ARTICLE   OPEN

Check for updates

DNA methylation as a potential mediator of the association between prenatal tobacco and alcohol exposure and child neurodevelopment in a South African birth cohort

Sarina Abrishamcar[1], Junyu Chen[1], Dakotah Feil[1], Anna Kilanowski [2,3,4], Nastassja Koen[5,6,7], Aneesa Vanker[8], Catherine J. Wedderburn [5,8,9], Kirsten A. Donald [5,8], Heather J. Zar[6,8], Dan J. Stein [5,6,7] and Anke Hüls [1,10 ✉]

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 ($N = 112$) and 24 months ($N = 184$) 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 450 K BeadChips. Neurodevelopment (cognitive, language, motor, adaptive behavior, socioemotional) was measured using the Bayley Scales 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 eight 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 of the associated 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.

INTRODUCTION

Pregnancy is a critical time window for neurodevelopment and a time when the fetus is most susceptible to adverse environmental and prenatal exposures [1]. 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 birth [2–5].

PTE and PAE have also been linked to the etiology of a range of neurodevelopmental disorders in children. For example, PTE has previously been associated with attention deficit hyperactivity disorder (ADHD) [6]. There is evidence that the risk of ADHD

increases with higher concentrations of PTE [7]. There is also evidence that maternal smoking during pregnancy is a risk factor for neuropsychiatric disorders such as schizophrenia (SZ) and bipolar disorder (BPD) [8, 9]. PAE is the cause of fetal alcohol spectrum disorder (FASD), a condition characterized by severe neurodevelopmental delay, and which has various manifestations, including growth deficiencies and both behavioral and cognitive deficits [10, 11].

While PTE and PAE have been implicated in the etiology of several neurodevelopmental disorders, the biological mechanisms have yet to be established. Epigenetic modifications, such as differential DNA methylation (DNAm), are dynamic and highly sensitive to external environmental factors [12]. More importantly, DNAm is potentially

reversible, indicating that the methylome could be a therapeutic target for disease treatment and prevention [13].

There is strong evidence from large-scale epigenome-wide association studies (EWAS) suggesting that PTE is associated with changes in the cord blood methylome [14–16]. The evidence is weaker for an association between PAE and the cord blood methylome [17], although there has been some evidence of epigenomic changes in other tissues such as placental or buccal tissue [18]. 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 general [21–23]. Few studies have investigated the mediation of PTE and neuropsychiatric outcomes by DNAm. A candidate gene study found that differential cord blood DNAm levels in GFI1 acts as a mediator of the association between PTE and ADHD in 6-year-old children [24]. There is also evidence of whole blood DNAm mediating the association between PTE and schizophrenia, as indicated by a meta-analysis of five prospective birth cohorts [25]. While there is evidence of DNAm as a mediator of this relationship with psychiatric outcomes, the epigenetic mechanisms connecting PTE and PAE, cord blood DNAm, and child neurodevelopment are largely unknown.

The majority of EWAS 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 non-European ancestries and low- and middle-income countries (LMICs) are still largely underrepresented [26]. Consequently, there is a large gap in the literature on these underrepresented populations in which risk factors such as low socioeconomic status, poor maternal health, maternal substance use, and adverse child health outcomes, including neurodevelopmental delay, are at a much higher prevalence [27].

Here, we investigated whether differential DNAm in cord blood 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) [28]. For the causal mediation analysis (CMA), 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 one of the first mediation analyses to examine this relationship in an underrepresented, at-risk, LMIC population, 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 African and mixed ancestry and has been previously described [28]. Our analysis sample consisted of 262 children from the DCHS with DNAm data from cord blood, genotype data, PTE and PAE measures, and important covariates. Of these,112 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, Newman and 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 either the Illumina Infinium HumanMethylation450 BeadChip (450 K;

$n = 115$) or the Infinium MethylationEPIC BeadChip (EPIC; $n = 147$) arrays [23]. Preprocessing was conducted separately for each array in R 3.5.1, using an identical preprocessing pipeline. The 450 K 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 arrays [29]. Background subtraction, color correction, and normalization were performed using the preprocessFunnorm function [30]. After sample and probe filtering, 273 samples and 409,033 probes remained for analysis. Of these samples, there were 262 complete cases with genotype data, PTE and PAE measures, and important covariates. Batch effects were removed using the ComBat function from the sva package [31]. Cell type composition estimates were calculated using the most recent cord blood reference dataset [32].

Smoking and alcohol measurements
Prenatal tobacco exposure was objectively measured using urine cotinine levels, which were taken within 4 weeks of enrollment. Urine cotinine was measured using the IMMULITE® 1000 Nicotine Metabolite Kit (Siemens Medical Solutions Diagnostics®, Glyn Rhonwy, United Kingdom) [33]. 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 [34].

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 pregnancy [35]. The ASSIST questionnaire was developed by the World Health Organization and has been validated for use in international settings [36, 37]. Details of the composite score calculation have been described elsewhere [35, 38, 39].

Neurodevelopment measurements
Neurodevelopment was assessed at 6 and 24 months of age using the BSID-III [40, 41]. The BSID-III assessment and its composite scores have been previously validated in South African populations [42, 43]. Trained assessors administered the BSID-III to children via direct observation to generate scores for the cognitive, language, and motor development domains. The same trained assessors administered the BSID-III to mothers to report and generate a score for the adaptive behavior and socio-emotional domains [35]. 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 were conducted using MRS and individual CpG sites as mediators. Significant mediators identified at 6 months of age were validated for neurodevelopment measured at the second timepoint of 24 months.

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) (Fig. S1). Potential confounders were selected based on existing literature [35]. All models in this analysis were adjusted for maternal age, maternal HIV status, maternal depression during pregnancy (assessed with the Beck Depression Inventory II (BDI-II)), maternal psychological distress during pregnancy (assessed with the Self Reporting Questionnaire (SRQ-20)), 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. Multicollinearity between covariates was evaluated by calculating the variance inflation factors for each predictor using the total effect regression model. There was no evidence of multicollinearity for PTE and PAE (Table S6).

