Environmental Science & Technology

pubs.acs.org/est

Article

Suspect Screening, Prioritization, and Confirmation of Environmental Chemicals in Maternal-Newborn Pairs from San Francisco

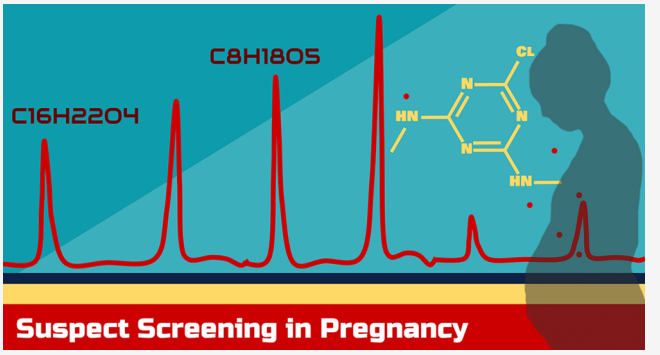
Aolin Wang,[+] Dimitri Panagopoulos Abrahamsson,[+] Ting Jiang, Miaomiao Wang, Rachel Morello-Frosch, June-Soo Park, Marina Sirota, and Tracey J. Woodruff*

Cite This: https://dx.doi.org/10.1021/acs.est.0c05984

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Our proof-of-concept study develops a suspect screening workflow to identify and prioritize potentially ubiquitous chemical exposures in matched maternal/cord blood samples, a critical period of development for future health risks. We applied liquid chromatography−quadrupole time-of-flight tandem mass spectrometry (LC-QTOF/MS) to perform suspect screening for ∼3500 industrial chemicals on pilot data from 30 paired maternal and cord serum samples (n = 60). We matched 662 suspects in positive ionization mode and 788 in negative ionization mode (557 unique formulas overall), and selected 208 of these for fragmentation analysis based on detection frequency, correlation in feature intensity between maternal and cord samples, and peak area differences by demographic characteristics. We tentatively identified 73 suspects through fragmentation spectra matching and confirmed 17 chemical features (15 unique compounds) using reference standards. We tentatively identified 55 compounds not previously reported in the literature, the majority which have limited to no information about their sources or uses. Examples include (i) 1-(1-acetyl-2,2,6,6-tetramethylpiperidin-4-yl)-3-dodecylpyrrolidine-2,5-dione (known high production volume chemical) (ii) methyl perfluoroundecanoate and 2-perfluorooctyl ethanoic acid (two PFAS compounds); and (iii) Sumilizer GA 80 (plasticizer). Thus, our workflow demonstrates an approach to evaluating the chemical exposome to identify and prioritize chemical exposures during a critical period of development.



KEYWORDS: *suspect screening, exposome, high-throughput, maternal blood, cord blood, pregnancy, biomonitoring*

INTRODUCTION

Prenatal exposure to environmental chemicals can lead to myriad health consequences throughout life.[1−4] Prior research using National Health and Nutrition Examination Survey (NHANES) data found that pregnant women in the U.S. are exposed to multiple different chemicals.[5,6] Most of these chemicals can cross the placenta into the fetal environment,[7,8] with sometimes higher exposure to the fetus compared to maternal blood measurements, such as mercury and polychlorinated biphenyls.[4] In a study of 65 pregnant women in San Francisco, we detected a median of ∼25 chemicals in maternal serum (out of 59 compounds tested), of which ∼80% were also detected in matched umbilical cord serum samples, with some compounds having higher concentrations than maternal levels.[9] Existing biomonitoring research mainly relies on targeted analytical methods that cover only a few hundred chemicals.[6] This is likely a small fraction of all the potential chemicals that humans are exposed to, as ∼8000 chemicals are manufactured or imported in large volume (>25 000 lbs/year) in the U.S.,[10] and chemical production totals at least 9.5 trillion pounds,[10,11] let alone the approximately 40 000 chemicals currently in commerce in the U.S.[12] A recent study reviewed over 700 chemicals from multiple chemical classes that have a high likelihood of exposure among mothers and children, have a potentially toxic structural moiety, but are not currently measured via biomonitoring or health effects in National Institutes of Health (NIH)'s Environmental influences on Child Health Outcomes (ECHO) initiative or NHANES.[13] The authors recommended 155 chemicals of high priority for future biomonitoring, suggesting an unmet need for characterizing exposures to these "known unknown" chemicals.

Recent advancements in high-resolution mass spectrometry (HRMS) paired with novel computational and statistical

Received: September 7, 2020
Revised: February 22, 2021
Accepted: February 22, 2021

approaches provide an opportunity for large-scale screening of chemicals from biological and environmental samples.[14−17] By leveraging the nontargeted chemical feature acquisition, suspect screening can efficiently identify chemicals of interest using HRMS and software-matching algorithms that map these features against user-defined chemical databases or existing chemical inventories.[18] This technology has gained increasing popularity in recent years as a new tool for environmental monitoring,[19−22] metabolite discovery,[23,24] and biomonitoring of industrial chemicals[25,26] to better characterize the chemisome,[27] the industrial chemical components of the human exposome.[28]

While there are numerous publications on HRMS in environmental monitoring and metabolite discovery, the application of this technique to biomonitoring of industrial chemicals remains limited. In a previous study, we leveraged this technology to identify novel chemicals never measured before in the blood of pregnant women,[26] and found that they are exposed to more chemicals than previously documented. As the first proof-of-concept study in applying suspect screening to detect chemicals in pregnant women's serum, we limited our search to a subset of environmental chemicals called environmental organic acids (EOAs), compounds with at least one ionizable proton, by using the negative ionization mode to optimize their detection.[26]

This paper builds upon our previous work[26] to demonstrate the application of a suspect screening method for characterizing exposure to a broader array of industrial chemicals in matched maternal and cord serum samples, a critical developmental period of health risk. We have developed and tested an analytical approach that uses HRMS to screen for multiple chemicals and a workflow to prioritize and identify ubiquitous endogenous chemicals that are differentially enriched in maternal/cord samples and/or across various demographic groups. Applying our approach to data from 30 paired maternal and cord serum samples (N = 60 total samples), we expand work in the field of suspect screening and nontargeted analysis of human blood samples in four ways: (1) using a chemical database of approximately 3500 high-production volume chemicals as well as chemicals of emerging concern including an expanded list of short-chain per- and polyfluoroalkyl substances;[29] (2) using both positive and negative ionization modes to facilitate detection of more chemical features; (3) evaluating cord serum matched to maternal serum allowing evaluation of differential enrichment of chemicals between the two; and (4) confirming chemical structures via matching of experimental MS/MS spectra against MS/MS spectra from existing reference libraries and analytical standards. Furthermore, to the best of our knowledge, this is the first study to characterize the chemical exposome to industrial chemicals in matched maternal and cord blood sample pairs using a suspect screening or a nontargeted analysis approach.

MATERIALS AND METHODS

Study Population and Sample Collection. The study population is part of the Chemicals in Our Bodies 2 Study (CiOB2), which consists of women seeking prenatal and delivery care at the Zuckerberg San Francisco General Hospital and UCSF Mission Bay Medical Center.[26] From April 2, 2014, we enrolled women from an economically and ethnically diverse population (47% Latina, 37% non-Hispanic whites, and 17% non-Hispanic Asians, Pacific Islanders, African Americans) who were English or Spanish-speaking, aged 18 through 40 years old, and had singleton pregnancies between 13 and 27 weeks gestation (second trimester) at the time of recruitment. Paired maternal and cord blood samples were collected at delivery for chemical analysis from participants who agreed to have their samples banked and included in supplemental studies. Maternal blood was collected during labor and delivery and umbilical cord blood after delivery and prior to umbilical cord clamping whenever possible. Blood was collected in BD Vacutainer Plus Serum Tubes and stored at −80 °C until analysis. We collected demographic information via interviewer-administered questionnaire and obtained information from maternal and infant medical with permission from participants. In this proof of concept study, we analyzed paired maternal-cord serum samples from 30 women. CiOB2 study protocols were approved by the Institutional Review Boards of the University of California, San Francisco (13-12160).

Chemical Analysis: Suspect Screening. *Chemical Suspect Database.* For our maternal/cord paired serum suspect screening study, we developed a ∼3500 chemical suspect database that combined data from an in-house Environmental Organic Acid (EOA) database we used in our earlier study[26] with additional high-production chemicals in the U.S. as described below (Supporting Information (SI) Figure S1).

1. In-House Industrial Chemical Database. Our in-house chemical database (SI Figure S1) consists of 714 chemical entries, including 369 chemicals from our previous published Environmental Organic Acid (EOA) database,[26] 207 less-studied per and polyfluoroalkyl substances (PFAS), 44 flame retardants (FR) including organophosphate flame retardants (OPFR), 30 quaternary ammonium compounds (QACs), and 64 other industrial chemicals widely used in everyday life (e.g., plasticizers and over-the-counter medications).

