

Ecotoxicology and Human Environmental Health



ACS Publications
Urine metabolites
signatures of multiple

Subscriber access provided by Imperial

Copyright © American
College London | Library
Chemical Society.

However, no copyright

environmental pollutants in pregnancy & complications exposome approach



is published by the
American Chemical
Society, 1115 Sixteenth
Street N.W., Washington,
D.C. 20036

Published by American

Chemical Society

Subscriber access provided by Imperial

Copyright © American
College London | Library
Chemical Society.

However, no copyright

— Léa Maitre, Oliver
Robinson, David Martinez,
Mireille B Toledano, Jesus
Ibarluzea, Loreto Santa Marina,
Jordi Sunyer, Cristina Villanueva,



ACS Publications

Subscriber access provided by Imperial
College London | Library
Copyright © American
Chemical Society.

However, no copyright

Hector C. Keun, Martine
Vrijheid, and Muireann P Coen



Environ. Sci. Technol. Just

Accepted Manuscript • DOI

10.1021/acs.est.8b02215 •

Publication Date (Web): 04/09/2018

Subscriber access provided by Imperial

Copyright © American
College London | Library
Chemical Society.

However, no copyright

Downloaded from <http://pubs.acs.org> on October 5, 2018



Just Accepted

ACS Publications

is published by the
American Chemical
Society, 1155 Sixteenth
Street N.W., Washington,
DC 20036

Published by American
Chemical Society

Subscriber access provided by Imperial

College London | Library
Copyright © American
Chemical Society.

However, no copyright

“Just Accepted” manuscripts have been posted online prior to technical editing, formatting and proof correction. The Royal Society of Chemistry provides “Just Accepted” as a service to help disseminate scientific material as soon as possible in full in PDF format accompanied by an HTML abstract. The material is peer reviewed, but should not be considered final. Subscriber access to this article is limited to the Digital Object Identifier (DOI®). Just Accepted

is published by the
American Chemical
Society, 1155 Sixteenth
Street N.W., Washington,
DC 20036

Published by American
Chemical Society, Imperial
College London Library
Copyright © American
Chemical Society.

However, no copyright

the "Just Accepted" Web site may not include a manuscript is technically edited and formatted and published as an ASAP article. Notwithstanding to the manuscript text and/or graphics which appear in this journal, the American Chemical Society does not warrant the consequences arising from the use of information contained in these articles.



ACS Publications

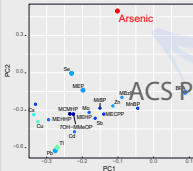
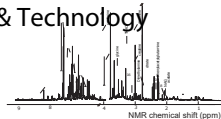
Subscriber access provided by Imperial College London | Library
Copyright © American Chemical Society.
However, no copyright

Environmental pollutants biomonitoring

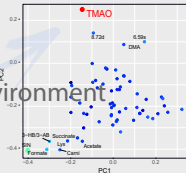
Arsenic
Lead
Phthalates
PCBs
Smoking and more...

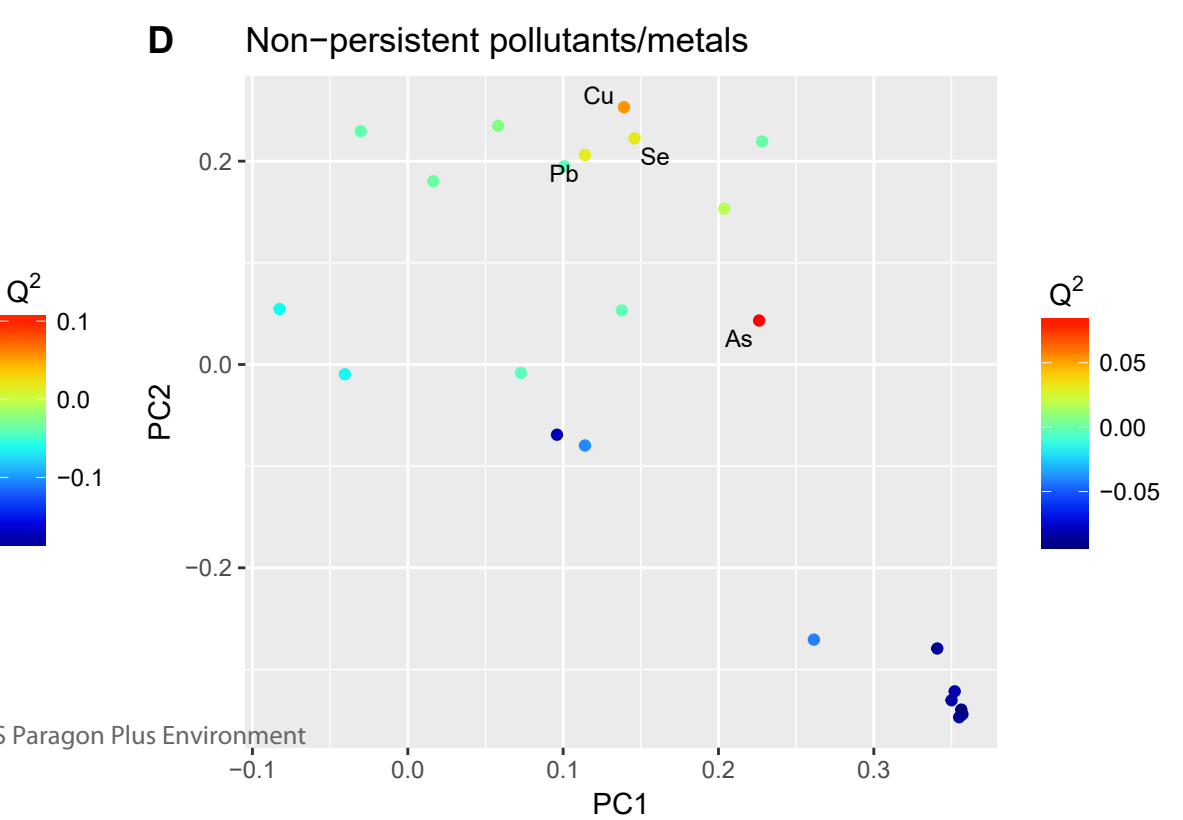
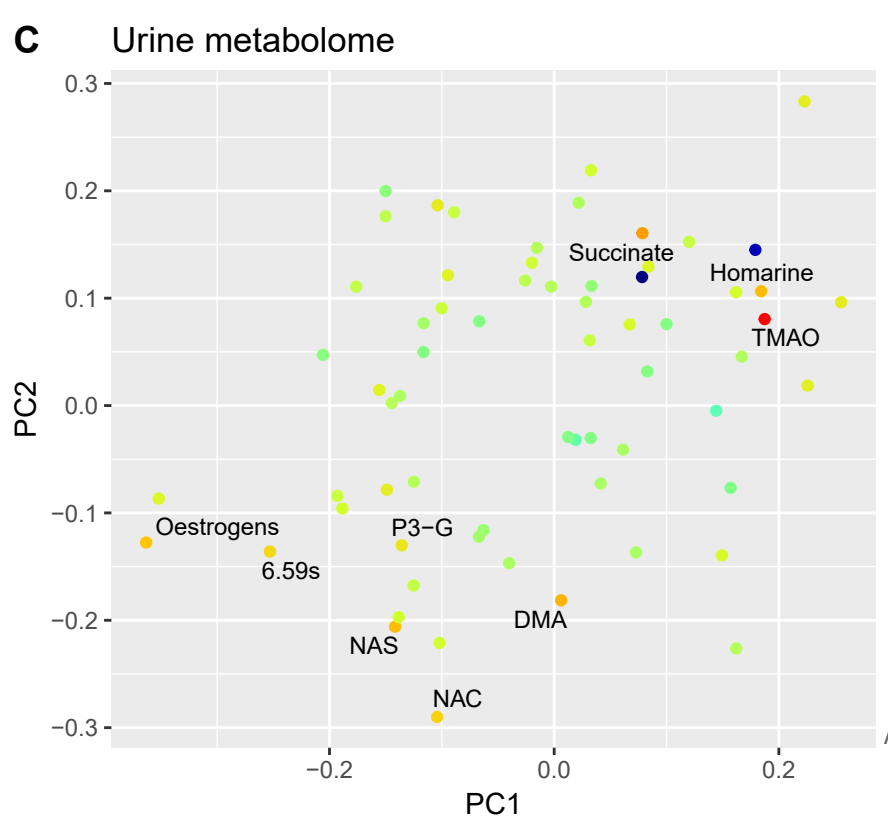
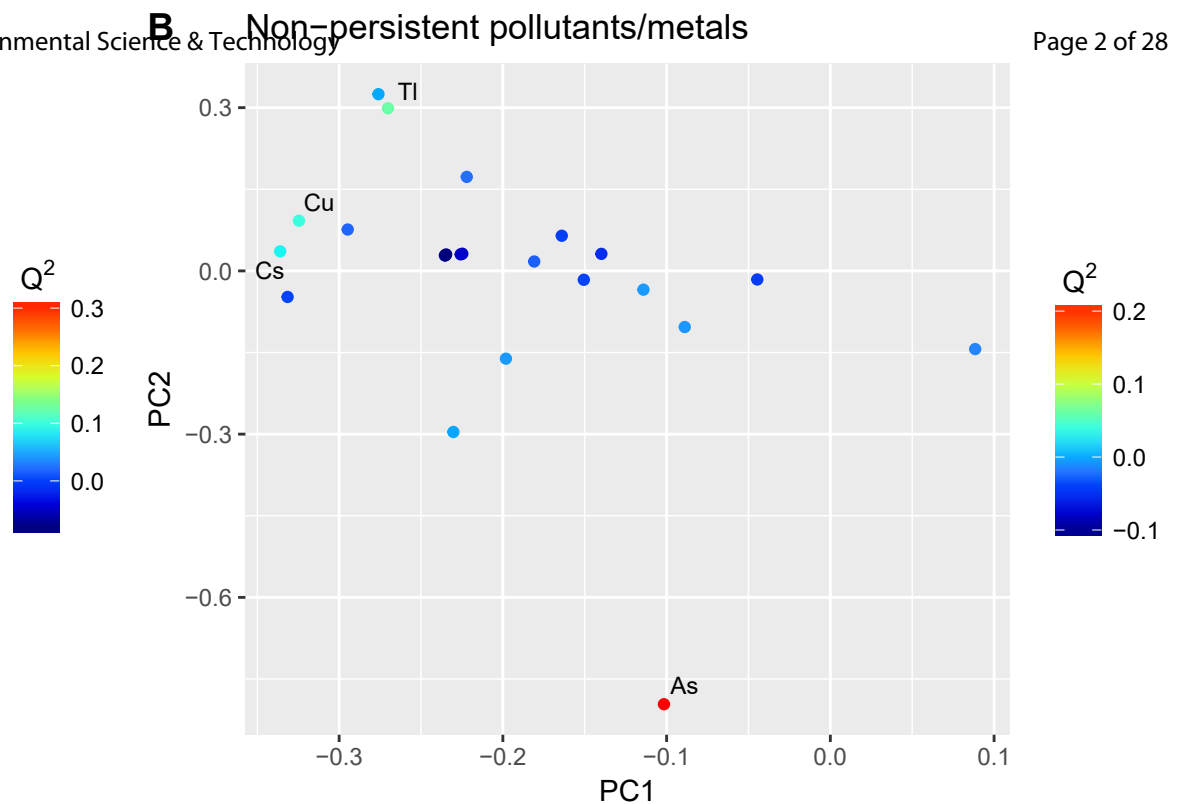
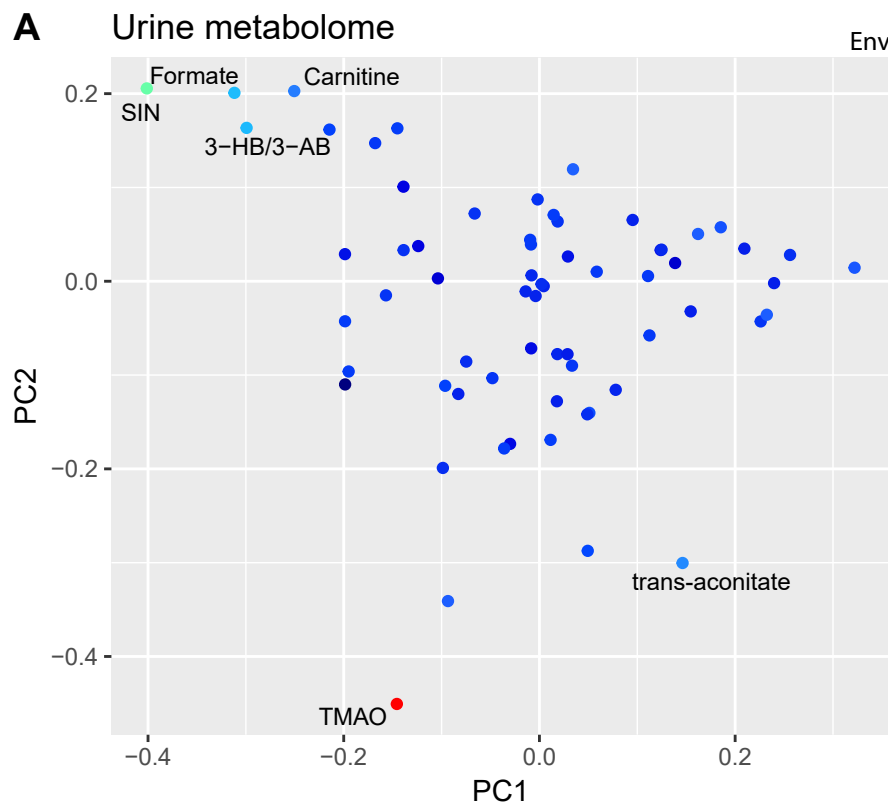


