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:

_____ _____

Dakotah N. Feil Date

DNA methylation as a potential mediator of the association between indoor air pollution and neurodevelopmental delay in a South African birth cohort

By

Dakotah N. Feil

Master of Public Health

Epidemiology

_____ [Chair's signature]

Anke Hüls, PhD

Committee Chair

DNA methylation as a potential mediator of the association between indoor air pollution and neurodevelopmental delay in a South African birth cohort

By

Dakotah Feil

Bachelor of Science

University of Michigan

2018

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 indoor air pollution and neurodevelopmental delay in a South African birth cohort

By Dakotah N. Feil

Exposure to indoor air pollution (IAP) during pregnancy has been linked to neurodevelopmental delay in toddlers. Epigenetic modification, particularly DNA methylation (DNAm), may help explain this link. In this study, we employ three high-dimensional mediation analysis (HIMA, DACt, and gHMA) methods followed by causal mediation analysis to identify differentially methylated CpG sites and genes that mediate the association between IAP and neurodevelopmental delay. Analyses were performed using data from 142 mother-child pairs from the South African Drakenstein Child Health Study (DCHS). DNAm from cord blood was measured using the Infinium MethylationEPIC and HumanMethylation450 arrays. Neurodevelopment was assessed at 2 years of age with the Bayley Scores of Infant and Toddler Development (BSID-III). Particulate matter with an aerodynamic diameter of 10μm or less (PM_{10}) was collected from devices placed in participants' homes during the second trimester of pregnancy. A total of 29 CpG sites and 4 genes (*GOPC*, *RP11-74K11.1*, *DYRK1A*, *RNMT*) were identified as significant mediators. Proportion mediated estimates (95%-confidence interval) ranged from 0.29 (0.014,0.86) for cg00694520 to 0.54 (0.11,1.56) for cg05023582. *DYRK1A* and several genes our CpG sites mapped to, including *CNKSR1, IPO13, IFNGR1, LONP2,* and *CDH1* are associated with biological pathways implicated in neurodevelopment. These findings suggest that DNAm might mediate the association between prenatal indoor PM_{10} exposure and cognitive neurodevelopment.

DNA methylation as a potential mediator of the association between indoor air pollution and neurodevelopmental delay in a South African birth cohort

By

Dakotah Feil

Bachelor of Science

University of Michigan

2018

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

2021

Acknowledgements

I would like to thank. Dr. Anke Hüls, my thesis advisor and mentor who taught me so much about genetic epidemiology and how to be a better public health practitioner. I would also like to thank the Hüls lab for all their input, advice, and assistance on this study.

I would like to thank all the co-authors for this study: Grace M. Christensen, Sarina Abrishamcar, Aneesa Vanker, Nastassja Koen, Anna Kilanowski, Nadia Hoffman, Kirsten A. Donald, Michael S. Kobor, Heather J. Zar, and Dan J. Stein.

Lastly, I would like to thank all of the researchers before me who have laid the foundation for this study.

Introduction

Certain detrimental effects of air pollution on pregnancy outcomes such as low birth weight and respiratory disease in infants are well-known and have been confirmed by many studies over the last several decades [1,2]. However, the current literature remains sparse regarding the role of prenatal air pollution exposure on other health outcomes such as neurodevelopment, and sparser yet for the biological mechanisms underpinning these associations. A handful of studies have reported significant associations between prenatal air pollution exposure and neurological conditions such as autism spectrum disorder (ASD), attention deficit hyperactivity disorder (ADHD), and neurodevelopmental delay [3–7] .

However, the bulk of these studies have been conducted in high income country (HIC) contexts and have focused on the effects of outdoor air pollution; therefore, findings may not be wholly generalizable to other settings [8,9]. It is equally important to address the impact of indoor air pollution, particularly in low- and middle-income country (LMIC) settings where burning solid fuels such as coal or wood for cooking and heat is common. In such settings, solid fuel burning can greatly increase indoor air pollutant (IAP) concentration and in turn impact infant neurodevelopment and other health outcomes [5,10,11] .

Epigenetic modification has long been discussed as the missing link to understanding how gene-environment interactions affect neurodevelopment [12,13]. As such, careful dissection of the relationship between air pollution, epigenetic modification, and neurological outcomes may allow us to better understand the complex mechanisms behind environment-associated neurological disorders and neurodevelopment. With the

rise of high-throughput genomics, the field of epigenetics has undergone rapid development. Epigenetic modification, specifically DNA methylation (DNAm), has been linked to a number of neurological outcomes such as severe neurodevelopmental delay, schizophrenia, ASD, and ADHD [14–20] . DNAm is known to play a key role in embryonic development and has been hypothesized to impact neural stem cell differentiation and maintenance [21] , thereby affecting neurological outcomes throughout the life course. It should be noted that DNAm is reversible and identification of causal, differentially methylated CpG sites may be useful in multiple contexts, including clinical therapy design and biomarker identification [22] .

DNAm levels are altered by a number of environmental exposures such as drugs, nutrition, stress and air pollution [8,9,23,24] . Data collected from the Pregnancy and Child Epigenetics (PACE) consortium have been used to study the effects of prenatal exposure to nitrogen dioxide (NO_2), airborne particulate matter with a diameter 10 microns or less (PM_{10}), and airborne particulate matter with a diameter of 2.5 microns or less ($PM_{2.5}$) on newborn and childhood DNAm [8,9] . Prenatal exposure to each of these pollutants was associated with differential DNAm in neonates which highlights the need for additional research to understand how environment-driven epigenetic changes impact fetal development and downstream health outcomes [8,9]. While there is evidence of an association between air pollution and DNAm as well as between DNAm and neurodevelopment, few studies have examined the interconnections between them.

To the best of our knowledge, only one study has examined the interconnections between prenatal IAP exposure, DNAm and neurodevelopment in a mediation analysis [5]. This study, which was based on the same South African birth cohort as our current study, focused on deviations of epigenetic gestational age from chronological gestational age (ΔGA) as potential mediator, which has been hypothesized to be an indicator for adverse fetal development. However, this previous study did not find evidence of mediation by ΔGA, leading us to take a more granular approach to understand the role of DNAm as potential mediator of the association between prenatal IAP and neurodevelopment [5]. In this study, we examine distinct, differentially methylated DNAm probes to investigate whether mediation by differential DNAm occurs on the pathway from prenatal IAP exposure to neurodevelopmental delay.

Innovation in the field of mediation analysis has expanded the scope of mediation analysis to allow for the testing of multiple mediators at a time, a tactic which up until recently was not widely practiced due to methodological constraints discussed later in this paper. Because these methods remain novel, there is no consensus as to which will serve as the gold standard [25]. As such, we have incorporated three of the most readily available and well-documented methods into our analyses: HIMA, DACT, and gHMA. Each of these methods tackles the multiple mediation question in a different way. HIMA uses screening to reduce dimensionality, followed by minimax concave penalty (MCP) estimation and joint significance testing of the exposure-mediator and mediator-outcome effects; the null hypothesis of no mediation is rejected only when both effects are significant [25,26]. DACT leverages epigenome-wide multiple testing to estimate the

proportions of the composite null hypothesis to improve power [25,27]. Gene-based HMA (gHMA) takes a slightly different approach and focuses on genes as functional regions as opposed to individual CpG sites. This method is comprised of a linear and non-linear component as the true nature of the relationship between the mediator and outcome is generally unknown; gHMA combines the nonlinear and linear components in a single omnibus test for mediation effects [25,28].

In this study, we aim to identify any differentially methylated CpG sites and gene regions that mediate the association between prenatal exposure to indoor PM_{10} and neurodevelopment measured at two years of age in the South African Drakenstein Child Health Study (DCHS) using a combination of high dimensional mediation analysis methods and traditional causal mediation analysis.

Materials & Methods

Study Population

The Drakenstein Child Health Study (DCHS) is a South African, population-based birth cohort that enrolls pregnant women from two primary health care clinics in peri-urban communities: TC Newman and Mbekweni. These clinics serve two demographically distinct populations, specifically a majority Black African ancestry community and a majority mixed ancestry community [29]. The DCHS follows infants starting at birth and continues follow-up until at least 5 years of age [29]. Our study population was composed of a total of 142 mother-child pairs enrolled in the DCHS with measures available for cord blood DNA methylation, genotype data, and Bayley Scales of Infant

and Toddler Development in at least one of the following domains: general cognitive function, general adaptive behavior, language, and motor control. Inclusion was also limited to mother-child pairs with measures available for relevant covariates which included race, maternal age, maternal smoking status, maternal alcohol use, birth weight, sex, and socioeconomic status (SES) score (Table 1). Smoking status was determined by maternal urine cotinine levels collected prenatally, while alcohol use was measured via the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST), a tool which was introduced by the World Health Organization (WHO) and which has shown good validity in LMIC settings [30] . Socioeconomic status was captured with a questionnaire and a corresponding socioeconomic score was designed to summarize four socioeconomic indicators: educational attainment, employment status, household income and assets and market access [30] .

The DCHS staff obtain written consent from mothers on an annual basis and the study was approved by the Ethics Committee of the Faculty of Health Sciences, University of Cape Town, by Stellenbosch University and the Western Cape Provincial Research committee [29].

