

Prenatal exposure to multiple air pollutants, mediating molecular mechanisms, and shifts in birthweight

Supporting Information

Jessica E. Laine^{1}, Barbara Bodinier¹, Oliver Robinson¹, Michelle Plusquin², Augustin Scalbert³, Pekka Keski-Rahkonen³, Nivonirina Robinot³, Roel Vermeulen⁴, Costanza Pizzi⁵, Federica Asta⁶, Tim Nawrot^{2,7}, John Gulliver⁸, Leda Chatzi⁹, Manolis Kogevinas^{10,11,12,13}, Mark Nieuwenhuijsen¹⁰, Jordi Sunyer^{10,11,12,13}, Martine Vrijheid^{10,11,12}, Marc Chadeau-Hyam¹, Paolo Vineis^{1,14}*

¹Department of Epidemiology and Biostatistics, MRC Centre for Environment and Health, Imperial College London, London, United Kingdom

²Center for Environmental Sciences, Hasselt University, Hasselt, Belgium

³International Agency for Research on Cancer (IARC), Nutrition and Metabolism Section, Biomarkers Group, Lyon, France

⁴Utrecht University, Institute for Risk Assessment Sciences, Environmental Epidemiology Division, Utrecht, Netherlands

⁵Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin and CPO-Piemonte, Turin, Italy

⁶Department of Epidemiology, Lazio Regional Health Service, ASL Roma 1, Rome, Italy

⁷Department of Public Health, Environment and Health Unit, Leuven University (KU Leuven), Leuven, Belgium

⁸Centre for Environmental Health and Sustainability & School of Geography, Geology and the Environment, University of Leicester, Leicester, United Kingdom

⁹Department of Social Medicine, Faculty of Medicine, University of Crete Heraklion, Crete Greece

¹⁰ISGlobal, Barcelona Institute for Global Health, 08003 Barcelona, Spain

¹¹CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

¹²Universitat Pompeu Fabra (UPF), Barcelona, Spain

¹³IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain

¹⁴Italian Institute of Technology, Genova, Italy

Supporting Information

Detailed Cohort Description:

The mother-child “Rhea” study in Crete is a prospective cohort examining a population sample of pregnant women and their children, at the prefecture of Heraklion (n = 1500). The study aims are to evaluate a) nutritional, environmental, biological and psychosocial exposures in the prenatal period and in early childhood, b) the association of these exposures with the development of the foetus and the child, c) mother’s health during and after pregnancy, and d) genetic susceptibility and the interactions between genetic and environmental factors affecting child health. A set of 100 children from the RHEA-cohort is included in the EXPOsOMICS children studies, for which data on cord blood DNA-methylation, metabolomics, proteomics and albumin adducts are available.

The INMA – Infancia y Medio Ambiente (Environment and Childhood) project follows up a population sample of around 3100 pregnant mothers and newborns. New and existing cohorts of pregnant women have been incorporated from seven different Spanish regions. Pregnant women are assessed at 12, 20 and 32 weeks of gestation to collect information about environmental exposures and foetal growth. Children are assessed at birth, at the age of 1 year and at the age of 4 years. The cohorts have been designed to evaluate the impact of environmental exposures and diet on children’s health. A set of 100 children from the INMA centre Sabadell is included in the EXPOsOMICS children studies, for which data on cord blood metabolomics, proteomics and albumin adducts are available.

Piccolipiù is a multicentric Italian birth cohort that recruited about 3000 new-borns and their mothers in 5 centres: Turin, Trieste, Florence, Viareggio and Rome (www.piccolipiu.it). Participants receive a follow-up interview 6, 12 and 24 months after delivery and a medical examination when the children turn 4 and 6. Growth trajectories and neurocognitive test-results are available. A set of 99 children from the Turin centre is

included in the EXPOsOMICS children studies, for which data on cord blood DNA methylation, metabolomics, proteomics and albumin adducts are available.

The ENVIRonAGE (ENVIRonmental influence ON AGEing in early life) cohort includes 1300 mother-infant pairs and further recruitment is ongoing. Data include mothers' lifestyle and socio-economic status, gestational history, measurements including the newborns' blood pressure (all healthy with gestational age 37–42 weeks), bio banked placental tissue and cord blood including RNA/DNA, toxic metals in cord blood and placenta, and in utero and early life exposure to fine particulates and NO₂ using a spatial temporal interpolation method. A set of 200 children from the ENVIRonAGE-cohort is included in the EXPOsOMICS children studies, for which data on cord blood DNA methylation, transcriptomics, metabolomics, proteomics and albumin adducts are available.

Exposure assessment:

Following the ESCAPE protocol the residential location of the mothers annual mean concentrations of PM_{2.5} and particulate matter with an aerodynamic diameter of less than 10 µm (PM₁₀), of between 2.5 µm and 10 µm (PM_{2.5–10}; coarse particulate matter), NO₂, and NO_x were estimated at the maternal home addresses with land use regression (LUR) models. New spatial models using LUR models for were developed to assess UFP exposure and oxidative potential. A detailed description of the models developed for UFP has been described elsewhere ¹. Using six European study areas we developed LUR models for predicting spatial patterns in UFP to assess the agreement in LUR model structure and performance within and between study areas. The performance of a model was evaluated using UFP concentration data from six study areas combined. Together this allowed for the evaluation of the robustness of model predictions at external residential sites, not included in model development in all six areas. A detailed description of the models developed for oxidative potential has been described elsewhere ². Briefly, we evaluated the spatial

variability of the OP of measured PM_{2.5} using ascorbate (AA) and (reduced) glutathione (GSH), and develop land use regression (LUR) models to explain this spatial variability. The estimated annual average values (m⁻³) of OPAA and OPGSH were derived.

