

**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 [ehpsubmissions@niehs.nih.gov](mailto:ehpsubmissions@niehs.nih.gov). Our staff will work with you to assess and meet your accessibility needs within 3 working days.

### **Supplemental Material**

#### **Development and Validation of a Novel Placental DNA Methylation Biomarker of Maternal Smoking during Pregnancy in the ECHO Program**

Lyndsey E. Shorey-Kendrick, Brett Davis, Lina Gao, Byung Park, Annette Vu, Cynthia D. Morris, Carrie V. Breton, Rebecca Fry, Erika Garcia, Rebecca J. Schmidt, T. Michael O'Shea, Robert S. Tepper, Cindy T. McEvoy, and Eliot R. Spindel, on behalf of program collaborators for Environmental influences on Child Health Outcomes

#### **Table of Contents**

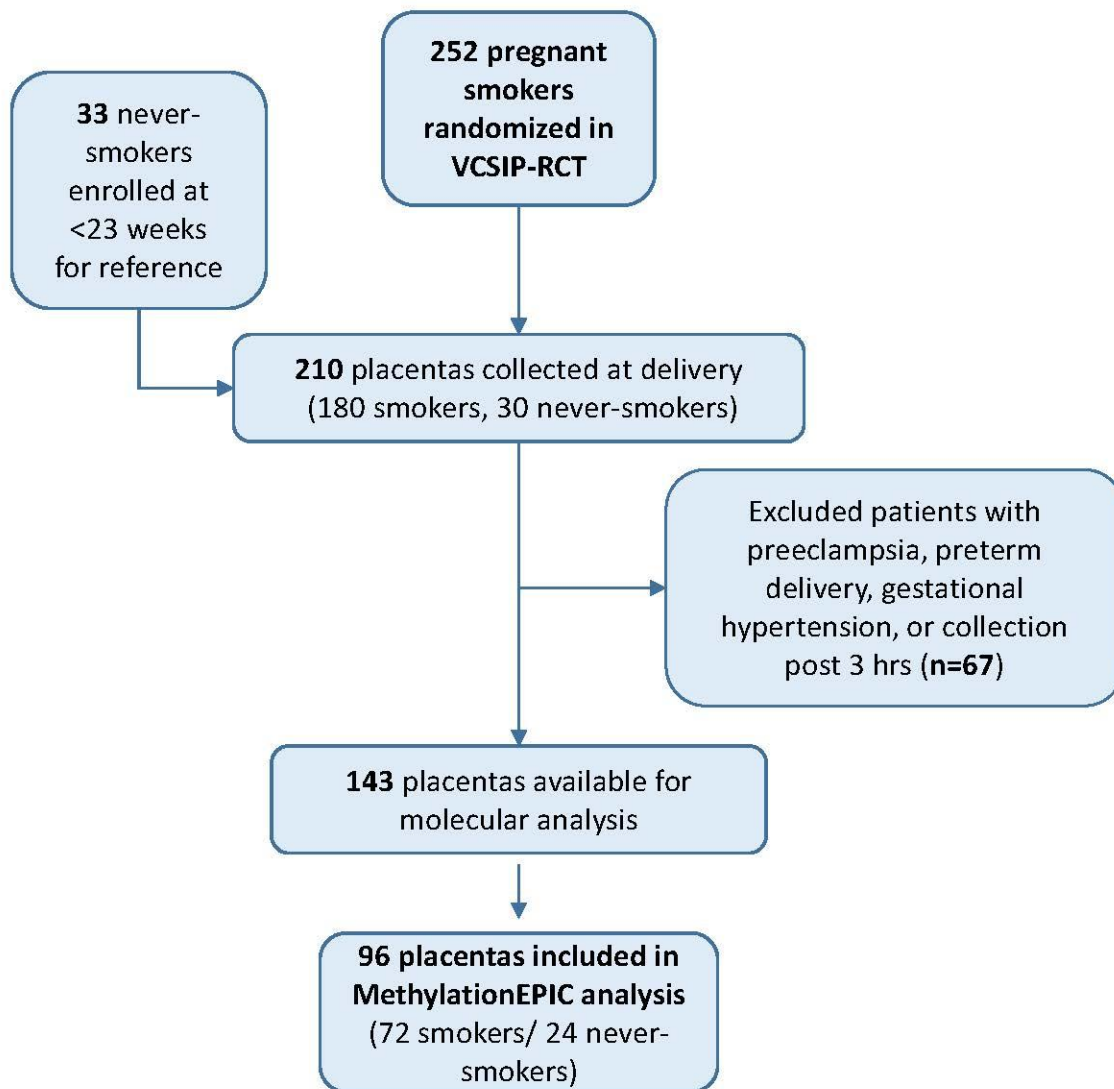
**Figure S1.** Consort diagram of samples used in the VCSIP training dataset.

