Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation

Carrie V. Breton, ScD[1], Hyang-Min Byun, PhD[2], Made Wenten, MS[1], Fei Pan, PhD[3], Allen Yang, MD, PhD[2,3], and Frank D. Gilliland, MD, PhD[1]

Online Data Supplement

Methods

Study Population

For the purposes of this study, 348 children were used for global methylation and 272 for the Illumina screen, between which there were approximately 180 in common. The main reason all 272 were not a subset of the 348 used for global methylation has to do with availability of adequate sample material for all assays, particularly because the Illumina platform required greater quantities of DNA.

Smoke Exposure

Assessment of prenatal tobacco smoke exposure and second hand smoke (SHS) exposure was based on parent/guardian written responses on a self-administered questionnaire. Prenatal smoke exposure was defined as an affirmative answer to the following question: "Did your child's biologic mother smoke while she was pregnant with your child (include time when she was pregnant but did not yet know that she was)?" SHS exposure was defined by the child's current exposure to smokers in the home, including regular visitors.

GSTP1 Haplotype creation

The LD patterns for GSTP1 in the CHS have previously been described (1). Briefly, two haplotype blocks with substantial interblock LD were defined based on the method using the method by Gabriel et al (2). The SNPs were then chosen using the TagSNPs program (download from http://www-rcf.usc.edu/~stram/tagSNPs.html).

We selected two SNPs for each of the two blocks (rs6591255 and rs4147581 for the first block and rs1695 and rs749174 for the second block). All four SNPs accounted

for >80% of the haplotypic variation in the GSTP1 locus without considering block structure. Haplotype frequencies for each ethnic group (non-Hispanic white and Hispanic white) were estimated separately. We coded the haplotype as the estimated number of copies of each haplotype a person carries.

DNA Methylation

Global methylation analysis of LINE1 and AluYb8 were performed by bisulphite-polymerase chain reaction (PCR) Pyrosequencing assay, using the HotMaster Mix (Eppendorf, Hamburg, Germany) consisting of HotMaster DNA polymerase and HotMaster Taq Buffer (45 mM KCl, 2.5 mM Mg^{2+}, 200 µM dNTP) (3-5). Since biotin-labeled reverse primers were used for PCR amplification, biotin-labeled PCR products can be purified using sepharose beads and made single-stranded to act as a template in a Pyrosequencing reactions as recommended by the manufacturer using the Pyrosequencing Vacuum Prep Tool (Pyrosequencing, Inc., Westborough, MA). Briefly, the PCR product was bound to Streptavidin Sepharose HP (Amersham Biosciences, Uppsala, Sweden) and the Sepharose beads were purified, washed, and denatured using a 0.2 M NaOH solution. Bisulfite treatment of DNA leads to the formation of a C (methylated) or T (unmethylated) single nucleotide polymorphism that can be quantitated by Pyrosequencing. The percentage of methylation was calculated using the ratio of C (methylated) / C+T (methylated+unmethylated).

For CpG gene-specific methylation analysis, the Illumina GoldenGate Cancer methylation panel I was used (Illumina, San Diego, CA). This panel spans 1,505 CpG loci selected from 807 genes. Many of the genes selected for the Illumina panel are involved in pathways potentially pertinent to the detoxification of tobacco smoke,

including DNA repair genes, genes regulated by various signaling pathways and/or responsible for altered cell growth, differentiation, and apoptosis, and previously reported differentially methylated genes. Other cancer-related genes included may be less relevant to our immediate hypotheses. Use of a larger and/or more targeted custom panel would have allowed us to potentially capture more targets but was cost prohibitive.

Normalization of Illumina data

A background normalization approach was employed to minimize the amount of variation in background signals between arrays. The signals of built-in negative controls were used to estimate the expected signal level in the absence of hybridization to a specific target. The average signal of the negative controls was subtracted from the Cy3 and Cy5 probe signal intensities. Outliers were removed using the median absolute deviation method. Detection p-values were also computed from the background model characterizing the chance that the target sequence signal was distinguishable from the negative controls. Loci with a detection p-value greater than 0.05 were eliminated and the corresponding beta values were replaced as missing values.

In order to minimize effects arising from scanner-to-scanner variation, the timed mean normalization approach was used to normalize beta values across plates (6). Briefly, the beta value of each sample was ranked. The average of beta values of all samples was calculated after excluding the 2.5% top and bottom ranked probes. The beta values for all samples were then linearly scaled by multiplying the ratio of the sample mean to the overall mean. Samples with more than 50% of loci having missing values were treated as having failed, resulting in four samples in run 1 and two samples in run 2 that were eliminated. In total, 1031 CpG loci remained for analysis after all quality

control steps were applied.

References

1. Li YF, Gauderman WJ, Conti DV, Lin PC, Avol E, Gilliland FD. Glutathione s-transferase p1, maternal smoking, and asthma in children: A haplotype-based analysis. *Environ Health Perspect* 2008;116:409-415.

2. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225-2229.

3. Choi SH, Byun HM, Kwan JM, Issa JP, Yang AS. Hydroxycarbamide in combination with azacitidine or decitabine is antagonistic on DNA methylation inhibition. *Br J Haematol* 2007;138:616-623.

