

Associations between Maternal Tobacco Smoke Exposure and the Cord Blood CD4⁺ DNA Methylome

Caitlin G. Howe,¹ Meng Zhou,² Xuting Wang,³ Gary S. Pittman,³ Isabel J. Thompson,³ Michelle R. Campbell,³ Theresa M. Bastain,¹ Brendan H. Grubbs,⁴ Muhammad T. Salam,^{1,5} Cathrine Hoyo,⁶ Douglas A. Bell,³ Andrew D. Smith,² and Carrie V. Breton¹

¹Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA

²Molecular and Computational Biology, University of Southern California, Los Angeles, California, USA

³Immunity, Inflammation and Disease Laboratory, Division of Intramural Research, National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health, Dept. of Health and Human Services, Research Triangle Park, North Carolina, USA

⁴Department of Obstetrics and Gynecology, Keck School of Medicine, Los Angeles, California, USA

⁵Department of Psychiatry, Kern Medical, Bakersfield, California, USA

⁶Department of Biological Sciences, North Carolina State University, Raleigh, North Carolina, USA

BACKGROUND: Maternal tobacco smoke exposure has been associated with altered DNA methylation. However, previous studies largely used methylation arrays, which cover a small fraction of CpGs, and focused on whole cord blood.

OBJECTIVES: The current study examined the impact of *in utero* exposure to maternal tobacco smoke on the cord blood CD4⁺ DNA methylome.

METHODS: The methylomes of 20 Hispanic white newborns ($n = 10$ exposed to any maternal tobacco smoke in pregnancy; $n = 10$ unexposed) from the Maternal and Child Health Study (MACHS) were profiled by whole-genome bisulfite sequencing (median coverage: $6.5 \times$). Statistical analyses were conducted using the Regression Analysis of Differential Methylation (RADMeth) program because it performs well on low-coverage data (minimizes false positives and negatives).

RESULTS: We found that 10,381 CpGs were differentially methylated by tobacco smoke exposure [neighbor-adjusted p -values that are additionally corrected for multiple testing based on the Benjamini-Hochberg method for controlling the false discovery rate (FDR) ($p_{FDR} < 0.05$). From these CpGs, RADMeth identified 557 differentially methylated regions (DMRs) that were overrepresented ($p < 0.05$) in important regulatory regions, including enhancers. Of nine DMRs that could be queried in a reduced representation bisulfite sequencing (RRBS) study of adult CD4⁺ cells ($n = 9$ smokers; $n = 10$ nonsmokers), four replicated ($p < 0.05$). Additionally, a CpG in the promoter of *SLC7A8* (percent methylation difference: -9.4% comparing exposed to unexposed) replicated ($p < 0.05$) in an EPIC (Illumina) array study of cord blood CD4⁺ cells ($n = 14$ exposed to sustained maternal tobacco smoke; $n = 16$ unexposed) and in a study of adult CD4⁺ cells across two platforms (EPIC: $n = 9$ smokers; $n = 11$ nonsmokers; 450K: $n = 59$ smokers; $n = 72$ nonsmokers).

CONCLUSIONS: Maternal tobacco smoke exposure in pregnancy is associated with cord blood CD4⁺ DNA methylation in key regulatory regions, including enhancers. While we used a method that performs well on low-coverage data, we cannot exclude the possibility that some results may be false positives. However, we identified a differentially methylated CpG in amino acid transporter *SLC7A8* that is highly reproducible, which may be sensitive to cigarette smoke in both cord blood and adult CD4⁺ cells. <https://doi.org/10.1289/EHP3398>

Introduction

Maternal smoking during pregnancy is one of the most prevalent and modifiable risk factors affecting newborn health (Curtin and Matthews 2016). In the United States, $>8\%$ of women smoke during pregnancy, with the highest rates occurring during the first trimester (Curtin and Matthews 2016). Prenatal exposure to tobacco smoke has been associated with adverse health outcomes for offspring, including overall fetal and infant mortality (Kleinman et al. 1988), preterm birth (Shah and Bracken 2000), altered fetal thyroid function (Meberg and Marstein 1986; Shields et al. 2009), obesity (Behl et al. 2013), childhood cancers (John et al. 1991), behavioral and cognitive effects (Clifford et al. 2012; He et al. 2017; Holz et al. 2014), and asthma (Gilliland et al. 2001).

As epigenetic modifications represent organismal flexibility in adapting to environmental stressors, epigenetic dysregulation

is one hypothesized mechanism by which exposure to tobacco smoke *in utero* may contribute to such a diverse array of adverse health outcomes later in life. Consistent with this, a growing body of evidence, including a meta-analysis of results from 13 cohorts (Joubert et al. 2016), suggests that maternal smoking during pregnancy alters newborn DNA methylation patterns, and some of these effects have been shown to persist throughout childhood (Bauer et al. 2016; Breton et al. 2014; Lee et al. 2015; Richmond et al. 2015), and even into adolescence (Lee et al. 2015) and adulthood (Flom et al. 2011; Richmond et al. 2015). However, most studies evaluating associations between maternal smoking during pregnancy and DNA methylation have focused on a handful of candidate genes or utilized methylation arrays, which cover a small fraction ($<5\%$) of CpG sites in the human genome, largely selected from promoter regions. A recent study, using whole-genome bisulfite sequencing (WGBS), observed that maternal smoking during pregnancy largely impacts DNA methylation levels within enhancer regions (Bauer et al. 2016), which until recently were underrepresented on methylation arrays. Thus, the most profound effects of maternal smoking during pregnancy may occur at distal regulatory regions, which have been largely unevaluated.

Another important limitation of previous studies is that most have measured DNA methylation in whole blood or total leukocyte samples, which consist of a mixture of different cell types, each with its own distinct DNA methylation profile. Findings from these studies therefore represent methylation differences averaged across multiple cell types. These studies also may have been susceptible to confounding by cell type heterogeneity. Although the importance of accounting for major cell types has

Address correspondence to Carrie V. Breton, 2001 N Soto St., Los Angeles, CA 90032, USA. Telephone: (323) 442-7383. Email: Breton@usc.edu
Supplemental Material is available online (<https://doi.org/10.1289/EHP3398>).

The authors declare they have no actual or potential competing financial interests.

Received 23 January 2018; Revised 20 February 2019; Accepted 5 March 2019; Published 30 April 2019.

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehponline@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

Table 1. Maternal and newborn characteristics of Maternal and Child Health Study (MACHS) participants by exposure group.

	WGBS subset			Larger MACHS study population
	Tobacco smoke exposed (<i>n</i> = 10) mean ± SD or <i>n</i> (%)	Unexposed (<i>n</i> = 10) mean ± SD or <i>n</i> (%)	<i>p</i> -Value ^a	All participants (<i>n</i> = 232) mean ± SD or <i>n</i> (%)
Matched characteristics				
Age, years	25 ± 6	27 ± 8	0.62	27 ± 6 ^b
Prepregnancy BMI, kg/m ²	29.6 ± 5.9 ^c	28.8 ± 5.6	1	28.7 ± 8 ^d
Prepregnancy diabetes	0 (0)	0 (0)	1	13 (5.6%) ^e
Gestational age, weeks	38 ± 2	38 ± 2	0.62	39 ± 2 ^f
Preterm	2 (20)	2 (20)	1	44 (19)
Adjustment characteristics				
Baby's sex, male	6 (60)	7 (70)	1	115 (50)
Mother worked during pregnancy	7 (70)	6 (60)	1	101 (44)

Note: BMI, body mass index; MACHS, Maternal and Child Health Study; SD, standard deviation; WGBS, whole-genome bisulfite sequencing.

^a*p*-value is from Wilcoxon rank-sum test (continuous variables), chi-square test (categorical variables with >5 counts per cell), or Fisher's exact test (categorical variables with <5 counts per cell), comparing participants in the maternal tobacco smoke exposed group with the unexposed group.

^bMaternal age information available for *n* = 228 participants.

^cOne participant from the maternal tobacco smoke exposed group was missing information on maternal prepregnancy BMI.

^dMaternal prepregnancy BMI available for *n* = 179 participants.

^eMaternal prepregnancy BMI available for *n* = 232 participants.

^fGestational age information available for *n* = 226 participants.

been emphasized by two recent studies, which observed differential effects of maternal smoking during pregnancy on DNA methylation in leukocyte subtypes (Bauer et al. 2016; Su et al. 2016), studies evaluating the influence of maternal smoking during pregnancy on DNA methylation patterns within specific cell types are still generally lacking. Additionally, most previous studies of prenatal tobacco smoke exposure and DNA methylation have focused on predominately Caucasian populations (e.g., Joubert et al. 2012; Küpers et al. 2015; Markunas et al. 2014; Richmond et al. 2015). Much less is known about the impact of *in utero* tobacco smoke exposure on newborn DNA methylation patterns in Hispanic populations.

The objective of the current study was therefore to investigate the impact of maternal smoking during pregnancy on the newborn methylome, using WGBS, within sorted cord blood CD4⁺ cells collected from a subset of Hispanic white participants enrolled in the Maternal and Child Health Study (MACHS), a Los Angeles birth cohort that was designed to identify early-life risk factors for adverse metabolic and respiratory outcomes in childhood. On average, CD4⁺ cells comprise 10–16% of white blood cells in cord blood. They are essential for host defense, as they coordinate immune responses to infections and malignancies by recruiting and activating other immune cells (Zhu and Paul 2008). They were selected for the current study because they are accessible in healthy newborns and relevant to tobacco smoke toxicity, as they have also been implicated in the pathogenesis of airway inflammation and asthma (Akbari et al. 2006; Ling and Luster 2016; Lloyd and Hessel 2010).

Methods

Maternal and Child Health Study

MACHS is a birth cohort in Los Angeles. Beginning in 2012, MACHS participants were recruited from the labor and delivery ward of the Los Angeles County and University of Southern California Medical Center, which serves a predominately low-income Hispanic population. Women were excluded from the study if they were HIV positive, currently incarcerated, or <18 years old. Women were also excluded if they had a multiple pregnancy or a physical, mental, or cognitive disability that would preclude their participation in the study. Written informed consent was obtained from all participants in accordance with the Health Sciences Institutional Review Board of the University of Southern California.

For the current study, 10 infants with exposure to maternal smoking *in utero* and 10 unexposed infants born to Hispanic white women were identified in MACHS and matched on fetal gestational age, maternal diabetes status, maternal prepregnancy body mass index (BMI), and maternal age (Table 1).

Assessing Maternal Smoking during Pregnancy

Maternal smoking during pregnancy was assessed by an interview-administered questionnaire, which was provided in either English or Spanish, depending on the primary language of the participant. Of the 10 mothers who smoked during pregnancy who were selected for this study, nine reported that they stopped smoking soon after learning that they were pregnant or at some point during the first trimester, and one reported that she had stopped smoking during the third trimester. The 10 women who did not smoke during pregnancy were lifetime nonsmokers. Although cigarette smoke exposure from other sources was not directly considered for this analysis, four of the women who reported smoking during pregnancy, and one woman who reported being a lifetime nonsmoker, shared a residence with a cigarette smoker.