'Omics assessment:

Whole blood samples were collected using venipuncture of cord vessels before the placenta was delivered and processed as follows in each cohort: In ENVIRONAGE, samples were collected into EDTA (BD, Franklin Lakes, NJ) vacutainers and within 20 min were centrifuged at 3200 rpm for 15 min into plasma. In Piccolipiù, samples were collected into BD EDTA vacutainers, stored at 4 °C for <24 h, and centrifuged for 10 min at 1300g into plasma. In Rhea, samples were collected into BD gel separator vacutainers and centrifuged within 2 h at 2500 rpm for 10 min into serum. In INMA, samples were collected into BD gel separator vacutainers, stored at 4 °C for <4 h, and centrifuged at 3000 rpm for 10–15 min into serum. Samples were immediately frozen at –20 °C (INMA) or –80 °C (all other cohorts) until analysis.

For the proteomics, those inflammatory markers measured on Luminex panel A included, TNF- α , IL-8, EGF, MIP1, GRO, IL-6, IL-17, and those measured on Luminex panel B included, MPO, IP-10, VEGF, IL-17, Periostin, MCP-1, IL-1rA, G-CSF, CRP was measured using ELSIA. Imputation was applied for inflammatory markers that had at least 40% detectable samples measurement per study group, based on maximum likelihood estimation procedure³. Samples with <40% detectable samples per study group were not imputed. To allow for plate to plate variation we imputed based on each plate-specific limit of quantification and included plate as predictor variable in the imputation model.

Sample analysis and pre-processing has been previously described ⁴ and is reiterated here. Samples were randomized and prepared by mixing a 30 μ L aliquot with 200 μ L of acetonitrile and filtering the precipitate with 0.2 μ m Captiva ND plates (Agilent

Technologies). The filtrate was collected into a polypropylene well plate that was sealed and kept refrigerated until analysis. A quality control (QC) sample was prepared by mixing small aliquots of 91 randomly selected study samples. Aliquots (30 μ L) of the QC sample were then processed along with the study samples, with each 96-well plate containing four individually extracted QCs. Samples were analyzed as a single uninterrupted batch with a UHPLC–MS system consisting of a 1290 Binary LC system, a Jet Stream electrospray ionization (ESI) source, and a 6550 QTOF mass spectrometer (Agilent Technologies). The autosampler tray was kept refrigerated, and 2 μ L of the sample solution was injected on an ACQUITY UPLC HSS T3 column (2.1 \times 100 mm, 1.8 μ m; Waters). Column temperature was 45 °C and mobile phase flow rate was 0.4 mL/min, consisting of ultrapure water and LC–MS-grade methanol, both containing 0.05% (v/v) of formic acid. The gradient profile was as follows: 0–6 min: 5% \rightarrow 100% methanol, 6–10.5 min: 100% methanol, 10.5–13 min: 5% methanol. The mass spectrometer was operated in positive polarity using the following conditions: drying gas (nitrogen) temperature 175 °C and flow 12 L/min, sheath gas temperature 350 °C and flow 11 L/min, nebulizer pressure 45 psi, capillary voltage 3500 V, nozzle voltage 300 V, and fragmentor voltage 175 V. Data were acquired in centroid format using an extended dynamic range mode, with a scan rate of 1.67 Hz and a mass range from 50 to 1000. For MS/MS analyses the isolation width was 1.3 Da and collision energies were 10, 20, and 40 V. Continuous mass axis calibration was performed using two reference ions (m/z 121.050873 and m/z 922.009798). The analytical run was initiated with priming injections of in-house human plasma extract to achieve a stable response, followed by the study samples and one QC sample after every 12 injections. Data were acquired using MassHunter Acquisition B.05.01 software.

Pre-processing of the acquired metabolomics data was performed using Qualitative Analysis B.06.00 SP1, DA Reprocessor, and Mass Profiler Professional 12.1 software

(Agilent Technologies). The initial processing was performed with Molecular Feature Extraction (MFE) algorithm for small molecules using a mass range of 50–1000. Thresholds for the mass and chromatographic peak heights were 1500 and 10 000 counts, respectively. Quality score threshold was 80. Only singly charged proton adducts ($[M+H]^+$) were included. Spacing tolerance for isotope peaks was 0.0025 m/z plus 7 ppm. The isotope model for common organic molecules was used, and features with indeterminate neutral mass were excluded. Feature alignment was performed with retention time and mass windows of 0.075 min and 15 ppm +2 mDa. A target list for a recursive extraction was created from features present in at least 2% of the samples. Find by Formula (FBF) algorithm was then employed with match tolerance for the mass and retention time ± 10 ppm and ± 0.04 min, respectively. Ion species was limited to $[M+H]^+$, and a threshold for chromatographic peak height was 2000 counts. The resulting features were aligned in Agilent Mass Profiler Professional using the same parameters as described above. For statistical analysis, metabolic features present in <60% of the samples were removed, data were log-transformed, and missing values were imputed using the impute.QRILC function within the impute.LCMD R package.(24)

