

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Sarina Abrishamcar

Date

**DNA Methylation as a Potential Mediator of the Association Between Prenatal Tobacco
and Alcohol Exposure and Child Neurodevelopment in a South African Birth Cohort**

By

Sarina Abrishamcar

Master of Public Health

Epidemiology

_____ [Chair's signature]

Anke Hüls, PhD

Committee Chair

**DNA Methylation as a Potential Mediator of the Association Between Prenatal Tobacco
and Alcohol Exposure and Child Neurodevelopment in a South African Birth Cohort**

By

Sarina Abrishamcar

Bachelor of Science

University of California, Santa Barbara

2020

Thesis Committee Chair: Anke Hüls, PhD

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Public Health

in Epidemiology

2022

Abstract

DNA Methylation as a Potential Mediator of the Association Between Prenatal Tobacco and Alcohol Exposure and Child Neurodevelopment in a South African Birth Cohort

By Sarina Abrishamcar

Prenatal tobacco exposure (PTE) and prenatal alcohol exposure (PAE) have been associated with an increased risk of delayed neurodevelopment in children as well as differential newborn DNA methylation (DNAm). However, the biological mechanisms connecting PTE and PAE, DNAm, and neurodevelopment are largely unknown. Here we aim to determine whether differential DNAm mediates the association between PTE and PAE and neurodevelopment at 6 (N=113) and 24 months (N=187) in children from the South African Drakenstein Child Health Study.

PTE and PAE were assessed antenatally using urine cotinine measurements and the ASSIST questionnaire, respectively. Cord blood DNAm was measured using the EPIC and 450K BeadChips. Neurodevelopment (cognitive, language, motor, adaptive behavior, socioemotional) was measured using the Bayley Scale of Infant and Toddler Development, Third Edition. We constructed methylation risk scores (MRS) for PTE and PAE and conducted causal mediation analysis (CMA) with these MRS as mediators. Next, we conducted a high-dimensional mediation analysis to identify individual CpG sites as potential mediators, followed by a CMA to estimate the average causal mediation effects (ACME) and total effect (TE).

PTE and PAE were associated with neurodevelopment at 6 but not at 24 months. PTE MRS reached a prediction accuracy (R^2) of 0.23 but did not significantly mediate the association between PTE and neurodevelopment. PAE MRS was not predictive of PAE ($R^2=0.006$). For PTE, 31 CpG sites and 8 CpG sites were identified as significant mediators (ACME and TE $P<0.05$) for the cognitive and motor domains at 6 months, respectively. For PAE, 16 CpG sites and 1 CpG site were significant mediators for the motor and adaptive behavior domains at 6 months, respectively. Several genes including *MAD1L1*, *CAMTA1*, and *ALDH1A2* have been implicated in neurodevelopmental delay, suggesting that differential DNAm may partly explain the biological mechanisms underlying the relationship between PTE and PAE and child neurodevelopment.

**DNA Methylation as a Potential Mediator of the Association Between Prenatal Tobacco
and Alcohol Exposure and Child Neurodevelopment in a South African Birth Cohort**

By

Sarina Abrishamcar

Bachelor of Science

University of California, Santa Barbara

2020

Thesis Committee Chair: Anke Hüls, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology

2022

Acknowledgements

First and foremost, I would like to thank my thesis advisor and mentor, Prof. Anke Hüls. Your guidance and mentorship these past two years have been incredibly important to me and changed the trajectory of my career. You truly care about your students and trainees, not only by preparing us to work at the highest caliber of research and public health, but also ensuring that we maintain work-life balance and take care of ourselves. I have always felt supported by you in every aspect and for that I am incredibly grateful.

Second, I would like to thank my co-authors who provided me with indispensable support and guidance and whom without, this project could not be completed: Junyu Chen, Dakotah Feil, Anna Kilanowski, Dr. Nastassja Koen, Dr. Aneesa Vanker, Dr. Kirsty Donald, Dr. Heather Zar, Dr. Dan Stein, and Dr. Anke Hüls. I would also like to thank the Drakenstein Child Health Study, our collaborators at University of Cape Town, and the families, mothers, and children that took part in this study.

Third, I would like to thank my previous mentors, for without their guidance and support I would not be where I am today: Prof. Douglas Thrower (UCSB), Dr. Mackenzie Wildman (Evidation Health), Dr. Luca Foschini (Evidation Health), Prof. Samantha Daly (UCSB), and Rae Hoyt (American Association of University Women). Thank you for believing in me and giving me a chance.

Finally, I would like to thank my family and friends. My parents, for immigrating from a war-ridden country with very little, in hopes of providing me with the opportunities that they did not have and for supporting me through all my endeavors. To Craig, for being at my side through the ups and downs of the last two years. And to my closest friends Lynn, Amanda, and Crystal who never stopped believing in me, even when I stopped believing in myself. To all, thank you for the unconditional love and support you all have given me.

Table of Contents

Introduction	1
Materials and Methods	3
Results	8
Discussion	10
References	18
Tables and Figures	28

Introduction

Pregnancy is a critical time window for neurodevelopment and a time when the fetus is most susceptible to adverse environmental and prenatal exposures¹. It is well established that prenatal tobacco exposure (PTE) and prenatal alcohol exposure (PAE) during pregnancy are associated with adverse fetal and childhood outcomes. Maternal smoking and alcohol consumption during pregnancy have been associated with adverse fetal outcomes such as low birth weight, sudden infant death syndrome (SIDS), and preterm birth²⁻⁵.

PTE and PAE have also been linked to the etiology of a range of neurodevelopmental disorders in children. For example, PTE has been previously associated with attention deficit hyperactivity disorder (ADHD)⁶. There is evidence that the risk of ADHD increases with higher concentrations of PTE⁷. There is also evidence that maternal smoking during pregnancy is a risk factor for neurological disorders such as schizophrenia (SZ) and bipolar disorder (BPD)^{8,9}. Additionally, PAE is one of the leading causes of fetal alcohol spectrum disorder (FASD), a form of severe neurodevelopmental delay that has various manifestations such as growth deficiencies and cognitive deficits^{10,11}.

While PTE and PAE have been implicated in the etiology of several neurodevelopmental disorders mentioned previously, the biological mechanisms have yet to be established.

Epigenetic modifications, such as differential DNA methylation (DNAm) are dynamic and highly sensitive to external environmental factors¹². More importantly, DNAm is potentially reversible, indicating that the methylome could be a therapeutic target for disease treatment and prevention¹³.

There is strong evidence from large-scale epigenome-wide association studies (EWAS) suggesting that PTE is associated with changes in the cord blood methylome¹⁴⁻¹⁶. The evidence is weaker for an association between PAE and the cord blood methylome¹⁷, although there has been some evidence of epigenomic changes in other tissues such as placental or buccal tissue¹⁸. While there is evidence that PTE and PAE influence DNA methylation in cord blood, not much is known about the subsequent adverse outcomes, such as delayed neurodevelopment^{19,20} or about the association between newborn DNAm and child neurodevelopment or related disorders in general²¹⁻²³. Thus, the biological mechanisms connecting PTE and PAE, DNA methylation, and child neurodevelopment are largely unknown.

The majority of the EWAS mentioned above have been conducted in high-income countries (HICs) and/or in children of primarily European ancestry. Despite the rapid proliferation of the field of genomics and translational science in the last decade, cohorts from low- and middle-income countries (LMICs) are still largely underrepresented²⁴. Subsequently, there is a large gap in the literature on these underrepresented populations in which risk factors such as low socioeconomic status, poor maternal health, and maternal substance use, and adverse child health outcomes including neurodevelopmental delay are at a much higher prevalence²⁵.

Here, we investigated whether differential DNAm mediates the relationship between PTE and PAE and neurodevelopment in children of 6 and 24 months of age using data from a well-characterized South African birth cohort, the Drakenstein Child Health Study (DCHS). For the mediation analysis, we utilized methylation risk scores (MRS) and high-dimensional mediation analysis (HDMA) to increase statistical power and reduce the dimensionality of the methylation data. To our knowledge, this study is the first mediation analysis to examine this relationship and is also one of the first to utilize HDMA and MRS for epigenetic mediation analysis.

Materials and Methods

Study population

The DCHS is a South African, population-based birth cohort of Black African and Mixed ancestry and has been previously described²⁶. Our analysis sample consisted of 266 children from the DCHS with DNAm data from cord blood, genotype data, PTE and PAE measures, and important covariates. Of these, 113 infants had measurements available from the Bayley Scales of Infant and Toddler Development, third edition (BSID-III) at 6 months of age, and 184 had measurements available from BSID-III at 24 months of age. Mothers were enrolled during their second trimester and followed through pregnancy at two primary clinics, TC Newman or Mbekweni. Mother-child pairs were then followed from birth until the child was at least 5 years of age. Ethical approval for human subjects' research was given by the Human Research Ethics Committee of the Faculty of Health Sciences of the University of Cape Town. Written consent for participation was obtained from the mother on behalf of herself and the infant.

