1 **Variable DNA methylation in neonates mediates the association between prenatal**
2 **smoking and birth weight.**
3

4 Eilis Hannon[1], Diana Schendel[2,3,4], Christine Ladd-Acosta[5,6], Jakob Grove[4,7,8,9], Christine
5 Søholm Hansen[4,10,11,], David Michael Hougaard[4,10], Michaeline Bresnahan[12], Ole Mors[4,13],
6 Mads Vilhelm Hollegaard[4,10], Marie Bækvad-Hansen[4,10] , Mady Hornig[12,14], Preben Bo
7 Mortensen[3,4,7,15], Anders D. Børglum[4,7,8], Thomas Werge[4,11,16] , Marianne Giørtz
8 Pedersen[3,4,15] , Merete Nordentoft[4,17] , iPSYCH-Broad[4], Joseph D. Buxbaum[18], M Daniele
9 Fallin[6,19] , Jonas Bybjerg-Grauholm[4,10], Abraham Reichenberg[18], Jonathan Mill[1].
10
11

[1] University of Exeter Medical School, University of Exeter, UK.

[2] Department of Public Health, Aarhus University, Aarhus, Denmark.

[3] National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark.

[4] The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark

[5] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA.

[6] Wendy Klag Center for Autism and Developmental Disabilities, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA.

[7] Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark

[8] Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark

[9] Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark

[10] Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark

[11] Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen, Denmark

[12] Center for Infection and Immunity, Columbia University Mailman School of Public Health, New York, USA

[13] Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark

[14] Department of Epidemiology, Columbia University Mailman School of Public Health, New York, USA

[15] Centre for Integrated Register-based Research, Aarhus University, Aarhus, Denmark

[16] Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

[17] Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, Copenhagen, Denmark

[18] Seaver Autism Center for Research and Treatment, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[19] Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA.

Corresponding author:
Jonathan Mill, RILD Building, Level 4, University of Exeter Medical School, Royal Devon & Exeter Hospital, Barrack Rd, Exeter. EX2 5DW. UK.
E-mail: J.Mill@exeter.ac.uk; Telephone: +44 (0)1392 408301

Keywords: DNA methylation, birth weight, gestational age, maternal smoking, epigenome-wide association study (EWAS), mediation analysis

Summary

There is great interest in the role epigenetic variation induced by non-genetic exposures may play in the context of health and disease. In particular, DNA methylation has previously been shown to be highly dynamic during the earliest stages of development and is influenced by *in utero* exposures such as maternal smoking and medication. In this study we sought to identify the specific DNA methylation differences in blood-associated prenatal exposures including birth weight, gestational age and maternal smoking. We quantified neonatal methylomic variation in 1,263 infants using DNA isolated from a unique collection of archived blood spots taken shortly after birth (mean = 6.08 days; sd = 3.24 days). An epigenome-wide association study (EWAS) of gestational age and birth weight identified 4,299 and 18 differentially methylated positions (DMPs) respectively, at an experiment-wide significance threshold of $P < 1x10^{-7}$. Our EWAS of maternal smoking during pregnancy identified 110 DMPs in neonatal blood, replicating previously reported genomic loci including *AHRR*. Finally, we tested the hypothesis that DNA methylation mediates the relationship between maternal smoking and lower birth weight, finding evidence that methylomic variation at three DMPs may link exposure to outcome. These findings complement an expanding literature on the epigenomic consequences of prenatal exposures and obstetric factors, confirming a link between the maternal environment and gene regulation in neonates.

Main Text

Introduction

Epigenetic mechanisms developmentally regulate gene expression via modifications to DNA, histone proteins, and chromatin. Because epigenetic processes can be influenced by exposure to a range of external environmental factors (1-4) and also by genetic variation(5, 6) , there is great interest in the role that epigenetic variation may play in the context of health and disease(7) . As epigenetic marks are inherited mitotically in somatic cell lineages, they provide a mechanism by which disruption early in life can be propagated through development, producing long-term phenotypic variation. DNA methylation is the best-characterized epigenetic modification, stably influencing gene expression via the disruption of transcription factor binding and recruitment of methyl-binding proteins that initiate chromatin compaction and gene silencing. Despite being often regarded as a mechanism of transcriptional repression, DNA methylation is associated with both increased and decreased gene expression (8), and also influences other genomic functions including alternative splicing and promoter usage (9).

The availability of high-throughput profiling methods for quantifying DNA methylation across the genome at single base resolution in large numbers of samples has enabled researchers to perform epigenome-wide association studies (EWAS) aimed at identifying methylomic variation associated with environmental exposures and disease (7). Although these studies are inherently more complex to design and interpret than genetic association studies (10-12), recent analyses have documented differences in DNA methylation in neonates and children following exposure to a wide range of environmental factors during gestation, including maternal smoking, maternal diet and pollution (2, 3, 13). There is also interest in the role that DNA methylation may play as mediator through which environment exposures can influence long-term health outcomes. For example, there is evidence that the causal relationship between maternal smoking during pregnancy and low birth weight is mediated through differences in DNA methylation at specific loci across the genome (14, 15). While noteworthy, these analyses have been based on moderate samples sizes, have generally not replicated the same loci, and may have overestimated mediation effects because of invalid assumptions and misclassification of the exposure (16).

Another active area of research concerns the utility of DNA methylation as a biomarker for clinical monitoring and screening. The potential of a DNA methylation based predictor has been most robustly demonstrated for age, with a number of algorithms available, referred to as “epigenetic clocks” (17-19). There is particular interest in how measures of age derived

from DNA methylation data correlate with actual chronological age, and also whether "accelerated" epigenetic age predicts aging phenotypes such as mortality, cancer and dementia (18, 20, 21). There is also interest in whether other exposures (or phenotypes) can be inferred from an epigenetic profile. The development of an epigenetic biomarker using neonatal blood samples might enable the evaluation of *in utero* exposures, which are hard to measure objectively, and could be a useful prospective predictor for future health outcomes. To this end, a biomarker of maternal smoking during pregnancy was recently developed using cord blood samples that demonstrates high specificity (97%) - but only moderate sensitivity (58%)(22) - demonstrating the potential application of such approaches. Because DNA methylation is known to be highly dynamic during the earliest stages of development (23, 24), insults during this period may have important functional consequences or impact upon disease susceptibility later in life. Of note, several psychiatric disorders are hypothesised to have important neurodevelopmental origins (25-27) and have been associated with a number of prenatal and perinatal risk factors. For example, epidemiological studies have reported a higher risk of autism in those born with a low birth weight (28, 29) or born pre-term (30). Although DNA methylation predictors for age (18) and smoking (31) developed in childhood or adult samples have been shown to work reasonably well, methods developed specifically in either cord blood or neonatal samples have superior performance at these ages (19).

