Title: Prenatal particulate matter exposure is associated with childhood saliva DNA methylation

Authors: Kelly M. Bakulski[1], Jonah D. Fisher[2], John F. Dou[1], Arianna Gard[3], Lisa Schneper[4], Daniel A. Notterman[4], Erin B. Ware[2], Colter Mitchell[2]

Affiliations:
[1]School of Public Health, University of Michigan, Ann Arbor, Michigan, USA
[2]Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA
[3]University of Maryland, College Park, Maryland, USA
[4]Princeton University, Princeton, New Jersey, USA

Abstract

Background: Exposure *in utero* to particulate matter (PM2.5 and PM10) is associated with maladaptive health outcomes. Although exposure to prenatal PM2.5 and PM10 have cord blood DNA methylation signatures at birth, signature persistence into childhood and saliva cross-tissue applicability has not been tested.

Methods: In the Fragile Families and Child Wellbeing Study, a United States 20-city birth cohort, average residential PM2.5 and PM10 during pregnancy was estimated using air quality monitors with inverse distance weighting. Saliva DNA methylation at ages 9 (n=749) and 15 (n=793) was measured using the Illumina HumanMethylation 450k BeadArray. Cumulative DNA methylation scores for particulate matter were estimated by weighting participant DNA methylation at each site by meta-analysis effect estimates from Gruzieva et al. 2019 and standardizing the sums. Using mixed effects regression analysis, we tested the associations between cumulative DNA methylation scores at ages 9 and 15 and PM exposure during pregnancy, adjusted for child sex, age, race/ethnicity, maternal income to needs ratio, nonmartial birth status, and saliva cell type proportions.

Results: Our study sample was 50.5% male, 56.3% non-Hispanic Black, and 19.8% Hispanic, with median income to needs ratio of 1.4. In the third trimester, mean PM2.5 exposure levels were 27.9 ug/m^3/day (standard deviation: 7.0, 23.7% of observations exceeded safety standards) and PM10 were 15.0 ug/m^3/day (standard deviation: 3.1). An interquartile range increase in PM2.5 exposure (10.73 μg/m^3/day) was associated with -0.0287 standard deviation lower cumulative DNA methylation score for PM2.5 (95% CI: -0.0732, 0.0158, p=0.20) across all participants. An interquartile range increase in PM10 exposure (3.20 μg/m^3/day) was associated with -0.1472 standard deviation lower cumulative DNA methylation score for PM10 (95% CI: -0.3038, 0.0095, p=0.06) across all participants. The PM10 findings were driven by the age 15 subset where an interquartile range increase in PM10 exposure was associated with -0.024 standard deviation lower cumulative DNA methylation score for PM10 (95% CI: -0.043, -0.005, p=0.012). Findings were robust to adjustment for PM exposure at ages 1 and 3.

Conclusion: *In utero* PM10 associated DNA methylation differences persist until age 15 and can be detected in saliva. Benchmarking the persistence and cell type generalizability is critical for epigenetic exposure biomarker assessment.

Keywords: DNA methylation, air pollution, particulate matter, saliva, biomarker

Introduction

Air pollution exposure *in utero* is associated with adverse pregnancy outcomes (Lamichhane et al. 2015) and postnatal health problems, such as impaired neurodevelopment (Volk et al. 2020), increased likelihood of autism spectrum disorder (Chun et al. 2020), and impaired lung function in children (Schultz et al. 2017). One component of air pollution is particulate matter. Particulate matter (PM) is classified based on the size of the particle. Smaller particles with diameter less than 2.5 μM (PM2.5) contain primary combustion particles and secondary particles (Kelly and Fussell 2012). Larger particles with diameters greater than 2.5 μM and less than 10 μM (PM10) are generally visible and include smoke, dust, and mechanically generated particles (Kelly and Fussell 2012). These particles differ in their typical chemical make up as well as their ability to penetrate the lungs (Brown et al. 1950). Both types of PM are capable of crossing the placenta (Bové et al. 2019) and thus directly impact the developing fetus. Characterizing the molecular consequences of air pollution exposure during the *in utero* period is critical to understanding environmentally-mediated health disparities that emerge early in life and predict lifelong outcomes.

In utero exposure to PM2.5 and PM10 have well-documented associations with infant DNA methylation in cord blood (Isaevska et al. 2021). In a meta-analysis of nine cohort studies, an interquartile range increase in PM2.5 exposure (2μg/m^3) was associated with 3% lower DNA methylation near the *PLXNA4* gene (Gruzieva et al. 2019). Similarly, an interquartile range increase PM10 exposure (5.6μg/m^3) was associated with 1% higher DNA methylation near the *GNB2L1* gene (Gruzieva et al. 2019). The prenatal period is a window of susceptibility for epigenetic changes such as DNA methylation (Faulk and Dolinoy 2011). Indeed, DNA methylation at birth has been shown to be an effective biomarker of prenatal environmental exposures (Ladd-Acosta and Fallin 2019). However, the persistence of these DNA methylation signatures of air pollution into childhood and the blood to saliva cross-tissue applicability has not been tested.

The goal of this study is to investigate the persistence of air pollution DNA methylation biomarkers into childhood and the cross-tissue applicability between blood and saliva. Specifically, we tested the associations between *in utero* air pollution exposure and saliva DNA methylation in the Fragile Families and Child Wellbeing Study. We hypothesized that *in utero* air pollution exposure would be associated with DNA methylation at ages 9 and 15 in the Fragile Families and Child Wellbeing Study. Further, we hypothesized that *in utero* air pollution exposure would be associated with cumulative DNA methylation scores for *in utero* air pollution exposure.