DNA methylation measurements

As described previously by Hüls et al. (2022) [14], DNA was extracted from cord blood collected at time of delivery [31]. DNA methylation measures were obtained with both the Illumina Infinium HumanMethylation450 BeadChips (n=156) and the MethylationEPIC BeadChips (n=160). Pre-processing and statistics were done using R

3.5.1 (R Core Team 2018). Raw iDat files were imported into Rstudio where intensity values were converted into beta values. The 450K and EPIC datasets were merged using the minfi R package [32]. Background subtraction, color correction and normalization were performed using the preprocessFunnorm function [33] . Following sample and probe filtering, 273 samples and 409,033 probes remained for downstream analysis. Of these samples, 142 had genotype data, at least one BSID-III score measured at 2 years of age and data available for all relevant covariates (Table 1). Batch effects were removed using ComBat from the R package sva [34] . Cord blood cell type composition was predicted using the most recent cord blood reference data set [35] .

Neurodevelopment measurements

The Bayley Scales of Infant and Toddler Development (BSID-III) is a widely used tool for assessing neurodevelopment in children between 0 and 2 years of age in four distinct domains: cognitive function, language, motor function, and adaptive behavior [30]. BSID-III has been validated in LMIC settings and previous research supports its use in South Africa specifically [36–38] . DCHS assesses neurodevelopment using BSID-III administered to the child at 2 years of age. The DCHS BSID-III assessment is conducted by a trained professional and incorporates direct observation of the child as well as caregiver input [30] . Each domain is scored according to the BSID-III manual using BSID-III specific software (BSID-III Scoring Assistant Update Version 2.02 with BSID-III PDA conduit) [30,39] . Composite scores for cognitive, motor, language, and

general adaptive behavior domains were scaled to have a mean of 100 and standard deviation of 15.

Assessment of Indoor Air Pollution Exposure

As described previously [40–42], PM_{10} was measured using a personal air sampling pump (AirChek 52; SKC, Eighty Four, PA, USA), connected to a styrene filter cassette (37 mm cassette blank; SKC) with a gravimetrically pre-weighted filter (PVC filter 37 mmx5 μm with support pad; SKC) left in the home for 24 hours [40,41]. Filters were weighed after sampling and National Institute for Occupational Safety and Health method 0600 was used to calculate an average PM_{10} concentration over 24 hours [42,43]. These 24-hour average PM_{10} measurements were used for our analyses.

Statistical Analysis

We used different high-dimensional mediation analysis approaches to investigate whether mediation by differential DNAm occurs on the pathway from prenatal IAP exposure to neurodevelopmental delay. Our primary outcome was the BSID-III cognitive domain, for which we have shown a significant total effect of PM_{10} in this subsample of the DCHS [5]. The other BSID-III neurodevelopment domains were used as secondary outcomes to evaluate consistency of results across domains.

Mediation analyses rely on the following three assumptions: 1) no exposure (PM_{10}) - mediator (DNAm) confounding, 2) no mediator (DNAm) - outcome (BSID-III Score) confounding and 3) no exposure (PM_{10}) -outcome (BSID-III Score) confounding [44]. To

fulfill these assumptions to the best of our knowledge, we constructed three directed acyclic graphs (DAGs) to visualize each of these three paths (Figure 1). Confounders were selected based on existing literature and a minimal sufficient adjustment set was identified for each pathway via the tracing of association directions and elimination of any potential confounders already associated with a precursory confounder. Exposure-mediator models were adjusted for SES score, genetic ancestry, and maternal smoking and mediator-outcome models were adjusted for maternal alcohol use, maternal age, SES score, sex, genetic ancestry, and maternal smoking. Birth weight as a proxy for gestational age was recognized as another possible mediator of the exposure-mediator association and is a possible mechanism through which prenatal IAP exposure could impact DNAm, therefore we did not control for birth weight in our analyses. Models were also adjusted for the first three cell type principal components (PCs), which explained >90% of cell type heterogeneity [45–47]. We adjusted for genetic ancestry by including the first five genotype PCs to account for population stratification [46].

To assess the role of DNAm as a potential mediator of the association between prenatal exposure to PM_{10} and neurodevelopment at age two years, we employed three well-documented methods for high-dimensional mediation analysis: HIMA (high-dimensional mediation analysis), DACT (divide-aggregate composite-null), and gHMA (gene-based high-dimensional mediation analysis) [26–28]. As these methods are quite novel, there is no consensus as to which is considered the gold standard [25]. Therefore, we have incorporated an analysis pipeline based on a combination of these methods with traditional causal mediation analysis in order to assess the robustness of results (Figure

S1). It should be noted that all models, used both for high-dimensional mediation analysis and causal mediation analysis, were adjusted for the appropriate confounders as defined above.

HIMA, a high-dimensional mediation analysis method introduced by Zhang et al. (2016) [26], employs a dimensionality reduction technique followed by minimax concave penalty (MCP) - penalized estimation of mediation effects and joint significance testing for mediation effects in order to identify significant mediators. Dimensionality reduction is performed using the sure independence screening (SIS) method which is built on a correlation learning framework that essentially filters out features that are weakly correlated with the response variable [48]. The HIMA joint significance testing procedure rejects the null hypothesis of no mediation only when both the exposure-mediator (α) and mediator-outcome effects (β) are significant [26] .

DACT leverages epigenome-wide multiple testing to estimate the proportions of the composite null hypothesis to improve power [25,27]. A preliminary step for the DACT method is to create two linear models pertaining to each CpG site: the first to model the exposure-mediator association (α) and the second to model the mediator-outcome association (β). While HIMA uses a screening technique to reduce dimensionality, DACT does not involve a screening step by default. As shown in Figure S1, we performed a pre-screen of CpG sites based on their association with the exposure (PM_{10}) and the outcome (BSID-III scores of neurodevelopment); only CpG sites with $p<0.05$ for both associations were included in the downstream analyses. Given previous

findings indicating a negative association between prenatal IAP exposure and neurodevelopment in the DCHS cohort [5], we chose to additionally filter our sites by only allowing a negative natural indirect effect (NIE) defined by $\alpha * \beta$ (acting in same direction as the association between IAP exposure and neurodevelopment). As we chose to pre-screen our CpG sites, we used the Efron correction feature of the DACT package to estimate the proportions of the composite null (Figure S2) as opposed to Jin and Cai correction which is recommended if performing epigenome-wide mediation effect testing with DACT [27].

gHMA was developed by Fang et al. (2020) [28] and focuses on genes as functional units as opposed to individual CpG sites. gHMA is primarily composed of three components: 1) linear mediation analysis, 2) nonlinear mediation analysis and 3) an omnibus test of mediation effects. Significance testing results for both the linear and nonlinear mediation analysis steps are combined using the gHMA omnibus (gHMA-O) test. As the true relationship between mediators and outcomes are often not well understood in practice, gHMA-O transforms and combines p-values from the linear and nonlinear analyses in order to construct the gHMA-O test statistic, which is used to assess mediation effects at the gene level [28] . CpG sites were annotated by closest gene using the Bioconductor package hiAnnotator (https://bioconductor.org/packages/release/bioc/html/hiAnnotator.html) and the Ensembl gene predictions (ensGene, version of Apr-06-2014; http://hgdownload.soe.ucsc.edu/goldenPath/ hg19/database/ensGene.txt.gz) as previously described in Hüls et al. 2022 [49].

P-values for CpG sites which passed the screening step and were tested using DACT and HIMA were corrected for multiple testing using the Benjamini-Hochberg false discovery rate correction (FDR) [50]. Due to the fact gHMA assesses differentially methylated gene regions as opposed to individual CpG sites, gHMA p-values were FDR corrected for the total number of gene regions tested, not the number of distinct CpG sites. HIMA and DACT CpG sites that remained significant at a false discovery rate of 0.05 were then funneled into a traditional causal mediation analysis using the function *mediate* from the R package *mediation* to obtain estimates of natural indirect effect (NIE), direct effect (DE), total effect (TE), and proportion mediated (PM) [51].

Results

Description of Study Participants

The final study sample consisted of 142 mother-child pairs with data available for genotype, cord blood methylation, PM_{10} concentration, scores for one or more BSID-III domains, and for all relevant covariates (Table 1). In total, 48.6% of infants were of Black African ancestry and 51.4% were of mixed-race ancestry; 40.1% of infants were female. The mean PM_{10} concentration was 64.5 $\mu g/m^3$ with a standard deviation of 96.8 $\mu g/m^3$. Mean composite BSID-III scores were 85.14 for the cognitive domain, 84.31 for the language domain, 94.04 for the motor function domain, and 83.72 for the general adaptive behavior domain. The prevalence of maternal smoking was high, with 40.1% of mothers classed as passive smokers and 33.1% classed as active smokers based on urine cotinine levels.

