A comprehensive non-targeted analysis study of the prenatal exposome

Dimitri Panagopoulos Abrahamsson[1], Aolin Wang[1], Ting Jiang[2], Miaomiao Wang[2], Adi Siddharth[1], Rachel Morello-Frosch[3], June-Soo Park[1,2], Marina Sirota[4,5+] and Tracey J. Woodruff[1]*[+]

[1] Department of Obstetrics, Gynecology and Reproductive Sciences, Program on Reproductive Health and the Environment, University of California, San Francisco, San Francisco, California, United States.

[2] California Environmental Protection Agency, Department of Toxic Substances Control, Environmental Chemistry Laboratory, 700 Heinz Ave # 200, Berkeley, CA, 94710, United States.

[3] Department of Environmental Science, Policy and Management and School of Public Health, University of California, Berkeley, Berkeley, California, United States.

[4] Bakar Computational Health Sciences Institute, University of California, San Francisco, California 94158, United States.

[5] Department of Pediatrics, University of California, San Francisco, California 94158, United States.

***Corresponding author: Tracey J. Woodruff** tracey.woodruff@ucsf.edu

[+]These two authors contributed equally to this work.

Abstract

Recent technological advances in mass spectrometry have enabled us to screen biological samples for a very broad spectrum of chemical compounds allowing us to more comprehensively characterize the human exposome in critical periods of development. The goal of this study was three-fold: 1) to analyze 590 matched maternal and cord blood samples (total 295 pairs) using non-targeted analysis (NTA); 2) examine the differences in chemical abundance between maternal and cord blood samples; and 3) examine the associations between exogenous chemicals and endogenous metabolites. We analyzed all samples with high-resolution mass spectrometry (HRMS) using liquid chromatography – quadrupole time-of-flight mass spectrometry (LC-QTOF/MS), in both positive and negative electrospray ionization modes (ESI+ and ESI-) and in soft ionization (MS) and fragmentation (MS/MS) modes for prioritized features. We confirmed 19 unique compounds with analytical standards, we tentatively identified 73 compounds with MS/MS spectra matching, and we annotated 98 compounds using an annotation algorithm. We observed 103 significant associations in maternal and 128 in cord samples between compounds annotated as endogenous and compounds annotated as exogenous. An example of these relationships was an association between 3 poly and perfluoroalkyl substances (PFAS) and endogenous fatty acids in both the maternal and cord samples indicating potential interactions between PFAS and fatty acid regulating proteins.

1. Introduction

The exposome describes the sum of all our exposures, both external and internal, throughout our lives from conception and onwards.[1,2] Humans are exposed to multiple and variable environmental contaminants in both the indoor and outdoor environments through inhalation, ingestion, and dermal absorption. Environmental exposures have been shown to play an important role in the development of human disease along with exposures to endogenous chemicals and genetic predisposition.[1,2]

Exposures to environmental contaminants during pregnancy are of critical importance due to the increased risk for adverse health outcomes that occur during periods of critical and unique susceptibility to biological perturbations, which can increase the risk of both maternal and child adverse health outcomes[3–6]. Prenatal exposures to industrial chemicals have been shown to increase the risk of complications during pregnancy, such as preterm birth, pregnancy-related hypertension, adverse birth outcomes, developmental and neurodevelopmental problems during infancy, and disease during adulthood.[3–6]

Approximately 40,000 chemicals are registered on the inventory of the Toxic Substances Control Act (TSCA) as actively used chemicals in the U.S.[7,8] This number does not include chemicals that are regulated by other U.S. statutes, such as pesticides, foods and food additives, drugs, cosmetics, tobacco and tobacco products, and nuclear materials and munitions.[7,8] The actual number of all chemicals used in the U.S. remains unclear but exceeds 40,000.

Conventional biomonitoring and human exposure research rely on targeted analytical chemistry techniques, in which one measures chemicals selected prior to the analysis. Up to now, with targeted techniques, only about 350 chemicals are biomonitored regularly via U.S. NHANES, constituting less than 1% of the chemicals used in the US. This limited number of measured targeted chemicals hinders our understanding of human exposure to chemicals and how they may impact human health. Considering the large number of chemicals that are not covered by these approaches, there is a need to develop more high-throughput approaches that cover a broader spectrum of human exposure to environmental contaminants.[9]

Recent advances in high-resolution mass spectrometry have brought non-targeted analysis (NTA) and suspect screening to the forefront of analytical chemistry. Non-targeted analysis techniques offer the possibility to screen biological and environmental samples for a very broad spectrum of chemicals that would previously remain undetected with conventional targeted analytical techniques. Such high-throughput analytical techniques enable a more holistic characterization of the exposome incorporating both internal (endogenous) and external (exogenous) exposures. Previous non-targeted and suspect screening studies[10–15] have demonstrated the value of NTA as an important screening tool for compound discovery in environmental applications. The compounds discovered through NTA can then inform more traditional targeted analytical approaches to further evaluate chemicals of interest with more stringent quality assurances that include further examination with analytical standards and quantification.

Our work builds upon previous NTA and suspect screening studies[11–13,16–18] of other scientific groups that have laid the groundwork for further analysis and have inspired further exploration. In our study, we developed an enhanced NTA workflow to screen human biological samples for a broad spectrum of chemicals that can be identified or tentatively identified, and then applied this approach to study exogenous and endogenous chemical exposures in a large racially and socioeconomically diverse population of pregnant women. The novelty of our work lies primarily in the analysis of a large cohort of maternal and cord blood samples and in the selection and combination of computational tools for the analysis and interpretation of non-targeted analysis data. Our study aims to explore the computational, analytical and environmental chemistry aspects of non-targeted analysis and explore the human exposome during pregnancy through the lens of chemistry.

The goal of this study was three-fold: 1) to analyze 590 matched maternal and cord blood samples (total 295 matched pairs) using NTA to characterize the maternal/fetal exposome; 2) examine the differences in chemical feature enrichment between maternal and cord blood samples; and 3) examine the associations between exogenous chemicals and endogenous metabolites in an attempt to understand the interplay between the exposome and the metabolome.

2. Materials and Methods

2.1 Study population

The study population consisted of 295 pregnant women recruited during the Chemicals in Our Bodies (CIOB) study (Table 1) at the University of California, San Francisco (UCSF). The CIOB study consists of about 700 (as of the time of this publication) English or Spanish-speaking pregnant women, aged 18 to 40 years old and with singleton pregnancies, recruited between March 1, 2014 and June 30, 2017 from the Mission Bay and San Francisco General Hospital (SFGH) hospitals at UCSF that serve a racially and socioeconomically diverse population. Our study population consists of 31.5% Non-Hispanic White women, 20.7% Hispanic/Latinx women and 33.6% earns less than $100,000/year. Additional demographic data and data from medical records are shown in Tables S1 and S2.

Table 1: Demographics of the CIOB cohort (N = 295) from San Francisco, CA. When a variable is shown as "missing", it indicates that the participant did not answer that question in the questionnaire. The numbers in the parentheses show the percentages (%) and standard deviations (std) as indicated in the table.

	Population
Baseline demographic, n (%)	**295 (100)**
Maternal age, y (std)	33.2 (5.1)
Gravidity, n (std)	2.4 (1.6)
Ethnicity group 1 (%)	
African American or Black	3.7
American Indian or Alaskan Native	1.4
Asian or Asian American	11.2
White	31.5
Other	15.6
Missing	36.6
Ethnicity group 2 (%)	
Hispanic/Latino	20.7
Non-Hispanic	50.5
Missing	28.8
Income (%)	
< $40,000	21.4
$40,000-$99,999	12.2
> $100,000	65.1
Missing	1.3

2.2 Non-targeted analysis workflow

Our non-targeted analysis workflow consisted of four main steps: i) chemical analysis, ii) database searching and annotations, iii) data clean-up and processing, and iv) data analysis (Fig. 1). Briefly, we analyzed serum samples with high resolution mass spectrometry and deduced chemical formulas from the detected molecular masses. We conducted MS/MS fragmentation for selected chemicals and tentatively confirmed the presence of a chemical by matching the experimental spectrum to database spectra, including experimental and *in silico* predicted spectra. We then used analytical standards for a select number of chemicals to confirm with the highest level of confidence. For our annotations, we employed the annotation scheme proposed by Schymanski et al.[19], where level 1 annotations are confirmed chemicals with analytical standards, level 2 annotations are tentative identifications with MS/MS spectra, level 3 annotations have some diagnostic evidence based on literature and data sources, and level 4 annotations are just molecular formulas without proposed structures. We examined the presence of the chemicals in chemical databases to search for potential matches to industrial uses. The details of the analytical method are described in the sections below.

In an attempt to navigate the complexity and high dimensionality of non-targeted analysis datasets, we selected and applied various software tools that helped us analyze our data and interpret our findings. The selection of the software packages was done based on the specific aims we tried to address in every step in our workflow (Fig. 1). When selecting software packages, we had to consider the capabilities of the software to address the aims of our study. For our purposes, we used i) commercial software (e.g., Agilent software packages) when available and suitable, ii) open-source tools if their application made an important contribution or offered a different approach compared to the commercial software (e.g., MS-Dial and different MS/MS databases), and iii) in-house built algorithms if we were not able to find an existing tool that could help us tackle a certain challenge in our study (e.g., level 3 annotations[19] for man-made/industrial chemicals). In the sections below we provide an explanation for the selection of each package.

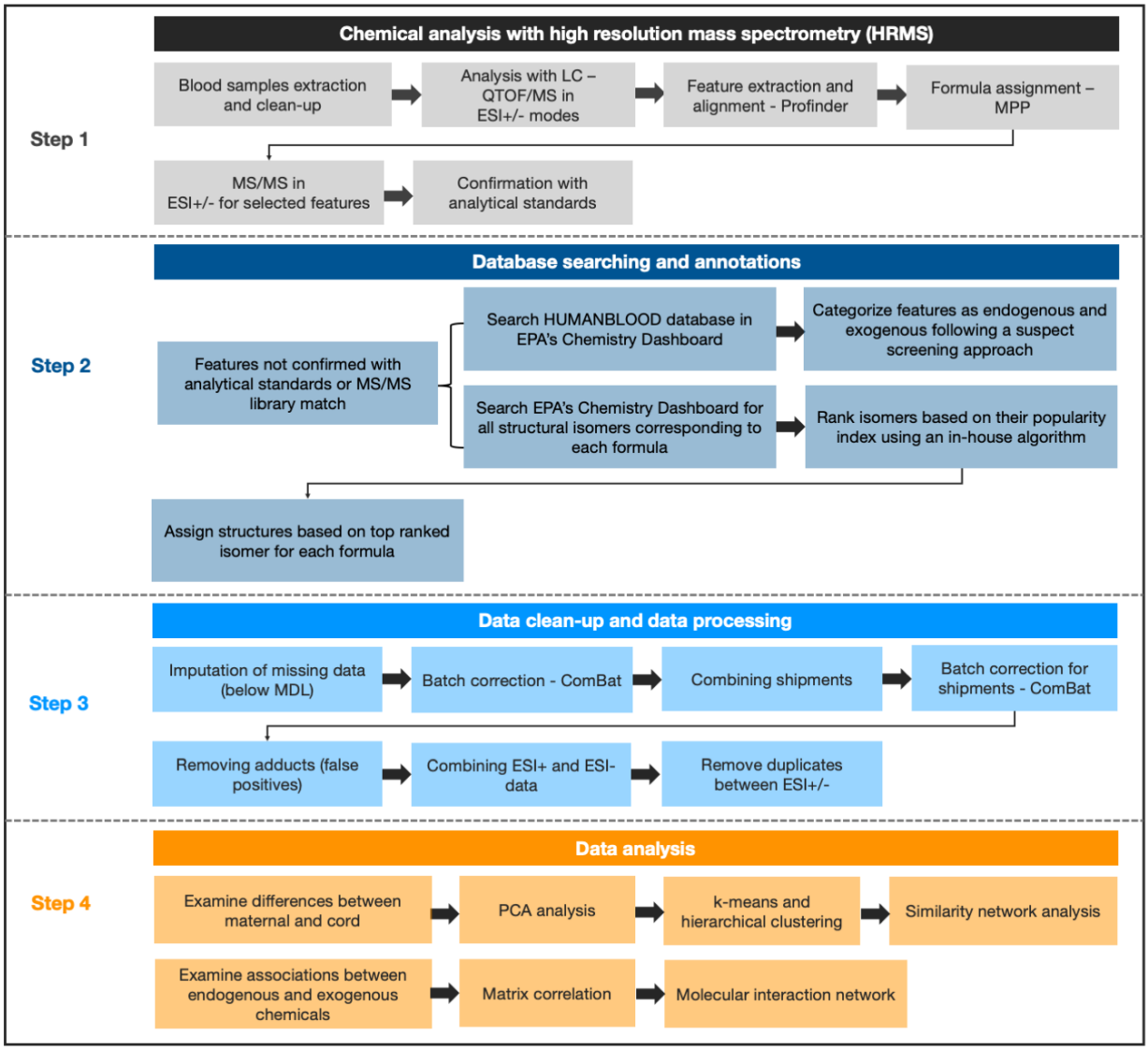

Figure 1: Flowchart describing the individual steps of analyzing the maternal and cord samples and processing the collected data from our LC/QTOF nontargeted analysis.

