UCSF
UC San Francisco Previously Published Works

Title

Exploring applications of non-targeted analysis in the characterization of the prenatal exposome.

Permalink

https://escholarship.org/uc/item/73h7m9nk

Authors

Bland, Garret
Abrahamsson, Dimitri
Wang, Miaomiao
et al.

Publication Date

2024-02-20

DOI

10.1016/j.scitotenv.2023.169458

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Sci Total Environ. Author manuscript; available in PMC 2025 February 20.

Published in final edited form as:

Sci Total Environ. 2024 February 20; 912: 169458. doi:10.1016/j.scitotenv.2023.169458.

Exploring applications of non-targeted analysis in the characterization of the prenatal exposome

Garret D. Bland[1], **Dimitri Abrahamsson**[1,*], **Miaomiao Wang**[2], **Marya Zlatnik**[1], **Rachel Morello-Frosch**[3], **June-Soo Park**[1,2], **Marina Sirota**[4], **Tracey J. Woodruff**[1,*]

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

[2]Department of Toxic Substances Control, California Environmental Protection Agency, Berkeley, California, United States

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

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

Abstract

Capturing the breadth of chemical exposures *in utero* is critical in understanding their long-term health effects for mother and child. We explored methodological adaptations in a Non-Targeted Analysis (NTA) pipeline and evaluated the effects on chemical annotation and discovery for maternal and infant exposure. We focus on lesser-known/underreported chemicals in maternal and umbilical cord serum analyzed with liquid chromatography-quadrupole time-of-flight mass spectrometry (LC-QTOF/MS). The samples were collected from a demographically diverse cohort of 296 maternal-cord pairs (n = 592) recruited in San Francisco Bay area. We developed and evaluated two data processing pipelines, primarily differing by detection frequency cut-off, to extract chemical features from non-targeted analysis (NTA). We annotated the detected chemical features by matching with EPA CompTox Chemicals Dashboard (n=860,000 chemicals) and Human Metabolome Database (n = 3,140 chemicals) and applied a Kendrick Mass Defect filter to detect homologous series. We collected fragmentation spectra (MS/MS) on a subset of serum samples and matched to an experimental MS/MS database within the MS-Dial website and other experimental MS/MS spectra collected from standards in our lab. We annotated ~72% of the features (total features = 32,197, levels 1-4). We confirmed 22 compounds with

*Corresponding authors: Tracey J. Woodruff, tracey.woodruff@ucsf.edu and Dimitri Abrahamsson, dimitri.abrahamsson@gmail.com.

Supporting Information

Supporting information word document: Demographics of cohort, Merging shipments schematic, Instrument parameters for sample measurement, Software MS-Dial parameters for MS/MS experiments, NTA pipeline, Merging shipment schematic, Number of features as a function of Feature Detection Frequency criteria in samples, Distribution of feature RT error in shipment 1 and shipment 2, Distribution of feature mass error in shipment 1 and shipment 2, Comparing Pipeline A and Pipeline B features, Kendrick Mass Defect flow diagram, Mass and RT thresholds, chemical list distribution, MS2 matching examples.

Repository code for the NTA Data Pipeline is available on Zenodo: https://zenodo.org/record/8310287

Supporting information excel document: Dataset Pipeline A, Dataset Pipeline B, Homologous Series, level 1 and 2 features not in blood exposome, spectral alignment for MS2 fragmentation ID for negative and positive ESI.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

analytical standards, tentatively identified 88 compounds with MS/MS spectra, and annotated 4,862 exogenous chemicals with an in-house developed annotation algorithm. We detected 36 chemicals that appear to not have been previously reported in human blood and 9 chemicals that were reported in less than five studies. Our findings underline the importance of NTA in the discovery of lesser-known/unreported chemicals important to characterize human exposures.

Keywords

non-targeted analysis; exposome; prenatal exposure; high resolution mass spectrometry; data pipeline; Kendrick Mass Defect

Introduction

Every day, people are exposed to a broad range of industrial chemicals from multiple sources including consumer products and environmental media.[1,2] Many classes of environmental organic contaminants have been measured both in people and found in multiple environmental media due to their persistence and long-range transport potential in the environment[3,4]. Examples include pesticides[5,6], flame retardants[7], bisphenols[8], perfluorinated compounds [9], siloxanes[3,10], plasticizers[11], pharmaceuticals[12], and parabens[13]. However, the number of chemicals measured in exposure and biomonitoring studies is a small fraction compared to the number of chemicals that are actively used.[14,15] The prenatal period is a critical window where humans are more susceptible to chemical exposures.[16-18] Chemical exposure during this life stage can negatively impact the health of the mother and influence children health outcomes throughout the life course.[19-21] There is, thus, a need to explore the breadth of chemical exposures to improve our understanding of how these exposures affect prenatal health.[22]