Other Maternal and Newborn Characteristics

Prepregnancy BMI was calculated based on prepregnancy height and weight values obtained from maternal medical records. Gestational age and baby's sex were acquired from pediatric medical records. Maternal age, prepregnancy diabetes status, and working status during pregnancy were determined by an interview-administered questionnaire, provided in the primary language of the participants (English or Spanish) at the time of delivery.

Cord Blood Collection, Cell Sorting, and DNA Isolation

Cord blood samples were drained into collection containers by hospital providers and then transferred by standard syringe into EDTA tubes BD Biosciences (Catalog Number: BD-366643). Samples were stored at room temperature until they were transported to the molecular biology laboratory at the Southern California Environmental Health Sciences Center, where they were processed. The median time from cord blood collection to sample processing was 14 h, which did not differ by exposure group (*p*_{Wilcoxon} = 0.80).

Peripheral blood mononuclear cells (PBMCs) were isolated by treating 5–10 mL of whole blood with Human Granulocyte Depletion Cocktail (STEMCELL Technologies) for 20 min. The

blood was diluted with an equal volume of EasySep™ Buffer (STEMCELL Technologies), which was mixed and overlaid into a 50-mL SepMate™ tube (STEMCELL Technologies) filled with 15 mL Lymphoprep™ (STEMCELL Technologies). The tube was centrifuged at 1,200 × g for 10 min, and the PBMC layer was transferred into a 50-mL centrifuge tube using a transfer pipette. Cells were washed with 20 mL of EasySep™ Buffer and centrifuged for 8 min at 300 × g. The cell pellet was resuspended in 2 mL of EasySep™ Buffer, then transferred into a 14-mL polystyrene tube. CD14⁺ monocyte cells and CD4⁺ T cells were then separated using the EasySep™ Human CD14 Selection Kit (STEMCELL Technologies), followed by the EasySep™ Human CD4 Selection Kit (STEMCELL Technologies), according to manufacturer protocols. Lysed buffy coat, CD14⁺, and CD4⁺ cells were suspended in Qiagen RLT + 1/2-mercaptoethanol and stored at -20°C until DNA extraction. This kit typically yields 97.4–99.5% CD4⁺ cell content.

DNA was isolated from sorted CD4⁺ cells using the Qiagen All Prep DNA/RNA/miRNA kit (80224; Qiagen), according to the manufacturer's instructions. A260:A280 ratios were determined by nanodrop and ranged between 1.80 and 1.94.

Whole-Genome Bisulfite Sequencing

Five hundred nanograms of DNA were shipped overnight on dry ice to BGI. Library preparation consisted of: *a*) sonication of the DNA to create 100–300 base pair fragments; *b*) DNA end repair, addition of 3' adenine overhangs, and ligation of methylated sequencing adaptors; *c*) bisulfite treatment using the EZ DNA Methylation-Gold Kit™ (Zymo Research); and *d*) desalting, size selection, polymerase chain reaction (PCR) amplification (10 cycles), and an additional round of size selection. Library fragment size (250–300 base pairs) was verified by a 2100 Bioanalyzer (Agilent). Library concentration (>5.39 nM) was verified by quantitative real-time PCR. 150 base pair, paired-end WGBS was performed by BGI, using Illumina's HiSeq 4000 System. Samples were multiplexed, and three samples were run per lane. The raw and processed WGBS data that were generated and analyzed for the current study are publicly available in Gene Expression Omnibus (GEO) (accession: GSE109212) (Edgar et al. 2002; Barrett et al. 2013).

Read Alignment

Trim Galore! (version 0.4.2, Babraham Bioinformatics) was used to assess read quality and to remove adapters from raw sequencing reads and conduct quality trimming, using the default Phred score cutoff of 20. Trimmed reads were mapped to human assembly GRCh37, using the Wildcard ALignment Tool (Chen et al. 2016). Eighty-four percent of the trimmed reads were mapped, and 93% of these were uniquely mapped. The MethPipe package (version 3.4.2) was used to remove duplicate reads and to calculate average CpG methylation levels and bisulfite conversion rates (Song et al. 2013). The median percent duplication rate was 19%. The median coverage for a CpG site was 6.5 ×. Mapping rates, percent duplication rates, and coverage were similar in the exposed and unexposed groups. Median bisulfite conversion rates were 0.99 in both groups.

Statistical Analysis

Differentially methylated CpG sites were identified using Regression Analysis of Differential Methylation (RADMeth) (Dolzhenko and Smith 2014), which utilizes beta-binomial regression and is part of the MethPipe pipeline (Song et al. 2013). We selected this method because it accounts for within-group variation in methylation and is therefore more appropriate for population studies, and has been shown to perform well on

low-coverage data (Dolzhenko and Smith 2014). The code for this method is publicly available (<https://github.com/smithlabcode/methpipe>). Although there were no significant differences between the exposed and unexposed groups for any measured maternal or newborn characteristics, there was one fewer male baby and one more mother who worked during pregnancy in the tobacco smoke-exposed group (Table 1). We therefore included baby's sex and maternal working status as covariates in regression models. The RADMeth program calculates three sets of *p*-values: *a*) raw *p*-values; *b*) *p*-values adjusted for their correlations with proximal CpGs (based on a 200-base pair window), which increases the statistical power to detect differential methylation, even at low-coverage sites, as described previously (Dolzhenko and Smith 2014); and *c*) neighbor-adjusted *p*-values that are additionally corrected for multiple testing based on the Benjamini-Hochberg method for controlling the false discovery rate (FDR) (Benjamini and Hochberg 1995) (hereafter referred to as *p*_{FDR}). The genomic inflation factor (λ) was calculated from the raw *p*-values. CpG sites with a *p*_{FDR} < 0.05 were considered differentially methylated. Differentially methylated regions (DMRs) were also called using RADmeth, which joins neighboring CpG sites that were identified as differentially methylated (*p*_{FDR} < 0.05). Resulting DMRs were filtered for those containing a minimum of at least three CpG sites with both a raw and FDR-adjusted *p* < 0.05 to capture potential methylation differences at regulatory regions that are CpG poor (e.g., enhancers) in addition to those that are CpG rich (e.g., promoters) (Jones 2012). Since MethPipe does not systematically check for outliers, we visually inspected for outliers in a subset of loci by plotting the raw methylation values for individual participants separately by exposure group. Given the low coverage of our data, we attempted to run a sensitivity analysis restricting to the 1.4 million CpGs in our dataset that had an average coverage ≥ 10 ×. However, the RADMeth program was unable to run on this restricted set of CpGs, likely because the data were too sparse. We were therefore also unable to run the subsequent DMR identification step for this restricted set of data.

Annotation of genomic regions was conducted using the Goldmine package in R (version 3.5.0, R Project) (Bhasin and Ting 2016). Significant enrichment of DMRs at genomic regions of interest was determined using Fisher's exact test. The drawGenomePool() function in the Goldmine package was used to create a length-matched pool of background sequences from GRCh37 that was 100 × the size of the query and did not *a*) overlap sequences in our query; or *b*) extend off chromosome ends or over assembly gaps (Bhasin and Ting 2016). Additionally, overlap between DMRs and FANTOM5 enhancers was examined (Andersson et al. 2014). We used a three-pronged approach and evaluated overlap with *a*) all known enhancers, *b*) enhancers expressed in T cells, and *c*) enhancers expressed ubiquitously across tissue types. Additionally, we examined overlap between DMRs and CD4⁺ DNase-sensitive regions and transcription factor binding sites. Coordinates for CD4⁺ DNase-sensitive regions were determined by merging three DNase-seq datasets for adult CD4⁺ T cells, acquired from the ENCODE Experimental Data Matrix (file accession numbers: ENCFF569GSL, ENCFF235AEA, and ENCFF907KBL, generated in the lab of John Stamatoyannopoulos, University of Washington) (ENCODE Project Consortium 2012). Overlapping features were merged using version 2.26.0, BEDTools (Quinlan and Hall 2010). Coordinates for CD4⁺ transcription factor binding sites were acquired from ReMap (Griffon et al. 2014). Additionally, CTCF CD4⁺ binding site coordinates were obtained from GEO (accession number GSE12889) (Cuddapah et al. 2009) and converted to GRCh37 using the University of California Santa Cruz Genome Browser liftOver tool (Haeussler et al. 2019). Gene set enrichment analyses for PANTHER pathways (Mi et al. 2017) and

gene ontology (GO) terms (Gene Ontology Consortium 2017) were conducted for all genes directly overlapping DMRs using the enrichR software (Kuleshov et al. 2016). Predicted targets of microRNAs (miRNAs) overlapping DMRs were identified using three different software programs: TargetScanHuman7.2 (Agarwal et al. 2015), miRWalk 2.0 (Dweep et al. 2014), and version 5.0, DIANA-microT-CDS tools (Paraskevopoulou et al. 2016). The intersection of predicted target genes identified by all three tools was then evaluated in downstream analyses. Potential enrichment of predicted target genes within PANTHER pathways (Mi et al. 2017) and GO terms (Gene Ontology Consortium 2017) associated with predicted targets were determined using enrichR (Kuleshov et al. 2016).

To identify single-nucleotide polymorphisms (SNPs) that might impact CpGs of interest, coordinates were obtained for all SNPs identified by the 1000 Genomes Project. Since the MXL population, which represents Mexican American individuals from the Los Angeles area, is the most similar to MACHS, we identified SNPs that are common (minor allele frequency >0.01) in this population using version 0.1.16, VCFtools (Danecek et al. 2011), and identified overlaps with differentially methylated CpGs using BEDTools (Quinlan and Hall 2010).

Replication studies were conducted for differentially methylated CpG sites and DMRs identified in MACHS, using CD4⁺ DNA methylation data from two different study populations. The first study, WakeMed Smoking Epigenetics Study (SMKE), profiled DNA methylation using the Illumina Infinium MethylationEPIC array in cord blood CD4⁺ cells isolated with CD4⁺ Dynabeads (Invitrogen) from a group of 30 newborns ($n=14$ exposed vs. $n=16$ unexposed to sustained maternal tobacco smoke), who were recruited from the WakeMed hospital in Raleigh, North Carolina. The second study, Epigenetic Biomarkers of Tobacco Smoke Exposure, profiled DNA methylation in adult CD4⁺ cells from a subset of smokers and nonsmokers using three different methods: *a*) Illumina's EPIC array ($n=9$ smokers; $n=11$ nonsmokers), *b*) Illumina's HumanMethylation450K ($n=59$ smokers; $n=72$ nonsmokers) array, and *c*) reduced representation bisulfite sequencing (RRBS) ($n=9$ smoking; $n=10$ nonsmoking women). These participants were recruited at the National Institute of Environmental Health Sciences (NIEHS) Clinical Research Unit (CRU) (Wan et al. 2018). Methods for replication analyses and participant demographics for the replication studies are described in more detail in Supplemental Materials (see "Replication Look-Up Analyses" and Tables S1–S4).