In this study, we first sought to identify specific patterns of DNA methylation in neonatal blood samples associated with three obstetric and neonatal influences measured in the same individuals: birth weight, gestational age and exposure to maternal smoking. We subsequently use our results to explore whether variable DNA methylation mediates the relationship between maternal smoking and low birth weight. We attempted to address these questions by quantifying methylomic variation in 1,263 infants using DNA isolated from archived blood spots taken shortly after birth (mean = 6.08 days; sd = 3.24 days) originally profiled in a case-control study of autism (32). Although we cannot exclude the role of DNA methylation changes occurring during the first few days after birth, our study extends previous research into the early-life epigenome that have either used samples collected later in childhood or from cord blood (13, 24, 33), which has the limitation that it may be contaminated by maternal blood (34, 35). Our findings complement the expanding literature on the epigenomic consequences of prenatal exposures and obstetric factors, confirming a link between the maternal environment and markers of gene regulation in neonates.

1 **Methods**
2 *Overview of the MINERvA cohort*
3 A description of the MINERvA cohort was recently published alongside extensive details of
4 the profiling of DNA methylation and data quality control steps (32). Briefly, MINERvA
5 contains a subset of 1,316 samples from the iPSYCH autism spectrum disorder case-control
6 sample (36). All perinatal data used for case-control sample matching, plus additional
7 information on birth weight and maternal smoking were obtained from the Danish Medical
8 Birth Register or the Central Person Register. An overview of the demographic
9 characteristics of the MINERvA cohort is given in **Supplementary Table 1**. Of note, cases
10 and controls were matched as closely as possible. Although rates of maternal smoking were
11 higher in autism cases, there was no significant difference in birth weight between autism
12 cases and controls.
13
14 *DNA methylation profiling in MINERvA*
15 Neonatal dried blood spot samples collected on standard Guthrie cards and stored within the
16 Danish Neonatal Screening Biobank (37) were retrieved as part of the iPSYCH study (36).
17 Neonatal DNA extractions and DNA methylation quantification was performed at the Statens
18 Serum Institut (SSI, Copenhagen, Capital Region, Denmark). Briefly, DNA was converted
19 with sodium bisulfite using the EZ-96 DNA Methylation Kit (Zymo Research, California,
20 United States of America) and DNA methylation was quantified across the genome using the
21 Infinium HumanMethylation450k array ("450K array") (Illumina, California, United States of
22 America). After a stringent quality control process (outlined in (32)), 1,263 samples (96.0%)
23 were included for subsequent analysis. Normalization of the DNA methylation data was
24 performed used the *dasen*() function in the *wateRmelon* package(38). For each sample, we
25 derived nine additional variables from the DNA methylation data using established
26 algorithms: DNA methylation age(18), gestational age(19), smoking(31), and six blood cell
27 composition variables(39, 40). Our previous publication on these data demonstrated that the
28 smoking score, despite being trained in adults who smoked, correlated with reported
29 maternal smoking status from the registry data (32). All quality control and statistical
30 analyses were performed using the R statistical environment version 3.2.1 (41).
31
32 *Epigenome-wide association analyses (EWAS)*
33 We performed an EWAS of three obstetric/neonatal factors that were robustly measured in
34 our cohort. First, to identify DNA methylation sites associated with birth weight (measured in
35 grams (g)) and gestational age (measured in weeks), a linear model was fitted for each DNA
36 methylation site with DNA methylation as the dependent variable, both birth weight and
37 gestational age as independent variables, and a set of possible confounders as covariates:

1 sex, experimental array number (i.e., chip), days to sampling, maternal smoking (using the
2 continuous variable estimated from the DNA methylation data) and six derived cell
3 composition variables. Given the strong concordance between the findings of our maternal
4 smoking EWAS and those of previous EWAS analyses of smoking (see Results), we used
5 the derived maternal smoking score to make up for missing data in the registry data and
6 maximize power for analyses. To compare results between autism spectrum disorder (ASD)
7 cases and controls, we tested for a heterogeneous effect by including an interaction term
8 between i) birth weight and case control status, and ii) gestational age and case control
9 status. In these interaction models, ASD case control status was also included as a main
10 effect. Second, to identify differentially methylated positions (DMPs) associated with registry-
11 reported maternal smoking exposure, a linear model was fitted for each DNA methylation
12 site with DNA methylation as the dependent variable, and a binary indicator variable for *in*
13 *utero* exposure to smoking in addition to a set of possible confounders as covariates: sex,
14 birth weight, gestational age, experimental array number (i.e. chip), and six derived blood
15 cell composition variables. Significant DMPs were identified at an experiment-wide multiple
16 testing adjusted threshold of $P < 1x10^{-7}$. Clustering of significant DMPs into loci was
17 performed by taking each significant site in turn, starting with the one with the smallest p
18 value (referred to as the index association), identifying all other significant sites within 5kb
19 upstream and downstream and merging these into a single locus. Any less significant (i.e.
20 larger p value) DMPs merged with an index site were then excluded from consideration as
21 an index association. This procedure was repeated until all significant DMPs were either
22 merged with a more significant association or considered as an index site. Conditional
23 analyses were performed within loci with at least two DMPs by repeating the original
24 association analysis for the secondary signal (i.e. the less significant site) including the most
25 significant DNA methylation site in that loci as an additional covariate.
26
27 *Replication dataset*
28 The Accessible Resource for Integrated Epigenomic Studies (ARIES;
29 http://www.ariesepigenomics.org.uk) cohort comprises of a sub-sample of 1018 ALSPAC
30 (http://www.bristol.ac.uk/alspac/) child–mother pairs with Illumina 450K array DNA
31 methylation data generated from cord blood (n = 914), and whole blood at two time points
32 during childhood (age 7 (n = 973) and age 15 or 17 years (n = 974)). The results used in this
33 manuscript are taken from the gestational age and birth weight EWAS performed by Simpkin
34 *et al*(24) and presented in Supplementary Table 1 and Table 3 published alongside the
35 manuscript.
36
37 *Mediation analyses*

Mediation analyses were performed using the criteria outlined by Baron and Kenny(42) and the Sobel test(43). We considered DMPs associated with registry-reported maternal smoking without controlling for birth weight in our dataset (n = 143 DMPs) and tested whether the following criteria were met for each site:

(i) smoking significantly correlated with DNA methylation level ($P < 1x10^{-7}$; sex, gestational age, batch and cell composition included as covariates).