High-resolution mass spectrometry (HRMS) and non-targeted analysis (NTA) enable the detection of a broad range of chemicals in a wide range of samples, including dust [23,24], drinking water[25,26], recycled products[27], sediment[28], and serum[29] by mass with greater precision compared to low-resolution mass spectrometers. After data measurement, data processing steps, such as merging batches, method blank subtraction, and identifying multiple adducts, are used to remove less abundant chemical features. NTA considers all chemical features after data processing and annotation of the chemicals features is based on multiple lines of evidence, such as MS/MS matching, full scan MS match to online databases, and confirmation with a standard. NTA can help uncover unknown exogenous chemicals and endogenous metabolites.[30,31]

Many studies have developed NTA pipelines to process raw HRMS data for their specific uses.[32-34] All studies utilized HRMS coupled with chromatography techniques (gas or liquid) which implements another dimension of separation for the chemical features. In these pipelines, there are numerous ways to combine and analyze the raw data output. This includes, but not limited to software choice for extraction of chemical features from raw HRMS data files, pretreatment processing such as batch correction (e.g., normalizing between measurement batches), feature cleanup (e.g., removing features based on specific criteria), adduct specification, and databases for feature annotation. The findings of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

chemicals from NTA depend on a series of decisions in the pipeline, so the analysis of the specific parameters and methods, such as feature cleanup and mass thresholds, is vital to answer the appropriate research question.

In our previous work, we developed a NTA pipeline to identify chemical exposures in pregnant women and their infants as part of the prenatal exposome, which included batch correction, an adduct specification and removal process, and an annotation algorithm to rank potential chemical candidates.[35] As a result, we extracted and processed 685 features with 19 compounds confirmed with analytical standards, 73 compounds annotated with MS/MS spectra matching and 98 compounds matched using an in-house algorithm.

In this study, we developed a computational pipeline and evaluated a set of methodological adaptations to the discovery of underreported and lesser-known chemicals in the prenatal and infant exposome. These methodological adaptations included using a larger chemical database (EPA CompTox Chemicals Dashboard), using an open-source software (MS-Dial) and additional computational approaches for structure elucidation using homologous series (Kendrick mass defect).

The objective of this study is to apply NTA strategies to discover previously underreported or lesser-understood chemicals that could be present in a diverse pregnancy cohort and their infants. The objective comprises the following steps: 1) to investigate a set of methodological adaptations for NTA data processing and chemical feature annotation and identification, 2) apply a feature annotation approach that matches the chemical features to larger online databases and apply Kendrick Mass Defect (KMD) analysis for homologous series extraction, and 3) summarize how these methodological changes influence the breadth of discovery of novel chemical exposures and report lesser-known or previously undetected chemicals in the blood exposome.

Methods

Study Population

The study population was based in San Francisco Bay Area and consisted of 296 pregnant women and their infants in the Chemicals in Our Bodies (CIOB) study at the University of California, San Francisco (Table S1). The CIOB study consisted of English and Spanish-speaking pregnant women, aged 18 to 40, recruited between March 2014 and June 2017. The study population consists of 45% non-Hispanic White women, 27% Hispanic/Latinx women, and 17% Asian. Additional details of the study population and recruitment including demographics and medical record data is described in our previous study.[35]

Sample measurement and analysis overview

In this study, maternal and matching cord blood samples were prepped in two different ways (Figure 1: 1A and 1B). The first sample preparation method involved all blood samples (two shipments of 592 samples, 296 maternal and 296 cord). Each shipment consists of 15 batches, with each batch containing matching 10 maternal and 10 cord samples. This method includes serum extraction after blood clotting and protein precipitation using cold methanol and centrifugation. The extraction and sample measurement details are described

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

in Abrahamsson 2021 et al. The samples were analyzed by acquiring soft ionization full-scan mass spectra (MS) and fragmentation mass spectra(MS/MS) based on collision-induced dissociation (CID) in positive and negative electrospray ionization modes using ultra-high pressure liquid chromatography-quadrupole time-of-flight mass spectrometry (UPLC-QTOF/MS) with a scan rate of four spectra per second and retention time window of 1 minute. The instrumental analysis was performed in a previous study (Abrahamsson 2021 et al.) and more details of instrument function and conditions are described later. The generated raw data files were reprocessed in this study using different software (MS-Dial) and exploring different settings that are detailed later. To investigate lesser abundant features, we employed a second sample preparation method that consists of randomly pooling 10 samples, combining, and concentrating them together. Sample measurement and analysis was performed in this study. Details of the methods are provided later in the manuscript.