2. High-Production Chemicals Obtained from EPA's Chemical Data Reporting 2016. We obtained a list of 8707 high-production (average national production volume over 25 000 lbs) chemicals from the U.S. EPA Chemical Data Reporting (CDR) 2016 database.[12] We queried their CASRN against the U.S. EPA CompTox Chemicals Dashboard[30] and kept 4963 chemicals that had molecular formulas. There were 3744 chemicals that were excluded because of unsuccessful matching of CASRN (n = 1370) and no matched molecular formula (potential mixtures, n = 2,374). We further restricted the Chemical Data Reporting list to include chemicals with formulas that were also included in the U.S. EPA suspect screening DSSTox desalted formula list to remove entries that were not LC amenable (e.g., metals). There were 3,380 Chemical Data Reporting chemicals remaining that corresponded to 2421 unique chemical formulas.

The final suspect database included 2421 unique chemical formulas and 3535 chemical entries after merging the in-house EOA database and Chemical Data Reporting lists and removing duplicated entries, chemicals with fewer than 100 units in mass or without formulas (e.g., chemical mixtures), and chemicals that are only gas-chromatography amenable. Gas-chromatography amenability was determined by examining the polarity of the chemicals. If a chemical did not have any polar groups, such as ROH or ROR, it was removed because it would not likely ionize in electrospray ionization. Structure information (SMILES and InChI keys) were obtained from PubChem search. This database was imported into the Agilent Mass Hunter Personal Compound Database and Library

https://dx.doi.org/10.1021/acs.est.0c05984
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

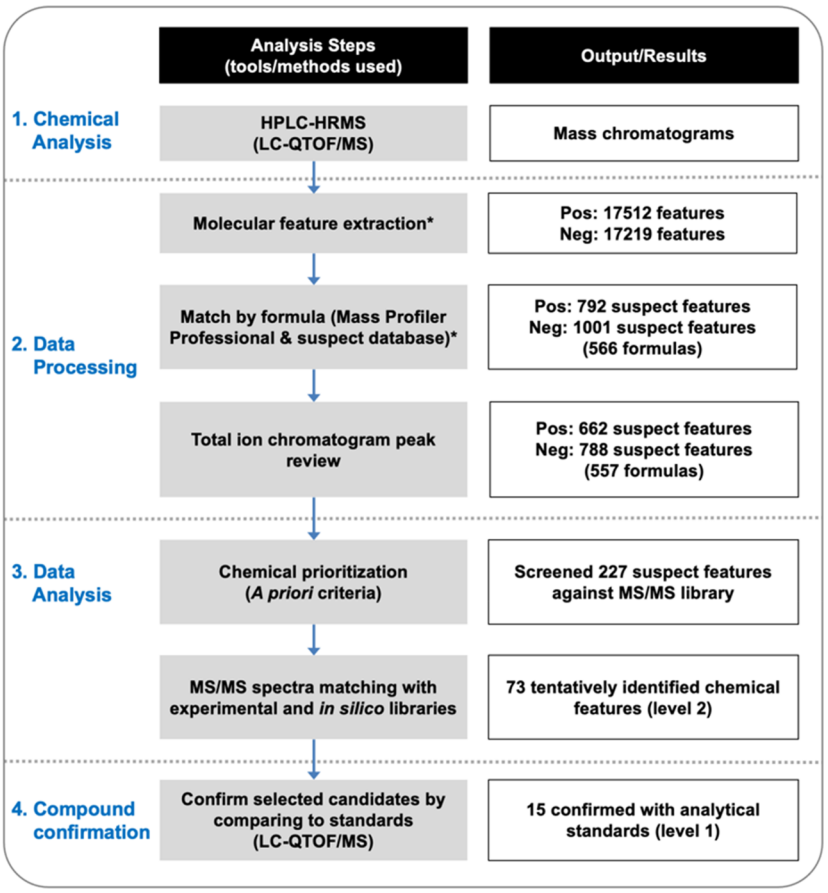


Figure 1. Suspect screen analysis workflow. *For detailed steps regarding feature extraction and formula matching, please refer to supplementary file (SI Figures S2 and S3). The annotation levels refer to the annotation scheme proposed by Schymanski et al.[31] for communicating confidence.

software (PCDL) for downstream suspect screening analysis. The suspect feature matching was done at the formula level, by matching an observed MS1 spectrum to theoretical spectra for MS-Ready formulas in PCDL. It is important to note that PCDL does not have retention times or reference MS1 or MS2 spectra.

LC-QTOF/MS Analysis. After compiling the final chemical suspect database, we performed HPLC/HRMS analysis using Agilent 1290 UPLC interfaced with Agilent 6550 QTOF/MS with electrospray ionization (ESI) in positive and negative mode (Agilent Technologies, Santa Clara, CA), data processing, data analysis, and compound confirmation with detailed steps listed below (Figure 1).

1. Chemical Analysis. Serum samples (250 μL) were extracted by protein precipitation with methanol. Ten μL of the serum extracts were then injected into the UPLC-QTOF/MS system. Both negative and positive ionization mode were studied. Agilent Eclipse Plus C18 (2.1 × 100 mm, 1.8 μm) column was used. Gradient A was made as 5 mM Ammonium Acetate in water (0.1% methanol). Gradient B was made as 5 mM ammonium acetate in methanol with 10% water. The gradient flow was set as 0.3 mL/min. The total ion chromatography (TIC) scan mass range was 100−1000 m/z. Quality control samples including blanks (LCMS grade water: Water, Burdick & Jackson for LC-MS, for HPLC, Burdick & Jackson, LC365-1; serum blank) and in-house laboratory control samples (matrix spike or LCS) were also analyzed together within one batch. Two technical replicates were analyzed for each sample.[32,33] The instrumental parameters are presented in the Supplementary Spreadsheet.

2. Quality Assurance/Quality Control. QA/QC samples were used to monitor the general performance of the injections, including retention time shifts, mass accuracy and peak intensity decay. Perfluoro-*n*-[1,2-13C2] octanoic acid (M2PFOA) was used as internal standard in negative ionization mode; triphenyl phosphate D15 and DL-cotinine (methyl D3) was used in the positive ionization mode. Blank samples were used to correct artificial features that might be introduced during sample preparation by removing features for which, abundances were no more than two times higher in the blanks compared to the samples. The blanks were made using ultraclean water (LCMS grade water: Water, Burdick & Jackson for HPLC, LC365-1) and the QCs were made using commercially available human AB serum (Corning Human AB Serum, 35060CI). QC serum is prepared using human AB serum spiked with 7 PFAS compounds and 6 OPFR compounds (SI Tables S2 and S3) (final concentration = 10 ng/mL in QC serum). The blank samples and the QC samples were treated the same way as the maternal and cord serum samples (SI Figure S2) following all the steps of the sample treatment and analysis.

For each batch, 10 pairs of maternal and cord matched serum samples, together with two water blanks, two blank serum samples, and two QC serum samples were extracted together and injected together in one batch. The samples were randomized, but each maternal and cord pair were run in the same batch to minimize any batch effect between maternal and

https://dx.doi.org/10.1021/acs.est.0c05984
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