Look-up analyses were conducted for DMRs identified in a previous study of maternal tobacco smoke exposure during pregnancy, which profiled DNA methylation in whole cord blood samples using WGBS (Bauer et al. 2016). Additionally, look-up analyses were conducted *a*) for 25 of the 26 CpGs that were identified as differentially methylated ($p_{\text{Bonferroni}} < 0.05$) by sustained maternal tobacco smoke exposure during pregnancy in whole cord blood in an Infinium 450K array study (Joubert et al. 2012), many of which have been widely replicated, that were also represented in the MACHS WGBS dataset; and *b*) for 3,258 CpGs on the 450K array that were identified as differentially methylated ($p_{\text{FDR}} < 0.05$) in relation to both sustained and any maternal smoking in meta-analyses conducted by the Pregnancy and Childhood Epigenetics (PACE) consortium (Joubert et al. 2016), which were also represented in the MACHS WGBS dataset.

Results

Characteristics of Study Participants

Characteristics of the subset of MACHS participants with WGBS data are shown separately by exposure group and are compared with the entire MACHS study population in Table 1.

For the subset of participants with WGBS data, the mean \pm standard deviation (SD) maternal age at pregnancy was 26 ± 7 y, and the mean \pm SD maternal prepregnancy BMI was 29.2 ± 5.6 . There were more male (65%) compared with female (45%) newborns in the current study, and the mean \pm SD gestational age was 38 ± 2 wk, with 20% of babies born preterm. Sixty-five percent of mothers included in the study reported that they worked during pregnancy. Measured covariates were similar between exposure groups (Table 1). The subset of participants with WGBS data were generally similar to the overall MACHS cohort, although $\sim 6\%$ of MACHS participants reported having diabetes prior to pregnancy, compared with 0% in the WGBS subset, which was by design (Table 1). There were also fewer male babies and fewer mothers who reported working during pregnancy in the overall MACHS cohort (Table 1).

Differentially Methylated CpG Sites and Regions

We identified 10,381 CpG sites that were differentially methylated in cord blood between newborns with and without exposure to maternal smoking *in utero* ($p_{\text{FDR}} < 0.05$) (Excel Table S1). None of the sites were statistically significant after applying a more conservative Bonferroni correction (number of tests = 26,674,173; Bonferroni threshold = 1.9×10^{-9} ; smallest $p = 5.7 \times 10^{-8}$). There was evidence of some genomic inflation ($\lambda = 1.25$). Of the 10,381 differentially methylated CpGs identified, 169 (1.6%) overlapped an SNP identified as common (minor allele frequency >0.01) in the 1000 Genomes MXL population (Excel Table S1). In our visual inspection of a subset of highly significant CpGs with moderate percentage methylation differences (10–11%), we did identify some outliers (Figure S1).

From the 10,381 differentially methylated CpGs, MethPipe identified 1,533 DMRs by merging neighboring CpGs that had a $p_{\text{FDR}} < 0.05$. Of these DMRs, 557 contained a minimum of three CpG sites with both a raw and FDR-adjusted $p < 0.05$ and were therefore retained for downstream analyses (Excel Table S2). Of these 557 DMRs, 249 DMRs were hypomethylated and 308 were hypermethylated in tobacco smoke exposed, compared with unexposed, newborns. The 10,381 differentially methylated CpGs and 557 DMRs are shown by chromosome in Figure S2. The median absolute methylation difference for the 557 DMRs was 12%, and the largest absolute methylation difference was 40% (Figure S3). These DMRs spanned from 8 to 568 base pairs, with a median length of 149 base pairs (Figure S2). Although the majority of DMRs contained fewer than 10 CpGs with both a raw p -value < 0.05 and $p_{\text{FDR}} < 0.05$, one DMR, located in the promoter of pseudogene *HSPA7*, contained as many as 31. These 31 CpGs comprise a subset of 57 CpGs that were found to be differentially methylated ($p_{\text{FDR}} < 0.05$) in the promoter region of *HSPA7* in individual CpG analyses. These 57 CpGs represent more than half of the 106 CpGs located in the *HSPA7* promoter that were represented in the MACHS dataset; all but one of these 57 CpGs was hypomethylated in the maternal tobacco smoke exposed, compared with unexposed, group. The 20 DMRs with the largest methylation differences are listed with their nearest genes in Table 2. The majority of these DMRs were hypermethylated and located within intergenic regions.

Genomic Context of Differentially Methylated Regions and Enrichment in Key Regulatory Regions

Compared with a random sample of similar-sized regions in the genome, DMRs were enriched at promoter, exon, and 3' end regions (Table 3). In contrast, there were significantly fewer DMRs within intergenic regions than would be expected by chance (Table 3). Results were similar when DMRs were evaluated separately by

Table 2. The twenty DMRs with the largest methylation differences.

Position	Percent methylation difference	No. of CpGs ^a	Genomic location	Nearest gene(s)	Distance from gene (base pairs)	Mean coverage per CpG in DMR
Hypermethylated						
Chr10:46775380-46775612	39.7	7	Intergenic	<i>GLUD1P7</i>	7,033	0.7
Chr17:7114926-7114953	39.2	3	Intron	<i>DLG4</i>	0	1.7
Chr2:144749232-144749290	36.6	5	Intron	<i>GTDC1</i>	0	0.8
Chr7:38308302-38308354	35.3	3	3' End	<i>TRGC2, TCRGC2, TARP, Z22690</i>	0, 0, 0, 0	4.0
Chr8:38533279-38533327	34.8	3	Intergenic	<i>TACC1</i>	52,376	5.8
Chr5:26725725-26725747	33.3	3	Intergenic	<i>CDH9</i>	154,961	0.7
Chr21:39004308-39004421	31.9	3	Intron	<i>KCNJ6</i>	0	4.9
Chr22:21663324-21663499	31.4	8	Intergenic	<i>POM121L8P</i>	11,308	1.1
Chr15:25200647-25200780	31.3	4	Promoter	<i>SNRPN, SNURF</i>	0	1.4
Chr12:58229294-58229355	30.8	3	Intron	<i>CTDSP2</i>	0	7.6
Chr10:123100131-123100324	30.5	5	Intergenic	<i>FGFR2</i>	137519	7.4
Chr16:57343206-57343353	29.9	6	Intergenic	<i>TRNA_Leu</i>	8,731	3.1
Chr17:27359874-27360049	29.0	4	Intergenic	<i>PIPOX</i>	9,868	5.0
Chr4:1607004-1607368	28.6	8	Intergenic	<i>AX748388</i>	29,037	3.2
Hypomethylated						
Chr6:171026627-171026778	-39.8	9	Intergenic	<i>BC036251</i>	20,591	1.4
Chr5:564378-564471	-38.5	5	Intergenic	<i>MIR4456</i>	28,380	0.8
Chr1:18343023-18343127	-34.4	3	Intergenic	<i>IGSF21</i>	91,112	8.9
Chr4:3295539-3295663	-30.5	3	Intron	<i>RGS12</i>	0	4.9
Chr9:36154747-36154804	-29.6	3	Intron	<i>GLIPR2</i>	0	7.6
Chr5:177209275-177209375	-29.4	4	Promoter	<i>FAM153A</i>	0	3.6

Note: Chr, chromosome; DMR, differentially methylated region.

^aNumber of CpGs in the DMR of interest with a raw and false discovery rate-adjusted *p*-value <0.05.

those that were hyper- vs. hypomethylated in the tobacco smoke exposed, compared with the unexposed, group, and median (range) percentage methylation differences were generally similar across genomic regions (Figure S4).

DMRs were also significantly enriched within FANTOM5 enhancers, CD4⁺ DNase-sensitive regions, and at CD4⁺ transcription factor binding sites, including CTCF binding sites (Table 3). Results were similar when DMRs were evaluated separately by those that were hypermethylated vs. hypomethylated in the exposed, compared with the unexposed, group (Table S5). Of the 557 DMRs identified, 11 overlapped active enhancer regions, two of which are active in T cells. Predicted targets of these enhancer regions and descriptive information for overlapping DMRs are shown in Table S6. The two DMRs overlapping T-cell enhancers were hypomethylated among the exposed, compared with unexposed, newborns. One DMR (chr8:141,108,929-141,109,987) had a methylation difference of -11.7%. The predicted target of

this enhancer was the promoter region of *KCNK9* (*r*=0.73) (Andersson et al. 2014; FANTOM Consortium 2014), which is imprinted in the brain (Ruf et al. 2007). This same DMR also overlapped a CD4⁺ DNase-sensitive region (chr8:141,109,195-141,109,405) and CTCF binding region (chr8:141,109,278-141,109,552). The second DMR (chr21:44,104,688-44,105,340) had a methylation difference of -15.0%. The enhancer overlapping this DMR had two predicted targets: the promoter region of *NDUFV3* (*r*=0.92), which codes for a mitochondrial enzyme, and also *ABCG1* (*r*=0.66) (Andersson et al. 2014; FANTOM Consortium 2014), which plays an important role in cellular lipid homeostasis (Kennedy et al. 2005). None of the DMRs overlapped ubiquitous enhancers.

Gene Set Enrichment Analyses

A total of 564 genes were annotated to the 557 DMRs (Excel Table S2). Of these genes, 385 directly overlapped a DMR (Excel Table S2). These genes were not significantly enriched in any PANTHER pathways or significantly associated with any GO biological processes after adjusting for multiple testing.

Predicted Targets of Differentially Methylated MicroRNA Genes

Of the 385 genes directly overlapping DMRs, four coded for miRNAs (*miR-29b-2*, *miR4750*, *miR1914*, and *miR646HG*). There were 1,065 predicted target genes identified for these miRNAs (Excel Table S3); these genes were associated with 15 GO biological process terms, which remained statistically significant after adjusting for multiple testing (Excel Table S4). Seven of these GO terms were related to the regulation of transcription or gene expression.

Replication Studies

Because we were unable to identify other WGBS studies of tobacco smoke exposure in CD4⁺ cells, we sought to replicate our results in populations with array-based DNA methylation

Table 3. Enrichment of DMRs within regions of interest.

	Observed	Expected	Enrichment <i>p</i> -Value ^a
Promoters	52	15	3.4×10^{-14}
Exons	33	14	8.0×10^{-6}
Introns	237	229	0.49
3' End	26	14	4.1×10^{-3}
Intergenic	209	285	1.3×10^{-10}
All enhancers	11	3	4.8×10^{-4}
T-cell enhancers	2	1	0.11
CD4 ⁺ DNase-sensitive regions (ENCODE)	35	5	$<2.2 \times 10^{-16}$
CD4 ⁺ transcription factor binding sites (ReMap)	34	6	1.8×10^{-17}
CD4 ⁺ CTCF binding sites	11	2	3.2×10^{-5}

Note: Five hundred fifty-seven DMRs were identified in cord blood CD4⁺ cells from newborns exposed, compared with newborns unexposed, to maternal tobacco smoke. These 557 DMRs were identified from a total of 10,381 differentially methylated CpGs using the MethPipe pipeline. Differentially methylated CpGs were identified using beta-binomial regression models, which were adjusted for baby's sex and maternal working status during pregnancy. DMR, differentially methylated region; ENCODE, Encyclopedia of DNA Elements.