There was a total of 296 pairs maternal and matching cord blood samples (n = 592). Samples were frozen at −80 °C prior to analysis and then thawed and centrifuged at 3000 rpm to separate serum from the red platelets. The method for serum extraction is explained in detail in previous studies.[35,36] Briefly, 250 μL of serum was extracted by protein precipitation by cold methanol addition. The final extracts were stored at 4 °C until they were analyzed by ultra-high pressure liquid chromatography-quadrupole time-of-flight mass spectrometry (UPLC-QTOF/MS). For analysis, 10 μL of the extract was injected in the instrument.

We implemented two different sample preparation methods for the MS/MS analysis. In the first method, all serum samples were prepared separately by adding methanol and centrifuging them to remove the precipitated proteins as described previously. The second method comprises of pooling a subset of 10 concentrated samples to detect chemicals that did not have enough signal abundance to be detected in the first method.

For this second method (e.g., referred to as the MTBE method), the extract was prepared as follows: i) 10 maternal serum samples (1 ml each) were randomly selected from the cohort, ii) 1 ml of cold methanol (JT Baker, LCMS grade, 7 °C) was added to each sample to initiate the protein precipitation, iii) after 5 min, the samples were then centrifuged for 5 min at 3000 rpm, iv) the supernatant of each sample was then transferred to a clean tube and 1 ml of methyl tert-butyl ether (MTBE) was added to each tube, the tubes were shaken for 2 mins and the organic phase of each mixture was transferred and pooled to 1 single tube, v) the combined extract was concentrated to near dryness with a gentle stream of nitrogen, and finally vi) the extract was reconstituted to 1 ml with methanol. Using MTBE as an extractant for organic chemicals in blood and serum is well-established and has been used in previously published studies.[37,38] The final extracts were analyzed in duplicate injections in both ESI+ and ESI− using the “Auto MS/MS” method for fragmentation and data collection (method details in the Supporting Information). The generated spectra were then processed with MS-Dial, which is an open-source software that was developed by the University of California, Davis and by RIKEN (Japan), and we matched to the experimental spectra in the databases found in the MS-Dial website. The instrument and software parameters of the processing method are shown in the Supporting Information (Table S2 and Table S3).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Instrument Analysis

We measured samples using an Agilent 1290 UPLC coupled with an Agilent 6550 QTOF system in both positive and negative electrospray ionization modes (ESI+ and ESI-). The full mass spectra (MS) scan was acquired in the range of 100 to 1,000 Da with a resolving power of 40,000 and a mass accuracy of < 5 ppm (Figure 1: 2A). The MS/MS fragmentations ion spectra (MS/MS) were collected in data dependent acquisition mode (DDA) at 10, 20, and 40 eV collision energies and a mass accuracy of 10 ppm. The QTOF system was calibrated before each batch (Figure 1: 2B).

An Agilent 1290 UPLC with an Agilent Zorbax Extend-C18 column (2.1 × 50 mm, 1.8 μm) was used for the chromatographic separation of the analytes. The mobile phase consisted of two gradient solutions: (A) 5 mM ammonium acetate in HPLC water with 0.1% MeOH and (B) 5 mM ammonium acetate in MeOH with 10% HPLC water, which were mixed under the following gradient program: 0 min 10% B and 90% A, 0-15 min gradual increase to 100% B, 16-20 min equilibration at 100% B. All samples were analyzed in duplicate injections and water blanks were analyzed in the beginning of each batch.

Quality Control and Assurance

Every batch that was analyzed by LC-QTOF/MS also included a water blank, a matrix blanks, and a matrix spike. The water blank is LCMS-grade clean water and is processed the same way as the samples. The matrix blank consisted of human AB serum and the QC samples were spiked with 7 poly- and per-fluoroalkyl substances and 6 organophosphate flame retardants at 10 ng/ml. A full assessment of the QA/QC process and the QC samples is described previously in Abrahamsson 2021 et al. For ESI+, we used triphenyl phosphate D15 and DL-cotinine (methyl D3) as internal standards, and for ESI−, we used perfluoro-n-[1,2–13C2] octanoic acid (isotopically labeled M2PFOA). These internal standards were spiked into the sample at 10 ng/mL before sample preparation and were monitored throughout the instrument analysis.

Confirmation with analytical standards

To confirm the presence of the detected compounds, we used a set of chemical mixtures that were prepared by the US Environmental Protection Agency (US EPA) during the EPA's Non-Targeted Analysis Collaborative Trial (ENTACT) (Figure 1: 1C). The mixtures and their preparation are described in Ulrich et al.[39] An identification was considered successful if the observed mass in the sample matched the one in the standard solution within 10 ppm and if the retention times did not differ by more than 0.5 min. In addition, MS/MS spectra were compared for those chemicals that generated more than two fragments in MS/MS mode (Auto MS/MS – Agilent). If a compound had at least 2 MS/MS fragments matching, and the MS mass and RT also matched, then it was considered a successful match.