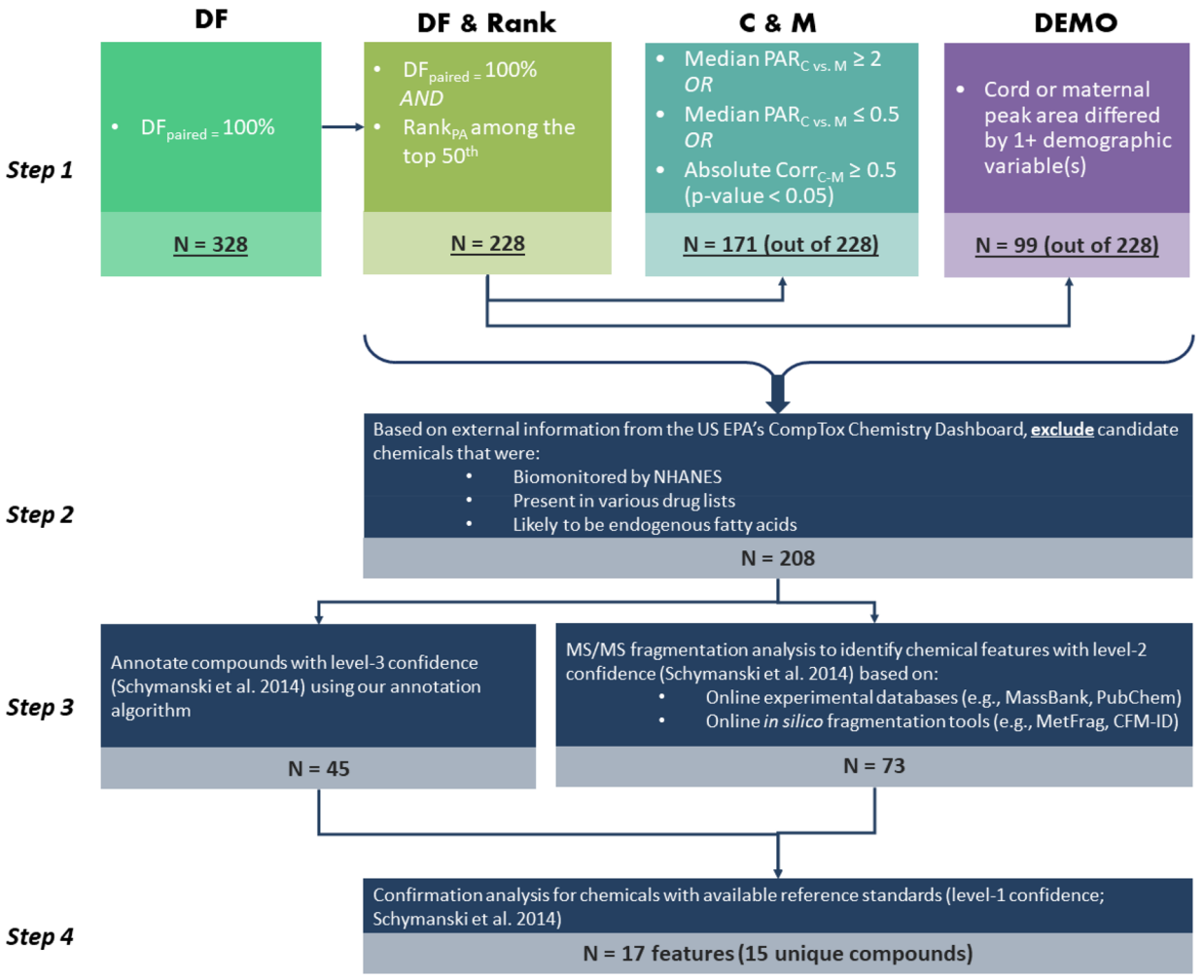


Figure 2. Steps for prioritizing chemicals of interest based on (1) a high detection frequency (suggesting ubiquitous exposure); (2) disproportionate distribution of peak area (relative concentrations) in fetal versus maternal serum (suggesting potentially different exposure concentration); (3) high correlation in peak area between fetal and maternal serum (suggesting that maternal concentration could be a proxy for fetal exposure); and/or (4) disproportionate distribution of peak area relative concentrations across maternal race/ethnicity or socioeconomic status (suggesting higher exposure to different demographic groups). The prioritized chemical features are then used to match against MS/MS spectral libraries and for confirmation with analytic standards. Abbreviations: DF: detection frequency; C&M: cord and maternal; PA: peak area; CvsM: cord compared to maternal peak area; C-M: cord-maternal; PAR: peak area ratio; Corr: correlation; DEMO: demographic differences; NHANES: National Health and Nutrition Examination Survey.

cord samples since we are interested in differences in peak areas between maternal and cord samples. Every sample was injected twice (instrumental replicate) to account for variability in peak areas originating from the instrument.

3. LC-QTOF/MS Data Processing. The obtained raw data files were processed following an optimized workflow described in detail elsewhere.[32,33] The workflow includes molecular feature extraction (MFE) to extract compound features across the batch data files and feature alignment using Agilent Masshunter Profinder software (version B.10.0) to align all features (identify and combine the same features by comparing their retention times and spectra) in each batch. For feature alignment across batches and formula matching we used Mass Profile Professional software (MPP version 12.06.01). The steps regarding feature extraction and formula matching is sketched in SI Figure S3. Each batch was composed of 10 pairs (N = 20 total) plus QC and blank samples for a total of three batches. After feature alignment, we kept only feature peaks with intensities at least two times higher than those in the water blank samples.

The chromatogram peak area, as integrated by the Agilent MassHunter Profinder software, is used a surrogate for chemical concentration allowing for comparisons of same chemical across samples and batches. This approach can only be used when studying the same chemical across samples and not when comparing two different chemicals due to potentially important differences in ionization efficiency. We used R (version 3.5.1) and Python (version 3.9.2) for our data processing and data analysis. The processing and analytical steps were (1) combining features obtained from all three batches, (2) averaging the peak area of the two technical replicates, (3) imputing values below the limit of detection, (4) performing batch correction, and (5) performing downstream analysis using batch-corrected peak areas.

Imputation of values below the limit of detection was conducted using a computational approach which assigned missing values based on the distribution of the data points. The measured abundances were log transformed and for each chemical across samples we calculated the median, the minimum and the standard deviation of the distribution.

https://dx.doi.org/10.1021/acs.est.0c05984
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

After fitting a normal distribution to the data points, the algorithm then generated random values between the measured minimum abundance (~5000) and the theoretical minimum (0) following the shape of the distribution. The algorithm is available on Github (https://github.com/dimitriabrahamsson/wangetal_maternal_cord.git).

Batch correction was conducted using a software package called ComBat,[34] which is commonly used in batch effect corrections in bioinformatics. One advantage of the ComBat package is that it can be used to correct for batch effect while preserving other differences across samples and that way avoiding overcorrection.

In addition to MassHunter Profinder, we expanded our MS/MS searching by employing MS-Dial,[35] which is an open source software package for analyzing nontargeted analysis data and has been developed by researchers at University of California, Davis and the RIKEN Center for Sustainable Resource Science (Japan).[35] Using the same parameters as for MassHunter (Supplementary Spreadsheet), we searched the "All public MS/MS" databases for positive MS/MS (13,303 unique compounds) and for negative MS/MS (12 879 unique compounds).

Descriptive and Statistical Analysis. We developed a workflow which uses descriptive and statistical analysis methods to prioritize chemical suspects in the large universe of chemical features that are detected with HRMS for further analysis. For our prioritization, chemicals of interest were those with (1) a high detection frequency (suggesting ubiquitous exposure); (2) disproportionate distribution of peak area (relative concentrations) in fetal versus maternal serum (suggesting potentially different exposure concentration); (3) high correlation in peak area between fetal and maternal serum (suggesting that maternal concentration could be a proxy for fetal exposure); and/or (4) disproportionate distribution of peak area relative concentrations across maternal race/ethnicity or socioeconomic status (suggesting higher exposure to different demographic groups). Accordingly, we derived different measures to evaluate these criteria of interest as described below.

First, we obtained the detection frequencies for the paired maternal and cord samples (DF_{paired}) as an indicator of how widespread chemical features may be among pregnant women and their newborns, ranging from zero to 30. DF_{paired} of one means that the feature was detected in both the maternal and cord samples obtained from the same participant. We also ranked the features according to their median peak area across both maternal and cord samples from largest to smallest peak area ($rank_{PA}$, smaller ranks corresponds to larger peak areas) as a proxy to identify features that may be of higher abundance.[20]

Second, we conducted two assessments of the relationship between maternal and cord peak areas: (1) the ratios of cord vs maternal peak areas ($PAR_{C\ vs\ M}$), and (2) the Spearman correlation between cord and maternal peak areas ($Corr_{C\text{-}M}$). A $PAR_{C\ vs\ M}$ greater than 1 indicated that the peak area of this feature was higher in cord serum than in maternal serum, whereas a value less than 1 means the peak area was higher in maternal than in cord serum. Features with an absolute $Corr_{C\text{-}M}$ value of at least 0.5 and a p-value of less than 0.05 were considered to have a statistically significant correlation between cord and maternal peak area.

Third, among those chemical features with detection frequencies of at least 80% in maternal or cord serum samples, we assessed separately whether the peak areas in cord or maternal serum samples differed by race/ethnicity, education, household income, and nativity (U.S.-born status). Linear regression with batch adjustment was used if the log-transformed peak area passed the Shapiro normality test (p-value being at least 0.05). Otherwise, logistic regression of the highest tertile of the peak area was used, adjusting for batch. When there is zero cell for the tabulation of peak area (highest tertile vs other) and the demographic variable, nonparametric Kruskal–Wallis test was used. A p-value less than 0.05 was considered statistically significant. The statistical analyses on the relationship between chemical features and demographic variables were not adjusted for multiple comparisons, as the main goal was to inform the prioritization of potential suspect chemicals.

Chemical Prioritization Criteria and Steps. Based on the criteria of interest and their corresponding measures, we used an iterative four-step approach to prioritize and select chemicals for confirmation using reference standards (Figure 2). We assigned confidence levels to features based on the scale developed by Schymanski et al.[31] All features extracted by MassHunter Profinder and/or with MS-Dial were at first considered level-5 annotations. The features that were assigned chemical formulas based on accurate mass, isotope patterns and abundance were assigned level-4 identification confidence. The ESI adducts that were used for matching formulas were H^+, Na^+, and NH_4^+ in positive mode and CH_3COO^- in negative mode.

The chemical candidates matched to suspect features by formula, and could be annotated with a tentative structure, were considered level-3 identification confidence. Due to the large variability and uncertainty in the level-3 annotations, we developed and applied a scoring algorithm for distinguishing between likely accurate and likely inaccurate level 3 annotations. As a first step, we collected all isomers for a given formula that could be found in EPA's CompTox Chemicals Dashboard.[30] We then calculated the probability of blindly picking the right isomer (called "blind probability") by dividing 1 by the number of available isomers. For example, if a formula had only 1 available isomer the probability of blindly picking the right isomer is 1, whereas if a formula had 100 available isomers the probability is 1/100 = 0.01. We then collected the number of Dashboard data sources, PubChem data sources, PubMed publications and CPDAT count for each isomer and normalized the data in each column (i.e., Dashboard data sources, PubChem data sources, etc.) from 0 to 1 for every group of isomers that corresponded to one formula. We then calculated the average source score (called "source score") for every isomer by taking the average of Dashboard, PubChem, PubMed and CPDAT scores. Finally, we calculated the overall score by taking the average of the blind probability and the source score. We decided to calculate the final score this way instead of taking the average of all numbers in order to give more weight to blind probability instead of the source score. The algorithm is available on Github (https://github.com/dimitriabrahamsson/wangetal_maternal_cord.git).

