Environ Health Perspect

DOI: 10.1289/EHP3398

Supplemental Material

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

Caitlin G. Howe, Meng Zhou, Xuting Wang, Gary S. Pittman, Isabel J. Thompson, Michelle R. Campbell, Theresa M. Bastain, Brendan H. Grubbs, Muhammad T. Salam, Cathrine Hoyo, Douglas A. Bell, Andrew D. Smith, and Carrie V. Breton

Table of Contents

Figure S1. Dot and boxplots for the 10 CpGs with the smallest p-values and absolute %methylation differences between 10-11%. The CpG position is shown on the x-axis and the %methylation level is shown on the y-axis. Horizontal lines within each boxplot indicate the median %methylation level for each CpG. Interquartile ranges are represented by the upper and lower boundaries of the boxplots. The vertical lines ("whiskers") at the top and bottom of the boxplots indicate the boundaries of 1.5 times the interquartile range. Points beyond these whiskers are outliers. Newborns who were exposed to maternal tobacco smoke *in utero* are indicated in blue, while unexposed newborns are indicated in pink.

Figure S2. Circos plot showing the 10,381 differentially methylated CpG sites and the 557 differentially methylated regions identified in cord blood $CD4^+$ samples from the Maternal and Child Health Study (n=10 exposed, n=10 unexposed to any maternal tobacco smoke during pregnancy) by chromosome and genomic location. The outermost ring is comprised of chromosome ideograms, which show the relative size of each chromosome, in megabases (MB), and its banding patterns (darker black and gray bands indicate heterochromatin, white bands indicates euchromatin, red bands indicate centromeres, and blue bands indicate stalks for acrocentric chromosomes). The middle ring shows the differentially methylated CpG sites and the innermost ring shows the differentially methylated regions. CpGs and regions that were hypermethylated in the maternal tobacco smoke exposed, compared with unexposed, group are shown in dark blue and dark red, respectively. CpGs and regions that were hypomethylated in the maternal tobacco smoke exposed, compared with unexposed, group are shown in light blue and pink, respectively. The height of each bar indicates the %methylation difference between groups.

Figure S3. Histograms showing distributions for the **A)** %methylation differences within differentially methylated regions (DMRs), **B)** base pair lengths of DMRs, and **C)** number of differentially methylated CpG sites (raw and false discovery rate-adjusted p<0.05) within DMRs.

Figure S4. (A) Proportion of differentially methylated regions (DMRs) that were hypermethylated (black) and hypomethylated (gray) in the maternal tobacco smoke exposed, compared with unexposed, group by genomic region, **(B)** corresponding enrichment tests (Fisher's exact test), comparing the number of DMRs overlapping each genomic region with a set of similar-sized regions randomly selected from the genome, and **(C)** median (range) %methylation differences by genomic region.

Figure S5. Dot and boxplots for the 33 CpGs that were identified as differentially methylated in MACHS **(A)**, which replicated (%methylation difference in the same direction and raw p-value<0.05) in the WakeMed SMKE EPIC array study of cord blood $CD4^+$ cells **(B)**. The name of the EPIC array CpG is shown on the x-axis, and the %methylation level is shown on the y-axis. Horizontal lines within each boxplot indicate the median %methylation value for each CpG. The interquartile range is represented by the upper and lower boundaries of the boxplot. The vertical lines ("whiskers") at the top and bottom of each boxplot indicate the boundaries of 1.5 times the interquartile range. Points beyond these whiskers are outliers. Newborns who were exposed to maternal tobacco smoke *in utero* are indicated in blue, while unexposed newborns are indicated in pink.

References

Additional File- Excel Document

Replication Look-Up Analyses

Replication analyses were conducted using data from two different studies of tobacco smoke exposure, which also profiled DNA methylation patterns in $CD4^+$ cells.

The first study, SMKE, profiled DNA methylation using Illumina's Infinium MethylationEPIC array in cord blood $CD4^+$ cells isolated from a subset of 30 newborns exposed (*n*=14) versus unexposed (*n*=16) to maternal tobacco smoke, who were recruited from the WakeMed hospital in Raleigh, North Carolina. Non-smoking mothers from this study were lifetime non-smokers. Cord blood mononuclear cells were isolated using Ficoll-Paque PLUS (Sigma-Aldrich), and $CD4^+$ antibody-coated magnetic beads (Invitrogen Dynabeads) were used to isolate $CD4^+$ T cells. DNA/RNA was extracted using the Qiagen All Prep DNA/RNA/miRNA kit (Qiagen, 80224), according to the manufacturer's instructions. Participant characteristics are shown in **Table S1**. DNA methylation was analyzed on Illumina's Infinium MethylationEPIC array as per manufacturer's instructions. Raw methylation image files were processed using the minfi package in R (Aryee et al. 2014). Background correction and dye-bias equalization was performed via the normal-exponential out-of-band (noob) correction method. The methylation level at each CpG was reported as the beta-value [β = intensity of the methylated allele (M) / (intensity of the unmethylated allele (U) + intensity of the methylated allele (M) + 100)]. Beta-values were then transformed to obtain the log ratio, defined as log[β/(1 – β)], or M. Robust linear regression was used to evaluate the association between DNA methylation (M) at each CpG and smoking status while adjusting for potential confounders, including gestational age, infant sex, mother's ethnicity (non-Hispanic black, non-Hispanic white, and Hispanic other), and sample batch. A total of 485 CpGs that were identified as differentially methylated (false discovery rate-adjusted $p<0.05$) in the Maternal and Child Health Study (MACHS) are represented on the EPIC array and could therefore be queried in the SMKE study.

