**Supplement**

**Pregnancy exposure to synthetic phenols and placental DNA methylation** — **An epigenome-wide association study in male infants from the EDEN cohort**

Paulina Jedynak, Jörg Tost, Antonia M. Calafat, Ekaterina Bourova-Flin, Florence Busato, Anne Forhan, Barbara Heude, Milan Jakobi, Sophie Rousseaux, Joel Schwartz, Rémy Slama, Daniel Vaiman, Claire Philippat, Johanna Lepeule

**Contents**

[**Supplementary** **methods** 3](#_Toc76041241)

[1. Study exclusion criteria 3](#_Toc76041242)

[2. Quantification of maternal concentrations of creatinine and phenols 3](#_Toc76041243)

[3. Quality control/quality assurance for the assessment of synthetic phenols in maternal urine 3](#_Toc76041244)

[4. Phenol concentrations standardization 4](#_Toc76041245)

[5. Placental tissue collection and DNA extraction 4](#_Toc76041246)

[6. Placental DNA methylation quality control, normalization and filtering of outliers 4](#_Toc76041247)

[7. Placental tissue heterogeneity estimation 5](#_Toc76041248)

[8. Mediation analysis 6](#_Toc76041249)

[**Appendix Tables** 7](#_Toc76041250)

[**Appendix Table 1:** Population characteristics for the mother-son pairs included in the study and recruited between 2003 and 2006 (n = 202). 7](#_Toc76041251)

[**Appendix Table 2:** Maternal urinary phenol concentrations measured between 22 and 29 gestational weeks in spot urine samples, measured and standardized on sampling conditions (n = 202). 8](#_Toc76041252)

[**Appendix Table 3**: Differentially methylated CpGs associated with prenatal concentrations of phenols in EWAS unadjusted and adjusted for placental cell heterogeneity (FDR p-value <0.05, n = 202, 379,904 CpGs). 9](#_Toc76041253)

[**Appendix Table 4:** DMRs associated with prenatal concentrations of phenols unadjusted and adjusted for placental cell heterogeneity (Šidák-corrected p-value <0.05, n = 202, 379,904 CpGs). 10](#_Toc76041254)

[**Appendix Table 5**: Data on the function of the genes associated with phenol concentrations. 11](#_Toc76041255)

[**Appendix Table 6**: Adjusted global analysis of methylation profiles (GAMP) associated with prenatal concentrations of phenols (n = 202, 379,904 CpGs). 12](#_Toc76041256)

[**Appendix Figures** 13](#_Toc76041257)

[**Appendix Figure 1**:Study flow chart. 13](#_Toc76041258)

[**Appendix Figure 2**: Spearman’s correlation coefficients (rho) between standardized urinary phenol concentrations (n = 202). 14](#_Toc76041259)

[**Appendix Figure 3**: Q-Q plots with genomic inflation factor (λ) and Bayesian inflation factor (BIF) for the association between phenol concentrations and DNA methylation sites in the EWAS (n = 202, 379,904 CpGs). 15](#_Toc76041260)

[**Appendix Figure 4:** Adjusted associations between prenatal concentrations of phenols and methylation levels of the repetitive elements *Alu* andLINE-1 (n = 201). 16](#_Toc76041261)

[**Appendix Figure 5**: Impact of data processing and normalization on the overall DNA methylation profile of placental samples (n = 767, 668 samples and replicates) using the ChAMP. 17](#_Toc76041262)

[**References** 18](#_Toc76041263)

# **Supplementary methods**

## **Study exclusion criteria**

Women with multiple fetuses, maternal diabetes before pregnancy, intention to deliver outside the university hospital or to move out of the study region within the next three years, and unable to speak French were excluded from the study.

## **Quantification of maternal concentrations of creatinine and phenols**

Quantification of maternal concentrations of creatinine and nine phenols (free plus conjugated species): bisphenol A, benzophenone-3, triclosan, 2,4- and 2,5-dichlorophenol and butyl-, ethyl-, methyl-, propylparaben was performed at the Centers for Disease Control and Prevention (CDC). In short, 0.1 mL of urine spiked with the appropriate reagents and standards were incubated to hydrolyze the biomarkers urinary conjugates. The procedure for extracting the deconjugated biomarkers from the urine involved concurrent online solid phase extraction and high performance liquid chromatography followed by isotope dilution tandem mass spectrometry as detailed before (Ye et al. 2005).

## **Quality control/quality assurance for the assessment of synthetic phenols in maternal urine**

The CDC laboratory quantified the phenols following the strict quality control/quality assurance requirements set forth in the Clinical Laboratory Improvement Amendments of 1988 (https://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/index.html). These requirements include evaluation of calibrators and successfully analyzing twice per year proficiency testing samples such as those provided by the University of Erlangen-Nuremberg’s German External Quality Assessment Scheme (G-EQUAS) and the Centre de Toxicologie du Québec’s External Quality Assessment Scheme for Organic Substances in Urine (OSEQAS). Furthermore, along with study samples, each analytical run included a set of calibrators, high- and low-concentration quality control materials, and reagent blanks to assure the accuracy and reliability of the data. CDC has used the method to quantify phenols since the early 2000s for the analyses of tens of thousands of biological specimens, including those collected as part of the ongoing U.S. National Health and Nutrition Examination Survey (NHANES). Precision (relative standard deviation from repeated measurements of quality control materials) varied from 4% to 10%, depending on the analyte.