DNA methylation measurements

As described previously, DNA methylation was measured from cord blood using the Illumina Infinium HumanMethylation450 BeadChip (450K; n=119) and Infinium MethylationEPIC BeadChip (EPIC; n=147) arrays²³. Pre-processing was conducted separately for each array in R 3.5.1, using an identical pre-processing pipeline. The 450K and EPIC arrays contained 426,378 probes and 781,536 probes, respectively. The 450k and EPIC arrays were combined using the *minfi* package, which resulted in 316 samples and 453,093 probes available in both arrays²⁷. Background subtraction, color correction, and normalization were performed using the *preprocessFunnorm* function²⁸. After sample and probe filtering, 273 samples and 409,033 probes remained for analysis. Of these samples, there were 266 complete cases with genotype

data, PTE and PAE measures, and important covariates. Batch effects were removed using the *ComBat* function from the *sva* package²⁹. Cell type composition estimates were calculated using the most recent cord blood reference dataset³⁰.

Smoking and alcohol measurements

Prenatal tobacco exposure was objectively measured using urine cotinine levels, which were taken within four weeks of enrollment. Urine cotinine was measured using the IMMULITE® 1000 Nicotine Metabolite Kit (Siemens Medical Solutions Diagnostics®, Glyn Rhonwy, United Kingdom)³¹. Urine cotinine levels were classified as non-smoker or passive smoker (<499 ng/ml) or active smoker (≥ 500 ng/ml). The continuous urine cotinine concentrations were used in this analysis.

Prenatal alcohol exposure was measured at the second antenatal study visit, using a dichotomous, composite score calculated from the Alcohol, Smoking, and Substance Involvement Screening Test (ASSIST), a self-report questionnaire, and from retrospectively collected data on alcohol consumption during pregnancy³². The ASSIST questionnaire was developed by the World Health Organization and has been validated for use in international settings^{33,34}. Details of the composite score calculation have been described elsewhere^{32,35,36}.

Neurodevelopment measurements

Neurodevelopment was assessed at 6 and 24 months of age using the BSID-III³⁷. The BSID-III assessment and its composite scores have been previously validated in South African populations^{38,39}. Trained assessors administered the BSID-III via direct observation to generate scores for the cognitive, language, and motor development domains. The same trained assessors administered the BSID-III to generate a score for the adaptive behavior and socioemotional domains³².

Composite scores generated for each domain were used in this analysis.

Statistical analysis

First, we estimated the association between PTE and PAE and neurodevelopment at age 6 and 24 months in adjusted linear regression analysis. Mediation analyses with differential DNAm as mediators (MRS and individual CpG sites) were conducted for any exposure-outcome combinations for which we identified indications of a total effect. Significant mediators were cross-validated for neurodevelopment measured at the second time point.

Confounding Assessment

Confounding was assessed by constructing directed acyclic graphs (DAGs). DAGs were created for each exposure-mediator (E-M) (PTE/PAE-DNAm), mediator-outcome (M-O) (DNAm-neurodevelopment), and exposure-outcome (E-O) relationship (PTE/PAE-neurodevelopment) (Figure S1). Potential confounders were selected based on existing literature³². All models in this analysis were adjusted for maternal age, maternal HIV status, maternal depression, maternal psychological distress, parental socioeconomic status (SES), gestational age, and cell-type proportions. Population stratification was adjusted for using the first five genetic principal components. Models evaluating PTE as the primary exposure were additionally adjusted for PAE and likewise, models evaluating PAE as the primary exposure were additionally adjusted for PTE.

Methylation Risk Scores

First, we constructed methylation risk scores (MRS) for PTE and PAE as described elsewhere⁴⁰. All 266 samples with DNAm data were included in the MRS calculation. MRS calculations were based on summary statistics from epigenome-wide association studies (EWAS) for DNAm probes in which the association between the exposure and DNAm was investigated. These external weights were used to calculate the methylation risk score in our cohort, described as the

weighted sum of methylation beta levels of the CpG sites. External summary statistics were acquired from EWAS conducted by the Pregnancy and Childhood Epigenetics (PACE) consortium. PTE summary statistics were acquired from an EWAS investigating the association between maternal smoking and cord-blood DNAm (N=5648) ¹⁶PAE summary statistics were acquired from an EWAS investigating the association between maternal alcohol use and cord-blood DNAm (N=1147) ¹⁷. To correct for correlated CpG sites, we estimated co-methylated regions (CMR) using the *CoMeBack* package ⁴¹. Then, we conducted “clumping”, in which one CpG site with the smallest p-value from the external summary statistics per CMR was included in the final MRS. MRS was calculated at several p-value thresholds for each EWAS, a procedure similar to “thresholding” in traditional polygenic risk score calculations. Finally, we calculated the correlation between the exposure and the resulting MRS at different p-value thresholds and chose the p-value threshold associated with the highest prediction accuracy for the subsequent analyses.

High-Dimensional Mediation Analyses (HDMA)

MRS are limited in that they only consider the E-M but not the M-O relationship. HDMA considers the E-M as well as the M-O relationship, which allows for a more holistic analysis of the mediating relationships. We performed two separate high-dimensional mediation analyses using the R packages *HIMA* and *DACt* ^{42,43}. The *HIMA* package utilizes a penalized-based regression and consists of three steps. First, it performs dimension reduction through sure independence screening to identify the $n/\log(n)$ CpG sites with the largest effect size in the mediator-outcome regression model. Second, the minimax concave penalty is applied to this subset of CpG sites for further dimension reduction ⁴⁴. Finally, a joint significance test is conducted to evaluate the significance of the mediation effects using a Benjamini-Hochberg FDR

correction for multiple testing. The divide-aggregate composite null test (DACT) is a more recent method for high-dimensional mediation analysis that is supposed to be better powered than HIMA⁴³. First, it utilizes the Efron empirical null framework to estimate the proportions of the three null cases across all epigenome mediators⁴⁵. Then, it performs the DACT for the composite null of no mediation effect in three cases in which the M-O effect is non-zero, the E-M effect is non-zero, or both effects are non-zero. Finally, the DACT p-value is calculated as the weighted sum of all three P values under the three null-hypothesis. We conducted an association analysis for the E-M and M-O associations with robust linear regression using the *rlm* function to input into DACT⁴⁶. We then pre-filtered CpG sites by two criteria to alleviate the burden of multiple testing. First, we filtered the CpG sites by a p-value threshold of 0.05 for the E-M and M-O models. Second, we filtered CpG sites by the direction of the E-M and M-O effects to achieve an overall negative indirect effect, in line with the hypothesized negative total and direct effects. After running DACT for the resulting subset of CpGs, we corrected for multiple testing using the Benjamini-Hochberg FDR method.

Causal Mediation Analysis (CMA)

Finally, we conducted a CMA for the MRS and significant CpG sites from HDMA as mediators. The *mediation* R package was used to estimate mediation effects⁴⁷. The *mediation* package estimates the average causal mediation effect (ACME), the average direct effect (ADE), the total effect (TE), and the proportion mediated (PM) in the population. 95% confidence intervals were constructed using a nonparametric, bootstrapped, quasi-Bayesian method⁴⁸.

Results

Study population characteristics

This analysis sample included 266 children with DNAm data, genotype data, and other relevant covariates, with a subset of 113 children and 184 children with information on neurodevelopment across 5 domains, respectively (Table 1). This is a cohort of children from Black African (55%) and Mixed (45%) ancestry. There was a high prevalence of maternal smoking with 47% of mothers classified as passive smokers and 30% classified as active smokers. The prevalence of maternal alcohol use among mothers was 17%. Additionally, 25% of mothers were classified as being above the Beck Depression Inventory Threshold and 30% being at elevated risk of psychological distress. The prevalence of mothers with an HIV diagnosis was 24%, which was not passed to the fetus in any case. Among the subset of children with information on neurodevelopment at 6 months of age, the prevalence of maternal smoking and maternal alcohol use was higher compared to the whole population and children at 24 months, with 35% of mothers classified as active smokers and 23% of mothers classified as alcohol consumers.

The total effect of prenatal smoking and alcohol exposure on child neurodevelopment

The total effects of PTE and PAE on child neurodevelopment across 5 domains were estimated using adjusted linear regression (Figure 1). There was a consistent negative association between PTE and neurodevelopment across all domains at 6 months of age, which was significant for the cognitive domain (-0.02; 95% CI: -0.028, 0.002; $P= 0.023$). There was also a consistent negative association between PAE and neurodevelopment across all domains at 6 months of age, which was significant for the motor domain (-9.36; 95% CI: -16.64, -2.08; $P= 0.012$). No associations were found between PTE and PAE and neurodevelopment at 24 months (Figure S2; Table S1-

S2). Therefore, the subsequent mediation analyses focused on neurodevelopment at 6 months as the primary outcome. Neurodevelopment at 24 months was used as validation for any significant mediators identified at age 6 months.

Mediation analyses with methylation risk scores as mediators

The MRS for PTE was well correlated with prenatal cotinine levels, with the highest R^2 being 0.23 at a p-value threshold of $5e-21$ (Figure 2A). As evident from Figure 2B, active smokers had a higher MRS than non-smokers and passive smokers combined. Additionally, there was a strong and significant association between PTE and the MRS (0.0022; 95% CI: 0.0016, 0.0027; $P=6.7E-15$) (Figure 2C). In contrast, the MRS for PAE was not correlated with alcohol use status, with the highest R^2 being 0.006 at a p-value threshold of $5e-06$ (Figure 2D). There was no difference between the MRS for non-alcohol users and alcohol users (Figure 2E) and there was no association between PAE and the MRS ($5.4e-03$; 95% CI: -0.26,0.27; $P=0.97$) (Figure 2F). Because of the poor performance of the PAE methylation risk score, we did not proceed with CMA for prenatal alcohol exposure using this approach.

