Prenatal diet buffers infant epigenetic changes linked to pollution and transient wheeze

Samantha Lee[1], Chaini Konwar[2], Robert Balshaw[3], Julia L. MacIsaac[2], Katia Ramadori[2], David T.S. Lin[2], Oscar Urtatiz[2], Kaja Z. LeWinn[5], Catherine J. Karr[6,7,8], Alicia K. Smith[9,10], Michael S. Kobor[2,11], Kecia N. Carroll[12], Nicole R. Bush[13,14], Meaghan J. Jones[1]

1. Department of Biochemistry and Medical Genetics, University of Manitoba, and Children's Hospital Research Institute of Manitoba, Winnipeg, MB
2. Department of Medical Genetics, University of British Columbia, Vancouver, BC
3. College of Community and Global Health, University of Manitoba, Winnipeg, MB
4. University of Tennessee Health Science Center, Memphis, TN
5. Department of Psychiatry and Behavioral Sciences, University of California, San Francisco, CA
6. Department of Epidemiology, University of Washington, Seattle, WA
7. Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA
8. Department of Pediatrics, University of Washington, Seattle, WA
9. Department of Gynecology and Obstetrics, Emory University School of Medicine, Atlanta, GA
10. Department of Psychiatry and Behavioral Science, Emory University School of Medicine, Atlanta, GA
11. BC Children's Hospital Research Institute, Vancouver, BC
12. Departments of Pediatrics and Environmental medicine, The Icahn School of Medicine at Mount Sinai, New York, Ny
13. Department of Pediatrics, School of Medicine, University of California, San Francisco, CA
14. Department of Psychiatry and Behavioral Sciences, School of Medicine, University of California, San Francisco, CA.

Corresponding author: Dr. Meaghan Jones
Email: Meaghan.Jones@umanitoba.ca

Author Contributions: SL conducted statistical analyses and wrote the first draft of the manuscript, CK performed data quality control, preprocessing, and estimation of cell type and genetic ancestry principal components, JLM, KR, DTSL, and OU performed nucleic acid extraction and microarray analysis of cord blood samples, RB provided guidance in statistical analysis. CJK estimated air pollutant exposures. NRB, KZL, KNC, and AKS supervised and coordinated sample and data collection and management in the CANDLE study. AKS, AJK, MSK, KNC and NRB provided guidance on study design and data analysis and reviewed manuscript. MJJ conceptualized the study, supervised data analysis, investigation, and reviewed and edited manuscript. All authors reviewed the final manuscript.

Competing Interest Statement: The authors have no competing interests to declare.

Keywords: DNA methylation, cord blood, air pollution, maternal diet

Abstract

Prenatal air pollution exposure is associated with childhood asthma, particularly among biological males. The mechanisms remain unclear, but may involve lasting epigenetic changes, such DNA methylation (DNAm), that occur during gestation in response to oxidative stress and inflammation. Higher maternal intake of “protective” micronutrients, like antioxidants, could buffer pollution-induced oxidative stress and inflammation to mitigate potentially adverse DNAm differences contributing to asthma. Using data from 515 CANDLE participants, we examined associations between prenatal NO_2, $PM_{2.5}$, and PM_{10} and cord blood DNAm, evaluated DNAm mediation of pollution associations with childhood wheeze phenotypes (transient, persistent, and late-onset), and assessed buffering of DNAm by maternal polyunsaturated fatty acid, vitamin C, or folate intake, and overall diet quality measured by the Alternative Healthy Eating Index-Pregnancy (AHEI-P). We identified 19, seven, and five regional DNAm differences associated NO_2, $PM_{2.5}$, and PM_{10}. Mediation analyses suggested a role for *HLA-DPA1/DPB1* DNAm in NO_2 and $PM_{2.5}$ associations with transient wheeze. To assess buffering, we fit pollutant-by-diet interaction models, defining buffering as an interaction opposite in sign to the main pollutant effect. One or more micronutrients or AHEI-P attenuated pollutant effects at 16 of 19 NO_2-associated DMRs, including *HLA-DPA1/DPB1*, and all $PM_{2.5}$- and PM_{10}-associated DMRs. However, attenuation of *HLA-DPA1/DPB1* DNAm did not significantly reduce the indirect effect of NO_2 on transient wheeze. In sex-stratified analyses, biological males exhibited lower $PM_{2.5}$-associated DNAm in *SERPINB9*, a gene linked to lung function. These findings suggest prenatal air pollution alters DNAm, which may contribute to transient wheeze, with some differences partially buffered by maternal diet.

Significance Statement

Prenatal air pollution exposure contributes to child wheeze and asthma, potentially through the oxidative stress response and subsequent changes to infant DNA methylomes. Here, we used data from the CANDLE cohort to identify cord blood DNAm differences associated with NO_2, $PM_{2.5}$, or PM_{10}. We examined if any alterations mediated the relationship between prenatal air pollution exposures and transient, persistent, or late-onset wheeze at age 4 to 6 years. Some of these DNAm differences appeared to be at least partially buffered by maternal micronutrients and/or overall diet quality.

Introduction

Ambient air pollution is a heterogeneous mixture of gaseous and particulate matter (PM) compounds and is a widespread environmental health hazard. PM is further classified according by aerodynamic diameter, with particles <2.5µm ($PM_{2.5}$) and <10µm (PM_{10}) receiving the most attention as they can penetrate deep into the lungs and enter the periphery(1, 2). Here, transition metals on the surface of PM can catalyze oxidative reactions(3) and activate antioxidant and inflammatory pathways(4, 5), resulting in oxidative stress and inflammation. Nitrogen dioxide (NO_2) is frequently used as a surrogate measure for overall air pollution exposure in epidemiological studies, but it also directly contributes to oxidative stress and inflammation, especially within the lungs(6, 7), where it oxidizes proteins and lipids, depletes antioxidants, activates cellular pathways involved in cell differentiation and death, and increases the release of inflammatory cytokines(7, 8). These effects can increase susceptibility to, or exacerbate, inflammatory conditions such as asthma(9, 10). Numerous investigations have reported positive associations between prenatal exposures to PM and NO_2 air pollution and childhood asthma, with biological males appearing to be at especially high risk(11–13).

Childhood asthma is a complex condition in which wheezing is a common phenotype. Wheeze in children can be classified as no/infrequent wheeze, transient wheeze, persistent wheeze, and late onset wheeze(14). A systematic review identified a limited number of studies investigating the relationship between prenatal air pollution exposures and specific wheeze phenotypes, with some reporting associations with transient, persistent, and late-onset wheeze; however, findings varied across studies, likely due to differences in statistical methodology, wheeze and asthma definitions, and/or genetics(15). Additionally, these analyses did not include sex-stratified investigations, so it remains unclear whether susceptibility to these phenotypes differs by sex following prenatal air pollution exposure. While the exact molecular mechanisms connecting prenatal air pollution exposure to child wheeze and asthma remain unclear, growing evidence suggests it may operate through the biological embedding of *in utero* and early life oxidative stress and inflammation.

Biological embedding refers to a form of cellular “memory” of past experiences or exposures that can influence later biological responses and ultimately, health outcomes(16). This process is hypothesized to involve epigenetic modifications, including DNA methylation (DNAm) (16, 17), which refers to the addition of a methyl group to 5’ position of a cytosine in a cytosine-guanine dinucleotide (CpG). Previous investigations have identified relationships between prenatal air pollution exposures, altered DNAm at birth, and child health outcomes. For example, one study linked prenatal $PM_{2.5}$ exposure to altered *UBE2S* DNAm at birth, with this DNAm difference positively correlated to asthma at six years of age(18). A meta-analysis reported associations between DNAm of the asthma-associated genes *NOTCH4* and *FAM13A* in cord

blood, prenatal PM_{10} exposure, and gene expression in peripheral blood during adolescence(19). More recently, we identified *TRIM41* DNAm differences associated with prenatal NO_2 exposure that may contribute to atopy in male children within the CHILD study(20). Collectively, these studies have increased our understanding of the molecular pathways connecting prenatal air pollution exposure to child health outcomes, while highlighting key opportunities for more comprehensive investigation. Specifically, previous studies have not examined how air pollution-associated DNAm differences relate to distinct wheeze phenotypes, leaving important heterogeneity in respiratory outcomes unexplored. Additionally, most participants in these studies were mainly of European ancestry and from socioeconomically advantaged populations, potentially limiting the generalizability of findings to more diverse populations who may have higher asthma-associated prevalence and morbidity(21) and underscoring the need to determine which DNAm alterations are robust across human populations. Finally, no studies have explored potential buffers to reduce or prevent the effects of prenatal air pollutants on cord blood DNAm, representing an opportunity to help inform interventions aimed at lowering childhood asthma risk.

Given that oxidative stress and resulting DNAm differences appear to play an important role in mediating the health effects of prenatal air pollution exposure(22, 23), reducing oxidative stress could help prevent or attenuate DNAm alterations, which could weaken the association between prenatal air pollution exposure and childhood wheeze and asthma by reducing DNAm-driven disruptions in cellular function(24). Dietary antioxidants such as vitamin C can counteract oxidative stress, but are themselves depleted after prolonged exposure to oxidizing agents. For example, cigarette smoke (a type of air pollutant) is associated with lower plasma vitamin C levels in both smokers and infants born to mothers who smoked during pregnancy(24). A single-blind crossover intervention trial demonstrated that B-vitamin (folic acid, B6, B12) supplementation attenuates increases in inflammatory white blood cells after $PM_{2.5}$ exposure in adults(25). In the CANDLE (Conditions Affecting Neurocognitive Development and Learning in Early Childhood) cohort, higher maternal plasma folate attenuated the negative association between prenatal PM_{10} exposure and child IQ measured at ages four to six (26). In another study, maternal vitamin C supplementation mitigated cord blood DNAm differences associated with prenatal cigarette smoking(27), with improved child respiratory outcomes, with higher lung function and decreased wheeze at age five years compared to the placebo group(28). Beyond individual nutrients, overall maternal diet quality may also influence oxidative stress and related biological pathways. One approach to measuring maternal diet quality is through the Alternate Healthy Eating Index for Pregnancy (AHEI-P), which captures intake of foods and micronutrients associated with lower chronic disease risk and those relevant to maternal and fetal health, including vitamin C, polyunsaturated fats, and folate(29). Despite this evidence, no study to date has delineated how overall maternal diet quality or micronutrient intake may modify the

relationship between maternal prenatal air pollution exposure, infant DNAm at birth, and child respiratory outcomes including wheeze and asthma.