Cross-omics Analysis:

Calibration of the sPLS models was done using fivefold cross-validation which was independently repeated 1000 times. Cross-validation procedure for the SPLS models were repeated for possible values of (i) the number of components to select and, (ii) the number of non-zero loadings coefficients (ie, the number of original variables contributing to the component). The number of components to be considered was determined using the average Q² statistic calculated across all folds and repeats and was defined as the maximal the number of components such that adding an additional component would yield a substantive drop in the Q² value. An example of a calibration plot is displayed in Figure S4. Sparsity of

the SPLS models was controlled by setting the number of variables included to the one minimising the cross-validated prediction error.

Table S1. Average level of exposure of each pollutant the year before birth, and posterior inclusion probability from the Bayesian Machine Kernal Regression models.

		3 pollutant model ^a	5 pollutant model ^b	5 pollutant model ^c	Sensitivity model 1 ^d	Sensitivity model 2 ^e	Sensitivity model 3 ^f
Air pollutant	Mean (Range)	Posterior inclusion probability					
PM _{2.5} (ug/m ³)	19 (9- 45)	0.5368	0.5882	0.5898	-	-	-
PM ₁₀ (ug/m ³)	35 (10- 45)	0.5906	0.5590	0.5486	-	-	-
NO ₂ (ug/m ³)	30 (8- 98)	0.7748	0.6824	0.7620	0.9880	-	0.8436
NO _x (ug/m ³)	73 (12- 236)	-	0.6474	-	-	0.7986	-
OP _{AA} (OP/m ³)	62 (-18- 131)	-	-	0.5044	-	-	0.5518
OP _{GSH} (OP/m ³)	5 (0.6- 14)	-	-	0.4964	-	-	0.4910
UFP (n/m ³)	13393.2 (1465- 30644)	-	0.5468	-	-	0.4366	-
PM _{coarse} (ug/m ³)		-	-	-	0.6182	0.4644	0.5560

Exposure levels represent the average level for each pollutant for the year before birth.

^aPM_{2.5}, PM₁₀, NO₂ (N=498)

^b3 pollutant + NO_x, UFP (N=299)

^c3 pollutant + OPAA, OPGSH (N=398)

^dPM_{coarse}, NO₂ (N=498)

^ePM_{coarse}, NO_x, UFP (N=299)

^fPM_{coarse}, NO₂, OPAA, OPGSH (N=398)

Table S2. Loadings coefficients for proteins identified from the SPLS_x models

Proteins	SPLS _x		
	Comp 1 (n=4)	Comp 2 (n=3)	Comp 3 (n=6)
TNF	0	0	0
IL8	0	0	0
EGF	0.075	-0.098	0.083
MIPI	0	0	0.026
GRO	0	0	0
IL6	0	0	-0.029
VEGF	0.53	0	-0.042
IL17	0.53	0	0
MCP1	0	0	0
GCSF	0	0	0
CRP	0	0	0.93
MPO	0.85	0	0
IP10	0	0	0.34
MDC	0	-0.63	0
Perios	-0.046	-0.78	0
ILRA	0	0	0

Table S3. Pathways and number of metabolites represented in each pathway identified from the 665 metabolites selected from the cross-omics assessments.

Pathway	Number of metabolites	p-value
Aspartate and asparagine metabolism	16	0.001
Arginine and Proline Metabolism	12	0.001
Tyrosine metabolism	11	0.001
Urea cycle/amino group metabolism	10	0.001
Glycerophospholipid metabolism	10	0.001
Fatty acid activation	9	0.001
Tryptophan metabolism	8	0.001
Arachidonic acid metabolism	8	0.001
Vitamin A (retinol) metabolism	8	0.001
Prostaglandin formation from arachidonate	8	0.001
Pyrimidine metabolism	7	0.001
Methionine and cysteine metabolism	6	0.001
Vitamin E metabolism	6	0.001
Linoleate metabolism	6	0.001
Purine metabolism	6	0.001
Glycine, serine, alanine and threonine metabolism	5	0.001
Glutathione Metabolism	5	0.001
Glutamate metabolism	5	0.001
Vitamin B3 (nicotinate and nicotinamide) metabolism	5	0.001
Alanine and Aspartate Metabolism	4	0.001