2.3 Sample preparation

We analyzed 295 maternal and 295 matched cord blood samples (n total = 590). The blood samples were stored in the freezer at -80 °C at the University of California, San Francisco (UCSF). Prior to analysis, the samples were centrifuged (3000 rpm) to separate the serum from the red platelets. The serum samples were transported on

dry ice to the Environmental Chemistry Laboratory (ECL) of the Department of Toxic Substances Control (DTSC) of California, in Berkeley, CA. The method is described in detail below and in our previous study.[14] Briefly, aliquots of 250 µL of serum were extracted by protein precipitation with methanol and the samples were mixed and stored at 4 ºC until they were analyzed with ultra-high pressure liquid chromatography – quadrupole time-of-flight / mass spectrometry (UPLC-QTOF/MS). At the time of analysis, 10 µL of extract were injected into the UPLC-QTOF/MS system.

2.4 Instrumental analysis

The extracts were analyzed with an Agilent UPLC coupled to an Agilent 6550 QTOF (Agilent Technologies, Santa Clara, CA) operated in both positive and negative electrospray ionization modes (ESI+ and ESI-). Full scan accurate mass spectra (MS) were acquired in the range of 100-1000 Da with resolving power of 40,000 and a mass accuracy of <5 ppm. The MS/MS fragmentation ion spectra (MS/MS) were collected at 10, 20 and 40 eV collision energies and a mass accuracy of 10 ppm. The QTOF was calibrated before each batch and the mass accuracy was regularly corrected with reference standards of reference masses 112.985587 and 1033.988109. The UPLC was operated with an Agilent Zorbax Extend-C18 column (2.1 x 50 mm, 1.8 µm) and a gradient solvent program of 0.3 mL/min with 5 mM ammonium acetate in 90% methanol/water increasing the organic phase from 10% to 100% over 15 min, following a 4 min equilibration at 100%.

The collected data from the total ion chromatograms (TIC) were processed with Agilent MassHunter Profinder for feature extraction. The features were then aligned using Mass Profiler Professional (MPP) across all batches and the features found in blanks were subtracted from the samples. The features were matched to formulas via screening with an in-house database of 2,420 unique formulas. The database was originally compiled to contain 3,535 structures of exogenous chemicals of interest based on a literature search and expert curation. Briefly, the database was compiled with the purpose of gathering manmade chemicals of high production volumes and chemicals of concern for environmental health scientists due to their potential for adverse health effects. The original database and the steps for its compilation are presented in our previous study.[14] However, in this study, we expanded our database by including all isomers corresponding to the 2,420 formulas and could be found on EPA's Dashboard[20]. After collecting all structural isomers, the updated version of the database contained 65,535

compounds (Supporting Spreadsheet 0-database). The updated version of the database contains both endogenous and exogenous compounds, however, the vast majority of the features are exogenous. Matched features were evaluated based on mass accuracy and isotopic pattern. Features of interests were prioritized for validation of identification with data dependent acquisition and with targeted MS/MS. The MS/MS spectra of the prioritized features were reviewed by empirical check of possible fragmentation peaks and were compared with spectra in online experimental MS/MS databases: MassBank of Europe and North America[21–23], Human Metabolome Database (HMDB)[24,25] and mzCloud[26] and with support from *in silico* fragmentation tools: CFM-ID[27,28] (Competitive Fragmentation Modeling for Metabolite Identification).

The acquired spectra were then used to search both experimental and in silico databases for potential matches with at least one fragment peak, aside from the molecular ion, and within a mass error of 10 ppm. We limited our search to chemical features for which we could observe a clear chromatographic peak for the molecular ion and for which the isotopic pattern match gave a score of 70 or higher. We then used the top candidate structure proposed by the software to annotate the chemical features for which we found potential matches.

In addition to MassHunter Profinder, we also utilized MS-Dial[29], which is an open source software for high-resolution mass spectrometry (HRMS) data processing and it was developed at University of California, Davis and the RIKEN Center for Sustainable Resource Science (Japan).[29] Adding MD-Dial to our search, enabled us to expand our search with additional databases. For MS-Dial, we used the same software parameters as for MassHunter Profiler (Supporting Spreadsheet 1). The databases we used were: “All public MS/MS databases for positive MS/MS” (13,303 unique compounds) and “All public MS/MS databases for negative MS/MS” (12,879 unique compounds).

Finally, matched chemical features were further compared with purchased reference standards for confirmation. The confirmation with chemical standards was done by comparing the retention times (RTs) and the MS/MS spectra of the chemical feature in the sample to the analytical standard. The selection of features for confirmation with analytical standards is described in detail in our earlier study.[14]

2.5 Quality assurance (QA) / Quality control (QC)

Extraction blanks, spike blanks and QC samples were included with each set of 20 extracted samples. Every batch analyzed with LC-QTOF/MS was accompanied by a water blank, a matrix blank and a matrix spike analyzed in the same sequence. QC samples were used to monitor the instrument's performance by inspecting RT shifts, changes in mass accuracy and changes in peak intensity. In ESI+, we used triphenyl phosphate D15 and DL-cotinine (methyl D3) as internal standards, while in ESI-, we used Perfluoro-n-[1,2-13C2] octanoic acid (M2PFOA). We used blank samples to correct the abundances of the chemical features and to remove features for which the abundances in the samples were not higher than two times that found in the blanks. The blanks consisted of LCMS grade ultraclean water (Water, Burdick & Jackson™ for HPLC, LC365-1) and were processed in the same way as the samples. The QC samples consisted of commercially available human AB serum (Corning™ Human AB Serum, 35060CI) spiked with 7 poly and perfluoroalkyl substances (PFAS) and 6 organophosphate flame retardants (OPFRs) (Supporting Spreadsheet 1: QC samples) at 10 ng/ml. The QC samples were treated in the same way as the real samples and followed the same process (Supporting Spreadsheet 1).

2.6 Database searching for feature annotation

We used a suspect screening approach for annotation. First, we searched the HUMANBLOOD database in EPA's Chemistry Dashboard[20], which contains chemicals that are endogenous and have been previously detected in human blood. The database is an aggregate from public resources, including the Human Metabolome Database (HMDB)[24], WikiPathways[30], Wikipedia[31] and literature articles[20]. The database excludes metals, metal ions, gases, drugs and drug metabolites. Screening this database allowed us to distinguish between features that are more likely to be endogenous and features that are more likely to be exogenous. To do that, we searched every formula in the database and marked the ones that had a hit in the database. Then, we labeled all features corresponding to these formulas as endogenous and the remaining as exogenous. The rationale behind this approach is that since we know we are analyzing blood samples and HUMANBLOOD is an extensive database about all endogenous compounds that have been previously detected in blood, if a detected feature in our samples has a formula that is present in the HUMANBLOOD database, then that feature is most likely an endogenous compound. We then searched the HUMANBLOOD database for all isomers corresponding to our endogenous

223 formulas and the remaining databases in EPA's Chemistry Dashboard for all isomers corresponding to our
224 exogenous formulas. We then applied an algorithm developed by the first author, Dr. Abrahamsson, to rank the
225 isomers of each formula based on (i) total number of available isomers on the Dashboard, (ii) the number of data
226 sources in the Chemistry Dashboard, (iii) number of PubChem data sources, and (iv) number of PubMed
227 publications. We then used the top ranked isomer to annotate the chemical features that were not confirmed with
228 MS/MS spectra matching or with analytical standards. For example, searching $C_8HF_{17}O_3S$ gives us two isomers:
229 perfluorodecanoic acid and perfluoro-3,7-dimethyloctanoic acid. If we were to randomly select one of the isomers
230 our probability of picking the right isomer would be 0.5. Then, making the assumption that more prevalent
231 isomers have a higher number of literature and data sources, we can adjust that probability by taking into account
232 that information after normalizing all numbers for (ii), (iii), and (iv) from 0-1. So, while the probability of
233 randomly picking the right isomer for $C_8HF_{17}O_3S$ is 0.5, perfluorodecanoic acid has a higher probability (0.73) of
234 being the right isomer because it has more literature and data sources than perfluoro-3,7-dimethyloctanoic acid
235 (0.27). It is important to acknowledge that these estimates are amenable to change as EPA's Chemistry Dashboard
236 is a dynamic project and keeps being updated with additional chemicals. Furthermore, these annotations may be
237 susceptible to the Matthew effect[32], where researchers prioritize chemicals to study mainly because other
238 researchers have prioritized the same chemicals. However, since these are just annotations and serve only in
239 providing diagnostic evidence for the identification of chemical compounds, we deemed them as sufficient for
240 that purpose. The code for the algorithm is available on GitHub
241 (https://github.com/dimitriabrahamsson/nontarget-maternalcord.git).
242 In order to evaluate the effectiveness of the algorithm, we compared the level 3 annotations of the
243 algorithm to the level 1 and 2 annotations and observed how many times the predictions of the algorithm agreed
244 with the level 1 and 2 annotations (Supporting Spreadsheet 1: algorithm validation). Although the level 3
245 annotations are just annotations and not confirmations, in some cases they can be very informative and help
246 compose a diagnostic picture for the underlying structure of a detected chemical feature. This is particularly
247 helpful for certain chemicals that are more targetable than others. For instance, the presence of fluorine in a
248 formula would indicate that this compound is an exogenous compound and it most likely belongs to the category

of poly and perfluoroalkyl substances (PFAS). Another example is when a chemical formula has only a limited number of potential isomers (e.g., 5-10 isomers) and all potential isomers are endogenous compounds with very similar function and properties (e.g. chenodeoxycholic acid).[20]

2.7 Data clean-up and data processing

2.7.1 Imputation of values below detection limit

To impute below detection limit values, we used a computational approach which assigned missing values based on the distribution of the data points. We log transformed the data from the MS analysis for each chemical across samples and calculated the median, the minimum and the standard deviation of the distribution. We then fit a normal distribution to the data points based on the median and the standard deviation that we calculated from the experimental data. The model then generated random values between the minimum measured experimental value (~5,000) and the absolute minimum (0). The minimum measured value is dependent on the cut-off point set in the software during the first processing steps of the chromatograms. Since in non-targeted analysis studies the true method detection limit in unknown, this cut-off point is set so that it represents a safe margin from the baseline of the chromatogram. So, for example, if the abundance for the baseline is 1,000 then the cut-off point is set as 5 x 1,000. The code for the imputation is available as supporting information on GitHub (https://github.com/dimitriabrahamsson/nontarget-maternalcord.git)

2.7.2 Batch correction

We analyzed 590 samples in total consisting of 295 maternal and 295 cord blood samples. The samples were analyzed in two shipments of approximately 300 samples (150 maternal samples and 150 cord samples) in each shipment. Within a shipment, the 300 samples were analyzed in 15 batches yielding 20 samples per batch (15 batches x 20 = 300). Each batch of 20 consisted of 10 maternal and 10 cord blood samples. Before the analysis, the samples were randomized, however, in every batch, the maternal samples were analyzed with their corresponding cord samples in order to avoid introducing additional batch effects between maternal and cord samples. To clarify even further, the maternal and cord samples within each batch were randomized and were not analyzed in pairs of maternal and cord. To correct the abundances of the chemicals measured in the samples for

batch effect, we employed the ComBat package for python[33]. ComBat uses a parametric and non-parametric Bayes framework to adjust the values for batch effects. The method requires that the batch parameter is known and that the data are log transformed (method is described in detail in Johnson et al.[34]). For our dataset, we first applied the ComBat package to each shipment separately to correct for batch effect within shipment. Then we applied the package again to correct for batch effect across shipments.