There was no evidence of mediation by MRS in any domain (Figure 3; Table S3). While the cognitive domain showed a significant total effect, a significant indirect effect (ACME) was not found (ACME=0.006; 95% CI: -3.84E-03, 0.002; $P=0.2$). The direct effect was significant, however (ADE=-0.02; 95% CI: -0.04, -0.01; $P=0.01$).

HDMA with individual CpG sites as mediators

After correcting for multiple testing, DACT identified a total of 381 CpG sites as significant mediators of the association between PTE and neurodevelopment across the cognitive (123 CpG sites), language (94 CpG sites), motor (68 CpG sites), adaptive behavior (61 CpG sites), and socioemotional (35 CpG sites). After validating these CpG sites with CMA, we identified 31

CpG sites to have a significant mediating effect (ACME P and TE $P < 0.05$) between PTE and cognitive development and 8 CpG sites to have a significant mediating effect between PTE and motor development (Table 2). Three of these CpG sites overlapped between both the cognitive and motor domains (cg22263591 [*SLC39A11*], cg25284031 [*HVCN1*], cg25857569 [*PPM1L*]).

DACT identified 395 CpG sites as significant mediators of the association between PAE and neurodevelopment across the cognitive (79 CpG sites), language (74 CpG sites), motor (100 CpG sites), adaptive behavior (93 CpG sites), and socioemotional (49 CpG sites) domains. After conducting CMA, this resulted in 16 CpG sites and 1 CpG site with a significant mediating effect between PAE and the motor and adaptive behavior domains, respectively (Table 3).

After correction for multiple testing, HIMA did not identify any CpG sites as potential mediators of the association between either PTE and PAE and neurodevelopment at 6 months of age.

None of the CpG sites identified as significant mediators for neurodevelopment at 6 months mediated the association between PTE or PAE and neurodevelopment at 24 months (Table S4-S5).

Discussion

In this South African cohort, we found a significant, negative association (total effect) between PTE and PAE and neurodevelopment at 6 months of age as well as evidence of epigenetic mediation of this association using high-dimensional mediation analysis (DACT). We identified 39 CpGs that mediated the association between PTE and child cognitive (31 CpGs) and motor (8 CpGs) development. We also identified 16 CpG sites and 1 CpG site that mediated the association between PAE and child motor and adaptive behavior development, respectively.

While the PTE MRS was highly predictive of PTE, it did not mediate the association between

PTE and neurodevelopment. We did not find any evidence of epigenetic mediation of PTE and PAE and neurodevelopment in children of 24 months of age, most likely due to the absence of a total effect of PTE and PAE on neurodevelopment at 24 months.

Several studies have investigated the association between PTE and PAE and neurodevelopment and related disorders⁴⁹⁻⁵¹. One study (N=446) conducted in a rural region of China in children with an average age of 15 months found that prenatal exposure to tobacco smoke was associated with neurodevelopmental delay in the cognitive and language domains, assessed using the BSID-III⁵². While we did find evidence of an association between PTE and cognitive development, associations with language development were not statistically significant. Language may be more difficult to assess using the BSID-III in the DCHS due to cultural differences and multiple language groups, which may decrease the sensitivity of the tool^{39,53}. Prenatal alcohol exposure has been previously associated with decreased gross motor function, but findings have been inconsistent^{54,55}. For example, a study (N=1324) that utilized the BSID-III to evaluate the relationship between PAE and gross motor development in infants of 12 months did not find evidence of an association⁵⁶. However, the prevalence of moderate to severe PAE was lower in this Australian cohort than in the DCHS. In our study, we found associations between PTE and PAE and neurodevelopment at age 6 months but not at age 24 months. Previous literature on neurodevelopment at 6 months of age is sparse, particularly in association with prenatal exposures. Most studies had an average infant age of at least 12 months. Unlike our present analysis, previous studies have found associations between PTE and PAE and neurodevelopment at or around 24 months of age or older^{50,51,57}. For example, a study conducted in a Polish birth cohort (N=461) identified a statistically significant association between prenatal tobacco smoke exposure and delayed development in the cognitive, language, and motor domains at 24 months

of age using the BSID-III⁵⁸. A potential explanation for this discrepancy is our relatively small sample size (N=113 at age 6 months and N=184 at age 24 months). Several of the CpG sites and associated genes that we identified through HDMA are well known for their association with sustained maternal smoking during pregnancy across two major studies from PACE⁵⁹.

Differential methylation at one of the CpG sites (cg23219570) and four genes (*MAD1L1*, *CAMTA1*, *LARP4B*, *FGF23*) that were identified as mediators for associations with cognitive development was previously identified in one PACE study which investigated the association of sustained prenatal smoking on changes in cord blood DNAm (450K) across 9 cohorts (N=5648)¹⁶. Differential DNAm at 2 CpG sites (cg02668773, cg12031962) and 7 genes (*MAD1L1*, *MPO*, *CAMTA1*, *COL11A2*, *LARP4B*, *ALDH1A2*, *SLC39A11*) in association with cognitive development and 1 CpG site (cg24671734) and 4 genes (*SKAP2*, *ZNF710*, *SLC39A11*, *BTBD11*) in association with motor development which we identified as mediators were previously identified in a more recent PACE study, which investigated the effects of sustained prenatal smoking on cord blood DNAm (450K) across 13 cohorts (N=6685)¹⁴. It is of note that the genes *MAD1L1*, *CAMTA1*, and *LARP4B* had significantly, differentially methylated CpG sites across both PACE studies and our study.

Four of these genes have further been shown to be associated with cognitive outcomes. Aldehyde Dehydrogenase 1A2 (*ALDH1A2*) is associated with autism spectrum disorder (ASD) in children⁶⁰. The *ALDH* family of genes plays a key role in the metabolism of vitamin A, an essential molecule for neuronal differentiation and development⁶¹. However, there has been inconsistent evidence of an association between maternal smoking and ASD⁶²⁻⁶⁴. Furthermore, brain tissue-based DNAm in two of these genes (*MAD1L1* and *COL11A1*) were associated with ASD, indicating their role in neurodevelopment⁶⁵.

Mitotic arrest deficient 1 like 1 (*MAD1L1*) is a key regulatory gene of the cell cycle. It has been implicated in the etiology of neurodevelopmental disorders such as SZ and BPD in several genome-wide association studies (GWAS) and candidate gene studies, including one study which found altered levels of expression of *MAD1L1* in human neural progenitor cells⁶⁶⁻⁶⁸. The biological mechanisms of both BPD and SZ are linked to changes in the mesolimbic reward pathway and high levels of *MAD1L1* expression have been found in brain regions such as the ventral tegmental area⁶⁹. Previous studies have found that maternal smoking was associated with a higher rate of SZ and BPD, which suggests that differential expression of *MAD1L1* may be a plausible pathway that links smoking to neurodevelopmental disorders⁷⁰.

Calmodulin-binding transcription activator 1 (*CAMTA1*) is also involved in cell-cycle regulation and is highly expressed in the brain⁷¹. Studies have found that *CAMTA1* plays a role in memory production and recall, dysfunction of which is implicated in disorders such as down syndrome, schizophrenia, and depression^{72,73}. However, these studies have been conducted in older adults, and more research needs to be done to corroborate memory dysfunction in children with differential expression of *CAMTA1*.

Two genes of note that we identified as significant mediators of the association between PAE and motor development are *SLC18A2* and *HNRNPU*⁷⁴. Homozygous mutations in *SLC18A2*, a vesicular monoamine transporter, have been implicated in the presentation of infantile movement disorder also known as infantile hypotonic parkinsonian disorder or brain dopamine-serotonin vesicular transport disease⁷⁵⁻⁷⁷. Mutations in *HNRNPU* have been associated with infantile spasms^{78,79}. However, there has been little to no research published on prenatal and early life exposures or pediatric movement disorders and related epigenetic modifications.

Methylation risk scores can be used as a potential biomarker for prenatal smoking and alcohol use and to reduce the dimensionality of methylation data, giving us more statistical power to detect associations⁴⁰. Although our methylation risk score for prenatal smoking was highly predictive of PTE, it did not significantly mediate the association between PTE and neurodevelopment. One possible explanation is that our MRS algorithm only considers the relationship between exposure and mediator, whereas HDMA considers both the E-M and M-O relationships. Additionally, it is possible that the difference in ethnicity between the external studies (primarily European ancestry) and the DCHS (Black African and Mixed ancestry) reduced the prediction accuracy of our MRS for both prenatal smoking and prenatal alcohol use⁸⁰.

Our study has a few limitations. First, unlike PTE which has been extensively studied^{14–16,81} in association with DNAm, the association between PAE and DNAm is not well understood, which makes the comparison of our findings to existing literature more challenging. The vast majority of epigenetic studies evaluating the effects of prenatal alcohol use are conducted in birth cohorts from HICs and tend to have a low prevalence of prenatal alcohol use^{17,59}. Because of the low prevalence of the exposure, these studies often do not have the statistical power to detect an association. A PACE study that investigated the association between PAE and cord blood DNAm (450K) across 6 cohorts (N=1147) did not find strong evidence of a relationship¹⁷. This limitation also in part explains the poor performance of our methylation risk score for PAE.