NTA Workflow

The overall NTA Workflow consists of three major parts: 1) sample and standard measurement, 2) data processing, and 3) chemical annotation approach (Figure 1).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

One important choice for NTA is the software used for alignment and extraction of chemical features from raw HRMS data files. Using open-source software is advantageous because it does not require a commercial license. In this study, we used MS-DIAL developed by UC Davis and by RIKEN (Japan) to extract chemical features (Figure 1: 2A and 2B) and align them across samples which has been used often in previous studies.[40]

NTA Data Processing Pipeline

Our NTA Data Processing pipeline consists of 10 data processing steps developed in Python (Figure 1: 2C, Figure S1). For the HRMS data processing, the following data processing steps were in this order: Align data by MS-Dial, water blank subtraction, detection frequency filter, below detection limit imputation, overlap feature cleanup, batch correction across batches within a shipment of samples, merging the two shipments of 15 batches, batch correction across both shipments, adduct specification (i.e., specifying adduct ion for each feature), and merging positive and negative mode. All samples were measured by LC-QTOF-MS in 15 batches within each of the 2 shipments. We developed two pipelines for the raw HRMS data to process through (Figure S1): Pipeline A for chemical identification and Pipeline B for signal abundance comparison across samples. Pipeline A allows all chemical features to be processed and excludes detection frequency filter (step 3) and below detection limit (DL) imputation (step 4) in the NTA process. In this case, a 'detected' feature within a sample is defined as a measurement value > 5000 signal intensity. Pipeline B restricts the number of chemical features passed through based on step 3's criteria and includes the steps previously mentioned.

Data Alignment: We used MS-Dial (version 4.80) to align each shipment of 15 batches (each batch containing 10 maternal and 10 cord matching samples) in positive and negative mode. Hence, we produced four data frames, each data frame contains a list of features within each shipment and ESI mode (e.g., positive Shipment 1, negative Shipment 1, positive Shipment 2, and negative Shipment 2). The number of features for each data processing step is listed in Figure S1. *Water blank subtraction*: All batches contained two process water blanks. For each batch, we subtracted the average signal peak area of the two process water blanks across all the features for each serum sample. *Detection Frequency Filter (Pipeline B only)*: The next step is to remove chemical features that were infrequently detected in samples. For Pipeline B, we set a removal criterion for any chemical feature that is detectable (> 5,000 signal counts) in less than 40% of all samples measured within each shipment and ESI mode. This threshold percentage of a given feature detected across samples, termed as "percentage threshold", can be adjusted depending on the use case. Therefore, we performed a sensitivity analysis of this criterion from 0 to 100% to see the possible effects downstream and evaluate how much this influences a number of features. *Below DL imputation (Pipeline B only):* The next step is below detection limit imputation which was done similarly in our previous study.[35] We imputed the signal assuming a log-normal distribution of each feature across shipment of samples for values below the detection limit (< 5,000 signal counts). For Pipeline A, select features may be detected in a small number of samples and the signal cannot be imputed. Therefore, for Pipeline A, all abundance values below the detection limit were set to 1,000 signal counts to ensure the values were below detection limit and data processes that occur later in the pipeline, like batch correction, can be properly performed.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Overlapping feature cleanup, batch correction, merging shipments, adduct specification and combining positive and negative step details are provided in Supporting Information.

Chemical Annotation Approach

We applied a chemical annotation approach and a framework proposed by Schymanski 2014 et al for communicating five different levels of confidence of a chemical structure and identification for a given feature. [41] We define the five different classification levels, as we have done in our past work[35,36,42], as follows: Level 1 features have exact chemical structure and isomer and is confirmed with an analytical standard, level 2 features have a proposed chemical structure using MS/MS spectral databases and matching to the experimental spectrum, level 3 are features that have a tentative candidate that was matched by an online database with an *m/z* value (MS1) and a sufficient annotation score in our in-house algorithm as discussed later, level 4 are features that have an assigned chemical formula based on spectral information but not enough evidence to assign tentative chemical structures, and level 5 chemical features are detected for specific mass and RT values but do not have any information to assign a chemical formula, therefore not annotated. After aligning the chemical features across samples (before data cleaning), we matched the tentatively identified compounds from the MS/MS experiments to the MS1 data. If multiple features were matched with one MS2 compound (level 2), they were flagged as a possible match. After alignment, each feature was compared to online databases to match for metabolites and exogenous chemicals based on *m/z*. We matched MS1 data with chemicals from Human Metabolome Database (HMDB) (hmdb.ca) containing 3,140 chemicals followed by EPA's Distributed Structure-Searchable Toxicity (DSSTox) Database (~860,000 reported chemicals, accessed May 19th, 2022) (Figure 1: 3A).[43,44] We prioritized metabolite annotation by screening through the HMDB database first followed by the DSSTox Database. We rank the chemicals based on metadata for each feature using an in-house algorithm (e.g., number of data sources, publications). The top ranked compound is then reported for that feature. If the annotation score is > 0.3, the annotation is categorized as level 3 and < 0.3 as level 4.[35] We chose the cutoff point of 0.3 because when evaluating the algorithm performance, it had a 73% accuracy for compounds with annotations scores ≥ 0.3 and the accuracy dropped significantly for features with an annotation score < 0.3.