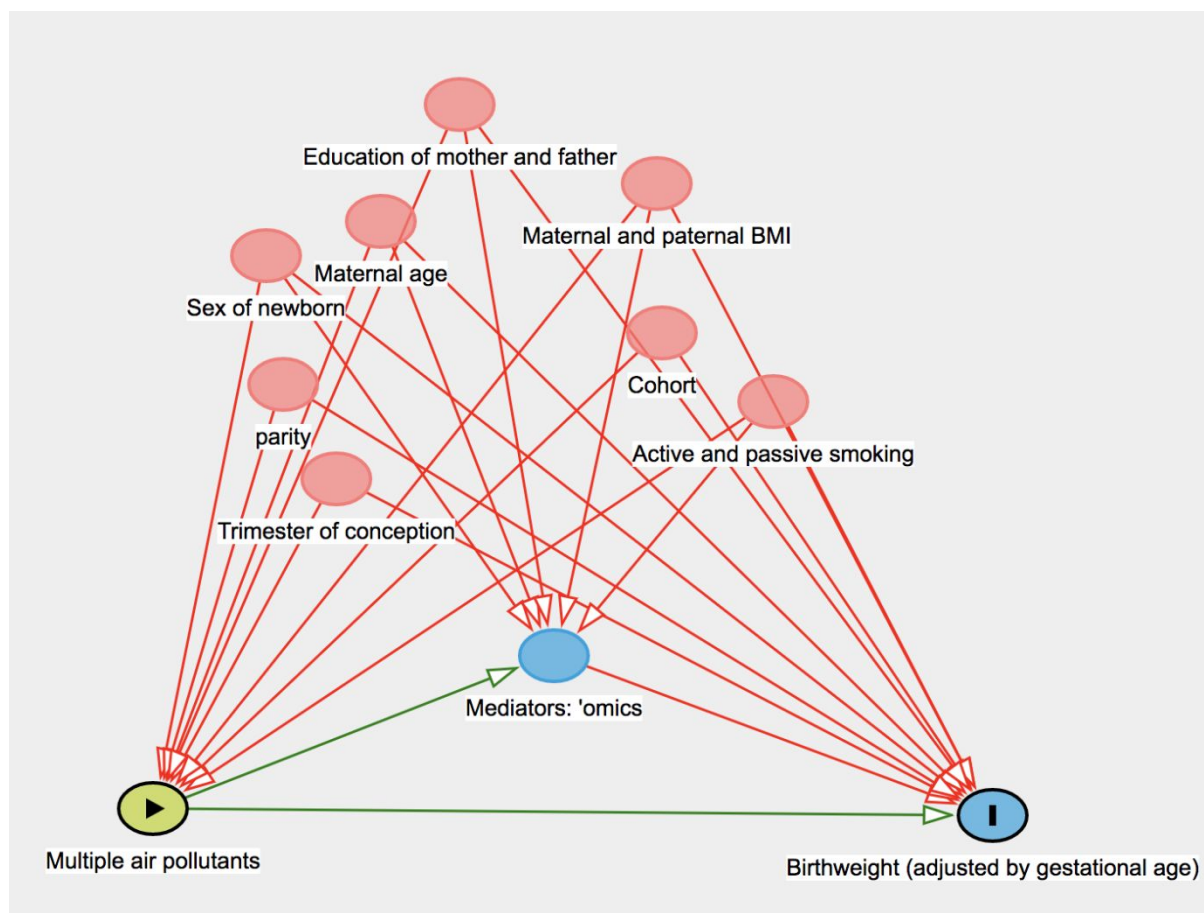


Figure S1. Directed Acyclic Graph to determine the set of potential confounders.

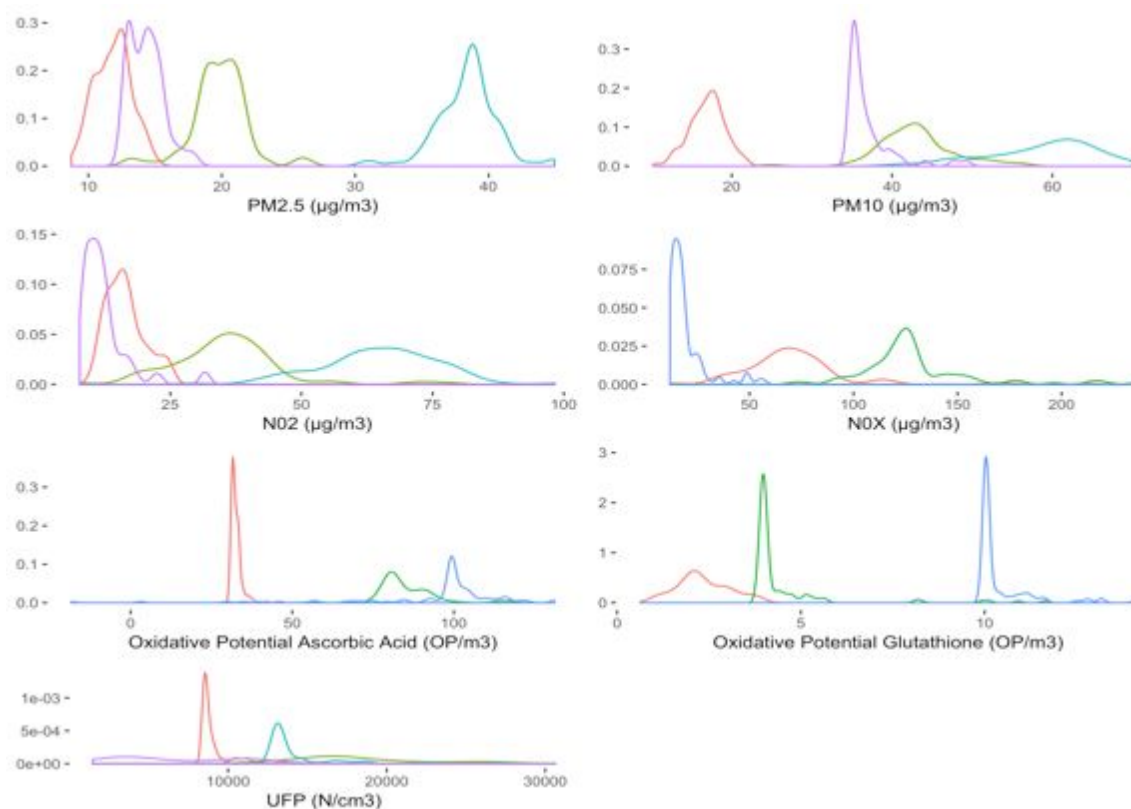


Figure S2. Levels of air pollutant by cohort (red lines represent ENVIRONAGE, green lines represent INMA, blue lines represent Piccolipiù, and purple lines represent Rhea).