The present study examines whether maternal intake of protective micronutrients or overall diet quality can buffer the effects of prenatal air pollution exposure on cord blood DNA methylation (DNAm) and, in turn, influence the relationship between air pollution and childhood wheeze phenotypes at four years of age. Using the socioeconomically and racially diverse CANDLE cohort, we first conducted our main epigenome-wide analyses to identify individual CpG and regional DNAm differences associated with prenatal exposure to NO_2, $PM_{2.5}$, and PM_{10}. We conducted secondary analyses, repeating the main analyses with biological sex interaction terms to explore whether DNAm differences varies between males and females to provide greater insight into the increased susceptibility of male children to wheeze and asthma with prenatal air pollution exposure. We also conducted secondary analyses, including air interaction terms for maternal intake of vitamin C, folate, or polyunsaturated fatty acids or overall maternal diet quality measured by AHEI-P score, to examine buffer buffering of DNAm differences identified in our main analysis. Finally, for DNAm differences that were associated with wheeze phenotypes, we explored the significance of the buffering effect of maternal micronutrient intake and overall diet quality using moderated mediation models.

Materials and Methods

Study participants

The Conditions Affecting Neurocognitive Development and Learning in Early Childhood (CANDLE) study is a U.S.-based observational and longitudinal birth cohort(30). A total of 1503 eligible pregnant women were recruited from Shelby County, Tennessee between 2006 and 2011. The CANDLE study was approved by review boards at the University of Tennessee Health Science Center (UTHSC) and three hospitals where eligible participants could deliver (06-08495-FB). This study uses a subset (N=515) of CANDLE participants with complete DNAm, genotyping, prenatal NO_2, $PM_{2.5}$, and PM_{10} exposure estimates, and covariate data and was approved by the University of Manitoba Bannatyne Research Ethics Board (HS23413). Covariates were selected *a priori* for inclusion in linear models to account for biological variation and included maternal prenatal tobacco use, adjusted household income, biological sex, gestational age, genetic ancestry, and estimated cell types. See ***SI Methods*** for a list of inclusion/exclusion criteria, information on genotyping preprocessing, quality control, and generation of genetic ancestry principal components (PCs), cell type estimations and PCs, and covariate descriptions.

Prenatal air pollution exposure estimates

This study examined the separate effects of estimated prenatal exposure to three air pollution components: NO_2, $PM_{2.5}$, and PM_{10}. Prenatal NO_2 and $PM_{2.5}$ exposure estimates were calculated using a spatiotemporal extension of the universal kriging model at the individual address level that incorporated monitoring data from local monitors and regulatory networks, and land-use characteristics(31, 32). Average pregnancy period NO_2 and $PM_{2.5}$ exposures were obtained by averaging region-specific biweekly estimates over the pregnancy period (date of conception to date of birth). Prenatal PM_{10} exposure estimates were calculated in a similar manner(33). However, PM_{10} models used a fixed year (2006) for all estimates to avoid temporal confounding or bias related to year-specific modelling errors. Address changes were accounted for using a time-weighted exposure average of reported addresses.

DNA methylation measurement, preprocessing, and quality control

Umbilical cord blood was collected at birth and processed to obtained cord blood mononuclear cells (CBMCs), which were stored at -80C until use. Genomic DNA was purified using the Qiagen Blood and Tissue Kit (Qiagen, USA), and bisulfite-converted using the EZ-96 DNA Methylation kit (Zymo Research, Irvine, CA, and subsequently hybridized to the Illumina MethylationEPIC array. All nucleic acid procedures followed manufacturer protocol. Multiple R

(v3.5) packages including *ewastools(34, 35)*, *minfi(36)*, and *conumee(37)* were used to perform sample quality checks quality control checks. Briefly, these tools were applied to assess technical parameters, to determine concordance between reported and estimated biological sex, to confirm clustering of samples based on the median intensities in both the methylated and unmethylated channels, to check maternal blood contamination(38), to assess sample identify using SNP probes and to remove samples that have bad detection p-values in >1% of probes, and <3 beads contributing to DNAm signal in >1% of probes. Afterwards, DNAm data were corrected for background noise and normalized for dye-bias using *preprocessNoob* from the *minfi* R package(36), followed by Beta-Mixture Quantile (BMIQ) normalization(39) from the *watermelon* R package(40) to adjust for type I-type II probe bias(36, 40). Poorly performing probes (detection p-value >0.01), probes with <3 beads in >5% of samples, probes on sex chromosomes, cross-hybridizing probes, polymorphic probes, and missing probes were removed, leaving 782,881 probes(41). Lastly, *combat* from the *sva* R package(42) was used to remove batch effects related to chip, chip position, and plate.

Age four to six years health outcomes

This study investigated the role of air-pollution induced DNAm alterations in wheeze phenotypes. Wheeze phenotypes at ages four to six years were created based on findings from the Tucson Children's Respiratory Study and included: never, transient, persistent, and late-onset wheeze(14). Participants with never wheeze were defined as those with no reports of wheeze at the year three or year four-six visit. Given that atopic dermatitis and childhood wheeze and asthma are thought to have a shared etiology(43), we also excluded participants that reported atopic dermatitis from the never wheeze group. See ***SI Methods*** for definitions of wheeze used in this study

Maternal diet

Mothers completed the Block Food Frequency Questionnaire (FFQ) at intake (16 to 26 weeks gestation) and responses were processed by NutritionQuest (Berkley, California) to quantify micronutrient intake, serving size, and frequency of food item consumption. This information was subsequently used to calculate Alternative Healthy Eating Index for Pregnancy (AHEI-P) scores(29). Participants reporting very low (<1000 kcal/day) or very high (>5000 kcal/day) daily energy intake were excluded from further analysis(44).

Statistical analysis

All statistical analyses were performed using R (v4.2-4.5).

Individual and regional DNAm changes. Epigenome-wide analyses were conducted using robust multivariable linear regressions. DMRs were identified by spatially-adjusting p-values obtained from epigenome-wide analyses using the *combp* function(45) from the *missMethyl* R package with seed equal to 0.01. DMRs were excluded from further analysis if they contained fewer than 3 CpGs or if composite CpGs within the DMR exhibited inconsistent directionality in effect sizes.

Association with health outcomes. Casual mediation analysis was performed using the *mediation* R package(46) with a quasi-Bayesian approach and 1000 iterations. Health outcomes were modelled as 0 (control) or 1 (affected) in multivariable linear models examining the total effect of prenatal air pollution exposure and treatment. Multivariable linear models also examined the effect of prenatal air pollution exposure on DNAm in mediation analysis. Treatment and control values were defined as mean ± standard deviation of air pollution exposure at the study population level.

Buffering by maternal diet micronutrients and overall quality. Robust linear multivariable regressions investigated buffering of DNAm differences. Specifically, an interaction term was included between each air pollutant, respectively, and vitamin C, the ratio of n-6 to n-3 PUFA consumption, folate, or AHEI-P score. We considered three potential indicators of buffering by diet: (1) Re-identification of DMRs from the main analysis within the interaction terms of the diet-adjusted models, suggesting these DNAm differences may depend on maternal nutrient intake or overall diet quality; (2) The emergence of new DMRs in the main air pollutant term of the diet interaction models, which may indicate DNAm differences that were previously masked by maternal diet in the primary analysis; (3) New or stronger causal mediation effects for DMRs identified in the main analysis that emerged only after including an interaction term between air pollutants and maternal intake of vitamin C, PUFAs, folate, or AHEI-P score (47). For DMRs exhibiting new or stronger causal mediation effects, we attempted to quantify the significance of buffering using the *mediation* R package(46) to conduct moderated mediation. Moderated mediation models were specified as described above, with the addition of an interaction term between air pollutant exposures and individual micronutrients or AHEI-P score. Casual mediation analysis was performed using the *mediation* R package(83) with a quasi-Bayesian approach and 10 iterations. Mediator treatment and control values were defined as mean ± standard deviation of air pollution exposure at the study population level. Moderated was tested using the *test.modmed()* function from the mediation R package with a quasi-Bayesian approach and 1000 iterations. Moderator levels were defined as mean ± standard deviation of micronutrients or AHEI-P score.

Statistical considerations. The MASS R package(48) was used to run multivariable linear regressions with the Huber's Proposal-2 M-estimator. For all analyses (unless otherwise

specified), models were run twice: once with DNAm m-values to obtain p-values, as m-values are homoscedastic and therefore more statistically appropriate, and again with β-values to obtain biologically-interpretable effect sizes(49). CpGs that failed to converge within 50 iterations were removed from further analysis in epigenome-wide investigations. All models included variables to adjust for biological variation and technical variation (see ***SI Methods***). Vitamin C, PUFA, and folate intake and AHEI-P scores were mean-centred prior to inclusion in robust linear models. P-values obtained from epigenome-wide investigations were adjusted using the Benjamini-Hochberg method. P-values obtained from (moderated) mediation analyses were adjusted using the Bonferroni method. Air pollutant effect sizes (β-values) were defined as the change in DNAm across the range of air pollution exposure estimates and were calculated by multiplying the coefficient of air pollution exposures obtained from robust multivariable linear models by the range of air pollution exposures in the study population. Effect sizes of micronutrients or overall diet quality buffering on air pollutant-associated DNAm differences (β-value/ppb or β-value/ug/m^3) were defined as the change in DNAm across the range of micronutrient or diet quality exposure and were calculated by multiplying the coefficient of the interaction term between a given micronutrient or overall diet quality obtained from robust multivariable linear models by the range of micronutrient intake or maternal diet quality across. Statistical significance was defined as (adjusted) $p<0.05$ and for epigenome-wide and biological significance was defined as an absolute effect size >0.01% for individual and region DNAm differences.