2.7.3 Combining shipments

As our samples were analyzed in two separate shipments of approximately 150 samples each, one of the challenges was to combine the two datasets of the two shipments, given the potential shifts in RT and differences in peak alignment. This step was done after batch correction for within shipment variability. In order to address this issue, we grouped all chemical features by their formulas and sorted them by ascending RTs. We then created an index for each group of formulas (1, 2, 3, etc.), which we then used to create an identifier based on the formula and the position of each isomer in the index. For example, if the formula $C_5H_{13}NO$ had three isomers, the first isomer was named C5H13NO_1, the second isomer as C5H13NO_2 and the third isomer as C5H13NO_3. We then merged the two datasets on the identifier and removed features that were present in only one of the datasets. We examined the difference in the RT and molecular mass and removed those features for which RT differed by more than 0.5 min or where the mass difference was more than 15 ppm. A limitation associated with this approach is that there could be cases where we are removing valid features if the molecular formula assigned in one shipment does not match the molecular formula assigned in the other shipment. This would then lead to false negatives and can result in underestimating the number of truly detected compounds. This would be more likely to happen in instances where multiple formulas can be assigned to a given chemical feature. This challenge warrants further exploration to ensure that we can leverage the full potential of NTA datasets.

2.7.4 Removing adducts

Electrospray ionization adducts are chemicals that are formed inside the instrument during analysis of the samples as the salts ions from the electrolytes used to enhance ionization bind to the ions of the organic molecules formed during electrospray ionization. We filtered out these chemicals by identifying the features that strongly

299 correlate ($r > 0.5$) with each other and have distinct mass differences corresponding to salt ions, such as sodium
300 (Na^+), potassium (K^+), formate ($HCOO^-$) and ammonium (NH_4^+). Na^+ and K^+ adducts particularly important in
301 serum analysis as these elements occur naturally in the human body and can form adducts with analytes during
302 ionization. For filtering out adducts, we used a mass accuracy filter of 15 ppm.

303 ## 2.8 Data Analysis

304 ### 2.8.1 Abundance and frequency calculations

305 We examined the relationship between chemical features in maternal samples and cord samples in terms
306 of abundances and detection frequencies. For the abundances, we used the mean log transformed abundance of
307 each chemical in maternal samples and compared it to the corresponding feature in the cord samples using a linear
308 regression model. For the detection frequencies, we used a universal abundance cutoff of 5,000, which is
309 comparable to the minimum measured value in the chemical features (~5000). We compared the detection
310 frequencies of the chemical features between maternal and cord samples both in terms of kernel density estimates
311 and in terms of absolute numbers. We also examined the differences in detection frequencies of endogenous and
312 exogenous chemical features.

313 ### 2.8.2 Unsupervised clustering

314 We conducted a principal component analysis (PCA) to examine the differences in the PCs between
315 maternal and cord samples. We then conducted a correlation analysis, where we examined the relationship of the
316 first 3 PC components with technical features and clinical covariates, i.e., batch, shipment, sample type
317 (maternal/cord) and gestational age group (preterm/full-term). We identified the features that were differentially
318 enriched in maternal and in cord blood samples by comparing the abundances of the chemical features in maternal
319 samples to those of cord samples and marking the features that showed a significant trend to be higher in maternal
320 and lower in cord and vice versa ($p < 0.05$) after correcting for multiple hypothesis testing using the approach of
321 Benjamini-Hochberg with a false discovery rate of 5%. We checked the cluster stability by comparing the PC1
322 values of the maternal samples to the PC1 values of the cord samples using a two-sided Mann-Whitney-Wilcoxon
323 test with Bonferroni correction.

2.8.3 Network analysis for maternal and cord samples

The purpose of the network analysis was to assess whether maternal samples are more similar in terms of chemical abundances to their corresponding cord samples than to other maternal samples. For this analysis, we considered two network-based approaches.

For the first approach, we conducted a matrix correlation of all samples using a linear regression model and calculated the correlation coefficients and p-values. We then adjusted the p-values by applying a multiple hypothesis correction using the Benjamini-Hochberg correction with a false discovery rate of 5% and we marked the maternal and cord sample pairs that remained significant after the multiple hypothesis correction. We then plotted the correlations as a correlation network using the NetworkX[35] package for Python. We then divided the network into four subnetworks i) correlations between matched maternal-cord pairs only, ii) correlations between unmatched maternal cord pairs and between maternal only and cord only, iii) correlations between maternal samples only, and iv) correlations between cord samples only. We then calculated the number of connections in each subnetwork and the averages correlation coefficient for each subnetwork and compared the subnetworks to each other.

For the second approach, we carried out permutation analysis randomly picking a matched pair of a maternal and cord samples (M1 and C1), and a random maternal sample (M2) 100 times. For each iteration, we then calculated the abundance ratios of all chemical features for every sample pair (M1-C1, M1-M2 and M2-C1). Chemical features with ratios in the range of 0.75 – 1.25 were considered "similar" chemical features between two samples. We calculated the number of chemicals for each pair and compared them to each other. We calculated the average number of similar chemicals for every pair and compared the pairs to each other. The code is available on GitHub (https://github.com/dimitriabrahamsson/nontarget-maternalcord.git).

2.8.4 Partitioning of chemical features between maternal and cord

As part of our analysis, we wanted to understand why different chemicals exhibit different partitioning behaviors between maternal and cord blood. We examined the partitioning behavior of the detected chemical features between maternal and cord by calculating the cord/maternal abundance ratio (R_{CM}) as:

$$R_{CM} = \frac{A_c}{A_m}$$

where, A_c is the abundance of a chemical feature in cord blood and A_m is the abundance of a chemical in maternal blood. R_{CM} has been previously described in environmental chemistry studies[36–38] as:

$$R_{CM} = \frac{C_C}{C_M}$$

where, C_C is the concentration of a given chemical in cord blood and C_M is the concentration in maternal blood. Since concentrations are not available for all chemical features, we replaced concentration with abundance as follows:

$$R_{CM} = \frac{C_C}{C_M} = \frac{\frac{A_c}{RRF}}{\frac{A_m}{RRF}} = \frac{A_c}{A_m}$$

where, RRF is the relative response factor used to calculate concentrations assuming a linear calibration curve.

It is important to note that R_{CM} does not describe an equilibrium partition ratio, such as the octanol-water equilibrium partition ratio (K_{OW}), but rather a concentration ratio representing the current state of a dynamic system. Considering that the placenta is a dynamic system, where chemicals are transported through passive diffusion and active transport to and from the system, it is unlikely that any chemicals will be at thermodynamic equilibrium. The partitioning of chemicals between cord and maternal blood has also been described as a concentration ratio in previous studies.[36–38]

Previous studies have shown that the partitioning behavior of chemicals between maternal and cord blood is related to the chemicals' physicochemical properties[39,40] and to certain physiological parameters that can affect the placenta, such placental aging[41] and gestational diabetes[42]. In an attempt to understand the parameters determining R_{CM} we used a linear regression model to assess its relationship to physicochemical properties and physiological parameters. The physicochemical properties we used are known as the Abraham descriptors[43–45] and commonly used in quantitative structure-activity relationships (QSARs). These descriptors were: i) E, which describes a chemical's ability to engage in London dispersion forces and dipole-induced dipole interaction; ii) S, which describes a chemical's ability to engage in dipole-induced dipole and dipole-dipole interactions, iii) A,

which describes hydrogen bond acidity; iv) B, which describes hydrogen bond basicity; v) V, which is the McGowan molecular volume; and vi) L, which is the hexadecane/air partition ratio. The Abraham descriptors were obtained from the UFZ-LSER database of the Helmholtz Centre for Environmental Research-UFZ[46] (Zentrum für Umweltforschung). In addition to the Abraham descriptors, we also collected the K_{OW} of the chemicals in the dataset and examined its relationship to R_{CM}. These calculations were only applied to chemical features whose structures that were annotated with level 1-3 annotations.

The physiological parameters we used were the body-mass index (BMI), maternal age at delivery, gestational age, birth weight and gestational diabetes (Table S2). Since R_{CM} is a chemical-specific parameter and not a participant-specific parameter, in order to access its relationship to physiological parameters we calculated R_{CM} for every chemical and every maternal-cord pair and then we calculated the average value per participant, as a hypothetical R_{CM} representing the average R_{CM} of all chemicals in each participant.

2.8.5 Associations between endogenous and exogenous compounds

After calculating the number of exogenous and endogenous chemicals, as described previously in the section for database searching, we examined the associations between endogenous and exogenous compounds using the approach of molecular interaction networks. It is important to note that although these types of networks are commonly known as "molecular interaction networks"[47–50], the term "interaction" can be interpreted as in that the chemical compounds are having an effect on one another or in the epidemiological sense that two parameters are having an effect on one outcome. However, in this context, "interaction" refers to the associations between chemical features. In NTA applications, the precise relationships are still speculative and the "interactions" shown by these networks are proposed associations that need to be further explored and validated with experimentation. One important advantage of these networks is that they allow for visualization of multiple endogenous and exogenous features at once together with their inter- and intra- associations.

As a first step for our exercise, we applied a matrix correlation and calculated the correlation coefficients and p-values between all endogenous and all exogenous chemical features after adjusting the p-values for multiple hypothesis testing using the Benjamini-Hochberg approach and a false discovery rate of 5%. We applied the approach of molecular interaction networks to visualize the associations and examine the relationships between

endogenous and exogenous compounds for the significant correlations between endogenous and exogenous chemical features separately for maternal and cord samples. To build the network, we used Cytoscape[51] with Metscape[52] as a plug-in. Cytoscape[51] is an established tool in the field of bioinformatics and -omics research for the visualization of networks and assisting in the discovery of underlying biological mechanisms. Due to the large number of relationships and the complexity of the network, we focused our comparison on the chemical features that had an annotation score > 0.3, or confirmed with MS/MS or analytical standards, and had a Pearson $|r| > 0.4$.

2.9 Statistical analyses

For all the correlations mentioned in the sections above we used Pearson r and we adjusted the calculated p-values for multiple hypothesis testing using the Benjamini-Hochberg approach with a false discovery rate of 5%. When comparing two groups for statistically significant differences, such as in unsupervised clustering, we used a two-sided Mann-Whitney-Wilcoxon test with Bonferroni correction.

3. Results

3.1 Chemical analysis with LC-QTOF/MS

The recursive feature extraction and formula matching for the 295 pairs of maternal and cord blood samples (n total = 590 samples) resulted in 824 features in ESI- and 731 features in ESI+ for shipment 1, and 707 features in ESI- and 576 features in ESI+ for shipment 2. After combining the datasets for the two shipments, the resulting dataset for ESI- summed up to 412 features and the dataset for ESI+ to 298 features (n total = 710 features) after filtering out the features that showed an RT difference of > 0.5 min or a mass difference of > 15 ppm. Combining the data from ESI- and ESI+, resulted in 712 features. This number is higher by 2 features compared to the total number of ESI- and ESI+ because 1 isomer from ESI- had more than 1 possible matches from ESI+ based on the criteria that we set for merging the two datasets (RT difference of 0.5 min and mass accuracy of 15 ppm). Ten features were identified as duplicates between ESI- and ESI+ and were removed from the dataset. Seventeen features were identified as adducts and were also removed from the dataset. The complete datasets before (n = 712) and after clean-up (n = 685) are presented in Supporting Spreadsheet 1 (sheets: dataset 1.0 and dataset 2.0). We confirmed 19 unique compounds with analytical standards, we tentatively identified 73

compounds with MS/MS spectra and annotated 98 compounds using our annotation algorithm (Supporting Spreadsheets 1: level 1-2 and level 3-4).

3.2 Database searching for feature annotation

We annotated 142 features as endogenous compounds and the remaining 543 features as exogenous compounds. Among the chemical compounds with the highest annotation scores, we found 5 PFAS: perfluorohexanesulfonic acid (PFHxS), perfluorooctanesulfonic acid (PFOS), perfluorodecanoic acid (PFDA), perfluoroundecanoic acid (PFUnDA) and perfluorononanoic acid (PFNA); and 2 cyclic volatile methylsiloxanes: octamethylcyclotetrasiloxane (D4) and decamethylcyclopentasiloxane (D5) (annotations with the individual scores in Supporting Spreadsheet 1: level 3-4). PFDA, PFNA, PFHxS and PFOS were also confirmed with analytical standards (Supporting Spreadsheet 1: level 1-2). When we evaluated the performance of the algorithm used for the level 3 annotations, we observed that for compounds with annotation scores from 1-0.3, the algorithm predicted correctly 16 out of the 22 formulas that were common between level 3 and level 1 and 2 annotations, corresponding to an accuracy of 73% (Supporting Spreadsheet 1: algorithm validation). For compounds with an annotation score of 0.3-0.1, the accuracy of the algorithm was 50% and for compounds with annotation score <0.1 the accuracy dropped to 8%. As anticipated, higher annotation scores were more likely to give a correct prediction. We, therefore, considered as level 3 annotations only the compounds that had an annotation score > 0.3.