Another limitation to our study is the possible presence of exposure misclassification bias. Underreporting is a common problem when attempting to ascertain socially undesirable behaviors such as smoking or alcohol consumption during pregnancy. In our case, alcohol use was measured using a self-report questionnaire and smoking was measured via urine cotinine

levels. Although using cotinine as a biomarker of smoking is not affected by social desirability bias, it was only measured at one time point in this cohort. Exposure misclassification can be an issue in epigenomic mediation analyses in that the DNA methylation estimates may be a more accurate indicator of exposure than the variables available in our cohort, which can lead to an overestimation of the mediation effect⁸².

Furthermore, while blood is the most common tissue used to evaluate the epigenomic mechanisms of a multitude of risk factors and outcomes, it is not yet clear whether blood tissue can be effectively used to evaluate neurological outcomes. Epigenetic signatures are tissue and cell-type specific, and as such, the selection of relevant tissues and cell-type is crucial for epigenetic studies. While brain tissue would be the most relevant tissue to study in association with neurodevelopment, cord blood has the advantage in that it is more easily accessible from living samples and could therefore be linked to future cognitive outcomes⁸³. Nevertheless, epigenetic signatures in cord blood may represent systematic and mechanistic changes through various peripheral pathways or may show blood-brain concordance, and thus are important biomarkers of neurological disease⁸⁴.

Our study has several strengths. To start, we have data at 6 months of age, which may allow us to infer a causal relationship more confidently because of the short temporality between birth and measurement. For example, a recent study from the PACE consortium that investigated the effects of DNA methylation at birth on childhood cognitive skills found weak evidence of an association at 4-9 years of age²¹. We also were not able to detect any evidence of differential DNAm as a mediator of this relationship at 24 months of age. Likely, the numerous, unmeasured, environmental, social, and behavioral factors that influence cognitive development and neurodevelopment in general from birth to 9 years of age likely obscured any associations

with newborn DNAm. Additionally, as Caramaschi et al. mentioned in their paper, cognitive skills may be affected through epigenetic changes by exposures such as prenatal smoking, and thus a mediation analysis such as ours may be more appropriate to elucidate a potential causal pathway²¹.

Although this analysis had a small sample size, the use of novel methods to increase our statistical power allowed us to detect more potentially meaningful associations and is an important strength. To our knowledge, this is one of the first epigenomic studies to employ HDMA. The HIMA method, which utilizes the joint significance test, was potentially too conservative and underpowered to detect significant mediators in this study⁴³. DACT has been shown to be more powered than HIMA. The use of methylation risk scores and high-dimensional mediation analysis in the field is still in its infancy. We recognize that more studies will be required to validate the robustness of these methods as the field advances, but this study provides some insight into the utility and potential of HDMA in real-life settings.

Finally, our analysis was conducted on a subset of a well-characterized, underrepresented, birth cohort from an LMIC. We were able to detect associations due to the high prevalence of behavioral risk factors - such as maternal smoking and alcohol consumption- in such vulnerable populations, a characteristic that cohorts from HICs typically do not exhibit. This highlights the importance of conducting large-scale multi-ethnic cohort studies to advance the field of (epi)genomics and reduce health disparities. As Wojcik et al. point out, while the burden of disease disproportionality lies with minority populations, the majority of studies are conducted in European populations, which limits the field's ability to investigate the burden of disease in vulnerable communities⁸⁵. As follows, investigating epigenomic associations with health

outcomes in a well-characterized cohort from an LMIC allows us to add to the sparse literature and shed light on the importance of prioritizing cohorts like the DCHS.

Overall, these findings suggest that epigenetic changes at the methylation level may in part explain the biological mechanisms underlying the relationship between maternal smoking and alcohol consumption during pregnancy and child neurodevelopment across the cognitive, motor, and adaptive behavior domains. Our study provides motivation to conduct larger mediation studies to replicate our findings. Mediation analyses in epigenomics are important for discerning causal mechanisms from exposure to disease. The identification of significant CpG sites could provide novel insights for the early detection of disease and potential prevention targets in translational research and community interventions in at-risk populations.

References

1. Fitzgerald, E., Hor, K. & Drake, A. J. Maternal influences on fetal brain development: The role of nutrition, infection and stress, and the potential for intergenerational consequences. *Early Human Development* **150**, 105190 (2020).
2. Wang, X. *et al.* Maternal Cigarette Smoking, Metabolic Gene Polymorphism, and Infant Birth Weight. *The Journal of the American Medical Association* **287**, 195–202 (2002).
3. Zhang, K. & Wang, X. Maternal smoking and increased risk of sudden infant death syndrome: A meta-analysis. *Legal Medicine* **15**, 115–121 (2013).
4. Kesmodel, U., Wisborg, K., Olsen, S. F., Henriksen, T. B. & Secher, N. J. Moderate Alcohol Intake during Pregnancy and the Risk of Stillbirth and Death in the First Year of Life. *American Journal of Epidemiology* **155**, 305–312 (2002).
5. Addila, A. E., Azale, T., Gete, Y. K. & Yitayal, M. The effects of maternal alcohol consumption during pregnancy on adverse fetal outcomes among pregnant women attending antenatal care at public health facilities in Gondar town, Northwest Ethiopia: a prospective cohort study. *Substance Abuse: Treatment, Prevention, and Policy* **16**, 1–15 (2021).
6. Marceau, K. *et al.* Within-family effects of smoking during pregnancy on ADHD: the importance of phenotype. *Journal of Abnormal Child Psychology* **46**, 685–699 (2018).
7. Huang, L. *et al.* Maternal Smoking and Attention- Deficit / Hyperactivity Disorder in Offspring : A Meta-analysis. *Pediatrics* **141**, 1–11 (2018).
8. Niemelä, S. *et al.* Prenatal Nicotine Exposure and Risk of Schizophrenia Among Offspring in a National Birth Cohort. *Am J Psychiatry* **173**, 799–806 (2016).

9. Talati, A. *et al.* Maternal smoking during pregnancy and bipolar disorder in offspring. *Am J Psychiatry* **170**, 1178–1185 (2013).
10. May, P. A. *et al.* Maternal alcohol consumption producing fetal alcohol spectrum disorders (FASD): Quantity, frequency, and timing of drinking. *Drug and Alcohol Dependence* **133**, 502–512 (2013).
11. Roozen, S. *et al.* Systematic literature review on which maternal alcohol behaviours are related to fetal alcohol spectrum disorders (FASD). *BMJ Open* **8**, (2018).
12. Feinberg, A. P. Phenotypic plasticity and the epigenetics of human disease. *Nature* **447**, 433–440 (2007).
13. Moosavi, A. & Ardekani, A. M. Role of epigenetics in biology and human diseases. *Iranian Biomedical Journal* **20**, 246–258 (2016).
14. Joubert, B. R. *et al.* DNA Methylation in Newborns and Maternal Smoking in Pregnancy : Genome-wide Consortium Meta-analysis. *The American Journal of Human Genetics* 680–696 (2016) doi:10.1016/j.ajhg.2016.02.019.
15. Richmond, R. C. *et al.* Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: Findings from the Avon Longitudinal Study of Parents and Children (ALSPAC). *Human Molecular Genetics* **24**, 2201–2217 (2015).
16. Sikdar, S. *et al.* Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking. *Epigenomics* **11**, 1487–1500 (2019).
17. Sharp, G. C. *et al.* Maternal alcohol consumption and offspring DNA methylation: Findings from six general population-based birth cohorts. *Epigenomics* **10**, 27–42 (2018).

18. Loke, Y. J. *et al.* Sex- And tissue-specific effects of binge-level prenatal alcohol consumption on DNA methylation at birth. *Epigenomics* **13**, 1921–1938 (2021).
19. Xu, R. *et al.* DNA methylation mediates the effect of maternal smoking on offspring birthweight : a birth cohort study of multi - ethnic US mother – newborn pairs. *Clinical Epigenetics* 1–13 (2021) doi:10.1186/s13148-021-01032-6.
20. Küpers, L. K. *et al.* DNA methylation mediates the effect of maternal smoking during pregnancy on birthweight of the offspring. *International Journal of Epidemiology* **44**, 1224–1237 (2015).
21. Caramaschi, D. *et al.* Meta-analysis of epigenome-wide associations between DNA methylation at birth and childhood cognitive skills. *Molecular Psychiatry* (2022) doi:10.1038/s41380-022-01441-w.
22. Neumann, A. *et al.* Association between DNA methylation and ADHD symptoms from birth to school age : a prospective meta-analysis. *Translational Psychiatry* **10**, (2020).
23. Hüls, A. *et al.* Newborn differential DNA methylation and subcortical brain volumes as early signs of severe neurodevelopmental delay in a South African Birth Cohort Study. *World Journal of Biological Psychiatry* **0**, 1–12 (2022).
24. Tekola-Ayele, F. & Rotimi, C. N. Translational Genomics in Low- and Middle-Income Countries: Opportunities and Challenges. *Public Health Genomics* **18**, 242–247 (2015).
25. Stein, D. J. *et al.* Investigating the psychosocial determinants of child health in Africa: The Drakenstein Child Health Study. *Journal of Neuroscience Methods* **252**, 27–35 (2015).

