medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Title Page

Title: Polymethylation Scores for Prenatal Maternal Smoke Exposure Persist Until Age 15 and Are Detected in Saliva

Authors: Freida A. Blostein[1], Jonah Fisher[2], John Dou[1], Lisa Schenper[3], Erin B. Ware[2] Daniel A. Notterman[3]*, Colter Mitchell[2]*, Kelly M. Bakulski[1]*[+]

Affiliations:

[1]Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, Michigan, United States of America

[2]Institute for Social Research, University of Michigan, Ann Arbor, Michigan, United States of America

[3]Department of Molecular Biology, Princeton University, Princeton, New Jersey, United States of America

*Co-senior authors

[+]Corresponding author: bakulski@umich.edu 1415 Washington Heights, Ann Arbor, MI, 48109

Abstract

Background: Prenatal maternal smoking has negative implications for child health. DNA methylation signatures can function as biomarkers of prenatal maternal smoking. However little work has assessed how DNA methylation signatures of prenatal maternal smoking vary across ages, ancestry groups, or tissues. In the Fragile Families and Child Wellbeing study, we tested

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

whether prenatal maternal smoking was associated with salivary polymethylation scores for smoking in participants. We assessed the associations' consistency at ages 9 and 15, their portability across participants from African, European, and Hispanic genetic ancestries and the accuracy of exposure classification using area under the curve (AUC) from receiver operating curve analyses.

Results: We created saliva polymethylation scores using coefficients from a meta-analysis of prenatal maternal smoke exposure and DNA methylation in newborn cord blood. In the full sample at age 9 (n=753), prenatal maternal smoke exposure was associated with a 0.52 (95%CI: 0.36, 0.67) standard deviation higher polymethylation score for prenatal smoke exposure. The direction of the association was consistent when stratified by genetic ancestries. In the full sample at age 15 (n=746), prenatal maternal smoke exposure was associated with a 0.46 (95%CI: 0.3, 0.62) standard deviation higher polymethylation score for prenatal smoke exposure, and the effect size was attenuated among the European and Hispanic genetic ancestry samples. The polymethylation score was reasonably accurate at classifying prenatal maternal smoke exposure (AUC age 9=0.78, *P* value comparing to base model<0.001, age 15=0.77, *P* value comparing to base model<0.001). The polymethylation score showed higher classification accuracy than using a single *a priori* site in the *AHRR* gene (cg05575921 AUC=0.74, *P* value comparing to polymethylation score=0.03; age 15=0.73, *P* value comparing to polymethylation score=0.01).

Conclusions: Prenatal maternal smoking was associated with DNA methylation signatures in saliva samples, a clinically practical tissue. Polymethylation scores for prenatal maternal smoking were portable across genetic ancestries and more accurate than individual DNA methylation sites. DNA polymethylation scores from saliva samples could serve as robust and practical clinical biomarkers of prenatal maternal smoke exposure.

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Keywords

DNA methylation, prenatal maternal smoking, salivary biomarker, longitudinal cohort

Background

Maternal prenatal smoking is a public health concern. Children exposed to cigarette smoking in utero are more likely to have low-birth weight, negative neurodevelopmental outcomes, and asthma [1]. In 2018, 11% of pregnant women in the United States aged 15-44 reported any past-month cigarette use [2]. Measuring prenatal maternal smoking behavior is challenging. Maternal prenatal smoking is underreported due to stigma [3, 4]. The gold-standard for smoking measurements, serum cotinine levels, have a half-life of nine hours in pregnant women, limiting the ability for accurate detection to a short window [5]. Further, serum cotinine measures during pregnancy are rarely available outside of birth cohorts, limiting clinical application as a marker of prenatal exposures. Developing a portable and reliable biomarker of prenatal maternal smoke exposure would have important implications for research and clinical practice.

In infant cord blood and placental tissues, maternal cigarette smoking during pregnancy is associated with reproducible DNA methylation signatures, though this has primarily been tested in European cohorts [6–8]. Prenatal maternal cigarette smoking exposure is associated with postnatal DNA methylation signatures in blood from cohorts of older children and adults [9, 10]. Few studies have considered the persistence of the association between maternal prenatal cigarette smoking and DNA methylation at different ages in the same individuals. Additionally, little work has considered the portability of the association between maternal prenatal cigarette smoking and DNA methylation outside of cord blood and placental tissue. DNA methylation drives cell differentiation and cell type proportions differ across tissue types. Epigenetic markers

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

of prenatal maternal smoking are known to differ between cord blood and placental samples [7]. Cord blood and placental samples are unlikely to be available in clinical settings. While peripheral blood is a more realistic clinical sample, saliva is even easier to collect. To our knowledge, no study has evaluated associations between maternal prenatal cigarette smoking and salivary DNA methylation.

Additionally, DNA methylation can vary by genetics [11–14], and DNA methylation signatures of own-smoking differ across genetic ancestry groups [15–18]. Only a few studies have examined associations between prenatal maternal cigarette smoking and DNA methylation in non-European ancestry populations. Among 954 infants, 70% of whom had mothers identifying as Black, prenatal maternal smoking was associated with cord blood DNA methylation at 38 CpG sites [19]. The direction of the associations between prenatal maternal cigarette smoking and DNA methylation at these sites was generally similar in children of Black and non-Black participants [19]. Among 572 3–5-year-old children, 186 of whom were of African or admixed genetic ancestry, prenatal maternal smoking was associated with DNA methylation in peripheral blood samples, but ancestry-stratified analyses were not performed [20]. Similarly, among 89 middle-aged women, of whom 28 reported African American or Hispanic ethnicity, prenatal maternal cigarette smoking was associated with DNA methylation at 17 of 190 tested CpG sites [9]. Adjusting for ancestry did not substantially affect the association between prenatal maternal smoking and DNA methylation at any of the sites, although the authors note that African American women had on average 2–5% higher mean DNA methylation on the absolute scale as compared with White and Hispanic women at 5 *CYP1A1* CpG sites [9]. Of 148 CpG sites selected based on their association with prenatal maternal smoking among primarily European children, 7 CpG sites were also associated with prenatal maternal cigarette smoking in a cohort

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

of 572 Latino children [6, 21]. While prenatal maternal cigarette smoking is associated with DNA methylation, the consistency of this association across genetic ancestry is understudied.

After association testing, an important next step is to evaluate DNA methylation as a biomarker of prenatal maternal smoke exposure by comparing DNA methylation-based classification of maternal smoking behavior to self-reported behavior. Biomarkers can consist of single DNA methylation sites or summary measures across multiple sites [17]. One option is a polymethylation score, analogous to a polygenic score. In the polymethylation score, CpG sites are weighted by the strength of their association with maternal smoking from a previous, independent sample and then summed to a single score [24, 25]. One such DNA methylation score calculated from cord blood classified maternal smoking behavior with an area under the curve (AUC) of 0.82 in a primarily European cohort [23]. Among 572 3–5-year-old children, 186 of whom were of African or admixed ancestry, prenatal maternal smoking behavior was classified with an AUC of 0.87 [20]. Among middle-aged adults, a score calculated using DNA methylation in peripheral blood samples from a single time point could even predict prenatal smoke exposure (30 years previously) with an AUC of 0.72 (95% confidence interval 0.69, 0.76) [10]. Ideally, a biomarker for prenatal maternal smoking would be consistent over the life course and portable across genetic ancestry groups. However, the consistency of DNA methylation as a biomarker for prenatal maternal smoking across age, tissue, and genetic ancestry has not been evaluated in the same study.

DNA methylation differences hold promise as a biomarker for prenatal maternal smoking. Before translation to clinical applications, we must understand how and if the signal varies across age, tissue sample, and genetic ancestry. In the Fragile Families and Child Wellbeing study, a diverse longitudinal birth cohort of children, we aimed to assess the potential of salivary

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

DNA methylation biomarkers for prenatal maternal cigarette smoking. We tested associations between prenatal maternal smoke exposure and saliva DNA methylation, including polymethylation scores for smoke exposure, individual *a priori* CpG sites, global DNA methylation, and epigenetic clocks. We then performed age- and ancestry-stratified analyses to test the hypothesis that DNA methylation could serve as a persistent and consistent biomarker of prenatal maternal smoking.

Results

Study sample descriptive statistics

In a sample of the Fragile Families and Child Wellbeing Study, saliva DNA methylation was measured on 1806 samples from 897 unique participants with the Illumina 450K array. Complete data on covariates of interest was available on 1499 samples from 809 participants who were included in the analysis (Figure 1). There were 690 participants with both age 9 and age 15 DNA methylation samples. Excluded samples were similar to included samples, except that included samples were slightly more likely to be from children of African genetic ancestry and less likely to be from children of Hispanic genetic ancestry (Supplemental Table 1). In the included sample, 20% percent of the mothers reported any prenatal maternal smoking, 12% reported prenatal alcohol use and 5% reported prenatal drug use. The mean income to poverty ratio of mothers at birth was 2.2. Of the children in the analytic sample, 50% were male, 60% were of African genetic ancestry, and 24% were of Hispanic genetic ancestry.

