# Comprehensive non-targeted analysis of the prenatal exposome reveals significant differences in chemical enrichment between maternal and fetal samples

Dimitri Abrahamsson[1], Aolin Wang[1], Ting Jiang[2], Miaomiao Wang[2], Adi Siddharth[1], Rachel Morello-Frosch[3], June-Soo Park[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 of America.

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

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

[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

The exposome has been recognized as an important dimension in understanding human disease and complementing the genome but remains largely uncharacterized. We analyzed 295 matched maternal and cord blood samples using non-targeted high-resolution mass spectrometry and characterized exposome features. We compared the chemical enrichment of the maternal and cord blood samples using a similarity network analysis and examined the interactions between the exogenous and the endogenous chemical features using a molecular interaction networks approach. We detected over 700 chemical features in the maternal and cord pairs and we found that maternal samples are more similar in terms of chemical enrichment to their corresponding cord samples compared to other maternal samples or other cord samples. We observed significant associations between 3 poly/perfluoroalkyl substances (PFAS) and endogenous fatty acids in both the maternal and cord samples indicating important interactions between PFAS and fatty acid regulating proteins. To our knowledge, this is the first non-targeted analysis study that uses such large cohort to characterize the prenatal exposome.

47 # 1. Introduction

48 The exposome is recognized as a critical dimension in understanding human disease by
49 complementing genetic predisposition with environmental influences. The exposome describes the sum
50 of all our exposures, both external and internal, throughout our lives from conception and onwards.[1,2]
51 Humans are exposed to multiple and variable environmental contaminants in both the indoor and
52 outdoor environments through inhalation, ingestion, and dermal absorption. Environmental exposures
53 have been shown to play an important role in the development of human disease along with exposures
54 to endogenous chemicals and genetic predisposition.[1,2]

55 Exposures to environmental contaminants during pregnancy are of critical importance due to the
56 increased risk for adverse health outcomes that occur during periods of critical and unique susceptibility
57 to biological perturbations, which can increase the risk of both maternal and child adverse health
58 outcomes[3]. Prenatal exposures to industrial chemicals have been shown to increase the risk of
59 complications during pregnancy, such as pregnancy-related hypertension, adverse birth outcomes,
60 developmental and neurodevelopmental problems during infancy, and disease during adulthood.[4–6]

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

66 Conventional biomonitoring and human exposure research rely on targeted analytical chemistry
67 techniques, in which one measures chemicals selected prior to the analysis. Up to now, with targeted
68 techniques, only about 350 chemicals are biomonitored regularly via U.S. NHANES, constituting less
69 than 1% of the chemicals used in the US. This limited number of measured targeted chemicals hinders
70 our understanding of human exposure to chemicals and how they may impact human health.

71 Considering the large number of chemicals that are not covered by these approaches, there is a need
72 to develop more high-throughput approaches that cover a broader spectrum of human exposure to
73 environmental contaminants.[9]

74 Recent advances in high-resolution mass spectrometry have brought non-targeted analysis
75 (NTA) and suspect screening to the forefront of analytical chemistry. Non-targeted analysis techniques
76 offer the possibility to screen biological and environmental samples for nearly all chemicals present in a
77 sample. Such high-throughput analytical techniques enable a more holistic characterization of the
78 exposome incorporating both internal (endogenous) and external (exogenous) exposures. However,
79 previous studies have indicated that only a small number of the detected features in a sample can be
80 confirmed with analytical standards.[10–12] The vast majority of the detected chemical features remain as
81 either detected masses or assigned formulas without information about their underlying chemical
82 structures. This obstacle significantly limits the ability of non-targeted analysis techniques to inform
83 biomonitoring studies and thus human exposures. Combining non-targeted analysis datasets with *in*
84 *silico* screening of databases for structures that correspond to detected formulas and prioritization of
85 hazardous chemicals can help enhance our ability to utilize NTA approaches.

86 We have developed an NTA method and workflow that screens human biological samples for a
87 broad spectrum of chemicals that can be identified or tentatively identified, and we apply this approach
88 to study exogenous and endogenous chemical exposures in a large racially and socioeconomically
89 diverse population of pregnant women. Our goal was three-fold: 1) to analyze 590 matched maternal
90 and cord blood samples (total 295 matched pairs) using NTA to characterize the maternal/fetal
91 exposome; 2) examine the differences in chemical feature enrichment between maternal and cord
92 blood samples; and 3) examine the interactions between exogenous chemicals and endogenous
93 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 UCSF Chemicals in Our Bodies (CIOB) study. The CIOB study consists of about 700 (recruitment is ongoing) 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.

118 Table 1: Demographics of the CIOB chemisome cohort (N = 295)

| Baseline demographic, n (%) | Population 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 |

119

120 ## 2.2 Non-targeted analysis workflow

121 Our non-targeted analysis workflow consisted of four main steps: i) chemical analysis, ii)
122 database searching and annotations, iii) data clean-up and processing, and iv) data analysis (Fig. 1).
123 Briefly, we analyzed serum samples with high resolution mass spectrometry and deduced chemical
124 formulas from the detected molecular masses. We selected a subset of chemicals for MS/MS
125 fragmentation and confirmed the presence of a chemical by matching the experimental spectrum to
126 database spectra, including experimental and *in silico* predicted spectra. The feature selection and
127 prioritization for MS/MS fragmentation is described in our previous study of Wang et al.[10] We examined
128 the presence of the chemicals in chemical databases to search for potential matches to industrial uses.
129 The details of the analytical method are described in the sections below.

130

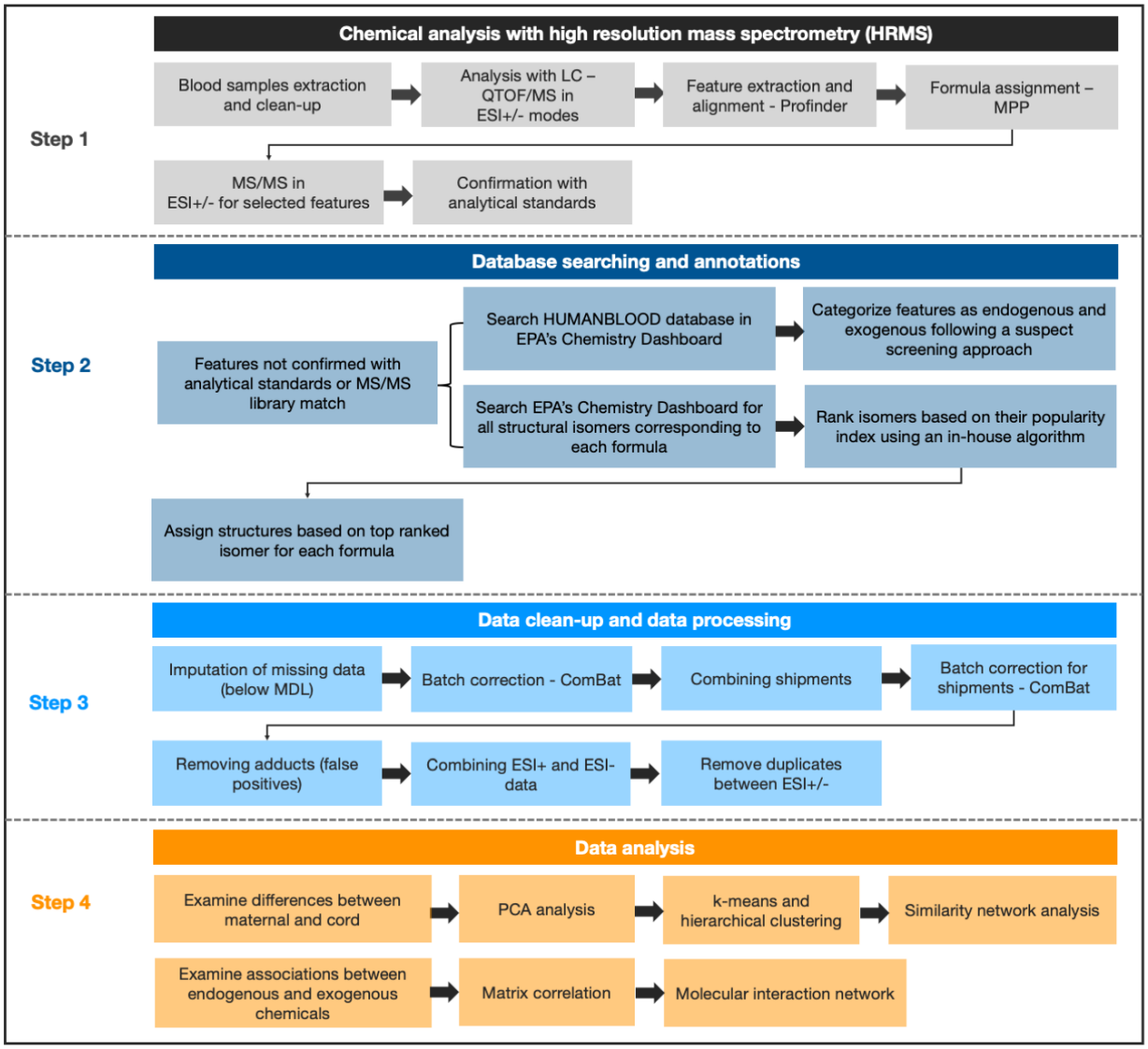



Figure 1: Flowchart describing the individual steps of analyzing the maternal and cord samples and processing the collected data from our LC/QTOF nontargeted analysis.