Methylation risk scores. First, we constructed methylation risk scores (MRS) for PTE and PAE as described elsewhere [44, 45]. All 262 samples with DNAm data were included in the MRS calculation. MRS are calculated

as the weighted sum of methylation beta levels of CpG sites. 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 summary statistics, which describe the effect size (beta) at each CpG site in the respective EWAS, were used as external weights to calculate the methylation risk score in our cohort. External summary statistics were acquired from EWASs 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 ($N = 5648$) [16]. PAE summary statistics were acquired from an EWAS investigating the association between maternal alcohol use and cord-blood DNAm ($N = 1147$) [17]. Both EWASs were comprised of cohorts from primarily European ancestry. To correct for correlated CpG sites, we estimated co-methylated regions (CMR) using the CoMeBack package [46]. 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 for any exposure-outcome combinations for which we identified indications of a total effect [47, 48]. 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 [49]. Finally, a joint significance test is conducted to evaluate the significance of the mediation effects using a Benjamini–Hochberg false discovery rate (FDR) correction for multiple testing. The divide-aggregate composite null test (DACT) is a more recent method for high-dimensional mediation analysis that is better powered than HIMA [48]. First, it utilizes the Efron empirical null framework to estimate the proportions of the three null cases across all epigenome mediators [50]. 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 DACT [51]. 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 effects [52]. 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. The 95% confidence intervals were constructed using a nonparametric, bootstrapped, quasi-Bayesian method and two-sided p values were provided for each effect. An α level of 0.05 was used to determine significance [53]. To examine trends across both time points, we cross-validated significant CpG sites identified at 6 months by determining whether they are also significant mediators of the association at 24 months of age.

Secondary analyses
To further support our findings, we conducted follow-up analyses for any CpG sites that were identified as significant mediators. This included gene ontology (GO) functional enrichment analysis using the missMethyl R package [54], blood–brain concordance analysis using the Blood–Brain

Epigenetic Concordance (BECon) tool [55], and methylation quantitative trait loci (mQTL) mapping using the GoDMC database [56]. Details are provided in the supplementary methods. Additionally, we compared our findings to previous studies using the publicly available EWAS catalog [57] and conducted further searches on PubMed.

RESULTS
Study population characteristics
This analysis sample included 262 children with DNAm data, genotype data, and other relevant covariates, with a subset of 112 children and 184 children with information on neurodevelopment across five domains at 6 months and 24 months, respectively (Table 1). The sample size at each timepoint varied slightly by neurodevelopmental domain and timepoint (Table 1). This is a cohort of children from African (55%) and mixed (45%) ancestry. There was a high prevalence of maternal smoking with 48% of mothers classified as passive smokers and 30% classified as active smokers. The prevalence of prenatal alcohol use among mothers was 17%. Additionally, 25% of mothers were classified as being above the Beck Depression Inventory-II threshold and 30% being at elevated risk of psychological distress on the Self-Reporting Questionnaire (SRQ-20). The prevalence of mothers with an HIV diagnosis was 24%, but all children in this population remained uninfected. Among the subset of children with information on neurodevelopment at 6 months of age, the prevalence of mothers who smoked and reported alcohol use was higher compared to the whole population with DNAm data and children at 24 months, with 35% of mothers classified as active smokers and 22% of mothers classified as consuming alcohol.

The total effect of prenatal smoking and alcohol exposure on child neurodevelopment
The total effects of PTE and PAE on child neurodevelopment across five domains were estimated using adjusted linear regression (Fig. 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 (β = −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 (β = −9.36; 95% CI: −16.64, −2.08; $P = 0.012$). No associations were found between PTE and PAE and neurodevelopment at 24 months in this sample (Fig. S2 and Tables 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 (Fig. 2A). As evident from Fig. 2B, active smokers had a higher MRS than non-smokers and passive smokers combined (β = 0.96, 95% CI: 0.72, 1.21; $P = 4.01E-13$). Additionally, there was a strong and significant association between PTE (continuous maternal urine cotinine levels) and the MRS (β = 0.0022; 95% CI: 0.0016, 0.0027; $P = 6.7E-15$) (Fig. 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 (Fig. 2D). There was no difference between the MRS for children with and without alcohol exposure (Fig. 2E) and there was no association between PAE and the MRS (β = 5.4e-03; 95% CI: −0.26, 0.27; $P = 0.97$) (Fig. 2F). Because of the poor performance of the PAE methylation risk score, we did not proceed with CMA for PAE using this approach.

There was no evidence of mediation by MRS in any domain (Fig. 3 and Table S3). While the cognitive domain showed a significant total effect, a significant indirect effect (ACME) was not

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	262	112	184
Maternal age (yrs) (median [IQR])	26 [22,31]	26 [22,31]	26 [22,32]
Gestational age (wks) (median [IQR])	39 [38,40]	39 [38,40]	39 [38,40]
Child sex (%)			
Female	116 (44.27)	54 (48.21)	79 (42.93)
Male	146 (55.73)	58 (51.79)	105 (57.07)
Parental SES[a] (%)			
Lowest SES	66 (25.19)	23 (20.53)	42 (22.83)
Low-moderate SES	64 (24.43)	32 (28.57)	46 (25.00)
Moderate-high SES	73 (27.86)	30 (26.79)	50 (27.17)
Highest SES	59 (22.52)	27 (24.11)	46 (25.00)
Ancestry (%)			
Black African	143 (54.58)	52 (46.43)	93 (50.54)
Mixed	119 (45.42)	60 (53.57)	91 (49.46)
Maternal alcohol use (%)			
Non-alcohol user	217 (82.82%)	87 (77.68)	150 (81.52)
Alcohol user	45 (17.18%)	25 (22.32)	34 (18.48)
Maternal smoking[b] (%)			
Non-smoker	59 (22.52)	23 (20.53)	43 (23.37)
Passive Smoker	125 (47.71)	50 (44.64)	79 (42.94)
Active Smoker	78 (29.77)	39 (34.82)	62 (33.70)
Maternal smoking (ng/ml) (median [IQR])	38.20 [11.4, 500.00]	57.35 [12.32, 500.00]	54.10 [10.90, 500.00]
Maternal HIV (%)			
Yes	64 (24.42)	29 (25.89)	43 (23.00)
No	198 (75.57)	83 (74.11)	144 (77.00)
Maternal depression during pregnancy[c] (%)			
Above threshold	66 (25.19)	36 (32.14)	48 (26.08)
Below threshold	196 (74.81)	76 (67.86)	136 (73.91)
Maternal psychological distress during pregnancy[d] (%)			
Above threshold	79 (30.15)	34 (30.36)	57 (30.98)
Below threshold	183 (69.85)	78 (69.64)	127 (69.02)
Bayley scales 6 months (median [IQR])			
Cognitive (n = 110)	-	105 [95,110]	-
Language (n = 109)	-	103 [91,112]	-
Motor (n = 110)	-	112 [100,118]	-
Adaptive behavior (n = 112)	-	102 [96,107]	-
Social-emotional (n = 112)	-	115 [100,125]	-

Table 1. continued

	Whole study sample	Children with BSID-III data at 6 mo	Children with BSID-III data at 24 mo
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]

[a]Socioeconomic status (SES) composite scores (see Supplemental for details).
[b]Urine cotinine levels (ng/ml) classified as non-smoker (<10 ng/ml); passive smoker (10–499 ng/ml); active smoker (≥500 ng/ml).
[c]Assessed with the Beck Depression Inventory II (BDI-II); threshold score of at least 20 indicated moderate depression.
[d]Assessed with the Self-Reporting Questionnaire (SRQ-20); threshold of at least 8 indicated high risk of psychological distress.