Methods

Study population

The Fragile Families and Child Wellbeing Study is a United States 20-city birth cohort that recruited children born between 1998 and 2000 from hospitals (Reichman et al. 2001). Unmarried mothers were oversampled by a ratio of 3:1, as the original aims of the study were to examine the downstream effects of families disproportionately likely to break up and live in poverty than more advantaged and historically privileged family structures. Assessments continued at ages 1, 3, 5, 9, and 15; additional follow up is ongoing. Data collection included medical records extraction, in home assessments, biosample collection, and surveys of the mother, father, primary caregiver, child, and teacher. This cohort has been extensively used to characterize pathways linking family structure, socioeconomic resources, and child and family outcomes (Waldfogel et al. 2010). Participants provided written informed consent for the study. These secondary data analyses were approved by the University of Michigan Institutional Review Board (IRB, HUM00129826).

Covariates and exposure measures

Demographic variables were derived from maternal self-report questionnaire data at baseline birth and included maternal race/ethnicity (Non-Hispanic Black, Non-Hispanic White, Hispanic, Other), household income to needs ratio, city of birth (to describe the sampling strategy), and child sex (male, female).

Air pollution exposure was estimated by the Fragile Families and Child Wellbeing Study, using methods described previously (Ailshire and Crimmins 2014). At the birth interview, mothers reported their current addresses, which were geocoded and assigned a United States Census tract according to the 2000 Dicennial Census (for more information see https://fragilefamilies.princeton.edu/restricted). Air pollution data was downloaded from the Inter-university Consortium for Political and Social Research (manifest # 27864) and prepared by the RAND Center for Population Health and Health Disparities (Escarce et al. 2011). Ambient air quality measures of particulate matter (PM2.5 and PM10; μg/m^3) were obtained from the US Environmental Protection Agency (EPA) Air Quality System database (US EPA 2018) spanning 1998 to 2000 (the year of the child's birth in the Fragile Families and Child Wellbeing Study). Quarterly PM2.5 and PM10 concentrations

were estimated per Census tract (based on 2000 Dicennial Census definitions) and matched to the Census tract where Fragile Families and Child Wellbeing Study mothers reported living at the birth of the target child. The date of the child's birth was used to infer PM concentrations during the third trimester. PM concentrations were based on a 24-hour mean of PM monitors within 60km of the Census tract, weighted by the inverse distance from the tract centroid to the PM monitors (i.e., the nearest PM monitoring are assigned a larger weight in the average PM estimate).

DNA methylation measures and cumulative DNA methylation scores

Child saliva samples at the age 9 and 15 home visits were collected in Oragene kits. DNA was extracted manually following DNA Genotek's purification protocol using prepIT L2P. DNA was bisulfite treated and cleaned using the Zymo Research EZ DNA Methylation kit. Saliva DNA methylation at ages 9 and 15 was measured using the Illumina HumanMethylation 450k BeadArray, imaged using the Illumina iScan system.

DNA methylation image data (IDAT) were processed in R statistical software (3.5) using the minfi package (Aryee et al. 2014). The IDAT pairs (n=1811) were read into an RGChannelset. The minfi *preprocessNoob* function was used to normalize dye bias and apply background correction before the beta matrix was derived. Further quality control was applied using the ewastools (Heiss and Just 2018) package. Samples that were dropped for QC reasons include: >10% of sites have detection p-value >0.01 (n=43), DNA methylation predicted sex discordant with recorded sex (n=20), and abnormal sex chromosome intensity (n=3). Probes were removed if they had detection p-value >0.01 in 5% of samples (n=26,830) or were identified as cross-reactive (n=27,782) (Chen et al. 2013). Relative proportions of immune and epithelial cell types were estimated from DNA methylation measures using a childhood saliva reference panel (Middleton et al. 2021).

Our primary cumulative DNA methylation scores were estimated by z-score standardizing participant DNA methylation at each site, weighting the site by the meta-analysis effect estimates (Gruzieva et al. 2019) for either PM2.5 or PM10, and taking the sum across all sites for each participant. Methods for cumulative DNA methylation scores are evolving, thus as sensitivity measures, we calculated two secondary cumulative DNA methylation scores. First, we used the direct participant DNA methylation levels (not transformed), weighted by meta-analysis effect estimates, and summed for each participant. Second, we mean-centered the participant DNA methylation levels, weighted by the meta-analysis effect estimates, and summed for each participant. All three DNA methylation scores for each exposure (PM2.5 and PM10) were then z-score standardized within our study sample for interpretability.

Statistical analyses

All analyses were conducted in R version 4.1.0. Code to complete analyses is available (https://github.com/bakulskilab). Distributions of covariates were described using mean and standard deviation for continuous variables with count and frequency for categorical variables. Samples were included with complete data on exposure, DNA methylation, and demographic information. The included sample was compared to the excluded sample using t-tests for continuous variables and Fisher's exact test for categorical variables. We described the sample distributions stratified by study visit (age 9 and age 15). We dichotomized exposures at the median for the study sample and tested for bivariate differences in covariates by exposure. We similarly dichotomized DNA methylation scores at the median for the study sample and tested for bivariate differences in covariates by exposure. In parallel models stratified for each DNA methylation study visit (age 9 or 15) and for each exposure (PM2.5, PM10), we used multivariable linear regression, to test cumulative DNA methylation scores for associations with pregnancy exposure levels, adjusted for child sex, child age at DNA methylation measure, maternal income to needs ratio, maternal marital status, maternal race/ethnicity, and cell type proportions. When jointly considering both DNA methylation study visits, we used mixed effects regression models, accounting for within-participant measures with a random intercept in the nlme packge (Pinheiro et al. 2021). Mixed effects models were also adjusted for child sex, child age at DNA methylation measure, maternal income to needs ratio, maternal marital status, maternal race/ethnicity, and cell type proportions. We reported the fixed effects estimates for an interquartile range increase in the relevant exposure measure, 95% confidence intervals, and p-values for the association.