## 2.3 Sample preparation

The blood samples were stored in the freezer at -80 °C at the University of California, San Francisco (UCSF). 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. Samples were centrifuged (3000 rpm) to separate the serum from the red platelets and the serum was

140 extracted by protein precipitation using a variation of a method described previously.[13,14] Aliquots of 250
141 uL from each sample were spiked with 25 uL 100 ng/L surrogate standard mixture during extraction.
142 The samples were mixed and stored at 4 °C until they were analyzed with liquid chromatography –
143 quadrupole time-of-flight / mass spectrometry (LC-QTOF/MS).

144 ## 2.4 Instrumental analysis

145 The extracts were analyzed with an Agilent UPLC coupled to an Agilent 6550 QTOF (Agilent
146 Technologies, Santa Clara, CA) operated in both positive and negative electrospray ionization modes
147 (ESI+ and ESI-). Full scan accurate mass spectra (MS) were acquired in the range of 100-1000 Da with
148 resolving power of 40,000 and a mass accuracy of <5 ppm. The and MS/MS fragmentation ion spectra
149 (MS/MS) were collected at 10, 20 and 40 eV collision energies and a mass accuracy of 10 ppm. The
150 QTOF was calibrated before each batch and the mass accuracy was regularly corrected with reference
151 standards of reference masses 112.985587 and 1033.988109. The UPLC was operated with an Agilent
152 Zorbax Extend-C18 column (2.1 x 50 mm, 1.8 µm) and a gradient solvent program of 0.3 mL/min with 5
153 mM ammonium acetate in 90% methanol/water increasing the organic phase from 10% to 100% over
154 15 min, following a 4 min equilibration at 100%.

155 The collected data from the total ion chromatograms (TIC) were processed with Agilent
156 MassHunter Profinder for feature extraction. The features were then aligned using Mass Profiler
157 Professional (MPP) across all batches and the features found in blanks were subtracted from the
158 samples. The features were matched to formulas via screening with an in-house database of 2,420
159 unique formulas. The database was originally compiled to contain 3,535 structures of exogenous
160 chemicals of interest based on a literature search and expert curation. The compilation process is
161 described in our previous study.[10] However, in this study, we expanded our database by including all
162 isomers found in the EPA's Dashboard corresponding to the 2,420 formulas, which resulted in 65,535
163 compounds (Supporting Information Spreadsheets). The updated version of the database contains both

164 endogenous and exogenous compounds. Matched features were evaluated based on mass accuracy
165 and isotopic pattern. Features of interests were prioritized for validation of identification with targeted
166 MS/MS spectrometry. The MS/MS spectra of the prioritized features were reviewed by empirical check
167 of possible fragmentation peaks, comparison with spectra in online experimental MS/MS databases,
168 and support from in-silico fragmentation tools. Matched formulas were further compared with purchased
169 reference standards for confirmation.

170 ## 2.5 Quality assurance / Quality control

171 Extraction blanks, spike blanks and matrix spike blanks were included with each set of 20
172 extracted samples. Every batch analyzed with LC-QTOF/MS was accompanied by a water blank, a
173 matrix blank and a matrix spike analyzed in the same sequence.

174 ## 2.6 Database searching for feature annotation

175 We used a suspect screening approach for annotation. First, we searched the HUMANBLOOD
176 database in EPA's Chemistry Dashboard[15], which contains chemicals that are endogenous and have
177 been previously detected in human blood. The database is an aggregate from public resources,
178 including the Human Metabolome Database (HMDB)[16], WikiPathways[17], Wikipedia[18] and literature
179 articles[15]. The database excludes metals, metal ions, gases, drugs and drug metabolites. Screening
180 this database allowed us to distinguish between features that are more likely to be endogenous and
181 features that are more likely to be exogenous. To do that, we searched every formula in the database
182 and marked the ones that had a hit in the database. Then, we labeled all features corresponding to
183 these formulas as endogenous and the remaining as exogenous. The rationale behind this approach is
184 that since we know we are analyzing blood samples and HUMANBLOOD is an extensive database
185 about all endogenous compounds that have been previously detected in blood, if a detected feature in
186 our samples has a formula that is present in the HUMANBLOOD database, then that feature is most
187 likely an endogenous compound. We then searched the HUMANBLOOD database for all isomers

188 corresponding to our endogenous formulas and the remaining databases in EPA's Chemistry
189 Dashboard for all isomers corresponding to our exogenous formulas. We then applied an algorithm
190 developed by Dr. Abrahamsson to rank the isomers of each formula based on (i) total number of
191 available isomers on the Dashboard, (ii) the number of data sources in the Chemistry Dashboard, (iii)
192 number of PubChem data sources, and (iv) number of PubMed publications. We then used the top
193 ranked isomer to annotate the chemical features that were not confirmed with MS/MS spectra matching
194 or with analytical standards. For example, searching $C_8HF_{17}O_3S$ gives us two isomers:
195 perfluorodecanoic acid and perfluoro-3,7-dimethyloctanoic acid. If we were to randomly select one of
196 the isomers our probability of picking the right isomer would be 0.5. Then, making the assumption that
197 more prevalent isomers have a higher number of literature and data sources, we can adjust that
198 probability by taking into account that information after normalizing all numbers for (ii), (iii), and (iv) from
199 0-1. So, while the probability of randomly picking the right isomer for $C_8HF_{17}O_3S$ is 0.5,
200 perfluorodecanoic acid has a higher probability (0.73) of being the right isomer because it has more
201 literature and data sources than perfluoro-3,7-dimethyloctanoic acid (0.27). It is important to
202 acknowledge that these estimates are amenable to change as EPA's Chemistry Dashboard is a
203 dynamic project and keeps being updated with additional chemicals. Furthermore, these annotations
204 may be susceptible to the Matthew effect[19], where researchers prioritize chemicals to study mainly
205 because other researchers have prioritized the same chemicals. However, since these are just
206 annotations and serve only in providing diagnostic evidence for the identification of chemical
207 compounds, we deemed them as sufficient for that purpose. The code for the algorithm is available on
208 GitHub (https://github.com/dimitriabrahamsson/nontarget-maternalcord.git ).
209 Although these are just annotations and not confirmations, in some cases they can be very
210 informative and help compose a diagnostic picture for the underlying structure of a detected chemical
211 feature. This is particularly helpful for certain chemicals that are more targetable than others. For
212 instance, the presence of fluorine in a formula would indicate that this compound is an exogenous

213 compound and it most likely belongs to the category of poly and perfluoroalkyl substances (PFAS).
214 Another example is when a chemical formula has only a limited number of potential isomers (e.g., 5-10
215 isomers) and all potential isomers are endogenous compounds with very similar function and properties
216 (e.g. chenodeoxycholic acid).[15]

217 ## 2.7 Data clean-up and data processing

218 ### 2.7.1 Imputation of values below detection limit

219 To impute below detection limit values, we used a computational approach which assigned
220 missing values based on the distribution of the data points. We log transformed the data from the MS
221 analysis for each chemical across samples and calculated the median, the minimum and the standard
222 deviation of the distribution. We then fit a normal distribution to the data points based on the median
223 and the standard deviation that we calculated from the experimental data. The model then generated
224 random values between the minimum measured experimental value (~5,000) and the absolute
225 minimum (0). The code for the imputation is available as supporting information on GitHub
226 (https://github.com/dimitriabrahamsson/nontarget-maternalcord.git )

227 ### 2.7.2 Batch correction

228 The samples were analyzed in two shipments of approximately 150 each and approximately 15
229 batches within each shipment. To correct the abundances of the chemicals measured in the samples
230 for batch effect, we employed the ComBat package for python[20]. ComBat uses a parametric and non-
231 parametric Bayes framework to adjust the values for batch effects. The method requires that the batch
232 parameter is known and that the data are log transformed (method is described in detail in Johnson et
233 al.[21]). For our dataset, we first applied the ComBat package to each shipment separately to correct for
234 batch effect within shipment. Then we applied the package again to correct for batch effect between
235 shipments.

236 ### 2.7.3 Combining shipments

237 As our samples were analyzed in two separate shipments of approximately 150 samples each,
238 one of the challenges was to combine the two datasets, given the potential shifts in retention time and
239 differences in peak alignment. This step was done after batch correction for within shipment variability.
240 In order to address this issue, we grouped all chemical features by their formulas and sorted them by
241 ascending retention times. We then created an index for each group of formulas (1, 2, 3, etc.), which
242 we then used to create an identifier based on the formula and the position of each isomer in the index.
243 For example, if the formula $C_5H_{13}NO$ had three isomers, the first isomer was named C5H13NO_1, the
244 second isomer as C5H13NO_2 and the third isomer as C5H13NO_3. We then merged the two datasets
245 on the identifier and removed features that were present in only one of the datasets. We examined the
246 difference in the retention time and molecular mass and removed those features for which the retention
247 time differed by more than 0.5 min or where the mass difference was more than 15 ppm.

248 ### 2.7.4 Removing adducts

249 Electrospray ionization adducts are chemicals that are formed inside the instrument during
250 analysis of the samples as the salts ions from the electrolytes used to enhance ionization bind to the
251 ions of the organic molecules formed during electrospray ionization. We filtered out these chemicals by
252 identifying the features that strongly correlate ($r > 0.5$) with each other and have distinct mass
253 differences corresponding to salt ions, such as sodium ($Na^+$), potassium ($K^+$), formate ($HCOO^-$),
254 ammonium ($NH_4^+$) and acetonitrile ($CH_3CN$). We used a mass accuracy filter of 15 ppm.