CpG-based High-Dimensional Mediation Analysis

After BH FDR adjustment for multiple testing, DACT identified a total of 123 distinct CpG sites across the cognitive (35 CpG sites, primary outcome), language (45 CpG sites), motor function (13 CpG sites), and general adaptive behavior (39 CpG sites) domains as significant mediators of the association between PM_{10} and neurodevelopment (Tables S1-S4). A total of 9 CpG sites were shared between at least two domains (Table S5) and one CpG site (cg26858414) was shared across the language, general adaptive behavior, and motor function domains. These 123 CpG sites were further examined via causal mediation analysis. Results for our primary outcome (cognitive development) are presented here (Table 2; Figure 2) and results for our secondary outcomes, for which we did not find a total effect of PM_{10} are presented in Supplementary Tables S6-S9.

Of the 35 CpG sites identified with DACT for the cognitive domain, 29 demonstrated significant natural indirect effects (NIE), significant estimates for proportion mediated, and significant estimates for total effect (TE) and two of them (cg13690126 and cg03234186) could be successfully validated for the language domain (Table 2; Figure 2). All effect estimates have been multiplied by the interquartile range (IQR) of PM_{10} observed in this cohort (58.78 $\mu g/m^3$) and therefore represent estimated effects per one IQR increase in PM_{10}. Estimated proportion mediated (95%-confidence interval) ranged from 0.29 (0.014,0.87) for cg00694520 to 0.54 (0.11,1.56) for cg05023582. Cg05023582 also showed the largest NIE estimate (95%-confidence interval) of -0.49 (-0.959, -0.146) per one IQR increase in PM_{10} (Figure 2).

After correction for multiple testing, HIMA did not identify any CpG sites that significantly mediated the effects of prenatal PM_{10} exposure on neurodevelopment in any domain (Table S10-S13). However, prior to multiple testing correction, one CpG site was identified as a significant mediator (cg05796992); this site was also identified with DACT and demonstrated a significant NIE estimate (95%-confidence interval) of -0.438 (-0.942, -0.0426) per one IQR increase in PM_{10} and a high estimated proportion mediated (95%-confidence interval) of 0.484 (0.0468, 1.58) in our causal mediation analysis (Figure 2, Table S6).

Gene-based High-Dimensional Mediation Analysis

Differential methylation in four gene regions (*GOPC*, *RP11-74K11.1*, *RNMT*, and *DYRK1A*) was identified as a significant mediator of the association between PM_{10} and cognitive development (Table 3). All of these differentially methylated gene regions could be validated for the other domains (Table 3). No CpG sites mapping to any of these 4 genes were identified for the cognitive domain with either of the CpG site-based methods (DACT or HIMA) (Tables S15-S18).

Additionally, we identified five differentially methylated genes as significant mediators for the secondary outcome “motor domain”, however, none of these were found to be significant for the other domains (Table 3).

Discussion

This study of 142 mother-child pairs from low SES communities in South Africa found a total of 29 distinct, differentially methylated DNAm probes to significantly mediate the effect of prenatal exposure to PM_{10} on cognitive neurodevelopment at age 2 years measured by BSID-III scores. Additionally, we found four differentially methylated gene regions which significantly mediate the effect of prenatal PM_{10} exposure on cognitive neurodevelopment using a gene-based high-dimensional mediation analysis technique.

Comparison with previous studies

To our knowledge, this study is the first to examine differential DNAm at individual probes as potential mediators of the association between prenatal PM_{10} exposure and neurodevelopment. However, a number of studies have examined the association between prenatal PM_{10} exposure and DNAm as well as the association between DNAm and neurodevelopment [8,9,14–20].

Many epigenetic-wide association studies (EWAS) have reported differentially methylated CpG sites associated with prenatal air pollution exposure [8,9,52–54] ; however, we did not identify any overlap between our findings and existing findings. It should be noted that our findings are not entirely comparable as, per the underlying assumptions of mediation analysis, probes must be associated with both PM_{10} exposure and cognitive neurodevelopment. Replication is a common problem in EWAS which often lack robust associations at single CpG sites across cohorts [52]. Further research is needed to validate our findings from high dimensional mediation analysis.

A recent DCHS study identified CpG sites significantly associated with severe neurodevelopmental delay in the cognitive (cg26971411, cg00490349, and cg15660740), language (cg26971411 & cg00490349), and motor (cg26971411 & cg00490349) domains [46] . We included these sites in our causal mediation analysis step; however, we did not observe significant evidence of mediation in any corresponding BSID-III domain for these CpG sites (Table S20). A recent meta-analysis examining epigenome-wide associations between DNAm at birth and childhood cognitive skills synthesizing data from eight pregnancy cohorts within the Pregnancy and Childhood Epigenetics (PACE) consortium (N=3300) did not find substantial evidence that differential cord blood DNAm at individual CpG sites is associated with cognitive skills [19]. We compared our findings to those from several EWAS investigating DNAm and cognitive development examined in the PACE study; however, no overlap was identified between our findings and those of previous studies [19,55] .

CpG-based High-Dimensional Mediation Analysis

We identified 29 differentially methylated CpG sites to significantly mediate the association between prenatal PM_{10} exposure and cognitive neurodevelopment. Of the 29 CpG sites, differential DNAm at 21 of these CpG sites has been associated with aging in EWAS examining DNAm trajectories occurring over the course of childhood [56,57]. Differential methylation at three CpG sites (cg23560546, cg22572779, cg15000966) has been associated with fetal brain development [58] . Differential

methylation at one CpG site (cg16975959) has been previously identified as a mediator of the association between maternal smoking and birth weight [59] (Table S21). Proportion mediated estimates for these 29 CpG sites are quite high, ranging from cg00694520 with 0.29 (0.014,0.87) to cg05023582 with 0.54 (0.11,1.56). However, such high PM estimates for each CpG site should be interpreted with caution due to the associated wide confidence intervals, our small sample size, and the fact that causal mediation analysis does not account for correlation between mediators, which we found to be present among these CpG sites (Figure S3). Several of these CpG sites are located within or adjacent to genes known to influence fetal development and/or neurological outcomes. Herein we discuss CpG sites that map to genes that have been previously linked to neurological outcomes.

Cg13690126 is located in *CNKSR1*, a protein-coding gene with low tissue expression specificity; defects in *CNKSR1* have been linked to syndromic autosomal recessive intellectual disability (ID) [60,61]. Najmabadi et al. (2011) [62] speculate its function as a scaffold protein mediating communication between Ras and Rho GTPase signaling pathways which have in turn been shown to play a role in neurodevelopmental disorders [62,63] . Cg07070893 is located in a promotor region for Importin 13 (*IPO13*), a gene showing tissue enhanced specificity in the brain and skeletal muscle [60,64]. *IPO13* is associated with agenesis of the corpus collosum and has been implicated in embryonic stem cell survival. *IPO13* has been proposed as integral to brain development, particularly for the purposes of neural cell-specific cargo trafficking [60,65,66] .

Cg23560546 is found in an enhancer region of Death Associated Protein Like 1 (*DAPL1*), a protein coding gene thought to be involved in early stages of epithelial differentiation and/or apoptosis [60]. *DAPL1* been identified as a significantly differentially methylated region (DMR) in a 2021 study comparing DNAm in blood cells of toddlers with Down syndrome to neurotypical toddlers [67] . Cg15007548 is located in the gene body of Tetratricopeptide Repeat, Ankyrin Repeat and Coiled-Coil Domain-Containing 1 (*TANC1*) [60]. *TANC1* is a protein-coding gene with low tissue specificity thought to regulate dendritic spines and excitatory synapses [60,64,68] . Dendritic spines are integral to synaptic function and loss of function in dendritic spines has been associated with a number of neurological disorders [69] . Cg26668632 is located in a promoter region for *IFNGR1* which belongs to the type II cytokine receptor family and encodes a ligand-binding chain of the gamma interferon receptor (IFN-γ) [60]. Several studies have found that IFN-γ signaling targets play a role in neuronal development and synaptic activity and a 2020 study [70] suggests that IFN-γ signaling is involved in neurodevelopmental disorder etiology [70–72]. Cg19187486 can be found in a promotor region for *ARID3B*, a member of the AT-rich interaction domain (ARID) family of DNA-binding proteins; *ARID3B* encodes a transcription factor associated with neuroblastoma growth and existing literature is sparse regarding additional gene function. However, the *ARID3B* complex has been implicated as playing a key role in regulation of genes related to placental development [73].

Cg05796992 occupies the same position as SNP rs78855787 in an intron of the gene *LRRK1*. *LRRK1* is a protein coding gene that shows low general tissue specificity and enhanced tissue specificity in the human brain in the cerebral cortex [60,64]. *LRRK1* expression has been found to interact significantly with parental warmth to impact executive function [74] ; variants in *LRRK1* have also been linked to Parkinson's disease [75]. Cg01085137 is located in a promotor region for Peroxisomal Lon Peptidase 2 (*LONP2*), a protein-coding gene encoding a protease integral to peroxisome function [60,76] which has been posited as a core component of human brain development [76,77] . Cg23989635 is located in the first exon of Cadherin 1 (*CDH1*), a protein coding gene showing enhanced expression in the parathyroid gland that has been implicated in neuronal differentiation and synaptic development in the central nervous system [60,64,78–80] . *CDH1* downregulation has been proposed to play a role in congenital neurodevelopmental disorders [81].