Sensitivity analyses

To assess the robustness of our findings, we performed several sensitivity analyses. First, we conducted parallel analyses on the alternative cumulative DNA methylation score calculation approaches (untransformed, centered). Second, we performed analyses mutually adjusted for both exposure types. Third,

we adjusted for postnatal air pollution exposure at ages 1 and 3. Fourth, we tested the specificity of the exposure cumulative DNA methylation score by testing the association of an NO_2 cumulative DNA methylation score with PM2.5 or PM10 exposure. Fifth, we tested single DNA methylation sites associated with PM10 in prior meta-analysis results in cord blood (Gruzieva et al. 2019). We tested DNA methylation levels at these sites (cg00905156, cg06849931, cg15082635, cg18640183, cg20340716, cg24127244) in saliva at age 15 for association with PM10 exposure at birth. We compared the effect estimates from our findings and prior results.

Results

Study sample descriptive statistics

Among 1811 study samples measured for DNA methylation, information on additional key covariates was available and the DNA methylation data passed quality control on 1542 observations (**Supplemental Figure 1**). Included observations were similar to the excluded observations, except the included observations were more likely to be from the age 15 study visit, and to be from participants that self-report as non-Hispanic Black (**Supplemental Table 1**). The included study sample had 749 participants from the age 9 study visit and 793 participants from the age 15 study visit (**Table 1**). There were 747 participants with measures in both study visits in the included sample. Children in the included sample were 50.5% male, 56.3% non-Hispanic Black, 19.8% Hispanic, and the mothers had a median income to needs ratio of 2.27 at birth of the child.

PM2.5 concentrations during pregnancy were available on 795 unique participants and PM10 concentrations were available on 736 participants (**Supplemental Figure 2**). During pregnancy, PM2.5 levels ranged from 14.3 to 45.0 $\mu g/m^3$/day with a mean of 27.9 $\mu g/m^3$/day (**Supplemental Figure 3A**). EPA standards state that 24-hour PM2.5 averages should not exceed 35 $\mu g/m^3$ (Environmental Protection Agency (EPA) 2021). During the third trimester of pregnancy, 23.7% of the age 15 analytic sample exceeded this standard. Pregnancy levels of PM2.5 were correlated with levels of PM2.5 at age 1 (Pearson correlation=0.54, p-value<$2*10^{-16}$, **Supplemental Figure** 4) and with levels of PM2.5 at age 3 (Pearson correlation=0.57, p-value<$2*10^{-16}$). During pregnancy, PM10 levels ranged from 7.5 to 20.2 $\mu g/m^3$/day with a mean of 15.0 $\mu g/m^3$/day (**Supplemental Figure 3B**). PM10 levels during the third trimester of pregnancy did not exceed EPA standards of a maximum 24-hours concentration of 150 $\mu g/m^3$ (Environmental Protection Agency (EPA) 2021). Pregnancy levels of PM10 were correlated with levels of PM10 at age 1 (Pearson correlation=0.71, p-value<$2*10^{-16}$) and with levels of PM10 at age 3 (Pearson correlation=0.69, p-value<$2*10^{-16}$). During pregnancy, PM2.5 and PM10 levels were moderately correlated (Pearson correlation=0.2, p-value=$7*10^{-8}$).

To calculate cumulative DNA methylation scores for air pollution exposure, we weighted our DNA methylation data using published individual CpG site regression effect estimates from cord blood DNA methylation associated with pregnancy air pollution exposure (Gruzieva et al. 2019). We generated separate scores for PM2.5, PM10, and NO_2 exposure (using weights from three separate epigenome-wide association tests) and scores were normally distributed within the sample (**Supplemental Figure 5**). We used 3 methods to calculate the cumulative DNA methylation scores and within each pollutant; the scores from these three different methods were highly correlated (Pearson correlations ranging from 0.79-1, **Supplemental Figure 6**). The cumulative DNA methylation scores across pollutants were less highly correlated (Pearson correlation ranging from 0.06-0.71).

Associations between exposure and DNA methylation scores

In bivariate testing, PM2.5 exposure levels and the cumulative DNA methylation score for PM2.5 were not associated in the combined study sample (p=0.12, **Supplemental Figure 7**), in the age 9 subset (p=0.13), nor in the age 15 subset (p=0.48). In mixed effects regression analyses adjusting for age at DNA methylation sample, child sex, maternal race-ethnicity, maternal income to needs ratio, proportion epithelial cells, and proportion immune cells, findings were consistent with the bivariate results (**Table 2**). An interquartile range increase in PM2.5 exposure (10.73 $\mu g/m^3$/day) was associated with -0.0287 standard deviation lower cumulative DNA methylation score for PM2.5 (95% CI: -0.0732, 0.0158, p=0.20) across all participants. Consistent null findings were observed with cross-sectional multivariable linear regression analyses in the age 9 and age 15 sample subsets and with all three methods for cumulative DNA methylation score calculation.

In bivariate testing, PM10 exposure levels and cumulative DNA methylation score for PM10 were not associated in the combined study sample (p=0.32, **Supplemental Figure 7**), in the age 9 subset (p=0.22), nor in the age 15 subset (p=0.78). In adjusted mixed effects regression analyses, we observed associations between PM10 exposure levels and cumulative DNA methylation score for PM10 (**Table 2**). An interquartile range increase in PM10 exposure (3.20 $\mu g/m^3$/day) was associated with -0.1472 standard deviation lower

cumulative DNA methylation score for PM10 (95% CI: -0.3038, 0.0095, p=0.06) across all participants. In all participants, consistent negative associations between PM10 exposure levels and PM10 cumulative DNA methylation score were observed across all three methods for cumulative DNA methylation score calculation. These findings were driven by the age 15 subset, where an interquartile range increase in PM10 exposure was associated with -0.024 standard deviation lower poly DNA methylation score for PM10 (95% CI: -0.043, -0.005, p=0.012).