^a*p*-Value from Fisher's exact test.

measures for CD4⁺ cells. However, <5% (485/10,381) of the differentially methylated CpGs identified in MACHS are covered on the Infinium EPIC array, and <4% (399/10,381) are covered on the Infinium 450K array. Results for the 485 differentially methylated CpGs identified in MACHS ($p_{FDR} < 0.05$), which are represented on the EPIC array, are shown for the WakeMed SMKE ($n = 14$ newborns exposed to sustained maternal smoking during pregnancy; $n = 16$ unexposed) and NIEHS CRU replication study ($n = 9$ smoking; $n = 11$ nonsmoking adults) in Excel Table S5. The 399 differentially methylated CpGs identified in MACHS, which are represented on the 450K array, are also shown in this table for the NIEHS CRU replication study ($n = 59$ smoking; $n = 72$ nonsmoking adults). Although only 60 of the 485 CpGs examined had a raw $p < 0.05$ in the WakeMed SMKE study, this was a larger number than would be expected by chance ($p_{Kolmogorov-Smirnov} = 5.3 \times 10^{-6}$). Of these 60 CpGs, 33 were differentially methylated in the same direction in both SMKE and MACHS (Table 4). Distributions of methylation levels within each of the 33 CpGs that replicated are shown for both MACHS and SMKE in Figure S5. The replication rate for CpGs with a mean coverage $\geq 10 \times$ (8.1%) did not differ (p -value from Fisher's exact test = 0.73) from the replication rate for CpGs with a mean coverage $< 10 \times$ (6.7%). One of the CpGs that replicated (located in the promoter of *SLC7A8*; mean coverage = $8.9 \times$) was found to be hypomethylated in the tobacco smoke exposed, compared with unexposed, group in both MACHS and the SMKE study, and was also hypomethylated in smokers compared with nonsmokers in both the EPIC and 450K NIEHS CRU studies of

adult CD4⁺ cells (Excel Table S5). Given the reproducibility of this result, we evaluated whether any other CpGs within *SLC7A8* that were identified as differentially methylated in MACHS replicated; of the 14 CpGs that could be queried, five were differentially methylated (raw $p < 0.05$) in the same direction in at least one of the replication studies (Excel Table S5). We also evaluated whether any of the CpGs contained within the 20 DMRs with the largest percentage methylation differences (Table 2) replicated in the WakeMed SMKE study. However, none of the six CpGs that could be queried replicated (Table S7).

Results for the nine DMRs that were identified in MACHS that could be queried in the NIEHS CRU RRBS study of adult CD4⁺ cells ($n = 9$ smoking; $n = 10$ nonsmoking women) are shown in Table 5. Four of these DMRs (shown in Figure 1) were significantly differentially methylated ($p < 0.05$) in the same direction in both studies (Table 5).

Comparisons with Previous Studies of Maternal Tobacco Smoke Exposure and DNA Methylation

A previous study by Bauer et al. examined associations between maternal smoking during pregnancy and whole cord blood DNA methylation, using WGBS (Bauer et al. 2016). Using a smoothing method, they identified 8,409 DMRs, with an FDR of 12.4%. Only six of the 557 DMRs ($p_{FDR} < 0.05$) identified in MACHS overlapped those identified by Bauer et al. (Table S8). The nearest genes to these six DMRs included *TRNA_VAL*, *RIN3*, *MIR646HG*, *ZNF890P*, *APBA1*, *ANK1*, and *NKX6-3* (the latter two genes were

Table 4. Differentially methylated CpGs identified in MACHS that were replicated in the SMKE cord blood CD4⁺ study ($n = 14$ exposed; $n = 16$ unexposed).

MACHS WGBS CpG	Percent methylation difference	Mean coverage	WakeMed EPIC CpG	Percent methylation difference	Raw p -value	Nearest gene	Distance (base pair)
Chr1:2112005	10.8	4.9	cg02066716	0.1	0.05	<i>PRKCZ</i>	0
Chr1:108113976	5.0	7.2	cg24950918	1.9	0.02	<i>VAV3</i>	0
Chr1:172582730	5.5	5.2	cg15874871	0.9	0.04	<i>SUCO</i>	-1,756
Chr3:194854341	8.9	11.1	cg11796219	1.1	1.6×10^{-3}	<i>XXYL1</i>	0
Chr5:8457720	-13.9	7.3	cg18371052	-6.0	7.3×10^{-3}	<i>LINC02226</i>	42
Chr6:33255589	6.1	6.5	cg00783244	0.7	0.05	<i>WDR46</i>	0
Chr6:155475641	39.9	9.5	cg05104189	0.4	0.04	<i>TIAM2</i>	0
Chr7:637221	3.9	5.1	cg14871771	1.9	0.02	<i>PRKAR1B</i>	0
Chr7:70258725	9.1	7.0	cg02327773 ^a	0.5	1.7×10^{-3}	<i>AUTS2</i>	839
Chr7:70258752	12.8	8.1	cg18542967	0.5	0.03	<i>AUTS2</i>	866
Chr8:2586007	-18.7	9.6	cg02210441	-8.1	5.3×10^{-3}	<i>LOC101927815</i>	49
Chr8:23179282	-17.4	8.5	cg12060669	-8.5	0.04	<i>LOXL2</i>	0
Chr8:74282930	24.8	9.8	cg20982046	2.5	7.0×10^{-3}	<i>RDH10-AS1</i>	14,233
Chr9:18493887	7.1	8.7	cg13552832	0.7	0.03	<i>ADAMTSL1</i>	0
Chr10:82296190	-8.6	5.7	cg09623377	-3.9	0.02	<i>SH2D4B</i>	1,467
Chr10:105420500	-20.1	11.3	cg19007269	-6.8	3.3×10^{-3}	<i>SH3PXD2A</i>	0
Chr11:73625665	8.3	5.5	cg08665076	1.6	0.02	<i>PAAF1</i>	0
Chr12:313103	-14.3	3.3	cg20188212	-0.3	0.05	<i>SLC6A12</i>	0
Chr12:133343427	7.7	9.9	cg25939853	4.3	3.2×10^{-3}	<i>GOLGA</i>	2,067
Chr14:23623727	-9.4	8.9	cg08104568	-2.4	0.03	<i>SLC7A8</i>	0
Chr16:88689493	19.6	4.9	cg01018360	0.5	0.04	<i>ZC3H18</i>	0
Chr17:8867337	11.5	10.7	cg25342409	5.2	0.02	<i>PIK3R5</i>	0
Chr17:76037363	-16.3	7.7	cg07787614	-7.6	0.04	<i>TNRC6C</i>	0
Chr18:77487393	0.6	5.0	cg13817822	3.4	0.03	<i>CTDP1</i>	0
Chr18:3412781	5.6	8.1	cg04431731	0.4	0.04	<i>TGIF1</i>	0
Chr19:613432	-12.8	3.9	cg00587228 ^a	-6.2	0.01	<i>HCN2</i>	0
Chr19:46915775	-16.9	4.6	cg17375267	-3.9	0.04	<i>CCDC8</i>	0
Chr21:44105264	-19.9	4.7	cg23590273	-6.3	0.02	<i>PDE9A</i>	0
Chr21:44105473	-23.2	4.9	cg16784985	-8.5	7.4×10^{-3}	<i>PDE9A</i>	0
Chr22:42548355	-4.3	4.4	cg17050807	-12.2	0.02	<i>TCF20</i>	7,663
ChrX:48685181	-17.1	3.4	cg09513996	-2.2	4.6×10^{-3}	<i>ERAS</i>	0
ChrX:48685302	-12.0	5.1	cg22656275	-7.6	0.02	<i>ERAS</i>	0
ChrX:134048711	-6.2	3.5	cg24138084	-3.4	0.01	<i>MOSPD1</i>	0

Note: CpGs were considered replicated if the raw p -value in the SMKE study was < 0.05 and the percent methylation difference was in the same direction as that observed in MACHS. The exposed group consisted of 14 newborns whose mothers smoked during pregnancy in SMKE. The unexposed group consisted of 16 newborns whose mothers were lifetime nonsmokers. DNA methylation was measured in cord blood CD4⁺ cells by the Infinium EPIC array for the SMKE study. Chr, chromosome; MACHS, Maternal and Child Health Study; WGBS, whole-genome bisulfite sequencing.

^aThese CpGs have previously been found to be polymorphic or to be targeted by a probe with a single-nucleotide polymorphism in the single base pair extension (Chen et al. 2013).

Table 5. DMR replication results.

MACHS WGBS DMR	MACHS percent methylation difference	NIEHS CRU RRBS DMR	NIEHS CRU percent methylation difference	Raw <i>p</i> -Value	Nearest gene	Distance (base pair)
Chr1:6852075-6852260	-14.4	Chr1:6852171-6852205	1.8	0.03	<i>CAMTA1</i>	0
Chr1:54701016-54701221	-13.3	Chr1:54701218-54701382	2.9	0.01	<i>SSBP3</i>	0
Chr2:237476491-237476692	-19.1	Chr2:237476624-23747642	15.6	0.01	<i>CXCR7</i>	0
Chr6:2764191-2764384	11.1	Chr6:2764287-2764557	4.1	0.02	<i>WRNIP1</i>	1,281
Chr7:150498649-150498694	-13.9	Chr7:150498243-150498694	6.7	0.01	<i>TMEM176A</i>	0
Chr9:72027018-72027281	17.8	Chr9:72026918-72027130	10.7	0.04	<i>APBA1</i>	15,167
Chr11:47952753-47952964	-7.8	Chr11:47952767-47952802	-2.6	0.01	<i>PTPRJ</i>	49,145
Chr12:7781003-7781099	17.1	Chr12:7780979-7781052	-19.7	0.02	<i>APOBEC1</i>	20,896
Chr19:47273681-4727802	6.7	Chr19:47273536-47273793	3.6	0.02	<i>SLCIA5</i>	4,337

Note: Chr, chromosome; DMR, differentially methylated region; MACHS, Maternal and Child Health Study; NIEHS CRU, National Institute of Environmental Health Sciences Clinical Research Unit; RRBS, reduced representation bisulfite sequencing; WGBS, whole-genome bisulfite sequencing.