4. Kawamoto K, Enokida H, Gotanda T, Kubo H, Nishiyama K, Kawahara M, Nakagawa M. P16ink4a and p14arf methylation as a potential biomarker for human bladder cancer. *Biochem Biophys Res Commun* 2006;339:790-796.

5. Yang AS, Doshi KD, Choi SW, Mason JB, Mannari RK, Gharybian V, Luna R, Rashid A, Shen L, Estecio MR, et al. DNA methylation changes after 5-aza-2'-deoxycytidine therapy in patients with leukemia. *Cancer Res* 2006;66:5495-5503.

6. Kroll TC, Wolfl S. Ranking: A closer look on globalisation methods for normalisation of gene expression arrays. *Nucleic Acids Res* 2002;30:e50.

Supplement Tables

Table E1. Haplotype frequencies of 4 GSTP1 SNPs for CHS populations used for CpG-site specific methylation

Haplotype*	All	Non-Hispanic white	Hispanic White
0100	0.40	0.48	0.31
1011	0.25	0.29	0.20
0010	0.14	0.01	0.30
1000	0.09	0.11	0.06
0000	0.08	0.08	0.08
other	0.04	0.03	0.05

*1: minor allele and 0: common allele, with SNPs in the following order: rs6591255, rs4147581, rs1695, and rs749174.

Table E2. Details for Bisulfite PCR Pyrosequencing reaction and primers

Primer Name	Forward primer	Reverse primer (biotin labeled)	Sequencing primer	Anneaning Temperature
LINE1	TTGAGTTAGGTGTGGGATATAG	AAACATTACCTCACCTAAAAAAC	GGGTGGGAGTGATT	61℃
Alu Yb8	AGATTATTTTGGTTAATAAG	AACTACYAACTACAATAAC	AATAACTAAAATTACAAA C	58℃
AXL_P223_F	GTTTGAGTGTGTTTGTGGGTTAGTA	CAAACCTCCTTAACCCTTCATT	TAGTATGTTTTTGTT	touchdown*
KLK11_P103_F	GTGGAAAGAATGTTA	AAACTTTTCAAAAATCCACAAAAAA	GTGGAAAGAATGTTA	56°C
MET_E333_F	ATTGTTGTTTTAGGGGGTAAAATTA	TCCCAACTTTAAAACTTACTAAACC	GGAAATTGAAGAGA	55.7°C
NBL1_P24_F	TTTTTTAATTTGGAGTTGGAAGAAT	CTCCCTAATATACTCCAACCTACTC	GGGTAGAGGGAAGGG	Failed
PTPRO_P371_F	AGTGGGGAAGGAGAGAAAATATTT	AAATTTCCCTAACCTCTCTTAAACAA	GGAATTGGGATTTGG	61.7°C
SNCG_E119_F	ATGTTTTTAAGAAGGGTTTTTTTAT	CCCCTCCTTAATCTTCTCAACTACTT	TTAAGTAGGGGGTGA	57°C
SPDEF_P6_R	TTTTTAGTTTAGGGTTGTTTGTTGG	AAACCCTAACCAACTCTTCATCTC	AGGAAGTTAGATAGT	63°C
TGFB3_E58_R	GATTGAGGTTTGGTAAGAAGGTGTAA	CTAAAAATCAAAACCCAACAAAACT	GGTAATTTTGAGGA	59.8°C

*The annealing temperature decreased in increments for every subsequent set of cycles.

Table E3. The effect of prenatal exposure to maternal smoking on methylation of CpG in promoters for nine genes, by Illumina run and pooled.

	Run 1 (N=182)			Run 2 (N=183)			Pooled (N=272)		
Gene*	β†	p-value	Wilcoxon p-value	β†	p-value	Wilcoxon p-value	β†	p-value	Wilcoxon p-value
AXL	1.02	0.002	0.0005	0.57	0.001	0.0003	0.58	0.0001	0.0002
PTPRO	0.37	0.40	0.65	0.47	0.01	0.004	0.37	0.01	0.02
KLK11	0.34	0.28	0.04	0.31	0.09	0.15	0.22	0.14	0.09
TGFB3	0.46	0.14	0.005	0.32	0.05	0.02	0.27	0.04	0.01
MET	0.59	0.07	0.01	0.65	0.01	0.002	0.39	0.01	0.001
SPDEF	0.58	0.10	0.33	0.41	0.05	0.03	0.40	0.01	0.02
SNCG	0.69	0.03	<0.0001	0.37	0.05	0.01	0.43	0.003	0.0003
HOXA5	0.12	0.71	0.004	0.17	0.30	0.006	0.12	0.34	0.02
NBL1	0.79	0.01	0.004	0.37	0.08	0.02	0.36	0.02	0.004

*CpG loci interrogated by Illumina were located at the following positions relative to the transcription start site: AXL(-223), PTPRO(-371), KLK11(-103), TGFB3(+58), MET(+333), SPDEF(-6), SNCG(+119), NBL1(-24) and HOXA5 (+187)

† Z-score transformed estimate is interpreted as a change in SD units. All models were adjusted for gender, race, Hispanic ethnicity and Illumina plate.