(ii) smoking significantly correlated with birth weight without adjusting the model for DNA methylation (sex and gestational age included as covariates).

(iii) DNA methylation significantly correlated with birth weight ($P < 1x10^{-7}$; sex, gestational age, batch and cell composition included as covariates).

(iv) the association between smoking and birth weight decreased upon addition of DNA methylation to the model (i.e. p value got larger; sex, gestational age, batch and cell composition included as covariates).

(v) the Sobel test gave $P < 3.50 \times 10^{-4}$ (corrected for 143 DMPs considered, implemented through the R *bda* package (https://cran.r-project.org/web/packages/bda/index.html).

DMPs meeting the criteria for mediation were taken forward for a sensitivity analysis that accounted for misclassification of the exposure following the method outlined in Valeri *et al*(16) using the SIMEX procedure(44). Our naive outcome regression model between birth weight and registry-reported maternal smoking exposure and naïve mediator regression model between DNA methylation and birth weight included covariates for gestational age, sex, cellular composition and experimental chip. The naive direct effect is then the coefficient from the outcome regression model for maternal smoking and the naïve indirect effect is the naive direct effect multiplied by the estimated coefficient for maternal smoking from the mediator regression. Applying SIMEX to outcome and mediator regressions, we obtained corrected estimates of the regression parameters which were then used to calculate the direct and indirect effects. We set the parameter for specificity to 1.0 as we assume that all smokers are likely to have reported correctly, whereas we assume that some non-smokers will have reported incorrectly therefore the sensitivity parameter was set to 0.6. A bootstrap method was used to estimate the standard errors of the estimated effects and their 95% confidence intervals.

1 **Results**

2 *Blood cell proportions derived from DNA methylation data correlate with birth weight and*
3 *gestational age in neonatal blood.*

4 Our first analyses aimed to explore whether measures derived from DNA methylation data at
5 birth (i.e. gestational age and blood cell composition estimates) are associated with birth
6 weight. We previously demonstrated the robustness of DNA methylation data derived from
7 neonatal blood spots by implementing two DNA methylation clock algorithms to derive
8 estimates for i) age in years(18) and ii) gestational age in weeks(19) for each sample. As
9 expected, we observed a strong positive correlation between estimated and actual
10 gestational age (r = 0.602, 95% CI = (0.566, 0.636), P = $3.80x10^{-125}$) and a weaker positive
11 correlation between estimated chronological age and actual gestational age (r = 0.139, 95%
12 CI = (0.0849, 0.193), P = $6.52x10^{-7}$)(32). We next extended these analyses to investigate
13 how birth weight correlates with these predicted ages and whether variable birth weight
14 explains the difference between predicted age (derived from DNA methylation data) and
15 actual age. Birth weight was significant correlated with predicted measures of both age (r =
16 0.119, 95% CI = (0.0638, 0.173), P = $2.40x10^{-5}$; **Supplementary Figure 1A**) and gestational
17 age (r = 0.333, 95% CI = (0.0849, 0.193), P = $5.15x10^{-34}$; **Supplementary Figure 1B**).
18 However, the age acceleration residuals - which are adjusted for reported gestational age -
19 are not significantly associated with birth weight (P > 0.05, **Supplementary Figure 1C, 1D**)
20 indicating that variation in birth weight does not explain the difference between reported
21 gestational age and predicted age and that other pregnancy and/or obstetric factors may be
22 influencing derived age estimates. Given the difficulties in collecting large volumes of blood
23 from neonates, little is known about blood cell-type variation at this stage of life. We
24 therefore explored how predicted cellular composition variables derived from the DNA
25 methylation data (see **Methods**) correlate with birth weight and reported gestational age.
26 Gestational age was positively correlated with the estimated proportions of CD8 T- cells (r =
27 0.140; 95% CI = (0.0857,0.194), P = $5.63x10^{-7}$) and natural killer cells (r = 0.0722, 95% CI =
28 (0.0171,0.127), P = 0.0103), and negatively correlated with the estimated proportion of B
29 cells (r = -0.231, 95% CI = (-0.282, -0.178), P = $1.06x10^{-16}$) (**Supplementary Figure 2**). Birth
30 weight was significantly negatively correlated with the estimated proportion of monocytes (r
31 = -0.0604, 95% CI = (-0.115, -0.00529), P = 0.0317) and positively correlated with the
32 estimated proportion of granulocytes (r = 0.0624, 95% CI = (0.00727, 0.117), P = 0.0266)
33 (**Supplementary Figure 3**). Given the potential confounding influence of cellular
34 heterogeneity in EWAS analyses using blood, these derived variables were included in all
35 subsequent analyses.

36

Birth weight and gestational age are associated with variable DNA methylation in neonatal blood