^aNine DMRs identified in MACHS could be queried in a reduced representation bisulfite sequencing study conducted in CD4⁺ cells from a subset of women from the NIEHS CRU study. DMRs identified in MACHS were considered replicated if a) they overlapped a DMR identified in the NIEHS CRU study; b) the overlapping DMR had a *p* < 0.05; and c) the percent methylation difference for the overlapping DMR was in the same direction as the DMR identified in MACHS.

associated with one DMR, located on chromosome 8). Two of these DMRs (one near *TRNA_VAL* and one near *APBA1*) were differentially methylated in the same direction (both were hypermethylated in the exposed, compared with unexposed, group), and

one of these DMRs (chr9:72,027,018-72,027,281) was also identified in the NIEHS CRU RRBS replication study of CD4⁺ cells from smoking and nonsmoking women. Across all three studies, this DMR was found to be hypermethylated in the tobacco smoke

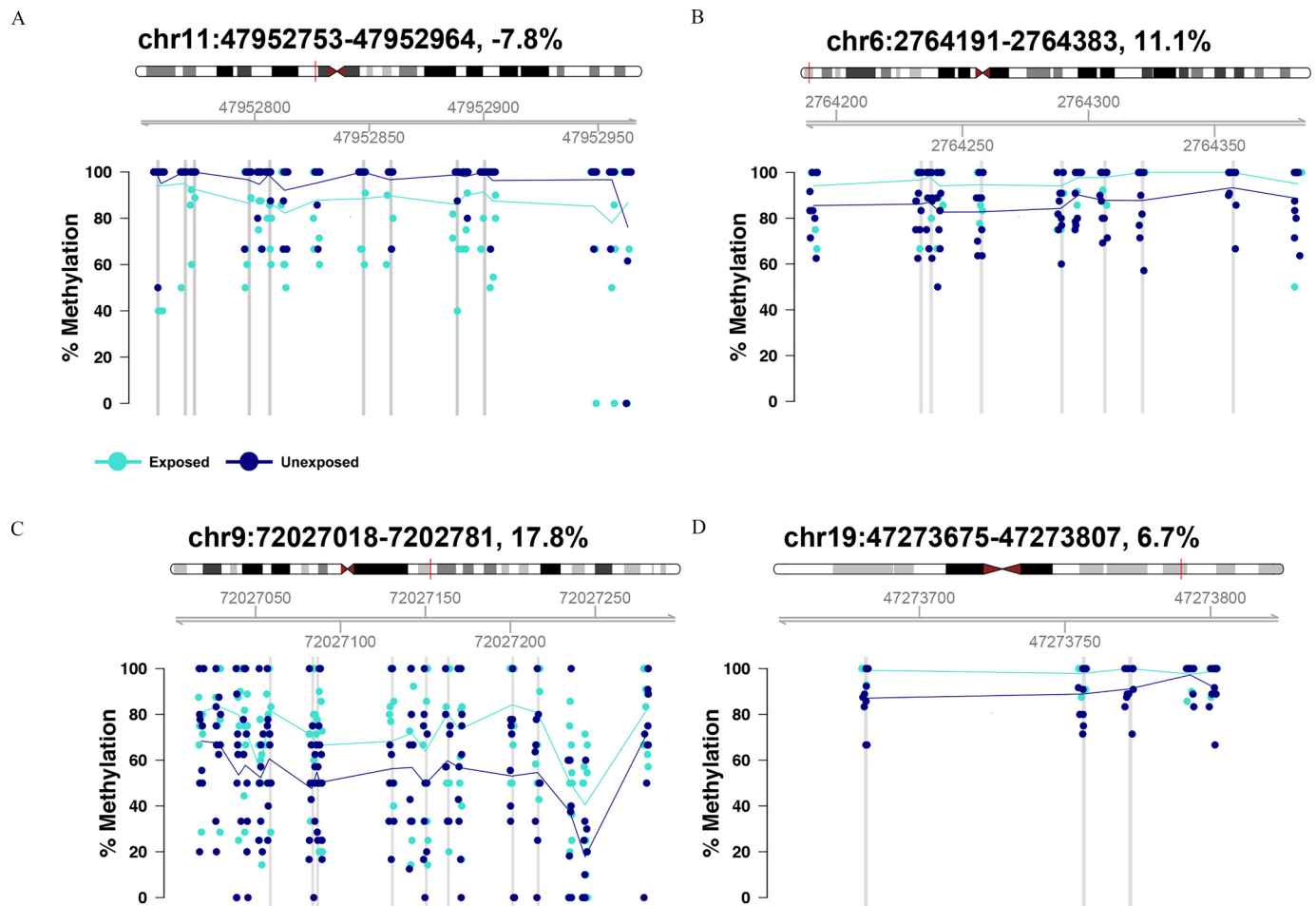


Figure 1. The four differentially methylated regions (DMRs) identified in the Maternal and Child Health Study (MACHS), which replicated in the National Institute of Environmental Health Sciences Clinical Research Unit (NIEHS CRU) reduced representation bisulfite sequencing study of adult CD4⁺ cells, are depicted in (A–D). DMRs were considered replicated if a) the chromosome coordinates overlapped between the two studies; b) the *p*-value for the NIEHS CRU study DMR was < 0.05; and c) the percent methylation difference was in the same direction. Coordinates for each DMR and the percent methylation difference (comparing maternal tobacco smoke exposed with unexposed newborns) are indicated at the top of each panel. The location of each DMR is indicated by a vertical red line in the ideogram of its respective chromosome. Plots show the percent methylation levels (*y*-axis) for each MACHS participant at each CpG contained in the DMR. Chromosome positions for CpGs are indicated in gray (*x*-axis) above the plot. Percent methylation levels for participants in the exposed group are shown as light blue dots, and mean percent methylation levels for each CpG are connected by light blue lines. Percent methylation levels for participants in the unexposed group are shown as dark blue dots, and mean percent methylation levels for each CpG are connected by dark blue lines. Vertical gray bars highlight CpGs that were identified as differentially methylated with both a raw *p* < 0.05 and a *p* < 0.05 after accounting for correlations with neighboring CpGs and multiple testing, using the Benjamini-Hochberg method to control the false discovery rate.

exposed, compared with unexposed, group (percent methylation differences: 17.8% in MACHS, 10.8% in the WGBS whole cord blood study by Bauer et al., and 10.7% in the NIEHS CRU RRBS study of adult CD4⁺ cells).

We also conducted a look-up analysis for 26 CpGs identified as differentially methylated ($p_{\text{Bonferroni}} < 0.05$) in cord blood by sustained maternal tobacco smoke exposure in a previous 450K array study, as many of these results have been widely replicated (Joubert et al. 2012). Twenty-five of the 26 CpGs were represented in the MACHS WGBS results and could therefore be queried. Sequencing coverage at these CpGs ranged from 1.9–10.6 \times . None of the 25 CpGs were found to be differentially methylated in MACHS (raw p -value and $p_{\text{FDR}} > 0.05$), although the directions of associations were similar for several sites (Table 6). For example, all four of the CpGs that have previously been identified as differentially methylated in *AHRR* were differentially methylated in the same direction in MACHS. To determine whether the inconsistencies between the Joubert et al. study and MACHS were due to the differences in the cell types examined, we conducted a look-up analysis for the same CpGs in the WakeMed SMKE EPIC array study of cord blood CD4⁺ cells. Of the 26 differentially methylated CpGs identified by Joubert et al., 23 are represented on the EPIC array; of these, 11 (47.8%) were significantly ($p < 0.05$) differentially methylated in the same direction in the SMKE study. This included one CpG that could not be examined in MACHS. For the 10 CpGs that were differentially methylated in SMKE but not MACHS, the median

coverage was 4.7 \times . This did not differ ($p_{\text{Wilcoxon}} = 0.31$) from the median coverage (7.8 \times) of the 12 CpGs that were not differentially methylated in either SMKE or MACHS.

We also conducted a look-up analysis for CpGs that were identified as differentially methylated ($p_{\text{FDR}} < 0.05$) in whole cord blood in relation to both sustained and any maternal smoking during pregnancy in meta-analyses conducted by the PACE consortium (Joubert et al. 2016). Of the 3,272 differentially methylated CpGs ($p_{\text{FDR}} < 0.05$) identified by PACE, 3,258 were represented in the MACHS WGBS dataset and could therefore be evaluated (Excel Table S6). Of these CpGs, 213 were differentially methylated in MACHS based on a raw $p < 0.05$, and 267 were differentially methylated after adjusting for correlations with proximal CpGs ($p_{\text{neighbor}} < 0.05$). These numbers are higher than would be expected by chance ($p_{\text{Kolmogorov-Smirnov}} = 3.9 \times 10^{-4}$ and 3.2×10^{-11} , respectively). Of the 213 CpGs with a raw $p < 0.05$, 123 were differentially methylated in the same direction in MACHS. Results did not differ for CpGs with an average coverage $\geq 10 \times$ vs. $< 10 \times$ (p -value from chi-square test = 1).

Discussion

Although earlier studies have evaluated the impact of maternal smoking during pregnancy on DNA methylation patterns, the majority used candidate gene approaches or methylation arrays, which cover a small fraction (<5%) of the >28 million CpG sites

Table 6. Look-up results for 26 CpGs identified as differentially methylated by maternal tobacco smoke exposure by Joubert et al. (2012) in MACHS and the SMKE study.