Results

A total of 515 mother-child dyads from the socioeconomically and racially heterogeneous cohort of CANDLE participants (30) were included in our main analyses of prenatal air pollution exposures. Most mothers self-identified as either Black/African American (53%) or White (40%), with the remainder identifying as Asian (1%), American Indian/Alaskan Native (<1%), as multiple races (5%), or another race not listed on the questionnaire (1%; **Table 1**). Median annual prenatal $PM_{2.5}$ and NO_2 exposure estimates of 10.5 μg/m^3 and 8.1 ppb were near or below the current primary annual standards (9 μg/m^3 and 53 ppb, respectively) for human health protection set by the United States Environmental Protection Agency (US EPA)(50, 51); there is no USA EPA annual standard for PM_{10} exposure but we noted a median PM_{10} exposure of 17.7 ppb(52). A total of 13%, 10%, and 4% of study participants exhibited transient, persistent, or late-onset wheeze, respectively, with 12% of biological males and 8% of female exhibiting persistent wheeze (**Table S1**). Caregiver reports of child atopic dermatitis in the study population were high (25%) compared to U.S. prevalence rates of 13.7% in children age zero to five years old(53), and even higher in those with transient (43%), persistent (45%), or late-onset (37%) wheeze (**Table 1 and Table S2**). Only 6% of those with transient wheeze received an asthma diagnosis by the age four to six year visit, compared to 82% of individuals with persistent wheeze and 58% with late-onset wheeze (**Table S2**). Participants with transient, persistent, or late-onset wheeze also exhibited consistent trends of higher prenatal exposure to NO_2, $PM_{2.5}$, and PM_{10} compared to individuals without wheeze, though differences did not reach statistical significance (one-sided t-test p>0.05; **Fig S1**).

The main analyses of this study examined associations between prenatal air pollutant exposures and cord blood DNAm among the 515 CANDLE participants. Specifically, we conducted epigenome-wide analyses of NO_2, $PM_{2.5}$, and PM_{10} to identify potential novel DNAm associations with broader applicability across human populations, and to assess whether findings previously reported in predominantly European ancestry populations replicate in a more diverse study population (**Fig S2**; **Files S1-3**). Models exhibited good genomic inflation factors between 0.93 and 1.04 (54). A single CpG (cg15994137) annotated to the body of *ZFYVE9* was significantly and negatively associated with prenatal $PM_{2.5}$ exposure (adjusted p=0.046, effect size=-0.06). No other statistically significant cord blood DNAm differences were identified in the epigenome-wide analyses of either prenatal NO_2 or prenatal PM_{10} exposure.

As part of our main analyses, we next examined differentially methylated regions (DMRs)(45), which provide greater power to detect small DNAm differences compared to epigenome-wide investigations. Prenatal exposures to NO_2, $PM_{2.5}$, and PM_{10} were associated with 19, seven, and five cord blood DMRs, respectively (Sidak p<0.05 and mean absolute effect size>0.01; **Table 2**). Although no DMRs were shared between air pollutants (**Fig 1**), each air

pollutant was associated with at least one DMR annotated to an immune-relevant gene including members of the *HLA* family (NO_2 and $PM_{2.5}$)(55), *IL9* ($PM_{2.5}$)(56), and *GCNT2* (PM_{10})(57). We further examined whether any of the observed regional DNAm differences might be under the influence of nearby genetic variations by crosschecking CpGs within each DMR with an online mQTL database (58). At least half of the component CpGs in three NO_2-associated DMRs, one $PM_{2.5}$-associate DMR, and one PM_{10}-associated DMR were previously annotated as mQTLs, suggesting that DNAm differences in these regions may be influenced by nearby genetic variants (**Table 2**). Allelic effect sizes (β-value) of potential mQTLs identified in the analyses of NO_2, $PM_{2.5}$, and PM_{10} exposure ranged from 0.02 to 0.28, 0.02 to 0.16, and 0.02 to 0.8, respectively.

The biological embedding of prenatal air pollution exposures via DNAm is hypothesized to contribute to child respiratory disease including wheeze and asthma(24). Although we consistently observed wheeze phenotypes were positively associated with air pollutant exposures (except for adjusted models of NO_2 and persistent wheeze, and $PM_{2.5}$ or PM_{10} and late-onset wheeze), the difference in prenatal air pollution exposures between non-wheezing and wheezing individuals did not reach statistical significance (defined as $p<0.05$) in one-sided t-tests (**Fig S1**). Nevertheless, we proceeded to investigate the role of mean DNAm across DMRs associated with prenatal NO_2, $PM_{2.5}$, and PM_{10} in wheeze outcomes using causal mediation analysis (**Files S4-S12**). One DMR within the promoter of chr6:33043574–33044013 (*HLA-DPA1/DPB1*) showed a suggestive positive mediation effect on transient wheeze associated with prenatal NO_2 (unadjusted p=0.032, ACME=0.040; **Fig 2A**). A nearby DMR in this same region (chr6:33041218-33041269) also showed a suggestive mediation effect on transient wheeze in relation to $PM_{2.5}$ exposure (unadjusted p=0.038, ACME=4.7E-02; **Fig 2B**). In both cases, the average direct effects of air pollution (-4.4E-02, and -8.2E-03, respectively) were in the opposite direction of the mediation effects. However, positive mediation effects did not remain significant after multiple testing correction (**Files S4-12**).

We next tested the hypothesis that maternal intake of protective micronutrients including vitamin C(27), or the ratio of n6 to n3 PUFAs (59, 60), or folate(26) could buffer the effects of prenatal exposure to NO_2, $PM_{2.5}$, or PM_{10} on DMRs identified in our main analyses(35). Because these beneficial micronutrients were correlated to one another (**Fig S3**), we also examined the potential buffering effect of overall diet quality using the AHEI-P score. This score ranges from 0 to 130, with higher values indicating greater intake of nutrients associated with chronic disease prevention(61). To test the above hypothesis, we conducted a secondary analysis, repeating epigenome-wide analyses (**Files S13-S24**), but this time including interaction terms between each air pollutant (NO_2, $PM_{2.5}$, PM_{10}) and either each micronutrient or the AHEI-P score. We focused on DMRs from our main analysis that had both a significant pollutant main effect term and a significant interaction term of the new models (**Files S25-S45**), with effect sizes in opposite

directions, indicating a potential buffering effect. Only one NO_2-associated DMR annotated to *C17orf64* (chr17:58499679-58499912) from our main analysis met this criterion when assessing the effect of maternal folate intake in secondary analyses. Specifically, higher maternal folate intake was associated with lower DNAm across this DMR (effect size = -1.23E-02 β-value/ppb), counteracting the higher DNAm linked to NO_2 exposure (**Fig 3A**). Though neither individual micronutrients nor overall diet quality significantly buffered other DMRs identified in the main analyses, we observed directionally consistent buffering effects for one or more micronutrients or for AHEI-P across 16 of the 19 NO_2-associated DMRs and across all $PM_{2.5}$- and PM_{10}-associated DMRs(**Fig 3B-D; Files 46-57**). Absolute interaction effect sizes were biologically significant ($\geq 1.00\times10^{-2}$ β-value/ppb or β-value/μg/m³) for two NO_2-associated DMRs, five $PM_{2.5}$-associated DMRs, and four PM_{10}-associated DMRs, suggesting buffering may be occurring across these DMRs even though they were not identified in any of the interaction effect terms.

We were also interested in DMRs that emerged in the air pollutant main effect term of the interaction models but were not identified in our main analyses (**Files S24-S36**). These DMRs likely represent pollutant-associated DNAm changes that were previously masked by the modifying effects of maternal micronutrient intake and/or overall diet quality. More than half of the DMRs associated with the main effect of $PM_{2.5}$ or PM_{10} in diet interaction models (vitamin C, PUFAs, folate, and AHEI-P) were not detected in the main analyses (**Fig 1**). Similarly, over a third of DMRs associated with the main effect of NO_2 in diet interaction models were not detected in main analyses. Among these, a NO_2-associated DMR annotated to *UPP1* (chr7:48129797-48130198) exhibited positive causal mediation effects (0.043 to 0.046) on the relationship between NO_2 exposure and persistent wheeze in diet interaction models that included PUFAs, folate, or AHEI-P score (**Files S58-S60**). These mediation effects were in the opposite direction of the direct effect of NO_2 on persistent wheeze and did not remain significant after Bonferroni correction (unadjusted p=0.024 to 0.032). No other newly emerged DMRs had significant positive causal mediation effects between any air pollutant and wheeze phenotype (**Files S61-S93**). Most (N=8 of 12) newly emerged NO_2-associated DMRs were shared across all diet interaction models (**Fig 1**), likely reflecting the high correlation between micronutrients (**Fig S4**). A similar pattern was observed for $PM_{2.5}$ DMRs where 7 of 9 newly emerged DMRs were shared across diet interaction models. In contrast, most newly identified PM_{10}-associated DMRS were shared only across diet interaction models adjusting for PUFA intake, folate, or AHEI-P score, but not vitamin C (**Fig 1**), which suggests maternal intake of vitamin C may have a less prominent role in buffering the effects of PM_{10} on cord blood DNAm.

Most NO_2 (≥13 of 19) and $PM_{2.5}$ (4 of 7) DMRs identified in our main analysis were also identified in the air pollutant effect term across all diet interaction models (**Fig 1**; **File S24-S30**). When we repeated mediation analyses with interaction terms to account for maternal dietary

effects, we identified new positive causal mediation effects between NO_2 and transient wheeze through a DMR annotated to *HLA-DMB* in models adjusting for vitamin C, folate, or AHEI-P score (**Files S61**, **S63**, and **S64**). This DMR also showed positive mediation effect between NO_2 and late wheeze when adjust for vitamin C, PUFAs, folate, or AHEI-P score (**Files S66-S69**). A second DMR annotated to *SYNE1* exhibited positive mediation effects between NO_2 and transient wheeze only after adjusting for vitamin C or PUFA intake (**Files S61** and **S62**). Additionally, we observed that the positive causal mediation effects of *HLA-DPA1/DPB1* on the relationship between NO_2 and transient wheeze increased from 0.04 to as much as 0.06 after adjustment for intake of vitamin C, PUFA, or folate or AHEI-P score (**Files S61-S64**). Although none of these causal mediation effects reached statistical significance after multiple testing correction, the emergence or amplification of these effects following dietary adjustment(47) supports our hypothesis that maternal intake of beneficial micronutrients and overall diet quality may buffer the effects of air pollutants observed in the main models. We further evaluated this hypothesis using moderated mediation analysis; however, the difference in causal mediation effects between ±1 standard deviation of vitamin C, PUFAs, folate, and AHEI-P did not reach statistical significance for any of the DMRs described above (**File S94-S98**).