To understand the effect of mass and RT error tolerance on chemical annotation, we applied three different conditions mass and RT matching to standards, MS2 spectrum, and online databases (Table S4, Figure S4). Further details and results are in Supporting Information. We chose the following mass and RT error tolerance thresholds: Standard based matching (level 1), 10 ppm mass error and 0.5 min RT error, MS2 spectrum to MS1 feature based matching (level 2), 15 ppm mass error and 0.5 min RT error, and match to an online database (level 3 and 4), 10 ppm mass error.

Kendrick Mass Defect Approach to Identify Homologous Series

We screened for homologous series (a series of chemicals that have the same functional group but differ in the number of repeated chemicals units) called Kendrick Mass Defect (KMD) analysis (Figure 1: 3B). Previous studies have implemented KMD as an analytical approach to extract homologous series in targeted and non-targeted analysis in many types

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of environmental and biological samples, but it has rarely been applied in blood/serum samples.[45-49]

A schematic is shown in the Supporting Information (Figure S5). To confidently annotate homologous series, the following criterion is as follows: 1) for each homologous series, the tentative compounds' chemical structure contains the same base structure and chemical units, 2) the features within a homologous series have similar KMDs in relation to the repeating chemical unit, and 3) mass and RT of features within a homologous series has a fixed positive correlation. If all criteria are met, all level 4 features are increased to level 3. Even though level 3, level 2, and level 1 features did not increase in level confidence, these features are marked separately to show that KMD provides an additional line of evidence for the matched chemical(s). Details of the KMD process are provided in Supporting Information ("KMD Method Details").

Chemical categorization

Once the dataset was annotated, the chemicals that were marked as "exogenous" (e.g., matched with the DSSTox Database but not matched with HMDB[50]), were categorized into groups of chemicals based on chemical lists in EPA CompTox Dashboard (thus, referred to as the dashboard from now on) (Figure 1: 3C). Categorization is further described in Supporting Information. PFAS category is broken down to two different categories (not mutually exclusive): **PFAS** is the list of PFAS compounds in EPA's Toxcast Inventory, and the other labeled as **PFAS Structure** is any chemical structure that contains a fluorinated carbon chain. Annotated level 1, 2 and 3 compounds (level 4 and 5 features do not have assigned chemicals) were also compared to the Blood Exposome generated by Barupel and Fiehn et al. (http://bloodexposome.org), and we report lesser-known chemicals. [51]

Results

NTA Pipeline Analysis Summary

At the start of the HRMS Data Processing, the total number of features ranged from ~60,000 – 86,000 across the maternal and cord samples depending on the shipment and ionization mode (Figure S1). Pipeline A did not filter any features based on detection frequency (step 3). For Pipeline B, we applied a 40% detection frequency to remove the lower abundant features. We evaluated visually the number of features detected by detection frequency filter, which shows an initial steep decrease of number of features between 0 to 40%, and a plateauing trend for > 40% threshold. This trend agrees with the distribution of detected features across samples in Pipeline A (Figure S3A), where most features are detected in < 20% or > 95% of the samples, and we chose 40% as a reasonable point in the trend. After the detection frequency filter, the number of features for Pipeline B decreased 70 – 80% to ~14,000 – 17,400 features. The threshold parameter for % detected in samples strongly influenced the number of features after the detection frequency filter step (Figure S3A). Because of the high number of features, there could be neighboring chemical features that are relatively close by mass and RT. Therefore, we need to remove the low abundant features with a 'feature cleanup' process (further explained in Supporting Information). After the overlapping feature cleanup, the feature count decreased to 38,000 – 47,500 features and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

8,000 – 11,000 features for Pipelines A and B, respectively. This step ensured all features within each shipment and ESI mode were separated by > 10 ppm mass error and > 0.5 min RT to ensure features can be spatially separated for annotation. After the merging of shipment 1 and shipment 2, the number of features decreased to 16,609 (+) and 17,456 (−) for Pipeline A and 3,638 (+) and 4,042 (−) for Pipeline B. After merging features between shipment 1 and 2, the RT difference between shipment 1 and shipment 2 features were <0.5 min and most features (> 99%) have a mass difference < 20 ppm.

After batch correction and adduct specification, the final number of aligned features for chemical annotation and analysis is 32,197 and 7,685 across all maternal and cord samples for Pipeline A and B, respectively. For chemical annotation summary and categorization in the Results section, the Pipeline A dataset is used.