3.3 MS data clean-up and data processing

In the original dataset before batch correction, we observed two distinct clusters that corresponded to the two shipments (Fig. S2 A-F). Following a matrix correlation, we observed strong correlations between the first 3 PCs and the parameters corresponding to batch number, shipment, and sample type (maternal vs cord) (Fig. S2 I). In addition, we observed significant differences in the PCs between shipment 1 and shipment 2 (Fig. S2G), and significant differences in the PCs between maternal and cord samples (Fig. S2H). Batch correction with ComBat removed the largest part of the effects related to batch and shipment (Fig. 2D), while maintaining the differences between maternal and cord (Fig. 2E). The updated plots after batch correction (Fig. 2) also showed that there were

two main clusters of samples (Fig. 2C and 2F) that corresponded to the maternal and cord sample groups (Fig. 2E).

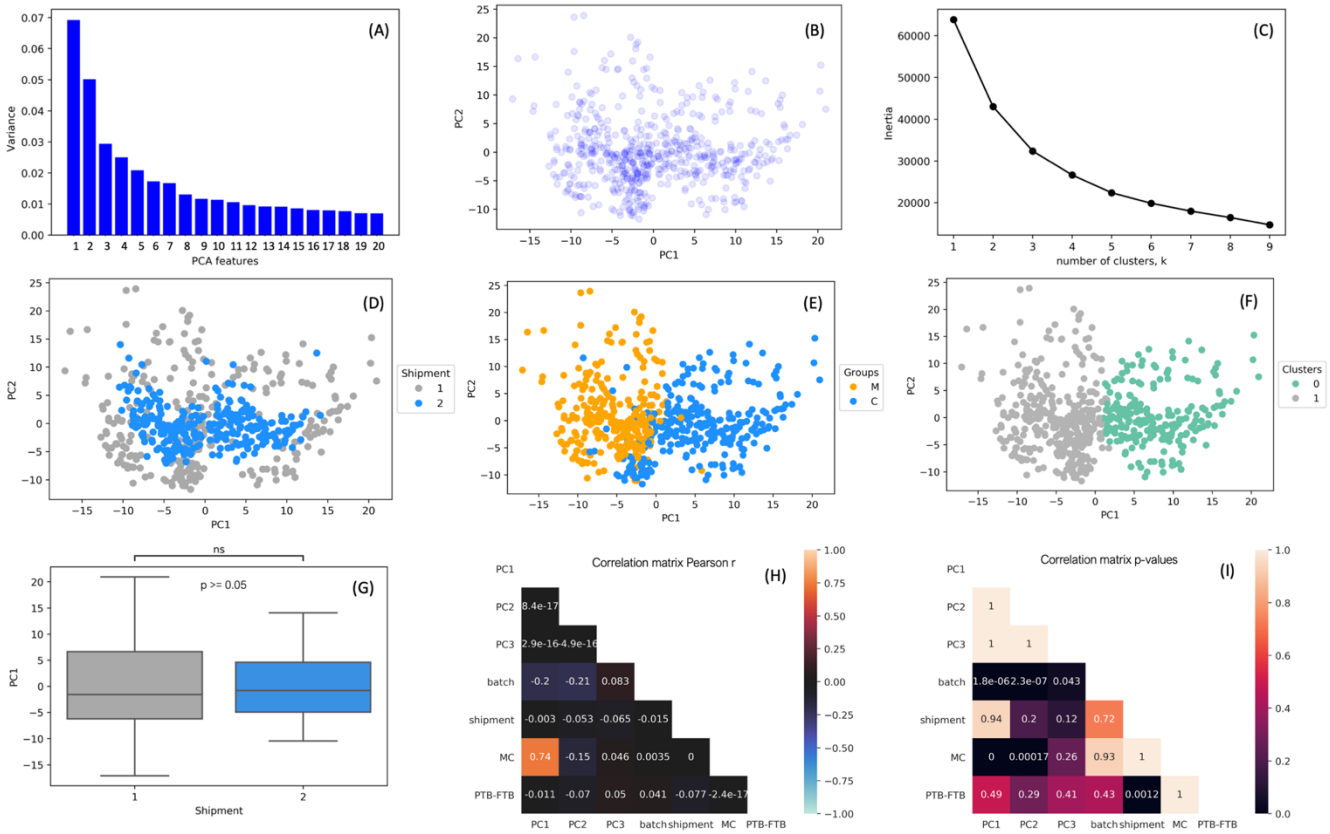

Figure 2: Results of the data analysis after batch correction with ComBat for the two shipments and the batches within each shipment. The samples were first corrected for the batches within shipment and then for the two shipments. (A): PCA features and the variance explained (%); (B) PC1 and PC2 as a scatterplot; (C) approximation of the optimal number of clusters in the dataset; (D) PC1 and PC2 color-coded by shipment; (E) PC1 and PC2 color-coded by sample type – maternal vs cord blood; (F) agnostically derived clusters using a k-means algorithm; (G) boxplot for PC1 by shipment (the error bars show the 10th and 90th percentiles, the boxes show the 25th and 75th percentiles and the middle line shows the median); (H) Pearson r values and p-values (I) for matrix correlation for PC1-3, batch, shipment, sample type maternal vs cord and full term vs preterm birth.

3.4 MS data analysis

3.4.1 Differences between maternal and cord

The maternal and cord samples showed similar profiles of detection frequency with the largest cluster of chemical features appearing at 80-100% frequency (Fig. 3B-C). We observed an overall good agreement (r = 0.93) between the mean log abundances of the chemical features in the maternal samples and the chemical features in the cord samples with some chemical features deviating from the regression line (Fig. 3A). In addition, in both maternal and cord samples the number of exogenous compounds was about 3 times higher than that of endogenous (Fig. 3D-E). This is expected considering that the vast majority of the chemicals in our database are exogenous.

We observed significant differences in PC1 between maternal and cord samples both before (Fig. S2E and S2H) and after batch correction (Fig. 2E and 2H). Removing the batch effect accentuated the differences between maternal and cord samples (Fig. 2E and 2H).

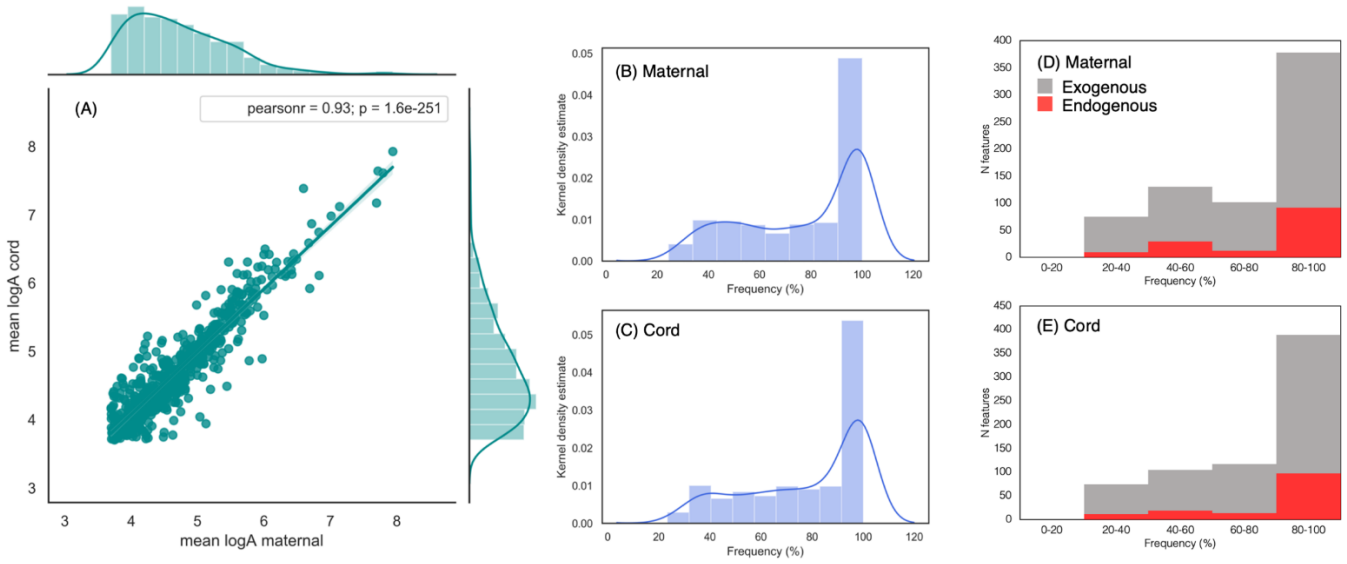

Figure 3: Correlation between maternal and cord abundances (A) (in log scale) and detection frequency calculations with kernel density curves for chemicals in maternal (B) and cord (C) blood samples (N=295 chord/maternal). The figure also displays the detection frequency for maternal (D) and cord (E) color-coded as endogenous and exogenous compounds.

Out of 685 chemical features detected in MS analysis after filtering (as described in the methods above), 450 showed a significant difference between maternal and cord samples (Fig. 4). We observed clear clustering between maternal and cord blood samples indicating a sufficient difference in the chemical composition between maternal and cord samples for them to be classified as two distinct clusters (p-value for PC1 between maternal and cord <= 0.0001; Fig. 4B).

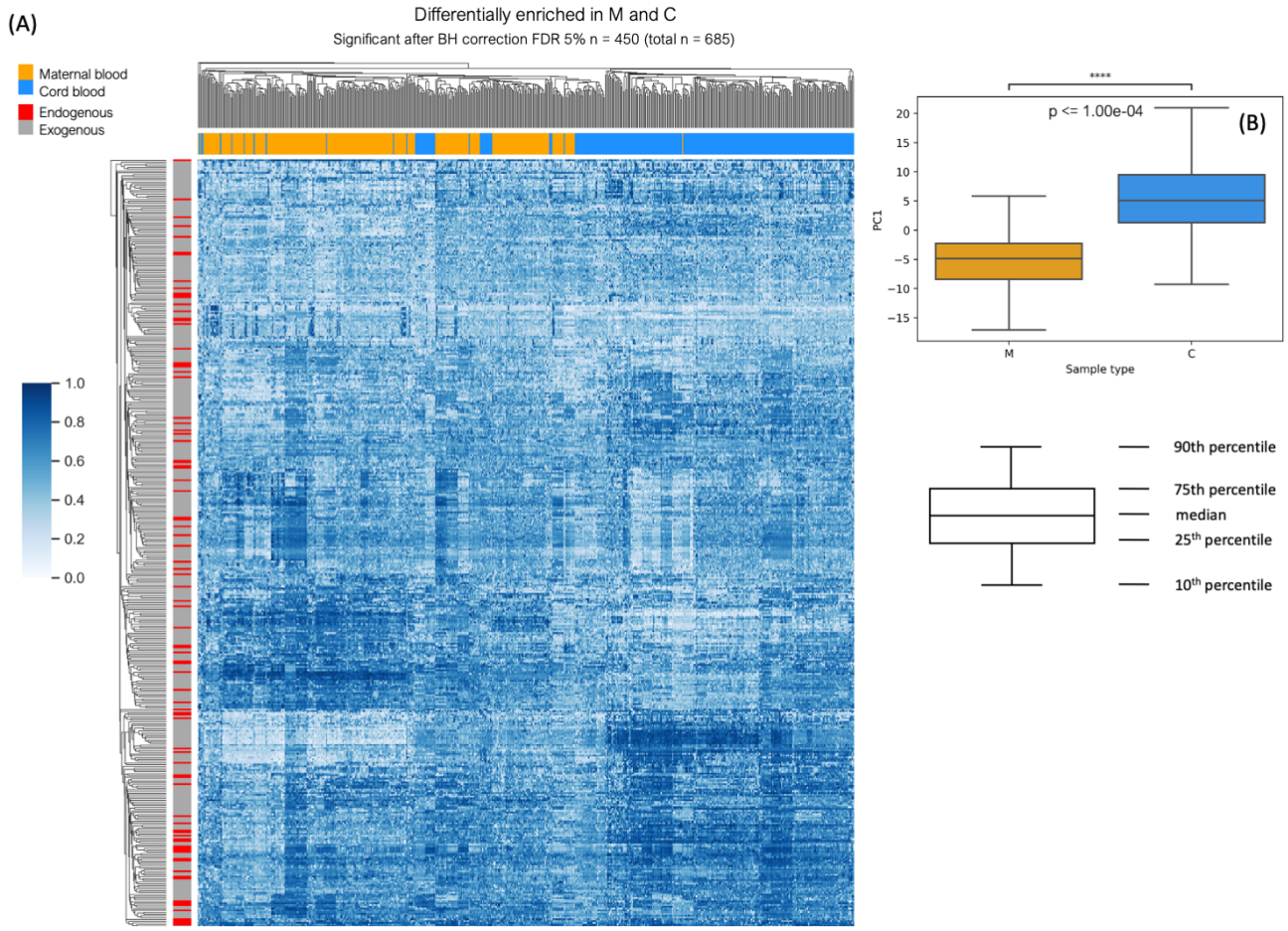

Figure 4: Clustering heatmap for maternal and cord blood samples and the chemical features that showed a significant trend to be higher in maternal or cord after multiple hypothesis correction (Benjamini-Hochberg test, 5% false discovery rate). Out of 685 chemical features in total, 450 showed a significant difference. The samples are color-coded by sample type (maternal vs cord). The features are color-coded by chemical type (endogenous vs exogenous). The error bars in the box-plot show the 10^{th} and 90^{th} percentiles, the boxes show the 25^{th} and 75^{th} percentiles and the middle line shows the median.