Previous research has shown that biological males are at greater risk of developing childhood asthma following prenatal exposure to air pollution(11–13). However, the role of sex-specific DNAm changes in this relationship remains unclear. To investigate this, we repeated the epigenome-wide analyses, this time including an interaction term between each air pollutant and biological sex (**File S99-101)**. Our primary focus was on DMRs identified in the interaction term, as these regions represent locations where the effects of air pollution differ between males and females. The effects of NO_2, $PM_{2.5}$, and PM_{10} on DNAm differed by biological sex across two, 11, and two DMRs respectively (**File S102-S104**). Males exhibited higher DNAm at both NO_2-associated DMRs and lower DNAm at six of the 11 $PM_{2.5}$- and both PM_{10}-associated DMRs compared to females. At least half of the component CpGs in the six $PM_{2.5}$-associated DMRs were previously annotated as putative mQLTs, with effect sizes ranging from 0.02 to 0.11, indicating potential genetic influence on DNAm across these regions. Several of the $PM_{2.5}$-associated DMRs containing putative mQTLs, including *SERPINB9*(62) *and WFDC1*(63), have previously been linked to lung function in genome- or epigenome-wide studies. However, in this study, none of the sex-differing DMRs showed evidence of mediating the relationships between air pollutants and transient, persistent, or late-onset wheeze. We did not examine mediation effects of sex-differing DNAm regions on late onset wheeze due to small sample size (N=9 males, 10 females).

Finally, we assessed consistency with prior epigenome-wide studies by examining whether CpGs previously associated with prenatal NO_2, $PM_{2.5}$, or PM_{10} exposure exhibited the

same direction of effect in our epigenome-wide investigations, even if they did not reach statistical significance (FDR <0.05). Several DNAm alterations associated with prenatal $PM_{2.5}$ and PM_{10} exposure in Precision Medicine Intervention in Severe Asthma (PRISM) participants, who have similar demographics to CANDLE, were replicated in this study, exhibiting the same direction of effect. These included lower DNAm with greater $PM_{2.5}$ exposure at cg01011943 (*PSG5*) and cg00348551 (*C7orf50*), and higher DNAm with PM_{10} exposure at cg00905156 (*FAM13A*) and cg06849931 (*NOTCH4*) (**Fig S15A-E**) (19, 20). We also observed higher DNAm at cg18684142 (*HOXB6*) with NO_2 in CANDLE participants, consistent with findings from the primarily European CHILD cohort study(20), implying this change is robust across populations. In contrast, several findings from studies using primarily European cohorts did not replicate in CANDLE. These include higher DNAm at cg10430077 (*TRIM41*) in males with NO_2(20), higher DNAm at cg00035623 (*UBE2S*) with $PM_{2.5}$(18), higher DNAm at cg02015529 (*TM9SF2*) with PM_{10}(18), and lower DNAm at cg22329831 (*TDRD6*) with PM_{10}(18), suggesting these associations may not be generalizable across more diverse populations (**Fig S5F-I**).

Discussion

In this study we identified DNAm alterations in cord blood associated with prenatal exposure to NO_2, $PM_{2.5}$, and/or PM_{10} in a subset of participants from the socioeconomically and racially diverse CANDLE cohort study. Most identified DMRs were unique to individual air pollutants, suggesting that different cellular pathways may be affected by prenatal exposure to NO_2, $PM_{2.5}$, and PM_{10}. Altered DNAm in *HLA-DPA1/DBP1* DNAm displayed suggestive positive mediation effects between prenatal exposure to NO_2 and $PM_{2.5}$ and transient wheeze. Additional DMRs associated with prenatal air pollutant exposures were identified only after adjusting for maternal intake of protective nutrients or overall diet quality in our secondary analyses, suggesting that the effects of these exposures on DNAm were likely masked by the buffering influence of maternal diet in the initial models. Two DMRs from the main analysis exhibited significant mediation effects only after considering diet interaction terms in the mediation analysis. Additionally, including diet interaction terms in the mediation analysis strengthened the previously observed mediation at *HLA-DPA1/DPB1*. Together, these findings suggest that prenatal air pollutant exposures result in unique alterations to cord blood DNAm, and that nutrient dense, high quality maternal diets may have a role in buffering some of these DNAm differences.

Lower DNAm in the promoter region of *HLA-DPA1/DPB1* displayed suggestive positive causal mediation effects between exposure to NO_2 or $PM_{2.5}$ and transient wheeze in our main models. These HLA genes code for MHC class II member proteins found on antigen presenting cells and play roles in immune response following inhalation of heavy metals(64) and in bronchodilator response(63). Genome-wide studies have also identified associations between *HLA-DPA1/DPB1* and atopic dermatitis(65) and asthma(66). The role of lower *HLA-DPA1/DPB1* DNAm in transient wheeze associated with NO_2 is unclear, but could involve similar mechanisms to its function in asthma, where heightened transcription has been shown to lead to increased T-cell activation and inflammation(67).

A primary focus of this work was whether maternal diet could mitigate the effects of prenatal air pollutant exposure on DNAm changes, and ultimately wheeze and asthma in childhood. Positive causal mediation effects of lower *HLA-DPA1/DPB1* DNAm associated with NO_2 further increased in diet interaction models (vitamin C, PUFA, folate, and AHEI-P score), suggesting micronutrient rich, high-quality diets may play a role in buffering the effects of NO_2 on DNAm(47). These observations are consistent with previous findings demonstrating greater maternal diet density of vitamin C can reduce the effects of maternal prenatal smoking on DNA and is protective towards child respiratory health at age five(27, 28), and that higher maternal folate can buffer the effects of prenatal PM_{10} exposure on childhood IQ(26). Though differences in the causal mediation effects of *HLA-DPA1/DPB1* DNAm between low and high maternal micronutrient intake did not reach significance in moderated mediation analyses, the above

findings support that more healthful maternal diets can help mitigate infant DNAm changes associated with prenatal air pollutant exposures, which could help reduce changes to children's immune function that contribute to wheeze and asthma.

This study extends prior work by examining the role of air pollution-induced DNAm differences in specific wheeze phenotypes including no/infrequent wheeze, transient wheeze, persistent wheeze, and late onset wheeze at age 4 to 6 years. We based our classifications on those outlined in the original Tucson Study(14). Our ability to classify wheeze phenotypes in CANDLE participants was facilitated by the availability of repeated questionnaire and clinical visit data from birth until the age four clinic visit. Prenatal air pollution exposures have been associated with transient, persistent, and late-onset wheeze phenotypes; however, evidence is limited and findings vary across studies(15). This work extends our knowledge of prenatal air pollution exposure and respiratory health by providing insight into the air pollution-induced DNAm differences associated with wheeze phenotypes. Specifically, these findings suggest that prenatal NO_2 and $PM_{2.5}$ may contribute to transient wheeze phenotypes in part through a shared pathway involving HLA-DP genes. HLA-DP genetic variants have been also associated with atopic dermatitis(43), and we noted a high rate of atopic dermatitis in participants with transient wheeze here and in the full study cohort compared to the Tucson Children's Respiratory Study participants(14). It remains possible that shared genetic variants are driving the development of wheeze and atopic dermatitis in this study, but is unlikely given that we adjusted data for genetic ancestry and HLA-DP gene CpGs did not constitute previously identified mQTLs listed on mQTLdb(58). Replication of these findings in other racially diverse cohorts will help confirm a shared role for the regulation of HLA-DP genes in the relationship between prenatal air pollutant exposures and childhood wheeze and asthma.

Biological males are more susceptible to asthma and the effects of prenatal air pollution exposure than females(12, 68, 69). This may be partly due to greater inflammatory signalling male fetal development(70, 71) and relatively slower large airway development in males compared to females(72). DNAm differences induced by air pollution that potentiate either of these effects could contribute to increase respiratory disease risk in males. In this study, prenatal $PM_{2.5}$ exposure was associated with higher DNAm in the body of *SERPINB9* in males and lower DNAm in females. *SERPINB9* is expressed in several immune cell types(52) and inhibits intracellular granzyme B, protecting releasing cells from apoptosis(73). It also inhibits caspase-1, preventing maturation of IL-1β and inflammation(74, 75). The *SERPINB9* DMR identified here overlaps with a sex-specific DMR previously associated with forced expiratory volume in one second (FEV_1) in males(76). No associations were observed between $PM_{2.5}$-induced *SERPINB9* DNAm differences and wheeze phenotypes in either males or females in this study. However, it remains plausible these DNAm differences could still contribute to sex-specific differences in lung

function. Further replication is needed to clarify the role of *SERPINB9* DNAm in human respiratory health.

Most epigenome-wide studies of prenatal air pollution exposure have been conducted in primarily European birth cohorts(18–20, 77, 78). While these studies have provided meaningful insights into infant DNAm differences, their findings may have limited generalizability to more racially diverse populations. For example, we were unable to replicate many of the air pollution-associated DNAm findings from these European cohorts in the subset of CANDLE participants, in which over half (53%) of mothers identified as Black or African American. In contrast, several DNAm associations observed in PRISM(19), where a larger proportion of mothers identified as Hispanic (42%) or African American (18%), were replicated here in CANDLE. Shared regional, sociodemographic, or cultural differences between PRISM and CANDLE may underlie the greater consistency in findings, compared to those from predominantly European cohorts. These discrepancies underscore the need for greater diversity in DNAm and air pollution research and replication across sociodemographic populations should be conducted to ensure results are robust.