26. Zar, H. J., Barnett, W., Myer, L., Stein, D. J. & Nicol, M. P. Investigating the early-life determinants of illness in Africa: The Drakenstein Child Health Study. *Thorax* **70**, 592–594 (2015).
27. Aryee, M. J. *et al.* Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
28. Fortin, J.-P. *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *bioRxiv* 002956 (2014) doi:10.1101/002956.
29. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
30. Gervin, K. *et al.* Systematic evaluation and validation of reference and library selection methods for deconvolution of cord blood DNA methylation data. *Clinical Epigenetics* **11**, 1–15 (2019).
31. Vanker, A. *et al.* Home environment and indoor air pollution exposure in an African birth cohort study. *Science of the Total Environment* **536**, 362–367 (2015).
32. Donald, K. A. *et al.* Risk and protective factors for child development: An observational South African birth cohort. *PLoS Medicine* **16**, 1–20 (2019).
33. Humeniuk, R. *et al.* The Alcohol, Smoking and Substance involvement Screening Test (ASSIST): manual for use in primary care / prepared by R. HumeniukU [et al].
34. Humeniuk, R. *et al.* Validation of the alcohol, smoking and substance involvement screening test (ASSIST). *Addiction* **103**, 1039–1047 (2008).

35. Myers, B. *et al.* Effect of Hazardous Alcohol Use During Pregnancy on Growth Outcomes at Birth: Findings from a South African Cohort Study. *Alcoholism: Clinical and Experimental Research* **42**, 369–377 (2018).
36. Donald, K. A. *et al.* Alcohol exposure in utero is associated with decreased gray matter volume in neonates. *Metabolic Brain Disease* **31**, 81–91 (2016).
37. Bayley, N. Bayley Scales of Infant and Toddler Development- Third Edition. San Antonio, Tx: Harcourt Assessment. *Journal of Psychoeducational Assessment* (2006).
38. Rademeyer, V. & Jacklin, L. A study to evaluate the performance of black South African urban infants on the Bayley Scales of Infant Development III. *SAJCH South African Journal of Child Health* **7**, 54–59 (2013).
39. Ballot, D. E. *et al.* Use of the Bayley Scales of Infant and Toddler Development, Third Edition, to Assess Developmental Outcome in Infants and Young Children in an Urban Setting in South Africa. *International Scholarly Research Notices* **2017**, 1–5 (2017).
40. Hüls, A. & Czamara, D. Methodological challenges in constructing DNA methylation risk scores. *Epigenetics* **15**, 1–11 (2020).
41. Gatev, E., Gladish, N., Mostafavi, S. & Kobor, M. S. CoMeBack: DNA methylation array data analysis for co-methylated regions. *Bioinformatics* **36**, 2675–2683 (2020).
42. Zhang, H. *et al.* Genetics and population analysis: Estimating and testing high-dimensional mediation effects in epigenetic studies. *Bioinformatics* **32**, 3150–3154 (2016).
43. Liu, Z. *et al.* Large-Scale Hypothesis Testing for Causal Mediation Effects with Applications in Genome-wide Epigenetic Studies. *J Am Stat Assoc* **0**, 1–39 (2021).

44. Zhang, C. H. *Nearly unbiased variable selection under minimax concave penalty. Annals of Statistics* vol. 38 (2010).
45. Zeng, P., Shao, Z. & Zhou, X. Statistical methods for mediation analysis in the era of high-throughput genomics: Current successes and future challenges. *Computational and Structural Biotechnology Journal* **19**, 3209–3224 (2021).
46. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S*. (Springer, 2002).
47. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. Mediation: R package for causal mediation analysis. *Journal of Statistical Software* **59**, 1–38 (2014).
48. Imai, K., Keele, L. & Tingley, D. A General Approach to Causal Mediation Analysis. *Psychological Methods* **15**, 309–334 (2010).
49. Subramoney, S., Eastman, E., Adnams, C., Stein, D. J. & Donald, K. A. The Early Developmental Outcomes of Prenatal Alcohol Exposure: A Review. *Frontiers in Neurology* **9**, (2018).
50. Banderali, G. *et al.* Short and long term health effects of parental tobacco smoking during pregnancy and lactation: a descriptive review. *Journal of Translational Medicine* **13**, 327 (2015).
51. Clifford, A., Lang, L. & Chen, R. Effects of maternal cigarette smoking during pregnancy on cognitive parameters of children and young adults: A literature review. *Neurotoxicology and Teratology* **34**, 560–570 (2012).
52. He, Y., Luo, R., Wang, T., Gao, J. & Liu, C. Prenatal Exposure to Environmental Tobacco Smoke and Early Development of Children in Rural Guizhou Province, China. *Int J Environ Res Public Health* **15**, 2866 (2018).

53. Donald, K. A. *et al.* Drakenstein Child Health Study (DCHS): investigating determinants of early child development and cognition. *BMJ Paediatr Open* **2**, e000282–e000282 (2018).
54. Bay, B. & Kesmodel, U. S. Prenatal alcohol exposure - a systematic review of the effects on child motor function. *Acta Obstet Gynecol Scand* **90**, 210–226 (2011).
55. Lucas, B. R. *et al.* Gross motor deficits in children prenatally exposed to alcohol: a meta-analysis. *Pediatrics* **134**, e192-209 (2014).
56. Hutchinson, D. *et al.* Correction to: Prenatal alcohol exposure and infant gross motor development: A prospective cohort study (BMC Pediatrics (2019) 19 (149) DOI: 10.1186/s12887-019-1516-5). *BMC Pediatrics* **19**, (2019).
57. Julvez, J. *et al.* Maternal smoking habits and cognitive development of children at age 4 years in a population-based birth cohort. *Int J Epidemiol* **36**, 825–832 (2007).
58. Polanska, K. *et al.* Environmental tobacco smoke exposure during pregnancy and child neurodevelopment. *International Journal of Environmental Research and Public Health* **14**, (2017).
59. Felix, J. F. *et al.* Cohort profile: Pregnancy and childhood epigenetics (PACE) consortium. *International Journal of Epidemiology* **47**, 22–23u (2018).
60. Gomez-Fernandez, A. *et al.* Children with autism spectrum disorder with regression exhibit a different profile in plasma cytokines and adhesion molecules compared to children without such regression. *Frontiers in Pediatrics* **6**, 1–9 (2018).
61. Xu, X. *et al.* Excessive UBE3A dosage impairs retinoic acid signaling and synaptic plasticity in autism spectrum disorders. *Cell Research* **28**, 48–68 (2018).

62. von Ehrenstein, O. S., Cui, X., Yan, Q., Aralis, H. & Ritz, B. Maternal Prenatal Smoking and Autism Spectrum Disorder in Offspring: A California Statewide Cohort and Sibling Study. *Am J Epidemiol* **190**, 728–737 (2021).
63. Jung, Y., Lee, A. M., McKee, S. A. & Picciotto, M. R. Maternal smoking and autism spectrum disorder: Meta-analysis with population smoking metrics as moderators. *Scientific Reports* **7**, 1–10 (2017).
64. Caramaschi, D. *et al.* Maternal smoking during pregnancy and autism: using causal inference methods in a birth cohort study. *Translational Psychiatry* **8**, (2018).
65. Nardone, S. *et al.* DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways. *Translational Psychiatry* **4**, 1–9 (2014).
66. Liu, X. *et al.* MAD1L1 and TSNARE gene polymorphisms are associated with schizophrenia susceptibility in the Han Chinese population. *BMC Medical Genomics* **14**, 1–10 (2021).
67. Su, L. *et al.* Genetic association of GWAS-supported MAD1L1 gene polymorphism rs12666575 with schizophrenia susceptibility in a Chinese population. *Neuroscience Letters* **610**, 98–103 (2016).
68. Hill, M. J. *et al.* Transcriptional consequences of schizophrenia candidate miR-137 manipulation in human neural progenitor cells. *Schizophrenia Research* **153**, 225–230 (2014).
69. Trost, S. *et al.* Investigating the impact of a genome-wide supported bipolar risk variant of MAD1L1 on the human reward system. *Neuropsychopharmacology* **41**, 2679–2687 (2016).
70. Quinn, P. D. *et al.* Association between maternal smoking during pregnancy and severe mental illness in offspring. *JAMA Psychiatry* **74**, 589–596 (2017).