Our similarity network analysis using a correlation network showed that paired maternal and cord samples had a higher number of significant correlations (N = 170; Fig. S3 A) compared to unpaired maternal and cord samples (N = 84; Fig. S3 B) and compared to maternal only (N=41; Fig. S3 C) and cord only (N=41; Fig. S3

D). No significant differences were observed in the average | r | values between the four groups. Our similarity network analysis using a permutation approach showed a very similar trend (Fig. S4). The average of 100 iterations showed that paired maternal and cord samples (M1-C1) shared more similar chemical features compared to maternal – maternal pairs (M1-M2) and unmatched maternal – cord samples (M2-C1) (Fig. S4).

We observed that the majority of R_{CM} values are concentrated around 1 indicating an even partitioning between maternal and cord blood (Fig. S5 A and S5 B). R_{CM} showed a weak but significant positive correlation with RT (S5 D). No significant correlation was found for R_{CM} and molecular mass (S5 C). We also observed a significant positive association between R_{CM} and E (Fig. S6 A), a significant negative association between R_{CM} and K_{OW} (Fig S6 G), and a significant positive association between K_{OW} and RT (Fig. S6 H). We observed a borderline significant association between R_{CM} and gestational age (p-value = 0.07) (Fig. S7) and the median of the overall R_{CM} values were slightly higher in preterm birth samples compared to full term and late term. A slightly elevated median value was also observed for the gestational diabetes samples, although there was no statistically significant difference between cases and controls (Fig. S7).

3.4.2 Correlations between endogenous and exogenous compounds

We observed 21,522 significant relationships between features that were annotated as endogenous and features that were annotated as exogenous in maternal samples and 19,846 in cord samples after multiple hypothesis correction (n total relationships = 77,106 in maternal and n = 77,106 in cord samples, Fig. S8). From the significant relationships, 103 relationships in maternal and 128 relationships in cord samples had an absolute Pearson r > 0.5, 5 relationships in maternal and 4 relationships in cord samples had an absolute Pearson r > 0.7 and 1 relationship in maternal and 1 relationship in cord samples had an absolute Pearson r > 0.8 (dataset with the calculated r and p-values in the Supporting Spreadsheet 2).

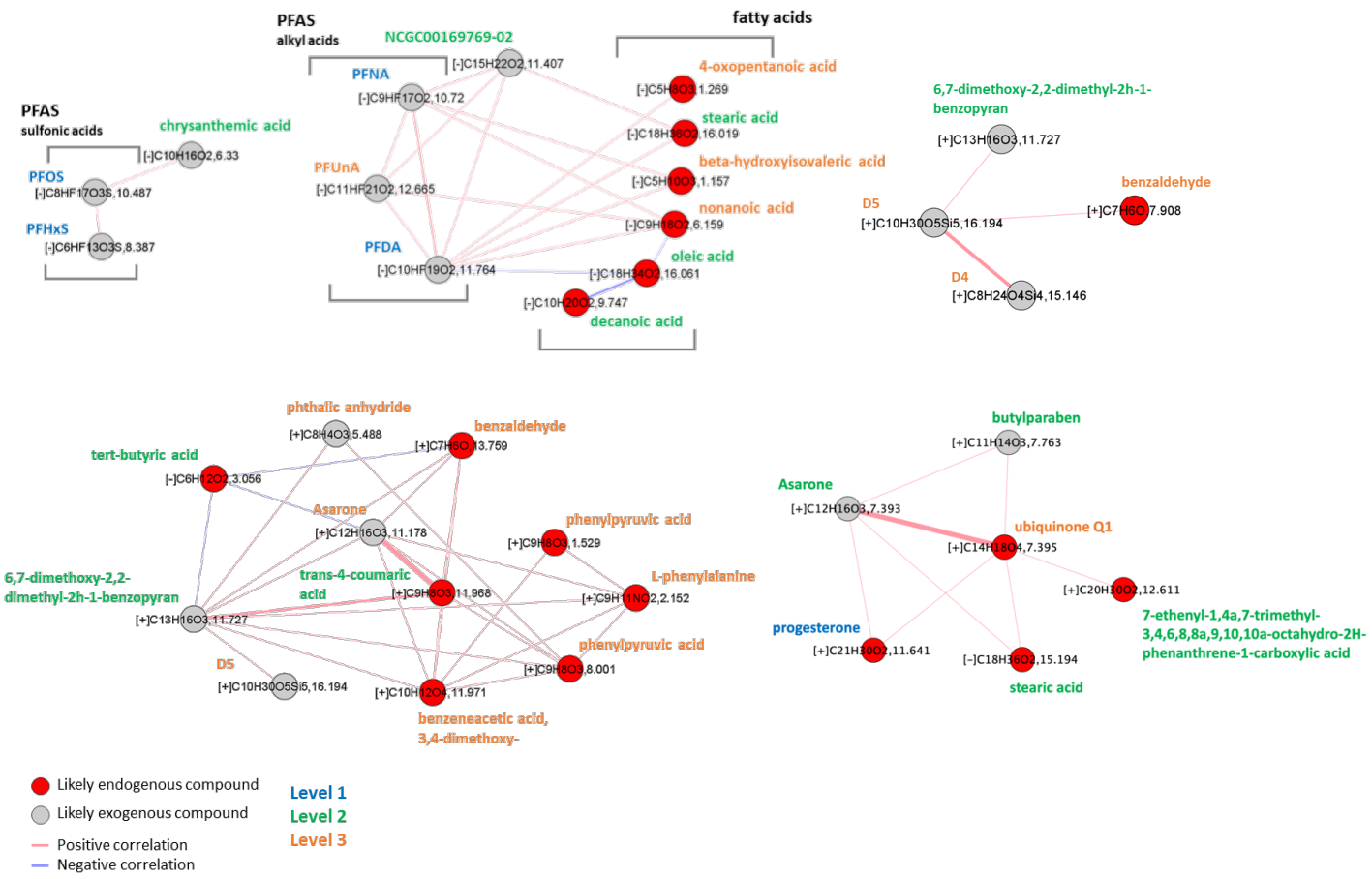

Figure 5: Molecular interaction networks for endogenous (red) and exogenous (gray) chemical features in the maternal blood samples (N = 295). The network shows the features which had an annotation score of > 0.3 or were identified with MS/MS or with analytical standards. The network shows the correlations with an absolute r > 0.4. The red lines indicate positive correlations and the blue lines indicate negative correlations. The thickness of each line indicates the strength of the correlation (|r| = 0.4 – 1). The different colors in the names of the chemicals correspond to the annotation levels of Schymanski et al.[19] showing confidence in annotation. Level 1 are compounds that have been confirmed with analytical standards, level 2 are compounds that have been tentatively identified with MS/MS spectra matching and level 3 are compounds for which we have a definitive formula and some diagnostic evidence based on our annotation algorithm described in materials and methods.

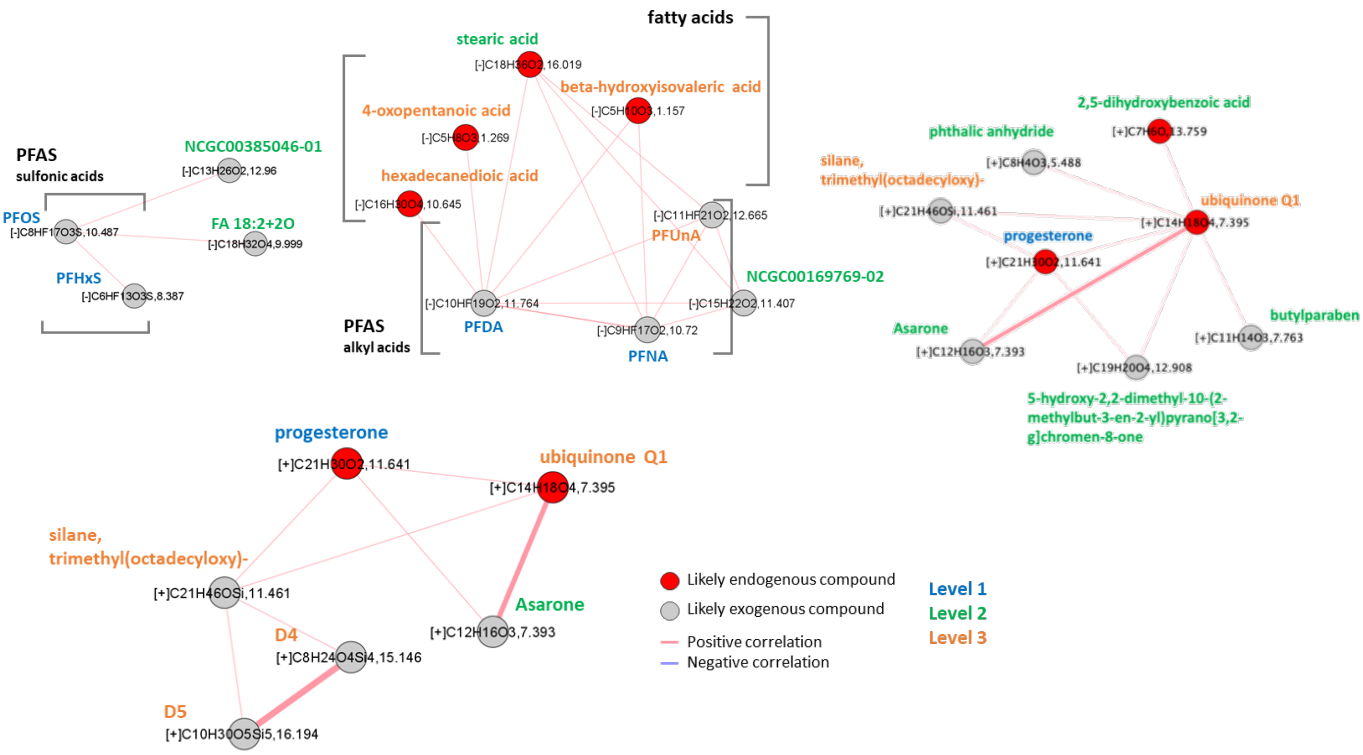

Figure 6: Molecular interaction networks for endogenous (red) and exogenous (gray) chemical features in the cord blood samples (N = 295). The network shows the features which had an annotation score of > 0.3 or were identified with MS/MS or with analytical standards. The network shows the correlations with an absolute r > 0.4. The red lines indicate positive correlations and the blue lines indicate negative correlations. The thickness of each line indicates the strength of the correlation (|r| = 0.4 – 1). The different colors in the names of the chemicals correspond to the annotation levels of Schymanski et al.[19] showing confidence in annotation. Level 1 are compounds that have been confirmed with analytical standards, level 2 are compounds that have been tentatively identified with MS/MS spectra matching and level 3 are compounds for which we have a definitive formula and some diagnostic evidence based on our annotation algorithm described in materials and methods.