Chemical annotation summary

Out of the total number of aligned features across all maternal and cord samples (N = 32,197), 24 features were confirmed with analytical standards and ENTACT mixture (level 1), 44 features were matched with MS2 spectrum from the MTBE method (level 2), 46 features were matched with MS2 spectrum from methanol precipitation (level 2, total n = 90), 23,031 features were matched with either the HMDB or DSSTox database (level 3 and 4) with 7,607 features having a sufficient annotation score (> 0.3) to be a tentative candidate (level 3), and 15,424 features having an assigned mass formula (level 4). There were 9,052 features that were not annotated (level 5). Because select features have the same mass but different RT, annotations of the same chemical compound can occur. Therefore, this repeated annotation may be a “false positive” annotation, or there also could be many potential stereoisomers for a specific chemical structure which produces multiple peaks. Out of the annotated features (Level 1 through 4), 12,145 features were different, meaning 11,000 features had a repeated annotation (34.2%). There were 3,210 compounds from the DSSTox database matched to features in the dataset, labeled as exogenous, and 767 compounds from the Human Metabolome database, labeled as endogenous, that are detected in any of the 592 samples. All information about chemical annotation including compound name, DTXSID, annotation score, and whether if a compound is considered as exogenous or endogenous is provided in Supporting Information Spreadsheet (Data Pipeline A).

There are 835 chemicals that were present in at least 1 chemical list (Figure S7). The categories with the highest number of chemicals include Pharmaceuticals, PFAS Structure, Cosmetics, Plastics, and Pesticides. Out of the total number of chemicals for each category, the percentage of chemicals that were annotated in the feature space ranged from 0.5% - 6.3%. (e.g., 0.5% for PFAS Structure; 6.3% for Plastics).

KMD Analysis

We found 61 different chemicals in 13 homologous series using KMD (Figure 2). There were 26 level 4 features that increased to level 3 and 32 level 3, 1 level 2, and 3 level 1 features marked as additional evidence for annotation confidence. For CF_2, our focus was on detecting perfluorinated chemicals. One notable series we identified was the perfluoro sulfonic homologous series. One level-1 chemicals, perfluorohexanesulfonic acid (PFHxS),

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and two level 3 perfluorinated compounds with a similar KMD, perfluoroheptanesulfonic acid (PFHpS) and perfluorobutanesulfonic acid (PFBS) (Figure 2A). All three compounds have a CF_2 backbone structure with a sulfonic acid group. Another CF_2 series to note is the perfluorinated carboxylic series. A total of three PFAS chemicals were identified with this series: perfluoroundecanoic acid (level 3), perfluorotridecanoic acid (level 3), and perfluorohexadecanoic acid (level 3) (Figure 2A). For C_2H_4O, the only series that was identified was polyethylene glycol series (PEG) which had eight different compounds (Figure 2B). For CH_2, various types of chemical series with a CH_2 repeating unit, including surfactants, glutamic acids, phosphorylcholines, and biphenyls. Full list of homologous series is in the Supporting Information Spreadsheet (Sheet Name = 'Homologous series').

MS/MS matching

A total of 90 compounds were tentatively identified with MS/MS spectrum from the MTBE method (43 for negative mode, and 47 for positive mode). For both shipments and ESI modes, 51% of the MS2 features were matched to at least 1 feature (mass error < 15 ppm and RT < 0.5 min). Of the 46 compounds that were newly detected in this study across maternal and cord samples, 11 of them were present in the HMDB (endogenous), 25 were in the DSSTOX Dashboard and not in the HMDB, and the remaining 10 compounds were manually inspected in PubChem and Google Scholar (6 endogenous and 4 exogenous) (Supporting Spreadsheet, Dataset Pipeline A).

Feature summary between maternal and cord pairs

The total number of detected features in maternal and cord samples were plotted as a function of percentage of total features (Figure 3A). For both maternal and cord samples, ~30% of total features were detectable on average, with the minimum and maximum being 20% and 4%, respectively. The similarity between maternal and cord pair was also calculated by determining if a specific feature was detected or not for both maternal and cord sample (Figure 3B). Most pairs (n = 280 pairs) were between 70 – 85%, which means 15 – 30% of features were detectable in either maternal or cord. There was a minor number of maternal/cord pairs (n = 11 pairs) that were 90-95% similar. We found 41 level 1 and 2 features that were detected more frequently in cord compared to maternal samples, including 1 long chain acid (9-stearolic acid), an aryl sulfate (2-Hydroxyacetanilide), a neuroactive alkaloid (Norharman), an herbicide (isoproturon), and a pharmaceutical (Azithromycin). There are 30 level 1 and 2 features that were detected more in maternal compared to cord samples, including 3 natural food compounds (caffeic acid, hispanolone, and daidzein), a saturated long-chain fatty acid (palmitic acid), a pharmaceutical (midodrine hydrochloride), a PFAS compound (perfluorononanoic acid), a dibutyl dialkylarylamine (N-Phenyldiethanolamine), a phenolic compound (4-Octylphenol), and an alpha-amino acid (DL-Proline).