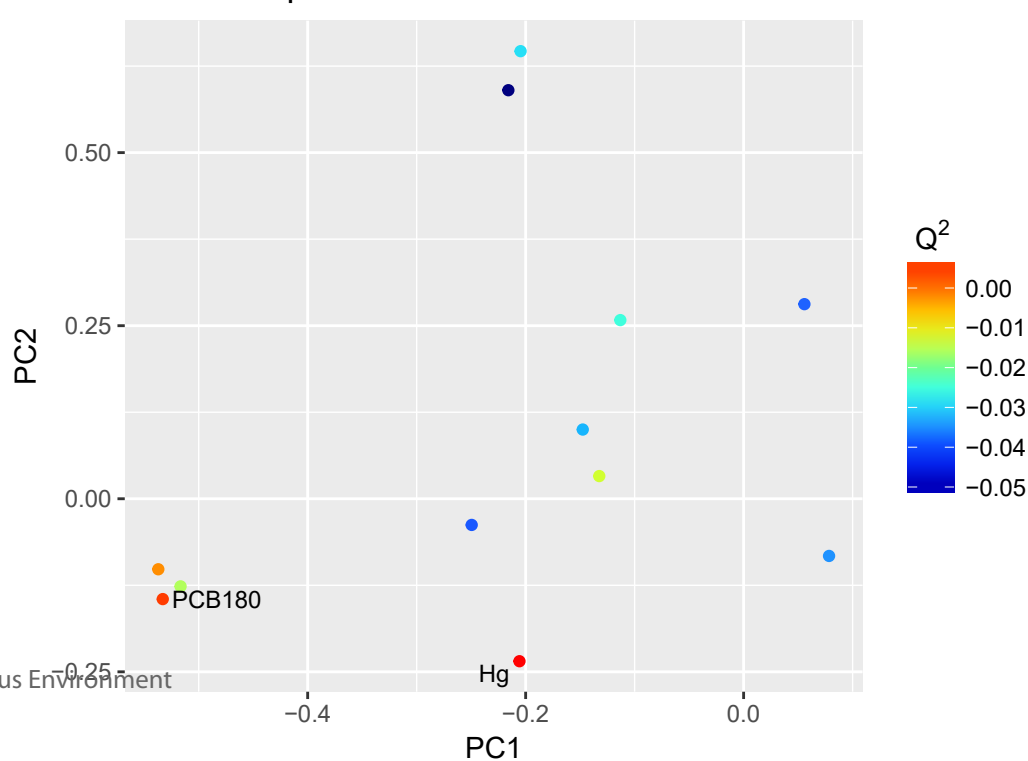
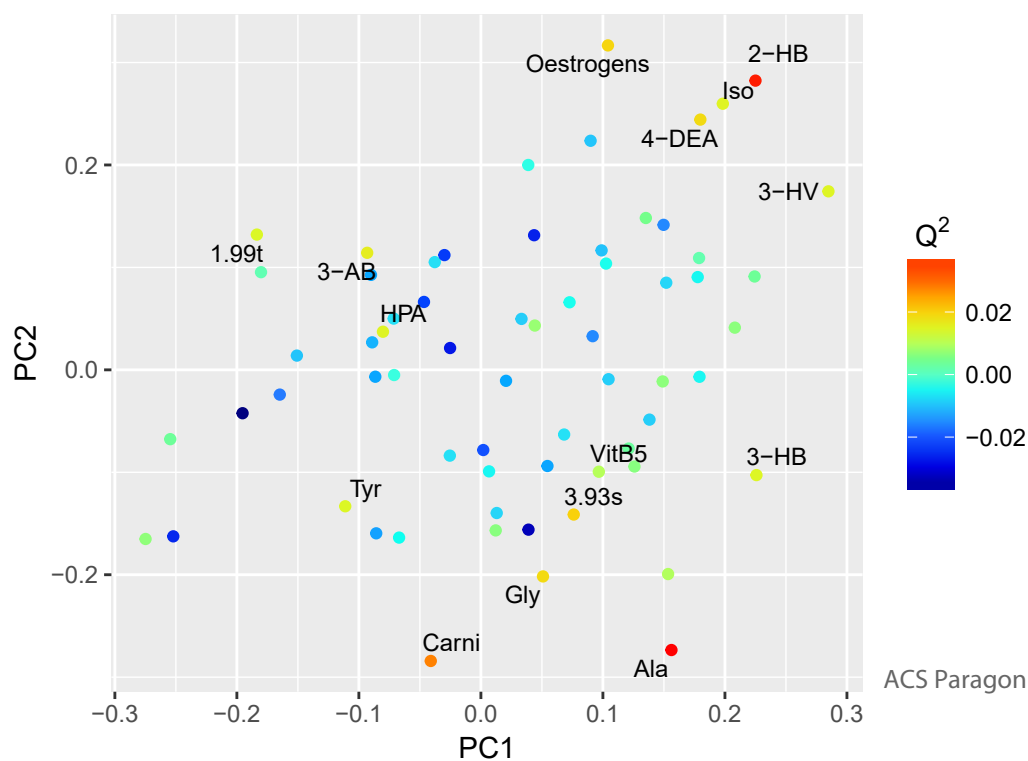
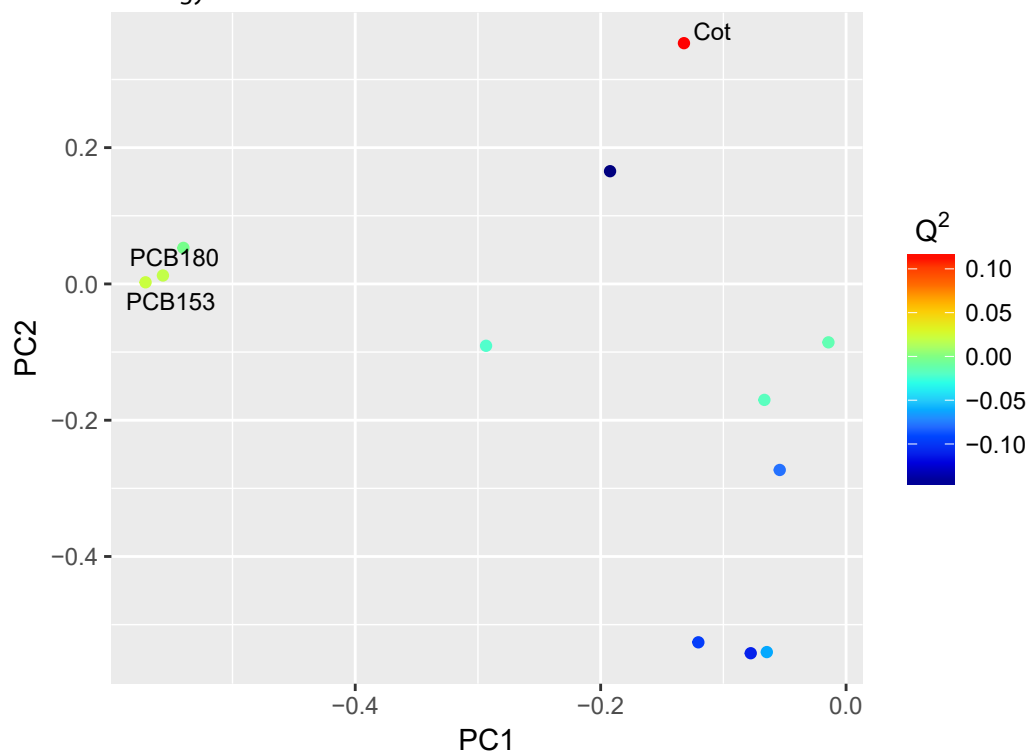
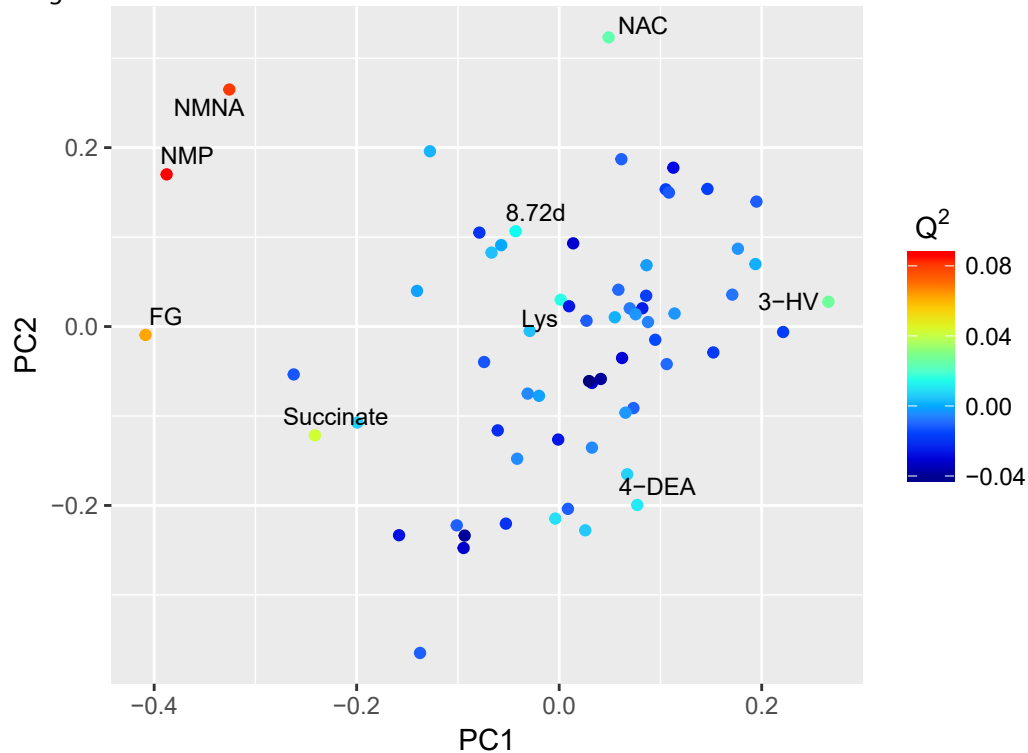
Urine endogenous metabolic footprint



An exposome-metabolome-wide association study







1 **Urine metabolic signatures of multiple environmental pollutants**

2 **in pregnant women - an exposome approach**

3 Léa Maitre*^{1,2,3}, Oliver Robinson^{1,2,3,4}, David Martinez^{1,2,3}, Mireille B. Toledano⁴, Jesús
 4 Ibarluzea^{3,5,6,7}, Loreto Santa Marina^{3,6,7}, Jordi Sunyer^{1,2,3,8}, Cristina M. Villanueva^{1,2,3}, Hector
 5 Keun⁹, Martine Vrijheid^{±1,2,3}, Muireann Coen^{±10,11}

6 § Shared first authorship

7 ±Shared last authorship

8 ***Corresponding author**

Léa Maitre, ISGlobal. Barcelona Institute for Global Health - Campus Mar, Parc de Recerca
 Biomèdica de Barcelona (PRBB), Doctor Aiguader, 88, 08003 Barcelona, Spain