## **Phenol concentrations standardization**

Phenol concentrations were standardized on sampling conditions (hour of sampling, day of sampling, year of sample analysis, gestational age at collection, duration of storage at room temperature before freezing) and creatinine concentration using a method based on regression residuals (Mortamais et al. 2012).

## **Placental tissue collection and DNA extraction**

Placental tissue from the fetal side, a few centimeters from the insertion of the cord, was sampled at delivery by the EDEN midwife or a technician of the study using a standardized procedure. Samples of around 5mm3 were collected and immediately frozen at −80°C. DNA was extracted by QIAGEN Genomic Services using the QIAsymphony instrument (QIAGEN GmbH, Germany) with the QIAsymphony DSP DNA Mini Kit according to the instruction. DNA concentration was determined by Nanodrop (ThermoFisher Scientific, USA) measurement and fluorescent quantification using PicoGreen (ThermoFisher Scientific, France). No samples were discarded due to low DNA concentration.

## **Placental DNA methylation quality control, normalization and filtering of outliers**

To reduce the influence of technical factors such as batch effects and to ensure the balance of individuals from each gender and each recruitment center (Poitiers/ Nancy), DNA samples from the main experiment (n = 668 containing children of both genders) were randomly allocated to assay chips. 24 samples were measured in replicates (from two to five replicates) across batches, sample plates and chips to detect technical issues, if any. Raw intensities of fluorescent signals were processed with the Chip Analysis Methylation Pipeline (ChAMP) V2.14 (Morris et al. 2014; Tian et al. 2017). All samples but one passed initial quality control with an average of >98% valid data points (p-value <0.01). Among the 667 samples that passed quality control, 202 were included in our study. Filtering included removal of probes with detection p-values above 0.01 (52,692 probes), low numbers of measured events [beadcount <3 in at least 5% of samples (44 probes)], probes not targeting a CpG (2,034 probes), probes associated with SNPs (50,829 probes) or unspecific probes (9 probes) (Nordlund et al. 2013). Methylation levels of individual CpGs were reported as continuous averaged β-values, representing the proportion of methylated alleles for each methylation site ranging from 0 (indicating that the site is completely unmethylated) to 1 (completely methylated), and were normalized in ChAMP via the Beta MIxture Quantile (BMIQ) normalization (Teschendorff et al. 2013) (Appendix Figure 5). To reduce the influence of outliers, methylation beta values above the 75th percentile + three interquartile ranges (IQRs) or below the 25th percentile - three IQRs for a CpG were removed (in total 0.39% of all methylation values in our subsample of 202 participants). 379,904 methylation sites remained after quality control, normalization and filtering of outliers.

## **Placental tissue heterogeneity estimation**

To estimate placental tissue heterogeneity, we followed the pipeline recommended by Decamps et al. (Decamps et al. 2020). First, we filtered out the probes correlated with factors affecting the DNA methylation but unlikely to affect the placental cell proportions, such as center of recruitment, delivery mode and technical factors (batch, plate and chip). To assess the associations between these factors and each probe we ran univariate linear regressions (*CF\_detection* function from the *medepir* R package, https://bcm-uga.github.io/medepir) with default settings. CpGs associated with at least one of these factors (p-value <0.15) were excluded from the cell mix estimation. For the 3,645 remaining CpGs we identified six putative constituent cell types using Cattell’s rule applied to the scree plot (Cattell 1966) and then estimated tissue heterogeneity for the same set of probes using a reference-free based method via the *RefFreeCellMix* function from the *RefFreeEWAS* package in R (Houseman et al. 2016).

## **Mediation analysis**

To explore whether placental cell heterogeneity could mediate the association between phenol exposure and DNA methylation levels, we first estimated the difference between regression estimates obtained for the models unadjusted and adjusted for the estimated placental tissue heterogeneity. The percentage difference was calculated using the following formula: [(βadjusted - βunadjusted) / βunadjusted] × 100%. Regression coefficient values showing ≥20% absolute difference between adjusted and unadjusted effect estimates were further tested for mediation. The six estimated placental cell types were reduced to the first principal component (PC1) explaining >50% of the variance of the cell heterogeneity. PC1 was then used in the mediation analysis using the *mediation* R package (Tingley et al. 2014).

