

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 ehp508@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.

Supplemental Material

DNA Methylation Score as a Biomarker in Newborns for Sustained Maternal Smoking during Pregnancy

Sarah E. Reese, Shanshan Zhao, Michael C. Wu, Bonnie R. Joubert, Christine L. Parr, Siri E. Håberg, Per Magne Ueland, Roy M. Nilsen, Øivind Midttun, Stein Emil Vollset, Shyamal D. Peddada, Wenche Nystad, and Stephanie J. London

Table of Contents

Figure S1: Density distribution of the smoking methylation score for the training set (A; N=1,057) and test set (B; N=221). Black solid line - all individuals; Blue dashed line – combined sustained smoking non-smokers; Red dotted line - combined sustained smoking smokers; Grey vertical dashed line indicates the score threshold.

Figure S2: Plot showing the combined sustained smoking model ROC curves for the training (red solid line) and test (black dotted line) datasets. Cutoff score of -0.37 developed in the training set is shown on both curves, along with the corresponding specificity and sensitivity.

Table S1: 28 CpGs selected by the logistic LASSO regression model using combined sustained smoking.

Table S2: Cross-tabulation of cotinine-based and self-reported sustained smoking

Table S3: 24 CpGs selected by the logistic LASSO regression model using cotinine-based sustained smoking.

Table S4: 12 CpGs selected by the logistic LASSO regression model using self-reported sustained smoking.

Table S5: 26 CpGs selected by the logistic LASSO regression model using combined sustained smoking restricted to CpGs that overlap with the IlluminaEPIC BeadChip.

Table S6: Logistic LASSO results for main model (combined sustained smoking), the main model using log ratios, and the main model using only the CpGs available on the IlluminaEPIC array– The number of CpGs (q) used to calculate the smoking methylation score, area under the curve (AUC) and 95% confidence interval (CI), smoking methylation score threshold, percent accuracy and CI, sensitivity and CI, specificity and CI, and the number and percentage of false negatives (FN) and false positives (FP).

Table S7: Descriptive statistics of smoking variables in the training data set.

Table S8: Descriptive statistics of birth weight and maternal and newborn characteristics in the training data set.

Table S9: Mean Methylation Values of the 28 CpGs selected by the main model (combined sustained smoking).

Table S10: 37 CpGs selected by the logistic LASSO regression model on the log ratios using combined sustained smoking.