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
HDMA was conducted for exposure-outcome combinations for which there was an indication of a total effect: PTE-cognitive, PTE-motor, PAE-motor, and PAE-adaptive behavior. After correcting for multiple testing, DACT identified a total of 191 CpG sites as significant mediators of the association between PTE and neurodevelopment across the cognitive (123 CpG sites) and motor (68 CpG sites) domains. 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 eight 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], and cg25857569 [PPM1L]).

DACT identified 193 CpG sites as significant mediators of the association between PAE and neurodevelopment across the motor (100 CpG sites) and adaptive behavior (93 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).

The second HDMA method we tested, 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 after correction for multiple testing.

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 (Tables S4, S5).

Secondary analyses
After correction for multiple testing (FDR <0.05), we did not identify any GO terms or KEGG pathways with an overrepresentation of genes containing significantly, differentially methylated CpGs that would indicate an enriched biological pathway. The top GO terms and KEGG pathways for each E-O model are included in the supplement (Table S7a, b).
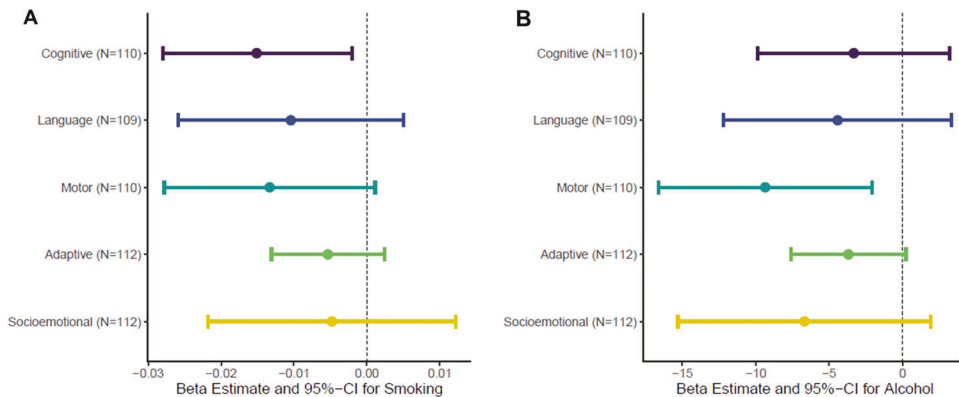
Fig. 1 Association of PTE and PAE on neurodevelopment domains at 6 months. Models were adjusted for parental SES, maternal depression, maternal psychological distress, gestational age, maternal age, maternal HIV status, cell-type proportions, and the first five genetic principal components. **A** Association of PTE on neurodevelopment domains, additionally adjusted for PAE. **B** Association of PAE on neurodevelopment domains, additionally adjusted for PTE.

Of the 31 CpG sites identified as significant mediators of the association between PTE and cognitive development, 19 were associated with at least one mQTL (Table S8). Seven CpG sites exhibited blood–brain concordance in accordance with the criteria listed in the methods (Table S9). CpG sites associated with the genes *ALDH1A2* (75–90%), *GPR6* (75–90%), and *HVCN1* (90%) exhibited the highest percentile mean correlations.

Of the eight CpG sites identified as significant mediators of the association between PTE and motor development, six were associated with at least one mQTL (Table S8). Here, six CpG sites exhibited blood–brain concordance, with the highest percentile mean correlations seen at CpG sites associated with the genes *FANCE* (75–90%), *HVCN1* (90%), *ZMIZ1-AS1* (75–90%), and *ZNF710* (90%) (Table S9).

For the 16 CpG sites identified as significant mediators of the association between PAE and motor development, six were associated with an mQTL. No CpG sites exhibited blood–brain concordance for the association between PAE and motor development (Table S9).

There were no mQTLs associated with the CpG site identified as a significant mediator of the association between PAE and adaptive behavior development (Table S8). Additionally, the probe did not exhibit blood–brain concordance (Table S9).

DISCUSSION
In this South African cohort, we found a significant negative association (total effect) between PTE and PAE and neurodevelopment of exposed children 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 (eight CpGs) development. We also identified 16 CpG sites and one 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 in this sample.

Several studies have investigated the association between PTE and PAE and neurodevelopment and related disorders [58–60]. 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-III [61]. While we did find evidence of an association between PTE and cognitive development, associations with language development were not statistically significant, although it may be difficult to detect language differences as early as 6 months [43, 62]. Prenatal alcohol exposure has been previously associated with decreased gross motor function, but findings have been inconsistent [63, 64]. 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 association [65]. However, the prevalence of moderate to severe PAE was lower in this 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. This is in line with findings from a previous study conducted in a larger subset of the DCHS ($N = 734$), in which PAE was associated with delayed motor development at 6 months, but not at 24 months [66]. 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 [59, 60, 67].