# **Appendix Tables**

## **Appendix Table 1:** Population characteristics for the mother-son pairs included in the study and recruited between 2003 and 2006 (n = 202).

|  |  |  |
| --- | --- | --- |
| **Characteristics** | **Distribution** | |
| **n (%)** | **Median [Q1, Q3]** |
| **Center of recruitment** |  |  |
| Nancy | 103 (51.0%) |  |
| Poitiers | 99 (49.0%) |  |
| **Season of conception** |  |  |
| January-March | 44 (21.8%) |  |
| April-June | 41 (20.3%) |  |
| July-September | 57 (28.2%) |  |
| October-December | 60 (29.7%) |  |
| **Maternal active smoking in the 3 months preceding pregnancy and during pregnancy** |  |  |
| Did not smoke | 127 (62.9%) |  |
| Smoked before pregnancy | 19 (9.4%) |  |
| Smoked before and during pregnancy | 26 (12.9%) |  |
| Other | 30 (14.9%) |  |
| **Parity** |  |  |
| Nulliparous | 88 (43.6%) |  |
| ≥ 1 child | 114 (56.4%) |  |
| **Maternal level of education** |  |  |
| <2 years after high school | 93 (46.0%) |  |
| high school + 2 years | 43 (21.3%) |  |
| ≥high school + 3 years | 66 (32.7%) |  |
| **Maternal age (years)** |  | 29.1 [25.6;33.0] |
| **Gestational age at delivery (weeks)**a |  | 40.0 [38.9;41.0] |

a Based on the date of the LMP or gestational duration assessed by the obstetrician if it differed from the LMP-based estimate by more than 2 weeks. Abbreviations: LMP = last menstrual period. Q = quantile.

## **Appendix Table 2:** Maternal urinary phenol concentrations assessed between 22 and 29 gestational weeks in spot urine samples, measured and standardized on sampling conditions (n = 202).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Phenol** | **LOD (µg/L)** | **% > LOD** | **Measured concentrations** | | | **Standardized concentrationsa** | | |
| **Percentiles (µg/L)** | | | **Percentiles (µg/L)** | | |
| **5th** | **50th** | **95th** | **5th** | **50th** | **95th** |
| **2,4-dichlorophenol** | 0.20 | 97.00 | 0.40 | 1.20 | 7.93 | 0.30 | 1.10 | 5.92 |
| **2,5-dichlorophenol** | 0.20 | 100.00 | 2.30 | 10.25 | 187.45 | 2.04 | 10.30 | 161.56 |
| **∑ dichlorophenols** | NA | NA | 0.02 | 0.08 | 1.19 | 0.02 | 0.07 | 1.03 |
| **Bisphenol A** | 0.40 | 100.00 | 0.71 | 2.65 | 11.39 | 0.94 | 2.15 | 9.64 |
| **Benzophenone-3** | 0.40 | 87.10 | 0.60 | 2.50 | 69.00 | 0.12 | 2.02 | 53.22 |
| **Triclosan** | 2.30 | 84.20 | 3.79 | 85.55 | 935.40 | 0.19 | 43.14 | 758.57 |
| **Butylparaben** | 1.00 | 81.20 | 0.40 | 3.35 | 87.39 | 0.09 | 1.93 | 72.38 |
| **Ethylparaben** | 1.00 | 72.30 | 1.30 | 7.45 | 68.12 | 0.09 | 2.91 | 47.77 |
| **Methylparaben** | 0.20 | 100.00 | 8.41 | 98.75 | 1135.50 | 10.26 | 101.60 | 980.16 |
| **Propylparaben** | 0.20 | 97.50 | 0.68 | 16.90 | 308.20 | 0.52 | 12.20 | 249.79 |

Phenol concentrations are displayed in µg/L for all compounds except for the ∑ dichlorophenols for which the concentrations are presented as µmol/L.

a Measured concentrations were standardized on sampling conditions using a method based on regression residuals [(Mortamais et al. 2012)](https://www.zotero.org/google-docs/?8m1MLp).

Abbreviations: LOD = limit of detection.

## **Appendix Table 3**: Differentially methylated CpGs associated with pregnancy concentrations of phenols in EWAS unadjusted and adjusted for placental cell heterogeneity (FDR p-value <0.05, n = 202, 379,904 CpGs).

PLEASE REFER TO THE EXTERNAL APPENDIX TABLE 3

EWAS regression models were adjusted for recruitment center, maternal active smoking in the three months preceding pregnancy and during pregnancy, maternal age, parity, maternal education level, season of conception, batch, plate, and chip. Sensitivity analysis was additionally adjusted for placental cell heterogeneity and the percentage difference between regression coefficient estimates was calculated using the following formula: [(βadjusted - β unadjusted) / β unadjusted] × 100%. ACME p-value was calculated using the first principal component representing six reference-free estimated placental cell types.

a UCSC.

ᵇ Mediation analysis was performed for CpGs associated with triclosan for which the absolute percentage difference between regression coefficient estimates obtained in models unadjusted and adjusted for placental tissue heterogeneity was ≥20%.

Abbreviations: ACME = average causal mediation effect. Chr = chromosome. CI = confidence interval. FDR = false discovery rate. UCSC = University of California, Santa Cruz.