Discussion

Pipeline A and Pipeline B comparison

While the way the annotated HRMS datasets were generated from both Pipeline A and Pipeline B were similar, the percentage threshold during the frequency detection filter step

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

provides a significant difference in the number of features. For Pipeline A, a significant portion of level 3 (52%) and level 4 (43%) features were detected only between 0 and 10% of samples (Figure S5A) which reveals more chemical information compared to Pipeline B. However, Pipeline A's limitation is the inability to compare abundance values across samples. After performing the batch correction between both shipments for pipeline A, the samples were likely under corrected because they can still be separated by shipment (Figure S5C). This makes sense because the features that are only present in < 10% of participants may be sparse or only present in one shipment. Therefore, for comparison of signal abundance values across samples, especially for samples performed in batches, removal of low detected features may be a more appropriate approach. Other future non-targeted MS studies that require separate batch measurements need to consider the pros and cons of including a detected feature removal step, depending on the specific use.

Lesser-known or underreported chemicals in blood exposome

We identified 7 level 1 chemicals in our samples that are not present in the Blood Exposome database (Table 1). 4,4'-Sulfonylbis[2-(prop-2-en-1-yl) phenol] is an industrial chemical that includes phenol groups and is primarily used to make resins.[52] N-Ethyl-4-menthane-3-carboxamide is a food additive that is also found in numerous of suspect screening lists and was detected in surface water.[53] Lithocholic acid is a bile acid formed by the dihydroxylation of chenodeoxycholic acid and is a rare example of a toxic endobiotic.[54] CI-1044 is an anti-inflammation agent, but has never been investigated in environmental media or in blood.[55] There were two level 1 chemicals that were lesser-known (i.e., were in less than 5 papers) in the Blood Exposome. N-Phenyldiethanolamine is a tertiary amino compound and is used as an intermediate for dyes.[56] Tridecanedioic acid is a dicarboxylic acid that is used for nylon production.[57]

There were 33 level 2 features that were not present in the Blood Exposome database (20 from the Methanol precipitation method, and 13 from the MTBE method) (Table 1). One compound of concern is 3-Methyl-4,6-di-tert-butylphenol.[58] Although there is sparse literature about this chemical, it is in many suspect lists, including EPA's chemical inventory for ToxCast and EPA High Production Volume List. However, it has potential cardiotoxicity, according to a recent study.[59] Another compound to note is N-Cyclohexyl-2-benzothiazol-amine.[60] This compound is present in the Norman Suspect Screening List but has no information on toxicity and exposure. A recent study also detected this contaminant in stormwater in Northern San Francisco Bay area released by wildfires.[61] The last compound to note is 2-hydroxbenzophenone, which is known to be an endocrine disruptor and a benzophenone derivative. The most probable source of exposure for this contaminant is food packaging and sunscreens.[62,63]

There were 6 compounds, 3 from Methanol method and 3 from MTBE method, that were in the blood exposome database but had limited literature available (< 5 human blood papers). One compound is benzothiazolone which is a benzothiazole and plays an important role in medicinal chemistry.[64] The last compound to note is 6,7-dimethoxy-2,2-dimethyl-2h-1-benzopyran. It is also known as Ageratochromene, an insecticide.[65]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

For the level 3 exogenous chemicals, 3,899 of them were not in the Blood Exposome list. There was 1 bisphenol, 71 cosmetic products, 13 flame retardants, 6 PFAS, 158 PFAS Structure, 22 pharmaceuticals, 74 plastics, and 18 surfactants. Also, we do not know all the sources because chemical manufacturers in the U.S. are not required to report these chemicals, thus there are likely sources not identified for all exogenous chemicals. There are 81 compounds that have a high confidence in level 3 (annotation score > 0.8). One compound to note is 2,4-Bis[(4-dodecylphenyl)azo]resorcinol which is in many lists including the NORMAN Suspect Screening list (SUSDAT) and EPA|Endocrine list (detected in 1 – 4 % of maternal and cord samples).[66] Another compound to note is Butanedioic acid, 2-octyl-, 1,4-bis(tributylstannyl) ester which is an organotin complex that is in SUSDAT list (2 annotated features, 0.3-16% in cord and 0-1.3% in maternal blood).[67] To our knowledge, this is the first study to report the potential presence of these chemicals in human blood samples.