To identify DNA methylation sites associated with reported gestational age and birth weight we next performed an EWAS across all Illumina 450K array sites on the autosomes and X-chromosome (n = 430,676 sites), undertaking the analyses simultaneously to minimise confounding resulting from the strong correlation between these two obstetric variables (r = 0.491,95% CI = (0.448,0.532), P = 1.14×10^{-77}; **Supplementary Figure 4**). In total, we identified 18 differentially methylated positions (DMPs) associated with birth weight (**Figure 1; Supplementary Figure 5** and **Supplementary Figure 6**) and 4,299 DMPs associated with gestational age (**Figure 1; Supplementary Figure 7** and **Supplementary Figure 8**) at an experiment-wide significance threshold (P < 1×10^{-7}) (**Supplementary Tables 2** and **3**). The associated DMPs were characterized by a median shift in DNA methylation of 1.40% (SD = 0.368%) per kg and 0.406% (SD = 0.275) per gestational week. Seven sites were significantly associated with both birth weight and gestational age (**Table 1**). Sensitivity analyses repeating the EWAS excluding samples from 'outlier' individuals born i) before 35 weeks (N = 23) or ii) before 32 weeks (N = 5) revealed high concordance with our primary analysis (**Supplementary Figure 9**), suggesting that the results are robust to the presence of premature individuals. Although the majority of DMPs associated with increased birth weight were associated with reduced DNA methylation (66.7%) there was not a significant bias (sign test P = 0.238) due to the small total number of DMPs identified. In contrast, there was a highly-significant bias towards increased DNA methylation at sites associated with older gestational age (73.2%, sign test P = 2.29×10^{-135}). Although this contradicts results from a previous study using cord blood derived DNA from the ARIES cohort (n = 914) which identified a smaller number of DMPs that were enriched for sites showing a decrease in DNA methylation with older gestational age (24), our most significant DMPs are characterised by reduced DNA methylation with age (**Supplementary Table 3**) and associations at sites showing this pattern of change are significantly stronger (Mann-Whitney P = 2.58×10^{-228}). Furthermore, we find a significant excess of consistent effects between studies indicating that age-associated changes are similar in cord and neonatal whole blood (**Supplementary Figure 10**). Of 148 DMPs associated with gestational age in the ARIES cohort, 146 (98.6%) had the same direction of effect (sign test P = 6.18×10^{-41}), with 110 being significantly associated (P < 1×10^{-7}) in both cohorts. We found similar consistency for our EWAS of birth weight; all 21 DMPs associated with birth weight in the ARIES cohort had the same direction of effect in our data (sign test P < 2×10^{-323}), with two DMPs being significantly associated (P < 1×10^{-7}) in both cohorts (**Supplementary Figure 10**). While the 18 DMPs associated with birth weight are annotated to distinct genomic loci, the 4,299 DMPs associated with gestational age are clustered into 3,550 distinct locations with up to 25 additional DMPs

1 located within 5kb of the index DMP characterized by the most significant association.
2 Conditional analyses within each genomic locus containing at least two DMPs (n = 483;
3 13.6%) revealed evidence for independent secondary signals for 240 DMPs in 193 of these
4 loci (conditional P < $5x10^{-5}$), while 100 DMPs within 66 loci were only associated as a result
5 of their correlation with the most significant DMP in that loci (conditional P > 0.05;
6 **Supplementary Table 4**). Finally, given the association between both low birth weight and
7 pre term birth and autism we tested whether the DNA methylation differences we identified
8 were consistent between individuals who later went on to develop a childhood diagnosis of
9 autism (n = 629) and matched controls (n = 634). There were no significant differences
10 between autism cases and controls for DMPs associated with birth weight (min P = 0.0446)
11 or for DMPs associated with gestation age (min P = 3.51x10-5) after correcting for the
12 number of DMPs significant in each analysis (birth weight: P < 0.00278 corrected for 18
13 DMPs; gestational age: P < 1.16x10-5 corrected for 4,299 DMPs) (**Supplementary Table 2**
14 and **Supplementary Table 3**).
15
16 *Maternal smoking influences DNA methylation in neonates at multiple loci*
17 Exposure to tobacco smoke is known to be associated with widespread alterations in DNA
18 methylation in whole blood (1), and previous analyses have demonstrated that these effects
19 can be detected in cord blood from neonates exposed to prenatal smoking (13). One
20 limitation of using cord blood is that DNA methylation estimates can be influenced by
21 contamination of maternal blood, and we therefore sought to test these whether these
22 associations were detectable in neonatal whole blood samples. Mothers were asked about
23 their smoking status at their first prenatal visit, early in the second trimester of pregnancy.
24 Amongst the mothers of neonates in our cohort, 294 (25.1%) reported smoking during
25 pregnancy and 879 (74.9%) reported not smoking during pregnancy; we excluded 36
26 mothers who reported giving up smoking at some time during the pregnancy and 54 mothers
27 for whom smoking data was not available. First, we assessed whether maternal smoking
28 influenced derived measures of age generated from DNA methylation data in these samples.
29 Neither age, gestational age nor age acceleration was associated with *in utero* exposure to
30 smoking (**Supplementary Figure 11**). Next, we performed an EWAS of maternal smoking
31 exposure (n = 1,173, controlling for sex, birth weight, gestational age, experimental batch,
32 and derived cell composition variables), identifying 110 neonatal blood DMPs associated
33 with maternal smoking (P < $1x10^{-7}$) representing 70 discrete genomic loci (**Figure 2;**
34 **Supplementary Figure 12, Supplementary Table 5**). Conditional analyses within each
35 genomic loci with at least two DMPs (n = 13) identified seven loci where a single DMP was
36 associated with maternal smoking (conditional P for other sites > 0.05) and four loci

1 characterized by secondary semi-independent effects (conditional P for other sites < $5x10^{-5}$;
2 **Supplementary Table 6**). There was no significant bias towards a particular direction of
3 effect (50.9% hypomethylated; 49.1% hypermethylated, sign test P = 0.924) and the median
4 effect was a difference of 2.28% DNA methylation (SD = 2.19%) (**Supplementary Figure**
5 **13**). There was considerable overlap with DMPs reported in a large EWAS of maternal
6 smoking performed in cord blood (13) (n = 6,685; **Supplementary Figure 14**). Of note,
7 4,847 (84.0%) of the 5,768 DMPs reported by Joubert *et al*. were characterized by the same
8 direction of effect (sign test P < $2x10^{-323}$) in our analysis of neonatal blood, with 102 meeting
9 criteria for experiment-wide significance (P < $1x10^{-7}$) in both studies. This included previously
10 reported DMPs associated with tobacco smoking in adults (1, 31) such as *AHRR*, where five
11 additional DMPs were clustered with the lead signal at cg05575921 none of which remained
12 significant in the conditional analysis, *GFI1* which had nine DMPs including multiple
13 independent associations, and *MYO1G* which had four DMPs with evidence of multiple
14 independent associations (**Supplementary Table 6**). These findings confirm that smoking
15 behaviour by mothers during pregnancy has a profound influence on DNA methylation in
16 their offspring at birth and in the first few days of life.