**Figure S2.** Important CpGs for each model (A-C) ranked by importance and overlap across models (D).

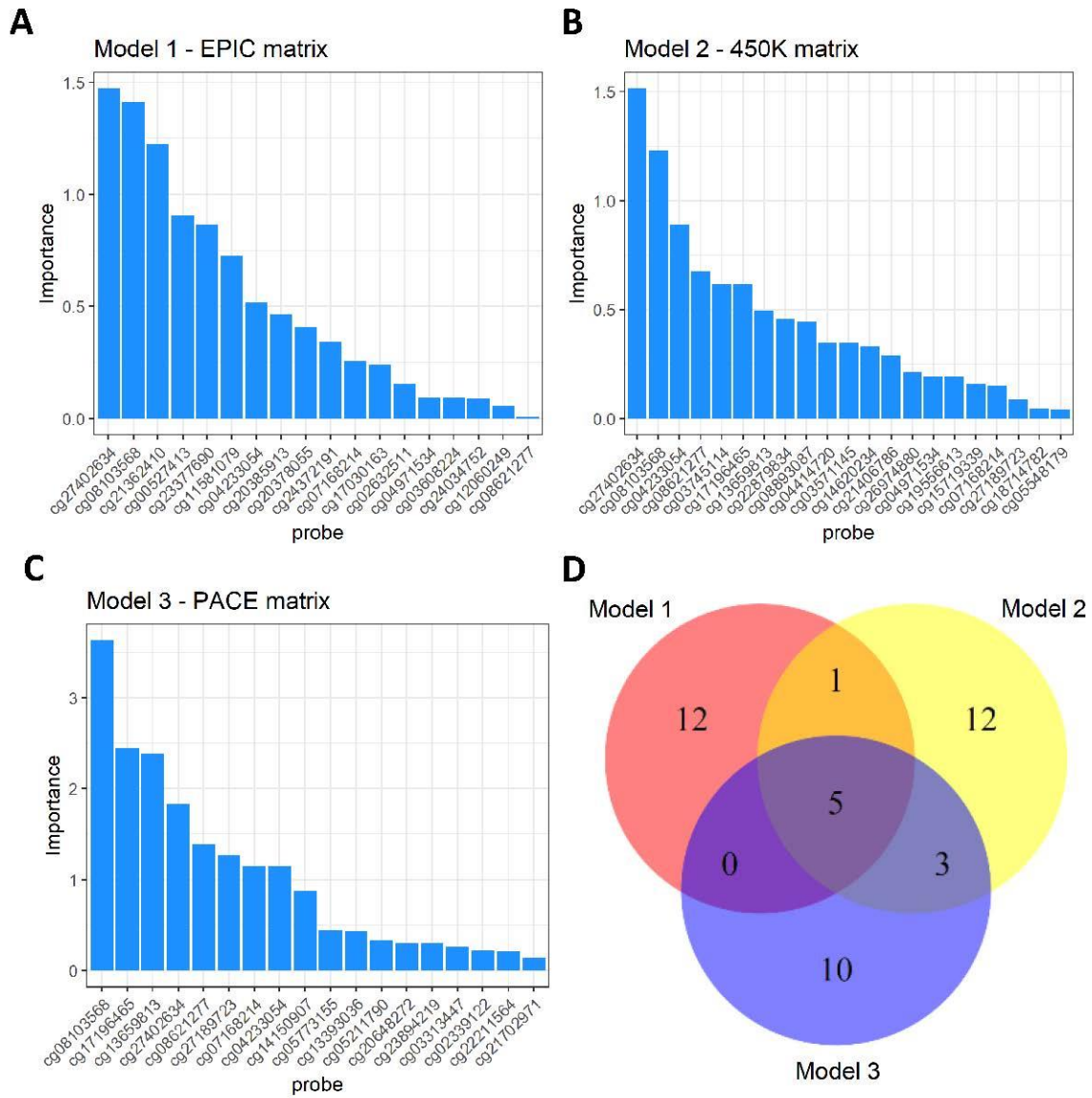
**Figure S3.** Boxplot distributions of PSI scores for each model in each dataset. Boxes indicate the interquartile range (IQR; 25<sup>th</sup> to 75<sup>th</sup> percentiles) of the data with horizontal lines across the median. Error bars indicate 1.5 \* IQR and points indicate outliers beyond that range. Nonsmokers consistently had lower PSI scores than smokers, and this is most apparent in the VCSIP training data, as expected.