This study had several strengths and limitations. One advantage was the availability of prenatal air pollution estimates, which allowed us to compare DNAm alterations across multiple pollutant exposures(79). Another advantage of this study is the use of cord blood to investigate the role of DNAm changes in early-life wheeze phenotypes. While cord blood is commonly used due to its accessibility and relevance to prenatal exposures, it also enables investigation of DNAm in white blood cells, including T-cells, which contribute to immune dysfunction underlying wheeze and asthma(80). Genes highlighted in this study, including *HLA-DPA1/DPB1* and *SERPINB9*, are expressed in both cord blood and lung tissue(81), supporting the plausibility that observed DNAm differences could influence immune pathways relevant to respiratory disease. Further research is needed to replicate and characterize the role of DNAm differences in observed in this study on immune function and respiratory health and to determine if any of these differences are shared in lung tissue(82, 83). Future studies could consider animal models to interrogate the effects of different air pollution components on infant DNAm, and lung DNAm and function directly. For instance, a mouse model of prenatal cigarette exposures used by our lab successfully reproduced DNAm changes observed in human cord blood and also identified novel DNAm changes in lung tissue(71)

In terms of limitations, we used genetic ancestry PCs to describe population variation along a continuum and incorporated them in the statistical models to account for genetic variation within ancestral populations. While genetic ancestry PCs are often correlated with racial and ethnic identity, they are distinct from those social constructs and thus, cannot be considered as proxies of self-identified racial and ethnic identities (84). We did not include self-identified racial or

ethnic identity in our models as we were interested in identifying DNAm alterations that were robust across populations. Therefore, we acknowledge that some of our associations may be influenced by residual confounding due to unaccounted racial and ethnic differences, or to ongoing racism and racist socioeconomic policies. Future studies thoughtfully designed to focus on these levels of inquiry may be of value(85). Finally, while this study includes more racial diversity than previous studies, CANDLE includes primarily Black/African American and White families. Replication of these findings across more populations is necessary before they can be generalized.

Together, the results from this study expand our understanding of how prenatal exposure to NO_2, $PM_{2.5}$, and PM_{10} affect DNAm. By leveraging the racially and socioeconomically diverse CANDLE cohort, we identified unique DNAm signatures associated with NO_2, $PM_{2.5}$, and PM_{10} exposure, and were able to replicate previous findings in cohorts with similar socioeconomic and racial structure. Importantly, the findings from this study highlight the potential for maternal diet micronutrients and overall diet quality to at least partially buffer some of these DNAm differences, supporting the role of a nutrient dense diet during pregnancy for child health.

Data availability

Supplementary files are available for download on Zenodo (10.5281/zenodo.19240961). Code used for this analysis is available on Github at https://github.com/mjoneslab/CANDLE_cord_blood.

Acknowledgments

The authors would like to acknowledge the participants of the CANDLE study.

Figures

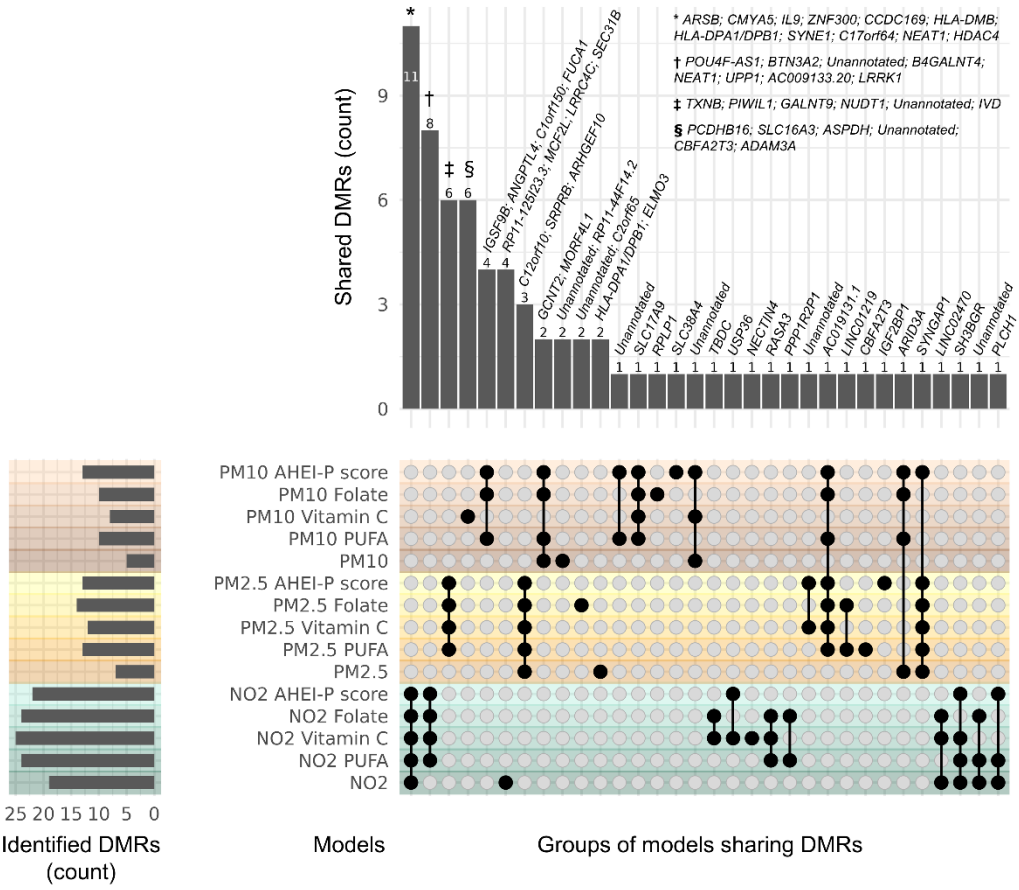

Figure 1. Overlap of differentially methylated regions (DMRs) identified in the main air pollution exposure analysis (N=515 participants) or in the main effect term of NO_2, $PM_{2.5}$, or PM_{10} in secondary diet interaction analyses (N=444 participants). Diet interaction analysis account for the maternal intake of vitamin C, the ratio of n-6 to n-3 polyunsaturated fatty acids (PUFAs), or overall diet quality using Alternative Healthy Eating Index Pregnancy (AHEI-P) scores. The bar chart on the left of the graph displays the total number of DMRs identified in each analysis. The dot matrix indicates which groups of DMRs belong to which analysis. An analysis that contains a group of DMRs is specified by a black circle. Black vertical lines joining plots indicate that a group of DMRs was identified in one or more analysis. The bar chart on the top of the graph indicates the number of DMRs in a given group. Bars are annotated with the genic annotations of DMRs within that group.

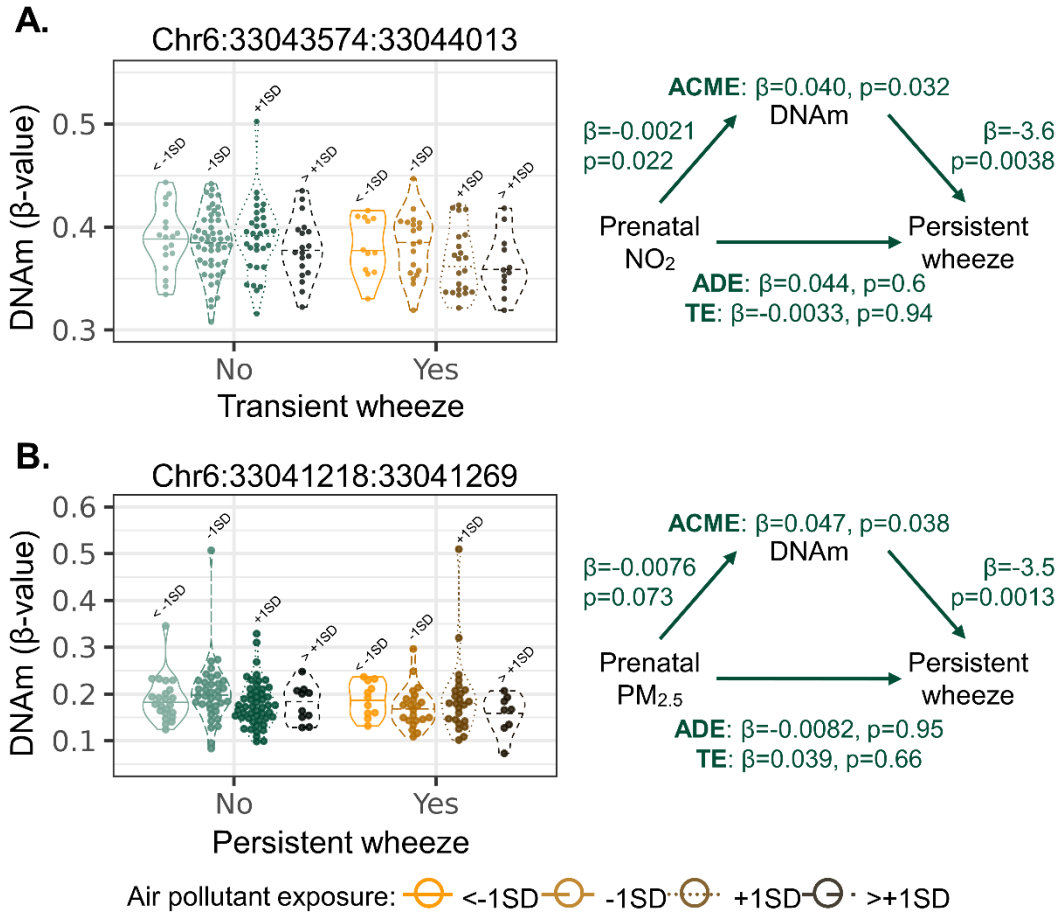

Fig 2. Causal mediation effects of cord blood differentially methylated regions (DMRs) between prenatal NO_2 or $PM_{2.5}$ exposures and transient wheeze (N=65 with, 119 without). (**A**) Relationship between mean DNAm of chr6:33043574-33044013 (*HLA-DPA1/DPB1*), prenatal NO_2 exposure, and transient wheeze (*left*). Mediation pathway of prenatal NO_2 exposure on transient wheeze via mean chr6:33043574-33044013 DNAm (*right*). (**B**) Relationship between mean DNAm of chr6:33041218-33041269 (*HLA-DPA1*), prenatal $PM_{2.5}$ exposure, and transient wheeze (*left*). Mediation pathway of prenatal $PM_{2.5}$ exposure on transient wheeze via mean chr6:33041218-33041269 DNAm (*right*). ACME = average causal mediation effects; ADE = average direct effects; TE = total effects. Participants with transient wheeze (N=65) are coloured in shades of orange/brown, and participants without transient wheeze (N=119) coloured in shades of green, where darker shades represent higher air pollution exposures. Data are stratified by NO_2 exposure quartiles for visualization. The first quarter is represented by open circles and solid lines, the second quarter by open triangles and long dashed lines, the third quarter by open squares and dotted lines, and the fourth quarter by crosses and short dashed lines.