The second set of replication analyses were conducted using results from a study of adult smokers (*n*=59) and lifetime non-smokers (*n*=72) who were recruited at the National Institute of Environmental Health Sciences Clinical Research Unit (NIEHS CRU), which has been described previously (Wan et al. 2018). DNA methylation was measured in $CD4^+$ cells for all participants using Illumina's Infinium HumanMethylation450 array, and was also measured in $CD4^+$ cells from a subset of participants using Illumina's EPIC array (*n*=9 smokers, *n*=11 non-smokers), and also in $CD4^+$ cells from a subset of female participants, using reduced representation bisulfite sequencing (RRBS) (*n*=9 smokers, *n*=10 non-smokers). The 450k and EPIC array data were processed and analyzed using the same methods as those described above for the SMKE study, and statistical models were similarly adjusted for age, sex, and ethnicity. RRBS was carried out as reported previously in Wan et al. (Wan et al. 2018). Briefly, DNA from $CD4^+$ cells was digested with MspI, bisulfite converted, made into RRBS libraries, and sequenced on the Illumina NextSeq platform. Bismark version 0.14.3 was used to align the reads to the hg19 assembly and methylation percentages were derived for each CpG, excluding any sites that had fewer than 10 reads or occurred at a single nucleotide polymorphism as reported previously (Su et al. 2016). RRBS DMRs were determined using the method of Wan et al. (Wan et al. 2018). Demographic characteristics for each set of NIEHS CRU participants are shown in **Tables S2-S4**.

CpG sites that were found to be differentially methylated ($p_{FDR}<0.05$) by maternal tobacco smoke exposure status in MACHS, which are represented on the 450K (*n*=399) and EPIC (*n*=485) arrays, were queried in the SMKE and NIEHS CRU EPIC array results using Practical Extraction and Reporting Language (PERL) script to match chromosomal positions. Raw *p*-values<0.05 were considered statistically significant for the replication study. Directions of

association were also compared. Similarly, DMRs that were represented in the NIEHS CRU RRBS results (n=9) were identified using chromosome position. Raw p-values<0.05 were considered statistically significant. Directions of effect were also compared.

Table S1. Demographic Characteristics of the WakeMed SMKE EPIC Study Participants

	Unexposed (*n*=16)	Smoke Exposed (*n*=14)
	Mean ± SD or *n* (%)	Mean ± SD or *n* (%)
Maternal Age, years	27.3 (5.2)	28.5 (4.8)
Gestational Age, days	274.6 (6.2)	274.6 (6.2)
Birth Weight, grams	3375.9 (281.8)	3113.9 (379.7)
Baby's Sex		
Female	8 (50.0)	5 (35.7)
Male	8 (50.0)	9 (64.3)
Ethnicity		
Non-Hispanic Black	8 (50.0)	5 (35.7)
Non-Hispanic White	8 (50.0)	8 (57.1)
Hispanic Other	0 (0.0)	1 (7.1)
Cord Blood Cotinine, ng/mL	2 (0.0)	104.3 (103.5)
Cigarettes Per Day	0 (0.0)	9.9 (5.6)
Years Smoked	0 (0.0)	12.3 (7.5)
Caesarean Section	8 (50.0)	10 (71.4)
Progesterone Use During Pregnancy	1 (6.3)	0 (0.0)
Any ETS Exposure[a]	2 (12.5)	12 (85.7)

[a] Environmental tobacco smoke (ETS) exposure is defined as any tobacco smoke exposure inside or outside the home.

Table S2. Demographic Characteristics of the NIEHS CRU EPIC Study Participants

	Non-Smokers (*n*=11)	Smokers (*n*=9)
	Mean ± SD or *n* (%)	Mean ± SD or *n* (%)
Age, years	42.1 (11.6)	40.7 (9.8)
Gender		
Female	6 (54.5)	5 (55.6)
Male	5 (45.5)	4 (44.4)
Ethnicity		
Non-Hispanic Black	6 (54.5)	5 (55.6)
Non-Hispanic White	5 (45.5)	4 (44.4)
Cotinine, ng/mL	2.0 (0.0)	178.5 (58.8)
Cigarettes Per Day	0 (0.0)	15.8 (10.0)
Pack-Years	0 (0.0)	20.2 (22.6)
Any ETS Exposure[a]	2 (18.2)	8 (88.9)

[a]Environmental tobacco smoke (ETS) exposure is defined as any tobacco smoke exposure inside or outside the home.