The features, for which there was some evidence to propose an exact structure based on experimental MS/MS spectra, or in silico MS/MS spectra, were considered level-2 annotations. Otherwise, they remained as level-3 or level-4 annotations. For a select number of prioritized features, we collected targeted MS/MS fragmentation spectra in both positive and negative electrospray ionization modes with collision energies of 10 eV,

https://dx.doi.org/10.1021/acs.est.0c05984
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

20 eV, and 40 eV with a scan rate of four spectra/s and a retention time window of ±1 min. The spectra for all three collision energies were collected simultaneously. The spectra were collected following data dependent acquisition (DDA) and a targeted MS/MS method for the prioritized chemical features.

The acquired spectra were then used to search for potential matches (at least one fragment peak with mass error <10 ppm) in available experimental MS/MS spectral libraries (MS-Dial databases,[35] MassBank of Europe and North America,[36] HMDB[37] and mzCloud[38]), and in in silico spectral computational tools (CFM-ID[39] and MetFrag[40]). For both the experimental databases and the in silico tools, we searched compounds for which we could observe a chromatographic peak for the molecular ion and for peaks which the isotopic pattern had a score of 70 out of 100 or higher. We then annotated the observed features with the top candidate ion suggested by the software's algorithm.

Suspect features that were confirmed using a reference standard with MS, MS/MS and retention time matching were assigned level-1 confidence in identification.

Step 1. Based on results from descriptive and statistical analysis, we selected chemical features that meet the following criteria (Figure 2):

a. DF_{paired} = 100%
b. $Rank_{PA}$ among the top 50th percentile
c. Having maternal and cord peak area relationship of interest: median $PAR_{C\ vs\ M} \geq 2$ or median $PAR_{C\ vs\ M} \leq 0.5$; absolute $Corr_{C\text{-}M} \geq 0.5$ (p-value <0.05). Median $PAR_{C\ vs\ M} \geq 2$ means that half of the cord samples had peak areas at least two times the median peak area among maternal samples, while median $PAR_{C\ vs\ M} \leq 0.5$ means that half of the maternal samples had peak area of at least two times of the median peak area among cord samples.
d. The peak area of cord or maternal samples were different across at least one demographic variable (race/ethnicity, education, household income, or nativity).

Step 2. For chemical features meeting the criteria in step 1, we merged back the candidate chemical names from our suspect chemical database based on formula, and then queried the U.S. EPA's CompTox Chemicals Dashboard[30] by CASRN to obtain additional information on the candidate chemicals, including whether they: are biomonitored by NHANES, are present in various drug lists (e.g., the DrugBank database from the University of Alberta), have associated ToxCast assay information, and are on the high production volume list or the chemical and products database. For the purposes of this paper, which focuses on exogenous chemicals, we further prioritized chemicals that were not biomonitored by NHANES, not pharmaceutical drugs, and not likely to be endogenous fatty acids (based on chemical structure). However, there are certain endogenous compounds, such as cortisol and bile acids, that have shown some associations with preterm birth in previous studies and might be of interest for future investigation. For that reason, we included four endogenous compounds in the prioritized list for MS/MS spectra matching: cortisol, progesterone, deoxycholic acid, and chenodeoxycholic acid (three unique formulas; deoxycholic acid and chenodeoxycholic acid share the same formula).

Step 3. To increase the likelihood of confirmation with reference standards given the limited volume of serum samples, we performed fragmentation analysis by checking the fragmentation peaks against various sources, including online experimental databases, such as the MS-Dial databases,[35] MassBank,[36] and mzCloud,[38] and spectral data generated by the online in silico fragmentation tools such as CFM-ID.[39] Chemical features with at least one matched fragment peak were assigned a level-2 confidence in identification as probable structures. All the remaining features were assigned a level-4 confidence in identification.[31]

Step 4. We further conducted confirmation analysis for chemicals with reference standards that were commercially available.

Chemical Confirmation Using Reference Standards. Among the level-2 identified chemical features with available reference standards, we confirmed the presence of chemical features by rerunning the LC-QTOF/MS analysis with their corresponding reference standard. A suspect feature was considered confirmed (present in maternal or cord serum) with level-1 confidence in annotation[31] if it had the same retention time (RT), accurate mass, and MS/MS spectral pattern as the LC- QTOF/MS results for the reference standard.

Database Searching for Previously Reported Structures and Chemical Uses. After collecting all the structural information on the detected features, we searched several databases to collect information on a chemical compound's reported chemical use and its presence in previous exposure studies. For this search we used all the chemicals in the top 3 levels of annotation (1−3) as proposed by Schymanski et al.[31] As a first step, we searched the Human Metabolome Database[37] to find which compounds were known endogenous compounds. We then searched EPA's CompTox Chemicals Dashboard[30] to find which chemicals have known uses as pharmaceuticals, pesticides, flame retardants, poly/perfluorinated alkyl substances (PFAS), plasticizers, cosmetics, consumer products, and which chemicals are registered as high production volume chemicals. Finally, we searched the Blood Exposome database[41,42] to find which chemicals had been previously reported in human blood samples in previous studies.

RESULTS

Participant Characteristics. The mean age of participants was 32 years (SD: 4.7, Table 1). Nearly half of the participants were Latinas, 37% were non-Hispanic whites, and 17% were non-Hispanic other race. Around one-third of the pregnant women were of higher socioeconomic status, with 40% having some postgraduate education and 30% having an annual household income ≥ $125,000. Half of the study participants were born outside of the U.S., and, on average, had lived in the U.S. for 22 years.

Suspects by Ionization Modes and Across Maternal Vs Cord Samples. After data processing, we detected in total 1,450 suspect features (herein referred to as "suspects") that were matched to 557 unique chemical formulas. Of the 1450 suspect features, we detected 662 suspects in the positive ion mode and 788 suspects in the negative ion mode, with 282 detected in both ion modes. We observed some limited batch effect related to how the samples were analyzed in the instrument (SI Figure S5). Correcting for that effect with ComBat resulted in small changes in the abundances of the samples (SI Figure S5). We also observed statistically significant differences in the abundance of some of the tracers

https://dx.doi.org/10.1021/acs.est.0c05984
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

Table 1. Demographics of the Current Analytical Sample (N = 30 Matched Maternal/Cord Samples)[a]

characteristics	mean (SD)	N (%)
age	32.4 (4.7)	
race/ethnicity		
Latinas		14 (47)
non-Hispanic whites		11 (37)
non-Hispanic Asians/Pacific Islanders/African Americans		5 (17)
Educational Attainment		
high school/GED or less		11 (37)
some college/AA/College completed		7 (23)
master's or doctoral degree		12 (40)
Household Income		
<$40,000		12 (40)
$40,000 − $124,999		9 (30)
≥$125,000		9 (30)
Nativity (Born in the U.S.)		
yes		14 (47)
no		15 (50)
DK/NA		1 (3)
years lived in the U.S.	22.0 (12.3)	
Infant Sex		
male		15 (50)
female		15 (50)

[a]Abbreviations: SD: standard deviation; GED: General Education Diploma; AA: Associate in Arts; DK: do not know; NA: not available.

across different batches (SI Figure S6 and S7). Even though these differences are relatively small and only three tracers showed significant differences (SI Figures S6 and S7), we chose to proceed with batch correction to remove any effect related to instrumental variability. This is particularly important for our statistical analyses since we use instrument abundances instead of concentrations, which would control for that effect. Median RT of all detected suspects was 8.9 min (range: 0.9−17.0) and the majority of suspects detected were compounds with mass values of 500 or less (98%).

When looking at the mass accuracy and retention time consistency across batches, the mass errors for the tracer compounds used in positive mode were all below 6 ppm and in negative mode below 5 ppm. The retention times for the tracer compounds in both modes showed only minor shifts approximating 0.2 min in positive mode and 0.3 min in negative mode in the worst cases (SI Tables S2 and S3).

When looking at the differences between maternal and cord samples, 1225 suspects (85%) were detected in at least one paired maternal-cord sample whereas 225 features (15%) were detected in either maternal or cord samples, but not in both pairs. (Figure 3). Three hundred and twenty-eight suspects (23%) were detected in all paired maternal-cord samples. Around half of the suspects (51%) had detection frequencies of 14 or greater among maternal-cord pairs. More suspects with a higher DF in cord relative to maternal samples were found in the negative mode and slightly more suspects with a higher DF in maternal relative to cord samples were found in the positive mode (SI Table S1 and Figure S4 for an overview of the suspects detected in the positive and the negative modes). It is important to note that Figure 3 shows only the features that were present in the suspect list. When looking at all the detected features regardless of their presence in the suspect list, there are approximately 1.5 times more positive ionization features than negative ionization features.