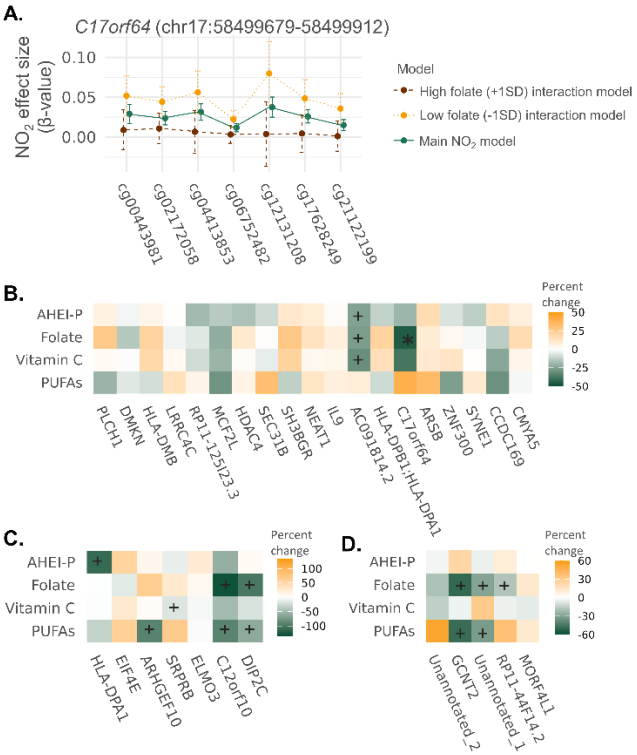

Fig 3. Effect sizes of the interaction term between maternal intake of polyunsaturates fatty acids (PUFAs), vitamin C, folate, or Alternative Healthy Eating Index - Pregnancy (AHEI-P) and air pollutants from secondary analyses relative to the main effect of NO_2, $PM_{2.5}$, or PM_{10} across cord blood differentially methylated regions (DMRs) identified in the main analyses of CANDLE participants (N=515). (**A**) A DMR associated with NO_2 and annotated to *C17orf64* in the main analysis was also identified in the interaction term of the model accounting for maternal folate intake, indicating a potential buffering effect. Across the DMR, mean DNAm associated with NO_2 was lower when accounting for higher maternal folate intake (+1SD) and increased when considering low maternal folate intake (-1SD). In interaction models, maternal intake of PUFAs, vitamin C, folate, and overall diet quality (AHEI-P) reduced the effect size of (**B**) NO_2, (**C**) $PM_{2.5}$, and (**D**) PM10 across several of the DMRs from the main analysis. Buffering effect sizes are displayed in percent (%) change of interaction effect sizes relative to the main effect sizes of air pollutants. Buffering is considered when the percent change is negative, representing an interaction effect in the opposite direction of the main air pollutant effect. A positive percent change occurs when the interaction effect and main effect are in the same direction and is not considered beneficial. Plus (+) symbols indicate interaction effect sizes that surpassing biological significance (≥0.01 β-value/ppb or β-value/μg/m³) but not identified in the interaction effect term of any of the interaction models. The asterisk (*) symbol indicates the NO_2 associated DMR annotated to *C17orf64* that was identified in both the main effect term and interaction effect term.

Tables

Table 1. Study population characteristics of mother-child dyads included in the main epigenome-wide prenatal air pollutant exposures (N=515) and in the secondary analysis examining moderation by maternal prenatal diet factors (N=444) analysis sets.

	Epigenome-wide analyses of cord blood (N=515)	**Moderation via maternal diet (N=444)**
Prenatal NO_2 (ppb) [mean±sd]	8.2 ± 2.6	8.00 ± 2.6
Prenatal $PM_{2.5}$ (μg/m^3) [mean±sd]	10.5 ± 0.8	10.4 ± 0.8
Prenatal PM_{10} (μg/m^3) [mean±sd]	17.8 ± 1.7	17.7 ± 1.6
Biological sex [count(%)]		
Male	249 (48%)	214 (48%)
Female	266 (52%)	229 (52%)
Maternal self-identified race [count(%)]		
Black/African American	274 (53%)	235 (53%)
White	205 (40%)	181 (41%)
Asian	6 (1%)	5 (1%)
American Indian/Alaska Native	1 (0%)	1 (0%)
Other	3 (1%)	3 (1%)
Multiple Race	26 (5%)	19 (4%)
Gestational Age (weeks) [mean±sd]	39 ± 1	39 ± 1
Adjusted income (USD) [mean±sd]	22600 ± 17800	23600 ± 17800
Maternal prenatal tobacco use [count(%)]		
Yes	41 (8%)	34 (8%)
No	474 (92%)	410 (92%)
Transient wheeze at age 4 to 6 years [count(%)]		
Yes	65 (13%)	59 (13%)
No	355 (69%)	309 (70%)
Missing	95 (18%)	76 (17%)
Persistent wheeze at age 4 to 6 years [count(%)]		
Yes	51 (10%)	45 (10%)
No	391 (76%)	339 (76%)
Missing	73 (14%)	60 (14%)
Late onset wheeze at age 4 to 6 years [count(%)]		
Yes	19 (4%)	18 (4%)
No	391 (76%)	341 (77%)
Missing	105 (20%)	85 (19%)
Ever Atopic Dermatitis at age 4 to 6 years [count(%)]		
Yes	129 (25%)	110 (25%)

No	283 (55%)	249 (56%)
Missing	103 (20%)	85 (19%)
Current asthma at age 4 to 6 years [count(%)]		
Yes	65 (13%)	57 (13%)
No	278 (54%)	246 (55%)
Missing	172 (33%)	76 (32%)
Vitamin C dietary density (mg/kcal) [mean±sd]	74.4 ± 30.7	74.1 ± 30.2
Missing [count(%)]	27 (5.2%)	
PUFA dietary density (gms/kcal) n6:n3 ratio [mean±sd]	9.1 ± 1.3	9.1 ± 1.3
Missing [count(%)]	27 (5.2%)	
Folate dietary density (mcg/kcal)	143 ± 54.1	144 ± 53.6
Missing [count(%)]	27 (5.2%)	
Alternative Healthy Eating Index Pregnancy (AHEI-P) score [mean±sd]	60.9 ± 12.5	61.1 ± 11.9
Missing [count(%)]	27 (5.2%)	
Estimated cord blood cell proportions [mean±sd]		
B-Cells	0.08 ± 0.04	0.08 ± 0.04
CD4 T-cells	0.15 ± 0.06	0.14 ± 0.06
CD8 T-cells	0.10 ± 0.04	0.10 ± 0.04
Granulocytes	0.44 ± 0.13	0.44 ± 0.13
Monocytes	0.12 ± 0.04	0.12 ± 0.04
NK cells	0.04 ± 0.03	0.04 ± 0.03
nRBC	0.12 ± 0.09	0.12 ± 0.09

Table 2. Differentially methylated regions in cord blood that are associated with prenatal NO_2 (N=19), $PM_{2.5}$ (N=7), and PM_{10} (N=5) exposures in CANDLE participants (N=515).

Air pollutant	Chr	Start	End	Pval	Sidak pval	Num Probes	Num probes as mQTLs	Gene annotation	Genic location	Mean effect size (β-value)
NO_2										
	5	78985425	78986161	2.4E-15	2.5E-12	11	7	*CMYA5*	TSS1500; TSS200; Body	5.3E-02
	13	36871646	36872347`	9.27E-15	1.03E-11	12	0	*CCDC169*	TSS1500; TSS200; 5'UTR; 1stExon 3'UTR	1.3E-02
	6	152958462	152958900	2.7E-12	4.9E-09	9	4	*SYNE1*	TSS1500; TSS200; 5'UTR; 1stExon	1.9E-02
	5	150284416	150284797	7.1E-12	1.5E-08	11	0	*ZNF300*	TSS1500; TSS200	2.7E-02
	5	78281684	78282100	6.4E-10	1.2E-06	8	4	*ARSB*	TSS1500; TSS200; 5'UTR; 1stExon	2.4E-02
	17	58499679	58499912	1.2E-08	4.0E-05	7	0	*C17orf64*	TSS1500; TSS200; 5'UTR; 1stExon	2.6E-02
	6	33043574	33044013	3.4E-08	6.0E-05	5	0	*HLA-DPB1; HLA-DPA*	TSS200; 5'UTR; 1stExon; Body	-2.4E-02