Abbreviations Used: NIEHS CRU, National Institute of Environmental Health Sciences Clinical Research Unit

Table S3. Demographic Characteristics of the NIEHS CRU 450K Study Participants

	Non-smokers (*n*=72)	Smokers (*n*=59)
	Mean ± SD or *n* (%)	Mean ± SD or *n* (%)
Age, years	37.6 (9.9)	42.0 ± 9.4
Gender		
Female	27 (45.8)	32 (54.2)
Male	45 (54.2)	27 (45.8)
Ethnicity		
Hispanic Black	2 (2.8)	0 (0.0)
Non-Hispanic Black	30 (41.7)	31 (52.5)
Hispanic White	0 (0.0)	1 (1.7)
Non-Hispanic White	34 (47.2)	23 (39.0)
Hispanic Other	0 (0.0)	1 (1.7)
Non-Hispanic Other	6 (8.3)	3 (5.1)
Serum Cotinine, ng/mL	2.3 (1.4)	231.2 (165.4)
Cigarettes per day	0 (0.0)	14.1 (7.7)
Pack-Years	0 (0.0)	17.3 (14.8)
Any ETS exposure[a]	3 (5.1)	42.0 (71.2)

[a]Environmental tobacco smoke (ETS) exposure is defined as any tobacco smoke exposure inside or outside the home.

Abbreviations Used: NIEHS CRU, National Institute of Environmental Health Sciences Clinical Research Unit

Table S4. Demographic Characteristics of the NIEHS CRU RRBS Study Participants

	Non-smokers (*n*=10)	Smokers (*n*=9)
	Mean ± SD or *n* (%)	Mean ± SD or *n* (%)
Age, years	46 ± 6	46 ± 8
Gender		
Female	10 (100)	9 (100.0)
Male	0 (0.0)	0 (0.0)
Ethnicity		
Black	5 (50.0)	5 (55.6)
White	5 (50.0)	4 (44.4)
Serum Cotinine, ng/mL	2 ± 0	217 ± 102
Cigarettes Per Day	0 (0.0)	22 ± 11
Pack-Years	0 (0.0)	34 ± 24

Abbreviations Used: NIEHS CRU, National Institute of Environmental Health Sciences Clinical Research Unit; RRBS, reduced representation bisulfite sequencing

Table S5. Enrichment of Hypomethylated and Hypermethylated DMRs Identified in MACHS in Regulatory Regions of Interest[a]

	Observed	Expected	*p*-value[b]
HYPERMETHYLATED			
All Enhancers	5	2	8.3×10^{-4}
T Cell Enhancers	0	0	>0.99
$CD4^+$ DNase Sensitive Regions (ENCODE)	17	3	0.01
$CD4^+$ Transcription Factor Binding Sites (ReMap)	3	13	0.02
$CD4^+$ CTCF Binding Sites	4	1	0.05
HYPOMETHYLATED			
All Enhancers	6	2	0.05
T Cell Enhancers	2	0	9.4×10^{-3}
$CD4^+$ DNase Sensitive Regions (ENCODE)	18	3	4.8×10^{-6}
$CD4^+$ Transcription Factor Binding Sites (ReMap)	3	21	7.4×10^{-11}
$CD4^+$ CTCF Binding Sites	7	1	1.7×10^{-4}

[a]Comparing maternal tobacco smoke-exposed to unexposed newborns

[b]*p*-value is from Fisher's Exact Test

Abbreviations Used: DMR, differentially methylated region; ENCODE, Encyclopedia of DNA Elements; Maternal and Child Health Study

Table S6. FANTOM5 Enhancers Overlapping DMRs Identified in MACHS[a] and Their Predicted Targets

Enhancer	%Methylation Difference	Tissues/Cells with Highest, Significant Overrepresentation of Enhancer	Genes	Distance	Correlation between Promoter and Enhancer Activity[b]
Chr1:173379915-173380349	-17.2	Natural Killer Cell, Lymphocyte from B Cell Lineage, Monocyte	*PRDX6, KLHL20*	65728/64, 304035	0.86/0.85, 0.79
Chr8:141108929-141109987*	-11.7	Brain, Olfactory Region, T Cell	*KCNK9*	394087	0.73
Chr9:36154326-36154976	-29.6	Basophil, Mast Cell, Monocytes	*GLIPR2, RNF38*	17650, 246318/454	0.64, 0.60/0.58
Chr10:134549634-134549842	-11.3	Spinal Cord, Brain	*KNDC1*	423706	0.77
Chr12:7781011-7781133	17.1	Testis	*CLEC4A*	495124/41	0.75/0.58
Chr16:1519780-1519909	10.5	Monocyte, Dendritic Cell	*BAIAP3, MAPK8IP3*	136156, 236394	0.58, 0.76
Chr16:85368480-85368758	8.7	Mast Cell	*GSE1*	278222/319419	0.71, 0.74
Chr17:154480-154575	24.1	Brain, Kidney	--	NA	NA
Chr17:14641682-14642082	8.1	Monocyte, Basophil, Amniotic Epithelial Cell	*HS3ST3B1*	437473	0.72
Chr17:76233617-76233864	-17.4	Spinal Cord, Brain	*TMEM235, CYTH1*	6029, 485879/907	0.83, 0.68/0.68
Chr21:44104688-44105340*	-15.0	Natural Killer Cell, T Cell	*ABCG1, NDUFV3*	465920, 194796	0.66, 0.92