The maternal and cord networks (Fig. S9 and S10) showed great overlap with most chemical compounds appearing in both networks and exhibiting similar relationships. Due to the complexity of the generated networks (Fig. S9 and S10), we extracted some example subnetworks (Fig. 5 and 6) that illustrated correlations between

endogenous and exogenous compounds. The strongest association we observed between an endogenous and an exogenous compound in both the maternal and cord networks was between ubiquinone q10 and Asarone (r = 0.82 in maternal network and r = 0.80 in cord network). We also observed two cyclic volatile methylsiloxanes (cVMS) (octamethylcyclotetrasiloxane; D4 and decamethylcyclotetrasiloxane; D5) that correlated strongly with each other (r = 0.77 in maternal network and r = 0.81 in cord network). In addition, in the maternal samples, D5 correlated with benzaldehyde (r = 0.41), while in the cord samples, D4 and D5 correlated with silane trimethyl(octadecyloxy)-, (r = 0.41 and 0.41) which in turn correlated with progesterone (r = 0.55) and ubiquinone (r = 0.45). Finally, three perfluoroalkyl acids PFAAs: perfluorononanoic acid (PFNA), perfluorodecanoic acid (PFDA) and perfluoroundecanoic acid (PFUnDA) correlated strongly with each other (r values in maternal: 0.66-0.74, r values in cord: 0.64-0.72) while 2 perfluorinated sulfonic acids (PFSA; perfluorohexanesulfonic acid, perfluorooctanesulfonic acid) formed their own group. Both groups of chemicals are poly/perfluoroalkyl substances (PFAS), a group of chemicals that has recently come under scrutiny due to their persistence, bioaccumulation potential and toxicity. The group of PFAA, in both networks, showed to correlate with certain fatty acids, such as stearic acid and 4-oxopentanoic acid (r = 0.4-0.5) (Fig. 5 and 6).

4. Discussion

Our chemical analysis of the maternal and blood samples with HRMS and a non-target analysis workflow provided important insights in the prenatal exposome, exposures to environmental pollutants, and their potential role in the development of human disease. To our knowledge, this is the largest dataset of the exposome of maternal and fetal exposures. We confirmed 19 with analytical standards (level 1), tentatively identified 73 compounds with MS/MS spectra matching (level 2) and annotated 98 features with our annotation algorithm (level 3) described in the materials and methods (Supporting Spreadsheet 1: level 1-2 and level 3-4).

Our data analysis showed that when analyzing large sample sets with non-targeted analysis, batch effects are substantial and they need to be adequately addressed before drawing any conclusions on the chemical, biological, and epidemiological importance of that collected data. ComBat[33,34] was able to remove batch effects for HRMS data for exposomics and metabolomics analyses.

Maternal and cord samples showed similarities in chemical feature enrichment (Fig. 3), but also important differences (Fig. 4) that allowed for these two groups to be classified as two distinct clusters (Fig. 4). Our similarity network analyses also showed that matched maternal and cord samples are more similar in terms of chemical feature enrichment compared to other maternal samples. These observations have important implications when studying the partitioning of chemical compounds between maternal and cord samples and when studying which chemicals show a stronger potential to cross the placenta and accumulate in the fetus. Previous studies have reported on the partitioning between maternal and cord blood,[53–56] however, the mechanism by which certain chemicals cross the placenta more readily than others requires further investigation. One interesting example of chemicals from our dataset that showed preferential partitioning for the maternal side were the five PFAS we detected. The log R_{CM} of the five PFAS ranged from -0.037 to -0.22 (Supporting Spreadsheet 1 and Fig S5 B; left tail of the distribution) indicating that the transfer of these chemicals to the fetus is to some degree inhibited by the placenta. This finding is in good agreement with previous biomonitoring studies where they examined the transplacental transfer of PFAS.[57,58] Due to their strong affinity for proteins, PFAS, bind to the proteins in the placenta and they are to some extend inhibited from reaching the fetus.[57,58]

We observed a significant positive association between R_{CM} and E, and a significant negative association between R_{CM} and K_{OW} indicating that R_{CM} is influenced by these two physicochemical properties. As E represents the ability of a chemical to engage in London dispersion forces and dipole-induced dipole interactions, its positive association with R_{CM} suggests that organic chemicals where large parts of the molecule are composed of C and H without highly electronegative atoms (e.g., Cl) are more likely to partition preferably to cord blood. The negative association of R_{CM} and K_{OW} suggests that hydrophobic molecules are likely to partition to maternal blood. This observation is in agreement with previous studies showing a negative correlation between R_{CM} and K_{OW}.[36] We observed a borderline significant relationship between R_{CM} for gestational age (0.07) (Fig. S6 C). Furthermore, when we grouped the R_{CM} values by gestational age group, we observed a slightly higher median R_{CM} for preterm birth samples compared to full term and late term (Fig. S6 E) indicating a higher overall transfer to the fetus in preterm birth. However, this also appears to depend on the chemicals and their physicochemical properties. In an earlier study on the transplacental transfer of PFAS Li et al.[41] noted the reverse trend, namely, that transfer of

PFAS was higher in full term compared to preterm birth samples. We observed a slightly elevated median for overall R_{CM} values in samples from patients with gestational diabetes. This finding, although, statistically not significant, is in agreement with the study of Eryasa et al.[42] that observed higher transplacental transfer in mothers with gestational diabetes. These observations are in agreement with the thermodynamic understanding in environmental chemistry that the behavior of chemicals is influenced by the chemicals' physicochemical properties and by the properties of their environment.[59]

We observed a weak but significant negative association between R_{CM} and RT (Fig. S5 D). As RT is a function of the chemicals' hydrophobicity (K_{OW}), with more hydrophobic chemicals exhibiting longer RTs (Fig. S6 H), its relationship with the R_{CM} indicates that more hydrophobic chemicals would show a preference to partition more to the maternal blood compared to cord blood. This finding suggests that RT could be used as a criterion for prioritizing chemical features for identification in maternal/cord blood studies and could potentially also be used in prioritization of chemicals for toxicity testing. Finally, considering that K_{OW} can vary significantly between structural isomers/isobaric features, the strong association we observed between log K_{OW} and RT (r=0.79, p=6.9e-33) gives an extra degree of confidence for our annotations of the level 1, 2 and 3 chemicals. If these annotations contained substantial errors one would expect to see greater variability in the data points for log K_{OW} and RT.

Our analysis of the associations between exogenous and endogenous exposures has provided a means to uncover chemicals potentially important to biological pathways. Such findings are particularly useful because they can be used to inform toxicological laboratory experiments to study the underlying molecular mechanisms. We observed thousands of significant relationships between exogenous and endogenous chemical features, hundreds of which showed an absolute $r > 0.5$. Many of these associations can be challenging to interpret in terms of molecular mechanisms. Thus, we focused our discussion on associations that were both strong in terms of correlation coefficient but also relatively easily interpretable.

The strongest association we observed between an endogenous and an exogenous compound in both the maternal and the cord networks was that of ubiquinone q10 and Asarone. Ubiquinone q10 occurs naturally in the human body in an oxidized (ubiquinone) and a reduced form (ubiquinol).[60] Ubiquinone acts as an electron and

proton carrier in mitochondrial electron transport connected to ATP synthesis. Ubiquinol acts as an antioxidant inhibiting lipid peroxidation, protecting mitochondrial inner-membrane proteins and protecting DNA damage due to oxidation.[60] Asarone is a chemical compound that occurs naturally in some plants, such as *Acorus calamus* and it is used as a pesticide and as an essential oil in perfumes and in alcoholic beverages.[61] Asarone is a carcinogenic compound whose epoxide metabolite of is suspected of causing DNA damage.[62] Based on the strong association we observed for these two compounds, we hypothesize that exposure to Asarone may trigger the upregulation of ubiquinone and ubiquinol. Despite its industrial applications, Asarone appeared to not be registered as a high production volume chemical and it was not included in the Chemical Data Reporting database (CDR) under the Toxic Substances Control Act (TSCA)[63]. This raises some concerns about the regulation of Asarone and similar toxic compounds that may have natural sources but are used in industrial applications.

Another group of exogenous chemicals that showed an interesting and pattern were three PFAS (PFNA, PFDA and PFUnA) that positively correlated strongly (r = 0.4-0.5) with endogenous fatty acids (Fig 5 and 6) indicating a potential interference with fatty acid metabolism. PFAS have been shown to interfere with fatty acid metabolism in *in vitro* toxicological studies by binding to fatty acid binding proteins.[64,65] Binding of PFAS to fatty acid binding proteins could reduce the available binding sites for endogenous fatty acids resulting in higher concentrations of fatty acids. This could explain the observed positive correlations between the three PFAS and endogenous fatty acids in our study. Similar associations between PFAS and fatty acids have been reported in previous metabolome/exposome studies[66,67], however, not for the exact same panel of PFAS and fatty acid compounds and not through an NTA workflow. Currently there are about 10,000 PFAS registered on EPA's Chemistry Dashboard, many of which do not have data on their toxicity potential in humans. Toxicological and epidemiological studies have shown that exposure to certain PFAS is associated with altered liver function[68,69], increased risk for preterm birth, low birth weight[70], and lower bone mineral density[71]. Our study corroborates the need for further experimental and modeling studies to assess the potential of the ever-increasing chemical library of PFAS and study how they interfere with human metabolism. High-throughput protein binding studies would help to elucidate some of these effects and help prioritize PFAS for biomonitoring and policy action.

687 Another group of chemicals that showed an interesting pattern were two cyclic volatile methylsiloxanes
688 (cVMS), octamethylcyclotetrasiloxane (D4) and decamethylcyclopentasiloxane (D5). cVMS are organosilicon
689 chemicals that are primarily used as carriers in personal care products, such as deodorants, and as intermediates in
690 the production of silicone polymers. Their strong positive correlation indicates a common source of exposure,
691 most likely due to use of personal care products. Their ubiquitous presence in personal care products makes it
692 very likely that these chemicals are from such applications. However, also because of their ubiquitous presence in
693 silicone polymers, there is a chance that these chemicals could be a result of contamination from inside the
694 analytical instrument. There is also a possibility that these chemicals could be also coming from personal care
695 products by people working in the lab, however, the physicochemical properties of D4 and D5, specifically their
696 equilibrium partition ratio between octanol and air (K_{OA}), indicates that partitioning from the air to an organic
697 solvent is very unlikely. D4 has a log K_{OA} of 4.97 and D5 has log K_{OA} of 3.94,[20] which indicate a strong
698 preference for the molecules to exist in the gas phase compared to other chemicals, such as polychlorobiphenyl
699 180 (PCB 180) which has a log K_{OA} of 9.94 and a much stronger preference to partition to octanol. Finally, all the
700 abundances in our data set were blank corrected which should minimize the potential of contamination. In the
701 maternal samples, D5 correlated with benzaldehyde which is a compound that occurs naturally in plants and in the
702 human body, and it is used as an additive in foods and personal care products.[72] The correlation with D5 indicates
703 a common source of exposure through personal care products. In the cord samples, D4 and D5 correlated with
704 silane trimethyl(octadecyloxy)-, which in turn correlated with progesterone and ubiquinone. Silane
705 trimethyl(octadecyloxy)- is an organosilicon compound used in personal care products[73] and its correlation with
706 D4 and D5 makes good sense given the applications of these chemicals. The correlation of silane
707 trimethyl(octadecyloxy)- with progesterone and ubiquinone is somewhat concerning considering the wide use of
708 that chemical in personal care products.
709

5. Limitations and future considerations

Our study illustrates the importance of broad screening using NTA in order to characterize the exposome and the mechanisms under which environmental exposures contribute to the development of human disease. While NTA is a powerful tool in compound discovery, it also has its limitations as it is still early in its development. One critical challenge with NTA is the small number of confirmed chemicals with analytical standards, which is usually in the 10s, compared to the total number of detected features, which is usually in the 1000s.[11,12,14,74]. This obstacle restricts the ability of non-targeted analysis to assist in prioritizing chemicals for biomonitoring and human exposure studies. Developing new computational tools for structure elucidation and expanding *in silico* screening of databases for structures that correspond to detected formulas and prioritization of hazardous chemicals can potentially help enhance our ability to utilize the potential of NTA.

A limitation of our study is that it uses only one analytical instrument, LC-QTOF/MS, which specializes in the analysis and identification of polar and involatile compounds. As a result, the chemical features that we detected are primarily from that physicochemical space. Complementing LC-QTOF/MS with gas chromatography / mass spectrometry, especially high-resolution mass spectrometry and multidimensional techniques, could help expand the spectrum of possible chemical features by including non-polar and volatile/semi-volatile chemicals.

Finally, our study focuses on the differences between maternal and cord blood as a surrogate for understanding fetal exposure and adverse fetal health outcomes. However, adverse fetal health outcomes depend not only on the amount of the chemical the fetus is exposed to, but also on the toxicity of the chemical. There is thus a need to develop high-throughput toxicity screening models to screen for chemicals found in fetal blood. Using NTA data to inform toxicity testing can provide unique insights in toxicology and environmental health and assist in preventing of exposure to toxic chemicals.