Cg0060655 occupies the same position as SNP rs147829886 and is found within an intron on the gene *FAM207A/SLX9* ribosome biogenesis factor*;* interestingly, *FAM207A* hypermethylation in umbilical cord tissue has been linked to pre-term birth, which in turn is associated with delayed neurodevelopment [60,82–84]. Cg14243899 is located at the same position as SNP rs10852772 in an intron of *ST6GALNA-C2*. *ST6GALNA-C2* is part of the sialyltransferase gene family and paralog to *ST6GALNA-C1*. A 2011 twin study found differential DNAm of this paralog to be linked to major psychoses such as schizophrenia (SZ) and bipolar disorder (BD); the study also found hypomethylation of the gene in an independent sample of postmortem brain tissue samples from psychosis

patients [60,85]. Cg23054321 is located in a promotor associated region proximal to Leucine Rich Repeat and Ig Domain Containing 3 (*LINGO3*). *LINGO3* is protein coding gene with tissue enriched in the brain [60,64]. The *LINGO* gene family has been found to show increased expression as an embryo develops whereas only low levels of these genes are found in adult brains with only *LINGO1* and *LINGO3* being detectable [86]. Epigenetic changes in *LINGO3* have been correlated with depression and a paralog to *LINGO3*, *LINGO1*, acts as a negative regulator of a number of processes key to cognitive function [87]. Cg05023582 is located in the Protein Rich and Gla Domain 2 (*PRRG2*) gene region and is associated with a promotor region for Proline Rich 12 (*PRR12*); existing literature is sparse regarding *PRRG2* potential function in neurodevelopment however several recent publications implicate *PRR12* haploinsufficiency as linked to neurodevelopmental outcomes such as intellectual disability (ID) [60,88,89]. The genes associated with the remaining CpG sites do not appear to be as well-represented in neuropathological and neurodevelopmental literature; additional research is needed to elucidate the roles of each of these differentially methylated sites on neurodevelopment.

Gene-based High-Dimensional Mediation Analysis

gHMA identified four differentially methylated gene regions associated with BSID-III cognitive neurodevelopmental scores (*GOPC*, *RP11-74K11.1*, *DYRK1A*, *RNMT*; Table 3). Golgi-Associated PDZ and Coiled-Coil Motif-Containing Protein (*GOPC*) is a protein coding gene; it has been linked to regulation of *GRID2* gene expression which has been shown to impact neurodegeneration [60,90]. *RP11-74K11.1* is a pseudogene on

chromosome 12 and although few studies have investigated *RP11-74K11.1*, its highest median expression is found in the brain, particularly the cerebellum [60,64,91]; additional research is needed to understand the role of *RP11-74K11.1* on neurodevelopment. Dual Specificity Tyrosine Phosphorylation Regulated Kinase 1A (*DYRK1A*) is located on chromosome 21 and encodes a protein kinase. *DYRK1A* is has been strongly linked to brain development and function across the life course [64,92]. Decreased expression of *DYRK1A* has been found in patients with autistic spectrum disorder (ASD) while elevated expression has been linked to Down syndrome (DS) [92]. Lastly, RNA Guanine-7 Methyltransferase (*RNMT*) is a protein coding gene found on chromosome 18; it is known to play a role in RNA binding and mRNA-methyltransferase activity [60]. The role of *RNMT* in neurodevelopment is not well-documented and further research is needed to better understand if such a link exists and its underpinning mechanism.

Although no individual CpG sites located in these gene regions were identified using DACT or HIMA, it is possible that differential methylation on the gene region scale plays a crucial role in neurodevelopment. Probes contained within or proximal to gene regions identified as significant mediators using the gHMA method were eliminated from the DACT analysis pipeline due to lack of significant exposure-mediator and/or mediator-outcome associations (Table S16-S19). The discrepancy between differentially methylated gene regions identified with gHMA and gene regions associated with CpG sites identified with DACT may be attributable to interaction effects between proximal, differentially methylated DNAm probes that were not captured via DNAm probe-specific mediation analysis. As both DACT and HIMA account for correlated CpG sites, HIMA

through its use of single multiple mediation models and DACT through integration of Efron's null empirical null inference framework, these methods may be unable to detect such interacting CpG sites on an individual scale [26,27].

Strengths and Limitations

This study has many strengths. It adds to the limited literature dedicated to investigating epigenetic modification and associated outcomes in LMIC contexts, particularly related to air pollution exposure. Additionally, the mothers and infants involved in the DCHS are of majority Black African or mixed-race ancestry, two populations underrepresented in epigenetic and genetic literature at large. To account for population stratification, which may play a role in DNAm variation, our study incorporated genome-wide genotype data which is the preferred approach to account for genetic ancestry. To our knowledge, this is the first study to investigate DNAm as a mediator of the association between prenatal IAP exposure and neurodevelopment. A unique feature of this study is our usage of three different methods of high-dimensional mediation analysis techniques. As no single high-dimensional mediation analysis method has been deemed the gold standard, it was important to us to employ several methods for high-dimensional mediation analysis in order to compare findings from different methods. Although we did not see good concordance between the methods (Table 2; Table S14-S15), we recognize that the probe-specific methods (HIMA and DACT) are differentially powered. A key advantage of the DACT method is the fact that it is better powered than the joint significance test used in HIMA, which tends to be overly conservative [27]. However, additional research

is needed to better understand why findings from these three mediation methods did not demonstrate substantial overlap.

Of course, our study also has several limitations. Our analyses were constrained by a small sample size (N=142), which may have limited the statistical power to detect mediation effects in neurodevelopment across domains (cognitive, general adaptive behavior, motor function, and language). Our sample size may have also limited statistical power to detect underlying associations (total effects) between prenatal PM_{10} exposure and neurodevelopment for all domains but cognition. As such, our results cannot be extended to non-cognitive neurodevelopmental measures. Secondly, as DNAm signatures are tissue and cell-type specific, our findings are limited in that we investigated DNAm in cord blood and not brain tissue. Although brain tissue DNAm data would have been more appropriate for a study of neurodevelopment, cord blood is far more feasible to collect from living study participants. Lastly, as previously mentioned, there is no gold standard for high dimensional mediation analysis and our results are limited in that there is no standard methodology to apply or with which to compare results.

Conclusion

Differential DNAm was found to significantly mediate the association between prenatal exposure to PM_{10} and cognitive neurodevelopment as measured by BSID-III at two years of age. Twenty-nine differentially methylated CpG sites as well as four differentially methylated gene regions were identified as significant mediators of this association in the DCHS cohort. Due to our small sample size and the general lack of

consensus on a gold standard high dimensional mediation analysis tool in the scientific community, this study should be regarded as a preliminary investigation. To our knowledge, these findings are novel and therefore, further research is necessary to replicate these results in order to better understand how DNAm and other biological pathways can help explain the impact of air pollution exposure on neurodevelopment.

References

[1] Korten I, Ramsey K, Latzin P. Air pollution during pregnancy and lung development in the child. Paediatr Respir Rev 2017;21:38–46. https://doi.org/10.1016/j.prrv.2016.08.008.

[2] Pope DP, Mishra V, Thompson L, Siddiqui AR, Rehfuess EA, Weber M, et al. Risk of low birth weight and stillbirth associated with indoor air pollution from solid fuel use in developing countries. Epidemiol Rev 2010;32:70–81. https://doi.org/10.1093/epirev/mxq005.

[3] Su X, Zhang S, Lin Q, Wu Y, Yang Y, Yu H, et al. Prenatal exposure to air pollution and neurodevelopmental delay in children: A birth cohort study in Foshan, China. Sci Total Environ 2021. https://doi.org/10.1016/j.scitotenv.2021.151658.

[4] Kerin T, Volk H, Li W, Lurmann F, Eckel S, McConnell R, et al. Association Between Air Pollution Exposure, Cognitive and Adaptive Function, and ASD Severity Among Children with Autism Spectrum Disorder. J Autism Dev Disord 2018;48:137–50. https://doi.org/10.1007/s10803-017-3304-0.

[5] Christensen GM, Rowcliffe C, Chen J, Vanker A, Koen N, Jones MJ, Gladish N,

Hoffman N, Donald K, Wedderburn CJ, Kobor MS, Zar H, Stein DJ HA. In-Utero Exposure to Indoor Air Pollution or Tobacco Smoke and Cognitive Development in a South African Birth Cohort Study. SSRN Prepr 2022. https://doi.org/https://dx.doi.org/10.2139/ssrn.4004759.

[6] Talbott EO, Arena VC, Rager JR, Clougherty JE, Michanowicz DR, Sharma RK, et al. Fine particulate matter and the risk of autism spectrum disorder. Environ Res 2015;140:414–20. https://doi.org/10.1016/j.envres.2015.04.021.

[7] Kalkbrenner AE, Windham GC, Serre ML, Akita Y, Wang X, Hoffman K, et al. Particulate matter exposure, prenatal and postnatal windows of susceptibility, and autism spectrum disorders. Epidemiology 2015;26:30–42. https://doi.org/10.1097/EDE.0000000000000173.