We calculated several summary measures of DNA methylation, including polymethylation scores for smoke exposure, global DNA methylation, and epigenetic clocks. Many of the methylation summary measures were correlated with each other (Supplemental Figure 1). For

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

example, as expected, the percent of estimated immune cells was perfectly inversely correlated with the percent of estimated epithelial cells (Pearson ρ=-1; *P* value<0.0001). Our polymethylation score for prenatal smoke exposure, which was calculated using coefficients from a cell-type corrected regression, was weakly correlated with estimated cell-type proportion of immune cells (Pearson ρ=0.05; *P* value=0.04).

We compared epigenetic measures between the age 9 and age 15 visits among the 690 individuals with data from both visits (Supplemental Figure 2). The correlation across ages was strongest for the polymethylation score for prenatal smoke exposure (Pearson ρ=0.9, *P* value<0.0001). Global DNA methylation (0.5, *P* value<0.0001) and epigenetic clocks were less strongly correlated across ages (Pediatric: 0.51; *P* value<0.0001, GRIM: 0.47; *P* value<0.0001) The distribution of epigenetic ages shifted higher between the age 9 and age 15 visits, as expected (Supplemental Figure 3). The pediatric clock more closely reflected chronological age than the GRIM clock. Among age 9 samples, the mean estimated age from the pediatric clock was 9 and the mean estimated age from the GRIM clock was 25. Among age 15 samples, the mean estimated age from the pediatric clock was 12 and the mean estimated age from the GRIM clock was 30. The distribution of estimated cell-type proportions also shifted between visits while the distribution of global methylation and polymethylation scores were more consistent (Supplemental Figure 3).

In bivariate analyses and multivariable models, we focused on the polymethylation score for prenatal smoke exposure and the single top CpG site from prior research, cg05575921 in the *AHRR* gene, as hypothesized biomarkers, and used global methylation and the pediatric clock as negative controls.

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Bivariate associations between prenatal maternal smoking and DNA methylation summary measures

Mothers who reported smoking during pregnancy had lower income-to-poverty ratios (1.49) than those who did not (2.38). Mothers who smoked were more likely to report prenatal alcohol use (30% vs 7%), prenatal drug use (18% vs 2%) and postnatal smoking (96% vs 26%) than mothers who did not report prenatal maternal smoking (Table 1). Children of mothers who reported smoking during pregnancy were more likely to be of European genetic ancestry (23%) than children of mothers who did not (14%, *P* value=<0.001). Children of mothers who reported smoking during pregnancy had higher prenatal maternal smoking polymethylation scores than children of mothers who did not at both the age 9 (mean-centered scores of 0.08 vs -0.04, *P* value<0.001) and age 15 (0.11 vs -0.01, *P* value<0.001) visits (Table 1, Supplemental Table 2). At age 9, children exposed to prenatal maternal smoking had lower DNA methylation at cg05575921 (76.82%) than children of non-smoking mothers (77.81%, *P* value=0.05). At age 15, children exposed to prenatal smoking had lower DNA methylation at cg05575921 (76.13%) than children of non-smoking mothers, although this difference was not significant (76.82% *P* value=0.24). Epigenetic age from the pediatric and GRIM clocks and global DNA methylation did not differ between children exposed vs unexposed to prenatal maternal smoking.

Multivariable associations between prenatal maternal smoking and DNA methylation summary measures

The association between prenatal maternal smoking and the polymethylation score for prenatal smoke exposure was also observed in multivariable models, adjusting for base model covariates of child sex, maternal income to poverty ratio at baseline, indicator variable for city of residence

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

being Detroit, Chicago or Toledo, proportion of salivary immune cells, sample plate from DNA methylation analysis, child age and the first two principal components of genetic ancestry (Figure 3; Supplemental Table 3). At age 9, prenatal maternal smoke exposure was associated with a 0.52 (95%CI: 0.36, 0.67) standard deviation higher polymethylation score for prenatal smoke exposure. At age 15, prenatal maternal smoke exposure was associated with a 0.46 (95%CI: 0.3, 0.62) standard deviation higher prenatal smoke exposure polymethylation score for prenatal smoke exposure. A consistent association was observed when stratifying by genetic ancestry. In the African genetic ancestry sample (n= 488) at age 9, prenatal maternal smoking was associated with a 0.55 (95%CI: 0.35, 0.75) standard deviation higher polymethylation score for prenatal smoke exposure. The direction of the association was similar in the European and Hispanic genetic ancestry samples, though the effect size was slightly attenuated and no longer statistically significant (Figure 3; Supplemental Table 4).

At age 9, prenatal maternal smoking was associated with 1 percent lower DNA methylation at cg055975921 (95%CI: -1.65, -0.34) after adjusting for base model covariates. Similarly, at age 15 prenatal maternal smoking was associated with 0.8 percent lower DNA methylation at cg055975921 (95%CI: (-1.58, -0.02)). Prenatal maternal smoking remained significantly associated with a decrease in cg05575921 DNA methylation in the African genetic ancestry sample at age 9 (-1.3 (95%CI: -2.15, -0.46)) although the association was not significant at age 15 (-0.97 (95%CI: -1.98, 0.04); Figure 3). In the European genetic ancestry sample, the association was consistent in direction at age 9 (-0.81 (95%CI: -2.45, 0.82)) and age 15 (-1.79 (95%CI: -3.43, -0.15)). In the Hispanic genetic ancestry sample, the association was no longer significant and not consistent in direction (Age 9: -0.05 (95%CI: -1.69, 1.59); Age 15: 0.82 (95%CI: -1.28, 2.92)) (Figure 3).

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Prenatal maternal smoking was not associated with global DNA methylation or epigenetic age (Figure 3).

Sensitivity analyses for association testing

After additionally adjusting for other prenatal exposures and postnatal smoke exposure, the association between prenatal maternal smoking and the polymethylation score was robust (Supplemental Figure 4). Similarly, after adjusting for other prenatal exposures and postnatal smoke exposure, the direction of the association between prenatal maternal smoking and cg055975921 was consistent, although no longer significant at age 15. Very few children who were exposed to maternal smoking prenatally were unexposed to postnatal smoke (Supplemental Figure 5). However, children exposed only to postnatal smoke did not have higher polymethylation scores than children unexposed to both postnatal and prenatal smoke (Supplemental Figure 5). Results from linear mixed effect models were similar to age-stratified and base covariate adjusted linear models (Supplemental Tables 3 & 4). Results were also similar when controlling for surrogate variables instead of known covariates (Supplemental Figures 4 and 6; Supplemental Table 5).

Results were also similar when using polymethylation scores constructed with alternative regression coefficients (see Methods, Supplemental Figure 7). For our main analysis we used coefficients from a regression of sustained smoking exposure and DNA methylation in newborn cord blood with cell-type control. As sensitivity analyses, we used coefficients from regressions of: sustained smoking exposure and DNA methylation in newborn cord blood without cell-type control, sustained smoking exposure and DNA methylation in peripheral blood from older children without cell-type control, and any smoking exposure and DNA methylation in newborn

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

cord blood without cell-type control [6]. The association between prenatal maternal smoking and the polymethylation score using regression coefficients from newborn cord blood with cell-type controls was the strongest.

In addition to cg055975921 in the *AHHR* gene, we tested the association of four other *a priori* probe sites identified from previous meta-analyses. We replicated the association of prenatal maternal smoking with these CpG sites (Supplemental Table 2, Supplemental Figure 8).

Accuracy of polymethylation scores as a biomarker of prenatal maternal smoking

Next, we compared the accuracy of different DNA methylation summary measures for classifying prenatal maternal smoking using receiver operating curves (Figure 4A; Supplemental Table 6). We estimated classification of prenatal maternal smoking when using DNA methylation summary measures alone. We also estimated prenatal maternal smoking classification when using DNA methylation summary measures in addition to base model covariates (child sex, maternal income to poverty ratio at baseline, child age at DNA methylation measurement, indicator variable for city of residence being Detroit, Chicago or Toledo, estimated immune cell proportion, plate from methylation processing and the first two genetic principal components). When used without DNA methylation measures, these base model variables had an area under the curve (AUC) of 0.73 at age 9 and 0.72 at age 15.