## **Appendix Table 4:** DMRs associated with pregnancy concentrations of phenols unadjusted and adjusted for placental cell heterogeneity (Šidák-corrected p-value <0.05, n = 202, 379,904 CpGs).

PLEASE REFER TO THE EXTERNAL APPENDIX TABLE 4

EWAS regression models on which the DMR analysis was based were adjusted for recruitment center, maternal active smoking in the three months preceding pregnancy and during pregnancy, maternal age, parity, maternal education level, season of conception, batch, plate, and chip. Sensitivity analysis was additionally adjusted for placental cell heterogeneity. DMRs with less than five CpGs are highlighted in grey.

a UCSC.

Abbreviations: Chr = chromosome; DMR = differentially methylated region. NS = not significant. SLK = Stouffer-Liptak-Kechris correction. UCSC = University of California, Santa Cruz.

## **Appendix Table 5**: Data on the function of the genes associated with phenol concentrations.

PLEASE REFER TO THE EXTERNAL APPENDIX TABLE 5

Information on genes encompassed by the DMRs identified as associated with phenol concentrations were retrieved from the GeneCards Human Gene Database (Stelzer et al. 2016).

## **Appendix Table 6**: Adjusted global analysis of methylation profiles (GAMP) associated with pregnancy concentrations of phenols (n = 202, 379,904 CpGs).

|  |  |  |
| --- | --- | --- |
| **Phenol** | **CDFᵃ p-value** | **Densityᵇ p-value** |
| **2,4-dichlorophenol** | 0.32 | 0.46 |
| **2,5-dichlorophenol** | 0.27 | 0.48 |
| **∑ dichlorophenols** | 0.25 | 0.47 |
| **Bisphenol A** | 0.67 | 0.84 |
| **Benzophenone-3** | 0.14 | 0.6 |
| **Triclosan** | 0.55 | 0.67 |
| **Butylparaben** | 0.57 | 0.9 |
| **Ethylparaben** | 0.24 | 1.00 |
| **Methylparaben** | 0.13 | 0.24 |
| **Propylparaben** | 0.56 | 0.51 |

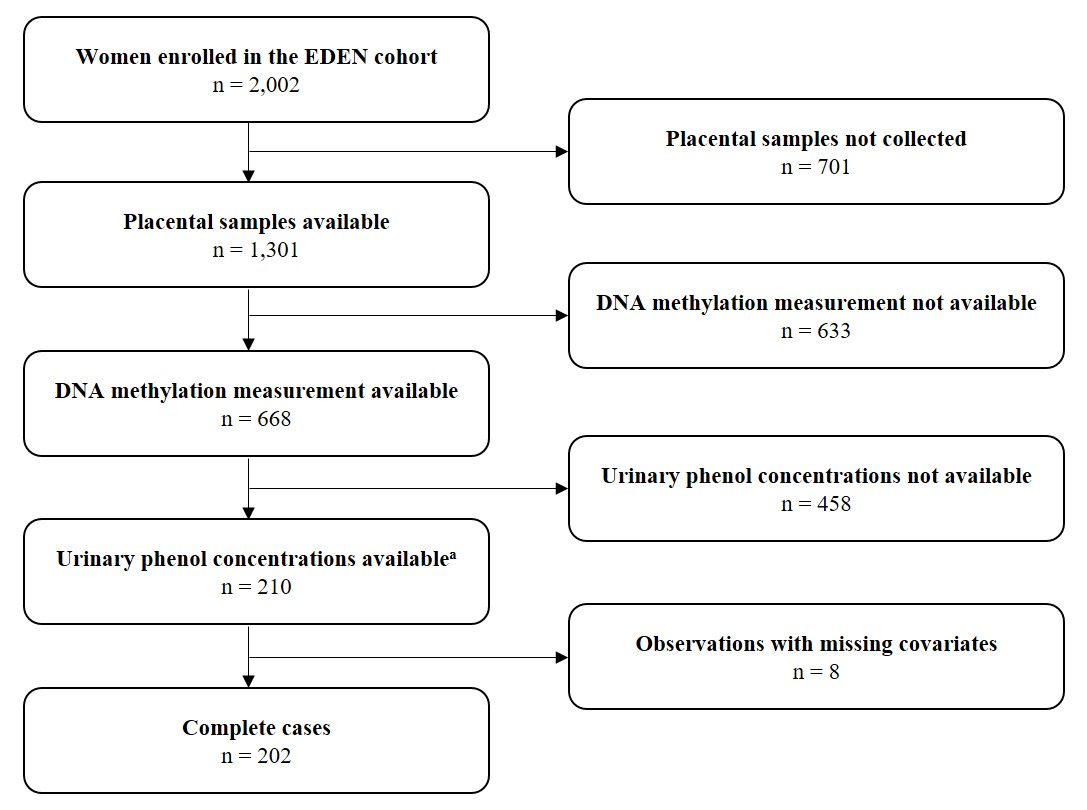
GAMP models were adjusted for recruitment center, maternal active smoking in the three months preceding pregnancy and during pregnancy, maternal age, parity, maternal education level, season of conception, batch, plate, and chip.

a Tests the association of the CDF of the observed methylation distributions for each individual with each exposure variable.

b Tests whether the densities of the observed methylation distributions for each individual are associated with an exposure variable.