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-III [68]. A potential explanation for this discrepancy is our relatively small sample size ($N = 112$ 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 the PACE consortium [69]. 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 (450 K) across nine cohorts ($N = 5648$) [16]. Differential DNAm at two CpG sites (cg02668773 and cg12031962) and seven genes (*MAD1L1, MPO, CAMTA1, COL11A2, LARP4B, ALDH1A2,* and *SLC39A11*) in association with cognitive development and one CpG site (cg24671734) and four genes (*SKAP2, ZNF710, SLC39A11,* and *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 (450 K) across 13 cohorts ($N = 6685$) [14]. It is notable
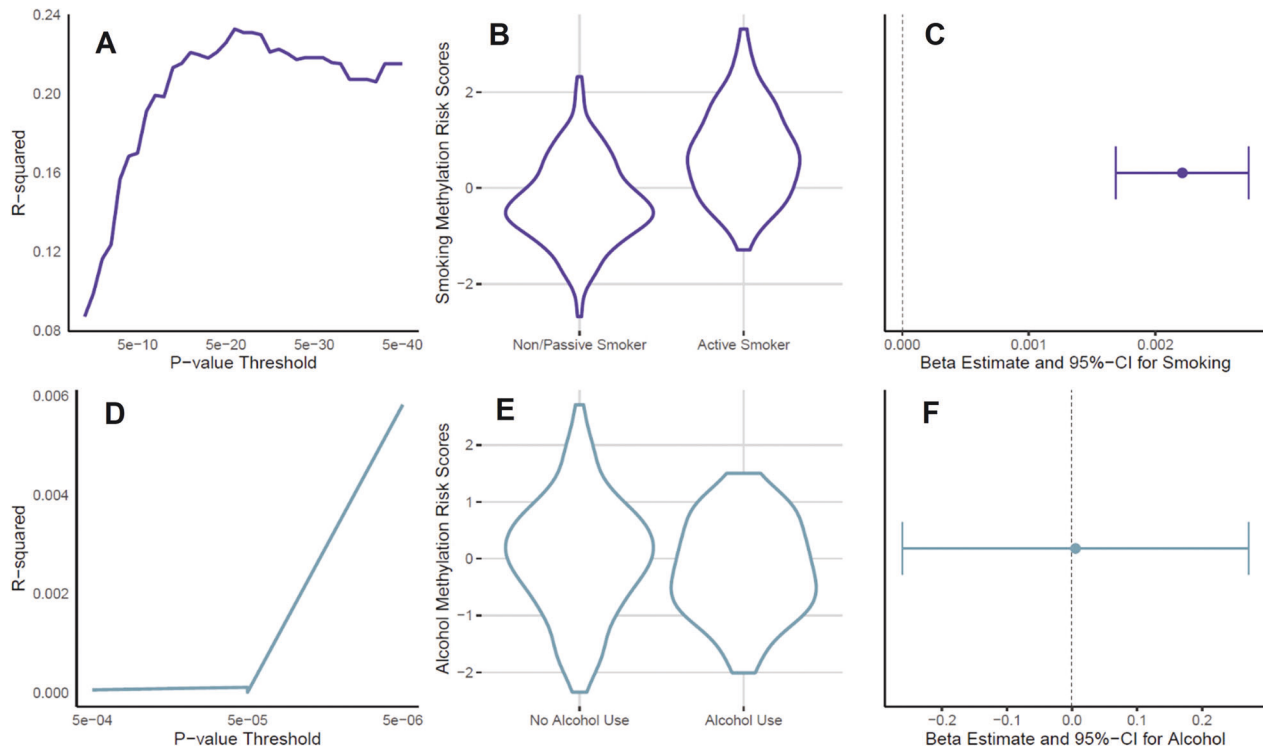
Fig. 2 Performance of the MRS for PTE (N = 262) and PAE (N = 262). A Correlation between PTE (continuous maternal urine cotinine levels) and MRS at several *p* value thresholds. **B** MRS (*p* value threshold of 5e-21) for non-/passive smoking mothers and active smoking mothers. **C** Effect of PTE (continuous maternal urine cotinine levels) on most highly correlated MRS (*p* value threshold of 5e-21). **D** Correlation between PAE and MRS at several *p* value thresholds. **E** MRS (*p* value threshold of 5e-06) for non-alcohol consuming mothers and alcohol consuming mothers. **F** Effect of PAE on most highly correlated MRS (*p* value threshold of 5e-06).

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 children [70]. The *ALDH* family of genes plays a key role in the metabolism of vitamin A, an essential molecule for neuronal differentiation and development [71]. Additionally, the identified probe (cg12031962) had high blood–brain concordance. However, there has been inconsistent evidence of an association between maternal smoking and ASD [72–74]. Furthermore, brain tissue-based DNAm in two of these genes (*MAD1L1* and *COL11A2)* was associated with ASD, indicating their role in neurodevelopment [75].

Mitotic arrest deficient 1 like 1 (*MAD1L1)* is a key regulatory gene of the cell cycle. It has been implicated in the etiology of neuropsychiatric 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 cells [76–78]. The biological mechanisms of both SZ and BPD 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 area [79]. However, the identified probe (cg01258793) did not exhibit blood–brain concordance using the BECon tool. Previous studies have found that maternal smoking was associated with a higher rate of SZ and BPD in offspring, suggesting that differential expression of MAD1L1 may be a plausible pathway that links smoking to neurodevelopmental disorders [80].

Calmodulin-binding transcription activator 1 (*CAMTA1*) is also involved in cell cycle regulation and is highly expressed in the

brain [81]. 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 [82, 83]. 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*.

Differential methylation at two CpG sites that we identified as significant mediators of the association between PAE and motor development mapped to *SLC18A2* (cg00512279) and *HNRNPU* (cg02230180) [84]. 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 disease [85–87]. Mutations in *HNRNPU* have been associated with infantile spasms [88, 89]. 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 associations [44]. Our MRS for prenatal smoking was highly predictive of PTE and is comparable to a previously established MRS for PTE by Reese et al. ($R^2 = 0.23$) and outperforms another established MRS for PTE by Richmond et al. ($R^2 = 0.14$) [45, 90, 91]. However, our MRS for PTE 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 ancestry between the external studies (primarily European ancestry) and the DCHS (African and
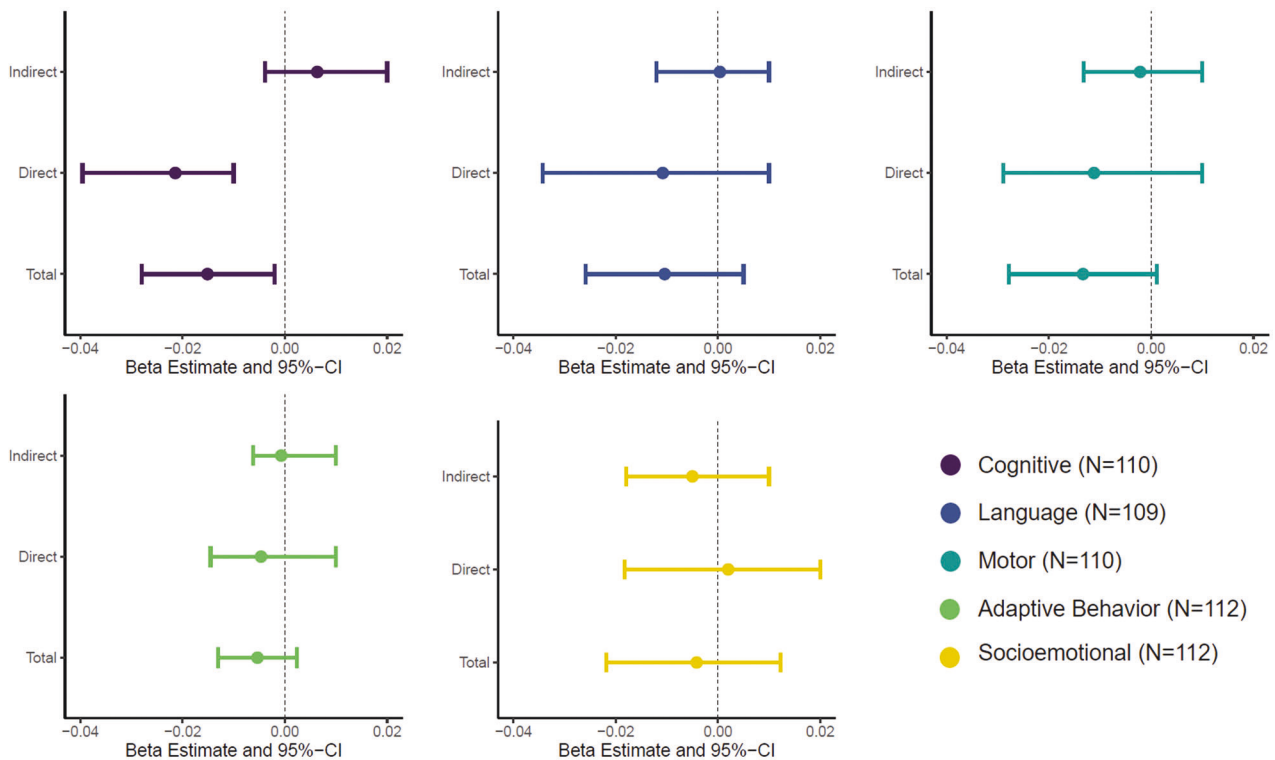
Fig. 3 Causal mediation analysis for PTE as the exposure, MRS as the mediator, and neurodevelopment across five domains at 6 months of age as the outcome. Estimated indirect effect, direct effect, and total effect with their corresponding 95% confidence intervals are shown.