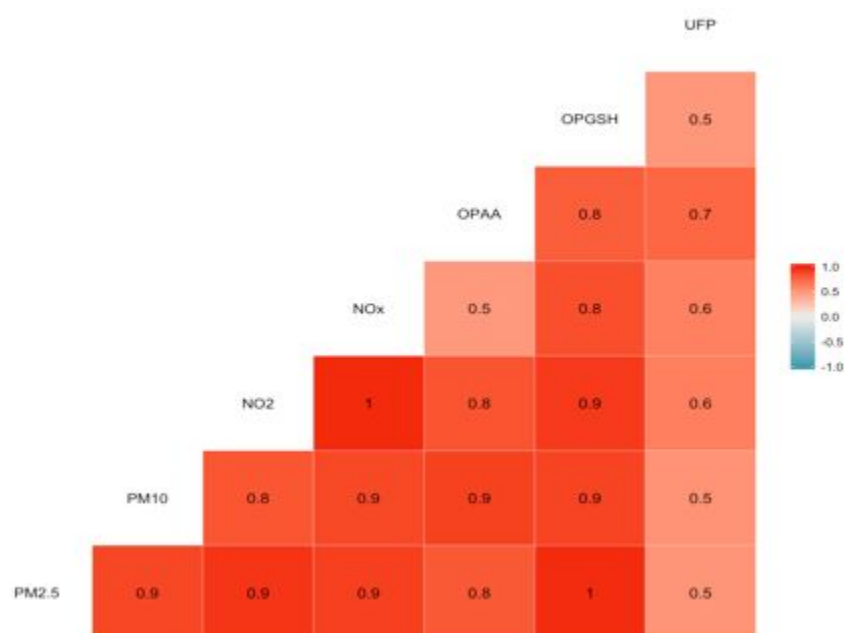


Figure S3. Correlation of air pollutants to one another, assessed using Pearsons correlation coefficient (r^2)