Among these 1225 suspects, the median $PAR_{C\ vs\ M}$(across all samples for a specific feature) for the 643 suspects detected in the negative mode was 1.1 (IQR: 0.7−1.7) and the median $PAR_{C\ vs\ M}$ among582 suspects detected in the positive mode was 0.9 (IQR: 0.5−1.5) (Figure 4A). Peak areas in maternal samples were numerically higher relative to the peak areas in cord samples among suspects detected in the positive mode but were numerically lower relative to the peak area in cord samples among suspects detected in the negative mode (Figure 4B). More suspects detected in the negative mode, compared to those in the positive mode, had a median cord peak area at least twice that of the median maternal peak area (median $PAR_{C\ vs\ M} \geq$ 2:15% vs 10%). On the contrary, more suspects detected in the positive mode, compared to those in the negative mode, had a median maternal peak area that was at least twice that of the median cord peak area (median $PAR_{C\ vs\ M} \leq$ 0.5:21% vs 10%).

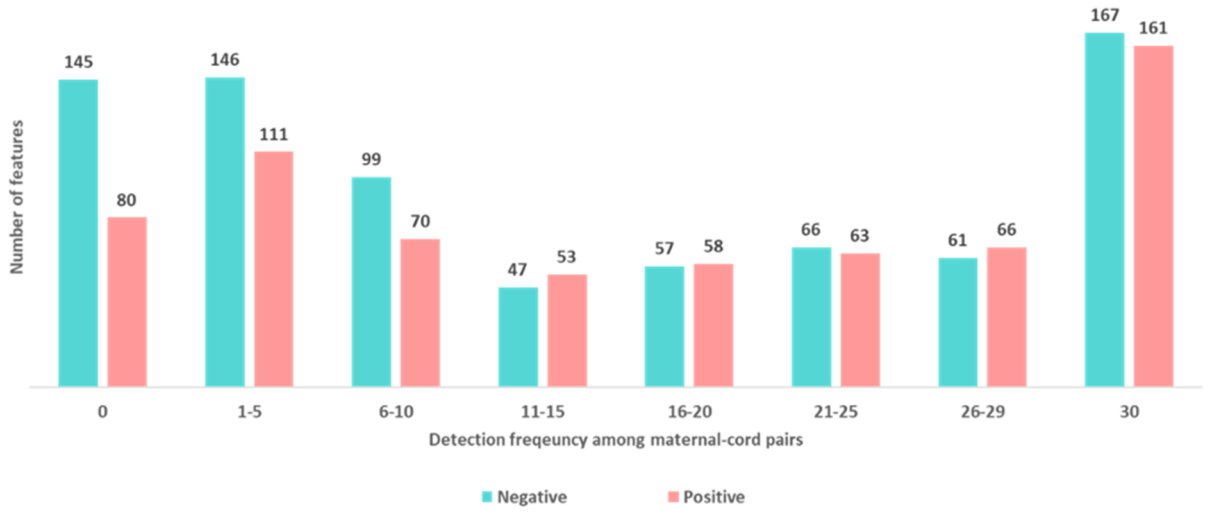


Figure 3. Number of suspects by detection frequency among the maternal-cord serum pairs (n = 30).

https://dx.doi.org/10.1021/acs.est.0c05984
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

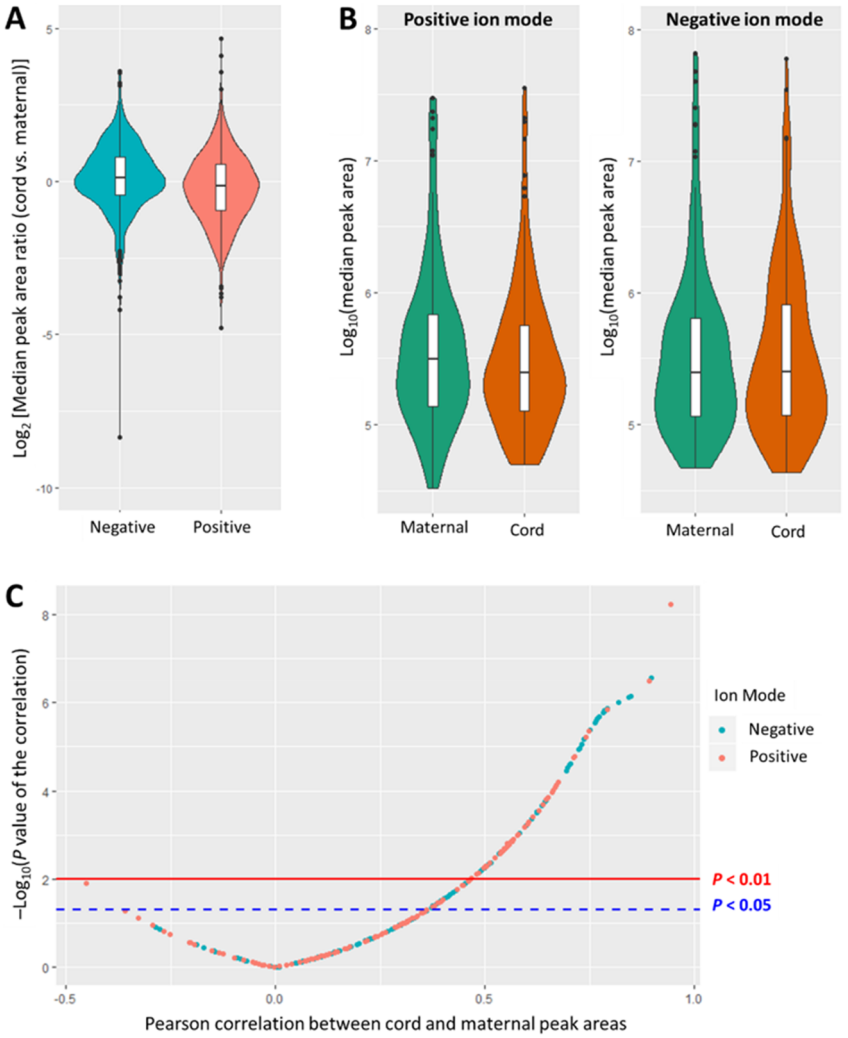


Figure 4. Relationship between cord and maternal peak area. A. Distribution of median peak area ratio (cord vs maternal) among 1225 suspects detected in at least 1 paired maternal-cord sample; B. Distribution ofmedian peak area by sample type and ion modes; C. Correlation between cord and maternal peak area among 328 features detected in all maternal-cord pairs.

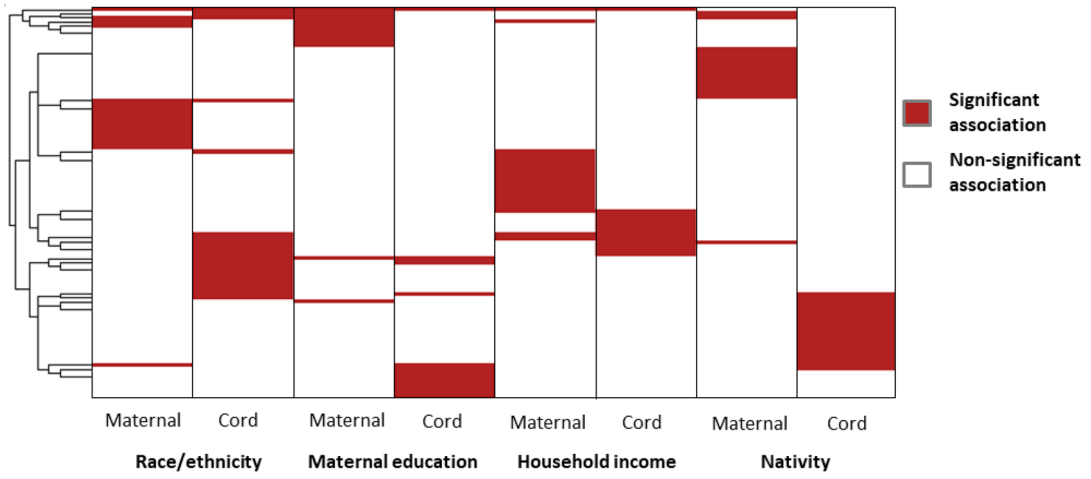


Figure 5. Clustering of suspects (row) in cord and maternal serum whose peak area significantly differed by at least one demographic variable (column).