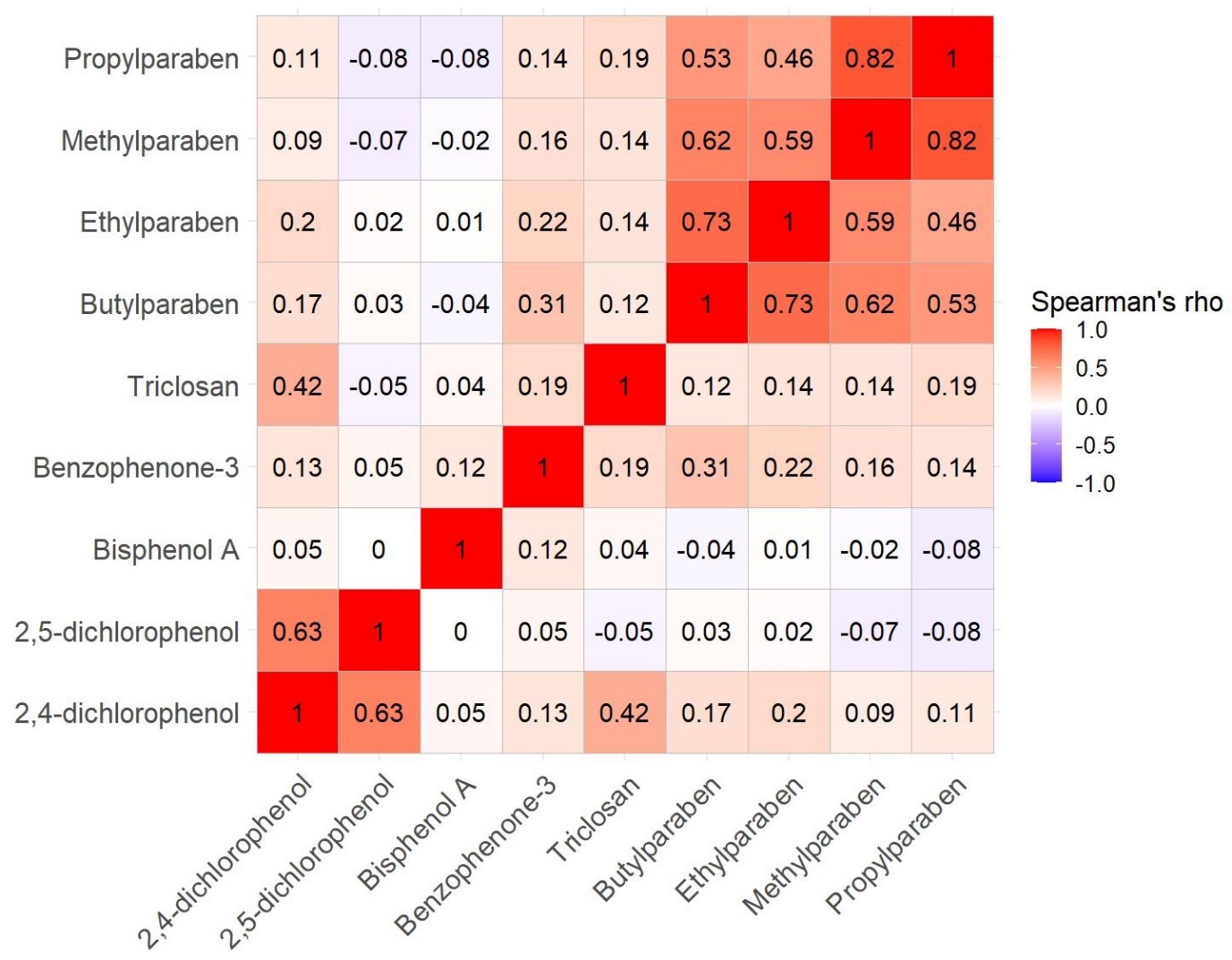
Abbreviations: CDF = cumulative distribution function. GAMP = global analysis of methylation profiles.