At age 9, including the polymethylation score for prenatal smoke exposure significantly improved the base model covariates classification (AUC: 0.78, *P* value comparing to base model<0.001). At age 15, similarly, including the polymethylation score for prenatal smoke exposure significantly improved the base model covariates classification (AUC:0.77, *P* value<0.001).

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

At age 9, the base model covariates with the polymethylation score for prenatal smoke exposure had a larger AUC than the base model covariates with cg05575921 (AUC base model covariates + cg05575921: 0.74; *P* value =0.02). At age 15, this was also true (AUC for base model covariates + cg05575921=0.73, *P* value=0.01; Supplemental Table 6). Accurate classification was not improved by including polymethylation scores from age 9 and age 15 together (Figure 4B).

Sensitivity analyses for biomarker accuracy assessment

Classification of prenatal maternal smoke exposure when using other coefficients as the weights in construction of the polymethylation scores was similar to the results when using coefficients from cell-type controlled regressions of sustained prenatal maternal smoking and DNA methylation in cord blood (Supplemental Table 7).

Discussion

In the longitudinal Fragile Families and Child Wellbeing birth cohort, we observed that prenatal maternal smoking was associated with several characterizations of DNA methylation in children's saliva samples from ages 9 and 15. Prenatal maternal smoking was associated with polymethylation scores for prenatal smoke exposure across strata of both child age and genetic ancestry. Global methylation and epigenetic clocks were not associated with maternal smoking exposure. Polymethylation scores for prenatal smoke exposure had reasonable accuracy for classifying prenatal maternal smoking (AUC > 0.7). Classification when using polymethylation scores for prenatal maternal smoke exposure was better than when using a single *a priori* CpG site, cg05575921 in the *AHRR* gene.

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Our findings are consistent with the previous literature on associations between prenatal maternal smoking and DNA methylation. We replicated the top hit from a previous epigenome wide association analysis [6]. Further, we found evidence of association in several additional hits from epigenome wide association studies of DNA methylation in cord and peripheral blood [6, 26]. There are no previous studies of prenatal maternal smoking and saliva DNA methylation. However, previous work has shown that the majority of CpG sites are similarly methylated in blood and saliva (reviewed in [27]).

We advance prenatal smoking - DNA methylation literature by evaluating the persistence of the association between prenatal maternal smoking and DNA methylation as children age and its portability across tissue and ancestry. Polymethylation scores built using coefficients from meta-analysis of cord blood DNA methylation from primarily European-ancestry newborns were still associated with prenatal maternal smoking in our independent saliva samples from a diverse cohort at ages 9 and 15 [6]. The portability of other risk scores, such as polygenic risk scores, across genetic ancestries is a complex research area [28] and the portability of epigenetic summary measures has been identified as a key area for evaluation [29]. In this case, the polymethylation score for prenatal maternal smoking appears to be portable across genetic ancestry groups. Effect estimates were consistent across ancestries at age 9, although Hispanic and European genetic ancestry samples had lower, non-significant effect estimates at age 15. The decreased precision may be the result of the reduced sample size of the Hispanic and European genetic ancestry samples. The reduction in effect estimate magnitude could reflect higher unreported smoking initiation in European and Hispanic ancestry teens in the United States than African ancestry teens. Genetic ancestry correlates with race, and White and Latino teens have much higher rates of teen smoking and earlier ages at initiation than Black children [30]. Though

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

we excluded children who reported own-smoking from our analytic sample, under-reporting of own-smoking could influence DNA methylation at age 15, creating outcome misclassification. This could also explain the observed attenuation of the effect on *AHRR*:cg05575921 methylation by age 15 in our sample, as DNA methylation at this probe is known to vary by personal cigarette smoking [22, 31].

Our findings are also consistent with previous work on DNA methylation as a biomarker for prenatal maternal smoke exposure. Prenatal maternal smoking classification using saliva DNA methylation in our sample of 9- and 15-year-olds performed comparably to classification using peripheral blood from older adults (AUC 0.72) [10]. However, previous prenatal maternal smoke exposure biomarkers created using LASSO regression and cord blood from newborns performed better (AUCs ranging from 0.82-0.97) [20] A support vector machine approach performed on peripheral blood from 3–5-year-old children also more accurately classified prenatal maternal smoke exposure (AUC=0.87) [23]. Differences in DNA methylation patterns across tissues and over time may influence the performance of DNA methylation biomarkers. Different methods for summarizing across multiple DNA methylation sites may also influence biomarker performance. The similarity in performance between saliva DNA methylation in our study and peripheral blood DNA methylation in older adults is encouraging for the use of saliva as a readily accessible clinical sample.

Our analysis contributes to the development of an accurate methylation biomarker for prenatal maternal smoking by evaluating the impact of specific methodological choices on accuracy of classification. DNA methylation biomarkers may be especially susceptible to confounding by subject age at methylation measurement and cell-type proportion [29]. Prenatal maternal smoke exposure classification accuracy of polymethylation score was similar when using coefficients

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

from cord blood vs peripheral blood samples from older children. Classification accuracy was not improved by using two time-points of methylation measurement. Classification accuracy was improved when polymethylation scores used coefficients which incorporated cell-type control. Our analysis therefore suggests that cell-type control when both generating (in the epigenome wide association study) and applying coefficients for polymethylation scores may positively influence outcome classification accuracy, although more work comparing coefficients from different populations is still needed.

Our analysis suggests that polymethylation scores may be more accurate than using single CpG sites as biomarkers. The site cg05575921 in the *AHRR* gene is a consistent marker of prenatal maternal smoke exposure in meta-analyses of newborn cord blood [6]. DNA methylation at cg05575291 in the *AHRR* gene can accurately classify own–smoking behavior in both blood (area under the curve 0.995) and saliva (area under the curve 0.971) [22, 31]. However, salivary *AHRR*:cg05575921 was not a persistent marker of prenatal maternal smoking in an analysis of middle-aged adult women [10]. In our sample, salivary *AHRR*:cg05575921 methylation was less accurate at predicting prenatal maternal smoke exposure than polymethylation scores. The accuracy of *AHRR*:cg05575921 and other single CpG site biomarkers may be influenced by time-since-exposure and new environmental exposures. Incorporating information across multiple sites of DNA methylation may yield a biomarker more robust to these influences.

Our analysis is not without its limitations. We used maternal self-report of prenatal maternal smoke exposure, as serum cotinine levels were not available. Due to social desirability bias, this could result in exposure misclassification. We would expect this to bias our results towards the null. In any analysis of a prenatal exposure and postnatal outcome, there is the possibility of selection bias into the cohort due to live birth bias. Selection bias is also possible due to loss-to-

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

follow-up between birth and age 15. Additionally, while we controlled for postnatal secondhand smoke exposure and excluded children who reported any own-smoking, we cannot exclude the possibility of residual confounding in this observational cohort.

However, our analysis also has several strengths. We analyzed samples from a large cohort of diverse participants underrepresented in genetic and epigenetic research [32, 33]. While our exposure measurement of prenatal maternal smoking was self-reported, it was assessed prospectively and preceded outcome measurements. We analyzed repeated measures of DNA methylation with reproducible array measures conducted in a single batch. We tested associations between prenatal maternal smoking and multiple DNA summary measures to evaluate the specificity of the polymethylation scores. In sensitivity analyses, we adjusted for other prenatal exposures and postnatal smoke exposure to examine the specificity of the biomarker to nature and timing of exposure.

Conclusions

In a large, prospective study of diverse participants, we showed that DNA methylation in children's saliva had strong associations with and reasonable classification accuracy for prenatal maternal smoke exposure. Further, we demonstrated that polymethylation scores could be applied as a biomarker of prenatal maternal smoke exposure across genetic ancestry groups, an important consideration for the equitable biomarker development. The development and application of biomarkers for prenatal maternal smoke exposure has important implications for epidemiological research and clinical practice. Given the difficulty of measuring prenatal maternal smoke exposure, such a biomarker could allow for confounder control in research areas where such control is currently impossible. Prenatal maternal smoke exposure is prevalent and

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

has negative health consequences, thus an exposure biomarker could be used to provide support and health interventions for children.