mixed ancestry) reduced the prediction accuracy of our MRS for both prenatal smoking and prenatal alcohol use [92].

Our study has a few limitations. First, unlike PTE, which has been extensively studied [14–16, 93] 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 exposure are conducted in birth cohorts from HICs and tend to have a low prevalence of PAE [17, 69]. 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 (450 K) across six cohorts ($N = 1147$) did not find strong evidence of a relationship [17]. This limitation also in part explains the poor performance of our MRS for PAE. Furthermore, although the DCHS has a higher prevalence of both PAE and PTE compared to HIC cohorts, this study had a relatively small sample size at both 6 months ($N = 112$) and 24 months ($N = 184$), which may have limited our statistical power to detect associations between PTE and PAE and neurodevelopment across other domains and at 24 months of age.

Another limitation of 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 timepoint 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 effect, an underestimation

of the direct effect, and an increased type I error rate [94]. Because of this potential for bias, our results should be interpreted with caution and methods to correct for misclassification should be considered in future studies.

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 individuals and could therefore be linked to future cognitive outcomes [95]. Furthermore, many of the CpG sites we identified as significant mediators of the association between PTE and neurodevelopment showed good blood–brain concordance (assessed using the BECon tool [55]).

Lastly, many of our significant CpG sites were associated with at least one known mQTL, which is an indicator of the genetic influence on DNAm levels [56]. While we cannot decompose the effects of genetic and environmental factors on DNAm levels, only a proportion of the variation in DNAm levels is explained by genetic effects. In fact, the joint effects of single nucleotide polymorphisms (SNP) and environmental factors have been found to be larger contributors to DNAm variation than SNPs alone [96, 97]. Since the environmental exposures, PTE and PAE, were ascertained before cord blood was collected, and neurodevelopment was ascertained at age 6 and 24 months, reverse causation is unlikely to be a problem in this study. 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. Further, while this analysis had a small sample size, the use of novel methods to increase our statistical power

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 were estimated by CMA.

Domain	CpG (Gene)	ACME (95% CI)	ADE (95% CI)	Total effect (95% CI)	Prop med. (95% CI)
Cognitive (*n* = 110)	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 (RXRB, 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)
	cg09643326 (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 (C4orf29, 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, 14)
	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, 12)
	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 (CERS1, 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 (−0.024, −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 (*n* = 110)	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, 14)
	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 are shown.

Domain	CpG (Gene)	ACME (95% CI)	ADE (95% CI)	Total effect (95% CI)	Prop med. (95% CI)
Motor (n = 110)	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, STRA13*)	−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 (*KLLN, 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 (*KAZN, 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 behavior (n = 112)	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)

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 study [48]. 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 marginalized populations, the majority of studies are conducted in populations of European ancestry, which limits the field's ability to investigate the burden of disease in vulnerable communities [98]. 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 in the methylome 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.

CODE AVAILABILITY

REFERENCES
1. Fitzgerald E, Hor K, Drake AJ. Maternal influences on fetal brain development: the role of nutrition, infection and stress, and the potential for intergenerational consequences. Early Hum Dev. 2020;150:105190.
2. Wang X, Zuckerman B, Pearson C, Kaufman G, Chen C, Wang G, et al. Maternal cigarette smoking, metabolic gene polymorphism, and infant birth weight. J Am Med Assoc. 2002;287:195–202.
3. Zhang K, Wang X. Maternal smoking and increased risk of sudden infant death syndrome: a meta-analysis. Leg Med. 2013;15:115–21.
4. Kesmodel U, Wisborg K, Olsen SF, Henriksen TB, Secher NJ. Moderate alcohol intake during pregnancy and the risk of stillbirth and death in the first year of life. Am J Epidemiol. 2002;155:305–12.
5. Addila AE, Azale T, Gete YK, 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. Subst Abus Treat Prev Policy. 2021;16:1–15.

6. Marceau K, Bidwell LC, Karoly HC, Evans AS, Todorov AA, Palmer RH, et al. Within-family effects of smoking during pregnancy on ADHD: the importance of phenotype. J Abnorm Child Psychol. 2018;46:685–99.

7. Huang L, Wang Y, Zhang L, Zheng Z, Zhu T, Qu Y, et al. Maternal smoking and attention-deficit/hyperactivity disorder in offspring: a meta-analysis. Pediatrics. 2018;141:1–11.

8. Niemelä S, Sourander A, Surcel H-M, Hinkka-Yli-Salomäki S, McKeague IW, Cheslack-Postava K, et al. Prenatal nicotine exposure and risk of schizophrenia among offspring in a national birth cohort. Am J Psychiatry. 2016;173:799–806.