In our future studies, we plan to conduct epidemiological analyses by further examining the correlations of exogenous compounds with endogenous metabolites and examine the influence of covariates on these associations. Furthermore, we plan to analyze additional samples from patients with adverse health outcomes to

734 enrich our dataset and investigate the role of endogenous and exogenous exposures to the development of adverse
735 health outcomes, such as gestational diabetes, preterm birth, birth weight, and preeclampsia, among others.
736

737 Data availability

738 All the datasets used are provided as supporting information. All the code is available on GitHub
739 (https://github.com/dimitriabrahamsson/nontarget-maternalcord.git)
740 All the MS and MS/MS files will be uploaded on the Metabolomics Workbench
741 (https://www.metabolomicsworkbench.org/) upon acceptance of the manuscript.

742 Acknowledgements

743 This study was funded by NIH/NIEHS grant numbers P30-ES030284, UG3OD023272, UH3OD023272,
744 P01ES022841, R01ES027051 and by the US EPA grant number RD83543301. We would like to thank the
745 clinical research coordinators for assisting in the collection of the samples and the data managers for their
746 assistance with the questionnaires and medical records. Finally, we would like to thank all the participants in the
747 Chemicals in Our Bodies cohort for their invaluable contribution.

748 Supporting Information

749 • Supporting Information: Figures S1-S10 and Tables S1 and S2
750 • Supporting Spreadsheet 0-database: The database of all chemical formulas and structures used in the
751 experimental analysis.
752 • Supporting Spreadsheet 1: Includes tables/spreadsheets referenced throughout the manuscript
753 • Supporting Spreadsheet 2-statistics: Includes correlation matrix data for endogenous and exogenous
754 compounds
755 • Supporting Spreadsheet 3-original data: Includes the original datasets before any processing

References

(1) Miller, G. W. *The Exposome*; Elsevier, 2014. https://doi.org/10.1016/C2013-0-06870-3.

(2) Wild, C. P. Complementing the Genome with an "Exposome": The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiol Biomarkers Prev* **2005**, *14* (8), 1847–1850. https://doi.org/10.1158/1055-9965.EPI-05-0456.

(3) Rager, J. E.; Bangma, J.; Carberry, C.; Chao, A.; Grossman, J.; Lu, K.; Manuck, T. A.; Sobus, J. R.; Szilagyi, J.; Fry, R. C. Review of the Environmental Prenatal Exposome and Its Relationship to Maternal and Fetal Health. *Reproductive Toxicology* **2020**, *98*, 1–12. https://doi.org/10.1016/j.reprotox.2020.02.004.

(4) Wang, A.; Padula, A.; Sirota, M.; Woodruff, T. J. Environmental Influences on Reproductive Health: The Importance of Chemical Exposures. *Fertil. Steril.* **2016**, *106* (4), 905–929. https://doi.org/10.1016/j.fertnstert.2016.07.1076.

(5) Gluckman, P. D.; Hanson, M. A. Living with the Past: Evolution, Development, and Patterns of Disease. *Science* **2004**, *305* (5691), 1733–1736. https://doi.org/10.1126/science.1095292.

(6) Stillerman, K. P.; Mattison, D. R.; Giudice, L. C.; Woodruff, T. J. Environmental Exposures and Adverse Pregnancy Outcomes: A Review of the Science. *Reprod Sci* **2008**, *15* (7), 631–650. https://doi.org/10.1177/1933719108322436.

(7) US EPA, O. TSCA Chemical Substance Inventory https://www.epa.gov/tsca-inventory (accessed Mar 18, 2020).

(8) US EPA, O. TSCA Inventory Notification (Active-Inactive) Rule https://www.epa.gov/tsca-inventory/tsca-inventory-notification-active-inactive-rule (accessed Mar 18, 2020).

(9) Wang Aolin; Gerona Roy R.; Schwartz Jackie M.; Lin Thomas; Sirota Marina; Morello-Frosch Rachel; Woodruff Tracey J. A Suspect Screening Method for Characterizing Multiple Chemical Exposures among a Demographically Diverse Population of Pregnant Women in San Francisco. *Environmental Health Perspectives 126* (7), 077009. https://doi.org/10.1289/EHP2920.

(10) Moschet, C.; Anumol, T.; Lew, B. M.; Bennett, D. H.; Young, T. M. Household Dust as a Repository of Chemical Accumulation: New Insights from a Comprehensive High-Resolution Mass Spectrometric Study. *Environ. Sci. Technol.* **2018**, *52* (5), 2878–2887. https://doi.org/10.1021/acs.est.7b05767.

(11) Newton, S. R.; McMahen, R. L.; Sobus, J. R.; Mansouri, K.; Williams, A. J.; McEachran, A. D.; Strynar, M. J. Suspect Screening and Non-Targeted Analysis of Drinking Water Using Point-of-Use Filters. *Environmental Pollution* **2018**, *234*, 297–306. https://doi.org/10.1016/j.envpol.2017.11.033.

(12) Rager, J. E.; Strynar, M. J.; Liang, S.; McMahen, R. L.; Richard, A. M.; Grulke, C. M.; Wambaugh, J. F.; Isaacs, K. K.; Judson, R.; Williams, A. J.; Sobus, J. R. Linking High Resolution Mass Spectrometry Data with Exposure and Toxicity Forecasts to Advance High-Throughput Environmental Monitoring. *Environment International* **2016**, *88*, 269–280. https://doi.org/10.1016/j.envint.2015.12.008.

(13) Phillips, K. A.; Yau, A.; Favela, K. A.; Isaacs, K. K.; McEachran, A.; Grulke, C.; Richard, A. M.; Williams, A. J.; Sobus, J. R.; Thomas, R. S.; Wambaugh, J. F. Suspect Screening Analysis of Chemicals in Consumer Products. *Environ. Sci. Technol.* **2018**, *52* (5), 3125–3135. https://doi.org/10.1021/acs.est.7b04781.

(14) Wang, A.; Abrahamsson, D. P.; Jiang, T.; Wang, M.; Morello-Frosch, R.; Park, J.-S.; Sirota, M.; Woodruff, T. J. Suspect Screening, Prioritization, and Confirmation of Environmental Chemicals in Maternal-Newborn Pairs from San Francisco. *Environ. Sci. Technol.* **2021**. https://doi.org/10.1021/acs.est.0c05984.

(15) Schultes, L.; van Noordenburg, C.; Spaan, K. M.; Plassmann, M. M.; Simon, M.; Roos, A.; Benskin, J. P. High Concentrations of Unidentified Extractable Organofluorine Observed in Blubber from a Greenland Killer Whale (Orcinus Orca). *Environ. Sci. Technol. Lett.* **2020**, *7* (12), 909–915. https://doi.org/10.1021/acs.estlett.0c00661.
(16) Sobus, J. R.; Grossman, J. N.; Chao, A.; Singh, R.; Williams, A. J.; Grulke, C. M.; Richard, A. M.; Newton, S. R.; McEachran, A. D.; Ulrich, E. M. Using Prepared Mixtures of ToxCast Chemicals to Evaluate Non-Targeted Analysis (NTA) Method Performance. *Anal Bioanal Chem* **2019**, *411* (4), 835–851. https://doi.org/10.1007/s00216-018-1526-4.
(17) Lai, A.; Singh, R. R.; Kovalova, L.; Jaeggi, O.; Kondić, T.; Schymanski, E. L. Retrospective Non-Target Analysis to Support Regulatory Water Monitoring: From Masses of Interest to Recommendations via in Silico Workflows. *Environmental Sciences Europe* **2021**, *33* (1), 43. https://doi.org/10.1186/s12302-021-00475-1.
(18) Singh, R. R.; Chao, A.; Phillips, K. A.; Xia, X. R.; Shea, D.; Sobus, J. R.; Schymanski, E. L.; Ulrich, E. M. Expanded Coverage of Non-Targeted LC-HRMS Using Atmospheric Pressure Chemical Ionization: A Case Study with ENTACT Mixtures. *Anal Bioanal Chem* **2020**, *412* (20), 4931–4939. https://doi.org/10.1007/s00216-020-02716-3.
(19) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **2014**, *48* (4), 2097–2098. https://doi.org/10.1021/es5002105.
(20) U.S. Environmental Protection Agency. Chemistry Dashboard https://comptox.epa.gov/dashboard/ (accessed Mar 9, 2021).
(21) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T. MassBank: A Public Repository for Sharing Mass Spectral Data for Life Sciences. *Journal of Mass Spectrometry* **2010**, *45* (7), 703–714. https://doi.org/10.1002/jms.1777.
(22) MassBank | MassBank Europe Mass Spectral DataBase https://massbank.eu/MassBank/ (accessed Jan 20, 2021).
(23) MassBank of North America https://mona.fiehnlab.ucdavis.edu/ (accessed Jan 20, 2021).
(24) Wishart, D. S.; Feunang, Y. D.; Marcu, A.; Guo, A. C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; Sayeeda, Z.; Lo, E.; Assempour, N.; Berjanskii, M.; Singhal, S.; Arndt, D.; Liang, Y.; Badran, H.; Grant, J.; Serra-Cayuela, A.; Liu, Y.; Mandal, R.; Neveu, V.; Pon, A.; Knox, C.; Wilson, M.; Manach, C.; Scalbert, A. HMDB 4.0: The Human Metabolome Database for 2018. *Nucleic acids research* **2018**, *46* (D1), D608–D617. https://doi.org/10.1093/nar/gkx1089.
(25) Human Metabolome Database http://www.hmdb.ca/ (accessed Feb 19, 2020).
(26) mzCloud - Advanced Mass Spectral Database https://www.mzcloud.org/.
(27) Allen, F.; Pon, A.; Wilson, M.; Greiner, R.; Wishart, D. CFM-ID: A Web Server for Annotation, Spectrum Prediction and Metabolite Identification from Tandem Mass Spectra. *Nucleic Acids Research* **2014**, *42* (W1), W94–W99. https://doi.org/10.1093/nar/gku436.
(28) Allen, F.; Greiner, R.; Wishart, D. Competitive Fragmentation Modeling of ESI-MS/MS Spectra for Putative Metabolite Identification. *Metabolomics* **2015**, *11* (1), 98–110. https://doi.org/10.1007/s11306-014-0676-4.
(29) CompMS | MS-DIAL http://prime.psc.riken.jp/compms/msdial/main.html (accessed Nov 13, 2020).
(30) WikiPathways - WikiPathways https://www.wikipathways.org/index.php/WikiPathways (accessed Sep 23, 2020).
(31) Wikipedia https://www.wikipedia.org/ (accessed Sep 23, 2020).