Methods

Cohort

The Fragile Families and Child Wellbeing Study is a birth cohort of nearly 5,000 children born in 20 cities in the United States between 1998 and 2000 [34]. Participants were selected at delivery using a three-stage stratified random sample design which oversampled unmarried mothers by a ratio of 3:1 [34]. Participants were excluded on the following criteria: those with parents who planned to place the child for adoption, those where the father was deceased, those who did not speak English or Spanish well enough to be interviewed, births where the mothers or babies were too ill to complete the interview, and those where the baby died before the interview could take place. Children were followed longitudinally with assessments at ages 1, 3, 5, 9 and 15; additional follow up is ongoing. Assessments included medical record extraction, biosample collection, in-home assessments, and surveys of the mother, father, primary caregiver, teacher and child. At ages nine and fifteen a saliva sample was taken from the child [34]. A subsample of the Fragile Families cohort was selected for saliva DNA and DNA methylation processing. To be eligible for the DNA methylation assessment children had to have participated and given saliva at age 9 and 15; after which a random sample of these eligible participants was then selected. One exception to this rule is that all participants in the Study of Adolescent Neurodevelopment (SAND) were assayed even if they only provided a sample at one time-point. All analyses account for this oversample of these participants.

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Covariates and exposure measurement

Demographic and prenatal maternal substance use variables were derived from maternal self-report questionnaire data at baseline (child's birth).

Maternal covariates included maternal income to poverty ratio at baseline, prenatal smoking, prenatal maternal drug and alcohol use, and postnatal maternal or primary caregiver smoking. Maternal income to poverty ratio is a constructed variable of the ratio of total household income (as self-reported by the mother) to the official poverty thresholds designated by the United States Census Bureau for the year preceding the interview. At baseline, mother's answered categorical questions about their prenatal smoking, drug, and alcohol use. For maternal prenatal smoking, mothers were asked

> During your pregnancy, how many cigarettes did you smoke? Did you smoke…:
>
> 2 or more packs a day
>
> 1 or more but less than 2 packs per day
>
> Less than 1 pack a day
>
> None

Few participants reported smoking a pack or more a day, thus we dichotomized to any vs no prenatal maternal smoking. For maternal prenatal drug and alcohol use the mothers were asked

> During your pregnancy how often did you use drugs/drink alcohol (respectively):
>
> Never
>
> Less than 1 time per month
>
> Several times per month

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Several times per week

Every day

When the child was 1, 5, 9 and 15 years of age, primary caregivers responded to questions about maternal and in-home smoking. To encapsulate general early childhood smoke exposure, we created a binary variable for postnatal exposure at ages 1 or 5. To encapsulate recent postnatal smoke exposure, we used a categorical variable for packs per day (no smoking, less than one pack/day, one or more packs/day) in the month prior to the age 9 and 15 interview. To control for the oversampling of the Detroit, Toledo and Chicago locations, an indicator variable denoting residence in one of these three cities was created.

Child covariates included child sex, child report of personal cigarette smoking, and child genetic ancestry. Mothers reported sex (male/female) of their child at baseline. At ages 9 children were asked if they had ever smoked a cigarette or used tobacco (yes/no) and at age 15 they were asked if they had ever smoked an entire cigarette (yes/no).

Child genetic ancestry was calculated. Grants R01 HD36916, R01 HD073352, and R01 HD076592 provided support for the collection, assay, and analysis of the genetic data in the Fragile Families cohort. Principal components (PC) of child genetic ancestry were calculated from genetic data using Illumina PsychChip_v1-1 with child saliva samples. Genetic ancestry was assigned by comparing PC loadings to 1000 Genomes super population clusters. Samples with $PC1 > 0.018$ and $PC2 > -0.0075$ were assigned to European ancestry. Samples with $PC1 < -0.005$ and $PC2 > 0.007+0.75(PC1)$ were assigned to African ancestry. Samples with $PC1 > 0.018$ and $-0.055 < PC2 < 0.025$ were assigned to Hispanic ancestry [35].

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

DNA methylation measurement

Salivary samples from the children were collected at ages nine and fifteen using the Oragene•DNA sample collection kit (DNA Genotek Inc., Ontario). Saliva DNA was extracted manually using DNA Genotek's purification protocol using prepIT L2P. DNA was bisulfite treated and cleaned using the EZ DNA Methylation kit (Zymo Research, California). Samples were randomized and plated across slides by demographic characteristics. Saliva DNA methylation was measured using the Illumina HumanMethylation 450k BeadArray [36] and imaged using the Illumina iScan system. All samples were run in a single batch to minimize technical variability.

DNA methylation image data were processed in R statistical software (3.5) using the minfi package [37]. The red and green image pairs (n=1811) were read into R and the minfi *preprocessNoob* function was used to normalize dye bias and apply background correction. Further quality control was applied using the ewastools packages [38]. We dropped samples with >10% of sites have detection p-value >0.01 (n=43), sex discordance between DNA methylation predicted sex and recorded sex (n=20), and abnormal sex chromosome intensity (n=3). CpG sites 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) (Figure 1) [39]. We used the November 2021 data freeze of the Fragile Families and Child Wellbeing DNA methylation data. Relative proportions of immune and epithelial cell types were estimated from DNA methylation measures using a childhood saliva reference panel [40].

We created polymethylation scores for prenatal maternal smoke exposure. From an independent meta-analysis of prenatal smoke exposure and newborn DNA methylation, we extracted the regression coefficients of 6,074 CpG sites associated with prenatal maternal smoking at a false

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

discovery ratio corrected *P* value <0.05 [6]. We mean-centered the DNA methylation beta values in our study, weighted them by the independent regression coefficients and took the sum. We calculated polymethylation scores using regression coefficients from 4 different regressions. For our main analysis we used coefficients from a regression of sustained smoking exposure and DNA methylation in newborn cord blood with cell-type control. As sensitivity analyses, we used sustained smoking exposure and DNA methylation in newborn cord blood without cell-type control, sustained smoking exposure and DNA methylation in peripheral blood from older children, any smoking exposure and DNA methylation in newborn cord blood) [6].

Global DNA methylation was calculated for each sample as the mean methylation value of each sample across the cleaned probe set. Mean DNA methylation restricted to probes in genomic regions (CpG island, shore, shelf or open sea, as identified in the R package IlluminaHumanMethylation450kanno.ilmn12.hg19 v 0.6.0) was also calculated.

Pediatric epigenetic age was calculated for each sample using the coefficients and methods provided by the creators (see <https://github.com/kobor-lab/Public-Scripts>) [41]. The GRIM age clock, including the smoking pack-years sub-scale, was calculated as previously described [42].

Single *a priori* CpG sites were selected from previous large meta-analyses of prenatal maternal smoking and DNA methylation in children's cord and peripheral blood samples [6, 26].

Surrogate variables were calculated from DNA methylation data using the function sva from the R package sva version 3.38.

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Statistical analyses

Inclusion exclusion criteria

From the 1685 samples from 882 unique individuals with quality-controlled DNA methylation data, we further excluded any samples missing data on: maternal prenatal smoking (0 samples), alcohol (4 samples) or other drug use (6 samples), maternal income to poverty ratio (0 samples), maternal postnatal smoking data (105 samples). We also excluded samples with missing child sex (0 samples), child age (0 samples) or genetic data (22 samples). Finally, if a child reported ever smoking a cigarette or using tobacco age 9, we excluded all of their available samples. Age 15 samples from children who reported ever smoking a cigarette at age 15 were also excluded. Children who were missing a response to the question at age 9 but answered that they had never smoked a whole cigarette at age 15 were kept in the sample.

Base model

In our base model we adjusted for maternal income to poverty ratio at baseline, child sex, child age, an indicator variable for residence in Detroit, Chicago or Toledo, plate from DNA methylation processing, and estimated immune cell proportion estimated from DNA methylation. In nonstratified models, we adjusted for the first two components of genetic ancestry from principal component analysis. In ancestry-stratified models, the first two principal components from principal component analysis run within each ancestry strata were used. While child sex, child age, plate from DNA methylation processing and immune cell proportions are not confounders (as they cannot casually affect prenatal maternal smoke exposure), these variables can strongly affect DNA methylation and so were adjusted for as precision variables.

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Receiver operator curve analysis

To evaluate the accuracy of the DNA methylation summary measures as biomarkers of prenatal maternal smoking, we used a receiver operating curve. First, we regressed exposure to prenatal maternal smoking (outcome) against the DNA methylation summary measures in individual logistic regressions, while adjusting for the base model variables listed above. Next, we calculated receiver operating curves (ROC) and area under the curves (AUCs) using the function roc from the R library pROC version 1.18.0. We compared ROC curves and AUC using the Delong method and the function roc.test from the R library pROC version 1.18.0.

Sensitivity analyses

In sensitivity analyses we performed additional adjustments: 1) models adjusting for prenatal drug and alcohol use, 2) models adjusting for prenatal drug and alcohol use and postnatal maternal/primary caregiver smoking, 3) models adjusting for surrogate variables calculated from DNA methylation data.

In addition to the cross-sectional models within visit-strata, we performed a linear mixed effect model with a random intercept for child.