255 ## 2.8 Data Analysis

256 ### 2.8.1 Abundance and frequency calculations

257 We examined the relationship between chemical features in maternal samples and cord
258 samples in terms of abundances and detection frequencies. For the abundances, we used the mean
259 log transformed abundance of each chemical in maternal samples and compared it to the

260 corresponding feature in the cord samples using a linear regression model. For the detection
261 frequencies, we used a universal abundance cutoff of 5,000, which is comparable to the minimum
262 measured value in the chemical features (~5000). We compared the detection frequencies of the
263 chemical features between maternal and cord samples both in terms of kernel density estimates and in
264 terms of absolute numbers. We also examined the differences in detection frequencies of endogenous
265 and exogenous chemical features.

266 ### 2.8.2 Unsupervised clustering

267 We conducted a principal component analysis (PCA) to examine the differences in the PCs
268 between maternal and cord samples. We then conducted a correlation analysis, where we examined
269 the relationship of the first 3 PC components with technical features and clinical covariates, i.e., batch,
270 shipment, sample type (maternal/cord) and gestational age group (preterm/full-term). We identified the
271 features that were differentially enriched in maternal and in cord blood samples by comparing the
272 abundances of the chemical features in maternal samples to those of cord samples and marking the
273 features that showed a significant trend to be higher in maternal and lower in cord and vice versa ($p <$
274 $0.05$) after correcting for multiple hypothesis testing using the approach of Benjamini-Hochberg with a
275 false discovery rate of 5%. We checked the cluster stability by comparing the PC1 values of the
276 maternal samples to the PC1 values of the cord samples using a two-sided Mann-Whitney-Wilcoxon
277 test with Bonferroni correction.

278 ### 2.8.3 Network analysis for maternal and cord samples

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

282 For the first approach, we conducted a matrix correlation of all samples using a linear
283 regression model and calculated the correlation coefficients and p-values. We then adjusted the p-

284 values by applying a multiple hypothesis correction using the Benjamini-Hochberg correction with a
285 false discovery rate of 5% and we marked the maternal and cord sample pairs that remained significant
286 after the multiple hypothesis correction. We then plotted the correlations as a correlation network using
287 the NetworkX[22] package for Python. We then divided the network into four subnetworks i) correlations
288 between matched maternal-cord pairs only, ii) correlations between unmatched maternal cord pairs and
289 between maternal only and cord only, iii) correlations between maternal samples only, and iv)
290 correlations between cord samples only. We then calculated the number of connections in each
291 subnetwork and the averages correlation coefficient for each subnetwork and compared the
292 subnetworks to each other.

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

301 ### 2.8.4 Partitioning of chemical features between maternal and cord

302 We examined the partitioning of the detected chemical features between maternal and cord by
303 calculating the maternal/cord abundance ratio as:

304 $$MC_{ratio} = \frac{A_m}{A_c}$$

305 where, $A_m$ is the mean abundance of a chemical feature across maternal samples and $A_c$ is the mean
306 abundance of a chemical feature across cord samples. We then used a linear regression model to

307 assess the relationship of the maternal/cord ratio to molecular mass and retention time of the chemical
308 features.

309 ### 2.8.5 Associations between endogenous and exogenous compounds

310 After we calculated the number of exogenous and endogenous chemicals, as described
311 previously in the section for database searching, we examined the associations between endogenous
312 and exogenous compounds using a molecular interaction network. First, we applied a matrix correlation
313 and calculated the correlation coefficients and p-values between all endogenous and all exogenous
314 chemical features after adjusting the p-values for multiple hypothesis testing using the Benjamini-
315 Hochberg approach and a false discovery rate of 5%. We applied the approach of molecular interaction
316 networks to visualize the associations and examine the relationships between endogenous and
317 exogenous compound for the significant correlations between endogenous and exogenous chemical
318 features separately for maternal and cord samples. For the molecular interaction network, we used
319 Cytoscape[23] with Metscape[24] as a plug-in. Cytoscape[23] is an established tool in the field of
320 bioinformatics and -omics research for the visualization of networks and assisting in the discovery of
321 underlying biological mechanisms. Due to the large number of relationships and the complexity of the
322 network, we focused our comparison on the chemical features that had an annotation score > 0.3, or
323 confirmed with MS/MS or analytical standards, and had a Pearson r > 0.4.

324 ### 2.9 Statistical analyses

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

329 # 3. Results

330 ### 3.1 Chemical analysis with LC-QTOF/MS

331 The recursive feature extraction and formula matching for the 295 pairs of maternal and cord
332 blood samples (n total = 590 samples) resulted in 824 features in ESI- and 731 features in ESI+ for
333 shipment 1, and 707 features in ESI- and 576 features in ESI+ for shipment 2. After combining the
334 datasets for the two shipments, the resulting dataset for ESI- summed up to 412 features and the
335 dataset for ESI+ to 298 features (n total = 710 features) after filtering out the features that showed a
336 retention time difference of > 0.5 min or a mass difference of > 15 ppm. Combining the data from ESI-
337 and ESI+, resulted in 712 features. This number is higher by 2 features compared to the total number
338 of ESI- and ESI+ because 1 isomer from ESI- had more than 1 possible matches from ESI+ based on
339 the criteria that we set for merging the two datasets (retention time difference of 0.5 min and mass
340 accuracy of 15 ppm). Ten features were identified as duplicates between ESI- and ESI+ and were
341 removed from the dataset. Seventeen features were identified as adducts and were also removed from
342 the dataset. The complete datasets before (n = 712) and after clean-up (n = 685) are presented in the
343 Supporting Information Spreadsheets. We confirmed 33 chemicals with MS/MS spectra match using
344 CFM-ID and 17 chemicals with analytical standards (Table 2 and Supporting Information
345 Spreadsheets).
346
347
348
349
350
351
352
353
354

355

356 Table 2: Chemical structures confirmed with analytical standards for 17 chemical features detected in

357 matched maternal (N=295) and cord samples (N=295). The table also shows some of the most

358 common uses [a] for the identified chemicals as well as their presence in databases from EPA's

359 Chemistry Dashboard[15] for endogenous, pharmaceuticals, pesticides, plastics, cosmetics,

360 poly/perfluorinated substances (PFAS) and high production volume chemicals.

361

| Chemical Name | Chemical Use [a] | Presence in databases | | | | | |
|---|---|---|---|---|---|---|---|
| Tridecanedioic acid | Fatty acid / polymers, lubricants, plastics | | | | | | |
| Isoquinoline | Dyes, paints, insecticides, antifungals | | | | | | |
| Eicosapentanoic acid | Omega-3 fatty acid | | | | | | |
| Caffeine | Beverages (e.g. coffee, soda), drugs | | | | | | |
| Tetraethylene glycol | Polyester resins, plasticizer, dyes | | | | | | |
| Mono(2-ethylhexyl) phthalate | Metabolite of DEHP | | | | | | |
| Phenylalanylphenylalanine | Human metabolite | | | | | | |
| Theobromine | Alkaloid in cacao / flavoring agent | | | | | | |
| Tetradecanedioic acid | Fatty acid | | | | | | |
| Progesterone | Hormone, drugs | | | | | | |
| Deoxycholic acid | Human metabolite, bile acid | | | | | | |
| Cortisone | Hormone, drugs | | | | | | |
| 1H-Indole-3-propanoic acid | Microbial metabolite of Tryptophan | | | | | | |
| 4-Nitrophenol | Air pollutant / drugs, dyes, fungicides, insecticides | | | | | | |
| Octadecanoic acid | Fatty acid / plastics, resins | | | | | | |

362

363 [a] The information on chemical uses was extracted from PubChem[25] and EPA's Chemistry Dashboard[15].

364 endogenous
pharmaceuticals
pesticides
365 plastics
cosmetics
poly/perfluoroalkyl substances
high production volume chemicals

366 ## 3.2 Database searching for feature annotation

367 We labeled 142 features as endogenous compounds and the remaining 543 features as

368 exogenous compounds. Among the chemical compounds with the highest annotation scores, we found

369 5 PFAS: perfluorohexanesulfonic acid (PFHxS), perfluorooctanesulfonic acid (PFOS),

370 perfluorodecanoic acid (PFDA), perfluoroundecanoic acid (PFUnDA) and perfluorononanoic acid

371 (PFNA); and 2 cyclic volatile methylsiloxanes: octamethylcyclotetrasiloxane (D4) and
372 decamethylcyclopentasiloxane (D5) (annotations with the individual scores in Supporting Information
373 Spreadsheets). PFHxS and PFOS were also confirmed with analytical standards.