CpG	Position	Gene	Joubert et al. 2012		MACHS				WakeMed SMKE			
			MoBa (450K)		NEST (450K)		WGBS			EPIC		
			Percent meth. diff.	p -Value	Percent meth. diff.	p -Value	Percent meth. diff.	p -Value	p_{FDR}^a	Ave. coverage	Percent meth. diff.	p -Value
cg10399789	Chr1:92945668	<i>GFII</i>	-3.7	1.1×10^{-10}	-3.4	0.02	-2.6	0.44	0.44	5.9	-0.5	0.44
cg09662411	Chr1:92946132	<i>GFII</i>	-6.6	3.0×10^{-17}	-6.8	2.3×10^{-3}	6.8	0.26	0.78	3.3	-1.6	0.61
cg06338710	Chr1:92946187	<i>GFII</i>	-5.8	5.0×10^{-14}	-5.2	2.6×10^{-3}	-4.3	0.57	0.66	4.0	-0.7	0.39
cg18146737	Chr1:92946700	<i>GFII</i>	-12.3	3.3×10^{-25}	-15.1	3.7×10^{-3}	0.8	0.66	0.64	4.3	-1.5	0.03
cg12876356	Chr1:92946825	<i>GFII</i>	-11.9	1.7×10^{-25}	-10.5	1.5×10^{-3}	-0.02	0.15	0.43	5.1	-2.1	0.01
cg18316974	Chr1:92947035	<i>GFII</i>	-7.1	3.2×10^{-20}	-9.0	2.3×10^{-3}	-4.3	0.50	0.54	3.9	-0.6	0.12
cg09935388 ^b	Chr1:92947588	<i>GFII</i>	-13.7	2.7×10^{-31}	-7.5	1.2×10^{-3}	NA	NA	NA	NA	-5.9	5.0×10^{-4}
cg14179389	Chr1:92947961	<i>GFII</i>	-8.6	2.6×10^{-25}	-8.2	4.3×10^{-3}	-3.4	0.43	0.74	4.1	-10.7	1.9×10^{-3}
cg23067299 ^c	Chr5:323907	<i>AHRR</i>	3.2	4.12×10^{-9}	3.7	3.6×10^{-3}	10.3	0.07	0.63	8.1	NA	NA
cg03991871 ^c	Chr5:368447	<i>AHRR</i>	-2.2	2.0×10^{-10}	-2.3	0.07	-6.9	0.52	0.86	8.2	NA	NA
cg05575921	Chr5:373378	<i>AHRR</i>	-7.5	8.0×10^{-33}	-7.7	3.0×10^{-4}	-1.4	0.79	>0.99	7.3	-4.3	2.1×10^{-8}
cg21161138	Chr5:399360	<i>AHRR</i>	-2.3	8.9×10^{-10}	-1.7	3.6×10^{-3}	-0.05	0.99	>0.99	9.3	-3.2	0.01
cg11715943	Chr6:33091841	<i>HLA-DPB2</i>	-1.8	3.6×10^{-8}	-0.3	0.80	1.7	0.69	0.96	7.8	1.2	0.11
cg19089201	Chr7:45002287	<i>MYO1G</i>	1.4	9.1×10^{-11}	2.2	9.2×10^{-3}	0.6	0.71	0.43	3.5	0.4	0.05
cg22132788 ^c	Chr7:45002486	<i>MYO1G</i>	2.8	4.8×10^{-18}	2.1	9.6×10^{-3}	4.3	0.81	0.42	2.6	NA	NA
cg04180046	Chr7:45002736	<i>MYO1G</i>	5.3	2.9×10^{-19}	4.9	2.7×10^{-3}	-5.1	0.80	0.69	1.9	2.7	0.01
cg12803068	Chr7:45002919	<i>MYO1G</i>	8.3	1.3×10^{-19}	3.8	4.1×10^{-3}	0.0	0.97	0.92	3.5	0.7	0.03
cg04598670	Chr7:68697651	<i>ENSG00000225718</i>	-3.0	1.3×10^{-9}	-5.3	3.6×10^{-3}	-0.3	0.96	>0.99	10.3	-0.5	0.89
cg25949550	Chr7:145814306	<i>CNTNAP2</i>	-1.8	1.0×10^{-26}	-2.5	2.5×10^{-3}	1.1	0.73	0.61	10.6	-1.5	6.7×10^{-3}
cg03346806	Chr8:119157879	<i>EXT1</i>	-1.5	9.3×10^{-8}	-0.1	0.12	5.8	0.21	0.81	6.9	0.6	0.17
cg18655025	Chr14:91008005	<i>TTC7B</i>	-1.2	6.8×10^{-8}	-1.1	0.27	6.0	0.21	0.82	9.0	-0.1	0.24
cg05549655	Chr15:75019143	<i>CYP1A1</i>	3.5	2.4×10^{-10}	3.8	6.0×10^{-4}	1.9	>0.99	0.90	7.6	3.9	6.9×10^{-3}
cg22549041	Chr15:75019251	<i>CYP1A1</i>	7.2	8.9×10^{-9}	8.9	4.4×10^{-3}	-8.7	0.07	0.98	7.8	4.9	0.41
cg11924019	Chr15:75019283	<i>CYP1A1</i>	3.2	4.8×10^{-8}	5.3	8.0×10^{-4}	-2.0	0.57	0.98	8.2	1.3	0.68
cg18092474	Chr15:75019302	<i>CYP1A1</i>	5.9	10.0×10^{-9}	5.3	4.4×10^{-3}	-4.5	0.37	0.99	7.9	5.1	0.22
cg12477880	Chr21:36259241	<i>RUNX1</i>	4.6	7.6×10^{-10}	4.0	0.09	3.2	0.21	0.98	8.1	1.6	0.21

Note: Twenty-six CpGs have previously been identified as differentially methylated (after adjusting for multiple testing using the Bonferroni correction) in whole cord blood by maternal tobacco smoke exposure in two study populations (MoBa and NEST), evaluated by Joubert et al. 2012, and have been widely replicated in subsequent studies. Look-up analyses were therefore conducted for these CpGs in both MACHS and the SMKE study. Percent methylation differences and raw p -values are presented for each CpG by study. False discovery rate-adjusted p -values are also presented for MACHS. Chr, chromosome; FDR, false discovery rate; MACHS, Maternal and Child Health Study; MoBa, Norwegian Mother and Child Cohort Study; NEST, Newborn Epigenetics Study; WGBS, whole-genome bisulfite sequencing.

^aRaw p -values from RADmeth beta-binomial regression models were adjusted for correlations with neighboring CpGs, as described previously (Dolzhenko and Smith 2014), and were subsequently adjusted for multiple testing using the Benjamini-Hochberg method to control the false discovery rate.

^bSince all reads in the exposed and unexposed group were methylated in MACHS, RADmeth did not calculate a p -value for this CpG.

^cThese CpGs are not represented on the Infinium EPIC array and therefore could not be examined in the SMKE study.

in the human genome, and have traditionally focused on promoter regions. Using WGBS with $6.5\times$ coverage as an alternative approach, we were able to interrogate >26 million CpG sites, including those located distally from genes, and identified 557 regions that were differentially methylated by exposure to maternal smoking *in utero*. While we observed enrichment of DMRs within promoter regions, we also observed enrichment in enhancer regions, which have been largely uninvestigated, and in $CD4^+$ DNase-sensitive regions and transcription factor binding sites, including binding sites for CTCF, which are methylation sensitive and regulate gene expression via alterations to the three-dimensional structure of the genome (Zuin et al. 2014).

To our knowledge, only one previous study, conducted by Bauer et al. (2016), has utilized WGBS to investigate the effects of maternal tobacco smoke exposure during pregnancy on the newborn methylome. Similar to our findings, Bauer et al. observed significant enrichment of DMRs within enhancer regions. However, only six of the DMRs identified by Bauer et al. overlapped those identified by MACHS. Since the smoothing method used by Bauer et al. has been shown to yield similar results as the method used in the current study (MethPipe) (Dolzhenko and Smith 2014), other study differences likely contributed to these discrepancies. Two potential explanations include the fact that Bauer et al. evaluated sustained maternal smoking during pregnancy and measured DNA methylation in whole blood, similar to most previous studies (Breton et al. 2014; de Vocht et al. 2015; Ivorra et al. 2015; Joubert et al. 2012; Ladd-Acosta et al. 2016; Markunas et al. 2014; Richmond et al. 2015; Rotroff et al. 2016; Rzehak et al. 2016), whereas MACHS evaluated any maternal smoking during pregnancy and focused on $CD4^+$ cells. We also conducted a look-up analysis of 3,258 CpGs that were identified as differentially methylated in cord blood in relation to both any and sustained maternal smoking during pregnancy in meta-analyses conducted by the PACE consortium (Joubert et al. 2016). While a larger number of these sites were differentially methylated in MACHS than would be expected by chance, only 123 (3.8%) were differentially methylated in the same direction with a raw $p < 0.05$.

To determine if the MACHS results were more reproducible in studies focusing on the same cell type, we conducted a series of replication analyses using data from studies that measured DNA methylation in $CD4^+$ cells from newborns or adults. We were able to query nine of the DMRs identified in MACHS in an RRBS study of $CD4^+$ cells from smoking and nonsmoking women (NIEHS CRU), four (44%) of which were significantly differentially methylated in the same direction. Of these four DMRs, one (chr9:72,027,018-72,027,281) was also identified in the whole cord blood WGBS study by Bauer et al. (2016), which suggests that this region may be sensitive to tobacco smoke exposure in multiple blood cell types. In contrast with the DMR replication results, very few of the individual CpGs that were identified as differentially methylated in MACHS that could be queried were replicated. For example, of the 485 differentially methylated CpGs identified in MACHS that could be evaluated in an EPIC array study (SMKE) of cord blood $CD4^+$ cells, only 33 (6.8%) replicated. This suggests that a large fraction of the differentially methylated CpGs identified in MACHS that are represented on the EPIC array may be false positives, although it is possible that differences in participant characteristics, such as the different racial/ethnic compositions of the two study populations, and assay differences may also play a role. Although replication rates were low for the individual CpG results, one CpG (chr14:23,623,727) was identified as hypomethylated among maternal tobacco smoke exposed, compared with unexposed, newborns in both MACHS and the SMKE study, and was also

found to be hypomethylated in smokers, compared with non-smokers, in the NIEHS CRU study of adult $CD4^+$ cells, which was confirmed across two different platforms (i.e., the EPIC array and the 450K array). This result is therefore highly reproducible and suggests that this CpG, located in the promoter of amino acid transporter *SLC7A8* (Ren et al. 2017), may be sensitive to tobacco smoke exposure in $CD4^+$ cells, irrespective of life stage.

Since a subset of CpGs on the Infinium 450K array have been consistently identified as differentially methylated by sustained maternal smoking during pregnancy in whole cord blood (Joubert et al. 2012), we were interested in evaluating whether these CpGs are also differentially methylated in $CD4^+$ cells. While some of these loci (e.g., all CpGs within *AHRR*) were differentially methylated in the same direction in MACHS, none were statistically significant. To determine if these discrepancies were driven by the different cell types examined, we conducted the same look-up analysis in the WakeMed SMKE EPIC array study of cord blood $CD4^+$ cells. Of the 23 CpGs that could be queried in SMKE, 11 (48%) were significantly differentially methylated in the same direction as that observed by Joubert et al., which suggests that about half of these CpGs are sensitive to tobacco smoke in $CD4^+$ cells, but the other half may not be. The latter result is not surprising, as there is evidence that exposure to maternal tobacco smoke *in utero* alters methylation patterns differently in certain leukocyte subpopulations (Bauer et al. 2016; Su et al. 2016). While it is currently unclear why the differentially methylated CpGs identified by Joubert et al. that replicated in SMKE were not also differentially methylated in MACHS, two possible explanations include the low coverage of the WGBS data, which may have reduced our statistical power, and the weaker exposure variable, as most of the MACHS mothers who smoked quit early in pregnancy.

Our study has some limitations that need to be considered. First, since MACHS participants were recruited at delivery, maternal samples were not collected during pregnancy, so exposure status could not be confirmed using maternal cotinine measures. However, women are more likely to underreport, rather than falsely report, that they smoked during pregnancy (Gorber et al. 2009), and there is no reason to believe that this would occur differentially by newborn methylation status. Therefore, any potential exposure misclassification would have likely biased results toward the null. Another important consideration is that 40% of the women who reported smoking during pregnancy resided in a household with another smoker. However, previous studies have observed that the effects of maternal smoking on newborn DNA methylation are much stronger than those of other secondhand smoke exposures (Bouwland-Both et al. 2015; Richmond et al. 2015). An additional limitation is that there may have been some confounding by acculturation and/or ethnic substructure, as this information was not collected for MACHS participants, but both have been related to smoking behaviors (Detjen et al. 2007; Kondo et al. 2016). Importantly, a genomic inflation factor (λ) of 1.25 was observed for the current study, suggesting possible population stratification. We therefore cannot rule out the possibility that there may have been some genetic differences between groups, particularly given our small sample size. This is important as certain SNPs can destroy a CpG site and thus impact methylation at that locus. However, only a small fraction (1.6%) of the differentially methylated CpGs identified in MACHS overlapped common SNPs identified in a similar population (i.e., MXL of 1000 Genomes), so this likely had a minimal impact on our results. While we focused our study on sorted $CD4^+$ cells to minimize potential confounding by cell type heterogeneity, it is possible that maternal smoking during pregnancy

may alter CD4⁺ subpopulations, which could influence our results. Furthermore, we cannot necessarily generalize our findings to tissues other than CD4⁺ cells. Another important limitation is that time from cord blood collection until sample processing varied between individuals. However, this did not differ by exposure status. While we observed significant enrichment of differentially methylated CpGs and DMRs within key regulatory elements, we did not measure gene expression and therefore cannot make definitive conclusions about the functional relevance of our findings. Follow-up studies will therefore be important to evaluate the impact of altered methylation levels at the sensitive regions identified in MACHS on related histone modifications and gene expression. Finally, it is possible that some of the observed associations between exposure to maternal smoking *in utero* and DNA methylation are driven by alterations in 5-hydroxymethylcytosine and related modifications, as these marks are indistinguishable from 5-methylcytosine after bisulfite conversion (Huang et al. 2010).