71. Bas-Orth, C., Tan, Y. W., Oliveira, A. M. M., Bengtson, C. P. & Bading, H. The calmodulin-binding transcription activator CAMTA1 is required for long-term memory formation in mice. *Learning and Memory* **23**, 313–321 (2016).
72. Ramanan, V. K. *et al.* FASTKD2 is associated with memory and hippocampal structure in older adults. *Molecular Psychiatry* **20**, 1197–1204 (2015).
73. Huentelman, M. J. *et al.* Calmodulin-binding transcription activator 1 (CAMTA1) alleles predispose human episodic memory performance. *Human Molecular Genetics* **16**, 1469–1477 (2007).
74. Yu, W., Gwinn, M., Clyne, M., Yesupriya, A. & Khoury, M. J. A navigator for human genome epidemiology. *Nature Genetics* **40**, 124–125 (2008).
75. Rilstone, J. J., Alkhater, R. A. & Minassian, B. A. Brain Dopamine–Serotonin Vesicular Transport Disease and Its Treatment. *New England Journal of Medicine* **368**, 543–550 (2013).
76. Zhai, H. *et al.* A case report of infantile parkinsonism–dystonia-2 caused by homozygous mutation in the SLC18A2 gene. *International Journal of Neuroscience* 1–4 (2021)
doi:10.1080/00207454.2021.1938036.
77. Jacobsen, J. C. *et al.* Brain dopamine-serotonin vesicular transport disease presenting as a severe infantile hypotonic parkinsonian disorder. *J Inherit Metab Dis* **39**, 305–308 (2016).
78. Du, X. *et al.* A genomic copy number variant analysis implicates the MBD5 and HNRNPU genes in Chinese children with infantile spasms and expands the clinical spectrum of 2q23.1 deletion. *BMC Medical Genetics* **15**, 1–12 (2014).
79. Allen, A. S. *et al.* De novo mutations in epileptic encephalopathies. *Nature* **501**, 217–221 (2013).

80. Heyn, H. *et al.* DNA methylation contributes to natural human variation. *Genome Research* **23**, 1363–1372 (2013).
81. Zhang, B. *et al.* Maternal smoking during pregnancy and cord blood DNA methylation: new insight on sex differences and effect modification by maternal folate levels. *Epigenetics* **13**, 505–518 (2018).
82. Valeri, L. *et al.* Misclassified exposure in epigenetic mediation analyses. Does DNA methylation mediate effects of smoking on birthweight? *Epigenomics* **9**, 253–265 (2017).
83. Bakulski, K. M., Halladay, A., Hu, V. W., Mill, J. & Fallin, M. D. Epigenetic Research in Neuropsychiatric Disorders: the “Tissue Issue.” *Current Behavioral Neuroscience Reports* **3**, 264–274 (2016).
84. Hannon, E., Lunnon, K., Schalkwyk, L. & Mill, J. Interindividual methylomic variation across blood, cortex, and cerebellum: Implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics* **10**, 1024–1032 (2015).
85. Wojcik, G. L. *et al.* Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019).

Tables and Figures

Figure 1	29
Figure 2	30
Figure 3	31
Table 1	32
Table 2	33
Table 3	34

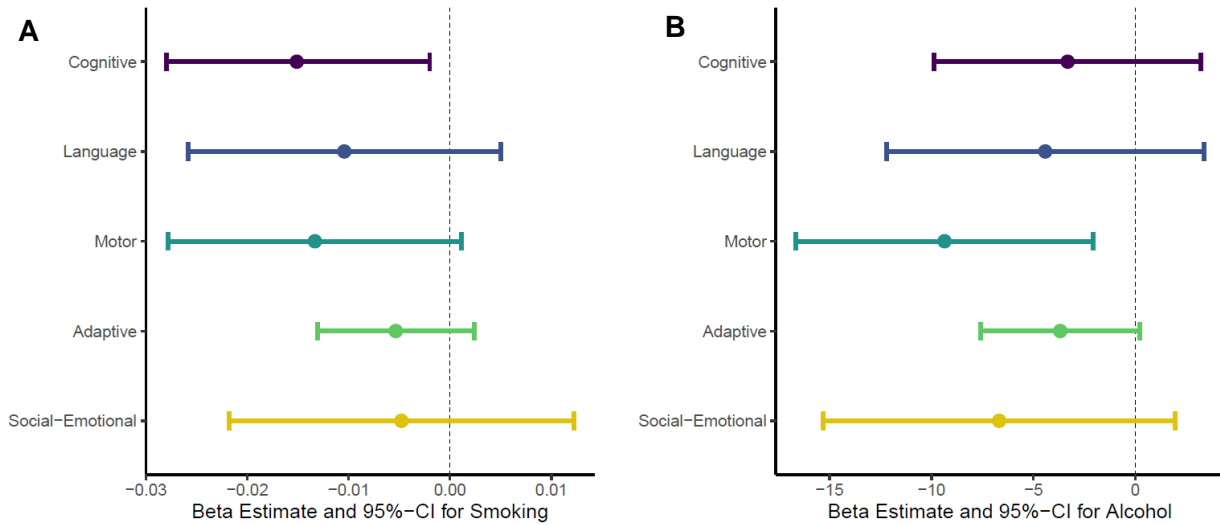


Figure 1: Association of PTE and PAE on neurodevelopment domains at 6 months. Models were adjusted for paternal SES, maternal depression, maternal psychological distress, gestational age, maternal age, maternal HIV status, cell-type proportions, and the first 5 genetic principal components. **1A)** Association of PTE on neurodevelopment domains, additionally adjusted for PAE. **1B)** Association of PAE on neurodevelopment domains, additionally adjusted for PTE.

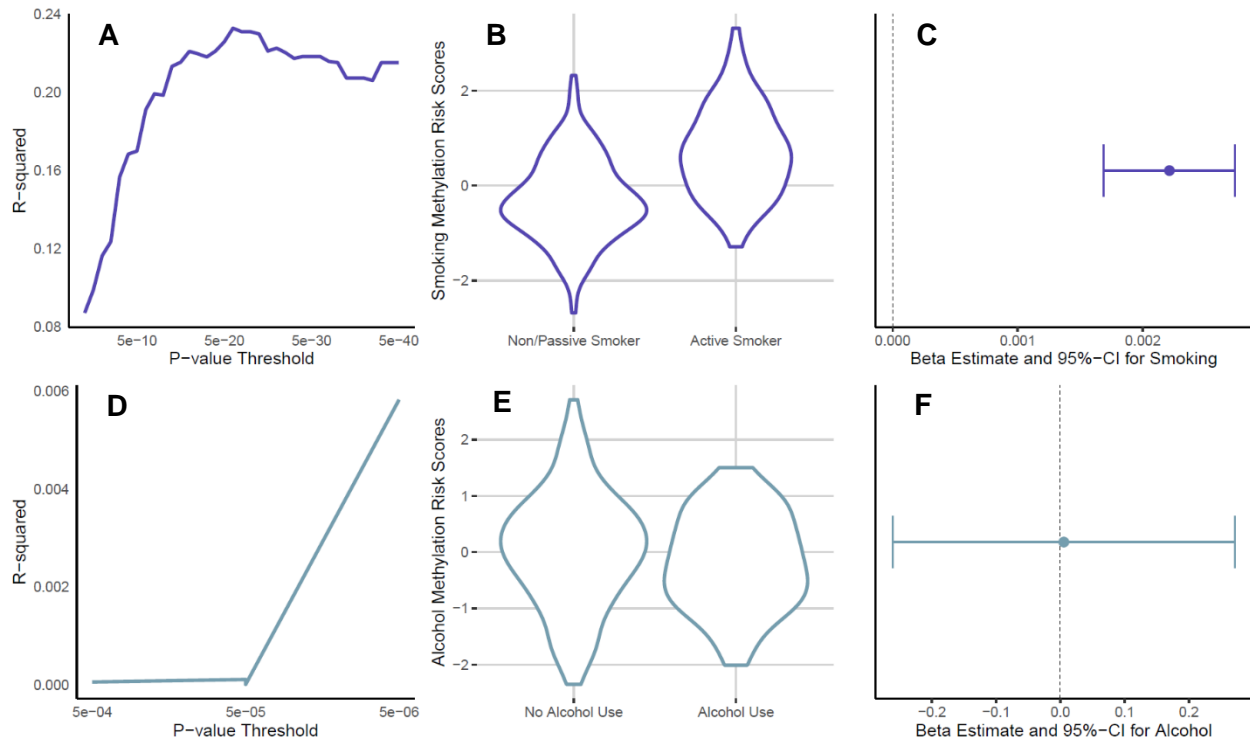


Figure 2: Performance of the MRS for PTE and PAE. 2A) Correlation between PTE and MRS at several p-value thresholds **2B)** MRS for non-/passive smoking mothers and active smoking mothers. **2C)** Effect of PTE on most highly correlated MRS. **2D)** Correlation between PAE and MRS at several p-value thresholds **2E)** MRS for non-alcohol consuming mothers and alcohol consuming mothers **2F)** Effect of PAE on most highly correlated MRS