374 ### 3.3 MS data clean-up and data processing

375 In the original dataset before batch correction, we observed two distinct clusters that
376 corresponded to the two shipments (Fig. S2 A-F). Following a matrix correlation, we observed strong
377 correlations between the first 3 PCs and the parameters corresponding to batch number, shipment, and
378 sample type (maternal vs cord) (Fig. S2 I). In addition, we observed significant differences in the PCs
379 between shipment 1 and shipment 2 (Fig. S2G), and significant differences in the PCs between
380 maternal and cord samples (Fig. S2H). Batch correction with ComBat removed the largest part of the
381 effects related to batch and shipment (Fig. 2D), while maintaining the differences between maternal and
382 cord (Fig. 2E). The updated plots after batch correction (Fig. 2) also showed that there were two main
383 clusters of samples (Fig. 2C and 2F) that corresponded to the maternal and cord sample groups (Fig.
384 2E).
385

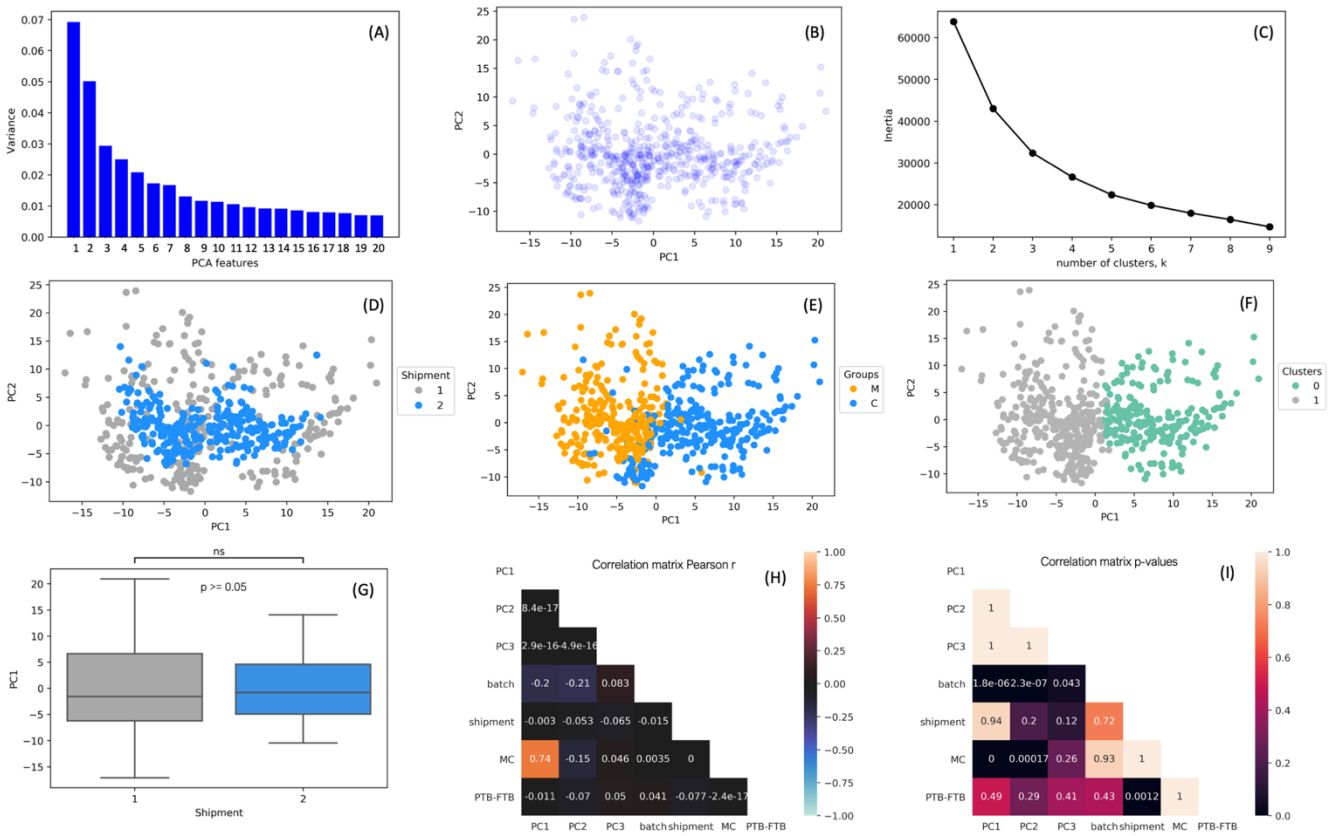



Figure 2: Results of the data analysis after batch correction with ComBat for the two shipments and the batches within each shipment. The samples were first corrected for the batches within shipment and then for the two shipments. (A): PCA features and the variance explained (%); (B) PC1 and PC2 as a 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.

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).

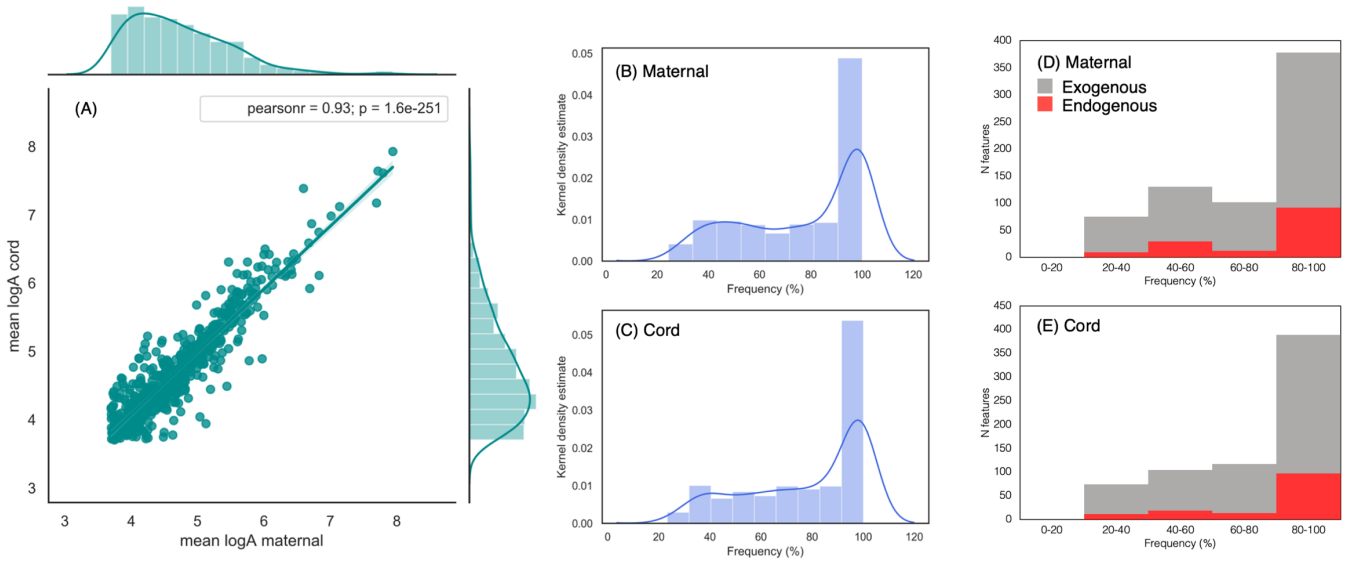



Figure 3: Correlation between maternal and cord abundances (A) (in log scale) and detection frequency calculations with kernel density curves for chemicals in maternal (B) and cord (C) blood samples (N=295 chord/maternal). The figure also displays the detection frequency for maternal (D) and cord (E)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

Out of 685 chemical features detected in MS analysis, 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).

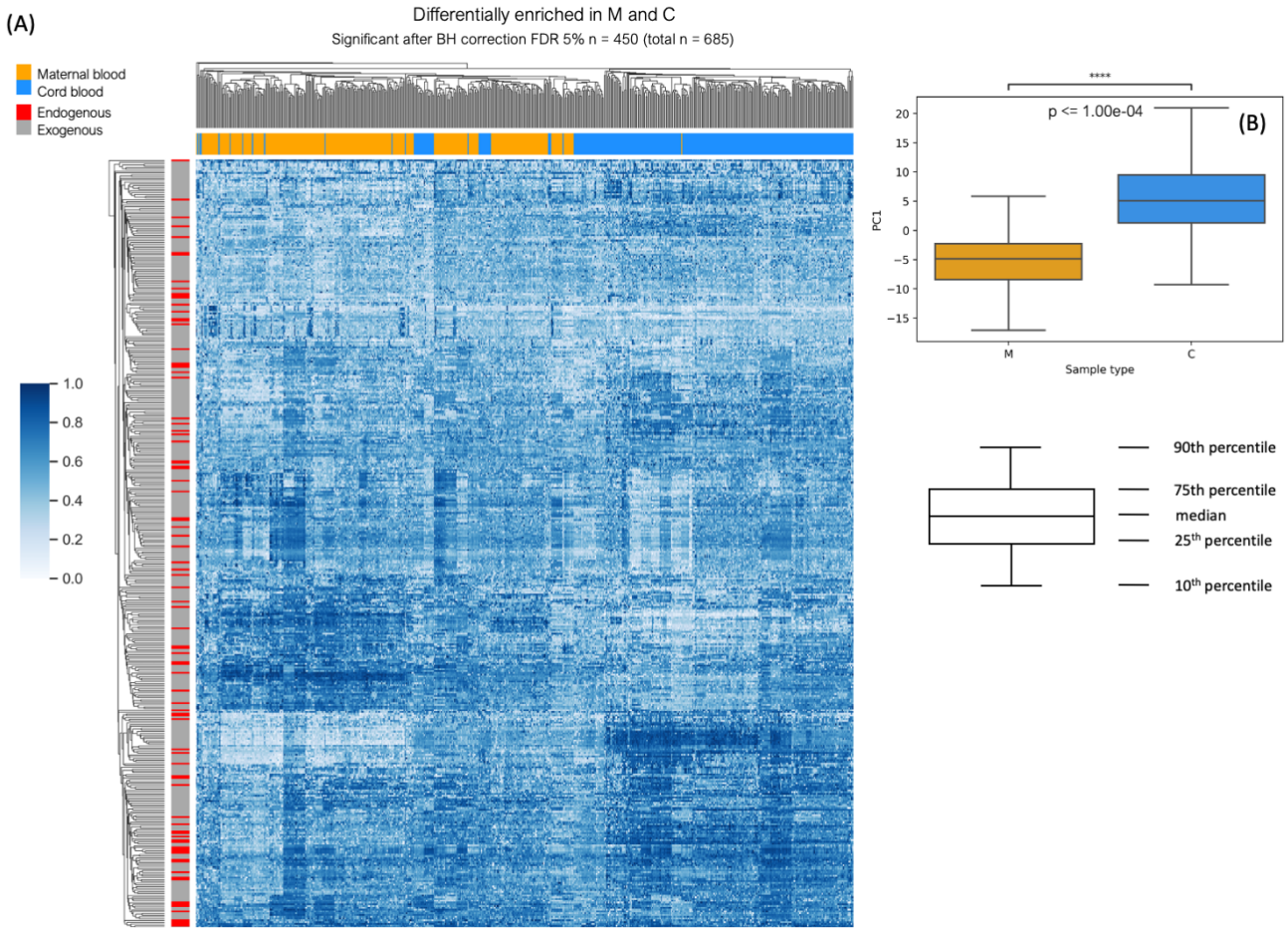



Figure 4: Clustering heatmap for maternal and cord blood samples and the chemical features that showed a significant trend to be higher in maternal or cord after multiple hypothesis correction