[a]DMRs were identified by merging neighboring CpGs with a false discovery rate-adjusted $p<0.05$, identified by beta-binomial regression models, adjusted for maternal working status and infant sex, using the RADMeth program

[b]The FANTOM5 consortium predicts targets of enhancers by examining correlations between enhancers and all robust FANTOM5 promoter pairs within 500 kb and then filters these pairs for correlations with p-values$<1.0\times10^{-5}$ after adjusting for the false discovery rate.

*Active in T cells

Abbreviations Used: DMR, differentially methylated region; FANTOM5, functional annotation of the mammalian genome; MACHS, Maternal and Child Health Study

Table S7. Replication Results for Six CpGs Contained within One of the 20 DMRs with the Largest %Methylation Differences that Could be Queried in the WakeMed SMKE Study

Position	EPIC CpG	MACHS %Methylation Difference	Average Coverage	SMKE %Methylation Difference	SMKE P-Value
chr4:1607110	cg08089543	21.0	6.5	-0.3	0.76
chr4:1607290	cg27207756	32.9	2.4	2.2	0.55
chr5:177209284	cg26673648	-47.0	2.0	-0.6	0.82
chr9:36154749	cg19097407	-34.7	7.6	8.7	0.44
chr10:123100180	cg07044115	42.4	7.5	-1.9	0.52
chr17:27359874	cg03597174	18.3	5.1	-3.1	0.03

Abbreviations Used: DMR, differentially methylated region; MACHS, Maternal and Child Health Study

Table S8. DMRs identified in MACHS[a] that overlap DMRs identified by Bauer et al. 2016

Bauer DMR	%Methylation Difference	#CpGs	MACHS DMR	%Methylation Difference	#CpGs[b]	Genomic Region	Nearest Gene	Distance
chr1:149293953-149294204	11.5	11	chr1:149293953-149293970	12.2	3	3'End	*TRNA_Val*	0
chr14:93153235-93153358	-32.9	9	chr14:93153343-93153502	14.2	5	Intron	*RIN3*	0
chr20:58713541-58714104	12.6	28	chr20:58713718-58713922	-15.0	8	Promoter	*MIR646HG*	0
chr7:5183707-5184308	13.7	27	chr7:5183954-5184155	-15.9	7	Promoter	*ZNF890P*	0
chr8:41593896-41594261	13.7	16	chr8:41593896-41594262	-15.1	8	Intron	*ANK1, NKX6-3*	0
chr9:72026853-72027406	10.8	29	chr9:72027018-72027281	17.8	8	Intergenic	*APBA1*	15,167

[a]DMRs were identified by merging neighboring CpGs with a false discovery rate-adjusted $p<0.05$, identified by beta-binomial regression models, adjusted for maternal working status and infant sex, using the MethPipe software

[b]Number of CpGs within the DMR that had both a raw and false discovery rate-adjusted $p<0.05$

Abbreviations Used: DM, differentially methylated; DMR, differentially methylated region; MACHS, Maternal and Child Health Study

