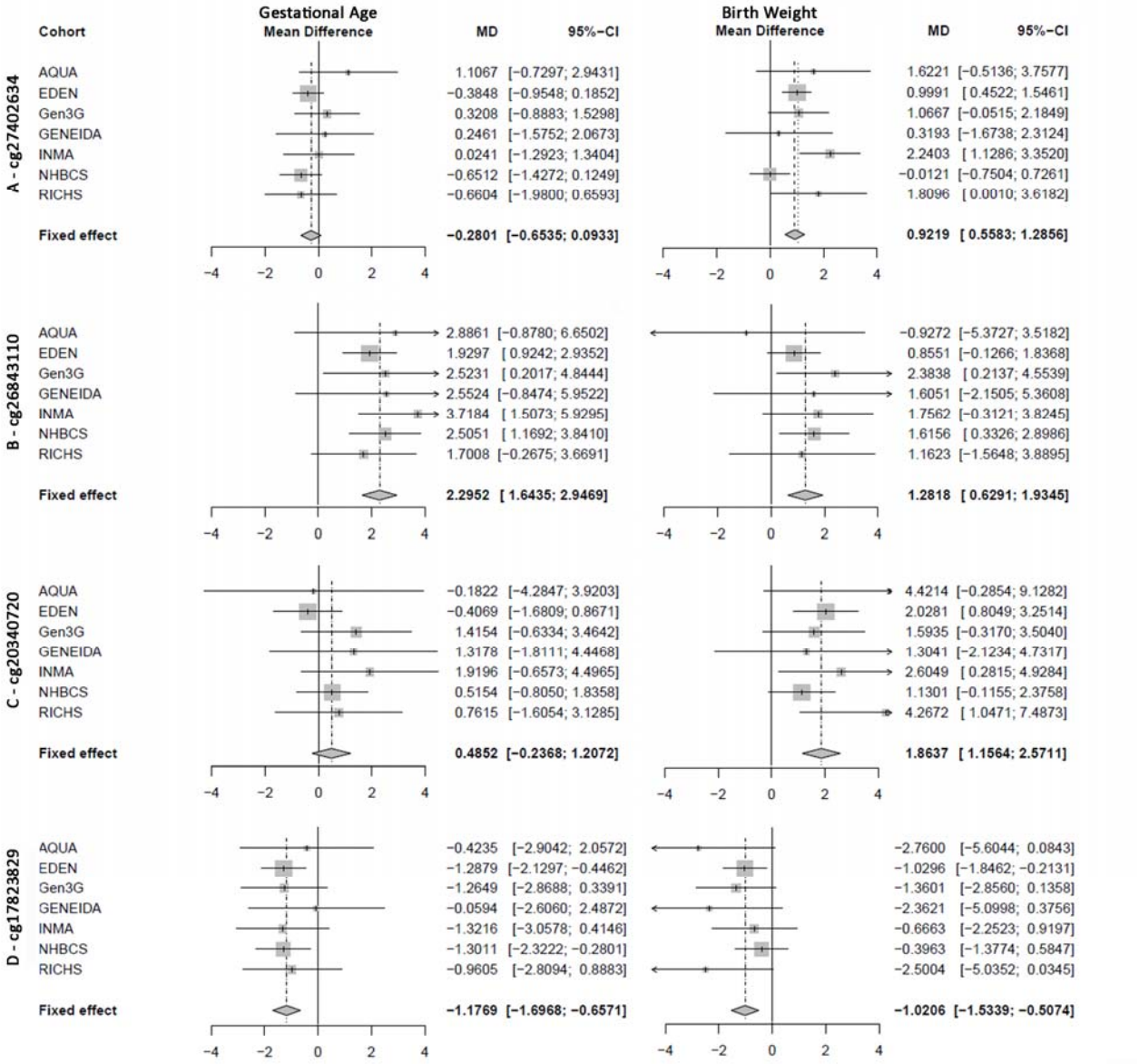

Figure 3. Forest plots of the cohort specific estimates of association and fixed-effect meta-analysis estimates of association between higher levels of placental DNAm at (A) cg27402634, (B) cg26843110, (C) cg20340720, and (D) cg17823829 with gestational age and birth weight z-scores; models adjusted for maternal age, parity, maternal education, and putative cellular heterogeneity.