	12	10095902	10096153	6.2E-08	1.9E-04	5	4	*AC091814.2*	TSS200	7.6E-02
	5	135231514	135231578	1.6E-07	1.9E-03	4	0	*IL9*	TSS500; 1stExon; Body	-1.9E-02
	11	65194933	65195040	4.4E-07	3.2E-03	4	0	*NEAT1*	5'UTR	-2.6E-02
	21	40823722	40824020	5.2E-07	1.4E-03	4	0	*SH3BGR*	TSS200; 5'UTR; 1stExon	1.8E-02
	10	102279690	102279955	8.8E-07	2.6E-03	10	0	*SEC31B*	TSS1500; TSS200	1.4E-02
	2	240230587	240230893	1.1E-06	2.8E-03	4	0	*HDAC4*	Body	2.9E-02
	13	113622539	113622751	1.6E-06	6.0E-03	7	0	*MCF2L*	TSS1500; TSS200	-2.4E-02
	13	27947286	27947412	1.8E-06	1.1E-02	3	0	*RP11-125I23.3*	TSS200	4.5E-02
	11	41481358	41481656	3.5E-06	9.2E-03	5	0	*LRRC4C*	TSS1500; TSS200	1.6E-02
	6	32908466	32908902	4.4E-06	7.9E-03	7	0	*HLA-DMB*	TSS200; 5'UTR; 1stExon; Body	-1.8E-02
	19	36004699	36004995	5.1E-06	1.3E-03	3	1	*DMKN*	TSS1500; TSS200	3.0E-02
	3	155421970	155422146	1.1E-05	5.0E-02	5	0	*PLCH1*	TSS200; 5'UTR; 1stExon	-3.0E-02
$PM_{2.5}$										
	10	729001	729957	2.2E-13	1.8E-10	9	0	*DIP2C*	Body	3.2E-02

	12	53693322	53693770	1.7E-11	3.0E-08	9	0	*C12orf10*	TSS200; 5'UTR; 1stExon; Body	-1.4E-02
	16	67233549	67233984	6.6E-09	1.2E-05	5	0	*ELMO3*	Body; ExonBnd	3.8E-02
	3	133502623	133502953	4.7E-08	1.1E-04	7	4	*SRPRB*	TSS1500; TSS200	5.7E-02
	8	1900353	1900779	3.7E-07	6.8E-04	8	1	*ARHGEF10*	Body	-2.4E-02
	4	99851003	99851282	4.8E-07	1.2E-03	7	0	*EIF4E*	TSS1500; TSS200; 5'UTR; 1stExon	3.5E-02
	6	33041218	33041269	3.3E-06	4.9E-02	4	0	*HLA-DPA1*	1stExon; Body	-2.7E-02
PM_{10}										
	15	79164541	79164807	4.3E-09	1.3E-05	3	3	*MORF4L1*	TSS1500	5.7E-02
	16	53407328	53407679	7.1E-09	1.6E-05	6	0	*RP11-44F14.2*	TSS200; 5'UTR; 1stExon	5.5E-02
	7	156400711	156400991	1.6E-08	4.5E-05	5	0	*Unannotated*		5.8E-02
	6	10555808	10556148	1.5E-07	3.5E-04	6	0	*GCNT2*	TSS200; 5'UTR; 1stExon; Body	-5.2E-02
	2	164204752	164205033	7.8E-06	2.2E-02	4	0	*Unannotated*		2.1E-02

638

References

1. K. E. Pinkerton, *et al.*, Distribution of particulate matter and tissue remodeling in the human lung. *Environ Health Perspect* **108**, 1063–1069 (2000).

2. D. Li, *et al.*, Fluorescent reconstitution on deposition of PM2.5 in lung and extrapulmonary organs. *PNAS* **116**, 2488–2493 (2019).

3. H. A. Jeng, Chemical composition of ambient particulate matter and redox activity. *Environ Monit Assess* **169**, 597–606 (2010).

4. T. Kennedy, *et al.*, Copper-dependent Inflammation and Nuclear Factor- κ B Activation by Particulate Air Pollution. *Am J Respir Cell Mol Biol* **19**, 366–378 (1998).

5. L. A. Jiménez, *et al.*, Activation of NF-kappaB by PM(10) occurs via an iron-mediated mechanism in the absence of IkappaB degradation. *Toxicol Appl Pharmacol* **166**, 101–110 (2000).

6. E. Goldstein, *et al.*, Fate and Distribution of Inhaled Nitrogen Dioxide in Rhesus Monkeys. *Am Rev Respir Dis* **115**, 403–412 (1977).

7. R. L. Persinger, M. E. Poynter, K. Ckless, Y. M. W. Janssen-Heininger, Molecular mechanisms of nitrogen dioxide induced epithelial injury in the lung. *Mol Cell Biochem* **234–235**, 71–80 (2002).

8. V. N. Ayyagari, A. Januszkiewicz, J. Nath, Pro-inflammatory responses of human bronchial epithelial cells to acute nitrogen dioxide exposure. *Toxicology* **197**, 148–163 (2004).

9. J. Qu, Y. Li, W. Zhong, P. Gao, C. Hu, Recent developments in the role of reactive oxygen species in allergic asthma. *Journal of Thoracic Disease* **9** (2017).

10. K. Liu, S. Hua, L. Song, PM2.5 Exposure and Asthma Development: The Key Role of Oxidative Stress. *Oxid Med Cell Longev* **2022**, 3618806 (2022).

11. A. Bettiol, E. Gelain, E. Milanesio, F. Asta, F. Rusconi, The first 1000 days of life: traffic-related air pollution and development of wheezing and asthma in childhood. A systematic review of birth cohort studies. *Environmental Health* **20**, 46 (2021).

12. H.-H. L. Hsu, *et al.*, Prenatal Particulate Air Pollution and Asthma Onset in Urban Children. Identifying Sensitive Windows and Sex Differences. *American Journal of Respiratory and Critical Care Medicine* **192**, 1052 (2015).

13. M. F. Hazlehurst, *et al.*, Maternal exposure to PM2.5 during pregnancy and asthma risk in early childhood: Consideration of phases of fetal lung development. *Environmental Epidemiology* **5**, e130 (2021).

14. F. D. Martinez, *et al.*, Asthma and Wheezing in the First Six Years of Life. *New England Journal of Medicine* **332**, 133–138 (1995).

673 15. N. Lau, A. Norman, M. J. Smith, A. Sarkar, Z. Gao, Association between Traffic Related Air
674 Pollution and the Development of Asthma Phenotypes in Children: A Systematic Review.
675 *International Journal of Chronic Diseases* **2018**, e4047386 (2018).

676 16. C. Hertzman, Putting the concept of biological embedding in historical perspective. *Proc*
677 *Natl Acad Sci U S A* **109**, 17160–17167 (2012).

678 17. J. F. Felix, C. a. M. Cecil, Population DNA methylation studies in the Developmental Origins
679 of Health and Disease (DOHaD) framework. *Journal of Developmental Origins of Health and*
680 *Disease* **10**, 306–313 (2019).

681 18. C. V. Breton, *et al.*, Particulate matter, the newborn methylome, and cardio-respiratory
682 health outcomes in childhood. *Environ Epigenet* **2**, dvw005 (2016).

683 19. O. Gruzieva, *et al.*, Prenatal Particulate Air Pollution and DNA Methylation in Newborns: An
684 Epigenome-Wide Meta-Analysis. *Environ. Health Perspect.* **127**, 57012 (2019).

685 20. S. Lee, *et al.*, Persistent DNA methylation changes associated with prenatal NO2 exposure in
686 a Canadian prospective birth study. [Preprint] (2023). Available at:
687 https://www.biorxiv.org/content/10.1101/2023.03.02.530668v1 [Accessed 6 March 2023].

688 21. A. Zanobetti, *et al.*, Childhood Asthma Incidence, Early and Persistent Wheeze, and
689 Neighborhood Socioeconomic Factors in the ECHO/CREW Consortium. *JAMA Pediatrics* **176**,
690 759–767 (2022).

691 22. J. Wang, *et al.*, Urban particulate matter triggers lung inflammation via the ROSMAPK- NF-
692 κB signaling pathway. *Journal of Thoracic Disease* **9**, 4398-4412–4412 (2017).

693 23. F. J. Kelly, T. Sandström, Air pollution, oxidative stress, and allergic response. *The Lancet*
694 **363**, 95–96 (2004).

695 24. S. Li, M. Chen, Y. Li, T. O. Tollefsbol, Prenatal epigenetics diets play protective roles against
696 environmental pollution. *Clin Epigenet* **11**, 82 (2019).

697 25. J. Zhong, *et al.*, B-vitamin Supplementation Mitigates Effects of Fine Particles on Cardiac
698 Autonomic Dysfunction and Inflammation: A Pilot Human Intervention Trial. *Scientific*
699 *Reports* **7**, 45322 (2017).

700 26. C. T. Loftus, *et al.*, Prenatal air pollution and childhood IQ: Preliminary evidence of effect
701 modification by folate. *Environmental Research* **176**, 108505 (2019).

702 27. L. E. Shorey-Kendrick, *et al.*, Vitamin C Prevents Offspring DNA Methylation Changes
703 Associated with Maternal Smoking in Pregnancy. *Am J Respir Crit Care Med* **196**, 745–755
704 (2017).

705 28. C. T. McEvoy, *et al.*, Effect of Vitamin C Supplementation for Pregnant Smokers on Offspring
706 Airway Function and Wheeze at Age 5 Years: Follow-up of a Randomized Clinical Trial. *JAMA*
707 *Pediatr* **177**, 16–24 (2023).

29. S. L. Rifas-Shiman, J. W. Rich-Edwards, K. P. Kleinman, E. Oken, M. W. Gillman, Dietary Quality during Pregnancy Varies by Maternal Characteristics in Project Viva: A US Cohort. *Journal of the American Dietetic Association* **109**, 1004–1011 (2009).

30. L. Sontag-Padilla, *et al.*, The Urban Child Institute CANDLE Study: Methodological Overview and Baseline Sample Description. (2015).

31. M. T. Young, *et al.*, Satellite-Based NO2 and Model Validation in a National Prediction Model Based on Universal Kriging and Land-Use Regression. *Environ Sci Technol* **50**, 3686–3694 (2016).

32. M. Wang, *et al.*, National PM2.5 and NO2 Spatiotemporal Models Integrating Intensive Monitoring Data and Satellite-Derived Land Use Regression in a Universal Kriging Framework in the United States: 1999-2016. *ISEE Conference Abstracts* (2018). https://doi.org/10.1289/isesisee.2018.O02.02.02.