For the 328 suspects detected in all paired samples, we further explored the correlation between cord and maternal peak area (Figure 4C). There were 104 features with a Spearman correlation of at least 0.5 and a p-value < 0.05. Despite that the majority of suspects were detected in at least one maternal and one cord sample, 133 suspects (9.2%) were

https://dx.doi.org/10.1021/acs.est.0c05984
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

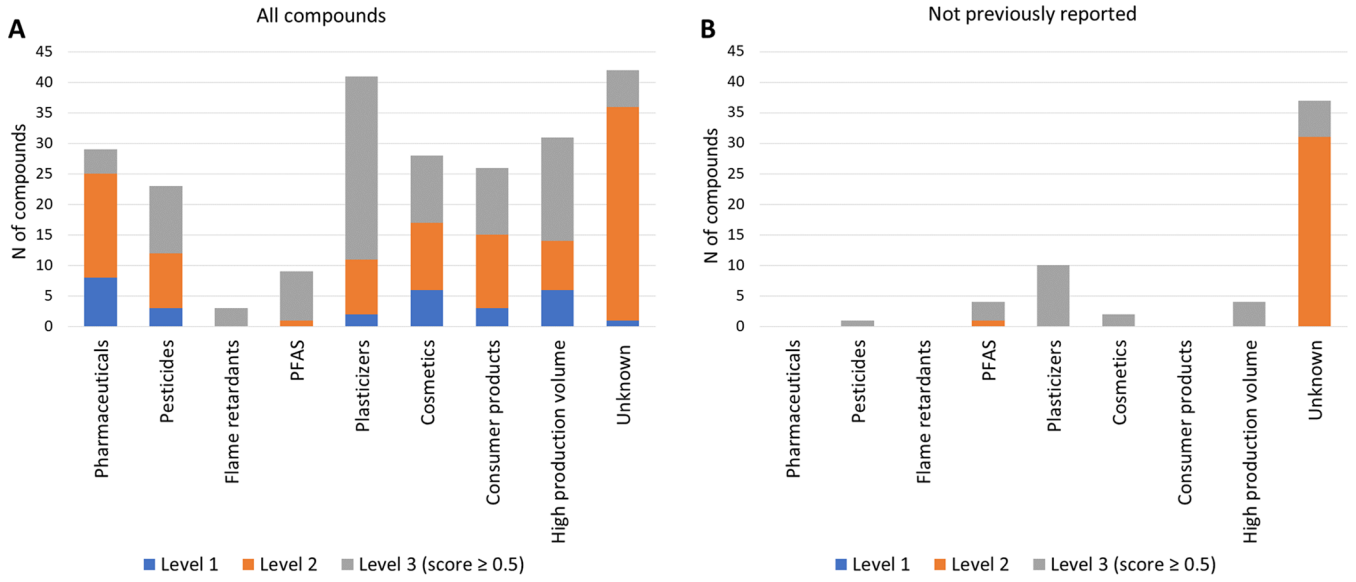


Figure 6. Chemical uses information for (A) all annotated compounds and (B) for compounds that were found to not have been previously reported in human exposure studies involving human blood or serum samples. The annotations are shown by confidence level as proposed by Schymanski et al.[31] The chemical use information was collected from databases on EPA's CompTox Chemicals Dashboard.[30] The Human Metabolome Database[37] was used to remove chemical features with endogenous sources. The Blood Exposome database[41,42] was used to determine if a compound had been previously reported in human exposure studies.

detected exclusively in maternal or cord serum samples. There were 666 suspects in maternal samples and 648 suspects in cord samples with detection frequencies of over 80% (n = 24). Among these, the peak areas of 114 and 102 suspects in maternal and cord samples, respectively differed across at least one of four demographic variables. There were 99 suspects that were detected in all 30 paired samples with peak areas in cord or maternal samples that differed across at least one demographic variables. Most of the suspects differed by a specific demographic variable either when examining maternal peak area or cord peak area but not both (Figure 5), suggesting that demographic differences in peak area of suspects may vary by sample type (maternal versus cord). Among features that significantly differed by each corresponding demographic variable, more features had a higher median peak area in maternal samples among women who were non-Latinas (relative to Latinas), had some college education or above (relative to those with a high school education or less), had a household income of $40,000 or more (relative to those with a household income of less than $40,000), and were born in the U.S. (relative to those who were not). Features' median peak area in cord samples showed a similar pattern except that more features had a higher median peak area among women who were not born in the U.S. (Supplementary Spreadsheet S2).

Features Selected for Fragmentation Analysis. Based on the chemical prioritization criteria and steps described above in the Materials and Methods Section (Figure 2), we selected 106 suspects detected in positive mode and 102 suspects (total n = 208) detected in negative mode for fragmentation analysis (Figure 2). After inspecting the MS/MS matches to the MS/MS libraries, we tentatively identified 73 chemical features (level 2 confidence) (Supplementary Spreadsheet: "Level 1−2 annotations").

Confirmed Features. After purchasing analytical standards and comparing the mass spectrum of the detected features and that of the corresponding standards, we confirmed the presence of 17 chemical features (Supplementary Spreadsheet: "Level 1−2 annotations"), which came down to 15 unique chemical compounds after removing duplicates between positive and negative ionization mode (cortisone) and after removing stereoisomers (chenodeoxycholic acid) (Supplementary Spreadsheet: "Level 1−2 annotations" and "Annotations summary").

Database Search. When looking at the top scored annotations 1, 2, and 3 (score ≥0.5), the largest group, with 42 annotated compounds, were chemical compounds for which there was limited to no available information on their chemical uses, their presence in consumer products and whether they were high production volume chemicals (Figure 6 and Supplemental Spreadsheet: "Annotations summary"). The majority of these chemicals (33/42) were annotated with MS/MS spectral libraries (level 2 annotations). The second largest group was plasticizers with 29 compounds. After removing the compounds that had been previously reported in human exposure studies, we found 55 chemical compounds that had not been previously reported. Also, in this case, the largest group consisted of chemicals with limited to no information (Unknowns; n = 37) and the second largest group consisted of plasticizers (n = 10). We also found four PFAS that, according to our method, appeared to not have been previously reported in human blood/serum: 4m perfluorooctanesulfonic acid, 6:2 fluorotelomer phosphate monoester, methyl perfluoroundecanoate, and 2-perfluorooctyl ethanoic acid. However, upon closer examination with literature review, we found that only methyl perfluoroundecanoate, and 2-perfluorooctyl ethanoic acid had not been previously reported, while 4m perfluorooctanesulfonic acid, 6:2 fluorotelomer phosphate monoester showed to have been reported in a very limited number of studies.

DISCUSSION

Suspect screening and nontargeted analysis approaches have been increasingly used for both environmental monitoring[19−21,43−45] and studying human exposure to known and

https://dx.doi.org/10.1021/acs.est.0c05984
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

unknown chemicals.[25,26,46] However, most studies evaluating human samples have focused on endogenous compounds and our study is the first—to our knowledge—that screens for a comprehensive database of industrial chemicals. Further, we have additionally expanded analytic capacity through MS/MS fragmentation analysis in both maternal and cord serum samples to assist in the identification of chemicals. With our study of focused screening of matched maternal and cord serum samples for high production volume industrial chemicals, our study provides valuable insights on fetal exposure to previously unreported chemicals.

While our study could be described as both "suspect screening" and "non-targeted analysis", we chose the terminology "suspect screening" because if fits better our focused search of industrial chemicals that are "suspected" to be present in human blood. In addition, while nontargeted analysis or untargeted metabolomics studies prioritize features for MS/MS fragmentation based on detection frequency and abundance,[22,47,48] we chose to prioritize features that showed some significance in terms of partitioning between maternal and cord blood and in terms of demographic variables, shifting our focus from the most abundant features to exogenous chemical features that are "suspected" to have some biological and/or demographic significance. This workflow can be used for methods prioritizing chemicals for further evaluation and adds to other approaches for prioritizing the chemical space for targeted biomonitoring.[13]

Following our suspect screening workflow, we found 42 chemical compounds that had limited to no information on their sources and use and could not be grouped under the categories of endogenous, pharmaceuticals, pesticides, flame retardants, PFAS, plasticizers, ingredients in cosmetics and consumer products or high production volume chemicals, as classified in EPA's Chemicals Dashboard.[30] After removing the chemical compounds that had been previously reported in human exposure studies, we found 37 chemical compounds that had limited to no information and could not be grouped in any of our categories (Figure 6 and Supplemental Spreadsheet: "Annotations summary" and "Not previously reported"). Some examples of these chemicals are pyrenophorol, thermopsine, and thymol-beta-D-glucoside. The identification of chemicals with unknown sources and uses is likely reflective of gaps in requirements for disclosing use of chemicals in consumer and industrial products.[49] Previous work on suspect screening of chemicals in consumer products has shown that only 30.5% of the chemicals used in consumer products are reported in chemical lists with known chemicals used in these applications.[49]

We tentatively identified a number of chemicals that had not been previously reported in other biomonitoring studies. Some examples of chemicals with known sources and uses but that had not been previously reported were (i) 1-(1-acetyl-2,2,6,6-tetramethylpiperidin-4-yl)-3-dodecylpyrrolidine-2,5-dione, which is a known high production volume chemical used in consumer products, such as fragrances; (ii) methyl perfluoroundecanoate, and 2-perfluorooctyl ethanoic acid, which are two PFAS; and (iii) Sumilizer GA 80, which is a plasticizer (Supplemental Spreadsheet: "Not previously reported"). It is important to note that although our database search for finding not previously reported chemicals is extensive, it may in some limited cases produce false positives. As illustrated by two PFAS (4m perfluorooctanesulfonic acid, 6:2 fluorotelomer phosphate monoester), there may be cases where less well-studied chemicals may appear as not-previously reported but they may be reported in human blood/serum by a very limited number of studies. Nevertheless, these chemicals require further investigation due to their very limited information in the literature.