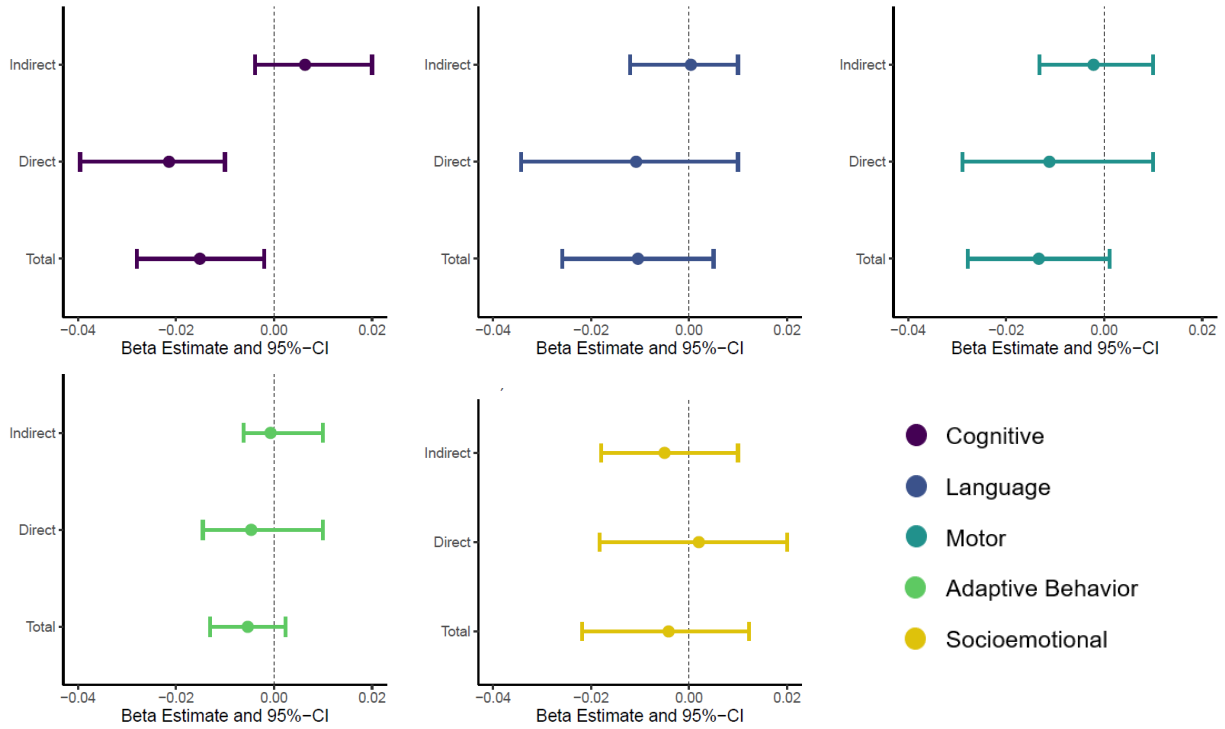


Figure 3: Causal mediation analysis for PTE as the exposure, MRS as the mediator, and neurodevelopment across 5 domains at 6 months of age as outcome. Estimated indirect effect, direct effect, and total effect with their corresponding 95% confidence intervals are shown.

Table 1: Study population characteristics for the whole sample, children with BSID-III measurements at 6 months of age, and children with BSID-III measurements at 24 months of age

	Whole study sample	Children with BSID-III data at 6 mo	Children with BSID-III data at 24 mo
N	266	113	184
Maternal Age (yrs) (median [IQR])	26 [22,31]	25 [22,31]	26 [22,32]
Gestational Age (wks) (median [IQR])	39 [38,40]	39 [38,40]	39 [38,40]
Child Sex (%)			
Female	119 (44.74)	54 (47.79)	82 (43.85)
Male	147 (55.26)	59 (52.21)	105 (56.14)
Parental SES¹ (%)			
Lowest SES	66 (24.81)	23 (20.35)	42 (22.46)
Low-moderate SES	65 (24.44)	33 (29.20)	47 (25.13)
Moderate-high SES	74 (27.82)	30 (26.55)	50 (26.74)
Highest SES	61 (22.93)	27 (23.89)	48 (25.67)
Ancestry (%)			
Black African	147 (55.26)	54 (47.79)	96 (51.34)
Mixed	119 (44.74)	59 (52.21)	91 (48.66)
Maternal Alcohol Use (%)			
Non-alcohol user	219 (82.33%)	87 (77.00)	151 (80.75)
Alcohol user	46 (17.29%)	26 (23.00)	35 (18.72)
Maternal Smoking² (%)			
Non-smoker	60 (22.56)	24 (21.24)	44 (23.53)
Passive Smoker	126 (47.37)	48 (42.48)	80 (42.78)
Active Smoker	79 (29.70)	40 (35.40)	63 (33.69)
Maternal Smoking (ng/ml) (median [IQR])	38.40 [11.4, 500.00]	57.35 [11.48, 500.00]	55.90 [10.80, 500.00]
Maternal HIV (%)			
Yes	64 (24.06)	29 (25.66)	43 (23.00)
No	202 (75.94)	84 (74.34)	144 (77.00)
Beck Depression Inventory Threshold (%)			
Above	67 (25.19)	36 (31.86)	48 (25.67)
Below	196 (73.68)	75 (66.37)	136 (72.73)
Self-Reporting Questionnaire (%)			
High Risk	80 (30.08)	34 (30.09)	58 (31.02)
Low Risk	184 (69.17)	78 (69.03)	127 (67.91)
Bayley Scales 6 months (median [IQR])			
Cognitive (n=113)	-	105 [95,110]	-
Language (n=109)	-	103 [90.5,112]	-
Motor (n=110)	-	112 [100,118]	-
Adaptive Behavior (n=112)	-	102 [96,107]	-
Social-Emotional (n=112)	-	115 [100,125]	-
Bayley Scales 24 months (median [IQR])			
Cognitive (n=184)	-	-	85 [80,90]
Language (n=173)	-	-	83 [74,91]
Motor (n=173)	-	-	91 [85,100]
Adaptive Behavior (n=184)	-	-	84 [71,92]
Social-Emotional (n=184)	-	-	115 [100,135]

¹Socioeconomic status (SES) composite scores (see Supplemental for details)

²Urine cotinine levels (ng/ml) classified as non-smoker (<10 ng/ml); passive smoker (10-499 ng/ml); active smoker (≥ 500 ng/ml)

Table 2: Significant CpG sites (ACME P and TE $P < 0.05$) resulting from HDMA and CMA of DNAm as a potential mediator of the association between PTE and neurodevelopmental delay at 6 months of age using the DACT method and after correction for multiple testing (BH FDR ≤ 0.05). ACME, ADE, TE, and PM effect sizes and 95% confidence intervals estimated by CMA.