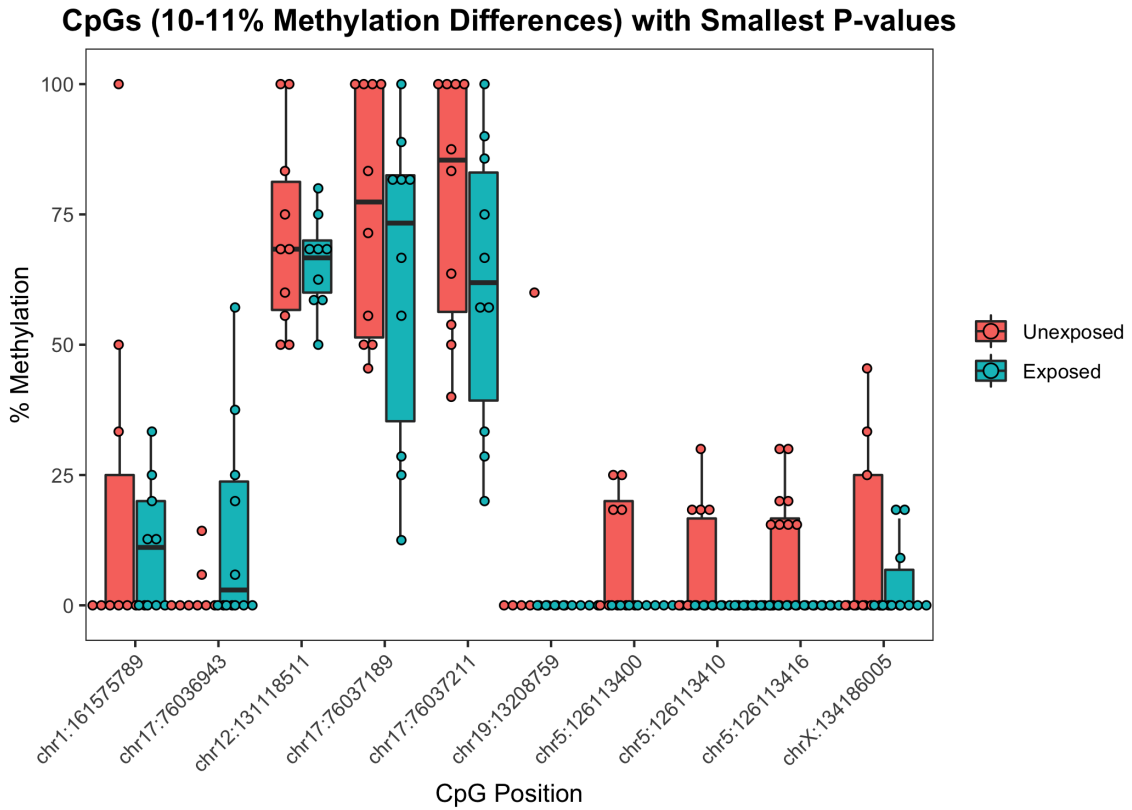


Figure S1. Dot and boxplots for the 10 CpGs with the smallest p-values and absolute %methylation differences between 10-11%. The CpG position is shown on the x-axis and the %methylation level is shown on the y-axis. Horizontal lines within each boxplot indicate the median %methylation level for each CpG. Interquartile ranges are represented by the upper and lower boundaries of the boxplots. The vertical lines ("whiskers") at the top and bottom of the boxplots indicate the boundaries of 1.5 times the interquartile range. Points beyond these whiskers are outliers. Newborns who were exposed to maternal tobacco smoke *in utero* are indicated in blue, while unexposed newborns are indicated in pink.