**Figure S4.** EWAS Atlas enriched traits, GO, and KEGG terms in CpGs included in PSI models. The  $-\log_{10}$  PValue is presented on the x-axis and the enriched categories are on the y-axis. The size of the dots indicates the number of loci in each category. Facets separate different datasets used for enrichment analysis.

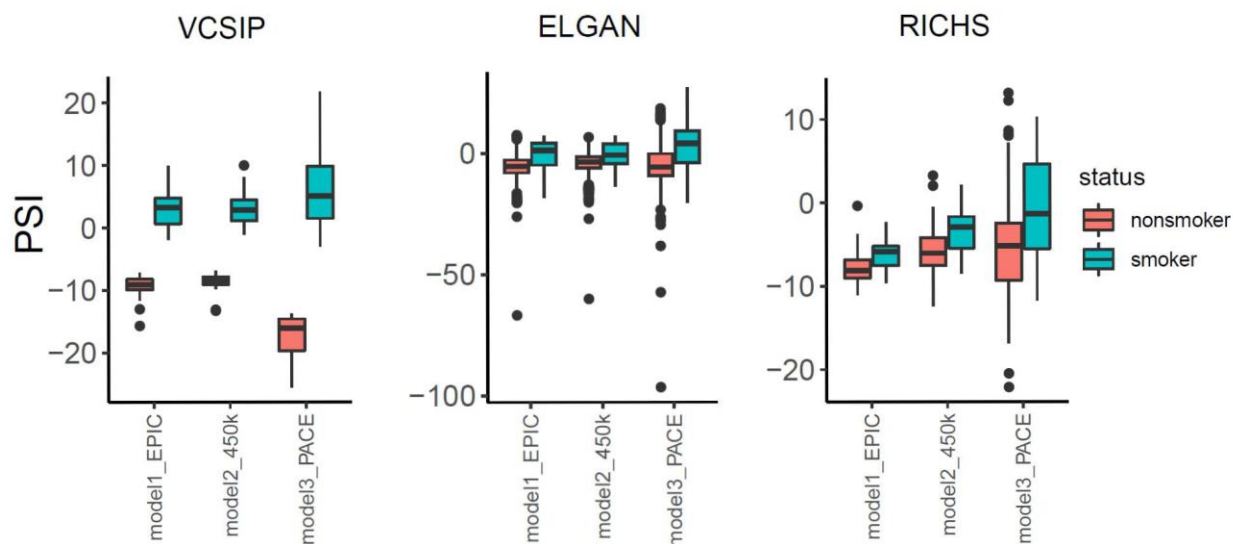
**Additional File-** Excel Document



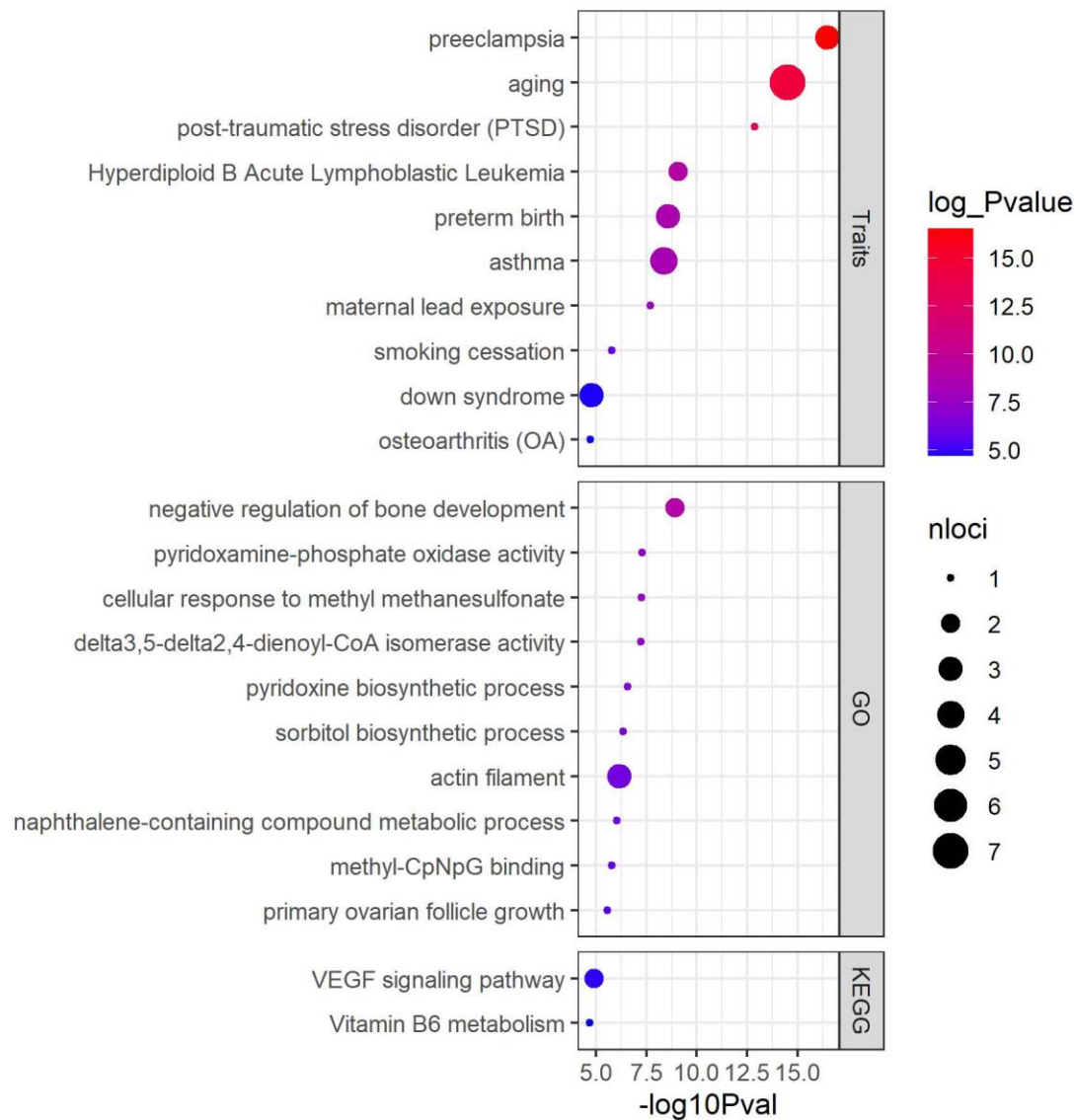
**Figure S1.** Consort diagram of samples used in the VCSIP training dataset. Note: VCSIP-RCT, Vitamin C to Decrease the Effects of Smoking in Pregnancy on Infant Lung Function – Randomized Clinical Trial.



**Figure S2.** Important CpGs for each model (A-C) ranked by importance and overlap across models (D). Note: EPIC, Infinium MethylationEPIC array; PACE, Pregnancy And Childhood Epigenetics (Importance = Absolute value of coefficient in **Supplemental Excel File Table 2**).



**Figure S3.** Boxplot distributions of PSI scores for each model in each dataset. Boxes indicate the interquartile range (IQR; 25<sup>th</sup> to 75<sup>th</sup> percentiles) of the data with horizontal lines across the median. Error bars indicate 1.5 \* IQR and points indicate outliers beyond that range. Nonsmokers consistently had lower PSI scores than smokers, and this is most apparent in the VCSIP training data, as expected. Note: ELGAN, Extremely Low Gestational Age Newborn; EPIC, Infinium MethylationEPIC array; PACE, Pregnancy And Childhood Epigenetics; PSI, placental smoking index; RICHs, Extremely Low Gestational Age Newborn ; VCSIP, Vitamin C to Decrease the Effects of Smoking in Pregnancy on Infant Lung Function (**Supplemental Excel File Table 3**).



**Figure S4.** EWAS Atlas enriched traits, GO, and KEGG terms in CpGs included in PSI models. The  $-\log_{10}$  PValue is presented on the x-axis and the enriched categories are on the y-axis. The size of the dots indicates the number of loci in each category. Facets separate different datasets used for enrichment analysis. Note: CpGs, cytosine-guanine dinucleotides; EWAS, epigenome-wide association study; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; VEGF, vascular endothelial growth factor.