The large presence of poorly characterized chemicals in maternal and cord blood samples warrants further investigation to understand where these chemicals might be coming from and how they may affect human health. We found that, in general, the levels of detected features were similar between cord and maternal samples (Figures 4A,B), indicating that the majority of the chemicals observed do not show differential partitioning between maternal and cord blood and that they can cross the placenta without being inhibited by filtering processes. It is important to acknowledge, however, that this finding could be an artifact of the analytical instrumentation (LC-QTOF/MS) used in this study, which is primarily focused on polar and involatile chemicals. Polar chemicals are generally hydrophilic and dissolve well in blood making it easy for them to cross the placenta as part of the blood flow from the mother to the fetus. An additional analysis of the samples with instruments that focus on nonpolar and volatile/semivolatile chemicals, such as gas chromatography (GC)−QTOF/MS, might present a different picture. Nonpolar chemicals may bind to lipids in the placenta which may slow down their transfer to the fetus. This is a hypothesis that could be explored further in future studies.

While the majority of chemicals that were detected in maternal samples were also detected in cord samples, 133 suspects (9.2%) were detected exclusively in maternal or cord serum samples. This finding indicates that there may be certain suspects that appear exclusively on the maternal or on the fetal side. However, it is important to note that the detection frequency is calculated based on the number of chemicals that were able to pass the detection threshold of the current method and that a "non-detect" does not necessarily mean "non-present." Thus, a more likely scenario is that these 133 features were present, but at low amounts that could not be detected with the current analytical method.

For several suspect features, we observed significant differences across socioeconomic and racial/ethnic groups indicating differential exposures to certain chemical compounds. We observed, for example, that among features that significantly differed by each corresponding demographic variable, more features had a higher median peak area in maternal samples among women with a household income of $40,000 or more. This finding could indicate important socioeconomic differences in the purchase and use of consumer products. This observation aligns with Montazeri et al.,[50] in their systematic review of multiple biomonitoring studies, in which they observed that environmental exposures are not exclusively associated with lower socioeconomic status, and that for many environmental contaminants, higher levels can occur in groups with higher socioeconomic status.

We found 23% of the detected features were matched with a chemical formula from our database (Figure 1). Given that we focused on high volume chemicals, we anticipated that we might find more matches. However, many suspects may be of relatively low concentration in the samples, as are most industrial chemicals, and in many cases, they may be below the detection limit of the analytical method. Targeted analysis with analytical methods of lower mass resolution but higher sensitivity, such as LC-triple quadrupole MS (LC-TQ/MS),

https://dx.doi.org/10.1021/acs.est.0c05984
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

could reveal the presence of additional compounds. This observation indicates that nontargeted analysis techniques could benefit from broad screening semitargeted methods, where hundreds or thousands of analytical standards are used to screen for specific chemical compounds. Also, there may be byproducts of metabolism that are generated through the activation, detoxification and elimination of exogenous synthetic chemical compounds. These industrial chemical metabolites can make up a large portion of the human chemisome of which more than 95% remain unknown or largely uncharacterized[51−53] and thus are not included in the current suspect database. Finally, some exposures may not be present due to biotransformation and metabolism inside human body. Future studies can consider including predicted metabolites from environmental chemicals of interest that are generated by recently developed computational tools such as the BioTransformer[51] in order to capture exposure to all possible forms of these chemicals.

Our study adds important information to a very scarce body of literature on suspect screening and nontargeted analysis of industrial chemical exposures in maternal and fetal pairs. Our results show that there are potential new chemical exposures that have not been adequately characterized and have not been previously of concern for environmental health scientists and regulators. Our study is an important methodological approach for future studies that will aim at characterizing the presence and toxicity of newly detected chemical compounds in the human body and assess the fate of these compounds in various human tissues, particularly between the mother and the fetus. Understanding these exposures and how they may contribute to adverse health outcomes is crucial in characterizing the human exposome and eventually preventing the development of disease.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.0c05984.

Figures S1−S7 and Tables S1−S3 (PDF)

Contains tables, spreadsheets and raw data referenced throughout the manuscript (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Tracey J. Woodruff − *Department of Obstetrics, Gynecology and Reproductive Sciences, Program on Reproductive Health and the Environment, University of California, California 94143, United States*; Email: tracey.woodruff@ucsf.edu

Authors

Aolin Wang − *Department of Obstetrics, Gynecology and Reproductive Sciences, Program on Reproductive Health and the Environment, University of California, California 94143, United States*

Dimitri Panagopoulos Abrahamsson − *Department of Obstetrics, Gynecology and Reproductive Sciences, Program on Reproductive Health and the Environment, University of California, California 94143, United States*; orcid.org/0000-0002-3402-7565

Ting Jiang − *California Environmental Protection Agency, Department of Toxic Substances Control, Environmental Chemistry Laboratory, Berkeley, California 94710, United States*

Miaomiao Wang − *California Environmental Protection Agency, Department of Toxic Substances Control, Environmental Chemistry Laboratory, Berkeley, California 94710, United States*

Rachel Morello-Frosch − *Department of Environmental Science, Policy and Management and School of Public Health, University of California, Berkeley, Berkeley, California 94710, United States*

June-Soo Park − *California Environmental Protection Agency, Department of Toxic Substances Control, Environmental Chemistry Laboratory, Berkeley, California 94710, United States*

Marina Sirota − *Bakar Computational Health Sciences Institute and Department of Pediatrics, University of California, San Francisco, California 94158, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.0c05984

Author Contributions

[+]These two authors contributed equally

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was funded by NIH/NIEHS grant numbers P30-ES030284, UG3OD023272, UH3OD023272, P01ES022841, R01ES027051 and by the U.S. EPA grant number RD83543301. We thank Courtney Cooper for assistance with the manuscript preparation, language editing, and proof reading. All raw data and spectra was submitted to the Metabolomics Workbench (https://www.metabolomicsworkbench.org/).

■ REFERENCES

(1) Diamanti-Kandarakis, E.; Bourguignon, J.-P.; Giudice, L. C.; Hauser, R.; Prins, G. S.; Soto, A. M.; Zoeller, R. T.; Gore, A. C. Endocrine-Disrupting Chemicals: An Endocrine Society Scientific Statement. *Endocr. Rev.* **2009**, *30* (4), 293−342.

(2) National Cancer Institute. *Reducing Environmental Cancer Risk: What We Can. Do Now. President's Cancer Panel 2008- 2009 Annual Report.*; Bethesda, MD, 2010.

(3) The American College of Obstetricians and Gynecologists. Exposure to Toxic Environmental Agents. Committee Opinion No. 575. *Obstet. Gynecol.* **2013**, *122* (4), 931−935.

(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.

(5) Woodruff, T. J.; Zota, A. R.; Schwartz, J. M. Environmental Chemicals in Pregnant Women in the United States: NHANES 2003−2004. *Environ. Health Perspect.* **2011**, *119* (6), 878−885.

(6) U.S. Centers for Disease Control and Prevention. The Fourth National Report on Human Exposure to Environmental Chemicals (Updated Tables, January 2017), 2017.

(7) Needham, L. L.; Grandjean, P.; Heinzow, B.; Jørgensen, P. J.; Nielsen, F.; Patterson, D. G.; Sjödin, A.; Turner, W. E.; Weihe, P. Partition of Environmental Chemicals between Maternal and Fetal Blood and Tissues. *Environ. Sci. Technol.* **2011**, *45* (3), 1121−1126.

(8) 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. *J. Toxicol. Environ. Health, Part B* **2014**, *17* (3), 175−203.

(9) 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. *Environ. Sci. Technol.* **2016**, *50* (22), 12464−12472.

(10) U.S. Environmental Protection Agency. Chemical Data Reporting Fact Sheet: Basic Information https://www.epa.gov/chemical-data-reporting/chemical-data-reporting-fact-sheet-basic-information (accessed 2017/6/19).

(11) Di Renzo, G. C.; Conry, J. A.; Blake, J.; DeFrancesco, M. S.; DeNicola, N.; Martin, J. N.; McCue, K. A.; Richmond, D.; Shah, A.; Sutton, P.; Woodruff, T. J.; van der Poel, S. Z.; Giudice, L. C. International Federation of Gynecology and Obstetrics Opinion on Reproductive Health Impacts of Exposure to Toxic Environmental Chemicals. *Int. J. Gynecol. Obstet.* **2015**, *131* (3), 219−225.

(12) U.S. Environmental Protection Agency. 2016 Chemical Data Reporting Results https://www.epa.gov/chemical-data-reporting/2016-chemical-data-reporting-results (accessed 2017/7/22).

(13) Pellizzari, E. D.; Woodruff, T. J.; Boyles, R. R.; Kannan, K.; Beamer, P. I.; Buckley, J. P.; Wang, A.; Zhu, Y.; Bennett, D. H. (Environmental influences on Child Health Outcomes). Identifying and Prioritizing Chemicals with Uncertain Burden of Exposure: Opportunities for Biomonitoring and Health-Related Research. *Environ. Health Perspect.* **2019**, *127* (12), 126001.

(14) Andra, S. S.; Austin, C.; Patel, D.; Dolios, G.; Awawda, M.; Arora, M. Trends in the Application of High-Resolution Mass Spectrometry for Human Biomonitoring: An Analytical Primer to Studying the Environmental Chemical Space of the Human Exposome. *Environ. Int.* **2017**, *100*, 32−61.

(15) Dennis, K. K.; Marder, E.; Balshaw, D. M.; Cui, Y.; Lynes, M. A.; Patti, G. J.; Rappaport, S. M.; Shaughnessy, D. T.; Vrijheid, M.; Barr, D. B. Biomonitoring in the Era of the Exposome. *Environ. Health Perspect.* 2017, *125* (4). DOI: 10.1289/EHP474.