[8] Gruzieva O, Xu CJ, Breton C V., Annesi-Maesano I, Antó JM, Auffray C, et al. Epigenome-wide meta-analysis of methylation in children related to prenatal NO2 air pollution exposure. Environ Health Perspect 2017;125:104–10. https://doi.org/10.1289/EHP36.

[9] Gruzieva O, Xu CJ, Yousefi P, Relton C, Merid SK, Breton C V., et al. Prenatal particulate air pollution and DNA methylation in newborns: An epigenome-wide meta-analysis. Environ Health Perspect 2019;127. https://doi.org/10.1289/EHP4522.

[10] Jafta N, Barregard L, Jeena PM, Naidoo RN. Indoor air quality of low and middle income urban households in Durban, South Africa. Environ Res 2017;156:47–56. https://doi.org/10.1016/j.envres.2017.03.008.

[11] Martin WJ, Glass RI, Araj H, Balbus J, Collins FS, Curtis S, et al. Household Air

Pollution in Low- and Middle-Income Countries: Health Risks and Research Priorities. PLoS Med 2013;10:1–8. https://doi.org/10.1371/journal.pmed.1001455.

[12] Tran NQV, Miyake K. Neurodevelopmental Disorders and Environmental Toxicants: Epigenetics as an Underlying Mechanism. Int J Genomics 2017;2017. https://doi.org/10.1155/2017/7526592.

[13] Jaenisch R, Bird A. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. Nat Genet 2003;33:245–54. https://doi.org/10.1038/ng1089.

[14] Hüls A, Wedderburn CJ, Groenewold NA, Gladish N, Jones M, 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 2021;0:1–31. https://doi.org/10.1080/15622975.2021.2016955.

[15] Grayson DR, Guidotti A. The dynamics of DNA methylation in schizophrenia and related psychiatric disorders. Neuropsychopharmacology 2013;38:138–66. https://doi.org/10.1038/npp.2012.125.

[16] van Mil NH, Steegers-Theunissen RPM, Bouwland-Both MI, Verbiest MMPJ, Rijlaarsdam J, Hofman A, et al. DNA methylation profiles at birth and child ADHD symptoms. J Psychiatr Res 2014;49:51–9. https://doi.org/10.1016/j.jpsychires.2013.10.017.

[17] Walton E, Pingault JB, Cecil CAM, Gaunt TR, Relton CL, Mill J, et al. Epigenetic profiling of ADHD symptoms trajectories: A prospective, methylome-wide study. Mol Psychiatry 2017;22:250–6. https://doi.org/10.1038/mp.2016.85.

[18] Mordaunt CE, Jianu JM, Laufer BI, Zhu Y, Hwang H, Dunaway KW, et al. Cord blood DNA methylome in newborns later diagnosed with autism spectrum disorder reflects early dysregulation of neurodevelopmental and X-linked genes. Genome Med 2020;12:1–25. https://doi.org/10.1186/s13073-020-00785-8.

[19] 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.

[20] 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;10. https://doi.org/10.1038/s41398-020-01058-z.

[21] Podobinska M, Szablowska-Gadomska I, Augustyniak J, Sandvig I, Sandvig A, Buzanska L. Epigenetic modulation of stem cells in neurodevelopment: The role of methylation and acetylation. Front Cell Neurosci 2017;11:1–16. https://doi.org/10.3389/fncel.2017.00023.

[22] Laird PW. The power and the promise of DNA methylation markers. Nat Rev Cancer 2003;3:253–66. https://doi.org/10.1038/nrc1045.

[23] Lam LL, Emberly E, Fraser HB, Neumann SM, Chen E, Miller GE, et al. Factors underlying variable DNA methylation in a human community cohort. Proc Natl Acad Sci U S A 2012;109:17253–60. https://doi.org/10.1073/pnas.1121249109.

[24] Olstad EW, Nordeng HME, Gervin K. Prenatal medication exposure and epigenetic outcomes: a systematic literature review and recommendations for

prenatal pharmacoepigenetic studies. Epigenetics 2021;00:1–24. https://doi.org/10.1080/15592294.2021.1903376.

[25] 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. https://doi.org/10.1016/j.csbj.2021.05.042.

[26] 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. https://doi.org/10.1093/bioinformatics/btw351.

[27] 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. https://doi.org/10.1080/01621459.2021.1914634.

[28] Fang R, Yang H, Gao Y, Cao H, Goode EL, Cui Y. Gene-based mediation analysis in epigenetic studies. Brief Bioinform 2021;22:1–11. https://doi.org/10.1093/bib/bbaa113.

[29] 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. https://doi.org/10.1136/thoraxjnl-2014-206242.

[30] DJ Stein, N Koen, KA Donald, CM Adnams, S Koopowitz, C Lund, A Marais B, Myers, A Roos, K Sorsdahl, M Stern, M Tomlinson, C van der Westhuizen B, Vythilingum, L Myer, W Barnett, K Brittain and HZ. Investigating the psychosocial determinants of child health in Africa: The Drakenstein Child Health Study. J

Neurosci Methods 2015. https://doi.org/10.1016/j.jneumeth.2015.03.016.

[31] Morin AM, Gatev E, McEwen LM, Macisaac JL, Lin DTS, Koen N, et al. Maternal blood contamination of collected cord blood can be identified using DNA methylation at three CpGs. Clin Epigenetics 2017;9:1–9. https://doi.org/10.1186/s13148-017-0370-2.

[32] 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. https://doi.org/10.1093/bioinformatics/btu049.

[33] Fortin J-P, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, Greenwood CM HK. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol 2014;15:503. https://doi.org/10.1186/s13059-014-0503-2.

[34] 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. https://doi.org/10.1093/bioinformatics/bts034.

[35] 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. https://doi.org/10.1186/s13148-019-0717-y.

[36] 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. https://doi.org/10.1155/2017/1631760.

[37] 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 J Child Heal 2013;7:54–9. https://doi.org/10.7196/SAJCH.547.

[38] Ranjitkar S, Kvestad I, Strand TA, Ulak M, Shrestha M, Chandyo RK, et al. Acceptability and reliability of the Bayley Scales of infant and Toddler Development-III Among Children in Bhaktapur, Nepal. Front Psychol 2018;9:1–10. https://doi.org/10.3389/fpsyg.2018.01265.

[39] N. B. Bayley scales of infant and toddler development 2006;San Antoni.

[40] Vanker A, Barnett W, Workman L, Nduru PM, Sly PD, Gie RP, et al. Early-life exposure to indoor air pollution or tobacco smoke and lower respiratory tract illness and wheezing in African infants: a longitudinal birth cohort study. Lancet Planet Heal 2017;1:e328–36. https://doi.org/10.1016/S2542-5196(17)30134-1.

[41] 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. https://doi.org/10.1016/j.scitotenv.2015.06.136.

[42] Hüls A, Vanker A, Gray D, Koen N, MacIsaac JL, Lin DTS, et al. Genetic susceptibility to asthma increases the vulnerability to indoor air pollution. Eur Respir J 2020;55:1–9. https://doi.org/10.1183/13993003.01831-2019.

[43] Health NI for OS and. Particulates not otherwise regulated, respirable 1988. www.%0Dcdc.gov/niosh/docs/2003-154/pdfs/0600.pdf.

[44] VanderWeele TJ. Mediation Analysis: A Practitioner's Guide. Annu Rev Public Health 2016;37:17–32. https://doi.org/10.1146/annurev-publhealth-032315-021402.

[45] Pawlowsky-Glahn V, Egozcue JJ. Compositional data and their analysis: An introduction. Geol Soc Spec Publ 2006;264:1–10. https://doi.org/10.1144/GSL.SP.2006.264.01.01.

[46] 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. https://doi.org/10.1080/15622975.2021.2016955.

[47] Drzymalla E, Gladish N, Koen N, Epstein MP, Kobor MS, Zar HJ, et al. Association between maternal depression during pregnancy and newborn DNA methylation. Transl Psychiatry 2021;11:1–8. https://doi.org/10.1038/s41398-021-01697-w.

[48] Fan J, Lv J. Sure independence screening for ultrahigh dimensional feature space. J R Stat Soc Ser B Stat Methodol 2008;70:849–911. https://doi.org/10.1111/j.1467-9868.2008.00674.x.

[49] Hüls A, Robins C, Conneely KN, Edgar R, De Jager PL, Bennett DA, et al. Brain DNA Methylation Patterns in CLDN5 Associated With Cognitive Decline. Biol Psychiatry 2022;91:389–98. https://doi.org/10.1016/j.biopsych.2021.01.015.

[50] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing Author (s): Yoav Benjamini and Yosef

Hochberg Source : Journal of the Royal Statistical Society . Series B (Methodological), Vol . 57 , No . 1 (1995), Publi. J R Stat Soc 1995;57:289–300.

[51] Dustin Tingley Teppei Yamamoto LK, Imai K. Mediation: R Package for Causal Mediation Analysis. R Package version 4.2.4. 2013;59.

[52] E I, V F, F A, M S, G M, M P, et al. Prenatal exposure to PM10 and changes in DNA methylation and telomere length in cord blood. Environ Res 2022:112717. https://doi.org/10.1016/j.envres.2022.112717.