Code to perform all analyses is available (www.github.com/bakulskilab)

Declarations

Ethics approval and consent to participate

Participants provided written informed consent for the study. The data used in this manuscript were prepared by the Fragile Families and Childhood Wellbeing Study administrators following

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

approval of the manuscript proposal. These secondary data analyses were approved by the University of Michigan Institutional Review Board (IRB, HUM00129826)

Consent for publication

Not applicable

Availability of data and materials

Many of the variables used in this analysis are publicly available in the Fragile Families dataset. Some of the data used in this analysis, including genetic and epigenetic data, is restricted use but available to researchers upon reasonable application.

Competing interests

The authors have no competing interests to declare.

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Funding

This research was made possible through several grants (Fragile Families Core Data Collection: R01 HD036916, genetic data processing R01 HD036916, R01 HD073352, R01 HD076592 and R01 MD011716). In addition, FB was supported by the National Institutes of Health National Institute of Dental and Craniofacial Research (F31 DE029992), EB was supported by R01 AG055406 and R01 AG067592, and CM was supported by R01 AG071071. KB and JD were supported by R01 AG067592 (MPI: Bakulski, Ware), R01 ES025531 (PI: Fallin), R01 ES025574 (PI: Schmidt), and R01 MD013299 (PI: Hicken).

Authors' contributions

FB performed the analysis with contributions from JD, JF, and EW. FB and KB wrote the paper with all authors contributing to revisions. CM, DN and LS contributed to the design and processing of the DNA methylation data subcohort. DN, CM, KM, EW, and FB all contributed to the design and conception of the analytic plan. All authors reviewed and revised the manuscript.

Acknowledgements

Not applicable

Table 1: Bivariate associations between prenatal maternal smoke exposure and selected DNA methylation summary measures and important covariates among a diverse sample of 809 children in the Fragile Families and Child Wellbeing study

Characteristic	Age 9	Age 15

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

	N	**Not exposed**, N = 598[1]	**Smoke exposed**, N = 155[1]	**p-value**[2]	N	**Not exposed**, N = 603[1]	**Smoke exposed**, N = 143[1]	**p-valu**
Child characteristics								
Polymethylation score for smoke exposure*	753	-0.04 (0.22)	0.08 (0.22)	<0.001	746	-0.01 (0.22)	0.11 (0.23)	<0.00
AHRR gene: percent cg05575921 methylation	753	77.8 (5.2)	76.8 (5.6)	0.047	746	77 (6)	76 (6)	0.2
Pediatric epigenetic clock (years)	753	8.98 (0.75)	8.94 (0.83)	0.6	746	11.62 (1.75)	11.59 (1.70)	0.9
Percent global DNA methylation	753	47.51 (0.74)	47.59 (0.78)	0.3	746	47.48 (0.85)	47.58 (0.83)	0.2
Immune cell proportion (saliva)	753	0.96 (0.10)	0.96 (0.08)	0.4	746	0.92 (0.15)	0.92 (0.14)	>0.9
Epithelial cell proportion (saliva)	753	0.04 (0.10)	0.04 (0.08)	0.4	746	0.08 (0.15)	0.08 (0.14)	>0.9
Ancestry categorization from child principal components of genetic data	753			0.001	746			0.001
African ancestry		355 (59%)	97 (63%)			364 (60%)	88 (62%)	
European ancestry		82 (14%)	35 (23%)			84 (14%)	34 (24%)	
Hispanic ancestry		161 (27%)	23 (15%)			155 (26%)	21 (15%)	
Child gender	753			0.8	746			0.9

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

	Age 9				Age 15			
Characteristic	**N**	**Not exposed**, N = 598[1]	**Smoke exposed**, N = 155[1]	**p-value**[2]	**N**	**Not exposed**, N = 603[1]	**Smoke exposed**, N = 143[1]	**p-valu**
Boy		300 (50%)	76 (49%)			295 (49%)	69 (48%)	
Girl		298 (50%)	79 (51%)			308 (51%)	74 (52%)	
Maternal characteristics								
Maternal prenatal alcohol use	753	44 (7.4%)	47 (30%)	<0.001	746	40 (6.6%)	42 (29%)	<0.00
Maternal prenatal any drug use	753	10 (1.7%)	26 (17%)	<0.001	746	9 (1.5%)	28 (20%)	<0.00
Maternal income to poverty threshold (at baseline)	753	2.41 (2.57)	1.52 (1.56)	<0.001	746	2.41 (2.58)	1.49 (1.57)	<0.00
Maternal city of residence (at baseline)	753			0.12	746			0.6
Detroit, Chicago, or Toledo		148 (25%)	48 (31%)			171 (28%)	44 (31%)	
Not Detroit, Chicago, or Toledo		450 (75%)	107 (69%)			432 (72%)	99 (69%)	
Postnatal maternal smoking at ages 1 or 5	753			<0.001	746			<0.00
No maternal smoking at age 1 and 5		447 (75%)	7 (4.5%)			448 (74%)	7 (4.9%)	
Maternal smoking at age 1 or age 5		151 (25%)	148 (95%)			155 (26%)	136 (95%)	

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

	Age 9				Age 15			
Characteristic	**N**	**Not exposed**, N = 598[1]	**Smoke exposed**, N = 155[1]	**p-value**[2]	**N**	**Not exposed**, N = 603[1]	**Smoke exposed**, N = 143[1]	**p-valu**
Postnatal primary caregiver smoking in past month prior to visit	753			<0.001	746			<0.00
No smoking		476 (80%)	23 (15%)			511 (85%)	44 (31%)	
Less than pack a day		95 (16%)	85 (55%)			80 (13%)	77 (54%)	
Pack or more a day		27 (4.5%)	47 (30%)			12 (2.0%)	22 (15%)	

Mean (SD); n (%)

Welch Two Sample t-test; Pearson's Chi-squared test

Poylmethylation score constructed using regression coefficients from a model of prenatal maternal smoking and DNA methylation in newborn cord blood as weights. DNA methylat eta values from Fragile Families and Child Wellbeing study were mean-centered, weighted by the regression coefficients and summed.

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 1 - Selection of samples from the Fragile Families and Child Wellbeing study into analytic subset. *N represents the number of individuals at each step in the selection procedure, M represents the number of samples. Individuals with repeated measures can have more than one sample.*

Figure 2 - Differences in selected DNA methylation summary measures by self-report of prenatal maternal smoking among 809 children in the Fragile Families and Child Wellbeing study at ages 9 and 15. *Samples from children exposed to prenatal maternal smoke in grey, samples from children unexposed to prenatal maternal smoke in black. From top-left, clockwise: Polymethylation scores for prenatal maternal smoke exposure, constructed using regression coefficients for prenatal smoke exposure predicting DNA methylation in newborn cord blood samples, accounting for cell-type control. DNA methylation values from samples in the Fragile Families and Child Wellbeing study were mean-centered, then multiplied by these regression coefficients and summed. Pediatric epigenetic clock (years). AHRR gene: percent cg05575921 methylation. Percent global DNA methylation. **** P value<0.0001 *P value<0.05, ns P value>0.05*

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 3 - Prenatal maternal smoke exposure is consistently associated with polymethylation scores at ages 9 and 15 and is portable across genetic ancestry groups in a sample of 809 children in the Fragile Families and Child Wellbeing study *All models shown controlled for: first two principal components of child genetic ancestry (from ancestry-stratified principal components for ancestry stratified models), child sex, child age (months), maternal income-to-poverty ratio at birth, indicator variable for residence in Detroit, Chicago, or Toledo, immune cell proportion estimated from methylation data, yes/no other maternal prenatal drug use, yes/no maternal prenatal alcohol use, postnatal maternal smoking when child 1 or 5 years of age, postnatal maternal/primary care give smoking packs/day in month prior to saliva sample. Poylmethylation score constructed using regression coefficients from a model of prenatal maternal smoking and DNA methylation in newborn cord blood as weights. DNA methylation beta values from Fragile Families and Child Wellbeing study were mean-centered, weighted by the regression coefficients and summed. The resulting scores were z-score standardized and used as outcomes in these models. Red dotted line denotes the null.*

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 4 - Polymethylation scores accurately classify prenatal maternal smoke exposure at ages 9 and 15 among 809 children in the Fragile Families and Child Wellbeing study *A) Receiver operator curve for select DNA methylation measures for predicting prenatal smoke exposure using no other variables (light colors) or using base model variables (dark colors, other variables included: child sex, child age, maternal income-poverty ratio at birth, indicator variable for residence in Detroit, Chicago, or Toledo, immune cell proportion and batch of methylation data processing). B) Receiver operator curve for including polymethylation scores individually at each visit (black & light grey) or jointly at both visits (dark grey). Poylmethylation score constructed using regression coefficients from a model of prenatal maternal smoking and DNA methylation in newborn cord blood as weights. DNA methylation beta values from Fragile Families and Child Wellbeing study were mean-centered, weighted by the regression coefficients and summed. The resulting scores were z-score standardized*

References

1. Banderali G, Martelli A, Landi M, Moretti F, Betti F, Radaelli G, Lassandro C, Verduci E (2015) Short and long term health effects of parental tobacco smoking during pregnancy and lactation: A descriptive review. J Transl Med 13:327

2. (2019) National survey on drug use and health 2018.

