1	Supplement	
---	------------	--

- 2 Pregnancy exposure to synthetic phenols and placental DNA methylation An epigenome-
- 3 wide association study in male infants from the EDEN cohort
- 4 Paulina Jedynak, Jörg Tost, Antonia M. Calafat, Ekaterina Bourova-Flin, Florence Busato, Anne
- 5 Forhan, Barbara Heude, Milan Jakobi, Sophie Rousseaux, Joel Schwartz, Rémy Slama, Daniel
- 6 Vaiman, Claire Philippat, Johanna Lepeule
- 7

8 Contents

9	Suppl	ementary methods
10	1.	Study exclusion criteria
11	2.	Quantification of maternal concentrations of creatinine and phenols
12	3.	Quality control/quality assurance for the assessment of synthetic phenols in maternal urine 3
13	4.	Phenol concentrations standardization
14	5.	Placental tissue collection and DNA extraction
15	6.	Placental DNA methylation quality control, normalization and filtering of outliers
16	7.	Placental tissue heterogeneity estimation
17	8.	Mediation analysis
18	Аррен	ndix Tables7
19 20	App recr	Dendix Table 1: Population characteristics for the mother-son pairs included in the study and uited between 2003 and 2006 ($n = 202$)
21 22 23	App gest 202	Dendix Table 2: Maternal urinary phenol concentrations measured between 22 and 29 ational weeks in spot urine samples, measured and standardized on sampling conditions (n =)
24 25 26	App phe 202	Dendix Table 3 : Differentially methylated CpGs associated with prenatal concentrations of nols in EWAS unadjusted and adjusted for placental cell heterogeneity (FDR p-value <0.05, n = , 379,904 CpGs)
27 28 29	App adju	pendix Table 4: DMRs associated with prenatal concentrations of phenols unadjusted and usted for placental cell heterogeneity (Šidák-corrected p-value <0.05, n = 202, 379,904 CpGs).
30	Арр	pendix Table 5: Data on the function of the genes associated with phenol concentrations11
31 32	App prer	Dendix Table 6 : Adjusted global analysis of methylation profiles (GAMP) associated with natal concentrations of phenols ($n = 202, 379,904$ CpGs)

33	Appendix Figures
34	Appendix Figure 1: Study flow chart
35 36	Appendix Figure 2 : Spearman's correlation coefficients (rho) between standardized urinary phenol concentrations (n = 202)
37 38 39	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)
40 41	Appendix Figure 4: Adjusted associations between prenatal concentrations of phenols and methylation levels of the repetitive elements Alu and LINE-1 (n = 201)
42 43	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 ChAMP17
44	References
45 46	

47 Supplementary methods

48 **1. 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.

52

2. Quantification of maternal concentrations of creatinine and phenols

53 Quantification of maternal concentrations of creatinine and nine phenols (free plus conjugated 54 species): bisphenol A, benzophenone-3, triclosan, 2,4- and 2,5-dichlorophenol and butyl-, ethyl-, 55 methyl-, propylparaben was performed at the Centers for Disease Control and Prevention (CDC). In 56 short, 0.1 mL of urine spiked with the appropriate reagents and standards were incubated to hydrolyze 57 the biomarkers urinary conjugates. The procedure for extracting the deconjugated biomarkers from 58 the urine involved concurrent online solid phase extraction and high performance liquid 59 chromatography followed by isotope dilution tandem mass spectrometry as detailed before (Ye et al. 2005). 60

G1 3. Quality control/quality assurance for the assessment of synthetic phenols in maternal urine

63 The CDC laboratory quantified the phenols following the strict quality control/quality assurance 64 requirements set forth in the Clinical Laboratory Improvement Amendments of 1988 65 (https://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/index.html). These requirements 66 include evaluation of calibrators and successfully analyzing twice per year proficiency testing samples 67 such as those provided by the University of Erlangen-Nuremberg's German External Quality 68 Assessment Scheme (G-EQUAS) and the Centre de Toxicologie du Québec's External Quality 69 Assessment Scheme for Organic Substances in Urine (OSEQAS). Furthermore, along with study 70 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.

76

4. 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).

81 5. Placental tissue collection and DNA extraction

82 Placental tissue from the fetal side, a few centimeters from the insertion of the cord, was sampled 83 at delivery by the EDEN midwife or a technician of the study using a standardized procedure. 84 Samples of around 5mm³ were collected and immediately frozen at -80°C. DNA was extracted by 85 QIAGEN Genomic Services using the QIAsymphony instrument (QIAGEN GmbH, Germany) with 86 the QIAsymphony DSP DNA Mini Kit according to the instruction. DNA concentration was 87 determined by Nanodrop (ThermoFisher Scientific, USA) measurement and fluorescent quantification 88 using PicoGreen (ThermoFisher Scientific, France). No samples were discarded due to low DNA 89 concentration.