427 (Benjamini-Hochberg test, 5% false discovery rate). Out of 685 chemical features in total, 450 showed
428 a significant difference. The samples are color-coded by sample type (maternal vs cord). The features
429 are color-coded by chemical type (endogenous vs exogenous). The error bars in the box-plot show the
430 $10^{th}$ and $90^{th}$ percentiles, the boxes show the $25^{th}$ and $75^{th}$ percentiles and the middle line shows the
431 median.
432
433

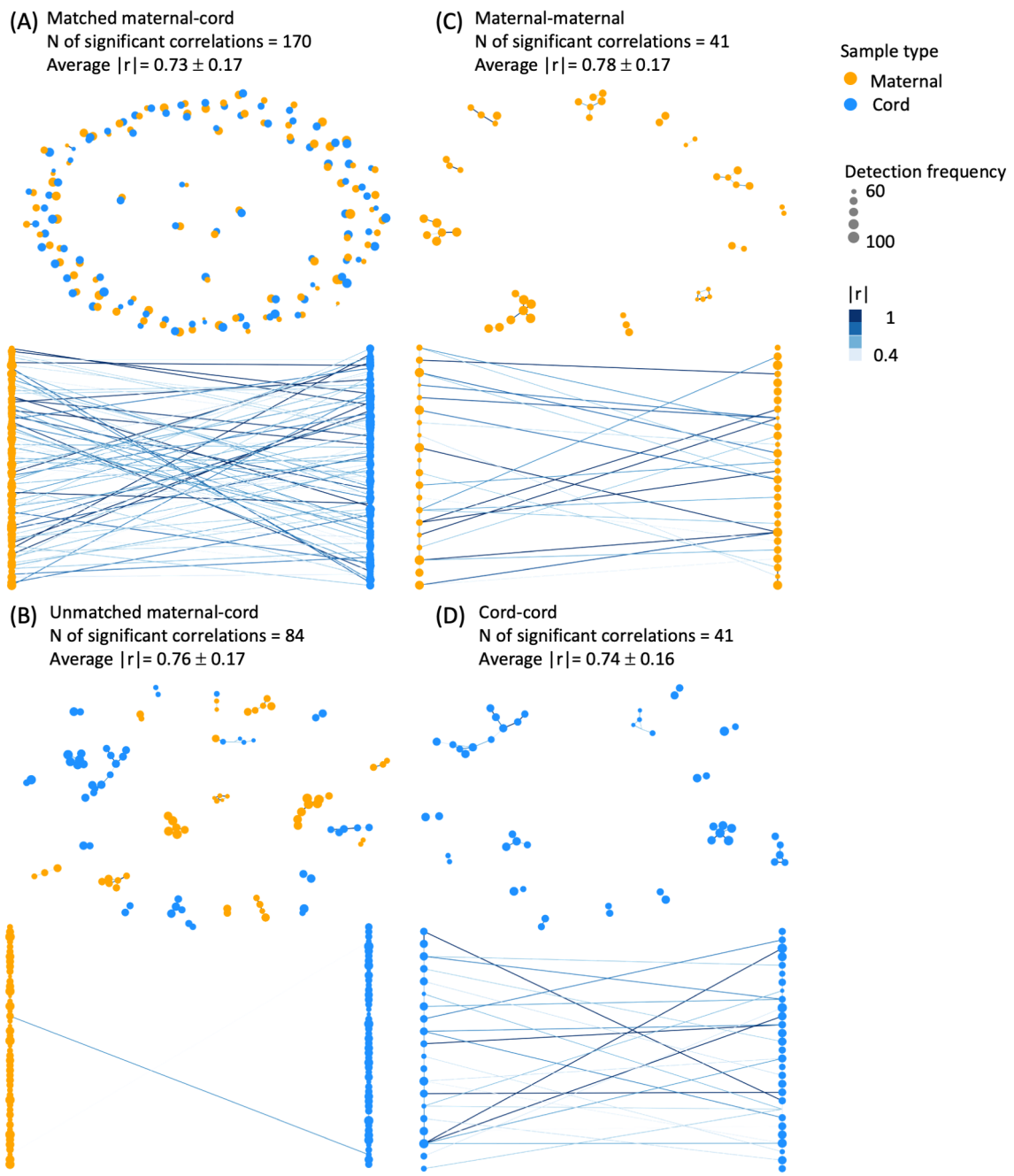



Figure 5: Similarity network analysis for matched maternal and cord samples (N = 590). Correlations for 85 maternal-cord pairs that remained significant after multiple hypothesis correction (Benjamini-Hochberg, 5% false discovery rate) as correlation networks in random positions and in bipartite graphs. (A): Showing correlations only between paired maternal and cord; (B): Showing remaining correlations

between maternal-maternal, cord-cord and unpaired maternal-cord; (C) Showing correlations between maternal samples only; (D) showing correlations between cord samples only.

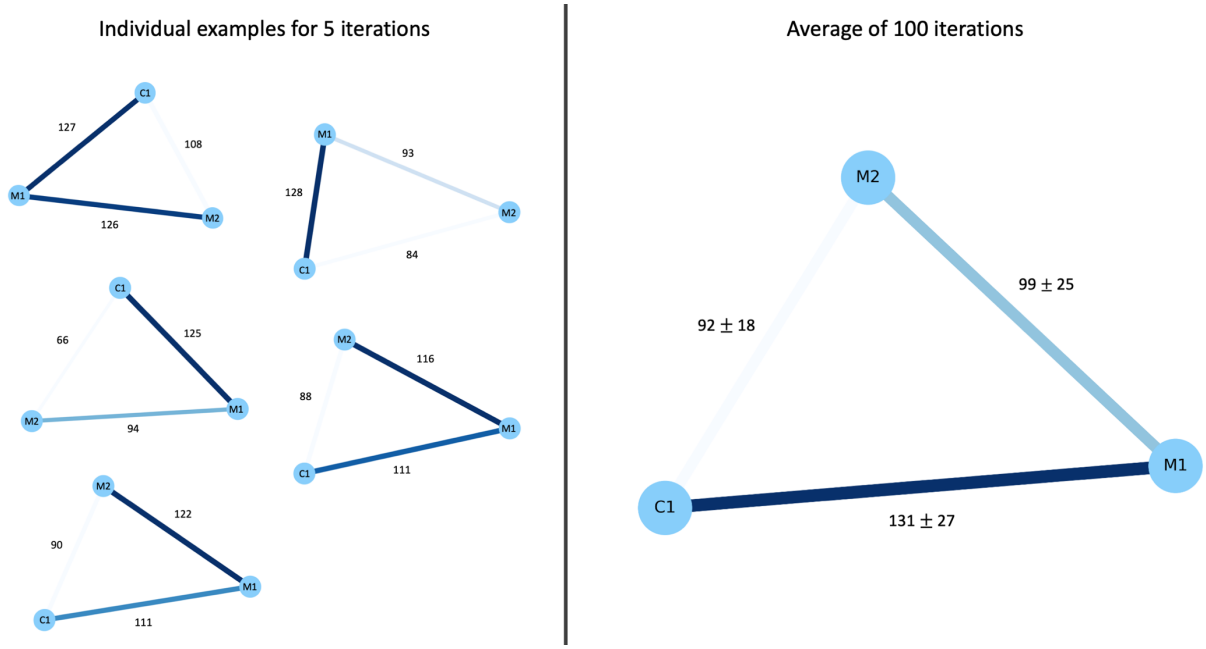
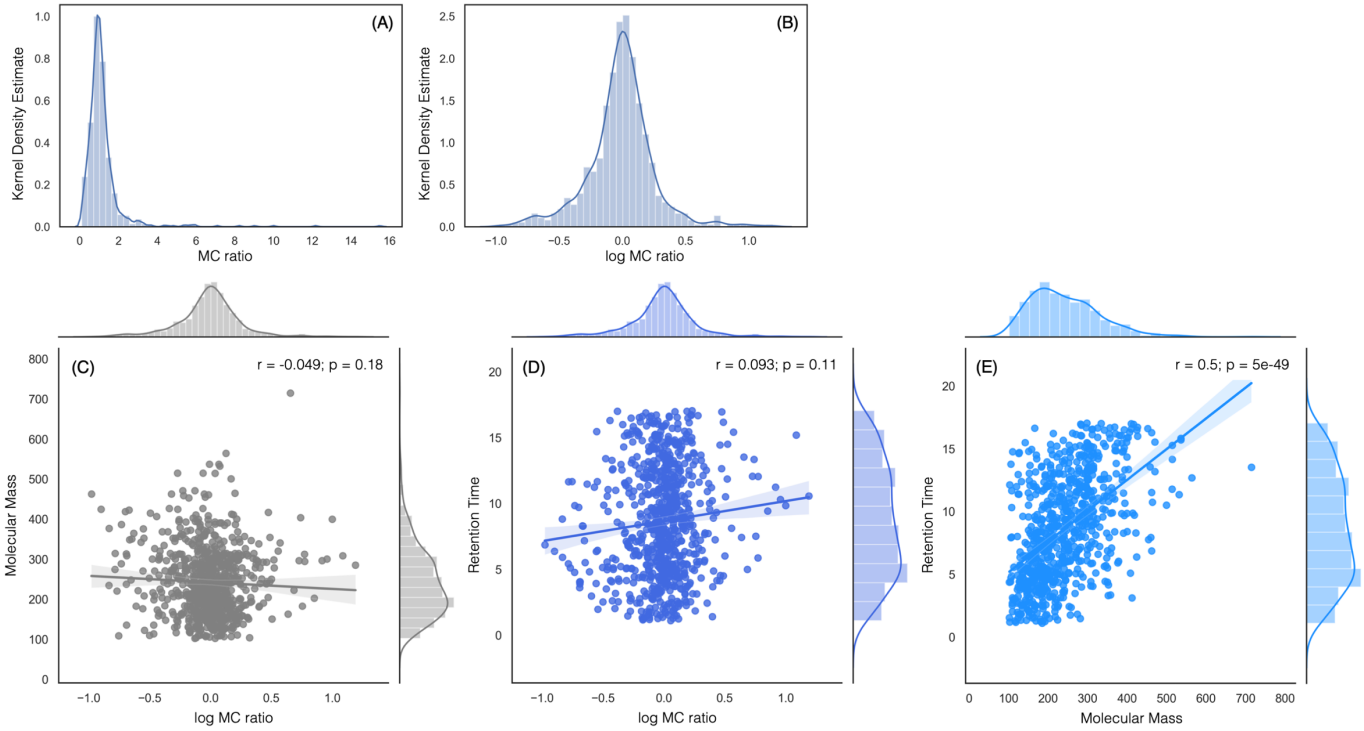