Tel: +34 932147362; email: lea.maitre@isglobal.org

9 **AFFILIATIONS**

- 10 1. ISGlobal, Institute for Global Health, Barcelona, Spain
- 11 2. Universitat Pompeu Fabra (UPF), Barcelona, Spain
- 12 3. CIBER Epidemiología y Salud Pública (CIBERESP), Spain
- 13 4. Medical Research Council–Public Health England (MRC-PHE) Centre for Environment and
 14 Health, Department of Epidemiology and Biostatistics, , School of Public Health, Imperial
 15 College London, London, UK
- 16 5. School of Psychology, University of the Basque Country UPV/EHU, San Sebastian, Basque
 17 Country, Spain.
- 18 6. Biodonostia Health Research Institute, San Sebastian, Basque Country, Spain.
- 19 7. Sub-Directorate for Public Health of Gipuzkoa, Department of Health, Government of the
 20 Basque Country , San Sebastian, Basque Country, Spain
- 21 8. Municipal Institute of Medical Research (IMIM-Hospital del Mar), Barcelona, Spain
- 22 9. Division of Cancer, Department of Surgery & Cancer, Faculty of Medicine, Imperial College
 23 London, UK.
- 24 10. Integrative Systems Medicine & Digestive Disease, Department of Surgery & Cancer,
 25 Faculty of Medicine, Imperial College London, UK.
- 26 11. Discovery Safety, Drug Safety and Metabolism, IMED Biotech Unit, AstraZeneca, 1 Francis
 27 Crick Avenue, Cambridge CB2 0RE, UK

28

29 **Short Running title:** Metabolome – exposome association during pregnancy

30 Abstract

31
32 Exposure to environmental pollutants, particularly during pregnancy, can have adverse consequences
33 on child development but little is known about the effects of pollutant mixtures on endogenous
34 metabolism in pregnant women. We aimed to identify urinary metabolic signatures associated with
35 low level exposure to multiple environmental pollutants in pregnant women from the INMA (INfancia y
36 Medio Ambiente) birth cohort (Spain, N=750).
37 35 chemical exposures were quantified in first trimester blood samples (organochlorine pesticides,
38 PCBs, PFAS), in cord blood (mercury), and twice in urine at 12 and 32 weeks of pregnancy (metals,
39 phthalates, bisphenol A). ¹H nuclear magnetic resonance (NMR) metabolic profiles were acquired in
40 the same two urine samples. We explored associations between exposures and metabolism through
41 an exposome-metabolome wide association scan and multivariate O2PLS modelling.
42 Novel and reproducible associations were found across two periods of pregnancy for 3 non-persistent
43 pollutants and across two sub-cohorts for 4 of the persistent pollutants. We found novel metabolic
44 signatures associated with arsenic exposure: TMAO and dimethylamine possibly related to gut
45 microbial methylamine metabolism and homarine related to fish intake, of tobacco smoke exposure
46 related to coffee metabolism and PCBs with 3-hydroxyvaleric acid, usually released under
47 ketoacidosis.
48 These findings will have implications for further understanding of maternal-fetal health, and health
49 across the life-course.

50 Introduction

51 Pregnancy is a key period during which to assess the exposome and its effects on both maternal and
52 child health, due to the sensitivity of the developing fetus and the potential lifecourse impact¹.
53 Exposure during pregnancy to chemical environmental pollutants may adversely affect child
54 development, even at low exposure levels, with evidence classified as good² for associations between
55 polychlorinated biphenyls (PCBs) and impaired fetal growth³ and between lead, methylmercury, PCBs
56 and organophosphate pesticides and neurodevelopment^{4–7}. For other ubiquitous pollutants, such as
57 perfluoroalkyl substances (PFASs) and bisphenol A (BPA), evidence of associations with
58 developmental effects has been classed as moderate or insufficient². The exposome concept
59 advocates systematic assessment of the totality of human environmental (i.e., nongenetic) exposures
60 from conception onward including both endogenous and exogenous exposure domains⁸. It holds
61 promise in improving the evidence base in this area. Measurement of multiple chemical exposures
62 allows consideration of the complex chemical environment to which individuals are exposed and
63 explicit correction for multiple testing when assessing associations with health effects⁹. In addition,
64 through linkage of multiple exposures to holistic measures of physiological processes, biological
65 pathways of effect can be uncovered. This may identify common pathways through which diverse
66 pollutant classes can act in concert on health¹⁰.and provide a mechanistic link between exposure and

developmental endpoints, strengthening the evidence for causality through the 'meet-the middle' approach¹¹.

Experimental studies have shown that endocrine disruptors such as dichlorodiphenyltrichloroethane (DDT), arsenic, BPA and phthalates can disrupt metabolism through processes such as steroid hormone interaction¹². It is therefore hypothesised that environmental pollutants may impair development through disruption of normal metabolic processes during pregnancy¹³, and as such may be expected to have a detectable influence on metabolic profiles measured in biological samples in pregnant women. For instance, work in animal models have found striking metabolic differences in offspring of exposed pregnant dams to BPA and a phthalate using NMR metabolomics^{14,15}. Others specifically looking at the urinary metabolome of exposed mice themselves have found more modest changes when exposed to BPA which were supported by similar changes observed at higher doses or enzymatic perturbation^{16,17}. A handful of studies have looked at the effects of environmental pollutants on the metabolome in humans. Cadmium exposure among the general UK population was associated with metabolic perturbations identified through ¹H NMR spectroscopic analysis of urine samples, demonstrating the utility of the platform in this context¹⁸. Associations between arsenic exposure and the metabolome have been reported in urine of Chinese males^{19,20}, and in cord blood of neonates²¹. High-resolution metabolomic platforms have been employed in studies of chlorinated by-product exposures and were able to simultaneously identify exogenous halogenated exposures and associated endogenous metabolic perturbations^{22,23}.

In this study, we sought to test the hypothesis that low-level environmental contaminants have detectable imprint on the metabolism of pregnant women, including pathways related to steroid hormone by-products, oxidative stress, lipid transport and oxidation. We have used data available among pregnant women enrolled in the INMA (INfancia y Medio Ambiente) population based birth cohort study, which aims to examine the role of environmental pollutants during pregnancy and early childhood in relation to child growth and development²⁴. INMA has previously assessed both the exposome through multiple exogenous chemicals measured in biological samples²⁵ and the metabolome through urinary ¹H NMR spectroscopy²⁶. Employing an agnostic exposome-wide association study approach, we aimed to link the exposome and metabolome domains to elucidate specific metabolic pathways that may be perturbed by exogenous exposure, thereby improving knowledge of the biological mechanisms through which exposures may exert their pathogenic effects. We performed separate analyses in two INMA study centres, Sabadell and Gipuzkoa, and at two time points during pregnancy, to assess the reproducibility of our findings.

Methods

Study population

Among the seven cohorts in the INMA study, the Sabadell subcohort from the Mediterranean area (Catalonia, 2004–2008) and the Gipuzkoa cohort from the Atlantic area (Basque Country, 2006–2008)

were selected based on sample and exposure data availability. In these two INMA sub-cohorts, pregnant women from the general population were recruited at the first trimester routine antenatal care visit in the main public hospital or health centre of reference, using the following inclusion criteria: women had to be at least 16 years old, intend to deliver in the reference hospital, have a singleton pregnancy with no assisted conception, and be able to read and write in Spanish or Catalan. A full description of the INMA project protocol and study characteristics have been previously described²⁴.

Women provided socio-demographic, medical history and lifestyle information and completed a validated food frequency questionnaire²⁷ during interviews with trained fieldworkers at the first (12.4±1.2 weeks pregnancy) and third (33.9±1.3 weeks) trimester visits. Spot urine samples were collected at both visits and blood samples were collected at the first trimester visit. Cord blood samples were collected at birth. The study was conducted with the approval of local hospital ethics committees, and written informed consent was obtained from all women. A total of 1500 urine samples were available, 410 for each 1st and 3rd trimesters for Gipuzkoa and 340 for each 1st and 3rd trimesters for Sabadell.

Exposure assessment

Concentrations of 35 different exposures were available, with additional description of methods and references available elsewhere²⁵. Exposures were selected on the basis of availability from current or ongoing INMA studies. Previous INMA studies have targeted exposures that are of current health concern and widely distributed within the population.

In the INMA Sabadell subcohort, pollutants and metals (in total 23 exposures) at two time points in 1st and 3rd trimester urine samples were available: BPA, phthalate metabolites (mono(2-ethyl-5-carboxypentyl) phthalate (MECPP), Monobenzyl phthalate (MBzP), Mono-2-ethyl-5-hydroxyhexyl phthalate (MEHHP), mono-2-ethylhexyl phthalate (MEHP), Mono-2-ethyl-5-oxohexyl phthalate (MEOHP), Monoethyl phthalate (MEP), Di-iso-butyl phthalate (DiBP) Mono-isobutyl phthalate (MiBP), Mono-*N*-butyl phthalate (MnBP), Mono-(4-methyl-7- hydroxy-octyl) phthalate (7OH-mMeOP), Mono[2-(carboxymethyl)hexyl] phthalate (MCMHP)) and metalloids (arsenic (As), cadmium (Cd), cobalt (Co), caesium (Cs), copper (Cu), molybdenum (Mo), nickel (Ni), lead (Pb), antimony (Sb), selenium (Se), thallium (Tl), zinc (Zn)).

In Sabadell and Gipuzkoa, the following persistent pollutants (in total 12) were available at one time point: organochlorines (b-hexachlorocyclohexane (bHCH), dichlorodiphenyldichloroethylene (DDE), hexachlorobenzene (HCB), PCB congeners 138, 153 and 180) and PFASs (perfluorohexane sulfonate (PFHxS), perfluorononanoic acid (PFNA), perfluorooctanoic acid (PFOA) and perfluorooctanesulphonate (PFOS)) in 1st trimester serum samples.. In addition, cotinine was measured in 3rd trimester urine samples and mercury was measured in cord blood samples. In the following sections, cotinine was classified as a persistent exposure since the habit (smoking) is often persistent although its half-life is only of a few days.

Organochlorine measurements were normalized to total serum lipid levels and measurements in urine were normalized to creatinine concentration (measured through a standard clinical chemistry assay) to correct for urine dilution.

NMR metabonomics

Urinary metabolic profiles were acquired twice during pregnancy (first and third trimesters) to capture potential changes in metabolism. Profiles were generated by ^1H NMR spectroscopy at 600 MHz²⁸ as fully described in **SI Methods**, and corrected for differential dilution of spot urine samples using probabilistic quotient normalisation²⁹. Relative concentrations of 65 metabolites, 54 of which were structurally identified, were obtained after manual spectral binning (integration within a fixed chemical shift window). These metabolites, represented by spectral integrals of single representative resonances, were selected on the basis of being present in a high proportion of the spectra, having a high signal-to-noise ratio, and exhibiting limited overlap with other resonances. A list of the 65 metabolites, their integration window, their metabolic pathways and details on assignments are presented in **Table S1**.

Statistical analysis

Treatment of missings in exposure variables

There were missing values in the exposure variables due to sample never been analysed ranging from 3.8% (arsenic and cotinine) to 38 and 49% (perfluorohexane sulfonate, in Sabadell and Gipuzkoa respectively) of participants. Only 16 and 17% (1st and 3rd trimesters respectively) of the Sabadell participants had complete data on all 23 non-persistent exposure variables and 45% of the participants in both subcohorts had complete data on all 12 persistent exposure variables. Overall, 18.0% of exposure data values were missing (4436 data points out of the 24640 total values for 23 non-persistent chemicals $\times$ 410 women $\times$ two time points and 12 persistent chemicals $\times$ 750 women). Therefore, to evaluate multiple pollutant exposures in one model, we adapted the multiple imputation approach to impute the missing exposure values³⁰ and selected at random one unique dataset. To determine which of the variables to include as predictors in the imputation process, we evaluated correlation, *t*-test, or chi-square test with each one of the imputed variables, as appropriate. We used imputation models that were more general than the analysis models and included the relative metabolite concentrations, the variables related to the missingness, and auxiliary variables that were associated with the exposure. The same approach was applied to impute missing values for the model covariates. Missing values due to values being under the limit of detection were imputed using the minimum value of the exposure variable divided by 2.

Identifying exposure-metabolite association pairs

We analyzed 1st and 3rd trimester samples separately and Gipuzkoa and Sabadell samples separately and compared associations between population and time points. Both cross-sectional and longitudinal associations were investigated (i.e. 1st trimester exposure with 1st trimester metabolic profile and 1st trimester exposure with 3rd trimester metabolic profile). We initially used an 'exposome-metabolome

wide association study' (EMWAS) approach by computing multiple pairwise correlations between exposures and metabolites. Spearman's partial correlations, adjusted for the time of the day of sampling (non-persistent analysis only), gestational week, age and body mass index (BMI), were calculated in the basic adjustment set. Prior to partial correlation analyses variables were log-transformed to give a normal distribution. EMWAS level statistical significance was set to 5% for adjusted p-value ("q-values" adjusted for 5% false discovery rate using the package "fdrtool" in R)^{31,32}.