For the current study, we used a method (RADmeth) that has been shown to perform well on very low-coverage WGBS data within a case-control context (Dolzhenko and Smith 2014). Nevertheless, we cannot rule out the possibility that the low coverage of the MACHS WGBS data may have still contributed to some false positives, both for the individual CpG and DMR results, as RADmeth's performance on low-coverage data has not been evaluated in the context of environmental exposures, which have generally been associated with smaller methylation differences. However, while the replication rate for the individual CpG results was very low (6.8%), this did not differ by sequencing coverage, which suggests that other factors likely contributed to the poor replication rate. Since MethPipe does not systematically evaluate outliers, it is possible that outliers may have contributed to some false positives in our results. It is also possible that the modest coverage and the small sample size of the current study may have contributed to some false negatives. For example, none of the CpGs identified as differentially methylated by Joubert et al. (2012) that have been widely replicated in studies of whole cord blood were significantly differentially methylated in MACHS. While some of these discrepancies may have been driven by differences in the cell types examined, a subset of these CpGs were found to be differentially methylated in the WakeMed study of cord blood CD4⁺ cells, which suggests that other factors, such as the low coverage of our data or the weaker maternal smoking exposure in MACHS, may have contributed to the null results. Future WGBS studies with higher coverage, which focus on specific cell types, are greatly needed to address some of these limitations. In particular, down-sampling simulations using higher coverage datasets are needed to determine the average sequencing coverage and sample sizes required to minimize false positives and false negatives in studies investigating environmental impacts on the DNA methylome.

Despite these limitations, our study also had several unique strengths. By using WGBS, we were able to interrogate millions of CpG sites that have been understudied in relation to maternal tobacco smoke exposure, including CpGs within critical regulatory regions, such as enhancers, which tend to be underrepresented in methylation arrays. Since we looked at a much more homogenous cell population than has previously been investigated, our findings are also less subject to potential confounding by cell type heterogeneity. Furthermore, since CD4⁺ cells have been implicated in the development of tobacco-related illnesses, such as asthma, they represent an ideal tissue to evaluate in the context of tobacco smoke exposure (Akbari et al. 2006; Ling and Luster 2016; Lloyd and Hessel 2010). Also, a novel aspect of the current study was its focus on Hispanic white individuals, a group

that has been largely understudied in this context. An additional strength of our study was the matched design. In addition to focusing on Hispanic white participants to reduce potential confounding by race or ethnicity, we matched on several important confounders, including maternal age, gestational age, maternal diabetes status, and maternal BMI, to try to isolate the effects of exposure to maternal smoking *in utero* on the newborn methylome. Another major strength of our study was the inclusion of replication studies that also profiled DNA methylation in CD4⁺ cells, which confirmed four regions and 33 CpGs that were identified as differentially methylated in MACHS. Finally, since MACHS was designed to identify early-life risk factors for adverse outcomes in childhood, regions found to be differentially methylated by maternal tobacco smoke exposure in the current study can be examined in relation to respiratory and other health outcomes in future studies.

Conclusions

Our findings contribute to growing evidence that exposure to maternal smoking *in utero* alters DNA methylation patterns in newborns, and present the first genome-wide description of the impact of maternal tobacco smoke exposure during pregnancy on the cord blood CD4⁺ DNA methylome. In particular, our results suggest that maternal tobacco smoke exposure during pregnancy alters DNA methylation levels within coding regions and in important regulatory regions, such as enhancers, which have been largely uninvestigated. Although we used a WGBS pipeline that performs well on low-coverage data, our results should be interpreted with some caution, as it is still possible that the low-coverage of the data contributed to false positives. Nevertheless, we observed that a subset of results that could be queried replicated in several external study populations. For example, four of the nine DMRs identified in MACHS that could be examined in a study of adult CD4⁺ cells were confirmed, and one differentially methylated CpG, located in the promoter region of *SLC7A8*, was validated in two different studies of CD4⁺ cells (one conducted in newborns and the other in adults), which suggests that these loci are sensitive to tobacco smoke exposure in CD4⁺ cells, and that these effects may span different life stages.

Acknowledgments

The authors would like to thank the field and laboratory staff for their dedication and hard work in collecting and processing cord blood samples, and would also like to acknowledge B. Decato and other members of A. Smith's lab for providing training and guidance in utilizing the WALT and MethPipe softwares. This study was funded by National Institutes of Health (NIH) grants R01 ES022216, T32 ES013678, P30 ES007048, P30 ES025128, and P50 ES026086, U.S. Environmental Protection Agency (EPA) grant 83615801-0, the NIH Center for Tobacco Research, and the Intramural Research Program of the National Institute of Environmental Health Science (Z01-ES100475).

References

- Agarwal V, Bell GW, Nam JW, Bartel DP. 2015. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* 4:e05005, PMID: 26267216.
- Akbari O, Faul JL, Hoyte EG, Berry GJ, Wahlström J, Kronenberg M, et al. 2006. Cd4+ invariant T-cell-receptor+ natural killer T cells in bronchial asthma. *N Engl J Med* 354(11):1117–1129, PMID: 16540612, <https://doi.org/10.1056/NEJMoa053614>.
- Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. 2014. An atlas of active enhancers across human cell types and tissues. *Nature* 507(7493):455–461, PMID: 24670763, <https://doi.org/10.1038/nature12787>.
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. 2013. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* 41(D1):D991–D995, <https://doi.org/10.1093/nar/gks1193>.

- Bauer M, Fink B, Thürmann L, Eszlinger M, Herberth G, Lehmann I. 2016. Tobacco smoking differently influences cell types of the innate and adaptive immune system—indications from CPG site methylation. *Clin Epigenetics* 8:83, PMID: 27493699, <https://doi.org/10.1186/s13148-016-0249-7>.
- Bauer T, Trump S, Ishaque N, Thürmann L, Gu L, Bauer M, et al. 2016. Environment-induced epigenetic reprogramming in genomic regulatory elements in smoking mothers and their children. *Mol Syst Biol* 12(3):861, PMID: 27013061, <https://doi.org/10.15252/msb.20156520>.
- Behl M, Rao D, Aagaard K, Davidson TL, Levin ED, Slotkin TA, et al. 2013. Evaluation of the association between maternal smoking, childhood obesity, and metabolic disorders: a national toxicology program workshop review. *Environ Health Perspect* 121(2):170–180, PMID: 23232494, <https://doi.org/10.1289/ehp.1205404>.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 57(1):289–300, <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- Bhasin JM, Ting AH. 2016. Goldmine integrates information placing genomic ranges into meaningful biological contexts. *Nucleic Acids Res* 44(12):5550–5556, PMID: 27257071, <https://doi.org/10.1093/nar/gkw477>.
- Bouwland-Both MI, van Mil NH, Tolhoek CP, Stolk L, Eilers PH, Verbiest MM, et al. 2015. Prenatal parental tobacco smoking, gene specific DNA methylation, and newborns size: the Generation R study. *Clin Epigenetics* 7:83, PMID: 26265957, <https://doi.org/10.1186/s13148-015-0115-z>.
- Breton CV, Siegmund KD, Joubert BR, Wang X, Qui W, Carey V, et al. 2014. Prenatal tobacco smoke exposure is associated with childhood DNA CpG methylation. *PLoS One* 9(6):e99716, PMID: 24964093, <https://doi.org/10.1371/journal.pone.0099716>.
- Chen YA, 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 HumanMethylation 450 microarray. *Epigenetics* 8(2):203–209, PMID: 23314698, <https://doi.org/10.4161/epi.23470>.
- Chen H, Smith AD, Chen T. 2016. WALT: fast and accurate read mapping for bisulfite sequencing. *Bioinformatics* 32(22):3507–3509, PMID: 27466624, <https://doi.org/10.1093/bioinformatics/btw490>.
- Clifford A, Lang L, Chen R. 2012. Effects of maternal cigarette smoking during pregnancy on cognitive parameters of children and young adults: a literature review. *Neurotoxicol Teratol* 34(6):560–570, PMID: 23022448, <https://doi.org/10.1016/j.ntt.2012.09.004>.
- Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K. 2009. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res* 19(1):24–32, PMID: 19056695, <https://doi.org/10.1101/gr.082800.108>.
- Curtin SC, Matthews T. 2016. Smoking prevalence and cessation before and during pregnancy: Data from the birth certificate, 2014. *Natl Vital Stat Rep* 65(1):1–14, PMID: 26905977.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158, PMID: 21653522, <https://doi.org/10.1093/bioinformatics/btr330>.
- de Vocht F, Simpkin AJ, Richmond RC, Relton C, Tilling K. 2015. Assessment of offspring DNA methylation across the lifecourse associated with prenatal maternal smoking using Bayesian mixture modelling. *Int J Environ Res Public Health* 12(11):14461–14476, PMID: 26580635, <https://doi.org/10.3390/ijerph121114461>.
- Detjen MG, Nieto FJ, Trentham-Dietz A, Fleming M, Chasan-Taber L. 2007. Acculturation and cigarette smoking among pregnant Hispanic women residing in the United States. *Am J Public Health* 97(11):2040–2047, PMID: 17901446, <https://doi.org/10.2105/AJPH.2006.095505>.
- Dolzhenko E, Smith AD. 2014. Using beta-binomial regression for high-precision differential methylation analysis in multifactor whole-genome bisulfite sequencing experiments. *BMC Bioinformatics* 15:215, PMID: 24962134, <https://doi.org/10.1186/1471-2105-15-215>.
- Dweep H, Gretz N, Sticht C. 2014. miRWalk database for miRNA-target interactions. *Methods Mol Biol* 1182:289–305, PMID: 25055920.
- Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30(1):207–210, PMID: 11752295, <https://doi.org/10.1093/nar/30.1.207>.
- Encode Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74, PMID: 22955616, <https://doi.org/10.1038/nature11247>.
- FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al. 2014. A promoter-level mammalian expression atlas. *Nature* 507(7493):462–470, <https://doi.org/10.1038/nature13182>.
- Flom JD, Ferris JS, Liao Y, Tehranifar P, Richards CB, Cho YH, et al. 2011. Prenatal smoke exposure and genomic DNA methylation in a multiethnic birth cohort. *Cancer Epidemiol Biomarkers Prev* 20(12):2518–2523, PMID: 21994404, <https://doi.org/10.1158/1055-9965.EPI-11-0553>.
- Gene Ontology Consortium. 2017. Expansion of the gene ontology knowledgebase and resources. *Nucleic acids research* 45:D331–D338, PMID: 27899567, <https://doi.org/10.1093/nar/gkw1108>.
- Gilliland FD, Li YF, Peters JM. 2001. Effects of maternal smoking during pregnancy and environmental tobacco smoke on asthma and wheezing in children. *Am J Respir Crit Care Med* 163(2):429–436, PMID: 11179118, <https://doi.org/10.1164/ajrccm.163.2.2006009>.
- Gorber SC, Schofield-Hurwitz S, Hardt J, Levasseur G, Tremblay M. 2009. The accuracy of self-reported smoking: a systematic review of the relationship between self-reported and cotinine-assessed smoking status. *Nicotine Tob Res* 11(1):12–24, PMID: 19246437, <https://doi.org/10.1093/ntr/ntn010>.
- Griffon A, Barbier O, Dalino J, van Helden J, Spicuglia S, Ballester B. 2014. Integrative analysis of public ChIP-seq experiments reveals a complex multi-cell regulatory landscape. *Nucleic Acids Res* 43(4):e27, PMID: 25477382, <https://doi.org/10.1093/nar/gku1280>.
- Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, et al. 2019. The ucsc genome browser database: 2019 update. *Nucleic Acids Res* 47(D1):D853–D858, PMID: 30407534, <https://doi.org/10.1093/nar/gky1095>.
- He Y, Chen J, Zhu LH, Hua LL, Ke FF. 2017. Maternal smoking during pregnancy and ADHD: results from a systematic review and meta-analysis of prospective cohort studies. *J Atten Disord*, PMID: 29039728, <https://doi.org/10.1177/1087054717696766>.
- Holz NE, Boecker R, Baumeister S, Hohm E, Zohsel K, Buchmann AF, et al. 2014. Effect of prenatal exposure to tobacco smoke on inhibitory control: neuroimaging results from a 25-year prospective study. *JAMA Psychiatry* 71(7):786–796, PMID: 24828276, <https://doi.org/10.1001/jamapsychiatry.2014.343>.
- Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. 2010. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One* 5(1):e8888, PMID: 20126651, <https://doi.org/10.1371/journal.pone.0008888>.
- Ivorra C, Fraga MF, Bayón GF, Fernández AF, García-Vicent C, Chaves FJ, et al. 2015. DNA methylation patterns in newborns exposed to tobacco in utero. *J Transl Med* 13:25, PMID: 25623364, <https://doi.org/10.1186/s12967-015-0384-5>.
- John EM, Savitz DA, Sandler DP. 1991. Prenatal exposure to parents' smoking and childhood cancer. *Am J Epidemiol* 133(2):123–132, PMID: 1822074, <https://doi.org/10.1093/oxfordjournals.aje.a115851>.
- Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 13(7):484–492, PMID: 22641018, <https://doi.org/10.1038/nrg3230>.
- Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, et al. 2016. DNA methylation in newborns and maternal smoking in pregnancy: genome-wide consortium meta-analysis. *Am J Hum Genet* 98(4):680–696, PMID: 27040690, <https://doi.org/10.1016/j.ajhg.2016.02.019>.
- Joubert BR, Häberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, et al. 2012. 450k epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ Health Perspect* 120(10):1425–1431, PMID: 22851337, <https://doi.org/10.1289/ehp.1205412>.
- Kennedy MA, Barrera GC, Nakamura K, Baldán A, Tarr P, Fishbein MC, et al. 2005. ABCG1 has a critical role in mediating cholesterol efflux to HDL and preventing cellular lipid accumulation. *Cell Metab* 1(2):121–131, <https://doi.org/10.1016/j.cmet.2005.01.002>.
- Kleinman JC, Pierre MB Jr, Madans JH, Land GH, Schramm WF. 1988. The effects of maternal smoking on fetal and infant mortality. *Am J Epidemiol* 127(2):274–282, PMID: 3337083, <https://doi.org/10.1093/oxfordjournals.aje.a114803>.
- Kondo KK, Rossi JS, Schwartz SJ, Zamboanga BL, Scalf CD. 2016. Acculturation and cigarette smoking in Hispanic women: a meta-analysis. *J Ethn Subst Abuse* 15(1):46–72, PMID: 26114872, <https://doi.org/10.1080/15332640.2014.1002878>.
- Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 44(W1):W90–W97, PMID: 27141961, <https://doi.org/10.1093/nar/gkw377>.
- Küpers LK, Xu X, Jankipersadsing SA, Vaez A, la Bastide-van Gemert S, Scholtens S, et al. 2015. DNA methylation mediates the effect of maternal smoking during pregnancy on birthweight of the offspring. *Int J Epidemiol* 44(4):1224–1237, PMID: 25862628, <https://doi.org/10.1093/ije/dyv048>.
- 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(Pt A):139–148, PMID: 26610292, <https://doi.org/10.1016/j.envres.2015.11.014>.
- Lee KW, Richmond R, Hu P, French L, Shin J, Bourdon C, et al. 2015. Prenatal exposure to maternal cigarette smoking and DNA methylation: epigenome-wide association in a discovery sample of adolescents and replication in an independent cohort at birth through 17 years of age. *Environ Health Perspect* 123(2):193–199, PMID: 25325234, <https://doi.org/10.1289/ehp.1408614>.