Sensitivity analyses

To assess the robustness of our findings, we performed several sensitivity analyses. In all sensitivity analyses, we again observed that prenatal PM2.5 exposure was not associated with PM2.5 cumulative DNA methylation score. However, we continued to observe that prenatal PM10 exposure was associated PM10 cumulative DNA methylation score, particularly in the age 15 sample. First, we repeated the regression analyses with additional adjustment for air pollution levels at age 1 (**Supplemental Table 2**). The association between prenatal PM10 exposure and age 15 PM10 cumulative DNA methylation score was robust to adjustment for postnatal exposure at age 1 (-0.0302, 95% CI: -0.0556, -0.0047, p=0.020). Second, we repeated the multivariable linear regression analyses with additional adjustment for air pollution levels at age 3 (**Supplemental Table 3**). The association between prenatal PM10 exposure with age 15 PM10 cumulative DNA methylation score was robust to adjustment for postnatal exposure at age 3 (-0.0343, 95% CI: -0.0604, -0.0082, p=0.010). Third, we repeated the regression analyses with additional adjustment for prenatal air pollution levels of the other type of particulate matter (**Supplemental Table 4**). The association between prenatal PM10 exposure with age 15 PM10 cumulative DNA methylation score was robust to adjustment for prenatal PM2.5 exposure (-0.0231, 95% CI: -0.0424, -0.0038, p=0.019). Fourth, we tested for adjusted associations between prenatal PM2.5 or prenatal PM10 exposure levels with cumulative DNA methylation scores for NO_2 (**Supplemental Table 5**). Prenatal PM10 exposure was associated with age 15 NO_2 cumulative DNA methylation score (0.1271, 95% CI: 0.0520, 0.2022, p=0.0009).

We next attempted to replicate six individual CpG sites previously associated with air pollution exposure in a cord blood meta-analysis at genome-wide significance levels. We observed that DNA methylation at two of these sites in saliva at age 15 was associated with PM10 at birth (**Table 3**). Specifically, at cg18640183 (associated with the *P4HA2* gene) an IQR-unit increase in PM10 exposure at birth was associated with 0.119 lower percent DNA methylation (p=0.027). At cg20340716 (associated with the *USP43* gene) an IQR-unit increase in PM10 exposure at birth was associated with 0.135 higher percent DNA methylation (p=0.015).

Discussion

In the nationwide, population-based Fragile Families and Child Wellbeing birth cohort, we observed prenatal PM10 exposure was associated with saliva DNA methylation in childhood. Previous meta-analyses have documented that prenatal air pollution exposure is associated with cord blood DNA methylation at birth (Gruzieva et al. 2019). We used effect estimates from these associations to weight measures of saliva DNA methylation at ages 9 and 15 to create cumulative DNA methylation scores for prenatal air pollution exposure. Using these cumulative DNA methylation scores as well as candidate DNA methylation sites, we observed that prenatal third trimester PM10 associated DNA methylation differences persist until age 15 and can be detected in saliva. Benchmarking the persistence and cell type generalizability of epigenetic exposure biomarker assessment is critical for application to epidemiologic application.

A recent systematic review of prenatal air pollution and infant DNA methylation identified 21 studies focusing on particulate matter (Isaevska et al. 2021). Most of these studies examined candidate genes or global DNA methylation. There were two prior particulate matter epigenome-wide association studies identified, including one in blood (Gruzieva et al. 2019) and one in placenta (Abraham et al. 2018). The particulate matter epigenome-wide association study conducted in blood (Gruzieva et al. 2019) was done by the PACE consortium and included 1949 participants in the PM10 discovery analysis and 1551 participants in the PM2.5 discovery analysis. Their findings replicated in an independent cord blood sample and postnatal blood showed persistence of the findings until ages 15 and 16. This study provided the weights for the cumulative DNA methylation scores in our current analysis. We observed an interquartile range increase in PM10 exposure (3.20 $\mu g/m^3$/day) was associated with -0.024 standard deviation lower poly DNA methylation score for PM10 (p=0.012) among age 15 participants. The observed direction of effect was opposite of our initial hypothesis

(higher exposure would be associated with higher cumulative DNA methylation score). However, the weights used to build the cumulative DNA methylation were from an analysis in blood and we measured DNA methylation in saliva, and prior research has shown cross-tissue differences in magnitude and direction of effects for other traits (Walton et al. 2016). Among six individual DNA methylation sites that were previously associated with PM10 exposure in cord blood (Gruzieva et al. 2019), we observed an association in our study at two of those sites (from the *P4HA2* and *USP43* genes). We observed associations between prenatal PM10 exposure and saliva DNA methylation at age 15, however the direction of association was opposite prior associations in cord blood at birth.

For several additional environmental exposures, epigenetic biomarkers in peripheral tissues have been shown to be specific and reproducible (Bakulski and Fallin 2014). The most well characterized environmental epigenetic biomarker is for smoke exposure. Prenatal smoke exposure associated DNA methylated sites have been documented in cord blood meta-analyses (Joubert et al. 2016b). These associations are persistent to age 5 (Ladd-Acosta et al. 2016) and adolescence (Rauschert et al. 2019). There is also evidence that folate or prenatal vitamin exposure during pregnancy has reproducible DNA methylation associations in cord blood (Bakulski et al. 2020; Joubert et al. 2016a), though the persistence of these associations postnatally has not yet been tested. In this study, we examined particulate matter, which is a broad exposure and particles can contain multiple types of toxicants that can vary geographically (Pérez et al. 2008). This variability across studies and within our US nationwide study may be part of why we did not observe an association between PM2.5 exposure and cumulative DNA methylation score for PM2.5. Similar to prior work on smoking and folate/prenatal vitamin exposures, this paper adds to a body of evidence suggesting that air pollution, specifically PM10, has a reproducible DNA methylation signature in peripheral tissues.