Greater confidence was given to observed associations passing FDR significance at two timepoints or in both sub-cohorts.

Mapping the pregnancy internal exposures by urinary metabolic profiles: a pattern recognition approach

Relative concentrations of 65 urinary metabolites were used in the analysis. O2PLS is an extension of a supervised multivariate analysis method—partial least squares—with an integrated orthogonal signal correction filter that allows prediction in both directions between multivariate matrices X and Y ^{33,34}. The exposure variables and metabolite data were mean centred and univariate-scaled. Sevenfold cross-validation was performed to calculate Q^2 -values. Like with the loadings from PCA, the loadings from O2PLS are interpreted in the same way, particularly, we look for clusters, outliers and interesting patterns in the line plots. Highly correlated exposure or metabolites have similar weights in the loading vectors and appear close together or in the same section in the loading plots of all dimensions (PCs). In addition, the individual variable Q^2 (colour scale in Figures 1 and 2) is interpreted as the strength of correlation/model predictability between the two blocks of data (urine metabolome and exposures). Overall statistical parameters are interpreted as follow: R^2X is the variation in the metabolome (X) explained by the model, the rest represents noise, R^2X_{corr} is the proportion of R^2X correlated to the exposome (joint variance). In the results, we discuss the product of R^2X and R^2X_{corr} . The same interpretation applies to R^2Y and R^2Y_{corr} which represent the variation in the exposures (Y). $Q^2X(cum)$ and $Q^2Y(cum)$ represent the overall predictive ability of the O2PLS model. Sample permutation ($n=1000$) was performed to estimate the significance of the obtained Q^2 -levels. By randomly permuting the metabolic profiles and exposures and generating 'null' models (Supplementary Figures S1A–D), we were able to identify from the model parameters (individual variable Q^2) a subset of exposures and metabolites that were significantly correlated. Analysis was performed in the R statistical environment using the open-source package ("pcaMethods") and an in-house script to perform O2PLS. Interpretation of the model parameters and plots were based on two publications^{35,36}.

Basic adjusted model comparison with lifestyle and clinical factors adjusted model.

We evaluated the change in the magnitude of the exposure-metabolite association due to the inclusion of lifestyle and clinical factors in the model, by comparing the estimated effects (p) of the basic adjusted partial correlations with those of the partial correlations including additional lifestyle and clinical factors. This was quantified by the percentage attenuation. Lifestyle and clinical factors

considered included characteristics known to affect endogenous or exogenous metabolism, in particular during pregnancy or that may be an intermediate factor, including: active tobacco smoking at week 12, total gestational weight gain, intake of coffee/tea, total seafood, meat, dairy food, dietary fats, folic acid, vitamin B12. Additional covariates were added to the basic adjusted correlations one-by-one. 95% confidence intervals were calculated for the percentage attenuation³⁷.

Results

A total of 1500 ¹H NMR spectra of urine were generated from a total of 750 women (340 in Sabadell and 410 in Gipuzkoa cohorts). Women were included in this analysis only if they had NMR metabolomics data acquired at both 1st and 3rd trimesters. Mothers had a median age (in Sabadell-Gipuzkoa respectively) of 30-31, were mainly born in Spain (90-97%) and reached secondary (42-35%) or university education (31-54%) in most cases (**Table S2**). **Table S3** and **S4** present the non-persistent/metal and persistent pollutant levels, respectively, in each sub-cohort for each trimester available. Median levels of non-persistent pollutants and metals in Sabadell population were similar in the 1st and 3rd trimesters, except for the following metals: cobalt, caesium, copper and molybdenum, as previously described³⁸. However, the 1st-3rd trimester correlations, indicating intra-individual variability during pregnancy, were low for BPA, MEHP and MEOHP ($p < 0.15$) and relatively high for metals including cadmium and zinc ($p > 0.5$), indicating a low intra-individual variability across trimesters. Differences in the levels of persistent chemicals between the two cohorts were observed with higher levels of β -HCH, mercury and PCBs in Gipuzkoa in comparison with Sabadell but lower levels of DDE, HCB and PFAS. Cotinine levels, related to both active and passive smoking habits, were in line with reported smoking level via questionnaire with 29-21% of the population smoking at least 1 to 10 cigarettes a day. Dietary patterns were also compared and Gipuzkoa women reported a less calorific and less fatty diet with a higher consumption of large oily fish and white fish but a lower consumption of shellfish in comparison to Sabadell women. Data on organic food or other dietary behaviours were not available.

Identifying exposure-metabolite association pairs

Non-persistent pollutants and metals. 81 and 71 cross-sectional associations were identified as significant (out of 1495 associations tested, after adjustment for multiple testing using FDR, the time of sampling, gestational week, maternal BMI and age) in the first and third trimesters respectively (**Figure S1**, **Table 1**). Of these, 19 associations were replicated in both trimesters and summarised in table 3 including the strongest ones for arsenic with trimethylamine oxide (TMAO, Spearman's rho ($\rho_{1^{st}}$)=0.36, $p=8.60E-12$ and $\rho_{3^{rd}}$ =0.25, $p=2.9E-06$), dimethylamine ($\rho_{1^{st}}$)=0.23, $p=2.2E-05$ and $\rho_{3^{rd}}$ =0.19, $p=4.7E-04$) and homarine ($\rho_{1^{st}}$)=0.25, $p=3.6E-06$ and $\rho_{3^{rd}}$ =0.29, $p=4.5E-08$). However, these associations were not observed longitudinally (1st trimester exposure with 3rd trimester urine metabolome). Four heavy metals, thallium, caesium, copper and lead presented similar metabolic association patterns including positive associations with scyllo-inositol, acetate, formate, dimethylamine and negative association with *N*-acetylserotonin (tentative assignment), *N*-acetylneuraminic acid and *trans*-aconitate. Several of these associations were also observed longitudinally, in particular for thallium with *N*-acetylserotonin and *N*-acetylneuraminic acid and caesium with 3-

hydroxyisovalerate. Lead, cadmium, thallium and copper also showed trimester specific associations with steroid hormones (**Table S5**). Positive associations with pregnanolone-3-glucuronide were particularly strong in the first trimester (from $p=0.27$ for thallium to 0.17 for copper, $p=5.9E-07$ to $1.3E-03$) and negative associations with oestrogen metabolites in the third trimester ($p=-0.28$ for thallium to -0.22 for caesium and cadmium, $p=1.6E-07$ to $4.5E-05$). These associations between thallium and copper and oestrogens were also observed in longitudinal associations between 1st trimester exposure measurement and 3rd trimester oestrogens. As a general pattern, all phthalates were negatively associated with kreb's cycle metabolites, succinate and citrate and the organic acid acetate, with stronger associations in the first trimester. No significant associations were found with BPA.

Persistent pollutants. Among this set of exposures which was measured only once during pregnancy, 5 metabolic associations were statistically significant in the first trimester and 12 in the third trimester in the Sabadell cohort (out of 780 associations tested, all the exposures were measured in the first trimester except for cotinine, in the third trimester, and mercury, in cord blood at birth) after adjustment for gestational week, maternal BMI, age and multiple testing using FDR (**Figure S2, Table 2**). Five associations in the third trimester were replicated in the Gizpukoa cohort after FDR correction (summarised in Table 3), including for PCBs 153 and 180 with 3-hydroxyisovalerate (p between -0.18 and -0.22). Mercury was associated with taurine in the third trimester ($p=0.21$) and cotinine was associated with *N*¹-methylnicotinic acid and *N*-methylpyridinium (p between 0.20 and 0.22). All these associations were also observed in the first trimester with at least $p<0.05$ (highlighted in red in Table 2). Additional patterns of associations were observed, although not common in both cohorts at FDR levels. 4-deoxyerythronic acid was negatively associated with all three PCBs in the first trimester (p between -0.20 and -0.22). This association was replicated in Gipuzkoa without reaching FDR threshold (p -values= 0.002 - 0.03). Cotinine was associated with lower citrate in both trimesters and cohorts but passing multiple testing correction only in the first trimester in Sabadell ($p=-0.22$). Cotinine was also associated to a lesser extent to furoylglycine ($p=0.19$) in the third trimester. Mercury was consistently associated with decreased oestrogen metabolites ($p=-0.16$) and positively associated with TMAO in both cohorts and at both trimesters with $p<0.05$ ($p=0.11$ - 0.2). Chlorinated pesticides, including bHCH, DDE and DDT did not associate significantly with urine metabolites. and perfluorinated compound (PFHxS, PFNA and PFOA) associations could not be replicated.

Table 3 displays those associations that were FDR-significant in both trimesters for the non-persistent pollutant analysis or in both sub-cohorts for persistent pollutant analysis. In Supplementary figures S3 and S4 are presented the complete case analysis for comparison (non-imputed datasets) which display very similar results.

Mapping the pregnancy internal exposures by urinary metabolic profiles: a pattern recognition approach

O2PLS models were constructed to identify patterns of co-variation between urinary metabolites and non-persistent pollutant levels complementing the first approach exploring exposure-metabolite association one by one and free of multicollinearity issue. For example, there is evidence from the results of the first approach that several metals affect the same group of metabolites and vice versa that groups of metabolites are modified in concert as the consequence of exposures. Statistics for all O2PLS models are shown in **Table S6**.

Non-persistent pollutants. Overall, in the first trimester, 3.4% of the variation in the metabolome modelled was correlated to the exposome and 24.3% of the exposome variation modelled was correlated to the metabolome. Although the overall predictive ability of the model was close to zero, the model performed significantly better than random models through permutation validation ($p=0.001$, **Figure S3-6**). The following patterns were identified based on the position of the variables on the loading plots and the strength of the correlation Q^2 (red indicate a strong correlation): TMAO and arsenic had both large estimated loadings (on the second principal component) in the first trimester and the highest correlation Q^2 (TMAO: -0.45 PC2, $Q^2=0.32$ and Arsenic: -0.80 PC2, $Q^2=0.23$), as observed in **Figure 1.A and B**, confirming that this is the strongest association (both $p<0.001$) in these datasets. Thallium, copper and caesium clustered at the top left corner of the loadings in **Figure 1.D** and the metabolites: *scyllo*-inositol, 3-hydroxybutyrate/3-aminoisobutyrate, carnitine and formate also at the top right corner in **Figure 1.B** ($Q^2=0.06-0.12$, $p<0.05$) indicating positive co-variation. In the third trimester, arsenic and TMAO had again the strongest correlation with homarine clustered with TMAO and copper, lead and selenium in **Figure 1.D** were clustered to the top of the exposure loading plot. Opposite to this cluster at the bottom of the metabolome plot were *N*-acetylserotonin, *N*-acetyl neuraminic acid, dimethylamine and oestrogens, indicating a negative correlation with the metals. Generally, the third trimester O2PLS model presented weaker associations based on individual variable Q^2 than the first trimester.

Persistent pollutants. Overall, 4.0% of the variation in the third trimester metabolome modelled was correlated to the persistent pollutants and 43.3% of the pollutant variation modelled was correlated to the metabolome. Very similar model parameters were found in Gipzukoia (see Table S3). Again model predictability was significantly better than random models through permutations although Q^2 was close to zero (permutation validation $p=0.001$ in both cohorts). In both cohorts, PCBs 180, was clustered on the left side of the loading plots (**Figure 2 A-D**) meaning that it is negatively correlated to urinary metabolites ($Q^2=0.03-0.05$), 3-hydroxyisovalerate is the metabolite with the highest weight on PC1, i.e. on the right side of the loading plot in both cohort. This result is consistent with the univariate analysis indicating a significant negative association between PCBs and 3-hydroxyisovalerate. In Sabadell only, cotinine, on the top of the loading plot displayed the strongest correlation ($Q^2=0.14$, $p=0.005$) in **Figure 2.A** and *N*-methylnicotinic acid, *N*-methylpyridinium and furoylglycine clustered on the left side of the metabolite loading plot in both cohorts (**Figure 2.C**, left quadrant, $Q^2=0.06-0.08$). These results also confirm the univariate analysis although none of the variables had a significant predictive ability (Q^2) compared to “null” models. Other associations observed in the univariate analysis such as mercury with taurine were not observed in the O2PLS analysis.