[53] Breton C V, Yao J, Millstein J, Gao L, Siegmund KD, Mack W. Associations with Newborn LINE1 and Alu Methylation and Childhood Blood Pressure and Carotid Intima-Media Thickness in the Children ' s Health Study. Env Heal Perspect 2016;124:1905–12.

[54] Ladd-Acosta C, Feinberg JI, Brown SC, Lurmann FW, Croen LA, Hertz-Picciotto I, et al. Epigenetic marks of prenatal air pollution exposure found in multiple tissues relevant for child health. Environ Int 2019;126:363–76. https://doi.org/10.1016/j.envint.2019.02.028.

[55] Marioni RE, McRae AF, Bressler J, Colicino E, Hannon E, Li S, et al. Meta-analysis of epigenome-wide association studies of cognitive abilities. Mol Psychiatry 2018;23:2133–44. https://doi.org/10.1038/s41380-017-0008-y.

[56] Mulder RH, Neumann A, Cecil CAM, Walton E, Houtepen LC, Simpkin AJ, et al. Epigenome-wide change and variation in DNA methylation in childhood: Trajectories from birth to late adolescence. Hum Mol Genet 2021;30:119–34. https://doi.org/10.1093/hmg/ddaa280.

[57] Xu CJ, Bonder MJ, Söderhäll C, Bustamante M, Baïz N, Gehring U, et al. The emerging landscape of dynamic DNA methylation in early childhood. BMC Genomics 2017;18:1–11. https://doi.org/10.1186/s12864-016-3452-1.

[58] Spiers H, Hannon E, Schalkwyk LC, Smith R, Wong CCY, O'Donovan MC, et al. Methylomic trajectories across human fetal brain development. Genome Res 2015;25:338–52. https://doi.org/10.1101/gr.180273.114.

[59] Hannon E, Schendel D, Ladd-Acosta C, Grove J, Hansen CS, Hougaard DM, et al. Variable DNA methylation in neonates mediates the association between prenatal smoking and birth weight. Philos Trans R Soc B Biol Sci 2019;374. https://doi.org/10.1098/rstb.2018.0120.

[60] Safran M, Rosen N, Twik M, BarShir R, Iny Stein T, Dahary D, Fishilevich S and LD. GeneCards – the human gene database. The GeneCards Suite Chapter 2022;Practical:27–56.

[61] Kazeminasab S, Taskiran II, Fattahi Z, Bazazzadegan N, Hosseini M, Rahimi M, et al. CNKSR1 gene defect can cause syndromic autosomal recessive intellectual disability. Am J Med Genet Part B Neuropsychiatr Genet 2018;177:691–9. https://doi.org/10.1002/ajmg.b.32648.

[62] Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, et al. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature 2011;478:57–63. https://doi.org/10.1038/nature10423.

[63] Reichova A, Zatkova M, Bacova Z, Bakos J. Abnormalities in interactions of Rho GTPases with scaffolding proteins contribute to neurodevelopmental disorders. J Neurosci Res 2018;96:781–8. https://doi.org/10.1002/jnr.24200.

[64] Mathias Uhlén, Linn Fagerberg, Björn M. Hallström, Cecilia Lindskog PO, Adil Mardinoglu, Åsa Sivertsson, Caroline Kampf, Evelina Sjöstedt AA, IngMarie Olsson, Karolina Edlund, Emma Lundberg SN, Cristina Al-Khalili Szigyarto, Jacob Odeberg DD, Jenny Ottosson Takanen, Sophia Hober, Tove Alm, Per-Henrik Edqvist HB, Hanna Tegel, Jan Mulder, Johan Rockberg, Peter Nilsson JMS, et al. Tissue-based map of the human proteome. Science (80-) 2015;347:1260419–1260419. https://doi.org/10.1126/science.1260419.

[65] Fatima S, Wagstaff KM, Lim SM, Polo JM, Young JC, Jans DA. The nuclear transporter importin 13 is critical for cell survival during embryonic stem cell differentiation. Biochem Biophys Res Commun 2021;534:141–8. https://doi.org/10.1016/j.bbrc.2020.11.099.

[66] You P, Peng Z, Wang Y, Tao T. Expression and subcellular distribution of imp13 are regulated in brain development. Vitr Cell Dev Biol - Anim 2013;49:346–53. https://doi.org/10.1007/s11626-013-9599-z.

[67] Naumova OY. Down syndrome Down syndrome 2021.

[68] Han S, Nam J, Li Y, Kim S, Cho SH, Cho YS, et al. Regulation of dendritic spines, spatial memory, and embryonic development by the TANC family of PSD-95-interacting proteins. J Neurosci 2010;30:15102–12. https://doi.org/10.1523/JNEUROSCI.3128-10.2010.

[69] Nimchinsky EA, Sabatini BL, Svoboda K. Structure and function of dendritic spines. Annu Rev Physiol 2002;64:313–53. https://doi.org/10.1146/annurev.physiol.64.081501.160008.

[70] Warre-Cornish K, Perfect L, Nagy R, Duarte RRR, Reid MJ, Raval P, et al.

Interferon-γ signaling in human iPSC–derived neurons recapitulates neurodevelopmental disorder phenotypes. Sci Adv 2020;6:1–17. https://doi.org/10.1126/sciadv.aay9506.

[71] Bilousova T, Dang H, Xu W, Gustafson S, Jin Y, Wickramasinghe L, et al. Major histocompatibility complex class I molecules modulate embryonic neuritogenesis and neuronal polarization. J Neuroimmunol 2012;247:1–8. https://doi.org/10.1016/j.jneuroim.2012.03.008.

[72] Needleman LA, Liu XB, El-Sabeawy F, Jones EG, McAllister AK. MHC class I molecules are present both pre- and postsynaptically in the visual cortex during postnatal development and in adulthood. Proc Natl Acad Sci U S A 2010;107:16999–7004. https://doi.org/10.1073/pnas.1006087107.

[73] Ali A, Bouma GJ, Anthony R V., Winger QA. The role of LIN28-let-7-ARID3B pathway in placental development. Int J Mol Sci 2020;21. https://doi.org/10.3390/ijms21103637.

[74] Chen C, Chen C, Xue G, Dong Q, Zhao L, Zhang S. Parental warmth interacts with several genes to affect executive function components: A genome-wide environment interaction study. BMC Genet 2020;21:1–11. https://doi.org/10.1186/s12863-020-0819-8.

[75] Schulte EC, Ellwanger DC, Dihanich S, Manzoni C, Stangl K, Schormair B, et al. Rare variants in LRRK1 and Parkinson's disease. Neurogenetics 2014;15:49–57. https://doi.org/10.1007/s10048-013-0383-8.

[76] Nordgren M, Fransen M. Peroxisomal metabolism and oxidative stress. Biochimie 2014;98:56–62. https://doi.org/10.1016/j.biochi.2013.07.026.

[77] Berger J, Dorninger F, Forss-Petter S, Kunze M. Peroxisomes in brain development and function. Biochim Biophys Acta - Mol Cell Res 2016;1863:934–55. https://doi.org/10.1016/j.bbamcr.2015.12.005.

[78] Almeida A. Regulation of APC/C-Cdh1 and its function in neuronal survival. Mol Neurobiol 2012;46:547–54. https://doi.org/10.1007/s12035-012-8309-2.

[79] Zhou Z, He M, Shah AA, Wan Y. Insights into APC/C: From cellular function to diseases and therapeutics. Cell Div 2016;11:1–18. https://doi.org/10.1186/s13008-016-0021-6.

[80] Li M, Shin YH, Hou L, Huang X, Wei Z, Klann E, et al. The adaptor protein of the anaphase promoting complex Cdh1 is essential in maintaining replicative lifespan and in learning and memory. Nat Cell Biol 2008;10:1083–9. https://doi.org/10.1038/ncb1768.

[81] Delgado-Esteban M, García-Higuera I, Maestre C, Moreno S, Almeida A. APC/C-Cdh1 coordinates neurogenesis and cortical size during development. Nat Commun 2013;4. https://doi.org/10.1038/ncomms3879.

[82] Wu Y, Lin X, Lim IY, Chen L, Teh AL, MacIsaac JL, et al. Analysis of two birth tissues provides new insights into the epigenetic landscape of neonates born preterm. Clin Epigenetics 2019;11:1–12. https://doi.org/10.1186/s13148-018-0599-4.

[83] Li C, Cao M, Zhou X. Role of epigenetics in parturition and preterm birth. Biol Rev 2021;10. https://doi.org/10.1111/brv.12825.

[84] Maxwell JR, Yellowhair TR, Oppong AY et al. Cognitive development in preterm infants: multifaceted deficits reflect vulnerability of rigorous neurodevelopmental

pathways. Minerva Pediatr 2017;69:298–313. https://doi.org/10.23736/s0026-4946.17.04905-2.

[85] Dempster EL, Pidsley R, Schalkwyk LC, Owens S, Georgiades A, Kane F, et al. Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. Hum Mol Genet 2011;20:4786–96. https://doi.org/10.1093/hmg/ddr416.