9. Talati A, Bao Y, Kaufman J, Shen L, Schaefer CA, Brown AS. Maternal smoking during pregnancy and bipolar disorder in offspring. Am J Psychiatry. 2013;170:1178–85.

10. May PA, Blankenship J, Marais AS, Gossage JP, Kalberg WO, Joubert B, et al. Maternal alcohol consumption producing fetal alcohol spectrum disorders (FASD): quantity, frequency, and timing of drinking. Drug Alcohol Depend. 2013;133:502–12.

11. Roozen S, Peters GJY, Kok G, Townend D, Nijhuis J, Koek G, et al. Systematic literature review on which maternal alcohol behaviours are related to fetal alcohol spectrum disorders (FASD). BMJ Open. 2018;8:e022578.

12. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. Nature. 2007;447:433–40.

13. Moosavi A, Ardekani AM. Role of epigenetics in biology and human diseases. Iran Biomed J. 2016;20:246–58.

14. 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;98:680–96.

15. Richmond RC, Simpkin AJ, Woodward G, Gaunt TR, Lyttleton O, McArdle WL, et al. Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the Avon longitudinal study of parents and children (ALSPAC). Hum Mol Genet. 2015;24:2201–17.

16. Sikdar S, Joehanes R, Joubert BR, Xu CJ, Vives-Usano M, Rezwan FI, et al. Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking. Epigenomics. 2019;11:1487–1500.

17. Sharp GC, Arathimos R, Reese SE, Page CM, Felix J, Küpers LK, et al. Maternal alcohol consumption and offspring DNA methylation: Findings from six general population-based birth cohorts. Epigenomics. 2018;10:27–42.

18. Loke YJ, Muggli E, Saffery R, Ryan J, Lewis S, Elliott EJ, et al. Sex- and tissue-specific effects of binge-level prenatal alcohol consumption on DNA methylation at birth. Epigenomics. 2021;13:1921–38.

19. Xu R, Hong X, Zhang B, Huang W, Hou W, Wang G, et al. DNA methylation mediates the effect of maternal smoking on offspring birthweight: a birth cohort study of multi - ethnic US mother – newborn pairs. Clin Epigenetics. 2021;13:47.

20. Küpers LK, Xu X, Jankipersadsing SA, Vaez A. la Bastide-van Gemert S, Scholtens S, et al. DNA methylation mediates the effect of maternal smoking during pregnancy on birthweight of the offspring. Int J Epidemiol. 2015;44:1224–37.

21. Caramaschi D, Neumann A, Cardenas A, Tindula G, Alemany S, Zillich L, et al. Meta-analysis of epigenome-wide associations between DNA methylation at birth and childhood cognitive skills. Mol Psychiatry. 2022. https://doi.org/10.1038/s41380-022-01441-w.

22. Neumann A, Walton E, Alemany S, Cecil C, González JR, Jima DD, et al. Association between DNA methylation and ADHD symptoms from birth to school age: a prospective meta-analysis. Transl Psychiatry. 2020. https://doi.org/10.1038/s41398-020-01058-z.

23. Hüls A, Wedderburn CJ, Groenewold NA, Gladish N, Jones MJ, Koen N, 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 J Biol Psychiatry. 2022;0:1–12.

24. Miyake K, Miyashita C, Ikeda-Araki A, Miura R, Itoh S, Yamazaki K, et al. DNA methylation of GFI1 as a mediator of the association between prenatal smoking exposure and ADHD symptoms at 6 years: the Hokkaido Study on Environment and Children's Health. Clin Epigenetics. 2021. https://doi.org/10.1186/s13148-021-01063-z.

25. Wiklund P, Karhunen V, Richmond RC, Parmar P, Rodriguez A, de Silva M, et al. DNA methylation links prenatal smoking exposure to later life health outcomes in offspring. Clin Epigenetics. 2019;11:1–16.

26. Tekola-Ayele F, Rotimi CN. Translational genomics in low- and middle-income countries: opportunities and challenges. Public Health Genomics. 2015;18:242–7.

27. Stein DJ, Koen N, Donald KA, Adnams CM, Koopowitz S, Lund C, et al. Investigating the psychosocial determinants of child health in Africa: the Drakenstein Child Health Study. J Neurosci Methods. 2015;252:27–35.

28. Zar HJ, Barnett W, Myer L, Stein DJ, Nicol MP. Investigating the early-life determinants of illness in Africa: the Drakenstein Child Health Study. Thorax. 2015;70:592–4.

29. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of infinium DNA methylation microarrays. Bioinformatics. 2014;30:1363–9.

30. Fortin J-P, Labbe A, Lemire M, Zanke B, Hudson T, Fertig E, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol. 2014;15:503.

31. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28:882–3.

32. Gervin K, Salas LA, Bakulski KM, Van Zelm MC, Koestler DC, Wiencke JK, et al. Systematic evaluation and validation of reference and library selection methods for deconvolution of cord blood DNA methylation data. Clin Epigenetics. 2019;11:1–15.

33. Vanker A, Barnett W, Nduru PM, Gie RP, Sly PD, Zar HJ. Home environment and indoor air pollution exposure in an African birth cohort study. Sci Total Environ. 2015;536:362–7.

34. Vanker A, Barnett W, Brittain K, Gie RP, Koen N, Myers B, et al. Antenatal and early life tobacco smoke exposure in an African birth cohort study. Int J Tuberculosis Lung Dis. 2016;20:729–37.

35. Donald KA, Wedderburn CJ, Barnett W, Nhapi RT, Rehman AM, Stadler JAM, et al. Risk and protective factors for child development: an observational South African birth cohort. PLoS Med. 2019;16:1–20.

36. Humeniuk R, Henry-Edwards S, Ali R, Poznyak V, Monteiro MG, Organization WH. The alcohol, smoking and substance involvement screening test (ASSIST): manual for use in primary care/prepared by R. Humeniuk̓ [et al]. https://apps.who.int/iris/handle/10665/44320.

37. Humeniuk R, Ali R, Babor TF, Farrell M, Formigoni ML, Jittiwutikarn J, et al. Validation of the alcohol, smoking and substance involvement screening test (ASSIST). Addiction. 2008;103:1039–47.

38. Myers B, Koen N, Donald KA, Nhapi RT, Workman L, Barnett W, et al. Effect of hazardous alcohol use during pregnancy on growth outcomes at birth: findings from a South African cohort study. Alcohol Clin Exp Res. 2018;42:369–77.