Domain	CpG (Gene)	ACME (95% CI)	ADE (95% CI)	Total Effect (95% CI)	Prop Med. (95% CI)
Cognitive	cg00480142	-0.004 (-8.9e-03, -5.0e-05)	-0.011 (-0.025, 2.4e-03)	-0.015 (-0.027, -2.4e-03)	0.24 (-0.012, 1.4)
	cg00613752 (GPR6)	-0.004 (-9.2e-03, -1.9e-04)	-0.012 (-0.024, 4.2e-04)	-0.016 (-0.029, -3.4e-03)	0.25 (3.2e-03, 0.88)
	cg01258793 (MAD1L1)	-0.005 (-0.012, -3.4e-04)	-0.011 (-0.024, 2.3e-03)	-0.016 (-0.030, -2.7e-03)	0.31 (5.5e-03, 1.3)
	cg02215141	-0.004 (-9.6e-03, -5.5e-05)	-0.011 (-0.024, 2.5e-03)	-0.016 (-0.028, -2.5e-03)	0.28 (-0.022, 1.3)
	cg02423105 (CDH26)	-0.005 (-0.012, -6.0e-05)	-0.014 (-0.026, -1.2e-03)	-0.018 (-0.031, -4.8e-03)	0.26 (1.1e-03, 0.83)
	cg02668773 (MPO)	-0.007 (-0.014, -7.7e-04)	-0.012 (-0.025, 1.4e-03)	-0.019 (-0.033, -4.8e-03)	0.36 (0.05, 1.1)
	cg02879122 (CCDC144A)	-0.004 (-9.3e-03, -3.8e-04)	-0.012 (-0.025, 1.2e-03)	-0.016 (-0.028, -2.9e-03)	0.26 (0.018, 1.2)
	cg03288340 (ZC3H4)	-0.005 (-0.012, -6.3e-04)	-0.011 (-0.023, 2.1e-03)	-0.016 (-0.029, -3.2e-03)	0.33 (0.039, 1.3)
	cg05129050 (CAMTA1)	-0.005 (-0.011, -6.9e-04)	-0.010 (-0.023, 4.5e-03)	-0.014 (-0.027, -1.4e-03)	0.35 (9.1e-03, 2.0)
	cg05331340 (ASGR1)	-0.010 (-0.020, -1.8e-03)	-0.012 (-0.025, 7.9e-04)	-0.022 (-0.037, -7.6e-03)	0.44 (0.093, 1.0)
	cg08481075 (COL11A2)	-0.005 (-0.011, -2.9e-04)	-0.012 (-0.024, 1.9e-03)	-0.016 (-0.029, -2.9e-03)	0.28 (2.5e-03, 1.1)
	cg08836481 (LARP4B)	-0.008 (-0.016, -1.2e-03)	-0.012 (-0.025, 4.2e-04)	-0.020 (-0.034, -6.2e-03)	0.39 (0.08, 1.0)
	cg09115275 (IL2RA)	-0.004 (-8.3e-03, -2.2e-04)	-0.012 (-0.024, 1.5e-03)	-0.015 (-0.027, -2.3e-03)	0.23 (-0.015, 1.1)
	cg09964326 (FGFBP3)	-0.004 (-0.010, -4.5e-04)	-0.012 (-0.025, 1.2e-03)	-0.017 (-0.030, -3.3e-03)	0.27 (0.024, 1.2)
	cg10149329 (MFSD8)	-0.004 (-9.2e-03, -3.7e-04)	-0.011 (-0.024, 3.0e-03)	-0.015 (-0.027, -1.9e-03)	0.27 (-2.7e-03, 1.4)
	cg10780097 (AMD1)	-0.003 (-8.0e-03, -1.6e-05)	-0.012 (-0.024, -5.9e-04)	-0.015 (-0.028, -2.6e-03)	0.20 (-5.9e-03, 0.76)
	cg12031962 (ALDH1A2)	-0.004 (-0.010, -3.5e-04)	-0.011 (-0.024, 3.1e-03)	-0.016 (-0.029, -2.5e-03)	0.26 (0.015, 1.4)
	cg12428160	-0.005 (-0.014, -4.3e-04)	-0.013 (-0.025, 7.9e-04)	-0.018 (-0.031, -4.0e-03)	0.28 (0.025, 1.1)
	cg13889963 (CALB1)	-0.004 (-9.4e-03, -8.2e-05)	-0.011 (-0.024, 2.1e-03)	-0.015 (-0.028, -2.2e-03)	0.28 (-0.039, 1.2)
	cg13918314 (ANAPC13)	-0.004 (-9.8e-03, -2.1e-05)	-0.013 (-0.025, -6.1e-04)	-0.017 (-0.030, -3.4e-03)	0.22 (-5.5e-03, 0.84)
	cg14323109 (KDR)	-0.004 (-9.1e-03, -8.4e-05)	-0.011 (-0.023, 3.1e-03)	-0.015 (-0.027, -9.5e-04)	0.24 (-0.01, 1.6)
	cg14896134	-0.007 (-0.016, -6.0e-04)	-0.010 (-0.023, 2.0e-03)	-0.017 (-0.030, -3.8e-03)	0.40 (0.035, 1.2)
	cg15624376 (HOXA4)	-0.004 (-0.010, -2.2e-04)	-0.012 (-0.025, 7.9e-04)	-0.016 (-0.029, -2.6e-03)	0.25 (1.7e-03, 1.1)
	cg18780259 (GDF1)	-0.005 (-0.012, -5.1e-04)	-0.010 (-0.022, 2.8e-03)	-0.015 (-0.028, -2.3e-03)	0.34 (0.022, 1.3)
	cg19611886	-0.009 (-0.021, -8.2e-04)	-0.012 (-0.025, 1.6e-03)	-0.021 (-0.035, -6.2e-03)	0.42 (0.034, 1.1)
	cg22263591 (SLC39A11)	-0.006 (-0.014, -3.9e-04)	-0.013 (-0.025, -4.9e-04)	-0.019 (-0.031, -5.1e-03)	0.33 (0.022, 0.89)
	cg23219570 (FGF23)	-0.007 (-0.016, -4.5e-04)	-0.011 (-0.024, 1.5e-03)	-0.019 (-0.032, -5.6e-03)	0.40 (0.021, 1.1)
	cg23643526 (HAO2)	-0.004 (-9.5e-03, -1.3e-04)	-0.011 (-0.024, 3.0e-03)	-0.015 (-0.027, -2.2e-03)	0.25 (-0.022, 1.4)
	cg25284031 (HVCN1)	-0.006 (-0.015, -2.2e-04)	-0.014 (-0.026, -5.0e-04)	-0.019 (-0.033, -5.8e-03)	0.30 (9.9e-03, 0.91)
	cg25823085 (ABCB4)	-0.004 (-9.1e-03, -2.0e-04)	-0.011 (-0.023, 2.2e-03)	-0.014 (-0.027, -1.5e-03)	0.27 (-0.021, 1.2)
cg25857569 (PPM1L)	-0.007 (-0.015, -1.6e-04)	-0.012 (-0.025, 5.7e-05)	-0.019 (-0.033, -5.1e-03)	0.36 (5.5e-03, 0.94)	
Motor	cg00328376 (SKAP2)	-0.010 (-0.03, -4.9e-03)	-0.010 (-0.023, 9.0e-03)	-0.023 (-0.042, -4.6e-03)	0.65 (0.26, 2.1)
	cg09490277 (FANCE)	-0.010 (-0.024, -5.3e-04)	-0.011 (-0.026, 4.4e-03)	-0.022 (-0.043, -3.1e-03)	0.49 (0.019, 1.4)
	cg14989316 (ZMIZ1-AS1)	-0.008 (-0.017, -1.6e-03)	-0.010 (-0.025, 6.5e-03)	-0.018 (-0.035, -1.9e-04)	0.46 (-7.2e-03, 2.9)
	cg15188623 (ZNF710)	-0.008 (-0.021, -6.8e-04)	-0.010 (-0.025, 6.0e-03)	-0.018 (-0.037, -8.3e-04)	0.46 (-0.015, 2.1)
	cg22263591 (SLC39A11)	-0.006 (-0.014, -2.8e-05)	-0.011 (-0.026, 4.8e-03)	-0.017 (-0.032, -3.6e-04)	0.35 (-0.074, 2.0)
	cg24671734 (BTBD11)	-0.007 (-0.019, -6.7e-04)	-0.011 (-0.027, 6.1e-03)	-0.018 (-0.034, -1.9e-03)	0.41 (-0.014, 2.0)
	cg25284031 (HVCN1)	-0.007 (-0.017, -3.1e-04)	-0.012 (-0.026, 4.6e-03)	-0.018 (-0.035, -1.9e-03)	0.36 (-0.037, 1.8)
	cg25857569 (PPM1L)	-0.010 (-0.023, -2.9e-03)	-0.010 (-0.024, 6.8e-03)	-0.020 (-0.037, -2.6e-03)	0.57 (0.17, 1.9)

Table 3: Significant CpG sites (ACME P and TE $P < 0.05$) resulting from HDMA and CMA of DNAm as a potential mediator of the association between PAE and neurodevelopmental delay at 6 months of age using the DACT method and after correction for multiple testing (BH FDR ≤ 0.05). ACME, ADE, TE, PM and corresponding 95% confidence intervals estimated by CMA shown.

Domain	CpG (Gene)	ACME (95% CI)	ADE (95% CI)	Total Effect (95% CI)	Prop Med. (95% CI)
Motor	cg00086871 (TBC1D14)	-1.8 (-4.4, -0.07)	-8.1 (-15.4, -0.83)	-9.9 (-17.1, -2.9)	0.18 (4.6e-03, 0.69)
	cg00512279 (SLC18A2)	-2.5 (-6.1, -0.21)	-7.3 (-14.5, 0.46)	-9.8 (-16.7, -2.9)	0.25 (0.02, 1.1)
	cg02230180 (HNRNPU)	-2.6 (-6.1, -0.14)	-7.4 (-14.8, 0.49)	-10.0 (-16.8, -2.9)	0.26 (0.02, 1.1)
	cg03113572 (ZNF331)	-3.5 (-7.4, -0.59)	-5.9 (-12.7, 1.0)	-9.3 (-16.5, -2.2)	0.37 (0.06, 1.3)
	cg04743947	-2.7 (-6.7, -0.03)	-6.7 (-14.3, 0.85)	-9.4 (-16.5, -2.6)	0.29 (2.8e-03, 1.2)
	cg04875987 (LRRC45)	-3.0 (-6.9, -0.29)	-6.3 (-13.5, 1.1)	-9.2 (-16.2, -2.2)	0.32 (0.03, 1.2)
	cg05141711 (AGMAT)	-2.8 (-6.7, -0.28)	-6.9 (-13.7, -0.11)	-9.7 (-16.5, -2.9)	0.29 (0.03, 0.9)
	cg09528884 (PTEN)	-4.0 (-8.2, -1.30)	-6.7 (-13.5, 5.3e-03)	-10.6 (-17.7, -3.8)	0.37 (0.13, 0.97)
	cg10137084 (CACNA2D3)	-3.3 (-7.9, -0.13)	-6.6 (-13.2, 0.99)	-9.8 (-16.5, -2.8)	0.33 (5.6e-03, 1.3)
	cg12653570	-2.7 (-6.5, -0.19)	-7.1 (-13.5, -0.49)	-9.8 (-16.9, -2.9)	0.27 (0.02, 0.78)
	cg12789343	-2.0 (-4.8, -0.04)	-8.0 (-15.1, -1.2)	-10.0 (-16.8, -3.1)	0.20 (5.9e-03, 0.64)
	cg19340420 (RCN3)	-2.9 (-6.1, -0.49)	-7.4 (-14.8, -0.64)	-10.3 (-17.1, -3.3)	0.28 (0.05, 0.86)
	cg22804663 (TDG)	-3.0 (-6.4, -0.09)	-6.6 (-13.9, -0.04)	-9.7 (-16.9, -2.5)	0.31 (-3.1e-03, 0.96)
	cg23238147 (TMEM51-AS1)	-2.3 (-5.4, -0.11)	-6.8 (-14.0, 0.23)	-9.2 (-16.1, -2.3)	0.25 (3.0e-03, 1.0)
	cg24716343 (PCNX2)	-2.8 (-6.2, -0.30)	-6.2 (-12.8, 1.20)	-9.0 (-16.0, -1.9)	0.31 (0.04, 1.3)
	cg26848718 (WT1)	-3.4 (-8.3, -0.23)	-5.4 (-11.9, 1.30)	-8.8 (-15.7, -2.0)	0.39 (0.02, 1.3)
Adaptive	cg21553656 (DAG1)	-2.1 (-4.4, -0.35)	-2.3 (-5.98, 1.90)	-4.3 (-8.2, -0.3)	0.48 (0.03, 2.0)