3. Kvalvik LG, Nilsen RM, Skjærven R, Vollset SE, Midttun O, Ueland PM, Haug K (2012) Self-reported smoking status and plasma cotinine concentrations among pregnant women in the norwegian mother and child cohort study. Pediatr Res 72:101–7

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

4. Dietz PM, Homa D, England LJ, Burley K, Tong VT, Dube SR, Bernert JT (2011) Estimates of nondisclosure of cigarette smoking among pregnant and nonpregnant women of reproductive age in the united states. Am J Epidemiol 173:355–9

5. Dempsey D, Jacob P 3rd, Benowitz NL (2002) Accelerated metabolism of nicotine and cotinine in pregnant smokers. J Pharmacol Exp Ther 301:594–8

6. Joubert BR, Felix JF, Yousefi P, et al (2016) DNA methylation in newborns and maternal smoking in pregnancy: Genome-wide consortium meta-analysis. Am J Hum Genet 98:680–696

7. Everson TM, Vives-Usano M, Seyve E, et al (2021) Placental DNA methylation signatures of maternal smoking during pregnancy and potential impacts on fetal growth. Nat Commun 12:5095

8. Cardenas A, Lutz SM, Everson TM, Perron P, Bouchard L, Hivert M-F (2019) Mediation by Placental DNA Methylation of the Association of Prenatal Maternal Smoking and Birth Weight. American Journal of Epidemiology 188:1878–1886

9. Tehranifar P, Wu H-C, McDonald JA, Jasmine F, Santella RM, Gurvich I, Flom JD, Terry MB (2018) Maternal cigarette smoking during pregnancy and offspring DNA methylation in midlife. Epigenetics 13:129–134

10. 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

11. Galanter JM, Gignoux CR, Oh SS, et al (2017) Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. eLife. https://doi.org/10.7554/elife.20532

12. Fraser HB, Lam LL, Neumann SM, Kobor MS (2012) Population-specificity of human DNA methylation. Genome Biology 13:R8

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

13. Moen EL, Zhang X, Mu W, Delaney SM, Wing C, McQuade J, Myers J, Godley LA, Dolan ME, Zhang W (2013) Genome-Wide Variation of Cytosine Modifications Between European and African Populations and the Implications for Complex Traits. Genetics 194:987–996

14. Rahmani E, Shenhav L, Schweiger R, et al (2017) Genome-wide methylation data mirror ancestry information. Epigenetics & Chromatin. https://doi.org/10.1186/s13072-016-0108-y

15. Barcelona V, Huang Y, Brown K, Liu J, Zhao W, Yu M, Kardia SLR, Smith JA, Taylor JY, Sun YV (2019) Novel DNA methylation sites associated with cigarette smoking among African Americans. Epigenetics 14:383–391

16. Sun YV, Smith AK, Conneely KN, et al (2013) Epigenomic association analysis identifies smoking-related DNA methylation sites in African Americans. Human Genetics 132:1027–1037

17. Philibert RA, Beach SRH, Lei M-K, Brody GH (2013) Changes in DNA methylation at the aryl hydrocarbon receptor repressor may be a new biomarker for smoking. Clinical Epigenetics. https://doi.org/10.1186/1868-7083-5-19

18. Philibert RA, Beach SRH, Brody GH (2012) Demethylation of the aryl hydrocarbon receptor repressor as a biomarker for nascent smokers. Epigenetics 7:1331–1338

19. Xu R, Hong X, Zhang B, Huang W, Hou W, Wang G, Wang X, Igusa T, Liang L, Ji H (2021) DNA methylation mediates the effect of maternal smoking on offspring birthweight: a birth cohort study of multi-ethnic US mothernewborn pairs. Clinical Epigenetics. https://doi.org/10.1186/s13148-021-01032-6

20. Ladd-Acosta C, Shu C, Lee BK, et al (2016) Presence of an epigenetic signature of prenatal cigarette smoke exposure in childhood. Environ Res 144:139–148

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

21. Neophytou AM, Oh SS, Hu D, Huntsman S, Eng C, Rodriguez-Santana JR, Kumar R, Balmes JR, Eisen EA, Burchard EG (2019) In utero tobacco smoke exposure, DNA methylation, and asthma in latino children. Environmental Epidemiology 3:

22. Philibert R, Dogan M, Beach SRH, Mills JA, Long JD (2019) AHRR methylation predicts smoking status and smoking intensity in both saliva and blood DNA. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics 183:51–60

23. Reese SE, Zhao S, Wu MC, et al (2017) DNA methylation score as a biomarker in newborns for sustained maternal smoking during pregnancy. Environ Health Perspect 125:760–766

24. Chen J, Zang Z, Braun U, et al (2020) Association of a Reproducible Epigenetic Risk Profile for Schizophrenia With Brain Methylation and Function. JAMA Psychiatry 77:628

25. Bakulski KM, Fisher JD, Dou JF, Gard A, Schneper L, Notterman DA, Ware EB, Mitchell C (2021) Prenatal Particulate Matter Exposure Is Associated with Saliva DNA Methylation at Age 15: Applying Cumulative DNA Methylation Scores as an Exposure Biomarker. Toxics 9:262

26. Wiklund P, Karhunen V, Richmond RC, et al (2019) DNA methylation links prenatal smoking exposure to later life health outcomes in offspring. Clinical Epigenetics. https://doi.org/10.1186/s13148-019-0683-4

27. Langie SAS, Moisse M, Declerck K, Koppen G, Godderis L, Vanden Berghe W, Drury S, De Boever P (2017) Salivary DNA Methylation Profiling: Aspects to Consider for Biomarker Identification. Basic & Clinical Pharmacology & Toxicology 121:93–101

28. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ (2019) Clinical use of current polygenic risk scores may exacerbate health disparities. Nature Genetics 51:584–591

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

29. Hüls A, Czamara D (2019) Methodological challenges in constructing DNA methylation risk scores. Epigenetics 15:1–11

30. El-Toukhy S, Sabado M, Choi K (2016) Trends in Susceptibility to Smoking by Race and Ethnicity. Pediatrics 138:e20161254

31. Dawes K, Andersen A, Vercande K, Papworth E, Philibert W, Beach SRH, Gibbons FX, Gerrard M, Philibert R (2019) Saliva DNA Methylation Detects Nascent Smoking in Adolescents. Journal of Child and Adolescent Psychopharmacology 29:535–544

32. Sirugo G, Williams SM, Tishkoff SA (2019) The Missing Diversity in Human Genetic Studies. Cell 177:26–31

33. Bentley AR, Callier S, Rotimi CN (2017) Diversity and inclusion in genomic research: why the uneven progress? Journal of Community Genetics 8:255–266

34. Reichman NE, Teitler JO, Garfinkel I, McLanahan SS (2001) Fragile families: Sample and design. Children and Youth Services Review 23:303–326

35. Ware EB, Fisher J, Schneper L, Notterman D, Mitchell C (2021) FFCWS polygenic scores - release 1. Survey Research Center, Institute for Social Research, University of Michigan; Department of Molecular Biology, Princeton University

36. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, Esteller M (2011) Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics 6:692–702

37. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA (2014) Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30:1363–1369

medRxiv preprint doi: https://doi.org/10.1101/2021.11.30.21267020; this version posted December 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