(16) Wambaugh, J. F.; Bare, J. C.; Carignan, C. C.; Dionisio, K. L.; Dodson, R. E.; Jolliet, O.; Liu, X.; Meyer, D. E.; Newton, S. R.; Phillips, K. A.; Price, P. S.; Ring, C. L.; Shin, H.-M.; Sobus, J. R.; Tal, T.; Ulrich, E. M.; Vallero, D. A.; Wetmore, B. A.; Isaacs, K. K. New Approach Methodologies for Exposure Science. *Current Opinion in Toxicology* **2019**, *15*, 76−92.

(17) Walker, D. I.; Valvi, D.; Rothman, N.; Lan, Q.; Miller, G. W.; Jones, D. P. The Metabolome: A Key Measure for Exposome Research in Epidemiology. *Current Epidemiology Reports* **2019**, *6* (2), 93−103.

(18) Gerona, R. R.; Schwartz, J. M.; Pan, J.; Friesen, M. M.; Lin, T.; Woodruff, T. J. Suspect Screening of Maternal Serum to Identify New Environmental Chemical Biomonitoring Targets Using Liquid Chromatography-Quadrupole Time-of-Flight Mass Spectrometry. *J. Exposure Sci. Environ. Epidemiol.* **2018**, *28* (2), 101−108.

(19) 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. *Environ. Pollut.* **2018**, *234*, 297−306.

(20) 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. *Environ. Int.* **2016**, *88*, 269−280.

(21) Wang, Y.; Yu, N.; Zhu, X.; Guo, H.; Jiang, J.; Wang, X.; Shi, W.; Wu, J.; Yu, H.; Wei, S. Suspect and Nontarget Screening of Per- and Polyfluoroalkyl Substances in Wastewater from a Fluorochemical Manufacturing Park. *Environ. Sci. Technol.* **2018**, *52* (19), 11007−11016.

(22) 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.

(23) Tautenhahn, R.; Cho, K.; Uritboonthai, W.; Zhu, Z.; Patti, G. J.; Siuzdak, G. An Accelerated Workflow for Untargeted Metabolomics Using the METLIN Database. *Nat. Biotechnol.* **2012**, *30* (9), 826−828.

(24) Petrick, L. M.; Schiffman, C.; Edmands, W. M. B.; Yano, Y.; Perttula, K.; Whitehead, T.; Metayer, C.; Wheelock, C. E.; Arora, M.; Grigoryan, H.; Carlsson, H.; Dudoit, S.; Rappaport, S. M. Metabolomics of Neonatal Blood Spots Reveal Distinct Phenotypes of Pediatric Acute Lymphoblastic Leukemia and Potential Effects of Early-Life Nutrition. *Cancer Lett.* **2019**, *452*, 71−78.

(25) Grashow, R.; Bessonneau, V.; Gerona, R. R.; Wang, A.; Trowbridge, J.; Lin, T.; Buren, H.; Rudel, R. A.; Morello-Frosch, R. Integrating Exposure Knowledge and Serum Suspect Screening as a New Approach to Biomonitoring: An Application in Firefighters and Office Workers. *Environ. Sci. Technol.* **2020**, *54* (7), 4344−4355.

(26) Wang, A.; Gerona, R. R.; Schwartz, J. M.; Lin, T.; Sirota, M.; Morello-Frosch, R.; Woodruff, T. J. A Suspect Screening Method for Characterizing Multiple Chemical Exposures among a Demographically Diverse Population of Pregnant Women in San Francisco. *Environ. Health Perspect.* **2018**, *126* (7), 077009.

(27) Wang, A.; Schwartz, J. M.; Gerona, R.; Lin, T.; Morello-Frosch, R.; Sirota, M.; Woodruff, T. The Pregnancy Chemisome in Relation to Birth Outcomes and Consumer Product Use: Suspect Screening of Industrial Chemicals. *ISEE Conference Abstracts* 2018, *2018* (1) DOI: 10.1289/isesisee.2018.O03.03.11.

(28) 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.

(29) Wang, Z.; Dewitt, J. C.; Higgins, C. P.; Cousins, I. T. A Never-Ending Story of Per- and Polyfluoroalkyl Substances (PFASs)? *Environ. Sci. Technol.* **2017**, *51*, 2508−2518.

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

(31) 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.

(32) 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.

(33) Wang, M.; Bastiaensen, M.; Kinyua, J.; Petreas, M.; Covaci, A.; Park, J.-S. Exposure Profile Comparison and Screening of Unknown Contaminants Using LC-QToF: A Cat Hyperthyroidism Study. *Dioxin 2019*: Kyoto, Japan, 2019.

(34) Leek, J. T.; Johnson, W. E.; Parker, H. S.; Jaffe, A. E.; Storey, J. D. The Sva Package for Removing Batch Effects and Other Unwanted Variation in High-Throughput Experiments. *Bioinformatics* **2012**, *28* (6), 882−883.

(35) CompMS | MS-DIAL http://prime.psc.riken.jp/compms/msdial/main.html#Demonstration (accessed 2020/5/19).

(36) 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. *J. Mass Spectrom.* **2010**, *45* (7), 703−714.

(37) 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 Res.* **2018**, *46* (D1), D608−D617.

https://dx.doi.org/10.1021/acs.est.0c05984
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

(38) mzCloud - Advanced Mass Spectral Database https://www.mzcloud.org/.
(39) 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 Res.* **2014**, *42* (W1), W94−W99.
(40) Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag Relaunched: Incorporating Strategies beyond in Silico Fragmentation. *J. Cheminf.* **2016**, *8*, 3.
(41) Barupal, D. K.; Fiehn, O. Generating the Blood Exposome Database Using a Comprehensive Text Mining and Database Fusion Approach. *Environ. Health Perspect.* 2019, *127* (9) 097008. DOI: 10.1289/EHP4713.
(42) The Blood Exposome Database https://bloodexposome.org/#/dashboard (accessed 2020/11/17).
(43) Ccanccapa-Cartagena, A.; Pico, Y.; Ortiz, X.; Reiner, E. J. Suspect, Non-Target and Target Screening of Emerging Pollutants Using Data Independent Acquisition: Assessment of a Mediterranean River Basin. *Sci. Total Environ.* **2019**, *687*, 355−368.
(44) Shin, H.; Moschet, C.; Young, T. M.; Bennett, D. H. Measured Concentrations of Consumer Product Chemicals in California House Dust: Implications for Sources, Exposure, and Toxicity Potential. *Indoor Air* **2020**, *30* (1), 60−75.
(45) Yu, N.; Guo, H.; Yang, J.; Jin, L.; Wang, X.; Shi, W.; Zhang, X.; Yu, H.; Wei, S. Non-Target and Suspect Screening of Per- and Polyfluoroalkyl Substances in Airborne Particulate Matter in China. *Environ. Sci. Technol.* **2018**, *52* (15), 8205−8214.
(46) Plassmann, M. M.; Brack, W.; Krauss, M. Extending Analysis of Environmental Pollutants in Human Urine towards Screening for Suspected Compounds. *Journal of Chromatography A* **2015**, *1394*, 18−25.
(47) Zhou, B.; Xiao, J. F.; Ressom, H. W. Prioritization of Putative Metabolite Identifications in LC-MS/MS Experiments Using a Computational Pipeline. *Proteomics* **2013**, *13* (2), 248−260.
(48) Hohrenk, L. L.; Vosough, M.; Schmidt, T. C. Implementation of Chemometric Tools To Improve Data Mining and Prioritization in LC-HRMS for Nontarget Screening of Organic Micropollutants in Complex Water Matrixes. *Anal. Chem.* **2019**, *91* (14), 9213−9220.
(49) 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.
(50) Montazeri, P.; Thomsen, C.; Casas, M.; de Bont, J.; Haug, L. S.; Maitre, L.; Papadopoulou, E.; Sakhi, A. K.; Slama, R.; Saulnier, P. J.; Urquiza, J.; Grazuleviciene, R.; Andrusaityte, S.; McEachan, R.; Wright, J.; Chatzi, L.; Basagaña, X.; Vrijheid, M. Socioeconomic Position and Exposure to Multiple Environmental Chemical Contaminants in Six European Mother-Child Cohorts. *Int. J. Hyg. Environ. Health* **2019**, *222* (5), 864−872.
(51) Djoumbou-Feunang, Y.; Fiamoncini, J.; Gil-De-La-Fuente, A.; Greiner, R.; Manach, C.; Wishart, D. S. BioTransformer: A Comprehensive Computational Tool for Small Molecule Metabolism Prediction and Metabolite Identification. *J. Cheminf.* **2019**, *11* (1), 2.
(52) Uppal, K.; Walker, D. I.; Liu, K.; Li, S.; Go, Y.-M.; Jones, D. P. Computational Metabolomics: A Framework for the Million Metabolome. *Chem. Res. Toxicol.* **2016**, *19*−*29* (12), 1956−1975.
(53) 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 Res.* **2018**, *46* (D1), D608−D617.

https://dx.doi.org/10.1021/acs.est.0c05984
Environ. Sci. Technol. XXXX, XXX, XXX−XXX