Cumulative DNA methylation scores are an emerging area of DNA methylation research. They are an approach to apply prior epigenome-wide discovery results to an independent cohort and aggregate epigenome-wide information into a single value. Previous research has documented cumulative DNA methylation score utility as a marker for smoking exposure, which was able to predict prenatal smoke exposure 30 years later in blood with an area under the curve of 0.72 (95% confidence interval: 0.69, 0.76) in the ALSPAC cohort (Richmond et al. 2018). cumulative DNA methylation scores are analogous to polygenic scores, which are widely used in genetic epidemiology (Lewis and Vassos 2020). Early findings suggest that cumulative DNA methylation scores for some traits may explain a proportion of the variance comparable to polygenic scores. For example, when predicting body mass index in the Lothian Birth Cohorts, the cumulative DNA methylation score explained 7% of the variance, the polygenic score predicted 8%, and the model containing both predicted 14% (Shah et al. 2015). This suggests that the DNA methylation and genetic components for that trait may be independent. Further testing of cumulative DNA methylation scores for additional exposures and traits will be needed to assess the generalizability of these findings.

There are several strengths and limitations of this study. First, the Fragile Families and Child Wellbeing Study is a well-characterized, large, diverse birth cohort with prospective DNA methylation sample collection at two time points. Much of epigenetic epidemiology is cross-sectional and focused on Non-Hispanic White participants (Collier et al. 2020). Particulate matter exposure levels were quantitated based on residential history, which is standard in the field (Xie et al. 2017). Participants may spend considerable time away from home, thus there is likely measurement error in the exposure estimates, which has been shown to bias estimates towards the null (Setton et al. 2011). Our air pollution exposure estimates were averaged for the third trimester of pregnancy. The meta-analysis used for cumulative DNA methylation score weights averaged exposure estimates over the entire pregnancy. Future studies may separate out exposures specific to trimester or months of pregnancy to investigate windows of susceptibility. DNA methylation was measured on reproducible genome-wide array using methods consistent with prior research. Our cumulative DNA methylation scores were calculated using effect estimates from a large consortium (Gruzieva et al. 2019) and importantly our study sample was independent from the sample that generated the weights (Wand et al. 2021). We performed multiple essential sensitivity analyses, including using three methods for cumulative DNA methylation score development, adjustment for postnatal exposure, and adjustment for alternate air pollution exposures. Together, these study design and analytic design elements contribute to rigorous research.

Particulate matter air pollution exposure is associated with global mortality (Daniels et al. 2000) and adverse pregnancy outcomes (Lamichhane et al. 2015). In particular, exposure during the *in utero* period has lasting health effects (Volk et al. 2020). Examining the DNA methylation consequences of *in utero* air pollution is useful to develop biomarkers of air pollution exposure, as well as to document potential molecular intermediates of health effects. Prior research documented *in utero* air pollution exposure was associated with cord blood DNA methylation. We newly showed the air pollution associated DNA methylation differences are detectable at age 15 and they are detectable in saliva. This study demonstrates the persistence and cross-tissue utility of DNA methylation as a biomarker of air pollution exposure, with important implications for future epidemiology studies.

Table 1. Univariate descriptive statistics in the analytic sample of the Fragile Families and Child Wellbeing Study. Participants are grouped by age at DNA methylation assessment. Particulate matter (PM).

	[ALL] N=1542	Age 9 visit N=749	Age 15 visit N=793	P-value	Number of observations
Child characteristics					
Sex				0.836	1542
Female	769 (49.9%)	371 (49.5%)	398 (50.2%)		
Male	773 (50.1%)	378 (50.5%)	395 (49.8%)		
Race/ethnicity				0.999	1542
Non-Hispanic White	256 (16.6%)	124 (16.6%)	132 (16.6%)		
Non-Hispanic Black	868 (56.3%)	420 (56.1%)	448 (56.5%)		
Hispanic	306 (19.8%)	150 (20.0%)	156 (19.7%)		
Other	44 (2.85%)	21 (2.80%)	23 (2.90%)		
Multi-Racial	68 (4.41%)	34 (4.54%)	34 (4.29%)		
Age at DNA methylation measure	12.4 (3.07)	9.30 (0.34)	15.4 (0.49)	-	1542
Maternal characteristics at birth					
Income/poverty	2.27 (2.49)	2.29 (2.51)	2.25 (2.48)	0.728	1542
Marital status				0.791	1542
Married	365 (23.7%)	180 (24.0%)	185 (23.3%)		
Not Married	1177 (76.3%)	569 (76.0%)	608 (76.7%)		
Race/ethnicity				0.998	1542
Non-Hispanic White	274 (17.8%)	133 (17.8%)	141 (17.8%)		
Non-Hispanic Black	902 (58.5%)	437 (58.3%)	465 (58.6%)		
Hispanic	312 (20.2%)	153 (20.4%)	159 (20.1%)		
Other	54 (3.50%)	26 (3.47%)	28 (3.53%)		
City of residence				1	1542
Oakland	114 (7.39%)	57 (7.61%)	57 (7.19%)		
Baltimore	97 (6.29%)	46 (6.14%)	51 (6.43%)		
Detroit	312 (20.2%)	148 (19.8%)	164 (20.7%)		
Newark	51 (3.31%)	27 (3.60%)	24 (3.03%)		
Philadelphia	120 (7.78%)	59 (7.88%)	61 (7.69%)		
Richmond	143 (9.27%)	73 (9.75%)	70 (8.83%)		
Corpus Christi	93 (6.03%)	44 (5.87%)	49 (6.18%)		
Indianapolis	92 (5.97%)	47 (6.28%)	45 (5.67%)		
Milwaukee	81 (5.25%)	38 (5.07%)	43 (5.42%)		
New York	30 (1.95%)	14 (1.87%)	16 (2.02%)		
San Jose	79 (5.12%)	40 (5.34%)	39 (4.92%)		
Boston	18 (1.17%)	9 (1.20%)	9 (1.13%)		
Nashville	25 (1.62%)	13 (1.74%)	12 (1.51%)		
Chicago	73 (4.73%)	31 (4.14%)	42 (5.30%)		
Jacksonville	20 (1.30%)	10 (1.34%)	10 (1.26%)		
Toledo	87 (5.64%)	39 (5.21%)	48 (6.05%)		
San Antonio	31 (2.01%)	15 (2.00%)	16 (2.02%)		
Pittsburgh	41 (2.66%)	21 (2.80%)	20 (2.52%)		
Norfolk	35 (2.27%)	18 (2.40%)	17 (2.14%)		