(32) Anna, S.; Sofia, B.; Christina, R.; Magnus, B. The Dilemma in Prioritizing Chemicals for Environmental Analysis: Known versus Unknown Hazards. *Environ. Sci.: Processes Impacts* **2016**, *18* (8), 1042–1049. https://doi.org/10.1039/C6EM00163G.
(33) GitHub - brentp/combat.py: python / numpy / pandas / patsy version of ComBat for removing batch effects. https://github.com/brentp/combat.py (accessed Apr 3, 2020).
(34) Adjusting batch effects in microarray expression data using empirical Bayes methods | Biostatistics | Oxford Academic https://academic.oup.com/biostatistics/article/8/1/118/252073 (accessed Apr 3, 2020).
(35) Software for Complex Networks — NetworkX 2.5 documentation https://networkx.github.io/documentation/stable/index.html (accessed Sep 11, 2020).
(36) Lancz, K.; Murínová, Ľ.; Patayová, H.; Drobná, B.; Wimmerová, S.; Šovčíková, E.; Kováč, J.; Farkašová, D.; Hertz-Picciotto, I.; Jusko, T. A.; Trnovec, T. Ratio of Cord to Maternal Serum PCB Concentrations in Relation to Their Congener-Specific Physicochemical Properties. *International Journal of Hygiene and Environmental Health* **2015**, *218* (1), 91–98. https://doi.org/10.1016/j.ijheh.2014.08.003.
(37) Aylward, L. L.; Hays, S. M.; Kirman, C. R.; Marchitti, S. A.; Kenneke, J. F.; English, C.; Mattison, D. R.; Becker, R. A. Relationships of Chemical Concentrations in Maternal and Cord Blood: A Review of Available Data. *Journal of toxicology and environmental health. Part B, Critical reviews* **2014**, *17* (3), 175–203. https://doi.org/10.1080/10937404.2014.884956.
(38) Morello-Frosch, R.; Cushing, L. J.; Jesdale, B. M.; Schwartz, J. M.; Guo, W.; Guo, T.; Wang, M.; Harwani, S.; Petropoulou, S.-S. E.; Duong, W.; Park, J. J.; Petreas, M. X.; Gajek, R.; Alvaran, J.; She, J.; Dobraca, D.; Das, R.; Woodruff, T. J. Environmental Chemicals in an Urban Population of Pregnant Women and Their Newborns from San Francisco. *Environmental Science & Technology* **2016**, *50* (22), 12464–12472. https://doi.org/10.1021/acs.est.6b03492.
(39) Li, J.; Sun, X.; Xu, J.; Tan, H.; Zeng, E. Y.; Chen, D. Transplacental Transfer of Environmental Chemicals: Roles of Molecular Descriptors and Placental Transporters. *Environ. Sci. Technol.* **2021**, *55* (1), 519–528. https://doi.org/10.1021/acs.est.0c06778.
(40) Jeong, Y.; Lee, S.; Kim, S.; Park, J.; Kim, H.-J.; Choi, G.; Choi, S.; Kim, S.; Kim, S. Y.; Kim, S.; Choi, K.; Moon, H.-B. Placental Transfer of Persistent Organic Pollutants and Feasibility Using the Placenta as a Non-Invasive Biomonitoring Matrix. *Science of The Total Environment* **2018**, *612*, 1498–1505. https://doi.org/10.1016/j.scitotenv.2017.07.054.
(41) Li, J.; Cai, D.; Chu, C.; Li, Q.; Zhou, Y.; Hu, L.; Yang, B.; Dong, G.; Zeng, X.; Chen, D. Transplacental Transfer of Per- and Polyfluoroalkyl Substances (PFASs): Differences between Preterm and Full-Term Deliveries and Associations with Placental Transporter MRNA Expression. *Environ. Sci. Technol.* **2020**, *54* (8), 5062–5070. https://doi.org/10.1021/acs.est.0c00829.
(42) Eryasa, B.; Grandjean, P.; Nielsen, F.; Valvi, D.; Zmirou-Navier, D.; Sunderland, E.; Weihe, P.; Oulhote, Y. Physico-Chemical Properties and Gestational Diabetes Predict Transplacental Transfer and Partitioning of Perfluoroalkyl Substances. *Environment International* **2019**, *130*, 104874. https://doi.org/10.1016/j.envint.2019.05.068.
(43) Abraham, M. H. Scales of Solute Hydrogen-Bonding: Their Construction and Application to Physicochemical and Biochemical Processes. *Chem. Soc. Rev.* **1993**, *22* (2), 73–83. https://doi.org/10.1039/CS9932200073.
(44) Abraham, M. H.; Ibrahim, A.; Zissimos, A. M. Determination of Sets of Solute Descriptors from Chromatographic Measurements. *Journal of Chromatography A* **2004**, *1037* (1), 29–47. https://doi.org/10.1016/j.chroma.2003.12.004.
(45) Abraham, M. H.; Smith, R. E.; Luchtefeld, R.; Boorem, A. J.; Luo, R.; Acree, W. E. Prediction of Solubility of Drugs and Other Compounds in Organic Solvents. *Journal of Pharmaceutical Sciences* **2010**, *99* (3), 1500–1515. https://doi.org/10.1002/jps.21922.

(46) Ulrich, N. ; E., S. ;. Brown, T. N. ;. Watanabe, N. ;. Bronner, G. ;. Abraham, M. H. ;. Goss, K. U. UFZ-LSER Database v 3.2 [Internet]. **2017**.
(47) Ideker, T.; Ozier, O.; Schwikowski, B.; Siegel, A. F. Discovering Regulatory and Signalling Circuits in Molecular Interaction Networks. *Bioinformatics* **2002**, *18* (suppl_1), S233–S240. https://doi.org/10.1093/bioinformatics/18.suppl_1.S233.
(48) Schramm, S.-J.; Jayaswal, V.; Goel, A.; Li, S. S.; Yang, Y. H.; Mann, G. J.; Wilkins, M. R. Molecular Interaction Networks for the Analysis of Human Disease: Utility, Limitations, and Considerations. *PROTEOMICS* **2013**, *13* (23–24), 3393–3405. https://doi.org/10.1002/pmic.201200570.
(49) Winterbach, W.; Mieghem, P. V.; Reinders, M.; Wang, H.; Ridder, D. de. Topology of Molecular Interaction Networks. *BMC Systems Biology* **2013**, *7* (1), 90. https://doi.org/10.1186/1752-0509-7-90.
(50) Interactome. *Wikipedia*; 2021.
(51) Cytoscape: An Open Source Platform for Complex Network Analysis and Visualization https://cytoscape.org/ (accessed Sep 24, 2020).
(52) MetScape 3 http://metscape.ncibi.org/ (accessed Sep 24, 2020).
(53) Kato, K.; Wong, L.-Y.; Chen, A.; Dunbar, C.; Webster, G. M.; Lanphear, B. P.; Calafat, A. M. Changes in Serum Concentrations of Maternal Poly- and Perfluoroalkyl Substances over the Course of Pregnancy and Predictors of Exposure in a Multiethnic Cohort of Cincinnati, Ohio Pregnant Women during 2003–2006. *Environ. Sci. Technol.* **2014**, *48* (16), 9600–9608. https://doi.org/10.1021/es501811k.
(54) Pan, Y.; Deng, M.; Li, J.; Du, B.; Lan, S.; Liang, X.; Zeng, L. Occurrence and Maternal Transfer of Multiple Bisphenols, Including an Emerging Derivative with Unexpectedly High Concentrations, in the Human Maternal–Fetal–Placental Unit. *Environ. Sci. Technol.* **2020**, *54* (6), 3476–3486. https://doi.org/10.1021/acs.est.0c00206.
(55) Chen, F.; Yin, S.; Kelly, B. C.; Liu, W. Isomer-Specific Transplacental Transfer of Perfluoroalkyl Acids: Results from a Survey of Paired Maternal, Cord Sera, and Placentas. *Environ. Sci. Technol.* **2017**, *51* (10), 5756–5763. https://doi.org/10.1021/acs.est.7b00268.
(56) Chen, A.; Park, J.-S.; Linderholm, L.; Rhee, A.; Petreas, M.; DeFranco, E. A.; Dietrich, K. N.; Ho, S. Hydroxylated Polybrominated Diphenyl Ethers in Paired Maternal and Cord Sera. *Environ. Sci. Technol.* **2013**, *47* (8), 3902–3908. https://doi.org/10.1021/es3046839.
(57) Mamsen, L. S.; Björvang, R. D.; Mucs, D.; Vinnars, M.-T.; Papadogiannakis, N.; Lindh, C. H.; Andersen, C. Y.; Damdimopoulou, P. Concentrations of Perfluoroalkyl Substances (PFASs) in Human Embryonic and Fetal Organs from First, Second, and Third Trimester Pregnancies. *Environment International* **2019**, *124*, 482–492. https://doi.org/10.1016/j.envint.2019.01.010.
(58) Mamsen, L. S.; Jönsson, B. A. G.; Lindh, C. H.; Olesen, R. H.; Larsen, A.; Ernst, E.; Kelsey, T. W.; Andersen, C. Y. Concentration of Perfluorinated Compounds and Cotinine in Human Foetal Organs, Placenta, and Maternal Plasma. *Science of The Total Environment* **2017**, *596–597*, 97–105. https://doi.org/10.1016/j.scitotenv.2017.04.058.
(59) Environmental Organic Chemistry, 3rd Edition | Wiley https://www.wiley.com/en-us/Environmental+Organic+Chemistry%2C+3rd+Edition-p-9781118767238 (accessed May 4, 2021).
(60) Ernster, L.; Dallner, G. Biochemical, Physiological and Medical Aspects of Ubiquinone Function. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1995**, *1271* (1), 195–204. https://doi.org/10.1016/0925-4439(95)00028-3.
(61) PubChem. PubChem https://pubchem.ncbi.nlm.nih.gov/ (accessed Aug 25, 2020).
(62) Cartus, A. T.; Schrenk, D. Metabolism of the Carcinogen Alpha-Asarone in Liver Microsomes. *Food and Chemical Toxicology* **2016**, *87*, 103–112. https://doi.org/10.1016/j.fct.2015.11.021.

(63) US EPA, O. Chemical Data Reporting under the Toxic Substances Control Act https://www.epa.gov/chemical-data-reporting (accessed May 4, 2021).
(64) Sheng, N.; Cui, R.; Wang, J.; Guo, Y.; Wang, J.; Dai, J. Cytotoxicity of Novel Fluorinated Alternatives to Long-Chain Perfluoroalkyl Substances to Human Liver Cell Line and Their Binding Capacity to Human Liver Fatty Acid Binding Protein. *Arch Toxicol* **2018**, *92* (1), 359–369. https://doi.org/10.1007/s00204-017-2055-1.
(65) Zhang, L.; Ren, X.-M.; Guo, L.-H. Structure-Based Investigation on the Interaction of Perfluorinated Compounds with Human Liver Fatty Acid Binding Protein. *Environ. Sci. Technol.* **2013**, *47* (19), 11293–11301. https://doi.org/10.1021/es4026722.
(66) Alderete, T. L.; Jin, R.; Walker, D. I.; Valvi, D.; Chen, Z.; Jones, D. P.; Peng, C.; Gilliland, F. D.; Berhane, K.; Conti, D. V.; Goran, M. I.; Chatzi, L. Perfluoroalkyl Substances, Metabolomic Profiling, and Alterations in Glucose Homeostasis among Overweight and Obese Hispanic Children: A Proof-of-Concept Analysis. *Environment International* **2019**, *126*, 445–453. https://doi.org/10.1016/j.envint.2019.02.047.
(67) Chen, Z.; Yang, T.; Walker, D. I.; Thomas, D. C.; Qiu, C.; Chatzi, L.; Alderete, T. L.; Kim, J. S.; Conti, D. V.; Breton, C. V.; Liang, D.; Hauser, E. R.; Jones, D. P.; Gilliland, F. D. Dysregulated Lipid and Fatty Acid Metabolism Link Perfluoroalkyl Substances Exposure and Impaired Glucose Metabolism in Young Adults. *Environment International* **2020**, *145*, 106091. https://doi.org/10.1016/j.envint.2020.106091.
(68) Bassler, J.; Ducatman, A.; Elliott, M.; Wen, S.; Wahlang, B.; Barnett, J.; Cave, M. C. Environmental Perfluoroalkyl Acid Exposures Are Associated with Liver Disease Characterized by Apoptosis and Altered Serum Adipocytokines. *Environmental Pollution* **2019**, *247*, 1055–1063. https://doi.org/10.1016/j.envpol.2019.01.064.
(69) Salihovic, S.; Stubleski, J.; Kärrman, A.; Larsson, A.; Fall, T.; Lind, L.; Lind, P. M. Changes in Markers of Liver Function in Relation to Changes in Perfluoroalkyl Substances - A Longitudinal Study. *Environment International* **2018**, *117,* 196–203. https://doi.org/10.1016/j.envint.2018.04.052.
(70) Li, M.; Zeng, X.-W.; Qian, Z. (Min); Vaughn, M. G.; Sauvé, S.; Paul, G.; Lin, S.; Lu, L.; Hu, L.-W.; Yang, B.-Y.; Zhou, Y.; Qin, X.-D.; Xu, S.-L.; Bao, W.-W.; Zhang, Y.-Z.; Yuan, P.; Wang, J.; Zhang, C.; Tian, Y.-P.; Nian, M.; Xiao, X.; Fu, C.; Dong, G.-H. Isomers of Perfluorooctanesulfonate (PFOS) in Cord Serum and Birth Outcomes in China: Guangzhou Birth Cohort Study. *Environment International* **2017**, *102*, 1–8. https://doi.org/10.1016/j.envint.2017.03.006.
(71) Khalil Naila; Chen Aimin; Lee Miryoung; Czerwinski Stefan A.; Ebert James R.; DeWitt Jamie C.; Kannan Kurunthachalam. Association of Perfluoroalkyl Substances, Bone Mineral Density, and Osteoporosis in the U.S. Population in NHANES 2009–2010. *Environmental Health Perspectives* **2016**, *124* (1), 81–87. https://doi.org/10.1289/ehp.1307909.
(72) PubChem. Benzaldehyde https://pubchem.ncbi.nlm.nih.gov/compound/240 (accessed May 4, 2021).
(73) PubChem. Silane, trimethyl(octadecyloxy)- https://pubchem.ncbi.nlm.nih.gov/compound/87778 (accessed May 4, 2021).
(74) Wang Aolin; Gerona Roy R.; Schwartz Jackie M.; Lin Thomas; Sirota Marina; Morello-Frosch Rachel; Woodruff Tracey J. A Suspect Screening Method for Characterizing Multiple Chemical Exposures among a Demographically Diverse Population of Pregnant Women in San Francisco. *Environmental Health Perspectives 126* (7), 077009. https://doi.org/10.1289/EHP2920.