Urine metabolites and pollutant associations in relation to main lifestyle, demographic and clinical parameters

We evaluated the change in the magnitude of the exposure-metabolite association due to the inclusion of clinical and lifestyle factors in the model, by comparing the estimated effects (p) of the basic adjusted correlations (adjusted for maternal BMI, age, gestational week and time of the day at sampling) with those of the partial correlations including different clinical and lifestyle factors (**Table S8** has the estimated effects (p) and **Table S9** the attenuation percentage). This analysis was conducted on selected associations replicated across the two trimesters in Sabadell or replicated across the cohorts in the third trimester (summarised in table 3). Addition in the models of most of the clinical and lifestyle factors did not result in a significant reduction (based on the 95% confidence intervals for the attenuation percentage) in the effect size for the metabolite-exposure associations. Attenuation for the arsenic-TMAO association correlation was 10% (95% CI: 2, 19) in the first trimester, suggesting that part of this association could be explained by reported seafood intake. We also observed a small attenuation in the third trimester (6% CI: -4, 17). The largest reduction in the effect size for a metabolite-exposure association was observed when including coffee/tea intake in the partial correlations for caesium with dimethylamine in both trimesters (1st trim.: 13% [0.4; 26] and 3rd trim.: 16% [2, 30]), and formate (1st: 10% [-0.1; 20] and 3rd: 18% [1, 34]). Effect size reductions due to the inclusion of coffee/tea intake for cotinine association with *N*¹-methylnicotinic acid (31%, 95% CI: 9; 53) and *N*-methylpyridinium (22%, 95% CI: 8; 37) were even larger. Finally, vitamin B12 showed a 20% effect reduction in the 3-hydroxyisovalerate/ PCB153 association in both cohorts.

Discussion

Using an approach based targeted pollutant measurements and untargeted NMR-based metabolic profiling, we have for the first time demonstrated reproducible associations between measurements of multiple pollutant exposures and endogenous metabolic processes in a highly relevant population. Our results have implications for general epidemiological and clinical practice by providing improved understanding of the metabolically mediated health effects of environmental exposures. Furthermore, the observed associations may reflect differences in the *in-utero* environment that could potentially impact on normal fetal development and subsequent child health². Exposure-metabolite associations were validated through two time points during pregnancy for non-persistent pollutants and in two cohorts for the persistent pollutants. In addition, two statistical approaches were applied, first univariate, aiming to characterise and quantify each metabolite-exposure association and second, O2PLS, to explore common exposure-metabolite patterns in a multivariate way. We found metabolic associations with arsenic exposure, potentially related to both seafood consumption and gut microbial methylamine metabolism, of tobacco smoke exposure related to coffee metabolism, of mercury with taurine and oestrogens indicating anti-oxidative stress and endocrine perturbation and of PCBs with 3-hydroxyisovalerate, an organic acid ketogenesis³⁹. While most of these associations are cross-sectional, and therefore difficult to confirm the direction of causal pathways, they demonstrate reproducible interaction between exogenous chemical exposures and endogenous metabolites.

367 *Urinary metabolic footprint of metal exposure during pregnancy*

368 Non-persistent pollutants, including metals and phthalates, showed distinctive patterns of association
369 with urinary metabolites in pregnancy and bisphenol A showed no associations. One group of metals:
370 thallium, caesium, copper and lead, showed similar patterns. These metals were mostly associated
371 with increased scyllo-inositol, acetate, formate, carnitine and decreased dimethylamine and *N*-
372 acetylated metabolites. They were detected in more than 90% of the samples except thallium which
373 was detected only in 15-20% of the women, and had similar levels to previous studies³⁸. Some of
374 these metals possess essential metabolic functionalities, such as copper (Cu) and others are toxic,
375 depending on their form (organic or inorganic) and dose, such as mercury (Hg), lead (Pb), arsenic
376 (As), thallium (Tl) or cadmium (Cd). They all have the capacity to bind to vital cellular components,
377 such as structural proteins, enzymes, and nucleic acids, and interfere with their function, as well as
378 provoke oxidative stress⁴⁰. The common metabolic signatures of these metals observed in this study
379 (but apparently different metabolic pathways) may arise from a common source of exposure since
380 they are correlated in our sample²⁵ or reflect common renal excretion pathways (glomerular filtration
381 and tubular reabsorption)^{41–43}. Urinary lead, cadmium, thallium and copper showed trimester specific
382 associations with steroid hormone by-products including positive associations with pregnanolone-3-
383 glucuronide in the first trimester and negative associations with oestrogen metabolites. Mercury as
384 measured in cord blood was also associated with lower urinary oestrogen metabolites in the third
385 trimester. In line with our results, the following metals have been previously reported as endocrine
386 disruptors: cadmium^{44,45}, mercury⁴⁶ and lead⁴⁷ and were more specifically shown to stimulate
387 progesterone synthesis at low levels. Supporting evidence from animal and human studies were
388 published indicating that mercury can impair endocrine function through its ability to reduce hormone-
389 receptor binding or through the inhibition of one or more key enzymes or steps in hormone
390 biosynthesis⁴⁸. Both oestrogen and progesterone by-products were previously shown to be
391 associated with fetal growth outcomes in this population²⁶. Mercury was consistently associated also
392 with increased taurine in the third trimester indicating a potential adaptive response of the antioxidant
393 system through glutathione repletion or more broadly liver damage^{49,50}.

394 Total urinary arsenic displayed among the strongest association with urinary metabolites (arsenic-
395 TMAO: Spearman's rho (ρ , 1st)=0.36, p =8.60E-12 and ρ (3rd)=0.25, p =2.9E-06), i.e. the two
396 methylamines: TMAO and dimethylamine. In contrast to other metalloids, the total arsenic urine
397 concentrations in the mothers of the Sabadell cohort, showed particularly high concentration (median
398 32 and 34 μ g/g creatinine in the 1st and 3rd trimesters respectively) when compared to other non-
399 exposed populations from European countries, Australia and US (populations without arsenic-
400 contaminated waters nor occupational exposure)⁵¹. The main source of arsenic in Sabadell women
401 was shown to be seafood, particularly lean fish, based on questionnaire data⁵¹. In an arsenic
402 speciation study in a subset of pregnant women from Valencia within the INMA cohort, arsenobetaine,
403 monomethylarsonic acid (MMA), dimethylarsonic acid (DMA) and inorganic arsenic were the main
404 species identified in urine⁵². The major source of urinary arsenobetaine is from dietary seafood intake,
405 whereas MMA and DMA arise from reduction and methylation of inorganic arsenic (both Arsenite (III)

and Arsenate (IV)). In addition, the gut microbiome has been shown to be capable of metabolism of inorganic arsenic via methylation and thiolation^{53,54}. Our results are in line with these previous findings. TMAO is a small organic metabolite found in high concentrations in fish which is rapidly absorbed and excreted in the urine^{55,56}. In addition, metabolic retroversion of TMAO to TMA by the gut microflora followed by subsequent oxidation in the liver by flavin-containing monooxygenases (FMO) has been shown to also contribute to urinary TMAO excretion⁵⁷. Urinary TMAO excretion has been shown to be highly variable in subjects following fish ingestion which has been attributed to gut microfloral differences⁵⁵. Indeed we have found that urinary TMAO excretion in HELIX children across a week is highly variable both within and between individuals⁵⁸. Our study is the first to highlight a strong and significant association between urinary total arsenic and TMAO in two cohorts. Arsenic, was also positively associated to a newly identified metabolite; homarine. Homarine (*N*-methylpicolinic acid) is widely distributed in marine invertebrates, in particular in molluscs and crustaceans⁵⁹ but has never been measured to our knowledge in humans. Therefore, it likely rises from the consumption of seafood although we could not completely confirm this based on dietary frequency questionnaires. Further adjustment for fish intake (dietary recall) only reduced the partial correlation coefficient between arsenic and TMAO by ca. 10% but the fish intake influence on TMAO might be underestimated due to the exposure error common in this type of dietary recall and the rapid clearance of TMAO.

Phthalates, PCBs and other pollutants during pregnancy

Phthalates were consistently associated with decreased urinary succinate and acetate in both cohorts and both trimesters (with stronger associations in the first trimester). Previous studies of phthalate exposure and endogenous metabolism are scarce and difficult to compare. One fertility study in 364 Chinese male subjects exposed to low-level environmental phthalates demonstrated a perturbed fatty acid beta-oxidation pathway and disturbed amino acid metabolism (alanine, taurine, tryptophan, ornithine, cystine and derivatives)⁶⁰. In line with our results, an animal study observed changes in pyruvate metabolism in rats exposed to DEP during the prepubertal period⁶¹. This type of metabolic perturbations, related with defects in mitochondrial metabolism (pyruvate, Kreb's cycle, acetyl-coA metabolism), should be better addressed in the future with metabolic flux studies in the context of phthalate exposure.

Among the persistent pollutants measured in maternal serum, only PCBs were consistently associated in both cohorts with urinary metabolic phenotypes. 3-hydroxyisovalerate, a product of the L-leucine mitochondrial catabolic pathway, was found to be decreased in women more exposed to PCBs. This metabolite was previously found decreased in a population with higher levels of cadmium exposure and smokers¹⁸.

Other pollutants such as BPA and PFAS didn't show any consistent associations with the urine metabolome during pregnancy. This is in line with the inconsistent evidence reported for these chemicals and associations with child health².

443 *Pollutant associations and coffee intake*

444 Urinary metabolic signatures of cotinine appeared to be related to coffee/tea consumption. Indeed, *N*-
445 methylnicotinic acid, (trigonelline) is known as the second most abundant alkaloid in coffee after
446 caffeine. In addition, two more metabolites associated with cotinine, i.e. *N*-methylpyridinium and
447 furoylglycine, are known to be dependent on the roasting procedure of the coffee bean and therefore
448 not influenced by consumption of tea, fruit and vegetables, or caffeine-containing energy drinks, which
449 make them highly specific to coffee⁶². The association between cotinine and coffee metabolites might
450 be further explained by an acceleration of caffeine metabolism in tobacco exposed women possibly
451 due to cigarette smoking activating cytochrome P450-mediated alkaloid metabolism⁶³. Indeed, a
452 previous study on chronic obstructive pulmonary disease (COPD) showed that downstream caffeine
453 metabolites, were elevated in the serum samples of COPD smokers as compared with both healthy
454 smokers and COPD non-smokers⁶⁴. Taken together, these coffee biomarkers in combination with
455 cotinine might be used to disentangle the health effect of smoking and coffee drinking habits during
456 pregnancy. Citrate was also consistently lower in women from both cohorts at both trimesters when
457 exposed to tobacco smoke. This result is in line with a previous study in a general human population
458 which demonstrated that smoking could induce a long-term change in citrate metabolism, possibly
459 reflecting mitochondrial perturbation and oxidative stress¹⁸.

460 As exposome research gathers increasingly richer data sets in population studies, researchers find
461 themselves confronted with complex exposure profiles that render identification of exposure–outcome
462 associations using conventional statistical analyses challenging. In our study, we applied two different
463 complementary statistical approaches: one conventional method based on single-exposure linear
464 regression models followed by multiple testing correction and the other one inspired by systems
465 biology, multivariate modelling and multi-omics integration: O2PLS³⁶, recently applied to exposure-
466 mixture models⁶⁵. Both these methods have limitations, the first one giving rise to potential false
467 positive results and the second one sometimes leading to overfitting of the data, lack of interpretability
468 and absence of specific adjustment for known confounders. However, to overcome some of these
469 problems in our study we used replication across two cohorts and time points to validate our findings,
470 interpreted exposure-metabolite association in light of potential intermediary factors in an attenuation
471 analysis and used cross-validation and model permutation to ensure we had statistically robust and
472 meaningful O2PLS models.

473 Our study had some limitations: although the results were replicated in two different cohorts for the
474 persistent pollutants, for the non-persistent analysis, data were only available in Sabadell. While we
475 were able to replicate some associations within this cohort in two different sampling time-points, the
476 urinary metabolic profile and overall xenobiotic metabolism vary greatly between the first and third
477 trimester of pregnancy^{26,38}. Therefore, we did not expect to find exactly the same associations in both
478 trimesters, as illustrated for steroid hormones and endocrine disrupting metals.

479 Recent advances in high-resolution metabolic phenotyping and metabolite identification have
480 provided the means for a full systems level integration of both the exposome and human metabolism

and will likely drive the continued development of personalised health assessment⁶⁶. It will be important to evaluate the additional prognostic value of metabolomic measurements compared to established exposure assessment and clinical visits. Finally, of course, it is impossible within the context of this cross-sectional study to determine the direction of associations between pollutant and metabolite measurements. For instance, it is possible that the existing metabolic phenotype may impact pollutant levels available for measurement in biological samples. In this regard, the strength of our study lies in the rich exposure, lifestyle, diet and metabolic data available, across multiple time points in two populations. Finally, the consequences of prenatal pollutant exposures on child metabolic phenotype was not assessed. These issues are currently being examined in follow-up studies (www.projecthelix.eu)⁶⁷.