90

6. Placental DNA methylation quality control, normalization and filtering of outliers

91 To reduce the influence of technical factors such as batch effects and to ensure the balance of
92 individuals from each gender and each recruitment center (Poitiers/ Nancy), DNA samples from the
93 main experiment (n = 668 containing children of both genders) were randomly allocated to assay
94 chips. 24 samples were measured in replicates (from two to five replicates) across batches, sample

95 plates and chips to detect technical issues, if any. Raw intensities of fluorescent signals were 96 processed with the Chip Analysis Methylation Pipeline (ChAMP) V2.14 (Morris et al. 2014; Tian et 97 al. 2017). All samples but one passed initial quality control with an average of >98% valid data points 98 (p-value <0.01). Among the 667 samples that passed quality control, 202 were included in our study. 99 Filtering included removal of probes with detection p-values above 0.01 (52,692 probes), low 100 numbers of measured events [beadcount <3 in at least 5% of samples (44 probes)], probes not 101 targeting a CpG (2,034 probes), probes associated with SNPs (50,829 probes) or unspecific probes (9 102 probes) (Nordlund et al. 2013). Methylation levels of individual CpGs were reported as continuous 103 averaged β -values, representing the proportion of methylated alleles for each methylation site ranging 104 from 0 (indicating that the site is completely unmethylated) to 1 (completely methylated), and were 105 normalized in ChAMP via the Beta MIxture Quantile (BMIQ) normalization (Teschendorff et al. 106 2013) (Appendix Figure 5). To reduce the influence of outliers, methylation beta values above the 107 75th percentile + three interquartile ranges (IQRs) or below the 25th percentile - three IQRs for a CpG 108 were removed (in total 0.39% of all methylation values in our subsample of 202 participants). 379,904 109 methylation sites remained after quality control, normalization and filtering of outliers.

110

7. Placental tissue heterogeneity estimation

111 To estimate placental tissue heterogeneity, we followed the pipeline recommended by Decamps et 112 al. (Decamps et al. 2020). First, we filtered out the probes correlated with factors affecting the DNA 113 methylation but unlikely to affect the placental cell proportions, such as center of recruitment, 114 delivery mode and technical factors (batch, plate and chip). To assess the associations between these 115 factors and each probe we ran univariate linear regressions (CF_detection function from the medepir 116 R package, https://bcm-uga.github.io/medepir) with default settings. CpGs associated with at least one 117 of these factors (p-value <0.15) were excluded from the cell mix estimation. For the 3,645 remaining 118 CpGs we identified six putative constituent cell types using Cattell's rule applied to the scree plot 119 (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 (Housemanet al. 2016).

122 8. Mediation analysis

123 To explore whether placental cell heterogeneity could mediate the association between phenol 124 exposure and DNA methylation levels, we first estimated the difference between regression estimates 125 obtained for the models unadjusted and adjusted for the estimated placental tissue heterogeneity. The 126 percentage difference was calculated using the following formula: $[(\beta_{adjusted} - \beta_{unadjusted}) / \beta_{unadjusted}] \times$ 127 100%. Regression coefficient values showing ≥20% absolute difference between adjusted and 128 unadjusted effect estimates were further tested for mediation. The six estimated placental cell types 129 were reduced to the first principal component (PC1) explaining >50% of the variance of the cell 130 heterogeneity. PC1 was then used in the mediation analysis using the mediation R package (Tingley et 131 al. 2014).

132 Appendix Tables

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

	Distribution		
Characteristics	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%)		
\geq 1 child	114 (56.4%)		
Maternal level of education			
<2 years after high school	93 (46.0%)		
high school $+ 2$ years	43 (21.3%)		
\geq 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 =

137 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).

			Measured concentrations		Standardized concentrations ^a			
	LOD	% >	Percentiles (µg/L)			Percentiles (µg/L)		
Phenol	(µg/L)	LOD	5th	50th	95th	5th	50th	95th
2,4-dichlorophenol	0.2	98	0.3	1.1	7.9	0.3	1.1	5.9
2,5-dichlorophenol	0.2	100	2.3	10.3	187.5	2.0	10.3	161.6
\sum dichlorophenols	NA	NA	0.0	0.1	1.2	0.0	0.1	1.0
Bisphenol A	0.4	100	0.7	2.7	11.4	0.9	2.2	9.6
Benzophenone-3	0.4	89.1	< LOD	2.2	65.2	< LOD	2.0	53.2
Triclosan	2.3	84.2	< LOD	38.0	889.1	< LOD	43.1	758.6
Butylparaben	0.2	86.1	< LOD	1.9	77.8	< LOD	1.9	72.4
Ethylparaben	1.0	72.8	< LOD	4.1	61.9	< LOD	2.9	47.8
Methylparaben	1.0	100	8.4	98.8	1135.5	10.3	101.6	980.2
Propylparaben	0.2	98	0.6	15.4	304.4	0.5	12.2	249.8

141

142 Phenol concentrations are displayed in $\mu g/L$ for all compounds except for the \sum dichlorophenols for

143 which the concentrations are presented as μ mol/L.