# **Appendix Figures**

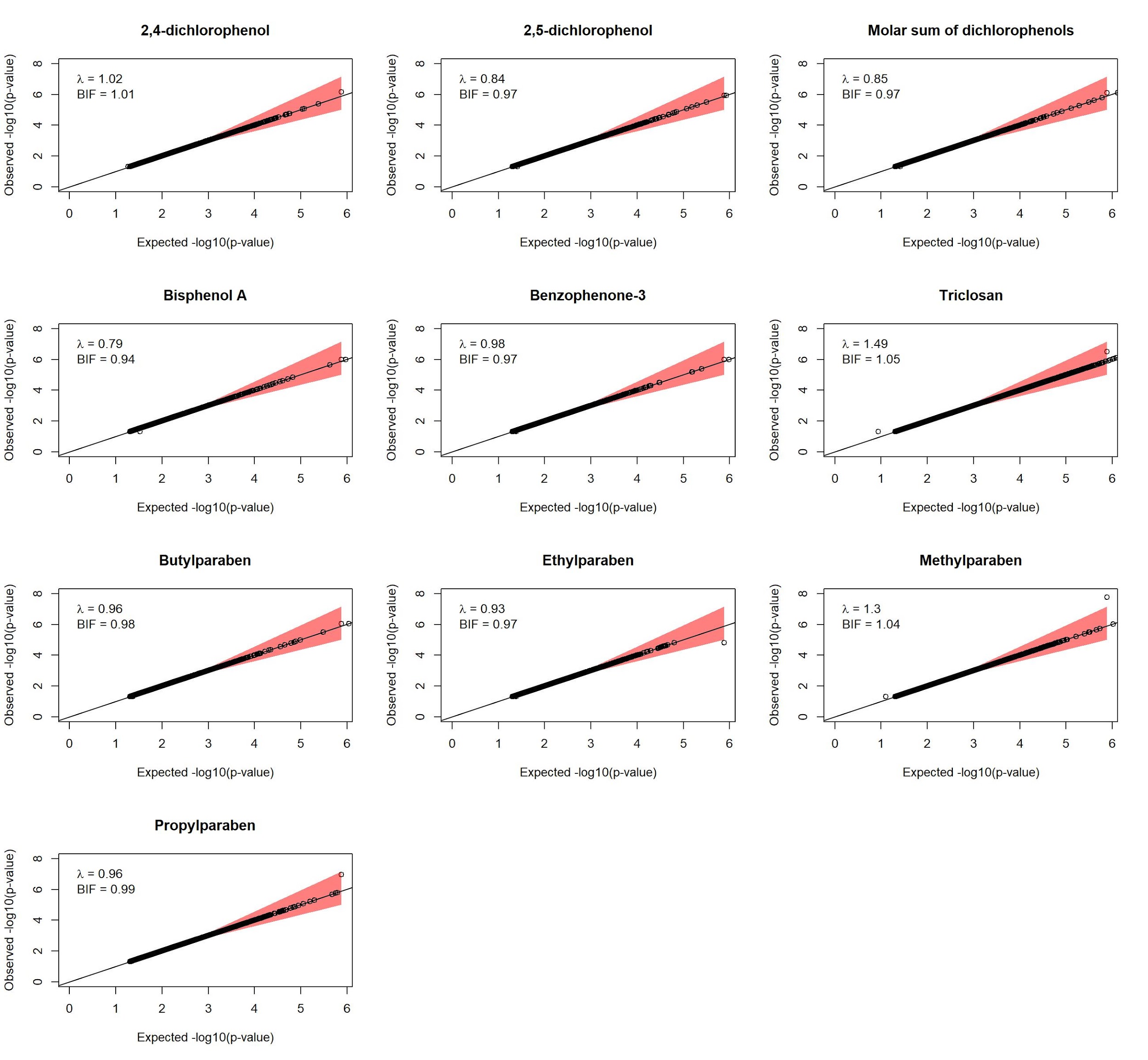


## **Appendix Figure 1**:Study flow chart.

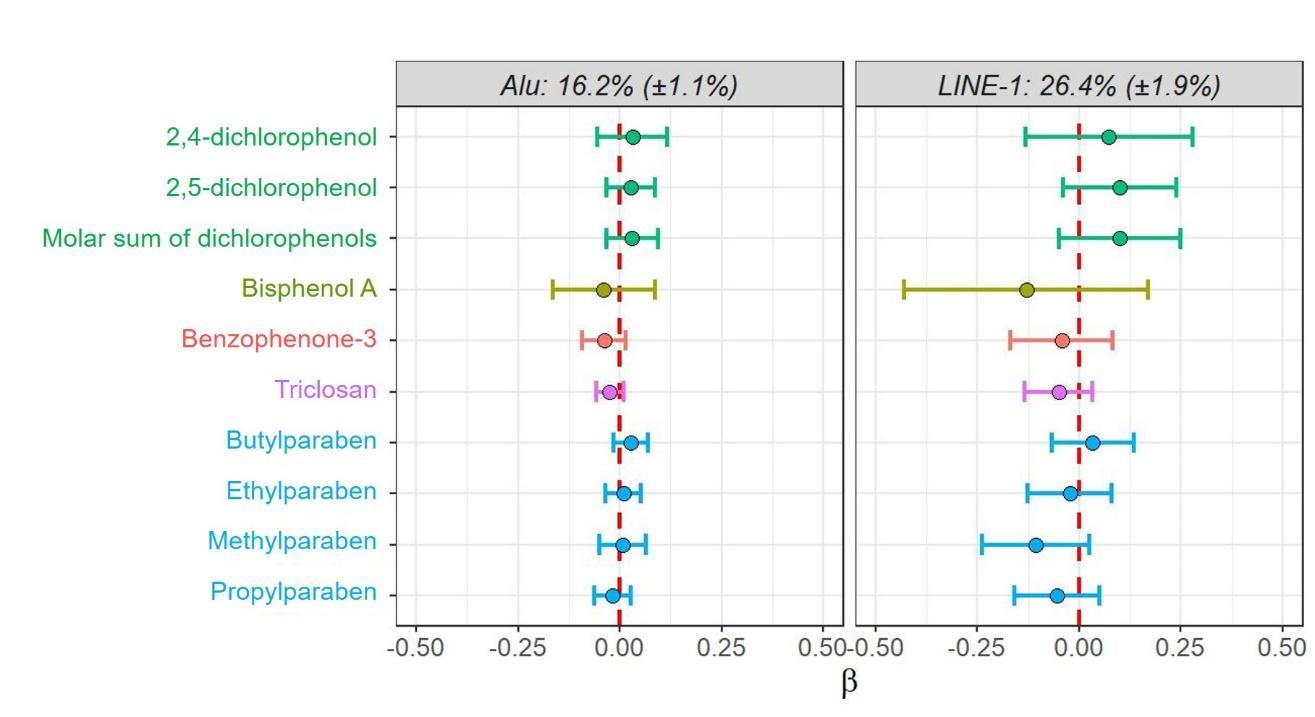
a Phenol concentrations were assessed in the framework of a previous study restricted to boys and with at least one maternal urine sample available for phenol measurements and complete data on prenatal and postnatal growth (Philippat et al. 2014).