- Ling MF, Luster AD. 2016. Allergen-specific CD4(+) T cells in human asthma. *Ann Am Thorac Soc* 13(Suppl 1):S25–S30, PMID: [27027948](https://doi.org/10.1513/AnnalsATS.201507-431MG), <https://doi.org/10.1513/AnnalsATS.201507-431MG>.
- Lloyd CM, Hessel EM. 2010. Functions of T cells in asthma: more than just T(H)2 cells. *Nat Rev Immunol* 10(12):838–848, PMID: [21060320](https://doi.org/10.1038/nri2870), <https://doi.org/10.1038/nri2870>.
- Markunas CA, Xu Z, Harlid S, Wade PA, Lie RT, Taylor JA, et al. 2014. Identification of DNA methylation changes in newborns related to maternal smoking during pregnancy. *Environ Health Perspect* 122(10):1147–1153, PMID: [24906187](https://doi.org/10.1289/ehp.1307892), <https://doi.org/10.1289/ehp.1307892>.
- Meberg A, Marstein S. 1986. Smoking during pregnancy—effects on the fetal thyroid function. *Acta Paediatr Scand* 75(5):762–766, PMID: [3564944](https://doi.org/10.1111/j.1651-2227.1986.tb10287.x), <https://doi.org/10.1111/j.1651-2227.1986.tb10287.x>.
- Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, et al. 2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res* 45(D1):D183–D189, PMID: [27899595](https://doi.org/10.1093/nar/gkw1138), <https://doi.org/10.1093/nar/gkw1138>.
- Paraskevopoulou MD, Vlachos IS, Hatzigeorgiou AG. 2016. DIANA-tarbase and DIANA suite tools: studying experimentally supported microRNA targets. *Curr Protoc Bioinformatics* 55:12.14.1–12.14.18, PMID: [27603020](https://doi.org/10.1002/cpbi.12), <https://doi.org/10.1002/cpbi.12>.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842, PMID: [20110278](https://doi.org/10.1093/bioinformatics/btq033), <https://doi.org/10.1093/bioinformatics/btq033>.
- Richmond RC, Simpkin AJ, Woodward G, Gaunt TR, Lyttleton O, McArdle WL, et al. 2015. Prenatal exposure to maternal smoking and offspring DNA methylation across the lifecourse: findings from the avon longitudinal study of parents and children (ALSPAC). *Hum Mol Genet* 24(8):2201–2217, PMID: [25552657](https://doi.org/10.1093/hmg/ddu739), <https://doi.org/10.1093/hmg/ddu739>.
- Rotroff DM, Joubert BR, Marvel SW, Häberg SE, Wu MC, Nilsen RM, et al. 2016. Maternal smoking impacts key biological pathways in newborns through epigenetic modification in utero. *BMC Genomics* 17(1):976, PMID: [27887572](https://doi.org/10.1186/s12864-016-3310-1), <https://doi.org/10.1186/s12864-016-3310-1>.
- Ruf N, Bähring S, Galetzka D, Pliushch G, Luft FC, Nürnberg P, et al. 2007. Sequence-based bioinformatic prediction and QUASEP identify genomic imprinting of the KCNK9 potassium channel gene in mouse and human. *Hum Mol Genet* 16(21):2591–2599, <https://doi.org/10.1093/hmg/ddm216>.
- Rzehak P, Saffery R, Reischl E, Covic M, Wahl S, Grote V, et al. 2016. Maternal smoking during pregnancy and DNA-methylation in children at age 5.5 years: epigenome-wide-analysis in the European Childhood Obesity Project (CHOP)-Study. *PLoS One* 11(5):e0155554, PMID: [27171005](https://doi.org/10.1371/journal.pone.0155554), <https://doi.org/10.1371/journal.pone.0155554>.
- Shah NR, Bracken MB. 2000. A systematic review and meta-analysis of prospective studies on the association between maternal cigarette smoking and preterm delivery. *Am J Obstet Gynecol* 182(2):465–472, PMID: [10694353](https://doi.org/10.1016/S0002-9378(00)70240-7), [https://doi.org/10.1016/S0002-9378\(00\)70240-7](https://doi.org/10.1016/S0002-9378(00)70240-7).
- Shields B, Hill A, Bilous M, Knight B, Hattersley AT, Bilous RW, et al. 2009. Cigarette smoking during pregnancy is associated with alterations in maternal and fetal thyroid function. *J Clin Endocrinol Metab* 94(2):570–574, PMID: [19017761](https://doi.org/10.1210/jc.2008-0380), <https://doi.org/10.1210/jc.2008-0380>.
- Song Q, Decato B, Hong EE, Zhou M, Fang F, Qu J, et al. 2013. A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One* 8(12):e81148, PMID: [24324667](https://doi.org/10.1371/journal.pone.0081148), <https://doi.org/10.1371/journal.pone.0081148>.
- Su D, Wang X, Campbell MR, Porter DK, Pittman GS, Bennett BD, et al. 2016. Distinct epigenetic effects of tobacco smoking in whole blood and among leukocyte subtypes. *PLoS One* 11(12):e0166486, PMID: [27935972](https://doi.org/10.1371/journal.pone.0166486), <https://doi.org/10.1371/journal.pone.0166486>.
- Wan M, Bennett BD, Pittman GS, Campbell MR, Reynolds LM, Porter DK, et al. 2018. Identification of smoking-associated differentially methylated regions using reduced representation bisulfite sequencing and cell type-specific enhancer activation and gene expression. *Environ Health Perspect* 126(4):047015, PMID: [29706059](https://doi.org/10.1289/EHP2395), <https://doi.org/10.1289/EHP2395>.
- Zhu J, Paul WE. 2008. CD4 T cells: fates, functions, and faults. *Blood* 112(5):1557–1569, PMID: [18725574](https://doi.org/10.1182/blood-2008-05-078154), <https://doi.org/10.1182/blood-2008-05-078154>.
- Zuin J, Dixon JR, van der Reijden MI, Ye Z, Kolovos P, Brouwer RW, et al. 2014. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A* 111(3):996–1001, PMID: [24335803](https://doi.org/10.1073/pnas.1317788111), <https://doi.org/10.1073/pnas.1317788111>.