^aMeasured concentrations were standardized on sampling conditions using a method based on

regression residuals (Mortamais et al. 2012).

146 Abbreviations: LOD = limit of detection.

- 148 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).

151 PLEASE REFER TO THE EXTERNAL APPENDIX TABLE 3

- 153 EWAS regression models were adjusted for recruitment center, maternal active smoking in the three
- 154 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
- 156 placental cell heterogeneity and the percentage difference between regression coefficient estimates
- 157 was calculated using the following formula: $[(\beta_{adjusted} \beta_{unadjusted}) / \beta_{unadjusted}] \times 100\%$. ACME p-value
- was calculated using the first principal component representing six reference-free estimated placentalcell types.
- 160 ^a UCSC.
- ^b Mediation analysis was performed for CpGs associated with triclosan for which the absolute
- 162 percentage difference between regression coefficient estimates obtained in models unadjusted and
- 163 adjusted for placental tissue heterogeneity was $\geq 20\%$.
- 164 Abbreviations: ACME = average causal mediation effect. Chr = chromosome. CI = confidence
- 165 interval. FDR = false discovery rate. UCSC = University of California, Santa Cruz.
- 166

- 167 Appendix Table 4: DMRs associated with pregnancy concentrations of phenols unadjusted and
- 168 adjusted for placental cell heterogeneity ($\check{S}id\acute{a}k$ -corrected p-value <0.05, n = 202, 379,904 CpGs).

169 PLEASE REFER TO THE EXTERNAL APPENDIX TABLE 4

- 171 EWAS regression models on which the DMR analysis was based were adjusted for recruitment
- 172 center, maternal active smoking in the three months preceding pregnancy and during pregnancy,
- 173 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.
- 176 ^a UCSC.
- 177 Abbreviations: Chr = chromosome; DMR = differentially methylated region. NS = not significant.
- 178 SLK = Stouffer-Liptak-Kechris correction. UCSC = University of California, Santa Cruz.
- 179

- 180 **Appendix Table 5**: Data on the function of the genes associated with phenol concentrations.
- 181 PLEASE REFER TO THE EXTERNAL APPENDIX TABLE 5
- 182 Information on genes encompassed by the DMRs identified as associated with phenol concentrations
- 183 were retrieved from the GeneCards Human Gene Database (Stelzer et al. 2016).

Appendix Table 6: Adjusted global analysis of methylation profiles (GAMP) associated with

186	pregnancy	concentrations	of phenols	(n = 202,	379,904	CpGs).
-----	-----------	----------------	------------	-----------	---------	--------

Phenol	CDF ^a p-value	Density ^b p-value
2,4-dichlorophenol	0.32	0.46
2,5-dichlorophenol	0.27	0.48
\sum 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

187

188 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 ofconception, batch, plate, and chip.

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

^bTests whether the densities of the observed methylation distributions for each individual are

194 associated with an exposure variable.

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



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).



206 Appendix Figure 2: Spearman's correlation coefficients (rho) between standardized urinary phenol 207 concentrations (n = 202).



211 Appendix Figure 3: Q-Q plots with genomic inflation factor (λ) and Bayesian inflation factor (BIF) 212 for the association between phenol concentrations and DNA methylation sites in the EWAS (n = 202, 213 379,904 CpGs). Red area marks the confidence interval of the fit. Regression models were adjusted 214 for recruitment center, maternal active smoking in the three months preceding pregnancy and during 215 pregnancy, maternal age, parity, maternal education level, season of conception, batch, plate, and 216 chip.



219 Appendix Figure 4: Adjusted associations between pregnancy concentrations of phenols and

220 methylation levels of the repetitive elements *Alu* and LINE-1 (n = 201). Circles represent β regression

coefficient estimates reported with 95% CIs and correspond to a change in the global DNA

222 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,

225 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.

235	Abbreviations: BMIQ = Beta MIxture	Quantile. ChAMI	P = Chip Analysis	Methylation Pipeline.
-----	------------------------------------	-----------------	-------------------	-----------------------

- 236 SNP = single nucleotide polymorphism.

239 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/s12859-0193307-2. 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 2step 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.00000000000132.
 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.