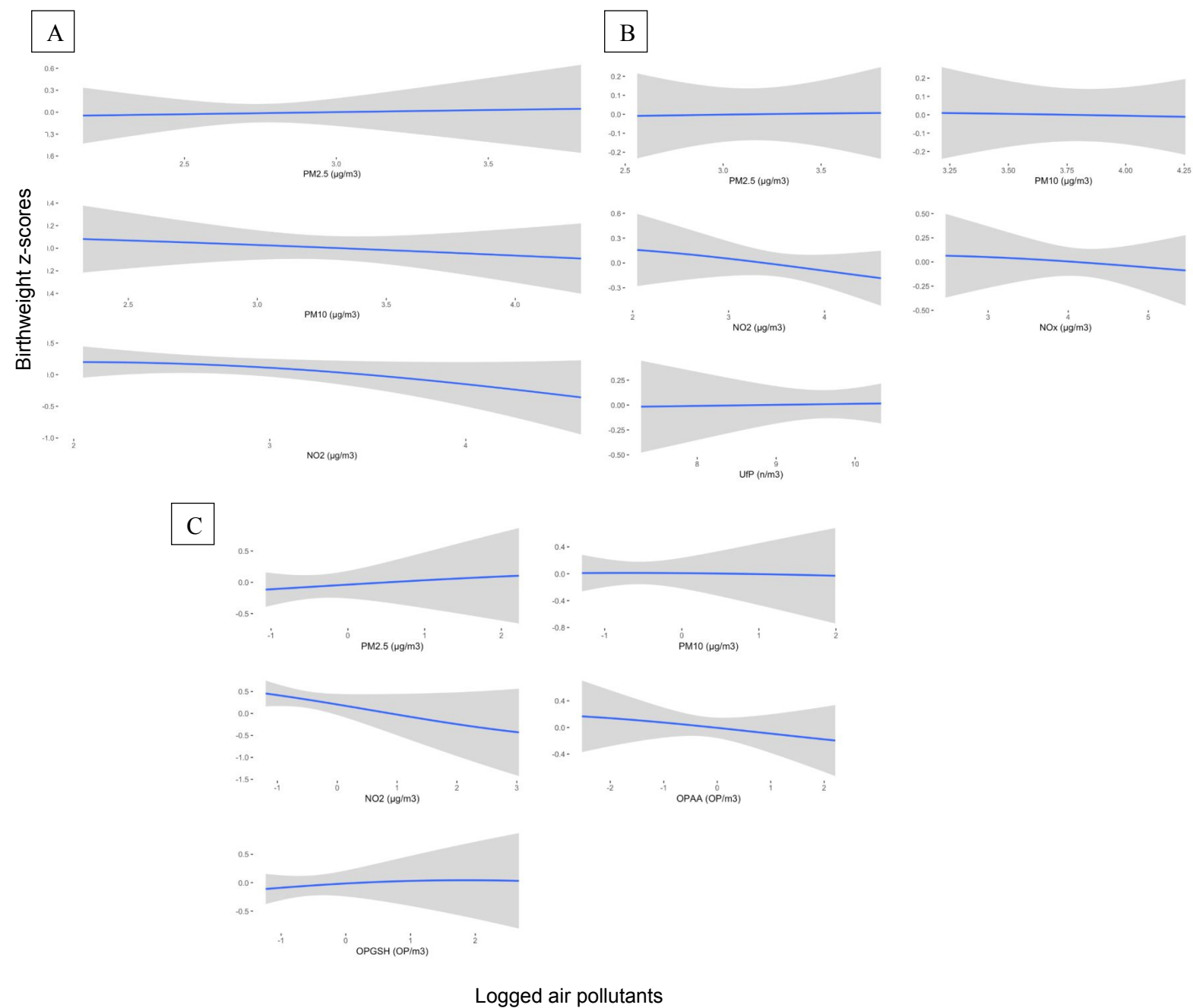


Figure S4. Cross-sectional univariate relationships between each air pollutant and birthweight z-scores, where other exposures are fixed to their median value. Panel A represents the three-pollutant model of PM_{2.5}, PM₁₀, and NO₂. Panel B represents the five-pollutant of PM_{2.5}, PM₁₀, and NO₂, NO_x, and UFP. Panel C represents the five-pollutant of PM_{2.5}, PM₁₀, NO₂, OP_{GSH}, and OP_{AA}.

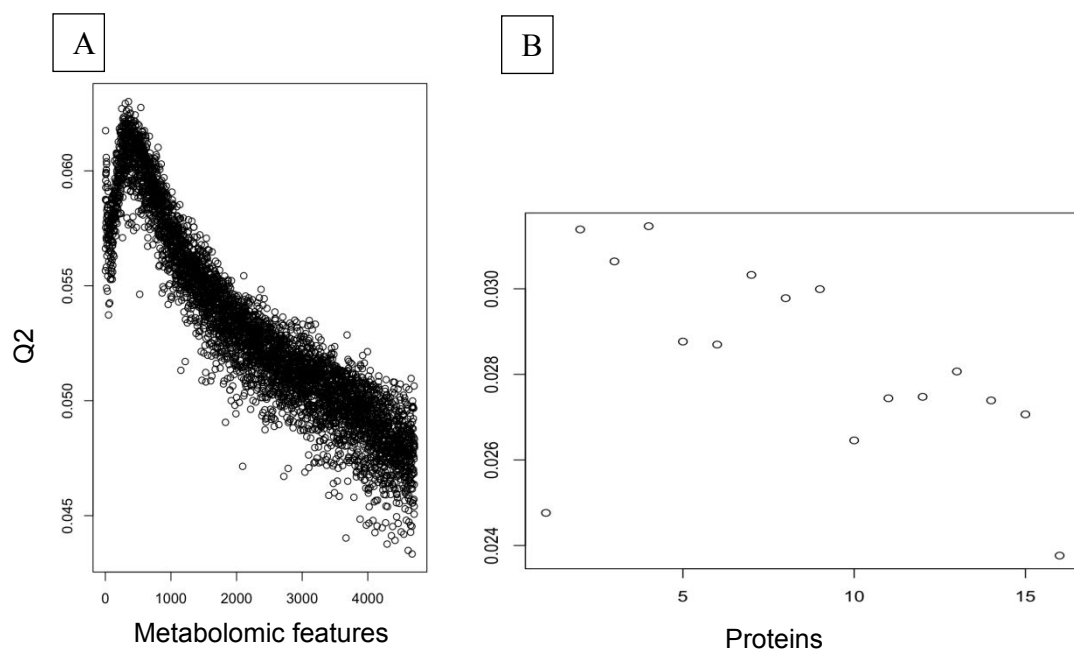
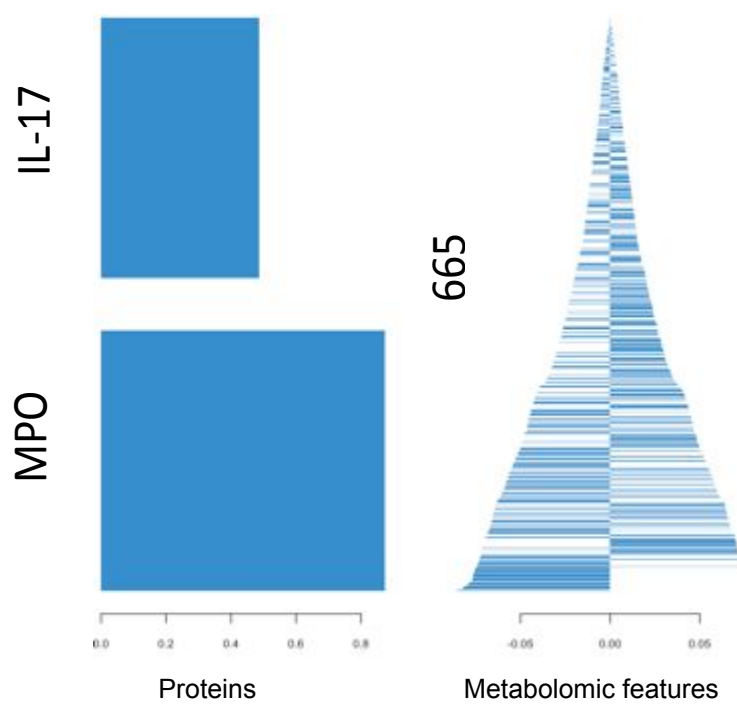