39. Donald KA, Fouche JP, Roos A, Koen N, Howells FM, Riley EP, et al. Alcohol exposure in utero is associated with decreased gray matter volume in neonates. Metab Brain Dis. 2016;31:81–91.

40. Bayley N. Bayley scales of infant and toddler development. 3rd edn. San Antonio, Tx: Harcourt Assessment; 2006.

41. Wedderburn CJ, Yeung S, Rehman AM, Stadler JAM, Nhapi RT, Barnett W, et al. Neurodevelopment of HIV-exposed uninfected children in South Africa: outcomes from an observational birth cohort study. Lancet Child Adolesc Health. 2019;3:803–13.

42. 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 Afr J Child Health. 2013;7:54–9.

43. Ballot DE, Ramdin T, Rakotsoane D, Agaba F, Davies VA, Chirwa T, 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. Int Sch Res Not. 2017;2017:1–5.

44. Hüls A, Czamara D. Methodological challenges in constructing DNA methylation risk scores. Epigenetics. 2020;15:1–11.

45. Chen J, Gatev E, Everson T, Conneely KN, Koen N, Epstein MP, et al. Pruning and thresholding approach for methylation risk scores in multi-ancestry populations. medRxiv: 2022.06.09.22276204 [Preprint]. 2022.

46. Gatev E, Gladish N, Mostafavi S, Kobor MS. CoMeBack: DNA methylation array data analysis for co-methylated regions. Bioinformatics. 2020;36:2675–83.

47. Zhang H, Zheng Y, Zhang Z, Gao T, Joyce B, Yoon G, et al. Genetics and population analysis: estimating and testing high-dimensional mediation effects in epigenetic studies. Bioinformatics. 2016;32:3150–4.

48. Liu Z, Shen J, Barfield R, Schwartz J, Baccarelli AA, Lin X. Large-scale hypothesis testing for causal mediation effects with applications in genome-wide epigenetic studies. J Am Stat Assoc. 2021;0:1–39.

49. Zhang CH. Nearly unbiased variable selection under minimax concave penalty. Ann Statist. 2010;38:894–942.

50. Zeng P, Shao Z, Zhou X. Statistical methods for mediation analysis in the era of high-throughput genomics: current successes and future challenges. Comput Struct Biotechnol J. 2021;19:3209–24.

51. Venables WN, Ripley BD. Modern applied statistics with S-plus. New York: Springer; 2002.

52. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. Mediation: R package for causal mediation analysis. J Stat Softw. 2014;59:1–38.

53. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychol Methods. 2010;15:309–34.

54. Phipson B, Maksimovic J, Oshlack A. MissMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics. 2016;32:286–8.

55. Edgar RD, Jones MJ, Meaney MJ, Turecki G, Kobor MS. BECon: a tool for interpreting DNA methylation findings from blood in the context of brain. Transl Psychiatry. 2017;7:e1187–10.

56. min JL, Hemani G, Hannon E, Dekkers KF, Castillo-Fernandez J, Luijk R, et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. Nat Genet. 2021;53:1311–21.

57. Battram T, Yousefi P, Crawford G, Prince C, Sheikhali Babaei M, Sharp G, et al. The EWAS catalog: a database of epigenome-wide association studies. Wellcome Open Res. 2022;7:41.

58. Subramoney S, Eastman E, Adnams C, Stein DJ, Donald KA. The early developmental outcomes of prenatal alcohol exposure: a review. Front Neurol. 2018. https://doi.org/10.3389/fneur.2018.01108.

59. Banderali G, Martelli A, Landi M, Moretti F, Betti F, Radaelli G, et al. Short and long term health effects of parental tobacco smoking during pregnancy and lactation: a descriptive review. J Transl Med. 2015;13:327.

60. Clifford A, Lang L, Chen R. Effects of maternal cigarette smoking during pregnancy on cognitive parameters of children and young adults: a literature review. Neurotoxicol Teratol. 2012;34:560–70.

61. 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. 2018;15:2866.

62. Donald KA, Hoogenhout M, du Plooy CP, Wedderburn CJ, Nhapi RT, Barnett W, et al. Drakenstein Child Health Study (DCHS): investigating determinants of early child development and cognition. BMJ Paediatr Open. 2018;2:e000282–e000282.

63. Bay B, Kesmodel US. Prenatal alcohol exposure - a systematic review of the effects on child motor function. Acta Obstet Gynecol Scand. 2011;90:210–26.

64. Lucas BR, Latimer J, Pinto RZ, Ferreira ML, Doney R, Lau M, et al. Gross motor deficits in children prenatally exposed to alcohol: a meta-analysis. Pediatrics. 2014;134:e192–209.

65. Hutchinson D, Youssef GJ, McCormack C, Wilson J, Allsop S, Najman J, et al. Correction to: prenatal alcohol exposure and infant gross motor development: a prospective cohort study. BMC Pediatr. 2019. https://doi.org/10.1186/s12887-019-1516-5.

66. Hendricks G, Malcolm-Smith S, Stein DJ, Zar HJ, Wedderburn CJ, Nhapi RT, et al. Prenatal alcohol exposure is associated with early motor, but not language development in a South African cohort. Acta Neuropsychiatr. 2019. https://doi.org/10.1017/neu.2019.51.

67. Julvez J, Ribas-Fitó N, Torrent M, Forns M, Garcia-Esteban R, Sunyer J. Maternal smoking habits and cognitive development of children at age 4 years in a population-based birth cohort. Int J Epidemiol. 2007;36:825–32.

68. Polanska K, Krol A, Merecz-Kot D, Ligocka D, Mikolajewska K, Mirabella F, et al. Environmental tobacco smoke exposure during pregnancy and child neurodevelopment. Int J Environ Res Public Health. 2017. https://doi.org/10.3390/ijerph14070796.

69. Felix JF, Joubert BR, Baccarelli AA, Sharp GC, Almqvist C, Annesi-Maesano I, et al. Cohort profile: pregnancy and childhood epigenetics (PACE) consortium. Int J Epidemiol. 2018;47:22–23u.

70. Gomez-Fernandez A, de la Torre-Aguilar MJ, Gil-Campos M, Flores-Rojas K, Cruz-Rico MD, Martin-Borreguero P, 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. Front Pediatr. 2018;6:1–9.

71. Xu X, Li C, Gao X, Xia K, Guo H, Li Y, et al. Excessive UBE3A dosage impairs retinoic acid signaling and synaptic plasticity in autism spectrum disorders. Cell Res. 2018;28:48–68.

72. von Ehrenstein OS, 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. 2021;190:728–37.