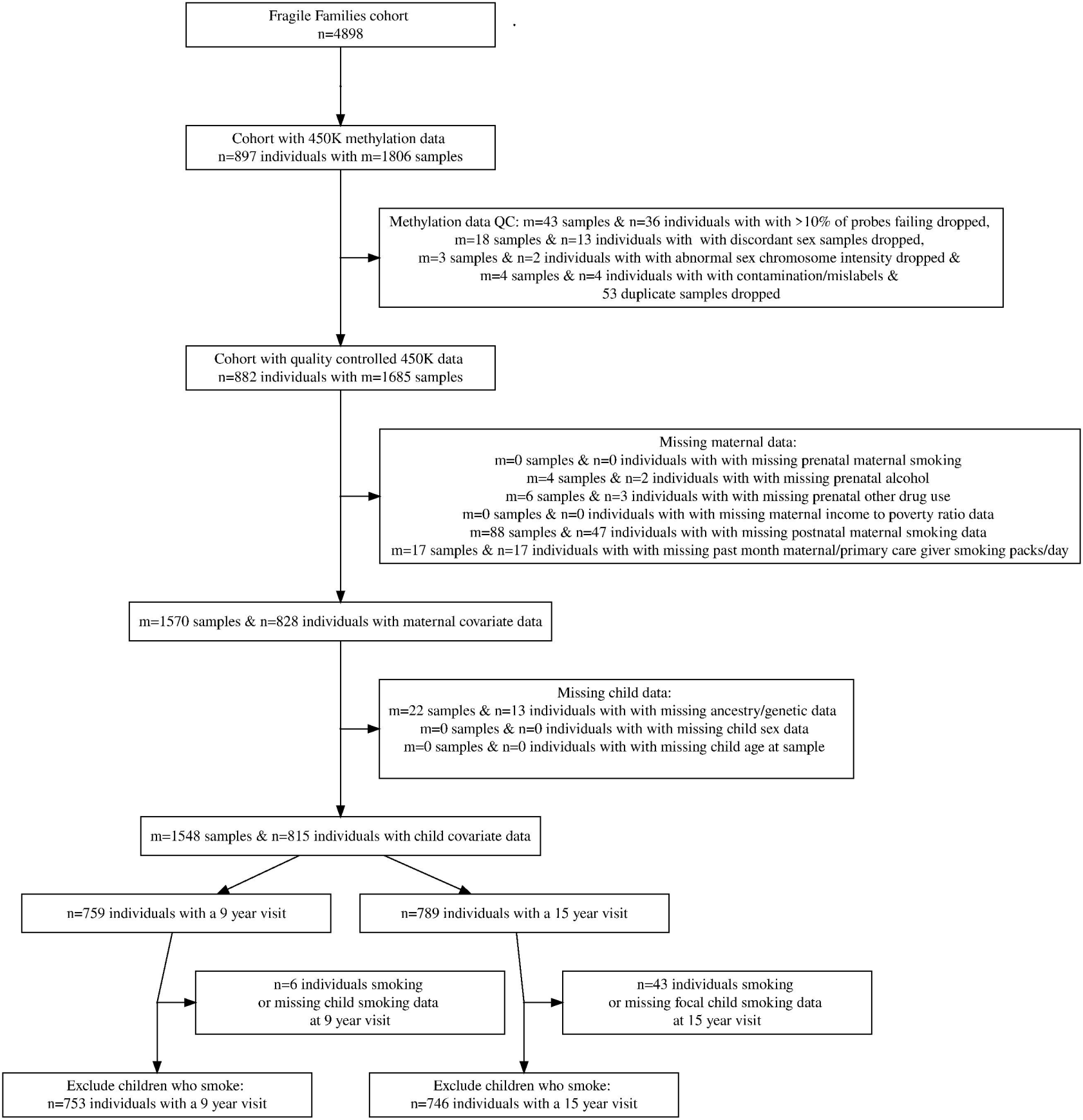
38. Heiss JA, Just AC (2018) Identifying mislabeled and contaminated DNA methylation microarray data: an extended quality control toolset with examples from GEO. Clinical Epigenetics. https://doi.org/10.1186/s13148-018-0504-1

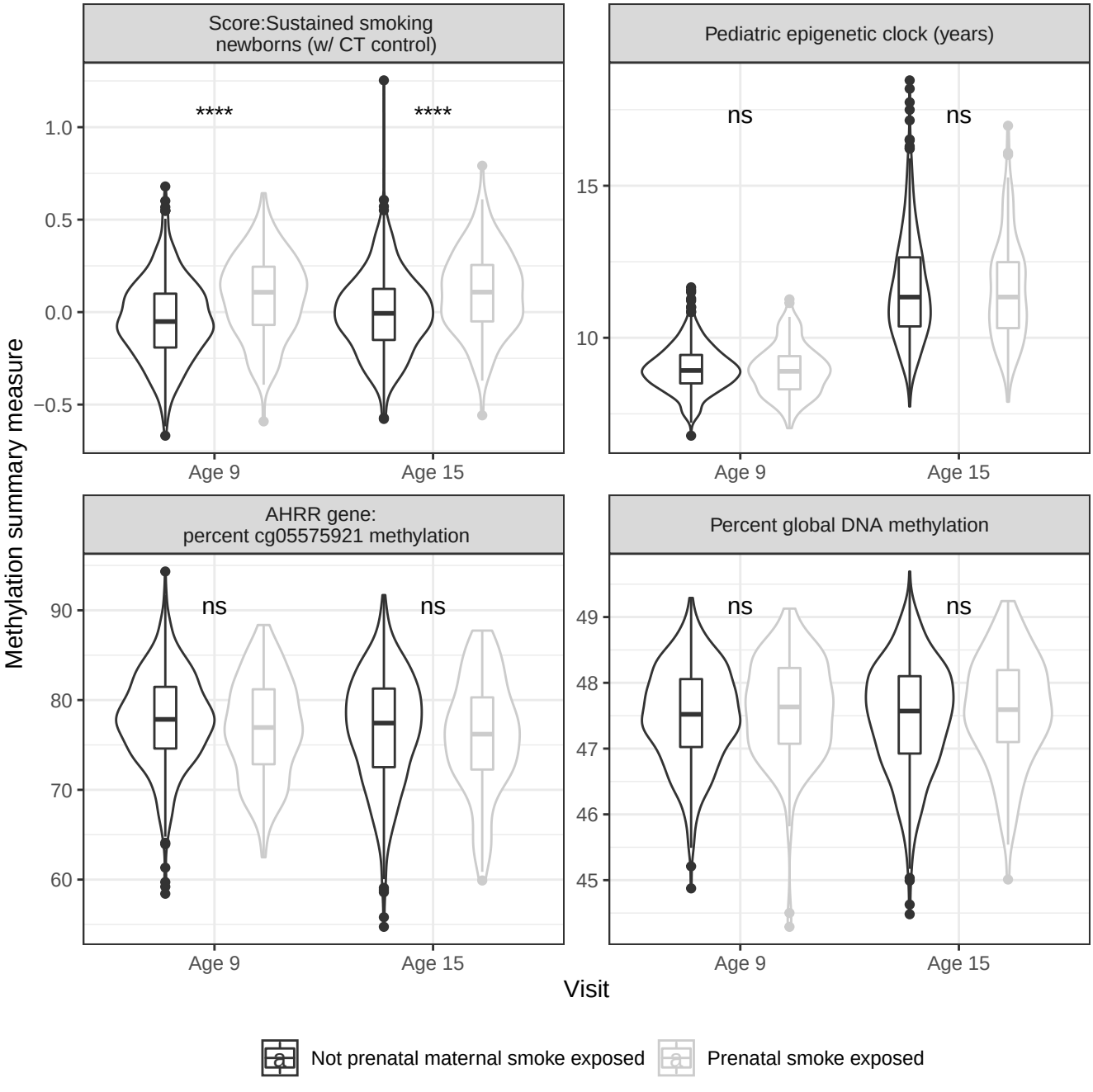
39. Chen Y, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R (2013) Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 8:203–209