[86] Haines BP, Rigby PWJ. Expression of the Lingo/LERN gene family during mouse embryogenesis. Gene Expr Patterns 2008;8:79–86. https://doi.org/10.1016/j.modgep.2007.10.003.

[87] Andrews JL, Fernandez-Enright F. A decade from discovery to therapy: Lingo-1, the dark horse in neurological and psychiatric disorders. Neurosci Biobehav Rev 2015;56:97–114. https://doi.org/10.1016/j.neubiorev.2015.06.009.

[88] Chowdhury F, Wang L, Al-Raqad M, Amor DJ, Baxová A, Bendová Š, et al. Haploinsufficiency of PRR12 causes a spectrum of neurodevelopmental, eye, and multisystem abnormalities. Genet Med 2021;23:1234–45. https://doi.org/10.1038/s41436-021-01129-6.

[89] Leduc MS, Mcguire M, Madan-Khetarpal S, Ortiz D, Hayflick S, Keller K, et al. De novo apparent loss-of-function mutations in PRR12 in three patients with intellectual disability and iris abnormalities. Hum Genet 2018;137:257–64. https://doi.org/10.1007/s00439-018-1877-0.

[90] Kalkan Z, Durasi IM, Sezerman U, Atasever-Arslan B. Potential of GRID2 receptor gene for preventing TNF-induced neurodegeneration in autism. Neurosci Lett 2016;620:62–9. https://doi.org/10.1016/j.neulet.2016.03.043.

[91] Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. Genome Res 2002;12:996–1006. https://doi.org/10.1101/gr.229102.

[92] Arbones ML, Thomazeau A, Nakano-Kobayashi A, Hagiwara M, Delabar JM. DYRK1A and cognition: A lifelong relationship. Pharmacol Ther 2019;194:199–221. https://doi.org/10.1016/j.pharmthera.2018.09.010.

Tables and Figures

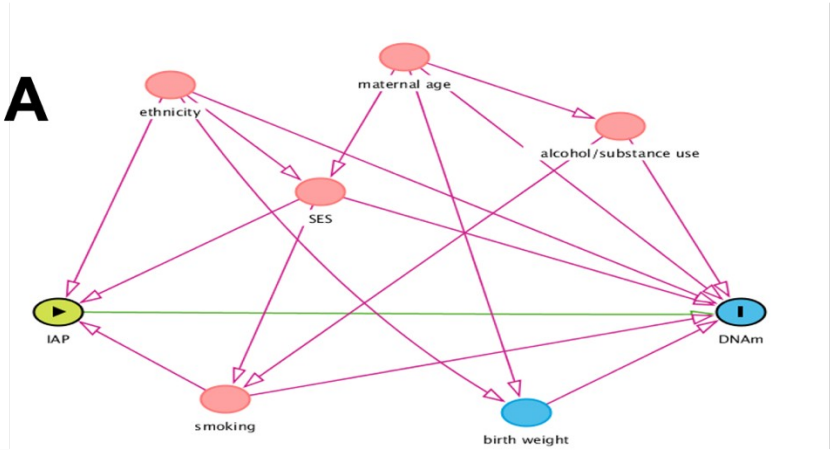
Table 1. Study Characteristics of the DCHS participants.

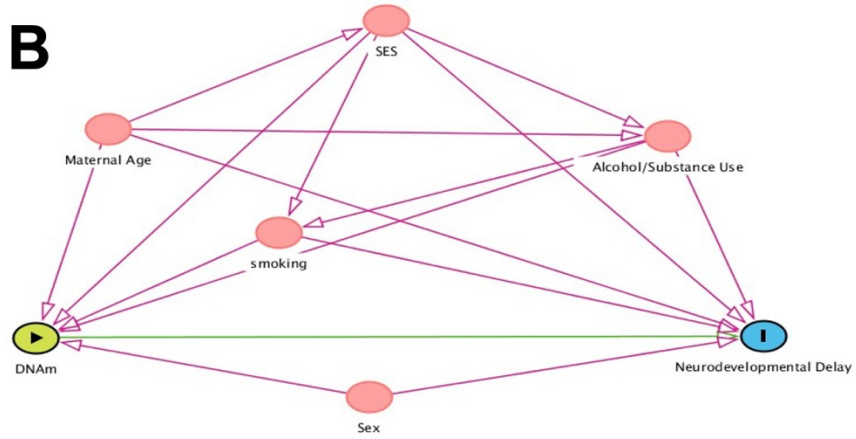
Characteristic	Participants (N=142) *
Race	
Black	69 (48.6%)
Mixed	73 (51.4%)
Sex	
Female	57 (40.1%)
Male	85 (59.9%)
Alcohol Use	
No Exposure	114 (80.3%)
Exposure	28 (19.7%)
Smoking[1]	
Non-Smoker	38 (26.8%)
Passive Smoker	57 (40.1%)
Active Smoker	47 (33.1%)
Birth Weight (grams)	3120 (512)
Maternal Age (years)	27.20 (5.98)
SES Score[2]	0.092 (2.25)
Bayley Score – Cognitive Composite	85.14 (8.65)
Bayley's Score – Language Composite	84.31 (12.22)
Bayley's Score – Motor Composite	94.04 (13.74)
Bayley's Score – General Adaptive Composite	83.72 (13.29)
PM10 Concentration ($\mu g/m^3$)	64.6 (96.8)

*The number of participants includes those with data available for all covariates and a composite Bayley's score in at least one of the relevant domains (Cognitive, General Adaptive, Language, and Motor)

[1]Smoking status was determined by urine cotinine levels (ng/ml), defined as follows: < 10 ng/ml non-smoker, 10-49 ng/ml passive-smoker, >= 500 ng/ml active smoker

[2]SES score description can be found in the supplement section


A
maternal age
ethnicity
alcohol/substance use
SES
IAP
DNAm
smoking
birth weight


B
SES
Maternal Age
Alcohol/Substance Use
smoking
DNAm
Neurodevelopmental Delay
Sex

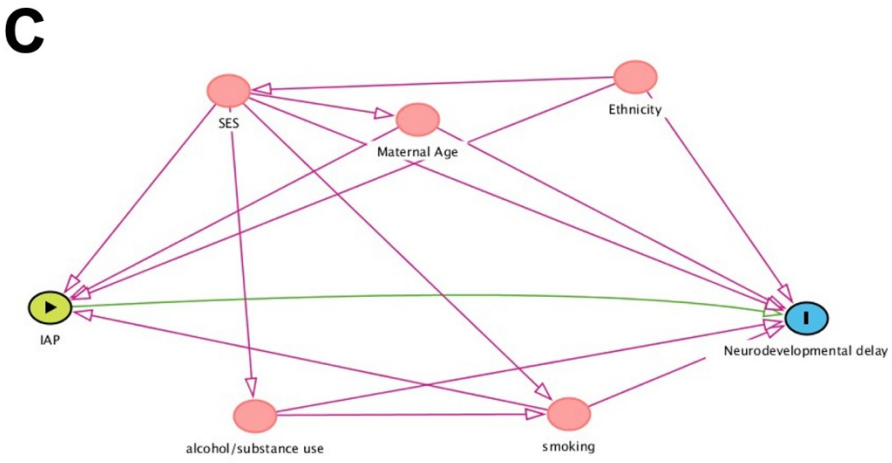

C
SES
Ethnicity
Maternal Age
IAP
Neurodevelopmental delay
alcohol/substance use
smoking

Figure 1. Directed acyclic graphs (DAGs) for exposure-mediator, mediator-outcome, and exposure-outcome associations. Below are DAGs for each causal pathway relevant to our mediation analysis.

Table 2. High-dimensional mediation analysis for the association between PM_{10} (exposure), DNAm (mediator) and neurodevelopment (outcome) using DACT for CpG sites identified via formal causal mediation analysis with a significant indirect effect and total effect. This table includes raw and corrected p-values from the DACT method for CpG sites identified across domains restricted to those for which we found with nominally significant indirect effects and total effects (domain-dependent) via causal mediation analysis following high-dimensional mediation analysis (BH FDR <= 0.05).