73. Jung Y, Lee AM, McKee SA, Picciotto MR. Maternal smoking and autism spectrum disorder: meta-analysis with population smoking metrics as moderators. Sci Rep. 2017;7:1–10.

74. Caramaschi D, Taylor AE, Richmond RC, Havdahl KA, Golding J, Relton CL, et al. Maternal smoking during pregnancy and autism: using causal inference methods in a birth cohort study. Transl Psychiatry 2018. https://doi.org/10.1038/s41398-018-0313-5.

75. Nardone S, Sharan Sams D, Reuveni E, Getselter D, Oron O, Karpuj M, et al. DNA methylation analysis of the autistic brain reveals multiple dysregulated biological pathways. Transl Psychiatry. 2014;4:1–9.

76. Liu X, Xie H, Fu Z, Yao Q, Han T, Zhan D, et al. MAD1L1 and TSNARE gene polymorphisms are associated with schizophrenia susceptibility in the Han Chinese population. BMC Med Genomics. 2021;14:1–10.

77. Su L, Shen T, Huang G, Long J, Fan J, Ling W, et al. Genetic association of GWAS-supported MAD1L1 gene polymorphism rs12666575 with schizophrenia susceptibility in a Chinese population. Neurosci Lett. 2016;610:98–103.

78. Hill MJ, Donocik JG, Nuamah RA, Mein CA, Sainz-Fuertes R, Bray NJ. Transcriptional consequences of schizophrenia candidate miR-137 manipulation in human neural progenitor cells. Schizophr Res. 2014;153:225–30.

79. Trost S, Diekhof EK, Mohr H, Vieker H, Krämer B, Wolf C, et al. Investigating the impact of a genome-wide supported bipolar risk variant of MAD1L1 on the human reward system. Neuropsychopharmacology. 2016;41:2679–87.

80. Quinn PD, Rickert ME, Weibull CE, Johansson ALV, Lichtenstein P, Almqvist C, et al. Association between maternal smoking during pregnancy and severe mental illness in offspring. JAMA Psychiatry. 2017;74:589–96.

81. Bas-Orth C, Tan YW, Oliveira AMM, Bengtson CP, Bading H. The calmodulin-binding transcription activator CAMTA1 is required for long-term memory formation in mice. Learn Mem. 2016;23:313–21.

82. Ramanan VK, Nho K, Shen L, Risacher SL, Kim S, McDonald BC, et al. FASTKD2 is associated with memory and hippocampal structure in older adults. Mol Psychiatry. 2015;20:1197–204.

83. Huentelman MJ, Papassotiropoulos A, Craig DW.Hoerndli FJ, Pearson J v., Huynh KD, et al. Calmodulin-binding transcription activator 1 (CAMTA1) alleles predispose human episodic memory performance. Hum Mol Genet. 2007;16:1469–77.

84. Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ. A navigator for human genome epidemiology. Nat Genet. 2008;40:124–5.

85. Rilstone JJ, Alkhater RA, Minassian BA. Brain dopamine–serotonin vesicular transport disease and its treatment. N. Engl J Med. 2013;368:543–50.

86. Zhai H, Zheng Y, He Y, Zhang Y, Guo Z, Cui W, et al. A case report of infantile parkinsonism-dystonia-2 caused by homozygous mutation in the SLC18A2 gene. Int J Neurosci. 2021:1–4.

87. Jacobsen JC, Wilson C, Cunningham V, Glamuzina E, Prosser DO, Love DR, et al. Brain dopamine-serotonin vesicular transport disease presenting as a severe infantile hypotonic parkinsonian disorder. J Inherit Metab Dis. 2016;39:305–8.

88. Du X, An Y, Yu L, Liu R, Qin Y, Guo 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 Med Genet. 2014;15:1–12.

89. Allen AS, Berkovic SF, Cossette P, Delanty N, Dlugos D, Eichler EE, et al. De novo mutations in epileptic encephalopathies. Nature. 2013;501:217–21.

90. Reese SE, Zhao S, Wu MC, Joubert BR, Parr CL, Håberg SE, et al. DNA methylation score as a biomarker in newborns for sustained maternal smoking during pregnancy. Environ Health Perspect. 2017;125:760–6.

91. Richmond RC, Suderman M, Langdon R, Relton CL, Smith GD. DNA methylation as a marker for prenatal smoke exposure in adults. Int J Epidemiol. 2018;47:1120–30.

92. Heyn H, Moran S, Hernando-Herraez I, Sayols S, Gomez A, Sandoval J, et al. DNA methylation contributes to natural human variation. Genome Res. 2013;23:1363–72.

93. Zhang B, Hong X, Ji H, Tang WY, Kimmel M, Ji Y, et al. Maternal smoking during pregnancy and cord blood DNA methylation: new insight on sex differences and effect modification by maternal folate levels. Epigenetics. 2018;13:505–18.

94. Valeri L, Reese SL, Zhao S, Page CM, Nystad W, Coull BA, et al. Misclassified exposure in epigenetic mediation analyses. Does DNA methylation mediate effects of smoking on birthweight? Epigenomics. 2017;9:253–65.

95. Bakulski KM, Halladay A, Hu VW, Mill J, Fallin MD. Epigenetic research in neuropsychiatric disorders: the "tissue issue". Curr Behav Neurosci Rep. 2016;3:264–74.

96. Czamara D, Eraslan G, Page CM, Lahti J, Lahti-Pulkkinen M, Hämäläinen E, et al. Integrated analysis of environmental and genetic influences on cord blood DNA methylation in new-borns. Nat Commun. 2019. https://doi.org/10.1038/s41467-019-10461-0.

97. Yuan V, Robinson WP. Epigenetics in development. In: Michels KB, editor. Epigenetic epidemiology. Cham: Springer International Publishing; 2022. p. 97–121.

98. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature. 2019;570:514–8.

ACKNOWLEDGEMENTS

12

AUTHOR CONTRIBUTIONS

SA: Formal analysis; interpretation of findings; visualization of findings; methodology; preparation of the original manuscript; review and editing of the manuscript. JC: Major contributor to the methylation risk score methodology; review and editing of the manuscript. DF: Methodology; review and editing the manuscript. AK: Methodology; review and editing the manuscript. NK, AV, CJW, and KAD: Investigation and data curation; review and editing of the manuscript. HJZ and DJS: Supervision and funding acquisition; review and editing of the manuscript. AH: Conceptualization; methodology; supervision; review and editing of the manuscript. All authors read, edited, and approved the final manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41398-022-02195-3.

Correspondence and requests for materials should be addressed to Anke Hüls.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.