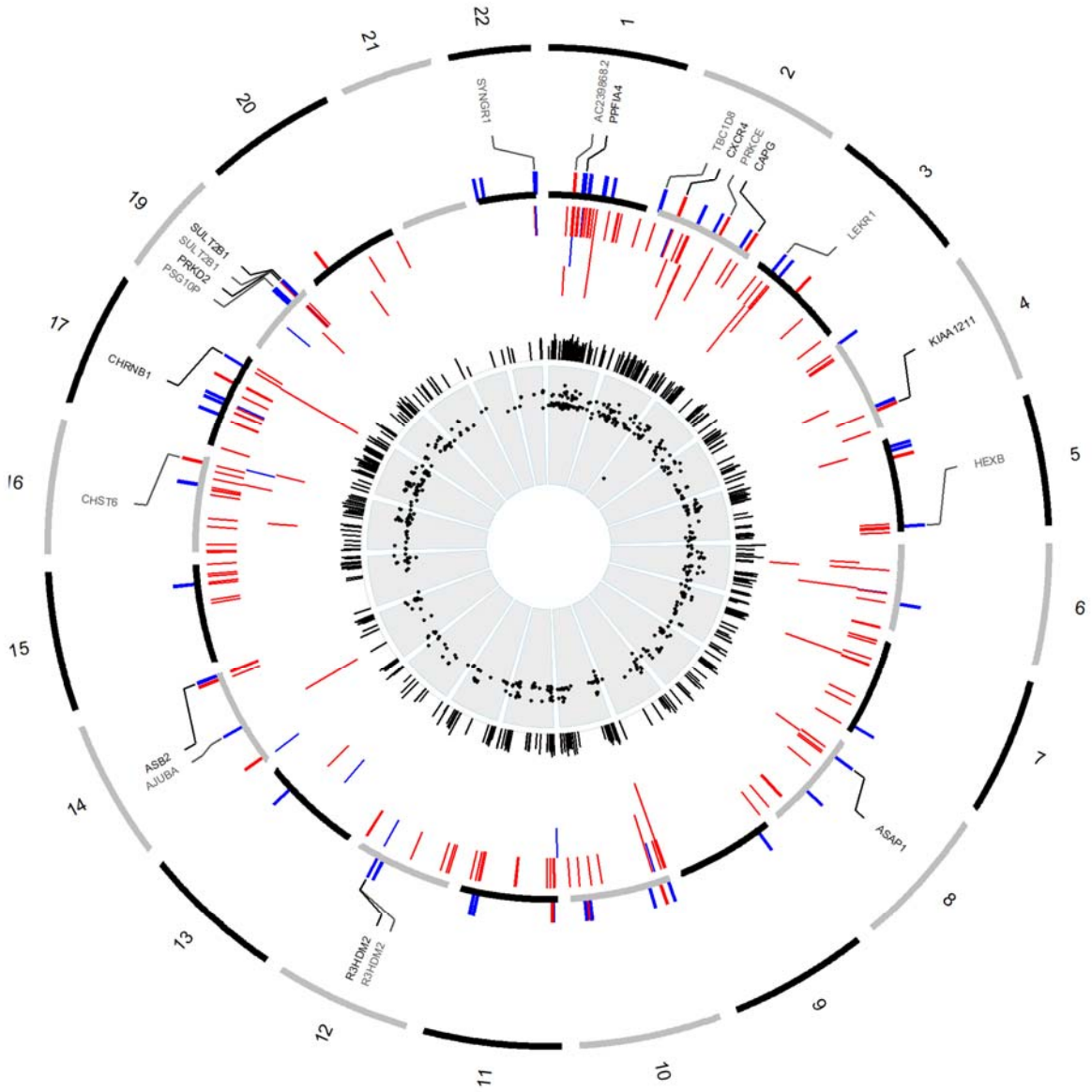

Figure 4. Circos plot summarizing all analyses for the 548 CpG sites that yielded Bonferroni-significant associations with both any and sustained MSDP. From inner to outer-band: Black points = differential DNAm between any- and no MSDP; Black Bars = percent increase in magnitude of association when comparing results from any MSDP to sustained MSDP; Color Bars = Positive (red) or Inverse (blue) associations between DNAm and HC, BL, BW, and GA, that yielded Bonferroni-significant associations; Color Bars = Positive (red) or Inverse (blue) associations between DNAm and nearby gene expression, that yielded an association at a 5% FDR; Names of eQTM genes were annotated to the outer band for CpGs that yielded an association between DNAm and expression with a 5% FDR and yielded an association between DNAm and at least one birth outcome after Bonferroni-adjustment.