Figure 6: Similarity network analysis using a permutation approach randomly selecting maternal and cord samples to compare the similarity between paired maternal and cord samples (M1-C1) compared to maternal – maternal (M1-M2) and unpaired maternal and cord (M2-C1). The numbers on the left side of the figure show the number of chemicals, for which the ratio of their abundance in the various pairs (M1-C1, M1-M2 and M2-C1) ranged from 0.75 to 1.25, with ratio = 1 indicating complete agreement. The numbers on the right show the average of these number after 100 iterations and their standard deviations.

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. 5A) compared to unpaired maternal and cord samples (N = 84; Fig. 5B) and compared to maternal only (N=41; Fig. 5C) and cord only (N=41; Fig. 5D). No significant differences were observed in the average | r | values between the

455 four groups. Our similarity network analysis using a permutation approach showed a very similar trend
456 (Fig. 6). The average of 100 iterations showed that paired maternal and cord samples (M1-C1) shared
457 more similar chemical features compared to maternal – maternal pairs (M1-M2) and unmatched
458 maternal – cord samples (M2-C1) (Fig. 6).
459
460

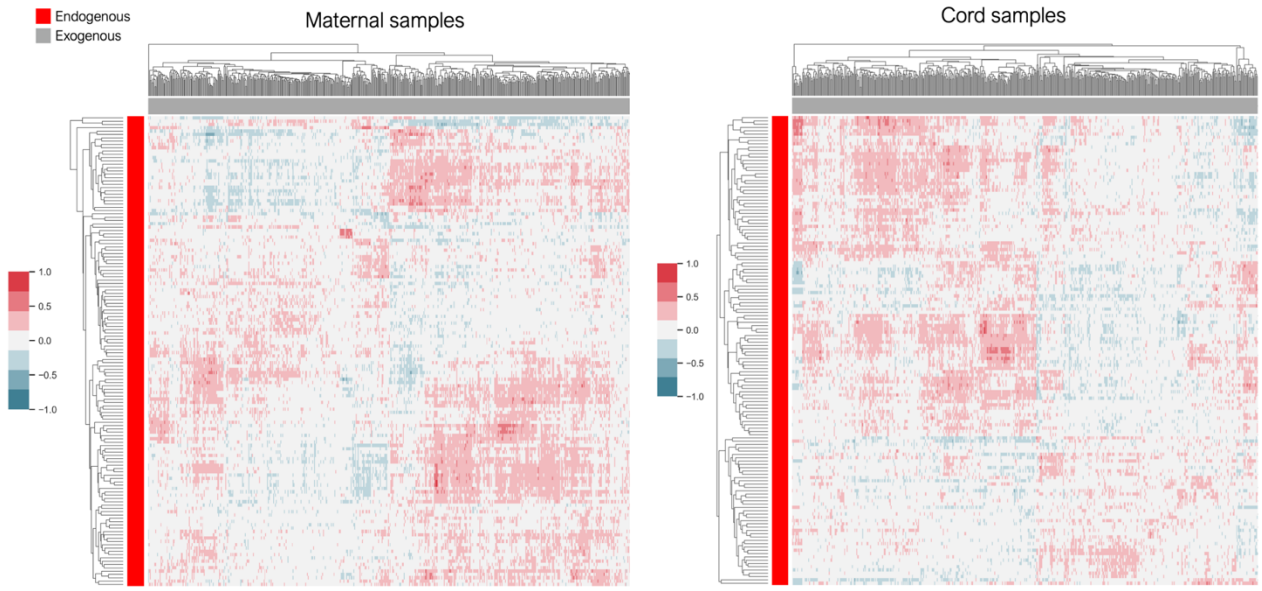



461
462 Figure 7: Maternal /cord abundance ratios (log MC ratio) for the chemical features detected in the
463 maternal blood (N=295) and in the cord blood (N=295) samples in linear (A) and logarithmic scale (B),
464 and its relationship to molecular mass (C) and retention time (D). (E): Retention time and its
465 relationship to molecular mass.
466
467 We observed that the majority of the maternal/cord abundance ratios are concentrated around 1
468 indicating an even partitioning between maternal and cord blood (Fig. 7A and 7B). The maternal/cord

469 abundance ratios showed a weak but significant positive correlation with retention time (7D). No
470 significant correlation was found for maternal/cord abundance ratio and molecular mass (7C).
471

472 ### 3.4.2 Correlations between endogenous and exogenous compounds

473 We observed 21,522 significant relationships between endogenous and exogenous features in
474 maternal samples and 19,846 in cord samples after multiple hypothesis correction (n total relationships
475 = 77,106 in maternal and n = 77,106 in cord samples, Figure 8). From the significant relationships, 103
476 relationships in maternal and 128 relationships in cord samples had an absolute Pearson r > 0.5, 5
477 relationships in maternal and 4 relationships in cord samples had an absolute Pearson r > 0.7 and 1
478 relationship in maternal and 1 relationship in cord samples had an absolute Pearson r > 0.8 (dataset
479 with the calculated r and p-values in the Supporting Information Spreadsheets).




480

481 Figure 8: Matrix correlation for endogenous (metabolites) and exogenous (industrial chemicals) in
482 maternal and cord blood samples separately (N maternal = 295 and N cord = 295)

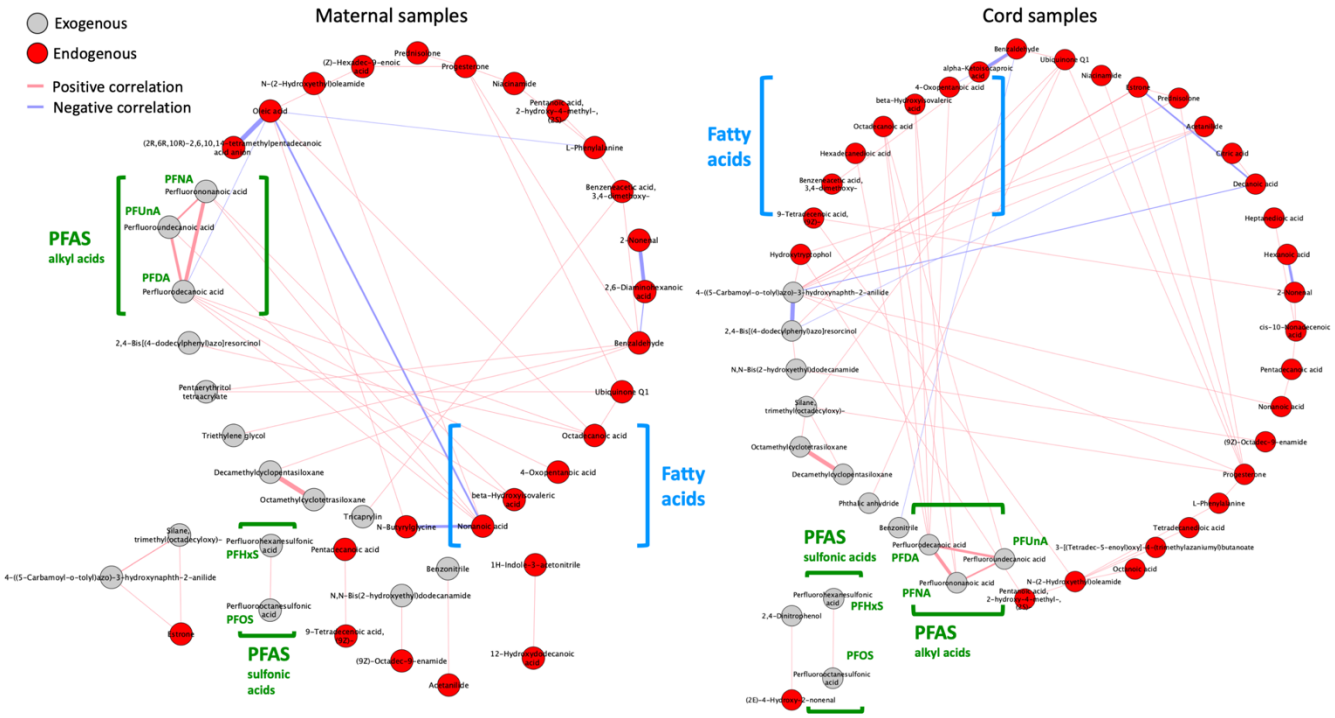



Figure 9: Molecular interaction networks for endogenous (red) and exogenous (gray) chemical features in the maternal blood (N = 295) and cord blood samples (N = 295). The network shows the features which had an annotation score of > 0.3 or were identified with MS/MS or with analytical standards. The 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 (0.4 – 1).