Acknowledgements

This work was supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 308333 – the HELIX project, the Medical Research Council Integrative Toxicology Training Partnership (ITTP) through a PhD studentship (recipient LM), an MRC-ITTP career development fellowship (recipient MC), an MRC early career fellowship (recipient OR), and the Medical Research Council–Public Health England (MRC-PHE) Centre for Environment and Health (MR/L01341X/1). The subcohort studies were funded by grants from Instituto de Salud Carlos III (Red INMA G03/176, FIS-PI06/0867, FIS-PS09/00090 and FIS-PI13/02187), Generalitat de Catalunya-CIRIT 1999SGR 00241, Department of Health of the Basque Government (2005111093, 2009111069 and 2013111089), and the Provincial Government of Gipuzkoa (DFG06/004 and DFG08/001). Convenios anuales con los ayuntamientos de la zona del estudio (Zumarraga, Urretxu, Legazpi, Azkoitia y Azpeitia y Beasain).

We would like to thank Olivier Cloarec and Ekaterina (Katya) Nevedomskaya for developing the in-house script to perform O2PLS models. The authors have no competing financial interests to declare.

Supporting Information.

- List of metabolite integrals and assignments obtained from NMR urine spectra and a detailed metabonomic protocol
- Tables of the characteristics of the study population and persistent and non-persistent chemical exposure/metal levels
- Cross-sectional significant associations between non-persistent pollutants and urine metabolome different in the first and third trimesters in the Sabadell cohort
- Tables and figures with further information on the O2PLS models
- Tables with attenuation by clinical and lifestyle factors of pollutants and metabolome associations and mutual adjustment for metals.

References

1. Robinson, O. & Vrijheid, M. The Pregnancy Exposome. *Curr. Environ. Heal. Reports* **2**, 204–213 (2015).
2. Vrijheid, M., Casas, M., Gascon, M., Valvi, D. & Nieuwenhuijsen, M. Environmental pollutants and child health-A review of recent concerns. *Int. J. Hyg. Environ. Health* **219**, 331–42 (2016).
3. Govarts, E. *et al.* Birth Weight and Prenatal Exposure to Polychlorinated Biphenyls (PCBs) and Dichlorodiphenyldichloroethylene (DDE): A Meta-analysis within 12 European Birth Cohorts. *Environ. Health Perspect.* **120**, 162–170 (2011).
4. Lanphear, B. P. *et al.* Low-level environmental lead exposure and children's intellectual function: an international pooled analysis. *Environ. Health Perspect.* **113**, 894–9 (2005).
5. Yoshimasu, K., Kiyohara, C., Takemura, S. & Nakai, K. A meta-analysis of the evidence on the impact of prenatal and early infancy exposures to mercury on autism and attention deficit/hyperactivity disorder in the childhood. *Neurotoxicology* **44**, 121–131 (2014).
6. Berghuis, S. A., Bos, A. F., Sauer, P. J. J. & Roze, E. Developmental neurotoxicity of persistent organic pollutants: an update on childhood outcome. *Arch. Toxicol.* **89**, 687–709 (2015).
7. González-Alzaga, B. *et al.* A systematic review of neurodevelopmental effects of prenatal and postnatal organophosphate pesticide exposure. *Toxicol. Lett.* **230**, 104–121 (2014).
8. Wild, C. P. Complementing the Genome with an "Exposome": The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiol. Biomarkers Prev.* **14**, 1847–1850 (2005).
9. Patel, C. J. & Ioannidis, J. P. A. Placing epidemiological results in the context of multiplicity and typical correlations of exposures. *J. Epidemiol. Community Health* **68**, 1096–1100 (2014).
10. Peters, A., Hoek, G. & Katsouyanni, K. Understanding the link between environmental exposures and health: does the exposome promise too much?: Figure 1. *J. Epidemiol. Community Health* **66**, 103–105 (2012).
11. Chadeau-Hyam, M. *et al.* Meeting-in-the-middle using metabolic profiling - a strategy for the identification of intermediate biomarkers in cohort studies. *Biomarkers* **16**, 83–88 (2011).
12. Heindel, J. J. *et al.* Metabolism disrupting chemicals and metabolic disorders. *Reprod. Toxicol.* **68**, 3–33 (2017).
13. Casals-Casas, C. & Desvergne, B. Endocrine Disruptors: From Endocrine to Metabolic Disruption. *Annu. Rev. Physiol.* **73**, 135–162 (2011).
14. Sumner, S. *et al.* Metabolomics in the assessment of chemical-induced reproductive and developmental outcomes using non-invasive biological fluids: application to the study of butylbenzyl phthalate. *J. Appl. Toxicol.* **29**, 703–714 (2009).
15. Cabaton, N. J. *et al.* Effects of Low Doses of Bisphenol A on the Metabolome of Perinatally Exposed CD-1 Mice. *Environ. Health Perspect.* (2013). doi:10.1289/ehp.1205588
16. Zeng, J. *et al.* Effect of Bisphenol A on Rat Metabolic Profiling Studied by Using Capillary Electrophoresis Time-of-Flight Mass Spectrometry. *Environ. Sci. Technol.* **47**, 7457–7465 (2013).
17. Chen, M. *et al.* Metabolomic Analysis Reveals Metabolic Changes Caused by Bisphenol A in Rats. *Toxicol. Sci.* **138**, 256–267 (2014).
18. Ellis, J. K. *et al.* Metabolic profiling detects early effects of environmental and lifestyle exposure to cadmium in a human population. *BMC Med.* **10**, 61 (2012).
19. Shen, H. *et al.* Urinary Metabolic Biomarkers Link Oxidative Stress Indicators Associated with General Arsenic Exposure to Male Infertility In a Han Chinese Population. *Environ. Sci. Technol.* 130722083038001 (2013). doi:10.1021/es402025n
20. Zhang, J. *et al.* Urinary metabolomics revealed arsenic internal dose-related metabolic alterations: A proof-of-concept study in a chinese male cohort. *Environ. Sci. Technol.* **48**, 12265–12274 (2014).
21. Laine, J. E. *et al.* Neonatal Metabolomic Profiles Related to Prenatal Arsenic Exposure. *Environ. Sci. Technol.* **51**, 625–633 (2017).
22. Walker, D. I. *et al.* High-resolution metabolomics of occupational exposure to trichloroethylene. *Int. J. Epidemiol.* **45**, 1517–1527 (2016).
23. van Veldhoven, K. *et al.* Effects of exposure to water disinfection by-products in a swimming pool: A metabolome-wide association study. *Environ. Int.* **111**, 60–70 (2018).
24. Guxens, M. *et al.* Cohort Profile: The INMA-INfancia y Medio Ambiente--(Environment and Childhood) Project. *Int. J. Epidemiol.* **054**, (2011).
25. Robinson, O. *et al.* The Pregnancy Exposome: Multiple Environmental Exposures in the INMA-Sabadell Birth Cohort. *Environ. Sci. Technol.* **49**, 10632–10641 (2015).
26. Maitre, L. *et al.* Maternal urinary metabolic signatures of fetal growth and associated clinical

- and environmental factors in the INMA study. *BMC Med.* **14**, 177 (2016).
27. Vioque, J. *et al.* Reproducibility and validity of a food frequency questionnaire among pregnant women in a Mediterranean area. *Nutr. J.* **12**, 26 (2013).
 28. Beckonert, O. *et al.* Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2**, 2692–2703 (2007).
 29. Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal. Chem.* **78**, 4281–90 (2006).
 30. White, I. R., Royston, P. & Wood, A. M. Multiple imputation using chained equations: Issues and guidance for practice. *Stat. Med.* **30**, 377–399 (2011).
 31. Liao, J. G., Lin, Y., Selvanayagam, Z. E. & Shih, W. J. A mixture model for estimating the local false discovery rate in DNA microarray analysis. *Bioinformatics* **20**, 2694–2701 (2004).
 32. Strimmer, K. fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* **24**, 1461–1462 (2008).
 33. Trygg, J. & Wold, S. O2-pls, a two-block (x–y) latent variable regression (lvr) method with an integral osc filter. *J Chemom.* **17**, (2003).
 34. Trygg, J. & Wold, S. Orthogonal projections to latent structures (o-pls). *J Chemom.* **16**, (2002).
 35. Koufaris, C. *et al.* Systematic integration of molecular profiles identifies miR-22 as a regulator of lipid and folate metabolism in breast cancer cells. *Oncogene* **35**, 2766–2776 (2016).
 36. Bouhaddani, S. el *et al.* Evaluation of O2PLS in Omics data integration. *BMC Bioinformatics* **17**, S11 (2016).
 37. Biswal, B. B., Taylor, P. A. & Ulmer, J. L. Use of jackknife resampling techniques to estimate the confidence intervals of fMRI parameters. *J. Comput. Assist. Tomogr.* **25**, 113–20 (2001).
 38. Fort, M. *et al.* Assessment of exposure to trace metals in a cohort of pregnant women from an urban center by urine analysis in the first and third trimesters of pregnancy. *Environ. Sci. Pollut. Res.* **21**, 9234–9241 (2014).
 39. Yu, W. M. *et al.* Increased urinary excretion of beta-hydroxyisovaleric acid in ketotic and non-ketotic type II diabetes mellitus. *Clin. Chim. Acta.* **188**, 161–8 (1990).
 40. Jaishankar, M., Tseten, T., Anbalagan, N., Mathew, B. B. & Beeregowda, K. N. Toxicity, mechanism and health effects of some heavy metals. *Interdiscip. Toxicol.* **7**, 60–72 (2014).
 41. Engelke, U. F. H. *et al.* N-acetylated metabolites in urine: proton nuclear magnetic resonance spectroscopic study on patients with inborn errors of metabolism. *Clin. Chem.* **50**, 58–66 (2004).
 42. Bell, J. D. D., Brown, J. C. C., Nicholson, J. K. & Sadler, P. J. Assignment of resonances for 'acute-phase' glycoproteins in high resolution proton NMR spectra of human blood plasma. *FEBS Lett.* **215**, 311 (1987).
 43. Wang, H., Dumont, X., Haufroid, V. & Bernard, A. The physiological determinants of low-level urine cadmium: an assessment in a cross-sectional study among schoolchildren. *Environ. Heal.* **16**, 99 (2017).
 44. Powlin, S. S., Keng, P. C. & Miller, R. K. Toxicity of Cadmium in Human Trophoblast Cells (JAR Choriocarcinoma): Role of Calmodulin and the Calmodulin Inhibitor, Zaldaride Maleate. *Toxicol. Appl. Pharmacol.* **144**, 225–234 (1997).
 45. Massányi, P., Uhrin, V., Sirotkin, A., ... K. P.-A. V. & 2000, undefined. Effects of cadmium on ultrastructure and steroidogenesis in cultured porcine ovarian granulosa cells. *actavet.vfu.cz*
 46. Mondal, S., Mukhopadhyay, B. & Bhattacharya, S. Inorganic mercury binding to fish oocyte plasma membrane induces steroidogenesis and translatable messenger RNA synthesis. *Biometals* **10**, 285–290 (1997).
 47. Srivastava, V. *et al.* The Effects of Low-level Pb on Steroidogenic Acute Regulatory Protein (StAR) in the Prepubertal Rat Ovary. *Toxicol. Sci.* **77**, 35–40 (2003).
 48. Iavicoli, I., Fontana, L. & Bergamaschi, A. The Effects of Metals as Endocrine Disruptors. *J. Toxicol. Environ. Heal. Part B* **12**, 206–223 (2009).
 49. Penttilä, K. E. Role of cysteine and taurine in regulating glutathione synthesis by periportal and perivenous hepatocytes. *Biochem. J.* **269**, 659–664 (1990).
 50. Waterfield, C. J., Turton, J. A., Scales, M. D. & Timbrell, J. A. Investigations into the effects of various hepatotoxic compounds on urinary and liver taurine levels in rats. *Arch. Toxicol.* **67**, 244–54 (1993).
 51. Fort, M., Grimalt, J. O., Casas, M. & Sunyer, J. Food sources of arsenic in pregnant Mediterranean women with high urine concentrations of this metalloid. (2013). doi:10.1007/s11356-014-2614-4
 52. Signes-Pastor, A. J. *et al.* Urinary Arsenic Speciation in Children and Pregnant Women from