Figure S5. Calibration plot for the SPLS_{XY} model for A) metabolomic features B) proteins



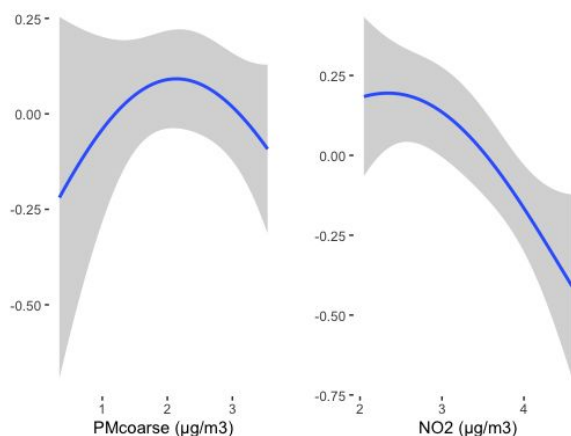
Loadings

Figure S6. Loading coefficients for the SPLS_{XY} model

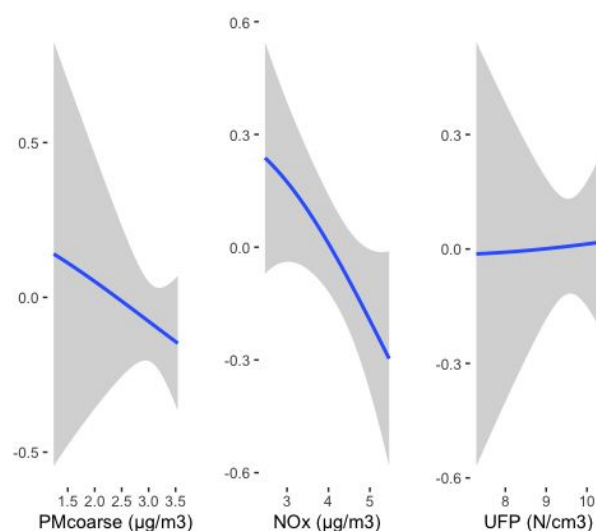
Sensitivity Analyses:

Birthweight z-scores

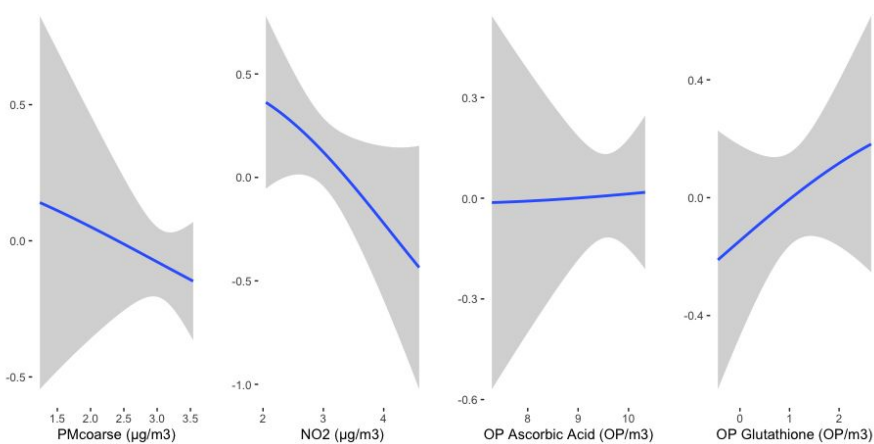
A



B



C



Logged air pollutants

Figure S7. Sensitivity results of cross-sectional univariate relationships between each air pollutant and birthweight z-scores, where other exposures are fixed to their median value. Panel

A represents a two-pollutant model of $\text{PM}_{\text{coarse}}$ and NO_2 . Panel B represents a three-pollutant model of $\text{PM}_{\text{coarse}}$, NO_x , and UFP. Panel C represents a four-pollutant model of $\text{PM}_{\text{coarse}}$, NO_2 , OP_{AA} , and OP_{GSH}