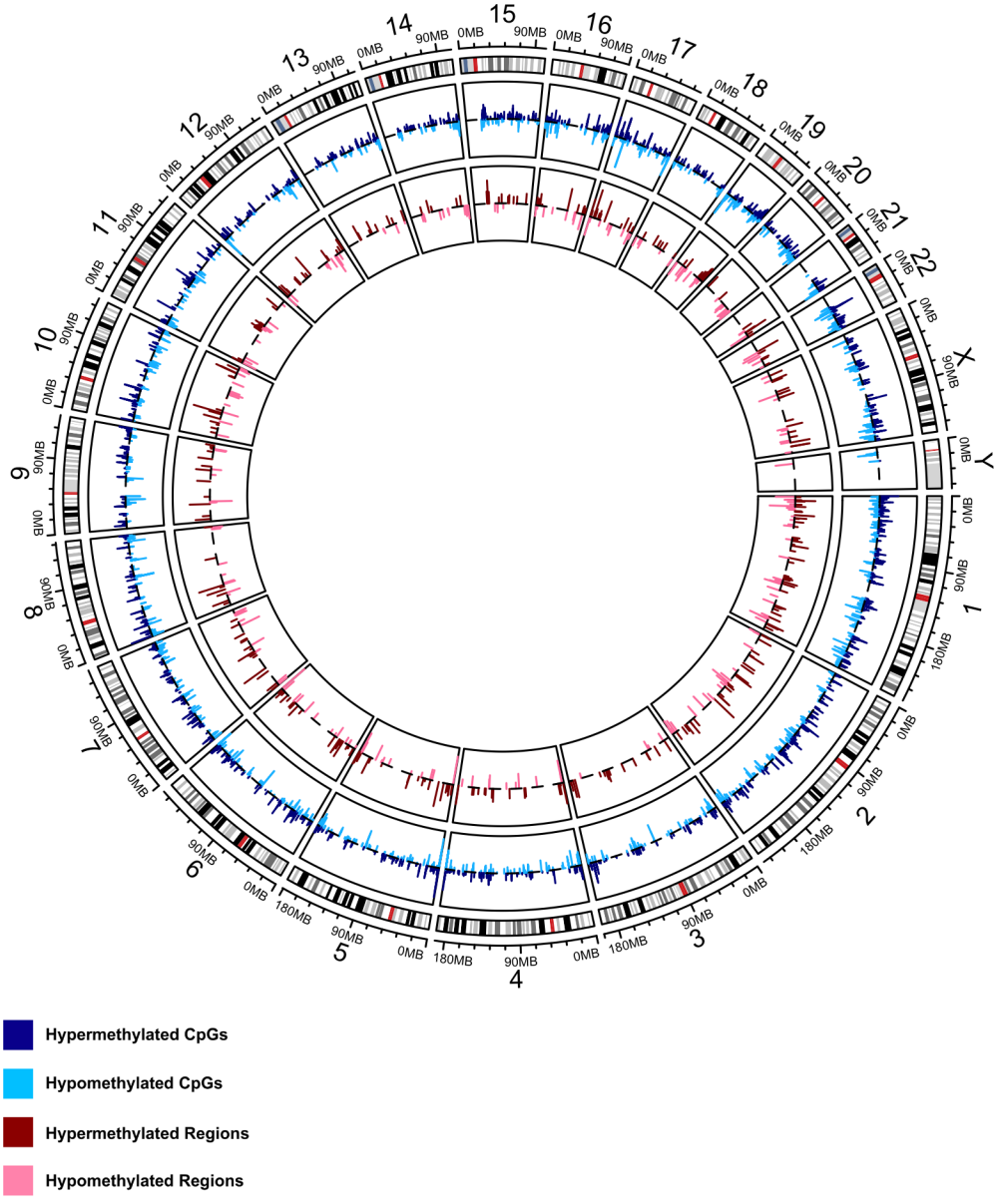


Figure S2. Circos plot showing the 10,381 differentially methylated CpG sites and the 557 differentially methylated regions identified in cord blood $CD4^+$ samples from the Maternal and Child Health Study (*n*=10 exposed, *n*=10 unexposed to any maternal tobacco smoke during pregnancy) by chromosome and genomic location. The outermost ring is comprised of chromosome ideograms, which show the relative size of each chromosome, in megabases (MB), and its banding patterns (darker black and gray bands indicate heterochromatin, white bands indicates euchromatin, red bands indicate centromeres, and blue bands indicate stalks for acrocentric chromosomes). The middle ring shows the differentially methylated CpG sites and the innermost ring shows the differentially methylated regions. CpGs and regions that were hypermethylated in the maternal tobacco smoke exposed, compared with unexposed, group are shown in dark blue and dark red, respectively. CpGs and regions that were hypomethylated in the maternal tobacco smoke exposed, compared with unexposed, group are shown in light

blue and pink, respectively. The height of each bar indicates the %methylation difference between groups.