- Spain. *Expo. Heal.* **9**, 105–111 (2017).
53. Van de Wiele, T. *et al.* Arsenic Metabolism by Human Gut Microbiota upon in Vitro Digestion of Contaminated Soils. *Environ. Health Perspect.* **118**, 1004–1009 (2010).
 54. DC.Rubin, S. S., Alava, P., Zekker, I., Du Laing, G. & Van de Wiele, T. Arsenic Thiolation and the Role of Sulfate-Reducing Bacteria from the Human Intestinal Tract. *Environ. Health Perspect.* (2014). doi:10.1289/ehp.1307759
 55. Zhang, A. Q., Mitchell, S. C. & Smith, R. L. Dietary precursors of trimethylamine in man: a pilot study. *Food Chem. Toxicol.* **37**, 515–20 (1999).
 56. Cho, C. E. *et al.* Trimethylamine- N -oxide (TMAO) response to animal source foods varies among healthy young men and is influenced by their gut microbiota composition: A randomized controlled trial. *Mol. Nutr. Food Res.* **61**, 1600324 (2017).
 57. Al-Waiz, M., Ayesh, R., Mitchell, S. C., Idle, J. R. & Smith, R. L. Disclosure of the metabolic retroversion of trimethylamine N-oxide in humans: a pharmacogenetic approach. *Clin. Pharmacol. Ther.* **42**, 608–12 (1987).
 58. Maitre, L. *et al.* Assessment of metabolic phenotypic variability in children's urine using 1H NMR spectroscopy. *Sci. Rep.* **7**, 46082 (2017).
 59. Carr, W., Netherton JC, I., Gleeson, R. A. & Derby, C. D. Stimulants of Feeding Behavior in Fish: Analyses of Tissues of Diverse Marine Organisms. *Biol. Bull.* **190**, 149–160 (1996).
 60. Zhang, J. *et al.* Low-Level Environmental Phthalate Exposure Associates with Urine Metabolome Alteration in a Chinese Male Cohort. *Environ. Sci. Technol.* **50**, 5953–5960 (2016).
 61. Houten, S. M. *et al.* Changes in the Metabolome in Response to Low-Dose Exposure to Environmental Chemicals Used in Personal Care Products during Different Windows of Susceptibility. *PLoS One* **11**, e0159919 (2016).
 62. Heinzmann, S. S., Holmes, E., Kochhar, S., Nicholson, J. K. & Schmitt-Kopplin, P. 2-Furoylglycine as a Candidate Biomarker of Coffee Consumption. *J. Agric. Food Chem.* **63**, 8615–8621 (2015).
 63. Grygiel, J. J. & Birkett, D. J. Cigarette smoking and theophylline clearance and metabolism. *Clin. Pharmacol. Ther.* **30**, 491–6 (1981).
 64. Chen, Q. *et al.* Serum Metabolite Biomarkers Discriminate Healthy Smokers from COPD Smokers. *PLoS One* **10**, e0143937 (2015).
 65. Jain, P. *et al.* A multivariate approach to investigate the combined biological effects of multiple exposures. *J. Epidemiol. Community Health* **72**, 564–571 (2018).
 66. Walker, D. I., Go, Y.-M., Liu, K., Pennell, K. D. & Jones, D. P. in *Metabolic Phenotyping in Personalized and Public Healthcare* 167–211 (Elsevier, 2016). doi:10.1016/B978-0-12-800344-2.00007-0
 67. Vrijheid M, Robinson O, Chatzi L, Coen M, van den Hazel P, Thomsen C, Wright J, Athersuch TJ, Avellana N, Basagaña X, Brochot C, Bucchini L, Bustamante M, Carracedo A, Casas M, Estivill X, Fairley L, van Gent D, Gonzalez JR, Granum B, Gražulevičienė R, Gu, S. R. The Human Early Life Exposome (HELIX): Project Rationale and Design. *Environ. Health Perspect.* **122**, 535–544 (2014).

Tables

Table 1. Cross-sectional and longitudinal associations between non-persistent pollutants/metals and the urine metabolome reproduced in the first and third trimesters of pregnancy . List of pairwise correlations with Spearman correlation coefficient, ρ and p-values that pass 0.05 threshold (in bold when they pass the FDR correction). Only the associations with at least one statistically significant association across all analyses after FDR correction are displayed.

		1 st trimester (cross-sectional associations)		3 rd trimester (cross-sectional associations)		1 st -3 rd trimester (longitudinal associations) ^a		
Exposures		Metabolites	Spearman correlation coefficient, ρ	p -value	Spearman correlation coefficient, ρ	p-value	Spearman correlation coefficient, ρ	p-value
Metals (toxic)	Arsenic	Dimethylamine	0.23	2.2E-05	0.19	4.7E-04	-0.11	1.1E-01
		Trimethylamine oxide	0.36	1.0E-11	0.25	2.9E-06	0.049	1.7E-01
		Homarine	0.25	3.6E-06	0.29	4.1E-08	-0.003	8.3E-01
	Cadmium	<i>N</i> -acetylserotonin	-0.17	1.3E-03	-0.16	3.2E-03	-0.06	6.7E-01
	Lead	Acetate	0.2	1.9E-03	0.13	1.4E-02	0.066	8.5E-01
		Carnitine	0.13	7.5E-04	0.25	2.8E-06	0.045	4.9E-01
		Lysine	0.16	1.7E-05	0.11	4.1E-02	0.095	1.4E-01
		<i>N</i> -acetylserotonin	-0.23	2.4E-04	-0.27	7.6E-07	-0.12	4.9E-01
		UA 1.15 (d)	0.17	1.9E-02	0.16	3.6E-03	0.088	1.7E-01
		<i>trans</i> -Aconitate	-0.22	3.0E-03	-0.16	3.8E-03	-0.048	9.8E-01
		Dimethylamine	-0.17	1.9E-05	-0.19	4.5E-04	-0.12	2.2E-01
	Nickel	<i>N</i> -acetylserotonin	-0.18	1.5E-03	-0.18	7.0E-04	-0.11	4.1E-01
		<i>N</i> ¹ -methyl-nicotinamide	-0.23	4.2E-05	-0.16	3.2E-03	0.074	3.2E-01
		Acetate	0.21	1.4E-04	0.18	7.2E-04	0.00094	7.9E-01
	Thallium	Carnitine	0.11	4.1E-02	0.4	2.8E-14	0.044	9.7E-01
		Dimethylamine	-0.16	3.8E-03	-0.15	5.5E-03	-0.077	7.7E-02
		Formate	0.27	4.3E-07	0.2	2.2E-04	0.036	5.4E-01
		Isoleucine	-0.12	2.5E-02	-0.3	2.4E-08	-0.066	5.7E-02
		<i>N</i> -acetyl neuraminic acid	-0.22	5.3E-05	-0.21	1.2E-04	-0.0021	1.6E-02

19

Metals (other, essentials and non-essentials)		<i>N</i> -acetylserotonin	-0.19	4.4E-04	-0.37	1.4E-12	-0.083	4.9E-02
		Oestrogen metabolites	-0.13	1.5E-02	-0.29	3.6E-08	-0.15	6.3E-03
		Scyllo-inositol	0.36	1.6E-11	0.21	8.4E-05	-0.0039	4.4E-01
		UA 3.8(s)	-0.14	1.1E-02	-0.16	2.6E-03	0.015	3.3E-01
		<i>trans</i> -Aconitate	-0.26	8.6E-07	-0.27	3.4E-07	-0.04	6.4E-01
	Antimony	<i>N</i> -acetylserotonin	-0.15	6.8E-04	-0.18	9.5E-04	-0.06	4.7E-01
	Caesium	3-hydroxyisovalerate	-0.18	1.4E-05	-0.15	4.4E-03	-0.26	2.8E-06
		Dimethylamine	-0.23	2.1E-06	-0.21	7.5E-05	-0.087	1.1E-01
		Formate	0.26	2.8E-04	0.16	2.8E-03	0.12	3.5E-02
		<i>N</i> -acetyl neuraminic acid	-0.2	5.0E-03	-0.2	2.7E-04	-0.016	5.9E-01
		Tyrosine	0.15	8.6E-03	0.19	5.3E-04	-0.0026	7.8E-01
		UA 3.93(s)	0.14	5.7E-03	0.19	6.5E-04	0.2	1.0E-03
		UA 6.48(d)	0.15	3.9E-03	0.21	1.5E-04	0.13	7.3E-03
		UA 7.68(s)	-0.16	7.3E-03	-0.14	9.1E-03	-0.0014	9.9E-01
	Copper	Carnitine	0.18	1.2E-03	0.19	4.5E-04	0.033	4.8E-01
		Formate	0.26	9.1E-07	0.12	3.2E-02	0.074	3.1E-02
		<i>N</i> -acetylserotonin	-0.19	5.2E-04	-0.14	9.9E-03	-0.098	2.3E-01
		Oestrogen metabolite	-0.12	2.8E-02	-0.17	1.8E-03	-0.15	7.7E-03
		Scyllo-inositol	0.37	1.5E-12	0.17	1.4E-03	-0.079	9.4E-01
		<i>trans</i> -Aconitate	-0.2	3.1E-04	-0.12	2.3E-02	-0.077	2.4E-01
	Molybdenum	<i>N</i> -acetyl neuraminic acid	-0.12	3.0E-02	-0.17	1.5E-03	-0.04	4.2E-01
	Zinc	3-hydroxyisobutyrate	0.2	2.6E-04	0.16	3.7E-03	0.098	1.5E-01
	Selenium	<i>N</i> -acetyl neuraminic acid	-0.14	9.8E-03	-0.17	1.5E-03	-0.033	9.1E-01
Phthalates	MiBP	Succinate	-0.11	1.1E-03	-0.19	4.8E-04	-0.055	2.7E-01
	MBzP	Acetate	-0.2	4.4E-02	-0.19	3.9E-04	-0.13	9.1E-02
		<i>N</i> -acetyl neuraminic acid	0.12	2.0E-04	0.14	8.3E-03	0.18	6.0E-05
		Succinate	-0.18	1.1E-03	-0.15	5.6E-03	-0.09	1.6E-01
	MCMHP	Citrate	-0.16	2.6E-03	-0.11	4.9E-02	-0.098	2.2E-01

19

20

MEHHP	Succinate	-0.19	5.7E-04	-0.13	2.0E-02	-0.018	1.5E-01
	Acetate	-0.17	2.0E-03	-0.18	1.0E-03	-0.072	9.0E-02
	Succinate	-0.24	8.4E-06	-0.14	1.0E-02	-0.041	2.0E-01
MEOHP	Acetate	-0.18	1.0E-03	-0.15	5.0E-03	-0.043	1.3E-01
	Succinate	-0.25	5.2E-06	-0.13	1.7E-02	-0.044	2.0E-01
	Valine	-0.12	2.3E-02	-0.16	2.7E-03	-0.15	8.1E-03

^a Association of exposure measured in first trimester urine and metabolite measured in third trimester urine

Abbreviations: 2Py, N-methyl-2-pyridone-5-carboxamide; 3HB/3AB, 3-hydroxybutyrate/3-aminoisobutyrate; TMAO, Trimethylamine oxide; UA, unassigned

Table 2. Associations between persistent pollutants and urine metabolome in first and third trimesters. List of pairwise correlations with Spearman correlation coefficient, ρ and p-values. Those that pass FDR correction are in bold. Only the associations with at least one statistical significant association across all analyses after FDR correction are displayed.