17

18 *DNA methylation mediates the relationship between maternal smoking and low birth weight*

19 Having established that the DNA methylation signatures associated with prenatal smoking
20 exposure are robustly detectable in neonatal blood, and that variable DNA methylation is
21 associated with an established outcome of maternal smoking (45, 46) we next asked
22 whether methylomic variation might *mediate* the relationship between maternal smoking in
23 pregnancy and lower birth weight. Previous attempts to explore this using data from cord
24 blood have been relatively inconsistent (14, 15), supporting a mediation role for DNA
25 methylation at non-overlapping DNA methylation sites. We repeated our EWAS of maternal
26 smoking, excluding birth weight as a covariate, and identified an extended set of 143 DMPs
27 (**Supplementary Table 7**) which contained 105 (95.5%) of the 110 DMPs we identified in
28 the analysis adjusting for birth weight. Mediation analyses, performed using the Sobel test as
29 (see **Methods**), showed that DNA methylation at three sites, annotated to three different
30 genomic loci, met the five criteria set out by Baron and Kenny (42) (see **Methods**) as
31 providing evidence for mediating the association between smoking and birth weight
32 (**Supplementary Figure 15; Supplementary Table 7**). This included one site (cg09935388)
33 annotated to *GFI1* that has been reported previously (15) and two novel mediation sites
34 (cg05575921 annotated to *AHRR*, and cg26889659 annotated to *EXOC2*) (**Table 2**).
35 Because smoking behaviour is prone to misclassification - not only due to smokers claiming

to be non-smokers, especially during pregnancy(47), but also because a complex behaviour is simplified into a dichotomous variable - we repeated the analysis for the three significant DNA methylation sites estimating the natural direct effect between maternal smoking and low birth weight (i.e. not via DNA methylation) and the natural indirect effect (i.e. via DNA methylation) using the SIMEX (simulation and extrapolation) procedure that incorporates misclassification (44). Stringently accounting for misclassification suggests that the estimated mediation effect via DNA methylation identified using the Sobel test is potentially overestimated; although the results robustly support a significant mediation effect for DNA methylation at cg26889659, under scenarios of more extreme misclassification (see **Methods**) the effects of mediation via DNA methylation at cg05575921 and cg09935388 are no longer significantly different from 0 (**Supplementary Table 8**).

Discussion

In this study we quantified neonatal variation in DNA methylation in 1,263 infants using samples isolated from archived blood spots taken shortly after birth. Our study finds that gestational age, birth weight and maternal smoking are all associated with significant DNA methylation differences in neonatal blood, with gestational age having the most effects widespread across the genome. These data add to a growing literature demonstrating that prenatal and obstetric exposures can influence epigenetic variation in early-life (3, 13, 24, 48-50), providing a potential mechanism linking them to altered gene function and long-term health and disease outcomes.

Our use of neonatal DNA samples means that we are uniquely positioned to identify epigenetic variation at birth, avoiding the confounding exposures that could influence the results from samples collected later in childhood (for example, health and disease, nutrition, medication, and stress). Although there have been larger studies of prenatal exposures, a strength of our study is that we profiled whole blood from neonatal infants rather than cord blood minimizing the contamination of our samples with maternal blood DNA and meaning our data can be more easily compared to the extensive blood DNA methylation datasets derived from samples later in life. A limitation of our sampling strategy, however, is that no blood cell reference DNA methylation datasets specifically for use on neonatal blood are yet available, likely reflecting the difficulties of obtaining sufficient volumes of neonatal blood for cell sorting and methylomic profiling. Although there has been much written on the importance of selecting an appropriate tissue to profile for epigenetic studies(10), the goal of this study was to identify biomarkers of exposure, and therefore our use of a peripheral tissue is justified. Furthermore, although blood spots were collected only a few days after birth (mean = 6.08 days; sd = 3.24 days), it is possible that some postnatal exposure to passive smoking during the first few days of life could also have influenced our results although the amount of exposure in ~6 days is likely to be negligible.

As well as identifying specific loci at which DNA methylation is associated with with early-life factors such as smoking exposure and gestational age, we tested the hypothesis that DNA methylation *mediates* established epidemiological relationships between exposures and outcomes. We explored the association between maternal smoking and lower birth weight, finding evidence that methylomic variation at several DMPs may be mechanistically involved in linking exposure (maternal smoking) to outcome (birth weight). While our results are consistent with previous reports, such analyses can be influenced by misclassification bias (16). Smoking, in particular is prone to misclassification not only due to participants claiming

to be non-smokers when they are in fact smokers, a circumstance known to be worse when reflecting smoking during pregnancy(47), but also as a result of simplifying a complex behaviour into a single dichotomous variable representing the entire period of pregnancy. Given our robust prenatal smoking exposure associations, it is plausible that the mediator in our analyses - DNA methylation - is in fact a better measure of smoking exposure than self-reported status(16). Of note, applying a methodology that accounts for misclassification of an exposure reduced the magnitude of the mediation effect at all three significant loci, suggesting that these results need validation using an alternative approach such as Mendelian Randomisation (51).

While we explored whether DNA methylation lies on the causal pathway between maternal smoking and low birth weight, results from EWAS analyses do not distinguish cause from effect. In fact, it is likely that the DNA methylation differences we report for birth weight and gestational age reflect other *in utero* exposures or processes. For example birth weight is known to be associated with maternal body mass index, blood pressure and fasting glucose levels (52-54) and the epigenetic changes we report may reflect the downstream influences of these pathways. It is also possible that the birth-weight-associated DMPs identified in our study reflect exposures or influences occurring in the immediate neonatal period before the blood samples were collected. Similarly, the DNA methylation differences observed in neonates exposed to prenatal smoking might also be influenced by exposure to passive smoking from the mother or father immediately after birth, although the amount of exposure in ~6 days is likely to be negligible. Another potential limitation of our design is that the analyses were performed within the context of an autism case-control study (32). Of note, however, although autism cases had a higher exposure to maternal smoking than non-autism controls, there was not significant difference in birth weight between infants who went on to develop autism than those who did not (**Supplementary Table 1**). Finally, the nature of the samples we profiled in this study (i.e. small amounts of neonatal blood) meant that additional DNA was not available for technical validation experiments. However, the Illumina 450K array has been shown to yield highly reproducible measures of DNA methylation, and the observed consistency with previous studies of maternal smoking, gestational age, and birth weight suggests our findings are robust.