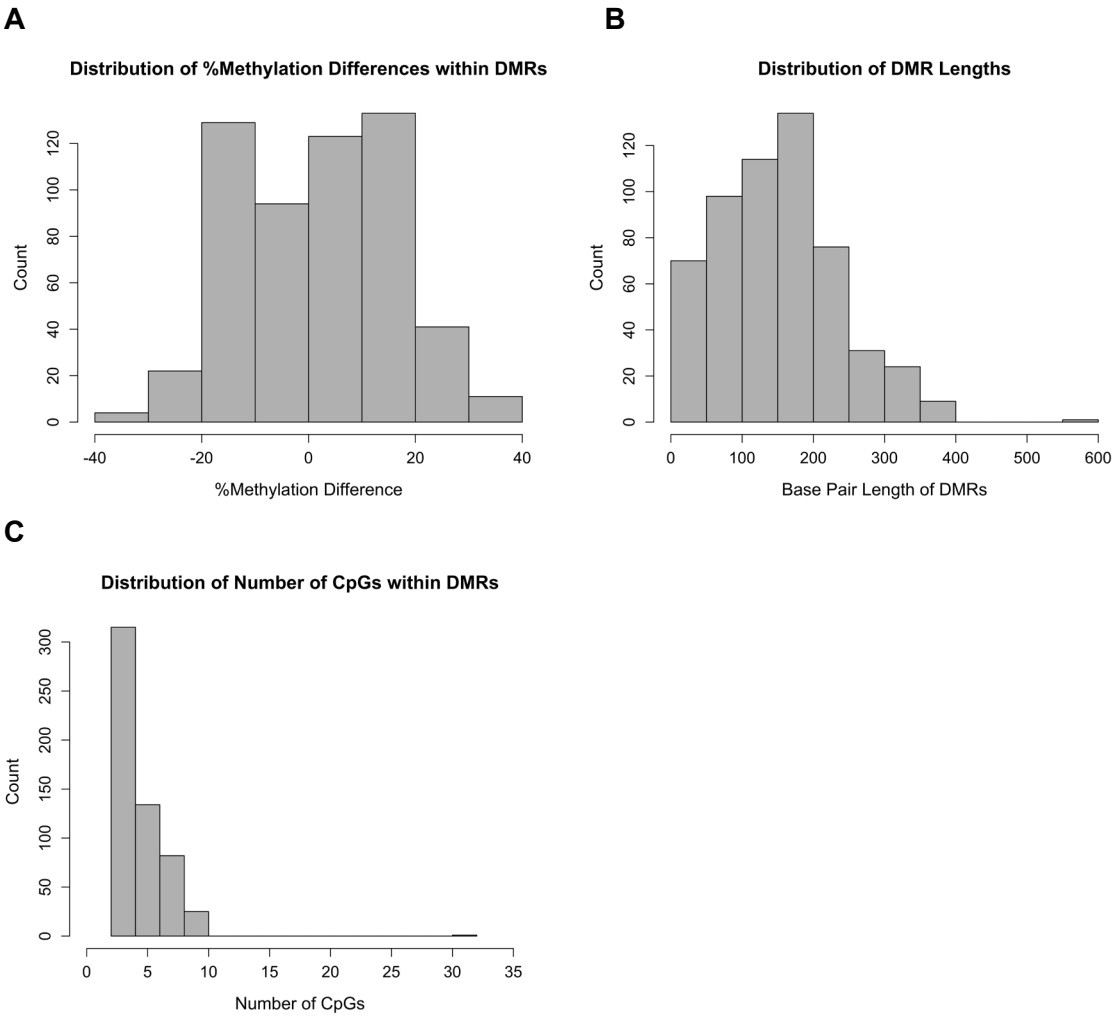
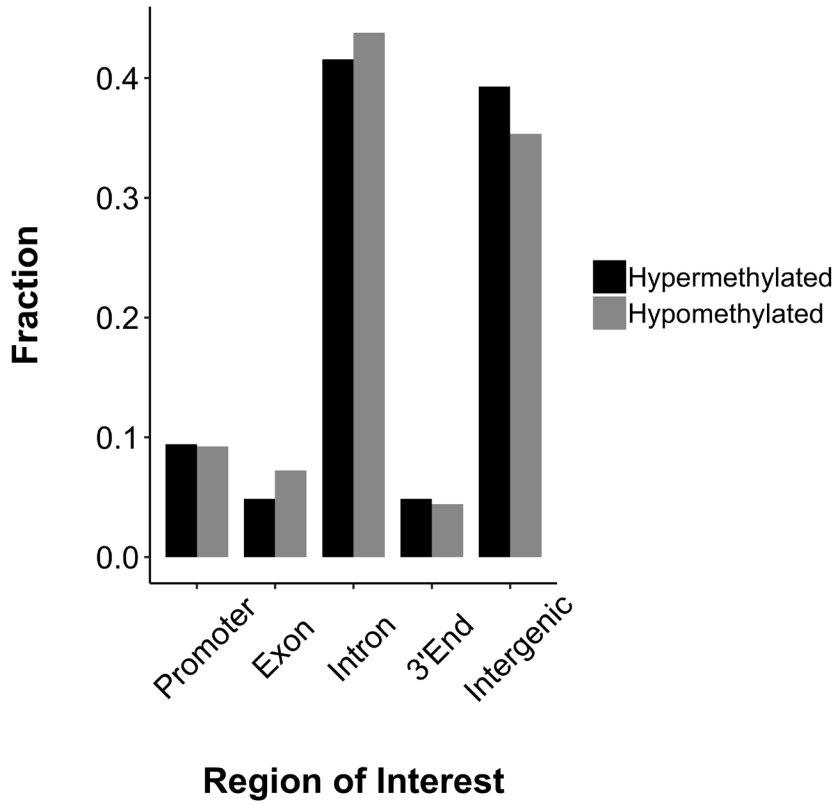


Figure S3. Histograms showing distributions for the **A)** %methylation differences within differentially methylated regions (DMRs), **B)** base pair lengths of DMRs, and **C)** number of differentially methylated CpG sites (raw and false discovery rate-adjusted $p<0.05$) within DMRs

A



B

	Obs	Exp	*p*-value
HYPERMETHYLATED			
DMRs			
Promoter	13	5	1.5×10^{-3}
Exon	8	5	0.15
Intron	74	77	0.65
3' End	8	5	0.15
Intergenic	82	93	0.10
HYPOMETHYLATED			
DMRs			
Promoter	13	4	2.6×10^{-4}
Exon	12	4	7.0×10^{-4}
Intron	65	63	0.81
3' End	9	4	0.02
Intergenic	56	80	1.4×10^{-4}

C

	HYPERMETHYLATED	HYPOMETHYLATED
DMRs		
Promoter	18.1 (10.6, 31.3)	-14.4 (-29.4, -10.5)
Exon	16.3 (12.6, 21.1)	-12.7 (-26.0, -10.1)
Intron	14.3 (10.0, 39.2)	-15.1 (-30.5, -10.2)
3'End	7.3 (12.2, 35.3)	-14.4 (-26.4, -10.1)
Intergenic	15.5 (10.0, 39.7)	-15.4 (-39.8, -10.2)

Figure S4. (A) Proportion of differentially methylated regions (DMRs) that were hypermethylated (black) and hypomethylated (gray) in the maternal tobacco smoke exposed, compared with unexposed, group by genomic region, **(B)** corresponding enrichment tests (Fisher's exact test), comparing the number of DMRs overlapping each genomic region with a set of similar-sized regions randomly selected from the genome, and **(C)** median (range) %methylation differences by genomic region