		1st trimester metabolome				3rd trimester metabolome				
		Sabadell, N=340		Gipuzkoa, N=410		Sabadell, N=340		Gipuzkoa, N=410		
Exposures	Metabolites	ρ	p-values	ρ	p-values	ρ	p-values	ρ	p-values	
Organochlorines ^a	PCB138	3-hydroxyisovalerate	-0.071	1.9E-01	-0.13	1.0E-02	-0.18	7.1E-04	-0.11	2.2E-02
		4-deoxyerythronic acid	-0.2	1.6E-04	-0.12	1.6E-02	-0.08	1.4E-01	-0.042	4.0E-01
		UA 3.95(s)	-0.085	1.2E-01	-0.033	5.1E-01	-0.18	1.2E-03	0.025	6.2E-01
		Succinate	0.067	2.2E-01	-0.0017	9.7E-01	0.18	8.0E-04	-0.022	6.5E-01
	PCB153	4-deoxyerythronic acid	-0.22	5.4E-05	-0.16	1.7E-03	-0.064	2.4E-01	-0.064	2.0E-01
		3-hydroxyisovalerate	-0.087	1.1E-01	-0.15	2.7E-03	-0.22	5.3E-05	-0.18	2.3E-04
		Succinate	0.033	5.5E-01	0.046	3.6E-01	0.22	3.5E-05	-0.008	8.7E-01
		UA 3.95(s)	-0.09	9.9E-02	-0.023	6.4E-01	-0.21	1.4E-04	-0.012	8.0E-01
	PCB180	3-hydroxyisovalerate	-0.076	1.6E-01	-0.18	2.4E-04	-0.18	1.0E-03	-0.19	1.2E-04
		4-deoxyerythronic acid	-0.22	4.9E-05	-0.11	3.0E-02	-0.05	3.6E-01	-0.062	2.1E-01

20

21

		Succinate	0.028	6.0E-01	0.013	7.9E-01	0.24	6.2E-06	-0.022	6.6E-01
PFAS ^a	PFHxS	UA 2.78(s)	-0.054	3.3E-01	0.19	1.7E-04	-0.00066	9.9E-01	0.14	4.1E-03
		UA 3.93(s)	0.021	7.0E-01	-0.08	1.1E-01	-0.019	7.3E-01	0.18	3.1E-04
		Pregnanolone-3G	0.087	1.1E-01	-0.027	5.9E-01	-0.0058	9.1E-01	-0.17	5.2E-04
		Acetone	-0.028	6.1E-01	0.0052	9.2E-01	0.039	4.7E-01	-0.21	1.9E-05
	PFOA	Succinate	-0.025	6.5E-01	-0.0039	9.4E-01	0.04	4.6E-01	0.17	7.3E-04
		Alanine	-0.0011	9.8E-01	-0.1	4.2E-02	0.015	7.8E-01	-0.2	3.7E-05
		Glycine	0.016	7.6E-01	-0.035	4.8E-01	0.004	9.4E-01	-0.16	8.4E-04
		3-hydroxybutyrate/3-aminoisobutyrate	0.02	7.2E-01	0.024	6.2E-01	-0.032	5.6E-01	0.18	3.7E-04
Tobacco ^b	Cotinine	Citrate	-0.22	5.9E-05	-0.16	9.2E-04	-0.15	5.9E-03	-0.12	1.6E-02
		N ¹ -methylnicotinic acid	0.2	1.8E-04	0.14	4.2E-03	0.24	1.1E-05	0.2	4.4E-05
		N-methylpyridinium	0.16	3.1E-03	0.11	2.7E-02	0.24	7.0E-06	0.25	5.3E-07
		Pantothenic acid	-0.028	6.1E-01	-0.11	2.8E-02	-0.19	5.0E-04	-0.026	6.1E-01
		UA 3.8(s)	-0.029	6.0E-01	-0.059	2.4E-01	-0.1	6.7E-02	-0.18	2.8E-04
		UA 8.72(d)	0.099	6.9E-02	-0.057	2.5E-01	0.22	3.5E-05	0.097	5.1E-02
		Furoylglycine	0.037	5.0E-01	0.034	4.9E-01	0.19	5.8E-04	0.15	2.4E-03
Heavy metal (cord blood) ^c	Hg	Trimethylamine oxide	0.11	3.8E-02	0.2	6.5E-05	0.11	4.6E-02	0.11	2.6E-02
		Oestrogen metabolites	-0.015	7.8E-01	-0.15	1.8E-03	-0.16	4.0E-03	-0.17	8.2E-04
		Pantothenic acid	0.076	1.6E-01	0.051	3.1E-01	0.03	5.8E-01	0.17	4.2E-04
		Taurine	0.084	1.2E-01	0.0088	8.6E-01	0.21	9.4E-05	0.17	4.8E-04

^a measured in serum during the first trimester^b Cotinine measured in urine during the third trimester^c measured in cord blood at birth

Abbreviations: Hg, mercury; PFHxS, perfluorohexane.sulfonate; PFOA, perfluorooctanoic.acid; PCB, polychlorinated biphenyl; pregnanolone-3G, pregnanolone-3-glucuronide; UA, unknown with chemical shift in ppm and multiplicity, i.e. singlet (s) or doublet (d) by ¹H NMR spectroscopy.

21

22

Table 3: Associations that pass FDR correction for statistical significance in both trimesters (non-persistent/metal analysis, Sabadell only) or in both Gipuzkoa and Sabadell (persistent pollutant analysis)

Arsenic	Dimethylamine
	Trimethylamine oxide
	Homarine
Lead	N-acetylserotonin
Nickel	Dimethylamine
	N-acetylserotonin
Thallium	Acetate
	Formate
	N-acetyl neuraminic acid
	N-acetylserotonin
	Scyllo-inositol
	trans-Aconitate
Caesium	Dimethylamine
	Formate
	N-acetyl neuraminic acid
Copper	Carnitine
	Scyllo-inositol
MBzP	Acetate
MEHHP	
PCB153	3-hydroxyisovalerate
Cotinine	N ¹ -methylnicotinic acid
	N-methylpyridinium
Hg	Taurine

Figure legends

Figure 1. O2PLS modelling of metabolite–non-persistent pollutant/metals associations in the first and third trimesters of pregnancy in the Sabadell subcohort. Models contain four predictive, one X-orthogonal and one Y-orthogonal components; $R^2X=0.24$ (1st trim.), 0.22 (3rd trim.) and $R^2Y=0.57$ (1st trim.), 0.58 (3rd trim.). (A) O2PLS model correlated component ‘loadings’ and individual variable cross-validated predicted explained variance (Q^2) for metabolites in the first trimester. (B) same for **non-persistent pollutant** exposures. (C) and (D) same for the third trimester. In this display the relative position of each point in the two loadings plots indicates if regression coefficients for a given pair of pollutant and metabolite correlated or anti-correlated to each other, while distance from the origin and colour indicate the magnitude of the association and magnitude of contribution to Q^2 respectively. Variables with Q^2 values above 0.02 are labelled.

Figure 2. O2PLS modelling of metabolite–persistent pollutant associations in the Sabadell and Gipuzkoa subcohorts in the third trimester of pregnancy. Models contain four predictive, one X-orthogonal and one Y-orthogonal components; with $R^2X=0.25$ (Gizpukoa), 0.23 (Sabadell) and $R^2Y=0.71$ (both). (A) O2PLS model correlated component ‘loadings’ and individual variable cross-validated predicted explained variance (Q^2) for metabolites in Sabadell 3rd trimester. (B) Same for non-persistent pollutant exposures, same for (C) and (D) for the in Gizpukoa 3rd trimester. In this display the relative position of each point in the two loadings plots indicates if regression coefficients for a given pair of pollutant and metabolite correlated or anti-correlated to each other, while distance from the origin and colour indicate the magnitude of the association and magnitude of contribution to Q^2 respectively. Variables with Q^2 values above 0.02 are labelled.

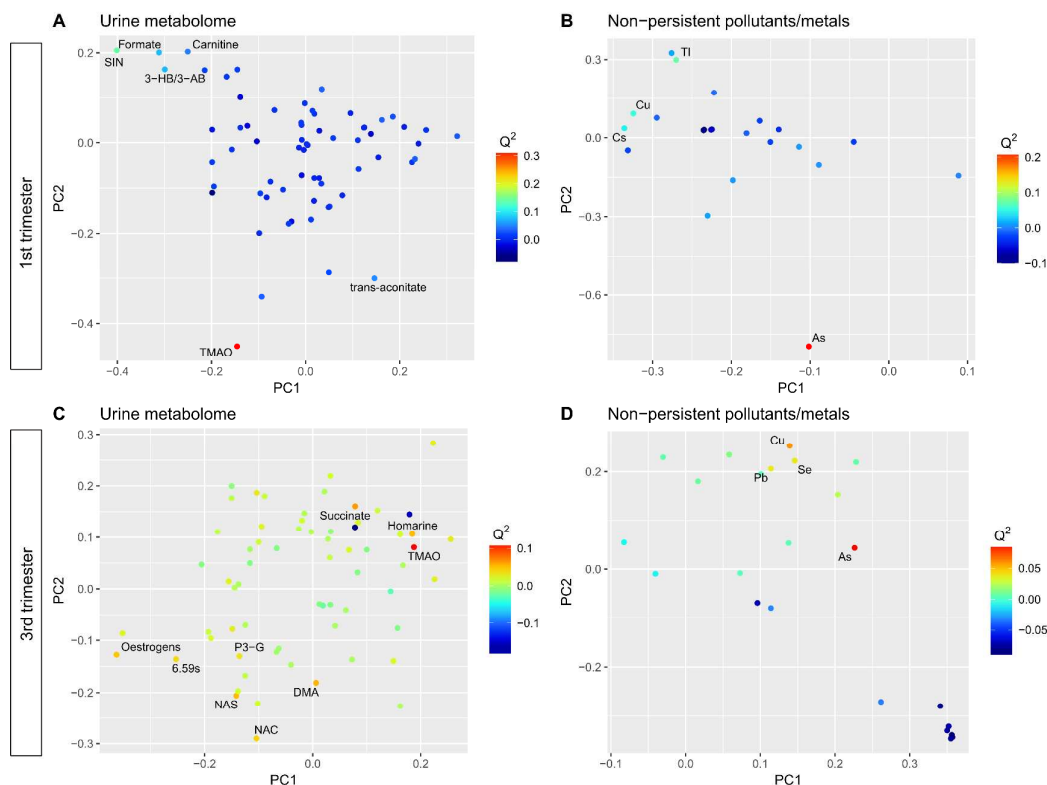


Figure 1. O2PLS modelling of metabolite–non-persistent pollutant associations in the first (A and B) and third trimesters (C and D) of pregnancy in the Sabadell subcohort.

In this display the relative position of each point in the two loadings plots indicates if regression coefficients for a given pair of pollutant and metabolite correlated or anti-correlated to each other, while distance from the origin and colour indicate the magnitude of the association and magnitude of contribution to Q^2 respectively. Variables with Q^2 values above 0.01 are labelled.

25

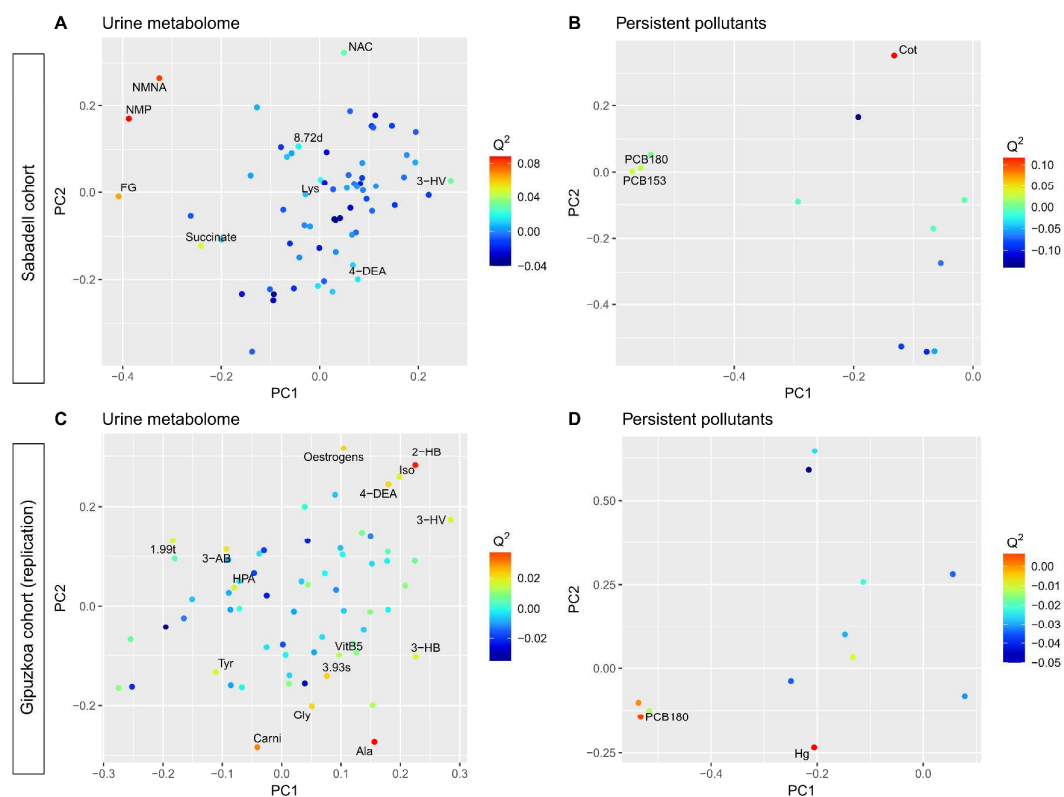


Figure 2. O2PLS modelling of metabolite–persistent pollutant associations in the Sabadell (A and B) and Gipuzkoa (C and D) subcohorts in the 3rd trimester of pregnancy.

25