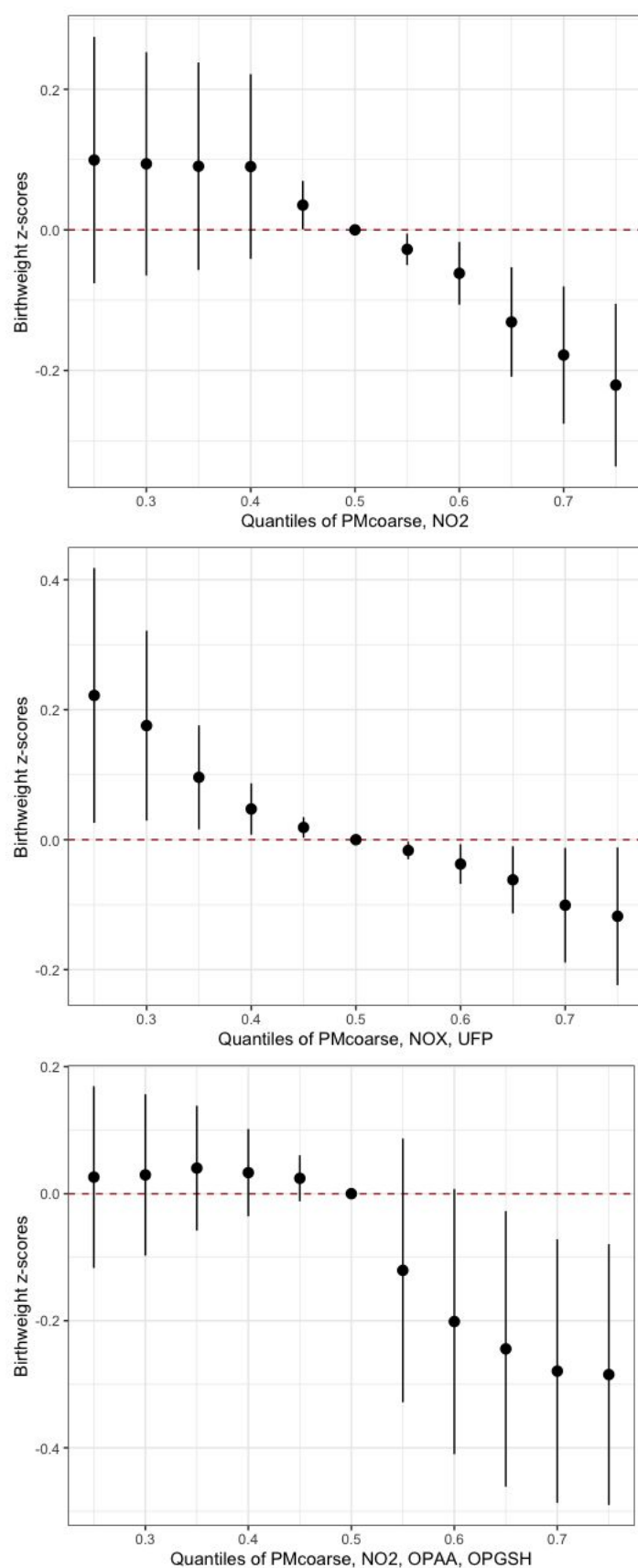


Figure S8. Results from the sensitivity analyses of the overall effects of mixtures of air pollutants on infant birthweight z-score. Estimates represent the predicted birthweight z-scores based on the overall effect of the mixture (estimates and 95% confidence intervals),

comparing birthweight z-scores when all exposures are at a particular quantile compared to the median value. Models adjusted for season of conception, sex, parity, maternal age, education of the mother and father, active and passive smoking during pregnancy, and maternal and paternal BMI and cohort.

References

1. van Nunen, E.; Vermeulen, R.; Tsai, M. Y.; Probst-Hensch, N.; Ineichen, A.; Davey, M.; Imboden, M.; Ducret-Stich, R.; Naccarati, A.; Raffaele, D.; Ranzi, A.; Ivaldi, C.; Galassi, C.; Nieuwenhuijsen, M.; Curto, A.; Donaire-Gonzalez, D.; Cirach, M.; Chatzi, L.; Kampouri, M.; Vlaanderen, J.; Meliefste, K.; Buijtenhuijs, D.; Brunekreef, B.; Morley, D.; Vineis, P.; Gulliver, J.; Hoek, G., Land Use Regression Models for Ultrafine Particles in Six European Areas. *Environ Sci Technol* **2017**, *51* (6), 3336-3345.
2. Gulliver, J.; Morley, D.; Dunster, C.; McCrea, A.; van Nunen, E.; Tsai, M. Y.; Probst-Hensch, N.; Eeftens, M.; Imboden, M.; Ducret-Stich, R.; Naccarati, A.; Galassi, C.; Ranzi, A.; Nieuwenhuijsen, M.; Curto, A.; Donaire-Gonzalez, D.; Cirach, M.; Vermeulen, R.; Vineis, P.; Hoek, G.; Kelly, F. J., Land use regression models for the oxidative potential of fine particles (PM. *Environ Res* **2018**, *160*, 247-255.
3. Lubin, J. H.; Colt, J. S.; Camann, D.; Davis, S.; Cerhan, J. R.; Severson, R. K.; Bernstein, L.; Hartge, P., Epidemiologic evaluation of measurement data in the presence of detection limits. *Environ Health Perspect* **2004**, *112* (17), 1691-6.
4. Robinson, O.; Keski-Rahkonen, P.; Chatzi, L.; Kogevinas, M.; Nawrot, T.; Pizzi, C.; Plusquin, M.; Richiardi, L.; Robinot, N.; Sunyer, J.; Vermeulen, R.; Vrijheid, M.; Vineis, P.; Scalbert, A.; Chadeau-Hyam, M., Cord Blood Metabolic Signatures of Birth Weight: A Population-Based Study. *J Proteome Res* **2018**, *17* (3), 1235-1247.