The maternal and cord networks (Fig. 9) showed great overlap with most chemical compounds appearing in both networks and exhibiting similar relationships. In both the maternal and cord, two cyclic volatile methylsiloxanes (cVMS) (octamethylcyclotetrasiloxane; D4 and decamethylcyclotetrasiloxane; D5) correlated strongly with each other (r = 0.77 in maternal network and r=0.81 in cord network) and both were part of the main network. In addition, three perfluoroalkyl acids PFAAs: perfluorononanoic acid (PFNA), perfluorodecanoic acid (PFDA) and perfluoroundecanoic acid

497 (PFUnDA) correlated strongly with each other (r values in maternal: 0.66-0.74, r values in cord: 0.64-
498 0.72) while 2 perfluorinated sulfonic acids (PFSA; perfluorohexanesulfonic acid, perfluorooctanesulfonic
499 acid) formed their own group outside the main networks. Both groups of chemicals are
500 poly/perfluoroalkyl substances (PFAS), a group of chemicals that has recently come under scrutiny due
501 to their persistence, bioaccumulation potential and toxicity. The group of PFAA, in both networks,
502 showed to correlate with certain fatty acids, such as octadecanoic acid.

## 503 4. Discussion

504 Our chemical analysis of the maternal and blood samples with HRMS and a non-target analysis
505 workflow provided unique insights in the prenatal exposome, exposures to environmental pollutants,
506 and their role in the development of human disease. To our knowledge, this is the largest dataset of the
507 exposome of maternal and fetal exposures. We identified 17 chemical structures with analytical
508 standards with mixed endogenous and exogenous sources (Table 2).

509 Our data analysis showed that when analyzing large sample sets with non-targeted analysis
510 batch effects are substantial and they need to be adequately addressed before drawing any
511 conclusions on the chemical, biological, and epidemiological importance of that collected data.
512 ComBat[20,21] was able to remove batch effects for HRMS data for exposomics and metabolomics
513 analyses.

514 Maternal and cord samples showed some similarities in chemical feature enrichment (Fig. 3),
515 but also important differences (Fig. 4) that allowed for these two groups to be classified as two distinct
516 clusters (Fig. 4). Our similarity network analyses also showed that matched maternal and cord samples
517 are more similar in terms of chemical feature enrichment compared to other maternal samples. These
518 observations have important implications when studying the partitioning of chemical compounds
519 between maternal and cord samples and when studying which chemicals show a stronger potential to
520 cross the placenta and accumulate in the fetus. Previous studies have reported on the partitioning

521 between maternal and cord blood,[26–29] however, the mechanism by which certain chemicals cross the
522 placenta more readily than others requires further investigation. One interesting example of chemicals
523 from our dataset that showed preferential partitioning for the maternal side were the five PFAS we
524 detected. The log $MC_{ratio}$ of the five PFAS ranged from 0.037 to 0.22 (Supporting Information
525 Spreadsheets and Fig 7B; right tale of the distribution) indicating that the transfer of these chemicals to
526 the fetus is inhibited by the placenta. This finding is in good agreement with previous biomonitoring
527 studies where they examined the transplacental transfer of PFAS.[30,31] Due to their strong affinity for
528 proteins, PFAS, bind to the proteins in the placenta and are inhibited from reaching the fetus.[30,31] On
529 the other hand, a compound that showed preferential partitioning for the fetal side was Triamcinolone,
530 which had a log $MC_{ratio}$ of -0.26. Triamcinolone is a pharmaceutical glucocorticoid used in human and
531 veterinary applications as an anti-inflammatory drug.[15,25] Triamcinolone is a highly water-soluble
532 substance with no particular affinity for lipids or proteins (equilibrium partition ratio between octanol and
533 water; log $K_{OW}$ = 0.967). These properties make it easily transferable across the placenta and
534 preferentially partition to cord blood due to its lower lipid content compared to maternal blood.[32–34]
535 We observed a weak but significant positive association between maternal/cord abundance
536 ratio and retention time (Fig. 7D). As retention time is a function of the chemicals' hydrophobicity, with
537 more hydrophobic chemicals exhibiting longer retention times, its relationship with the maternal/cord
538 ratio would indicate that more hydrophobic chemicals would show a preference to partition more to the
539 maternal blood compared to cord blood. This observation is in agreement with previous studies
540 showing a positive correlation between the maternal/cord ratio and $K_{OW}$.[35] This finding suggests that
541 retention time could be used as a criterion for prioritizing chemical features for identification in
542 maternal/cord blood studies and could potentially also be used in prioritization of chemicals for toxicity
543 testing. With regards to the endogenous compounds, the partitioning between maternal and cord blood
544 is more complicated. Many of them could be originating from the maternal side, the fetal side or both. In
545 order to draw a conclusion on the partitioning behavior of the endogenous compounds, we would need

546 to know the production rates of these compounds on each side and adjust the calculated partition
547 ratios. This is certainly an aspect that warrants further investigation.
548 Our analysis of the interactions between exogenous and endogenous exposure revealed
549 important insights into how environmental chemicals disrupt biological pathways. We observed
550 thousands of significant relationships between exogenous and endogenous chemical features,
551 hundreds of which showed an absolute $r > 0.5$. One group of chemicals that showed an interesting
552 pattern were two cyclic volatile methylsiloxanes (cVMS), octamethylcyclotetrasiloxane (D4) and
553 decamethylcyclopentasiloxane (D5). cVMS are organosilicon chemicals that are primarily used as
554 carriers in personal care products, such as deodorants, and as intermediates in the production of
555 silicone polymers. Their strong correlation indicates a common source of exposure, most likely due to
556 use of personal care products. Their ubiquitous presence in personal care products makes it very likely
557 that these chemicals are from such applications. However, also because of their ubiquitous presence in
558 silicone polymers, there is a chance that these chemicals could be a result of contamination from inside
559 the analytical instrument. There is also a possibility that these chemicals could be also coming from
560 personal care products by people working in the lab, however, the physicochemical properties of D4
561 and D5, specifically their equilibrium partition ratio between octanol and air ($K_{OA}$), indicates that
562 partitioning from the air to an organic solvent is very unlikely. D4 has a log $K_{OA}$ of 4.97 and D5 has log
563 $K_{OA}$ of 3.94, which indicate a strong preference for the molecules to exist in the gas phase compared to
564 other chemicals, such as polychlorobiphenyl 180 (PCB 180) which has a log $K_{OA}$ of 9.94 and a much
565 stronger preference to partition to octanol. Finally, all the abundances in our data set were blank
566 corrected which should minimize the potential of contamination.
567 Another group of exogenous chemicals that showed an interesting pattern were three PFAS
568 (PFNA, PFDA and PFUnA) that correlated strongly with endogenous fatty acids. PFAS have been
569 shown to interfere with fatty acid metabolism in *in vitro* studies by binding to fatty acid binding
570 proteins.[36,37] Binding of PFAS to fatty acid binding proteins could reduce the available binding sites for

571 endogenous fatty acids resulting in higher concentrations of fatty acids. This could explain the observed
572 correlations between the three PFAS and endogenous fatty acids. Currently there are about 5,000
573 PFAS registered on EPA's Chemistry Dashboard, many of which do not have data on their toxicity
574 potential in humans. Our study corroborates the need for further experimental and modeling studies to
575 assess the potential of the ever-increasing chemical library of PFAS and study how they interfere with
576 human metabolism. High-throughput protein binding studies would help to elucidate some of these
577 effects and help prioritize PFAS for biomonitoring and regulatory action.

578 ### 4.1 Limitations and other considerations

579 Our study illustrates the importance of broad screening using NTA in order to characterize the
580 exposome and the mechanisms under which environmental exposures contribute to the development of
581 human disease. As these techniques are powerful in detecting thousands of chemical features there
582 are still some challenges remaining to be addressed. One of the main shortcomings of current NTA
583 approaches is that the number of identified chemicals is very small compared to the number of detected
584 features with only 1-5% of chemicals often being confirmed with analytical standards.[11,12,38] Thus, there
585 is a need to develop novel computational tools for structure elucidation or structure annotation without
586 analytical standards that can help us circumvent that problem and leverage the full potential of NTA.

587 Another limitation of our study is that it uses only one analytical instrument, LC-QTOF/MS,
588 which specializes in the analysis and identification of polar and involatile compounds. As a result, the
589 chemical features that we detected are primarily from that physicochemical space. Complementing LC-
590 QTOF/MS with Gas Chromatography-QTOF/MS, which specializes in non-polar and volatile/semi-
591 volatile chemicals could help expand the spectrum of possible chemical features.

592 Finally, our study focuses on the differences between maternal and cord blood as a surrogate
593 for understanding fetal exposure and adverse fetal health outcomes. However, adverse fetal health
594 outcomes depend not only on the amount of the chemical the fetus is exposed to, but also on the

595 toxicity of the chemical. There is thus a need to develop high-throughput toxicity screening models to
596 screen for chemicals found in fetal blood. Using NTA data to inform toxicity testing can provide unique
597 insights in toxicology and environmental health and assist in preventing of exposure to toxic chemicals.