****

## **Appendix Figure 2**: Spearman’s correlation coefficients (rho) between standardized urinary phenol concentrations (n = 202).

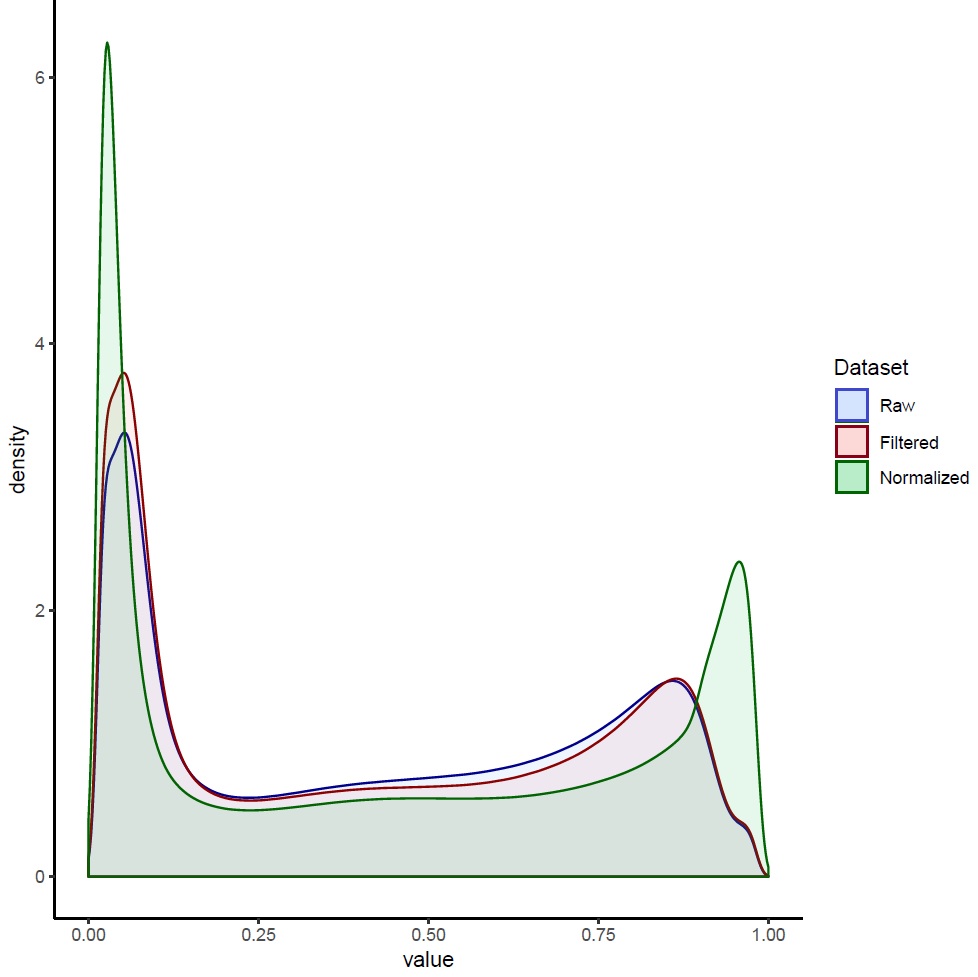


## **Appendix Figure 3**: Q-Q plots with genomic inflation factor (λ) and Bayesian inflation factor (BIF) for the association between phenol concentrations and DNA methylation sites in the EWAS (n = 202, 379,904 CpGs). Red area marks the confidence interval of the fit. Regression models were adjusted for recruitment center, maternal active smoking in the three months preceding pregnancy and during pregnancy, maternal age, parity, maternal education level, season of conception, batch, plate, and chip.