				Cognition		Language[ii]		General Adaptive[ii]	
	Chr	Pos	Closest Gene	Raw P-Value	FDR	Raw P-Value	FDR	Raw P-Value	FDR
cg13690126	1	26502524	CNKSR1	0.000129	0.0033	0.00035	0.0056	-	-
cg07070893	1	44412074	RP11-7O11.3	0.00042	0.0075	-[i]	-	-	-
cg23560546	2	159652019	DAPL1	0.00072	0.011	-	-	-	-
cg15007548	2	159950276	TANC1	0.0037	0.039	-	-	-	-
cg16975959	2	223184510	AC010980.2	0.0018	0.022	-	-	-	-
cg00694520	2	223916687	KCNE4	0.0043	0.044	-	-	-	-
cg08967927	5	30346164	RP11-136H13.2	4.29E-05	0.0014	-	-	-	-
cg22572779	6	10434761	RP1-290I10.7	0.00045	0.0077	-	-	-	-
cg15074838	6	32406521	HLA-DRA	0.00026	0.0055	-	-	-	-
cg26668632	6	137540814	IFNGR1	0.00029	0.0055	-	-	-	-
cg01550799	7	114561804	MDFIC	0.0018	0.022	-	-	-	-
cg25405984	10	31074039	ZNF438	0.0029	0.031	-	-	-	-
cg25544551	12	104531891	NFYB	1.51E-05	0.00064	-	-	-	-
cg19187486	15	74833838	ARID3B	3.30E-13	5.87E-11	-	-	-	-
cg05796992	15	101547339	LRRK1	4.39E-05	0.0014	-	-	-	-
cg06233301	16	1500359	CLCN7	0.0023	0.027	-	-	-	-
cg01085137	16	48278085	ABCC11	0.0011	0.015	-	-	-	-
cg23989635	16	68771203	CDH1	0.0011	0.015	-	-	-	-
cg26756782	16	89922218	SPIRE2	5.90E-05	0.0016	-	-	-	-
cg26894552	17	5323715	RPAIN	0.0026	0.029	-	-	-	-
cg13465852	17	27942115	CORO6	2.36E-10	2.10E-08	-	-	-	-
cg10159032	17	30347989	LRRC37B	0.00022	0.0053	-	-	-	-
cg14714923	17	43238186	HEXIM2	2.34E-07	1.66E-05	-	-	-	-
cg15000966	17	73128123	NT5C	1.62E-05	0.00064	-	-	-	-
cg14243899	17	74580397	ST6GALNAC2	0.00090	0.013	-	-	-	-
cg23054321	19	2289682	LINGO3	3.36E-07	1.99E-05	-	-	6.97E-09	6.76E-07
cg05023582	19	50093541	PRRG2	2.42E-11	2.87E-09	-	-	-	-
cg03234186	19	58220657	ZNF551	8.94E-16	3.18E-13	3.05E-26	1.26E-23	-	-
cg00660655	21	46368151	FAM207A	0.00072	0.011	-	-	-	-

Missingness in this table is resultant of our DACT filtering process which is described in more detail in the methods section.

35 CpG sites were considered for multiple testing correction in the cognition domain, 45 in the language domain,

39 in the general adaptive behavior domain, and 13 in the motor function domain.

[ii] No CpG sites with both a significant indirect effect and significant total effect were identified in the language, motor, or general adaptive behavior domains. Two significant CpG sites identified in the cognitive domain overlap with two in the language domain and one in the general adaptive domain and therefore, information two CpG sites identified with DACT are depicted for the language domain and one for the general adaptive domain alongside the cognition domain.

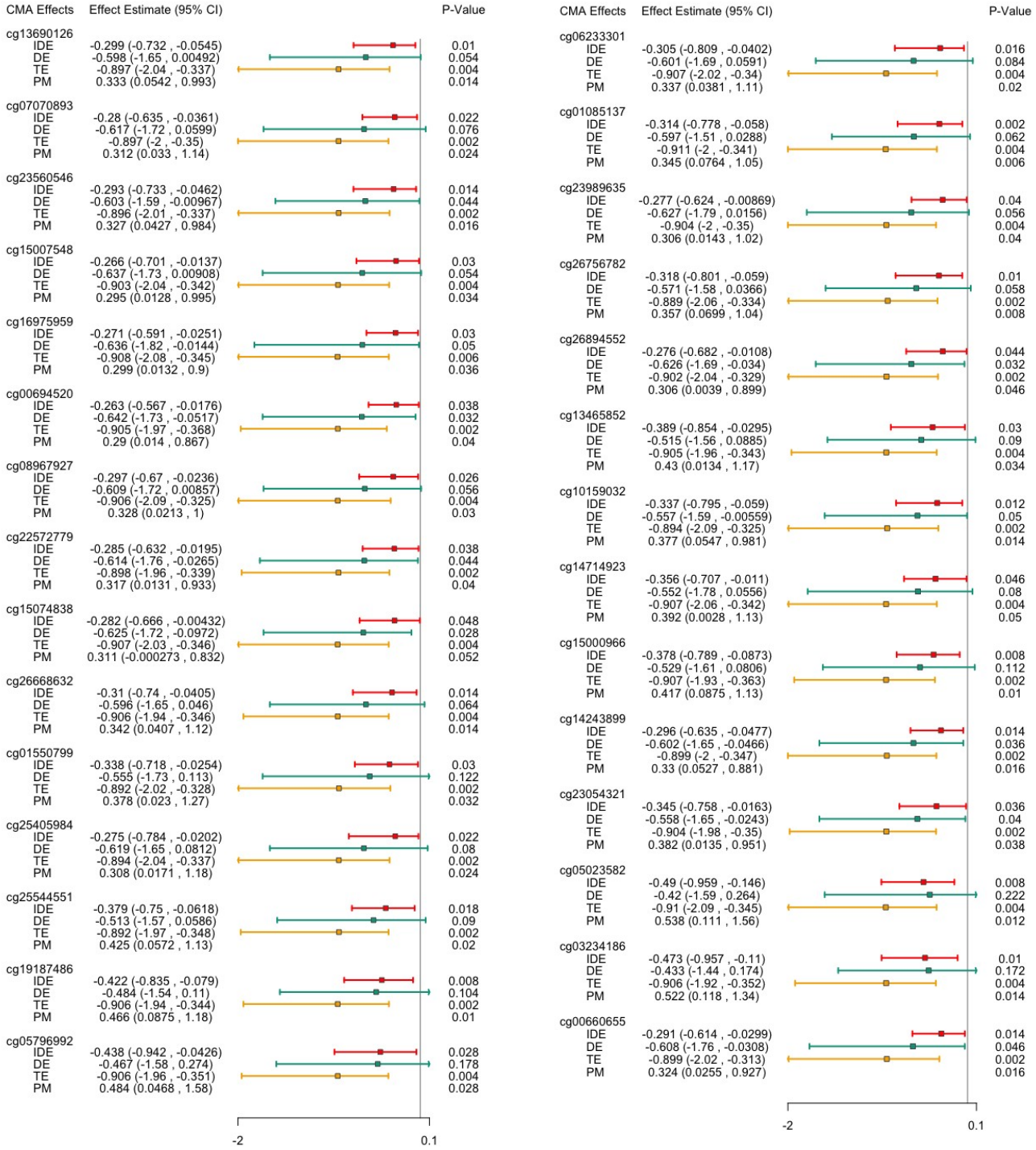


Figure 2. Causal mediation analysis for the association between PM_{10} (exposure) and cognitive neurodevelopment (outcome) using CpG sites identified with high-dimensional mediation analysis methods. This figure presents estimates for indirect

effect (IDE), direct effect (DE), total effect (TE), and proportion mediated (PM) for CpG sites with both significant IDE and TE. Significant total effects were found only for the cognitive neurodevelopmental domain (no significant total effects were identified for general adaptive behavior, language, or motor neurodevelopmental domains). IDE, DE, and TE effect estimates have been multiplied by the PM_{10} interquartile range (IQR) (58.78 mug/m3) observed in this cohort and thus these effect estimates represent the estimated effects on BSID-III cognitive score per one IQR increase in prenatal PM_{10} exposure.

Table 3. Gene-based High-dimensional mediation analysis for the association between PM_{10} (exposure), DNAm (mediator) and neurodevelopment (outcome) using gHMA. Presented associations were significant for at least one neurodevelopmental domain (BH FDR <0.05). FDR values with bold values indicate significance at a threshold of 0.05.

		Cognitive		General Adaptive Behavior		Language		Motor	
Closest Gene	**CpG Count**	**Raw P-Value**	**FDR**	**Raw P-Value**	**FDR**	**Raw P-Value**	**FDR**	**Raw P-Value**	**FDR**
A. Significant mediation for the primary outcome (cognitive domain) and cross-validation across domains									
GOPC	15	1.19e-09	**4.12e-05**	9.73e-03	0.83844	8.57e-03	0.98395	3.59e-06	**0.0381**
RP11-74K11.1	20	1.13e-07	**1.95e-03**	1.13e-07	**0.00391**	1.05e-07	**0.00363**	5.48e-02	0.8878
DYRK1A	19	1.21e-06	**1.39e-02**	1.20e-06	**0.02080**	1.49e-06	**0.02569**	3.78e-03	0.6598
RNMT	8	2.86e-06	**2.47e-02**	2.87e-06	**0.03303**	8.94e-06	0.09061	1.22e-02	0.8332
B. Significant mediation for the secondary outcomes (other domains) and cross-validation across domains									
DCAF13	7	8.92e-05	1.60e-01	8.86e-05	0.17015	3.65e-04	0.34329	6.52e-06	**0.0381**
TNN	13	4.44e-04	3.49e-01	4.40e-04	0.34394	1.86e-04	0.30610	8.11e-07	**0.0280**
TAL1	26	4.61e-03	8.13e-01	1.72e-03	0.57783	9.60e-03	0.99984	4.06e-06	**0.0381**
AC011648.1	3	1.48e-02	1.00e+00	1.49e-02	0.91517	1.05e-02	0.99984	5.76e-06	**0.0381**
SPINK2	5	9.85e-02	1.00e+00	4.14e-01	0.99987	4.52e-01	0.99984	6.62e-06	**0.0381**