	[ALL] N=1542	Age 9 visit N=749	Age 15 visit N=793	P-value	Number of observations
Air pollution exposure (μg/m^3/day)					
PM2.5 at birth	27.9 (7.04)	27.8 (7.07)	28.0 (7.02)	0.546	1542
Missing	.	.	.	.	0
PM10 at birth	15.0 (3.06)	15.0 (3.09)	15.0 (3.03)	0.927	1425
Missing	117 (100%)	59 (100%)	58 (100%)	.	117
PM2.5 at age 1	25.9 (5.29)	25.8 (5.32)	26.0 (5.27)	0.46	1454
Missing	88 (100%)	39 (100%)	49 (100%)	.	88
PM10 at age 1	14.6 (3.05)	14.6 (3.06)	14.6 (3.04)	0.892	1452
Missing	90 (100%)	41 (100%)	49 (100%)	.	90
PM2.5 exposure at age 3	26.7 (7.72)	26.6 (7.77)	26.7 (7.67)	0.812	1405
Missing	137 (100%)	65 (100%)	72 (100%)	.	137
PM10 exposure at age 3	14.2 (3.28)	14.2 (3.29)	14.3 (3.27)	0.664	1414
Missing	128 (100%)	61 (100%)	67 (100%)	.	128
DNA methylation score					
PM2.5 methylation score (raw)	-0.05 (0.75)	-0.09 (0.71)	-0.02 (0.77)	0.058	1542
PM2.5 methylation score (centered)	-0.05 (0.75)	-0.09 (0.71)	-0.02 (0.77)	0.058	1542
PM2.5 methylation score (z-score)	-0.07 (0.52)	-0.08 (0.51)	-0.06 (0.54)	0.538	1542
PM10 methylation score (raw)	-0.08 (0.51)	-0.15 (0.47)	-0.02 (0.55)	<0.001	1542
PM10 methylation score (centered)	-0.08 (0.51)	-0.15 (0.47)	-0.02 (0.55)	<0.001	1542
PM10 methylation score (z-score)	-0.09 (0.22)	-0.12 (0.21)	-0.06 (0.22)	<0.001	1542
NO2 methylation score (raw)	-0.02 (0.89)	-0.05 (0.87)	0.01 (0.91)	0.188	1542
NO2 methylation score (centered)	-0.02 (0.89)	-0.05 (0.87)	0.01 (0.91)	0.188	1542
NO2 methylation score (z-score)	-0.05 (0.71)	-0.05 (0.69)	-0.06 (0.73)	0.734	1542
Saliva cell composition					
Percent immune cells	93.9 (13.6)	95.3 (11.8)	92.5 (14.9)	<0.001	1542
Percent epithelial cells	6.15 (13.6)	4.69 (11.8)	7.52 (14.9)	<0.001	1542
Site specific DNA methylation					
cg00905156	2.48 (1.52)	2.30 (1.40)	2.64 (1.60)	<0.001	1542
cg06849931	73.4 (13.5)	74.8 (12.1)	72.0 (14.5)	<0.001	1542
cg15082635	1.91 (0.77)	1.74 (0.60)	2.07 (0.88)	<0.001	1542
cg18640183	4.82 (1.21)	4.79 (1.21)	4.84 (1.20)	0.380	1542
cg20340716	92.8 (1.37)	92.7 (1.46)	92.9 (1.27)	0.002	1542
cg24127244	2.46 (0.65)	2.34 (0.57)	2.57 (0.71)	<0.001	1542

Table 2. Adjusted associations between cumulative DNA methylation score for prenatal particulate matter (PM) exposure and levels of prenatal particulate matter exposure in the Fragile Families and Child Wellbeing Study. All ages models are mixed effects regression models with random intercepts for participants. Age stratified models are linear regression models. Models are adjusted for age at DNA measurement, child sex, maternal race, maternal income to poverty ratio, proportion epithelial cells, and proportion immune cells. Effect estimates and confidence intervals are provided for an interquartile increase in exposure (PM2.5: 10.73 ug/m3/day; PM10: 3.20 ug/m3/day).

				Raw DNA methylation				Centered DNA methylation				Centered & scaled DNA methylation			
Exposure	Age	N_{indiv}	N_{obs}	Effect estimate	Lower confidence interval	Upper confidence interval	P-value	Effect estimate	Lower confidence interval	Upper confidence interval	P-value	Effect estimate	Lower confidence interval	Upper confidence interval	P-value
PM2.5	All	787	1542	-0.029	-0.073	0.016	0.206	-0.029	-0.073	0.016	0.206	-0.017	-0.051	0.018	0.345
PM2.5	9	749	749	-0.021	-0.070	0.028	0.399	-0.021	-0.070	0.028	0.399	-0.014	-0.054	0.026	0.478
PM2.5	15	793	793	-0.017	-0.065	0.030	0.475	-0.017	-0.065	0.030	0.475	-0.008	-0.047	0.031	0.675
PM10	All	728	1425	-0.147	-0.304	0.010	0.066	-0.147	-0.304	0.010	0.066	-0.133	-0.274	0.008	0.065
PM10	9	690	690	-0.004	-0.023	0.015	0.701	-0.004	-0.023	0.015	0.701	-0.005	-0.021	0.012	0.573
PM10	15	735	735	-0.024	-0.043	-0.005	0.012	-0.024	-0.043	-0.005	0.012	-0.023	-0.039	-0.007	0.005