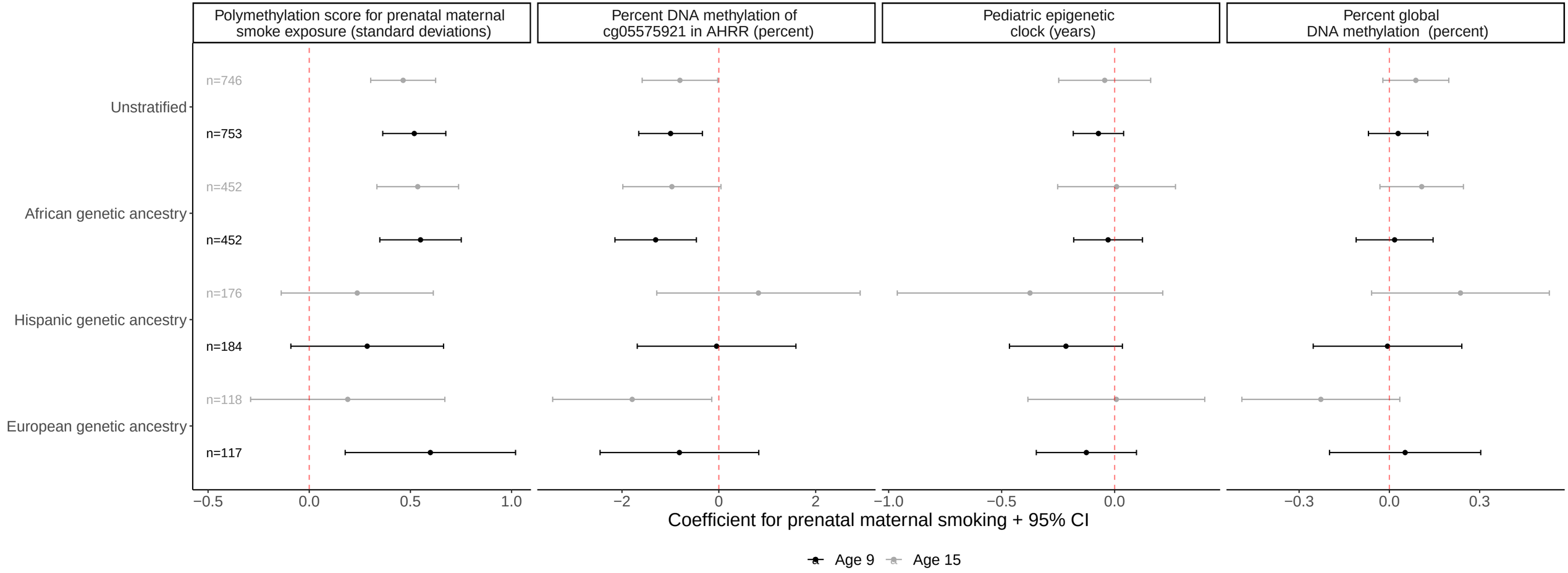
40. Middleton LYM, Dou J, Fisher J, et al (2021) Saliva cell type DNA methylation reference panel for epidemiological studies in children. Epigenetics 1–17

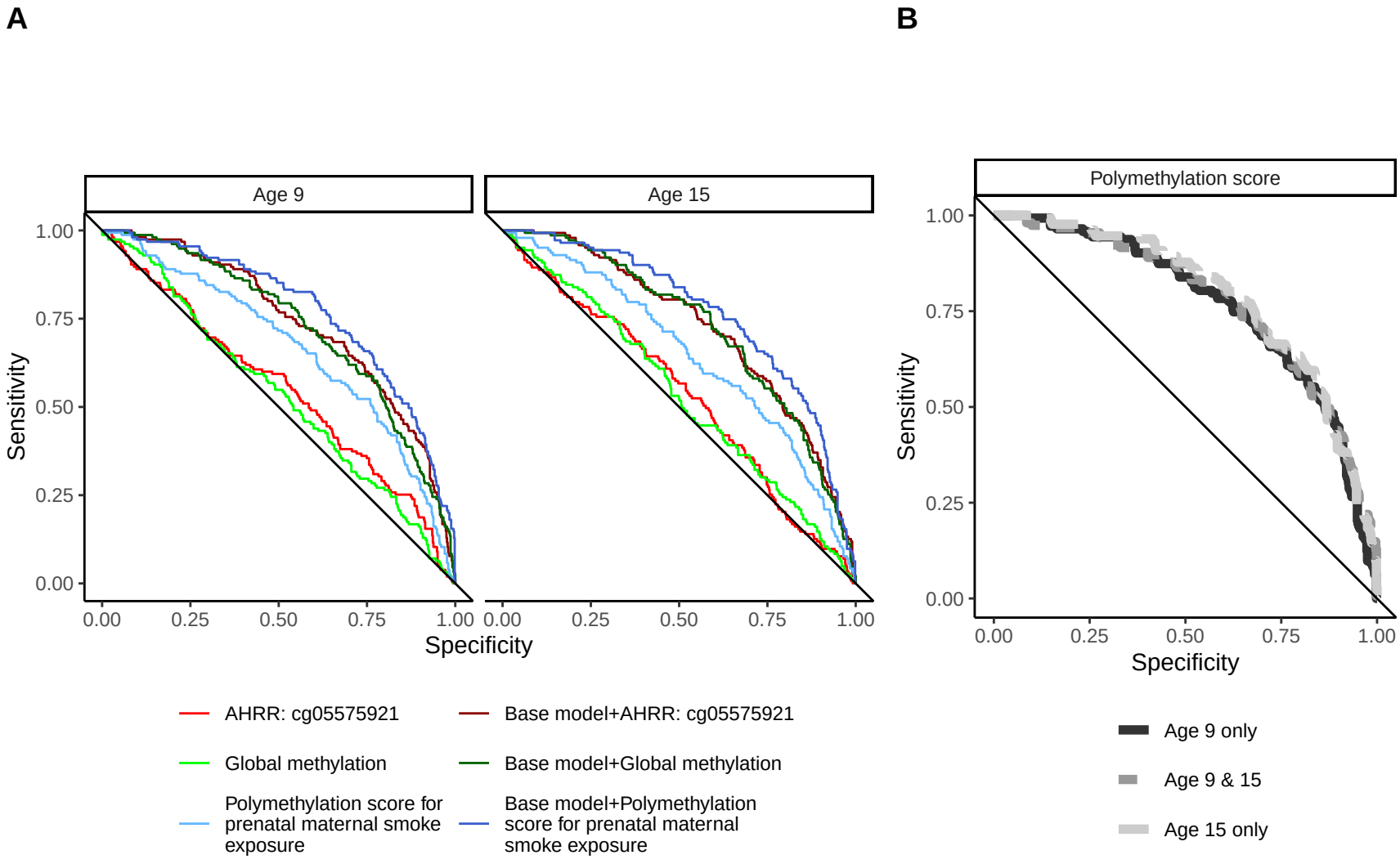
41. McEwen LM, O'Donnell KJ, McGill MG, et al (2019) The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. Proceedings of the National Academy of Sciences 117:23329–23335

42. Lu AT, Quach A, Wilson JG, et al (2019) DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging 11:303–327


Fragile Families cohort
n=4898
Cohort with 450K methylation data
n=897 individuals with m=1806 samples
Methylation data QC: m=43 samples & n=36 individuals with with >10% of probes failing dropped,
m=18 samples & n=13 individuals with with discordant sex samples dropped,
m=3 samples & n=2 individuals with with abnormal sex chromosome intensity dropped &
m=4 samples & n=4 individuals with with contamination/mislabels &
53 duplicate samples dropped
Cohort with quality controlled 450K data
n=882 individuals with m=1685 samples
Missing maternal data:
m=0 samples & n=0 individuals with with missing prenatal maternal smoking
m=4 samples & n=2 individuals with with missing prenatal alcohol
m=6 samples & n=3 individuals with with missing prenatal other drug use
m=0 samples & n=0 individuals with with missing maternal income to poverty ratio data
m=88 samples & n=47 individuals with with missing postnatal maternal smoking data
m=17 samples & n=17 individuals with with missing past month maternal/primary care giver smoking packs/day
m=1570 samples & n=828 individuals with maternal covariate data
Missing child data:
m=22 samples & n=13 individuals with with missing ancestry/genetic data
m=0 samples & n=0 individuals with with missing child sex data
m=0 samples & n=0 individuals with with missing child age at sample
m=1548 samples & n=815 individuals with child covariate data
n=759 individuals with a 9 year visit
n=789 individuals with a 15 year visit
n=6 individuals smoking
or missing child smoking data
at 9 year visit
n=43 individuals smoking
or missing focal child smoking data
at 15 year visit
Exclude children who smoke:
n=753 individuals with a 9 year visit
Exclude children who smoke:
n=746 individuals with a 15 year visit


Score:Sustained smoking newborns (w/ CT control)
Pediatric epigenetic clock (years)
AHRR gene: percent cg05575921 methylation
Percent global DNA methylation
****
****
ns
ns
ns
ns
ns
ns
Methylation summary measure
Visit
Age 9
Age 15
Not prenatal maternal smoke exposed
Prenatal smoke exposed


Polymethylation score for prenatal maternal smoke exposure (standard deviations)
Percent DNA methylation of cg05575921 in AHRR (percent)
Pediatric epigenetic clock (years)
Percent global DNA methylation (percent)
Unstratified
African genetic ancestry
Hispanic genetic ancestry
European genetic ancestry
n=746
n=753
n=452
n=452
n=176
n=184
n=118
n=117
−0.5
0.0
0.5
1.0
−2
0
2
−1.0
−0.5
0.0
−0.3
0.0
0.3
Coefficient for prenatal maternal smoking + 95% CI
Age 9
Age 15


A
Age 9
Age 15
Sensitivity
Specificity
1.00
0.75
0.50
0.25
0.00
AHRR: cg05575921
Base model+AHRR: cg05575921
Global methylation
Base model+Global methylation
Polymethylation score for prenatal maternal smoke exposure
Base model+Polymethylation score for prenatal maternal smoke exposure
B
Polymethylation score
Age 9 only
Age 9 & 15
Age 15 only