598 ### 4.2 Future directions

599 Non-targeted analysis can play an important role in deep phenotyping for precision medicine
600 and advanced patient care.[39] Precision medicine aims to provide the best possible patient care by
601 categorizing and subcategorizing patients with a certain disease using computational methods that
602 combine information from genomics, proteomics, metabolomics, and additional clinical data.[39] Deep
603 phenotyping is crucial in understanding the underlying mechanisms of adverse health outcomes in and
604 in developing strategies for prevention and treatment.[40] Finally, deep phenotyping can provide
605 important insights on the role of environmental exposures in the development of adverse health
606 outcomes during pregnancy. In that endeavor, we will further our studies by utilizing our high-
607 dimensional datasets to agnostically investigate the role of endogenous and exogenous exposures to
608 the development of adverse health outcomes, such as gestational diabetes, preterm birth, birth weight,
609 and preeclampsia, among others.

610 ## Data availability

611 All the datasets used are provided as supporting information. All the code is available on GitHub
612 (https://github.com/dimitriabrahamsson/nontarget-maternalcord.git )

613 ## Acknowledgements

614 This study was funded by NIH/NIEHS grant numbers UG3OD023272, UH3OD023272, P01ES022841,
615 R01ES027051 and by the US EPA grant number RD83543301.

## 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) 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.

(4) 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.

(5) 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.

(6) Woodruff; Zota AR.; Schwartz JM. Environmental Chemicals in Pregnant Women in the United States: NHANES 2003–2004. *Environmental Health Perspectives* **2011**, *119* (6), 878–885. https://doi.org/10.1289/ehp.1002727.

(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) Wang, A.; Abrahamsson, D.; Jiang, T.; Wang, M.; Morello-Frosch, R.; Park, J.-S.; Sirota, M.; Tracey J., W. Suspect Screening, Prioritization and Confirmation of Environmental Chemicals in Maternal-Newborn Pairs from San Francisco. *Environ Sci Technol In Press*.

(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) Houtz, E. F.; Sutton, R.; Park, J.-S.; Sedlak, M. Poly- and Perfluoroalkyl Substances in Wastewater: Significance of Unknown Precursors, Manufacturing Shifts, and Likely AFFF Impacts. *Water Research* **2016**, *95*, 142–149. https://doi.org/10.1016/j.watres.2016.02.055.

(14) Houtz, E.; Wang, M.; Park, J.-S. Identification and Fate of Aqueous Film Forming Foam Derived Per- and Polyfluoroalkyl Substances in a Wastewater Treatment Plant. *Environ. Sci. Technol.* **2018**, *52* (22), 13212–13221. https://doi.org/10.1021/acs.est.8b04028.

(15) U.S. Environmental Protection Agency. Chemistry Dashboard https://comptox.epa.gov/dashboard/.

(16) 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.

(17) WikiPathways - WikiPathways https://www.wikipathways.org/index.php/WikiPathways (accessed Sep 23, 2020).

(18) Wikipedia https://www.wikipedia.org/ (accessed Sep 23, 2020).

(19) 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.

(20) 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).

(21) 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).

(22) Software for Complex Networks — NetworkX 2.5 documentation https://networkx.github.io/documentation/stable/index.html (accessed Sep 11, 2020).

(23) Cytoscape: An Open Source Platform for Complex Network Analysis and Visualization https://cytoscape.org/ (accessed Sep 24, 2020).

(24) MetScape 3 http://metscape.ncibi.org/ (accessed Sep 24, 2020).

(25) PubChem. PubChem https://pubchem.ncbi.nlm.nih.gov/ (accessed Aug 25, 2020).

(26) 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.

(27) 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

689 the Human Maternal–Fetal–Placental Unit. *Environ. Sci. Technol.* **2020**, *54* (6), 3476–3486.
690 https://doi.org/10.1021/acs.est.0c00206.
691 (28) Chen, F.; Yin, S.; Kelly, B. C.; Liu, W. Isomer-Specific Transplacental Transfer of Perfluoroalkyl
692 Acids: Results from a Survey of Paired Maternal, Cord Sera, and Placentas. *Environ. Sci.*
693 *Technol.* **2017**, *51* (10), 5756–5763. https://doi.org/10.1021/acs.est.7b00268.
694 (29) Chen, A.; Park, J.-S.; Linderholm, L.; Rhee, A.; Petreas, M.; DeFranco, E. A.; Dietrich, K. N.; Ho,
695 S. Hydroxylated Polybrominated Diphenyl Ethers in Paired Maternal and Cord Sera. *Environ. Sci.*
696 *Technol.* **2013**, *47* (8), 3902–3908. https://doi.org/10.1021/es3046839.
697 (30) Mamsen, L. S.; Björvang, R. D.; Mucs, D.; Vinnars, M.-T.; Papadogiannakis, N.; Lindh, C. H.;
698 Andersen, C. Y.; Damdimopoulou, P. Concentrations of Perfluoroalkyl Substances (PFASs) in
699 Human Embryonic and Fetal Organs from First, Second, and Third Trimester Pregnancies.
700 *Environment International* **2019**, *124*, 482–492. https://doi.org/10.1016/j.envint.2019.01.010.
701 (31) Mamsen, L. S.; Jönsson, B. A. G.; Lindh, C. H.; Olesen, R. H.; Larsen, A.; Ernst, E.; Kelsey, T. W.;
702 Andersen, C. Y. Concentration of Perfluorinated Compounds and Cotinine in Human Foetal
703 Organs, Placenta, and Maternal Plasma. *Science of The Total Environment* **2017**, *596–597*, 97–
704 105. https://doi.org/10.1016/j.scitotenv.2017.04.058.
705 (32) Needham, L. L.; Grandjean, P.; Heinzow, B.; Jørgensen, P. J.; Nielsen, F.; Patterson, D. G.;
706 Sjödin, A.; Turner, W. E.; Weihe, P. Partition of Environmental Chemicals between Maternal and
707 Fetal Blood and Tissues. *Environmental science & technology* **2011**, *45* (3), 1121–1126.
708 https://doi.org/10.1021/es1019614.
709 (33) Nakamura, T.; Nakai, K.; Matsumura, T.; Suzuki, S.; Saito, Y.; Satoh, H. Determination of Dioxins
710 and Polychlorinated Biphenyls in Breast Milk, Maternal Blood and Cord Blood from Residents of
711 Tohoku, Japan. *Science of The Total Environment* **2008**, *394* (1), 39–51.
712 https://doi.org/10.1016/j.scitotenv.2008.01.012.

713 (34) Park, J.-S.; Bergman, Å.; Linderholm, L.; Athanasiadou, M.; Kocan, A.; Petrik, J.; Drobna, B.;
714 Trnovec, T.; Charles, M. J.; Hertz-Picciotto, I. Placental Transfer of Polychlorinated Biphenyls,
715 Their Hydroxylated Metabolites and Pentachlorophenol in Pregnant Women from Eastern
716 Slovakia. *Chemosphere* **2008**, *70* (9), 1676–1684.
717 https://doi.org/10.1016/j.chemosphere.2007.07.049.
718 (35) Lancz, K.; Murínová, Ľ.; Patayová, H.; Drobná, B.; Wimmerová, S.; Šovčíková, E.; Kováč, J.;
719 Farkašová, D.; Hertz-Picciotto, I.; Jusko, T. A.; Trnovec, T. Ratio of Cord to Maternal Serum PCB
720 Concentrations in Relation to Their Congener-Specific Physicochemical Properties. *International*
721 *Journal of Hygiene and Environmental Health* **2015**, *218* (1), 91–98.
722 https://doi.org/10.1016/j.ijheh.2014.08.003.
723 (36) Sheng, N.; Cui, R.; Wang, J.; Guo, Y.; Wang, J.; Dai, J. Cytotoxicity of Novel Fluorinated
724 Alternatives to Long-Chain Perfluoroalkyl Substances to Human Liver Cell Line and Their Binding
725 Capacity to Human Liver Fatty Acid Binding Protein. *Arch Toxicol* **2018**, *92* (1), 359–369.
726 https://doi.org/10.1007/s00204-017-2055-1.
727 (37) Zhang, L.; Ren, X.-M.; Guo, L.-H. Structure-Based Investigation on the Interaction of
728 Perfluorinated Compounds with Human Liver Fatty Acid Binding Protein. *Environ. Sci. Technol.*
729 **2013**, *47* (19), 11293–11301. https://doi.org/10.1021/es4026722.
730 (38) Phillips, K. A.; Yau, A.; Favela, K. A.; Isaacs, K. K.; McEachran, A.; Grulke, C.; Richard, A. M.;
731 Williams, A. J.; Sobus, J. R.; Thomas, R. S.; Wambaugh, J. F. Suspect Screening Analysis of
732 Chemicals in Consumer Products. *Environ. Sci. Technol.* **2018**, *52* (5), 3125–3135.
733 https://doi.org/10.1021/acs.est.7b04781.
734 (39) Robinson, P. N. Deep Phenotyping for Precision Medicine. *Human Mutation* **2012**, *33* (5), 777–
735 780. https://doi.org/10.1002/humu.22080.

736 (40) Paquette, A. G.; Hood, L.; Price, N. D.; Sadovsky, Y. Deep Phenotyping during Pregnancy for
737 Predictive and Preventive Medicine. *Science Translational Medicine* **2020**, *12* (527).
738 https://doi.org/10.1126/scitranslmed.aay1059.
739