Table 3. Adjusted associations between single DNA methylation sites and levels of prenatal particulate matter 10 μM exposure at birth in the Fragile Families and Child Wellbeing Study. Multivariable linear regression models have been adjusted for age at DNA measurement, child sex, maternal race, maternal income to poverty ratio, proportion epithelial cells, and proportion immune cells (n=735). Effect estimates and confidence intervals are for an interquartile range increase in exposure (3.20 μg/m^3/day). Published cord blood DNA methylation is from (Gruzieva et al. 2019).

				Saliva DNA methylation age 15				Published cord blood DNA methylation	
DNA methylation site	Nearest Gene	Chr	Position	Effect estimate	Lower confidence interval	Upper confidence interval	P-value	Effect estimate	P-value
cg00905156	*FAM13A*	4	89744363	-0.048	-0.191	0.094	0.506	0.001	3.55E-07
cg06849931	*NOTCH4*	6	32165893	0.160	-0.228	0.547	0.420	−0.001	1.72E-06
cg15082635	*GNB2L1; SNORD96A*	5	180670110	-0.009	-0.085	0.068	0.821	0.001	8.29E-08
cg18640183	*P4HA2*	5	131563610	-0.119	-0.224	-0.014	0.027	0.001	1.80E-06
cg20340716	*USP43*	17	9559558	0.135	0.026	0.244	0.015	−0.002	1.50E-07
cg24127244	*SRPRB*	3	133524572	-0.015	-0.076	0.046	0.627	0.001	7.33E-07

Acknowledgements

Research reported was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) of the National Institutes of Health under award numbers R01HD36916, R01HD39135, R01HD40421, R01 HD076592, and R01 MH103761, as well as a consortium of private foundations. Analyses were supported by R01 MD011716 (PI: Mitchell), R01 AG067592 (MPI: Bakulski, Ware), R01 ES025531 (PI: Fallin), and R01 ES025574 (PI: Schmidt). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References cited

Abraham E, Rousseaux S, Agier L, Giorgis-Allemand L, Tost J, Galineau J, et al. 2018. Pregnancy exposure to atmospheric pollution and meteorological conditions and placental DNA methylation. Environment International 118:334–347; doi:10.1016/j.envint.2018.05.007.

Ailshire JA, Crimmins EM. 2014. Fine particulate matter air pollution and cognitive function among older US adults. Am J Epidemiol 180:359–366; doi:10.1093/aje/kwu155.

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. 2014. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30:1363–1369; doi:10.1093/bioinformatics/btu049.

Bakulski KM, Dou JF, Feinberg JI, Brieger KK, Croen LA, Hertz-Picciotto I, et al. 2020. Prenatal Multivitamin Use and MTHFR Genotype Are Associated with Newborn Cord Blood DNA Methylation. Int J Environ Res Public Health 17; doi:10.3390/ijerph17249190.

Bakulski KM, Fallin MD. 2014. Epigenetic epidemiology: promises for public health research. Environ Mol Mutagen 55:171–183; doi:10.1002/em.21850.

Bové H, Bongaerts E, Slenders E, Bijnens EM, Saenen ND, Gyselaers W, et al. 2019. Ambient black carbon particles reach the fetal side of human placenta. Nature Communications 10:3866; doi:10.1038/s41467-019-11654-3.

Brown JH, Cook KM, Ney FG, Hatch T. 1950. Influence of Particle Size upon the Retention of Particulate Matter in the Human Lung. Am J Public Health Nations Health 40:450–480; doi:10.2105/ajph.40.4.450.

Chen Y, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. 2013. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 8:203–209; doi:10.4161/epi.23470.

Chun H, Leung C, Wen SW, McDonald J, Shin HH. 2020. Maternal exposure to air pollution and risk of autism in children: A systematic review and meta-analysis. Environmental Pollution 256:113307; doi:10.1016/j.envpol.2019.113307.

Collier AY, Ledyard R, Montoya-Williams D, Qiu M, Dereix AE, Farrokhi MR, et al. 2020. Racial and ethnic representation in epigenomic studies of preterm birth: a systematic review. Epigenomics; doi:10.2217/epi-2020-0007.

Daniels MJ, Dominici F, Samet JM, Zeger SL. 2000. Estimating Particulate Matter-Mortality Dose-Response Curves and Threshold Levels: An Analysis of Daily Time-Series for the 20 Largest US Cities. American Journal of Epidemiology 152:397–406; doi:10.1093/aje/152.5.397.

Environmental Protection Agency (EPA). 2021. Criteria Air Pollutants: National Ambient Air Quality Standards (NAAQS) Table. Available: https://www.epa.gov/criteria-air-pollutants/naaqs-table [accessed 23 July 2021].

Escarce JJ, Lurie N, Jewell A. 2011. RAND Center for Population Health and Health Disparities (CPHHD) Data Core Series: Disability, 2000 [United States].; doi:10.3886/ICPSR27862.v1.

Faulk C, Dolinoy DC. 2011. Timing is everything: the when and how of environmentally induced changes in the epigenome of animals. Epigenetics 6:791–797; doi:10.4161/epi.6.7.16209.

Gruzieva O, Xu C-J, Yousefi P, Relton C, Merid SK, Breton CV, et al. 2019. Prenatal Particulate Air Pollution and DNA Methylation in Newborns: An Epigenome-Wide Meta-Analysis. Environ Health Perspect 127:57012; doi:10.1289/EHP4522.