## **Appendix Figure 4:** Adjusted associations between pregnancy concentrations of phenols and methylation levels of the repetitive elements *Alu* andLINE-1 (n = 201). Circles represent β regression coefficient estimates reported with 95%CIs and correspond to a change in the global DNA methylation level for doubling of the urinary phenol concentration. Same color represents phenols belonging to the same family. Regression models were adjusted for recruitment center, maternal active smoking in the three months preceding pregnancy and during pregnancy, maternal age, parity, maternal education level, season of conception, batch, and plate.

Abbreviations: CI = confidence interval.



## **Appendix Figure 5**: Impact of data processing and normalization on the overall DNA methylation profile of placental samples (n = 767, 668 samples and replicates) using the ChAMP. Blue color represents raw data as extracted from the .idat files; red color represents filtered data (removal of probes with detection p-values >0.01, with a beadcount <3 in at least 5% of samples, probes not targeting CG, associated with SNPs or unspecific probes); green color represents BMIQ normalized data.

Abbreviations: BMIQ = Beta MIxture Quantile. ChAMP = Chip Analysis Methylation Pipeline.

SNP = single nucleotide polymorphism.

# **References**

Botton J, Philippat C, Calafat AM, Carles S, Charles M-A, Slama R, et al. 2016. Phthalate pregnancy exposure and male offspring growth from the intra-uterine period to five years of age. Environ Res 151:601–609; doi:10.1016/j.envres.2016.08.033. PMID: 27596487.

Cattell RB. 1966. The scree test for the number of factors. Multivariate Behav Res 1:245–276; doi:10.1207/s15327906mbr0102\_10. PMID: 26828106.

Decamps C, Privé F, Bacher R, Jost D, Waguet A, Achard S, et al. 2020. Guidelines for cell-type heterogeneity quantification based on a comparative analysis of reference-free DNA methylation deconvolution software. BMC Bioinformatics 21:16; doi:10.1186/s12859-019-3307-2. PMID: 31931698.

Houseman EA, Kile ML, Christiani DC, Ince TA, Kelsey KT, Marsit CJ. 2016. Reference-free deconvolution of DNA methylation data and mediation by cell composition effects. BMC Bioinformatics 17:259; doi:10.1186/s12859-016-1140-4.

Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. 2014. ChAMP: 450k Chip Analysis Methylation Pipeline. Bioinformatics 30:428–430; doi:10.1093/bioinformatics/btt684. PMID: 24336642.

Mortamais M, Chevrier C, Philippat C, Petit C, Calafat AM, Ye X, et al. 2012. Correcting for the influence of sampling conditions on biomarkers of exposure to phenols and phthalates: a 2-step standardization method based on regression residuals. Environ Health 11:29; doi:10.1186/1476-069X-11-29. PMID: 22537080.

Nordlund J, Bäcklin CL, Wahlberg P, Busche S, Berglund EC, Eloranta M-L, et al. 2013. Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. Genome Biol 14:r105; doi:10.1186/gb-2013-14-9-r105. PMID: 24063430.

Philippat C, Botton J, Calafat AM, Ye X, Charles M-A, Slama R. 2014. Prenatal exposure to phenols and growth in boys. Epidemiology 25:625–635; doi:10.1097/EDE.0000000000000132. PMID: 25061923.

Philippat C, Heude B, Botton J, Alfaidy N, Calafat AM, Slama R, et al. 2019. Prenatal exposure to select phthalates and phenols and associations with fetal and placental weight among male births in the EDEN cohort (France). Environ Health Perspect 127:17002; doi:10.1289/EHP3523. PMID: 30624098.

Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. 2016. The GeneCards Suite: from gene data mining to disease genome sequence analyses. Curr Protoc Bioinformatics 54:1.30.1-1.30.33; doi:10.1002/cpbi.5. PMID: 27322403.

Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, et al. 2013. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics 29:189–196; doi:10.1093/bioinformatics/bts680. PMID: 23175756.

Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, et al. 2017. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. Bioinformatics 33:3982–3984; doi:10.1093/bioinformatics/btx513. PMID: 28961746.

Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. 2014. mediation: *R* Package for causal mediation analysis. J Stat Soft 59; doi:10.18637/jss.v059.i05.

Ye X, Kuklenyik Z, Needham LL, Calafat AM. 2005. Quantification of urinary conjugates of bisphenol A, 2,5-dichlorophenol, and 2-hydroxy-4-methoxybenzophenone in humans by online solid phase extraction–high performance liquid chromatography–tandem mass spectrometry. Anal Bioanal Chem 383:638–644; doi:10.1007/s00216-005-0019-4. PMID: 16132150.