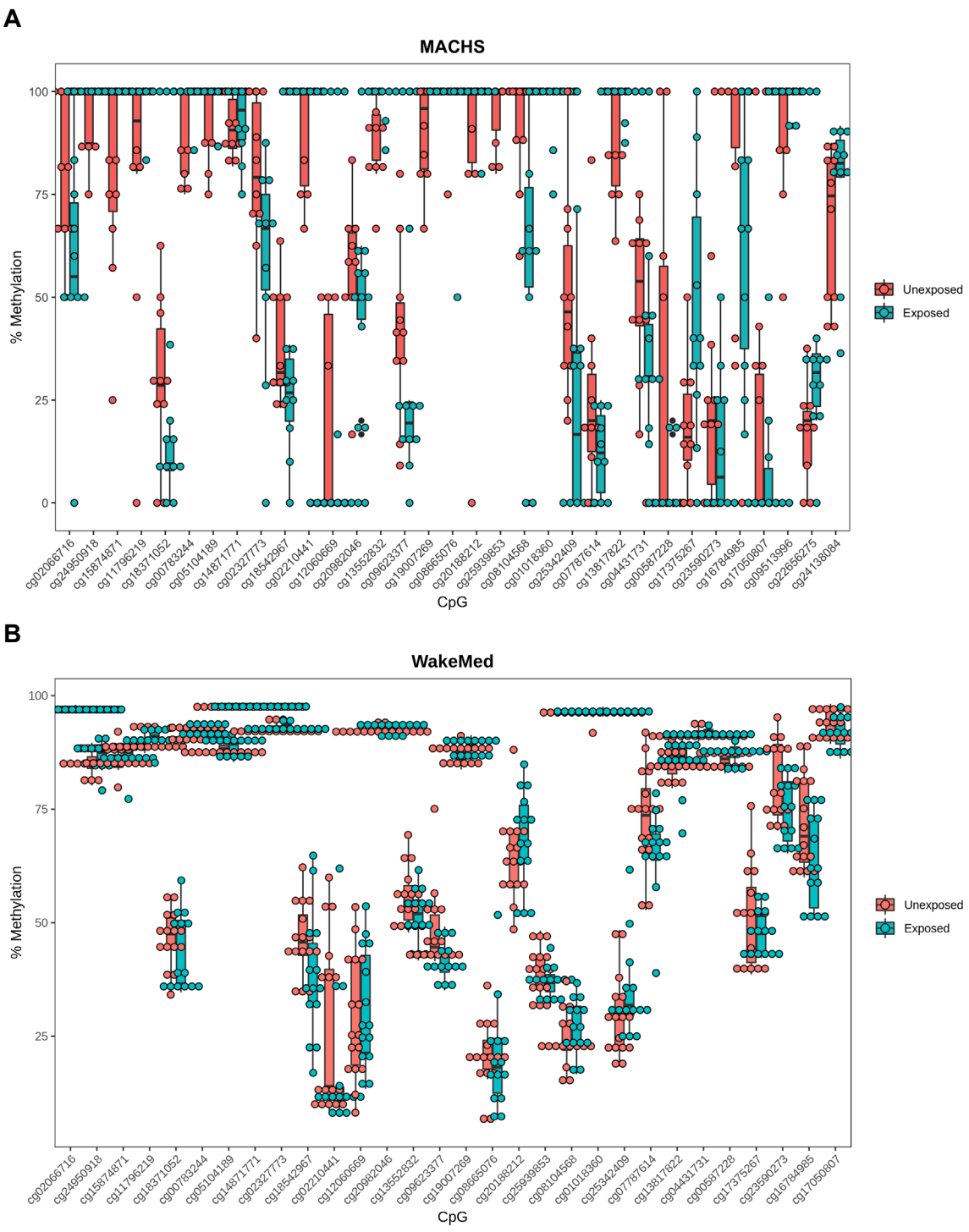


Figure S5. Dot and boxplots for the 33 CpGs that were identified as differentially methylated in MACHS **(A)**, which replicated (%methylation difference in the same direction and raw p-value<0.05) in the WakeMed SMKE EPIC array study of cord blood $CD4^+$ cells **(B)**. The name of the EPIC array CpG is shown on the x-axis, and the %methylation level is shown on the y-axis. Horizontal lines within each boxplot indicate the median %methylation value for each CpG. The interquartile range is represented by the upper and lower boundaries of the boxplot. The vertical lines ("whiskers") at the top

and bottom of each boxplot indicate the boundaries of 1.5 times the interquartile range. Points beyond these whiskers are outliers. Newborns who were exposed to maternal tobacco smoke *in utero* are indicated in blue, while unexposed newborns are indicated in pink.

References:

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. 2014. Minfi: A flexible and comprehensive bioconductor package for the analysis of infinium DNA methylation microarrays. Bioinformatics 30:1363-1369.

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. Molecular systems biology 12:861.

Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. 2013. Champ: 450k chip analysis methylation pipeline. Bioinformatics 30:428-430.

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:e0166486.

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. Environmental Health Perspectives (Online) 126.