Conclusions

Our data demonstrate that *in utero* exposures are associated with detectable patterns of DNA methylation in neonatal blood samples, highlighting the role that the prenatal environment plays in influencing gene regulation in neonates. While previous studies have

1 shown that maternal smoking effects persist into later childhood (13), these have found that
2 the effects are attenuated, suggesting that obtaining a biomarker as close to birth as
3 possible will have maximal sensitivity about exposures during gestation.
4
5 **Acknowledgments**
6 We acknowledge iPSYCH and The Lundbeck Foundation for providing samples and funding.
7
8 **Ethics**
9 The MINERvA study was approved by the Regional Scientific Ethics Committee in Denmark
10 and the Danish Data Protection Agency.
11
12 **Data Accessibility**
13 Given the nature of the MINERvA cohort, access to data can only be provided through
14 secured systems which comply with the current Danish and EU data standards. To comply
15 with the study's ethical approval, access to the raw data is only available to qualified
16 researchers upon request. All summary statistics and analysis scripts are available directly
17 from the authors (please contact Jonas Bybjerg-Grauholm at JOGR@ssi.dk). R scripts used
18 to perform the quality control of these data are available on GitHub
19 (https://github.com/ejh243/MinervaASDEWAS.git) and have been archived in Zenado at
20 https://zenodo.org/badge/latestdoi/116149862 and scripts for the analyses reported in this
21 manuscript are available on GitHub (https://github.com/ejh243/MinervaNeonatalEWAS.git)
22 and have been archived in Zenado at https://doi.org/10.5281/zenodo.1303340.
23
24 **Authors' Contributions**
25 EH, JM, AR and DS designed and coordinated the study. GB-G, DMH, MVH, M-BH and
26 CSH led generation of DNA methylation data from dried neonatal bloodspots. EH led and
27 AR, DS, JM, JB-G, JG, CL-A and MDF oversaw implementation of the data analyses. DMH,
28 OM, PBM, ADB, TW, and MN are principal investigators of the iPSYCH study. EH and JM
29 drafted the manuscript, with input from AR, DS, CL-A, JG, and JB-G. All co-authors read and
30 approved the final manuscript.
31
32 **Competing Interests**
33 TW has acted as advisor and lecturer to H. Lundbeck A/S. None of the other authors report
34 any potential conflict of interest.
35
36

Funding

This study was supported by grant HD073978 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institute of Environmental Health Sciences, and National Institute of Neurological Disorders and Stroke; and by the Beatrice and Samuel A. Seaver Foundation. The iPSYCH (The Lundbeck Foundation Initiative for Integrative Psychiatric Research) team acknowledges funding from The Lundbeck Foundation (grant no R102-A9118 and R155-2014-1724), the Stanley Medical Research Institute, the European Research Council (project no: 294838), the Novo Nordisk Foundation for supporting the Danish National Biobank resource, and grants from Aarhus and Copenhagen Universities and University Hospitals, including support to the iSEQ Center, the GenomeDK HPC facility, and the CIRRAU Center. This research has been conducted using the Danish National Biobank resource, supported by the Novo Nordisk Foundation. JM and EH are supported by funding from the UK Medical Research Council (K013807).

References

1. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic Signatures of Cigarette Smoking. Circ Cardiovasc Genet. 2016;9(5):436-47.
2. Tobi EW, Goeman JJ, Monajemi R, Gu H, Putter H, Zhang Y, et al. DNA methylation signatures link prenatal famine exposure to growth and metabolism. Nat Commun. 2014;5:5592.
3. Gruzieva O, Xu CJ, Breton CV, Annesi-Maesano I, Antó JM, Auffray C, et al. Epigenome-Wide Meta-Analysis of Methylation in Children Related to Prenatal NO2 Air Pollution Exposure. Environ Health Perspect. 2017;125(1):104-10.
4. Panni T, Mehta AJ, Schwartz JD, Baccarelli AA, Just AC, Wolf K, et al. A Genome-Wide Analysis of DNA Methylation and Fine Particulate Matter Air Pollution in Three Study Populations: KORA F3, KORA F4, and the Normative Aging Study. Environ Health Perspect. 2016.
5. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, et al. Systematic identification of genetic influences on methylation across the human life course. Genome Biol. 2016;17:61.
6. Hannon E, Spiers H, Viana J, Pidsley R, Burrage J, Murphy TM, et al. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. Nat Neurosci. 2015.
7. Murphy TM, Mill J. Epigenetics in health and disease: heralding the EWAS era. Lancet. 2014;383(9933):1952-4.
8. Wagner JR, Busche S, Ge B, Kwan T, Pastinen T, Blanchette M. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. Genome Biol. 2014;15(2):R37.
9. Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature. 2010;466(7303):253-7.
10. Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. Nat Rev Genet. 2013;14(8):585-94.
11. Relton CL, Davey Smith G. Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment. PLoS Med. 2010;7(10):e1000356.
12. Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. Nat Rev Genet. 2011;12(8):529-41.
13. Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, et al. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. Am J Hum Genet. 2016;98(4):680-96.