Heiss JA, Just AC. 2018. Identifying mislabeled and contaminated DNA methylation microarray data: an extended quality control toolset with examples from GEO. Clin Epigenetics 10:73; doi:10.1186/s13148-018-0504-1.

Isaevska E, Moccia C, Asta F, Cibella F, Gagliardi L, Ronfani L, et al. 2021. Exposure to ambient air pollution in the first 1000 days of life and alterations in the DNA methylome and telomere length in children: A systematic review. Environ Res 193:110504; doi:10.1016/j.envres.2020.110504.

Joubert BR, den Dekker HT, Felix JF, Bohlin J, Ligthart S, Beckett E, et al. 2016a. Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns. Nat Commun 7:10577; doi:10.1038/ncomms10577.

Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, et al. 2016b. DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. Am J Hum Genet 98:680–696; doi:10.1016/j.ajhg.2016.02.019.

Kelly FJ, Fussell JC. 2012. Size, source and chemical composition as determinants of toxicity attributable to ambient particulate matter. Atmospheric Environment 60:504–526; doi:10.1016/j.atmosenv.2012.06.039.

Ladd-Acosta C, Fallin MD. 2019. DNA Methylation Signatures as Biomarkers of Prior Environmental Exposures. Curr Epidemiol Rep 6:1–13; doi:10.1007/s40471-019-0178-z.

Ladd-Acosta C, Shu C, Lee BK, Gidaya N, Singer A, Schieve LA, et al. 2016. Presence of an epigenetic signature of prenatal cigarette smoke exposure in childhood. Environ Res 144:139–148; doi:10.1016/j.envres.2015.11.014.

Lamichhane DK, Leem J-H, Lee J-Y, Kim H-C. 2015. A meta-analysis of exposure to particulate matter and adverse birth outcomes. Environ Health Toxicol 30:e2015011; doi:10.5620/eht.e2015011.

Lewis CM, Vassos E. 2020. Polygenic risk scores: from research tools to clinical instruments. Genome Med 12:44; doi:10.1186/s13073-020-00742-5.

Middleton LYM, Dou J, Fisher J, Heiss JA, Nguyen VK, Just AC, et al. 2021. Saliva cell type DNA methylation reference panel for epidemiological studies in children. Epigenetics 1–17; doi:10.1080/15592294.2021.1890874.

Pérez N, Pey J, Querol X, Alastuey A, López JM, Viana M. 2008. Partitioning of major and trace components in PM10–PM2.5–PM1 at an urban site in Southern Europe. Atmospheric Environment 42:1677–1691; doi:10.1016/j.atmosenv.2007.11.034.

Pinheiro J, Bates D, DebRoy S, Sarkar D. 2021. nlme: Linear and nonlinear mixed effects models. R package version 31-152.

Rauschert S, Melton PE, Burdge G, Craig JM, Godfrey KM, Holbrook JD, et al. 2019. Maternal Smoking During Pregnancy Induces Persistent Epigenetic Changes Into Adolescence, Independent of Postnatal Smoke

Exposure and Is Associated With Cardiometabolic Risk. Front Genet 10; doi:10.3389/fgene.2019.00770.

Reichman NE, Teitler JO, Garfinkel I, McLanahan SS. 2001. Fragile Families: sample and design. Children and Youth Services Review 23:303–326; doi:10.1016/S0190-7409(01)00141-4.

Richmond RC, Suderman M, Langdon R, Relton CL, Davey Smith G. 2018. DNA methylation as a marker for prenatal smoke exposure in adults. Int J Epidemiol 47:1120–1130; doi:10.1093/ije/dyy091.

Schultz ES, Litonjua AA, Melén E. 2017. Effects of Long-Term Exposure to Traffic-Related Air Pollution on Lung Function in Children. Curr Allergy Asthma Rep 17:41; doi:10.1007/s11882-017-0709-y.

Setton E, Marshall JD, Brauer M, Lundquist KR, Hystad P, Keller P, et al. 2011. The impact of daily mobility on exposure to traffic-related air pollution and health effect estimates. J Expo Sci Environ Epidemiol 21:42–48; doi:10.1038/jes.2010.14.

Shah S, Bonder MJ, Marioni RE, Zhu Z, McRae AF, Zhernakova A, et al. 2015. Improving Phenotypic Prediction by Combining Genetic and Epigenetic Associations. Am J Hum Genet 97:75–85; doi:10.1016/j.ajhg.2015.05.014.

US EPA O. 2018. AQS User Guide. US EPA. Available: https://www.epa.gov/aqs/aqs-user-guide [accessed 24 March 2021].

Volk HE, Perera F, Braun JM, Kingsley SL, Gray K, Buckley J, et al. 2020. Prenatal air pollution exposure and neurodevelopment: A review and blueprint for a harmonized approach within ECHO. Environmental Research 110320; doi:10.1016/j.envres.2020.110320.

Waldfogel J, Craigie T-A, Brooks-Gunn J. 2010. Fragile Families and Child Wellbeing. Future Child 20: 87–112.

Walton E, Hass J, Liu J, Roffman JL, Bernardoni F, Roessner V, et al. 2016. Correspondence of DNA Methylation Between Blood and Brain Tissue and Its Application to Schizophrenia Research. Schizophr Bull 42:406–414; doi:10.1093/schbul/sbv074.

Wand H, Lambert SA, Tamburro C, Iacocca MA, O'Sullivan JW, Sillari C, et al. 2021. Improving reporting standards for polygenic scores in risk prediction studies. Nature 591:211–219; doi:10.1038/s41586-021-03243-6.

Xie X, Semanjski I, Gautama S, Tsiligianni E, Deligiannis N, Rajan RT, et al. 2017. A Review of Urban Air Pollution Monitoring and Exposure Assessment Methods. ISPRS International Journal of Geo-Information 6:389; doi:10.3390/ijgi6120389.