33. P. D. Sampson, *et al.*, A regionalized national universal kriging model using Partial Least Squares regression for estimating annual PM2.5 concentrations in epidemiology. *Atmos Environ (1994)* **75**, 383–392 (2013).

34. J. A. Heiss, A. C. Just, Identifying mislabeled and contaminated DNA methylation microarray data: an extended quality control toolset with examples from GEO. *Clin Epigenetics* **10**, 73 (2018).

35. J. A. Heiss, A. C. Just, Improved filtering of DNA methylation microarray data by detection p values and its impact on downstream analyses. *Clinical Epigenetics* **11**, 15 (2019).

36. M. J. Aryee, *et al.*, Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).

37. V. Hovestadt, M. Zapatka, conumee: Enhanced copy-number variation analysis using Illumina DNA methylation arrays. (2022). https://doi.org/10.18129/B9.bioc.conumee. Deposited 2022.

38. A. M. Morin, *et al.*, Maternal blood contamination of collected cord blood can be identified using DNA methylation at three CpGs. *Clinical Epigenetics* **9**, 75 (2017).

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

40. R. Pidsley, *et al.*, A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**, 293 (2013).

41. R. Pidsley, *et al.*, Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biology* **17**, 208 (2016).

42. J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, J. D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).

43. S. K. Bantz, Z. Zhu, T. Zheng, The Atopic March: Progression from Atopic Dermatitis to Allergic Rhinitis and Asthma. *J Clin Cell Immunol* **5**, 202 (2014).

44. E. Völgyi, *et al.*, Dietary Patterns in Pregnancy and Effects on Nutrient Intake in the Mid-South: The Conditions Affecting Neurocognitive Development and Learning in Early Childhood (CANDLE) Study. *Nutrients* **5**, 1511–1530 (2013).

45. B. S. Pedersen, D. A. Schwartz, I. V. Yang, K. J. Kechris, Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics* **28**, 2986–2988 (2012).

46. D. Tingley, T. Yamamoto, K. Hirose, L. Keele, K. Imai, mediation: R Package for Causal Mediation Analysis. *Journal of Statistical Software* **59**, 1–38 (2014).

47. K. J. Preacher, D. D. Rucker, A. F. Hayes, Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions. *Multivariate Behavioral Research* **42**, 185–227 (2007).

48. W. N. Venables, B. D. Ripley, *Modern Applied Statistics with S* (Springer, 2002).

49. P. Du, *et al.*, Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587 (2010).

50. O. US EPA, Primary National Ambient Air Quality Standards (NAAQS) for Nitrogen Dioxide. (2016). Available at: https://www.epa.gov/no2-pollution/primary-national-ambient-air-quality-standards-naaqs-nitrogen-dioxide [Accessed 21 January 2025].

51. O. US EPA, National Ambient Air Quality Standards (NAAQS) for PM. (2020). Available at: https://www.epa.gov/pm-pollution/national-ambient-air-quality-standards-naaqs-pm [Accessed 21 January 2025].

52. O. US EPA, NAAQS Table. (2014). Available at: https://www.epa.gov/criteria-air-pollutants/naaqs-table [Accessed 11 September 2023].

53. S. Choragudi, G. Yosipovitch, Trends in the Prevalence of Eczema Among US Children by Age, Sex, Race, and Ethnicity From 1997 to 2018. *JAMA Dermatology* **159**, 454–456 (2023).

54. C. J. Williams, *et al.*, Genome wide association study of response to interval and continuous exercise training: the Predict-HIIT study. *J Biomed Sci* **28**, 37 (2021).

55. K. Shi, *et al.*, HLA-DPA1 overexpression inhibits cancer progression, reduces resistance to cisplatin, and correlates with increased immune infiltration in lung adenocarcinoma. *Aging (Albany NY)* **15**, 11067–11091 (2023).

775 56. L. Jia, *et al.*, Detection of IL-9 producing T cells in the PBMCs of allergic asthmatic patients.
776 *BMC Immunology* **18**, 38 (2017).

777 57. N. Giovannone, *et al.*, Galectin-9 suppresses B cell receptor signaling and is regulated by I-
778 branching of N-glycans. *Nat Commun* **9**, 3287 (2018).

779 58. T. R. Gaunt, *et al.*, Systematic identification of genetic influences on methylation across the
780 human life course. *Genome Biology* **17**, 61 (2016).

781 59. Y. Chen, *et al.*, The Associations of Prenatal Exposure to Fine Particulate Matter and Its
782 Chemical Components with Allergic Rhinitis in Children and the Modification Effect of
783 Polyunsaturated Fatty Acids: A Birth Cohort Study. *Environmental Health Perspectives* **132**,
784 047010 (2024).

785 60. H.-S. Lee, *et al.*, Modulation of DNA methylation states and infant immune system by
786 dietary supplementation with ω-3 PUFA during pregnancy in an intervention study. *The*
787 *American Journal of Clinical Nutrition* **98**, 480–487 (2013).

788 61. A. K. Poon, E. Yeung, N. Boghossian, P. S. Albert, C. Zhang, Maternal Dietary Patterns during
789 Third Trimester in Association with Birthweight Characteristics and Early Infant Growth.
790 *Scientifica (Cairo)* **2013**, 786409 (2013).

791 62. S. K. Sunny, *et al.*, Sex-specific longitudinal association of DNA methylation with lung
792 function. *ERJ Open Research* **7** (2021).

793 63. J. D. Gereige, *et al.*, A genome-wide association study of bronchodilator response in
794 participants of European and African ancestry from six independent cohorts. *ERJ Open*
795 *Research* **8** (2022).

796 64. M. Amicosante, *et al.*, Beryllium binding to HLA-DP molecule carrying the marker of
797 susceptibility to berylliosis glutamate β69. *Human Immunology* **62**, 686–693 (2001).

798 65. D. J. Margolis, *et al.*, A combination of HLA-DP α and β chain polymorphisms paired with a
799 SNP in the DPB1 3' UTR region, denoting expression levels, are associated with atopic
800 dermatitis. *Frontiers in Genetics* **14** (2023).

801 66. P. Takejima, *et al.*, Allergic and Nonallergic Asthma Have Distinct Phenotypic and Genotypic
802 Features. *International Archives of Allergy and Immunology* **172**, 150–160 (2017).

803 67. S. H.-D. L. Mata, *et al.*, Cytotoxic CD4+ tissue-resident memory T cells are associated with
804 asthma severity. *Med* **4**, 875-897.e8 (2023).

805 68. É. Lavigne, *et al.*, Effect modification of perinatal exposure to air pollution and childhood
806 asthma incidence. *European Respiratory Journal* **51** (2018).

807 69. A. Lee, *et al.*, Prenatal fine particulate exposure and early childhood asthma: effect of
808 maternal stress and fetal gender. *J Allergy Clin Immunol* **141**, 1880–1886 (2018).

809 70. K. J. Baines, R. C. West, Sex differences in innate and adaptive immunity impact fetal,
810 placental, and maternal health†. *Biology of Reproduction* **109**, 256–270 (2023).

811 71. S. K. Hunter, *et al.*, Male fetus susceptibility to maternal inflammation: C-reactive protein
812 and brain development. *Psychol Med* **51**, 450–459 (2021).

813 72. J. Mead, Dysanapsis in normal lungs assessed by the relationship between maximal flow,
814 static recoil, and vital capacity. *Am Rev Respir Dis* **121**, 339–342 (1980).

815 73. H. Huang, Y. Mu, S. Li, The biological function of Serpinb9 and Serpinb9-based therapy.
816 *Front Immunol* **15**, 1422113 (2024).

817 74. R. van der Burgh, *et al.*, Reduced serpinB9-mediated caspase-1 inhibition can contribute to
818 autoinflammatory disease. *Oncotarget* **7**, 19265–19271 (2016).

819 75. R. R. Annand, *et al.*, Caspase-1 (interleukin-1beta-converting enzyme) is inhibited by the
820 human serpin analogue proteinase inhibitor 9. *Biochem J* **342**, 655–665 (1999).

821 76. S. K. Sunny, *et al.*, Sex-specific longitudinal association of DNA methylation with lung
822 function. *ERJ Open Res* **7**, 00127–02021 (2021).

823 77. O. Gruzieva, *et al.*, Epigenome-Wide Meta-Analysis of Methylation in Children Related to
824 Prenatal NO2 Air Pollution Exposure. *Environ Health Perspect* **125**, 104–110 (2017).

825 78. C. Ladd-Acosta, *et al.*, Epigenetic marks of prenatal air pollution exposure found in multiple
826 tissues relevant for child health. *Environment International* **126**, 363–376 (2019).

827 79. J. T. Leek, J. D. Storey, Capturing Heterogeneity in Gene Expression Studies by Surrogate
828 Variable Analysis. *PLOS Genetics* **3**, e161 (2007).

829 80. W. F. Davidson, *et al.*, Report from the National Institute of Allergy and Infectious Diseases
830 workshop on "Atopic dermatitis and the atopic march: Mechanisms and interventions."
831 *Journal of Allergy and Clinical Immunology* **143**, 894–913 (2019).

832 81. M. C. M. Strik, *et al.*, Human mast cells produce and release the cytotoxic lymphocyte
833 associated protease granzyme B upon activation. *Molecular Immunology* **44**, 3462–3472
834 (2007).

835 82. Y.-S. Lee, *et al.*, Epigenome-scale comparison of DNA methylation between blood leukocytes
836 and bronchial epithelial cells. *Epigenomics* **13**, 485–498.

837 83. M. de Vries, *et al.*, From blood to lung tissue: effect of cigarette smoke on DNA methylation
838 and lung function. *Respir Res* **19**, 212 (2018).

839 84. M. H. Chan, S. M. Merrill, C. Konwar, M. S. Kobor, An integrative framework and
840 recommendations for the study of DNA methylation in the context of race and ethnicity.
841 *Discov Soc Sci Health* **3**, 9 (2023).

842 85. National Academies of Sciences, Engineering, and Medicine, *Rethinking Race and Ethnicity*
843 *in Biomedical Research* (The National Academies Press, 2025).

844