1 14. Witt SH, Frank J, Gilles M, Lang M, Treutlein J, Streit F, et al. Impact on birth weight
2 of maternal smoking throughout pregnancy mediated by DNA methylation. Bmc Genomics.
3 2018;19(1):290.
4 15. Küpers LK, Xu X, Jankipersadsing SA, Vaez A, la Bastide-van Gemert S, Scholtens
5 S, et al. DNA methylation mediates the effect of maternal smoking during pregnancy on
6 birthweight of the offspring. Int J Epidemiol. 2015;44(4):1224-37.
7 16. Valeri L, Reese SL, Zhao S, Page CM, Nystad W, Coull BA, et al. Misclassified
8 exposure in epigenetic mediation analyses. Does DNA methylation mediate effects of
9 smoking on birthweight? Epigenomics. 2017;9(3):253-65.
10 17. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, et al. Genome-wide
11 methylation profiles reveal quantitative views of human aging rates. Mol Cell.
12 2013;49(2):359-67.
13 18. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol.
14 2013;14(10):R115.
15 19. Knight AK, Craig JM, Theda C, Bækvad-Hansen M, Bybjerg-Grauholm J, Hansen
16 CS, et al. An epigenetic clock for gestational age at birth based on blood methylation data.
17 Genome Biol. 2016;17(1):206.
18 20. Chen BH, Marioni RE, Colicino E, Peters MJ, Ward-Caviness CK, Tsai PC, et al.
19 DNA methylation-based measures of biological age: meta-analysis predicting time to death.
20 Aging (Albany NY). 2016;8(9):1844-65.
21 21. Levine ME, Lu AT, Bennett DA, Horvath S. Epigenetic age of the pre-frontal cortex is
22 associated with neuritic plaques, amyloid load, and Alzheimer's disease related cognitive
23 functioning. Aging (Albany NY). 2015;7(12):1198-211.
24 22. Reese SE, Zhao S, Wu MC, Joubert BR, Parr CL, Håberg SE, et al. DNA Methylation
25 Score as a Biomarker in Newborns for Sustained Maternal Smoking during Pregnancy.
26 Environ Health Perspect. 2017;125(4):760-6.
27 23. Spiers H, Hannon E, Schalkwyk LC, Smith R, Wong CC, O'Donovan MC, et al.
28 Methylomic trajectories across human fetal brain development. Genome research.
29 2015;25(3):338-52.
30 24. Simpkin AJ, Suderman M, Gaunt TR, Lyttleton O, McArdle WL, Ring SM, et al.
31 Longitudinal analysis of DNA methylation associated with birth weight and gestational age.
32 Hum Mol Genet. 2015;24(13):3752-63.
33 25. Murray RM, Lewis SW. Is schizophrenia a neurodevelopmental disorder? Br Med J
34 (Clin Res Ed). 1987;295(6600):681-2.
35 26. Weinberger DR. Implications of normal brain development for the pathogenesis of
36 schizophrenia. Arch Gen Psychiatry. 1987;44(7):660-9.

1 27. Ecker C, Bookheimer SY, Murphy DG. Neuroimaging in autism spectrum disorder:
2 brain structure and function across the lifespan. Lancet Neurol. 2015;14(11):1121-34.
3 28. Maimburg RD, Vaeth M. Perinatal risk factors and infantile autism. Acta Psychiatr
4 Scand. 2006;114(4):257-64.
5 29. Gardener H, Spiegelman D, Buka SL. Perinatal and neonatal risk factors for autism:
6 a comprehensive meta-analysis. Pediatrics. 2011;128(2):344-55.
7 30. Larsson HJ, Eaton WW, Madsen KM, Vestergaard M, Olesen AV, Agerbo E, et al.
8 Risk factors for autism: perinatal factors, parental psychiatric history, and socioeconomic
9 status. Am J Epidemiol. 2005;161(10):916-25; discussion 26-8.
10 31. Elliott HR, Tillin T, McArdle WL, Ho K, Duggirala A, Frayling TM, et al. Differences in
11 smoking associated DNA methylation patterns in South Asians and Europeans. Clin
12 Epigenetics. 2014;6(1):4.
13 32. Hannon E, Schendel D, Ladd-Acosta C, Grove J, Hansen CS, Andrews SV, et al.
14 Elevated polygenic burden for autism is associated with differential DNA methylation at birth.
15 Genome Med. 2018;10(1):19.
16 33. Ladd-Acosta C, Shu C, Lee BK, Gidaya N, Singer A, Schieve LA, et al. Presence of
17 an epigenetic signature of prenatal cigarette smoke exposure in childhood. Environ Res.
18 2016;144(Pt A):139-48.
19 34. Masuzaki H, Miura K, Miura S, Yoshiura K, Mapendano CK, Nakayama D, et al.
20 Labor increases maternal DNA contamination in cord blood. Clin Chem. 2004;50(9):1709-11.
21 35. Bauer M, Orescovic I, Schoell WM, Bianchi DW, Pertl B. Detection of maternal
22 deoxyribonucleic acid in umbilical cord plasma by using fluorescent polymerase chain
23 reaction amplification of short tandem repeat sequences. Am J Obstet Gynecol.
24 2002;186(1):117-20.
25 36. Pedersen CB, Bybjerg-Grauholm J, Pedersen MG, Grove J, Agerbo E, Bækvad-
26 Hansen M, et al. The iPSYCH2012 case-cohort sample: new directions for unravelling
27 genetic and environmental architectures of severe mental disorders. Mol Psychiatry. 2017.
28 37. Nørgaard-Pedersen B, Hougaard DM. Storage policies and use of the Danish
29 Newborn Screening Biobank. J Inherit Metab Dis. 2007;30(4):530-6.
30 38. Pidsley R, Wong CCY, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven
31 approach to preprocessing Illumina 450K methylation array data. Bmc Genomics. 2013;14.
32 39. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson
33 HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC
34 Bioinformatics. 2012;13:86.
35 40. Koestler DC, Christensen B, Karagas MR, Marsit CJ, Langevin SM, Kelsey KT, et al.
36 Blood-based profiles of DNA methylation predict the underlying distribution of cell types: a
37 validation analysis. Epigenetics. 2013;8(8):816-26.

1 41. R Development Core Team. R: A Language and Environment for Statistical
2 Computing. . Vienna, Austria: R Foundation for Statistical Computing; 2008.
3 42. Baron RM, Kenny DA. The moderator-mediator variable distinction in social
4 psychological research: conceptual, strategic, and statistical considerations. J Pers Soc
5 Psychol. 1986;51(6):1173-82.
6 43. Sobel ME. Asymptotic Confidence Intervals for Indirect Effects in Structural Equation
7 Models. Sociological Methodology. 1982;13:290-312.
8 44. Küchenhoff H, Mwalili SM, Lesaffre E. A general method for dealing with
9 misclassification in regression: the misclassification SIMEX. Biometrics. 2006;62(1):85-96.
10 45. Durmus B, Kruithof CJ, Gillman MH, Willemsen SP, Hofman A, Raat H, et al.
11 Parental smoking during pregnancy, early growth, and risk of obesity in preschool children:
12 the Generation R Study. Am J Clin Nutr. 2011;94(1):164-71.
13 46. Windham GC, Hopkins B, Fenster L, Swan SH. Prenatal active or passive tobacco
14 smoke exposure and the risk of preterm delivery or low birth weight. Epidemiology.
15 2000;11(4):427-33.
16 47. Dietz PM, Homa D, England LJ, Burley K, Tong VT, Dube SR, et al. Estimates of
17 nondisclosure of cigarette smoking among pregnant and nonpregnant women of
18 reproductive age in the United States. Am J Epidemiol. 2011;173(3):355-9.
19 48. Agha G, Hajj H, Rifas-Shiman SL, Just AC, Hivert MF, Burris HH, et al. Birth weight-
20 for-gestational age is associated with DNA methylation at birth and in childhood. Clin
21 Epigenetics. 2016;8:118.
22 49. Non AL, Binder AM, Kubzansky LD, Michels KB. Genome-wide DNA methylation in
23 neonates exposed to maternal depression, anxiety, or SSRI medication during pregnancy.
24 Epigenetics. 2014;9(7):964-72.
25 50. Schroeder JW, Conneely KN, Cubells JC, Kilaru V, Newport DJ, Knight BT, et al.
26 Neonatal DNA methylation patterns associate with gestational age. Epigenetics.
27 2011;6(12):1498-504.
28 51. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy
29 for establishing the causal role of epigenetic processes in pathways to disease. Int J
30 Epidemiol. 2012;41(1):161-76.
31 52. Tyrrell J, Richmond RC, Palmer TM, Feenstra B, Rangarajan J, Metrustry S, et al.
32 Genetic Evidence for Causal Relationships Between Maternal Obesity-Related Traits and
33 Birth Weight. JAMA. 2016;315(11):1129-40.
34 53. Lawlor DA, Relton C, Sattar N, Nelson SM. Maternal adiposity--a determinant of
35 perinatal and offspring outcomes? Nat Rev Endocrinol. 2012;8(11):679-88.
36 54. Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, Coustan DR, et al.
37 Hyperglycemia and adverse pregnancy outcomes. N Engl J Med. 2008;358(19):1991-2002.

1

2

Tables

Table 1. DNA methylation in neonatal blood is associated with birth weight and gestational age. In total, seven differentially methylated positions (DMPs) were associated with both birth weight and gestational age at an experiment-wide threshold of $P < 1x10^{-7}$. Genomic locations are based on hg19.

ProbeID	EWAS of birth weight (g)		EWAS of gestational age (weeks)		Chr	Pos	Gene Annotation	
	P-value	Regression coefficient	P-value	Regression coefficient			UCSC Gene Name	UCSC Genic Region
cg04411893	9.88E-08	-8.41E-06	2.28E-12	-0.003758699	chr3	185300709		
cg05937055	4.66E-09	-1.20E-05	5.20E-10	-0.004302303	chr1	181128764		
cg06870470	2.97E-09	-1.68E-05	6.93E-28	-0.01068899	chr19	11315767	DOCK6	Body
cg13066703	2.50E-08	-1.11E-05	3.68E-36	-0.008742825	chr1	211526705	TRAF5	Body
cg19744173	1.87E-10	-1.41E-05	8.41E-20	-0.006889143	chr2	112913178	FBLN7	Body
cg20068209	5.19E-08	-1.37E-05	2.00E-41	-0.011959147	chr6	75988568	TMEM30A	Body
cg20076442	1.18E-10	-1.96E-05	3.23E-25	-0.010843225	chr8	72745197		

Table 2. Three DNA methylation sites meet the criteria for significant mediation between birth weight and maternal smoking.

ProbeID	Chr	Pos	Gene Annotation	Birth weight ~ maternal smoking		EWAS of maternal smoking		EWAS of birth weight		EWAS of birth weight adjusted for maternal smoking		Sobel test P-value
				P-value	Regression coefficient	P-value	Regression coefficient	P-value	Regression coefficient	P-value	Regression coefficient	
cg09935388	chr1	92947588	GFI1	3.21E-14	-282.418	1.39E-55	-0.1208	2.58E-08	4.18E-05	2.72E-08	-229.556	3.58E-09
cg05575921	chr5	373378	AHRR	3.21E-14	-282.418	8.29E-119	-0.08196	1.17E-10	2.37E-05	0.000151	-178.772	8.52E-14
cg26889659	chr6	684090	EXOC2	3.21E-14	-282.418	7.42E-13	-0.03981	3.84E-08	2.89E-05	2.41E-10	-238.068	1.12E-06

Figure Captions

Figure 1. Methylomic variation associated with birth weight and gestational age in neonates. Manhattan plots of P-values from an epigenome-wide association study (EWAS) of A) birth weight (g) and B) gestational age (weeks). The red horizontal line indicates experiment-wide significance ($P < 1x10^{-7}$), the blue horizontal line indicates a 'discovery' significance threshold ($P < 5x10^{-5}$). Shown are scatterplots of the top-ranked differentially methylated positions (DMPs) associated with C) birth weight (cg20076442) and D) gestational age (cg11932158).

Figure 2. DNA methylation at birth is associated with maternal smoking during pregnancy. The red horizontal line indicates experiment-wide significance ($P < 1x10^{-7}$), the blue horizontal line indicates a 'discovery' significance threshold ($P < 5x10^{-5}$). A) Manhattan plots from an EWAS of maternal smoking. Boxplots of the three top-ranked DNA methylation sites associated with maternal smoking: B) cg05575921, C) cg09935388 and D) cg12803068. Shown are unadjusted DNA methylation values.

Supplementary Table Captions (see Excel file for Supplementary Tables).

Supplementary Table 1. Characteristics of samples included in the MINERvA cohort. * = primary characteristics used to match cases and controls. ** = secondary characteristics used to match cases and controls as closely as possible.

Supplementary Table 2. Eighteen birth weight associated DMPs were identified at a threshold of P < 1x10-7. Genomic locations are based on hg19.

Supplementary Table 3. 4,299 gestational age associated DMPs were identified at a threshold of P < 1x10-7. Genomic locations are based on hg19.

Supplementary Table 4. Conditional analysis within 3,550 distinct loci associated with gestational age.

Supplementary Table 5. 110 maternal smoking associated DMPs were identified at a threshold of P < 1x10-7. Genomic locations are based on hg19.

Supplementary Table 6. Conditional analysis within 70 distinct loci associated with maternal smoking.

Supplementary Table 7. 143 DMPs associated with maternal smoking selected for mediation analysis.

Supplementary Table 8. SIMEX adjusted mediation analysis for 3 DNA methylation sites identified as having significant mediation effects between maternal smoking and birth weight.