Workflow for chemical feature extraction and annotation

The NTA workflow modifications employed in this study enabled us to annotate more chemical features in both Pipeline A (n = 32,197) and B (n = 7,655) compared to our previous work (n = 685) in the largest sample of matched maternal/cord samples to date. The main reason for this difference is the use of a larger database (e.g., DSSTox Database) which enabled a higher percentage of exogenous feature annotation (level 3). Other modifications that increased the number of confidence level 1, 2, and 3 include KMD analysis which identified homologue series, the implementation of chemical standard mixtures from ENTACT, and implementing multiple sample preparation methods. In addition, it should be noted there are many repeated annotations in the final dataset since many chemical features have the same *m/z* but different RTs. Repeated annotations are not necessarily false positives as many chemicals have multiple stereoisomers and, in many cases, can be resolved in this type of instrumentation appearing as multiple peaks with the same mass but different RTs. One strategy to reduce these repeated annotations is to implement RT predictions for individual compounds.[68] Using similar KMDs is an effective tool to identify and increase the confidence of annotations for groups of chemical features in non-targeted analysis, but in this case, affected a small portion of the total number of features (< 0.5%).

This NTA pipeline provided an expansive feature set for chemical discovery and the number of features is comparable to other recent studies of non-targeted analysis in the blood exposome. Chao et al. also found up to 6,000 chemical features from extracts of 35 human placenta samples and used the DSSTox database for chemical annotation to identify drug and non-drug exogenous features.[29] Hallberg et al., measured human serum samples by LC-MS Orbitrap and extracted a total of 13,740 chemical features with 4,254 of them being detected in only one patient.[69] They applied a multi-tier procedure to prioritize chemical features and used two different databases for chemical matching: Swedish Chemicals Agency Market List and an in-house PFAS list. Other blood exposome studies have applied NTA to discover chemicals but focused on specific chemical classes such as fluorinated surfactants or identifying biomarkers for specific chemical exposure.[70,71]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

For the MS/MS experiments, we obtained 67 compounds at level 2 for the Methanol precipitation method (all samples) and 65 compounds for the MTBE method (10 samples pooled together). Example MS/MS matches are provided in SI (Figure S8). There are 11 compounds in common. Hence, the different sample preparation and selection processes for MS/MS measurement can deliver additional MS2 fragmentation information. Concentrating a group of samples into one aliquot provided new level 2 annotated features that have not been detected in the non-concentrated method, likely because of low abundance, but other chemical features could be lost during the liquid-liquid extraction procedure, depending on the chemical properties (i.e., hydrophobicity). Therefore, the implementation of both sample preparation methods is ideal to maximize chemical feature identification.

Limitations and future considerations

In this NTA study, we expanded the number of chemical features and matched it with a larger database. This greatly improved the number of level 3 and 4 annotated features, but there are some limitations to note.

One limitation is the removal of features based on the detection frequency. While a higher percentage threshold would increase removal of less abundant features, this may eliminate important chemical features for select participants. On the other hand, a lower percentage threshold would allow more features to pass through but may introduce more false positives to the HRMS dataset that could interfere with downstream data processes such as batch correction. To maximize the number of chemical features and enable the comparison of chemical abundance across multiple samples, a combination of pipelines is needed.

Another issue that requires further examination is the high number of repeated annotations for level 3 because features had a similar *m/z* with different RT (48% for Pipeline A and 42% for Pipeline B). Repeated annotations do not always equate to false positives because many chemicals can have multiple stereoisomers that exhibit multiple peaks in LCMS. However, there may be cases where repeated annotations may be false positives that could be further investigated. This observation also shows the limitations of current instrumentation and methods as we are not yet able to distinguish between stereoisomers, structural isomers, and isobars. There is a need to further explore both experimental and *in silico* approaches to disaggregate detected isomers and isobars.[72]

Lastly, many compounds we detected in our study have limited information about their toxicity and possible exposure routes. As many of these compounds are from detected features that have a high annotation confidence (levels 1 and 2) and are detectable in serum, it raises great concern about their potential to cause harm to humans and the environment. One important limitation when confirming chemical structures is that there is a lack of analytical standards available for many detected compounds. Chemical manufacturers in the US are not required to provide analytical standards for the chemicals that they produce and release to the environment except for pesticides.[73] This is an important obstacle for environmental scientists because it prevents the confirmation and quantification of many chemicals that are present in the environment and to which humans are exposed to.[74] Regulators need to implement effective policy for industry manufacturers to provide

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

analytical standards and additional information on toxicity, persistence, and bioaccumulation of all chemicals in commerce and use.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This study was funded by NIH/NIEHS (Grant Nos. R00ES032892, K99ES032892, P30ES030284, UG3OD023272, UH3OD023272, P01ES022841, and R01ES027051) and by the U.S. EPA (Grant Nos. RD83543301 and RD83564301). We would also like to thank Elin Ulrich (U.S. EPA) for her assistance with obtaining the ENTACT mixtures from the U.S. EPA. Finally, we would like to thank the U.S. for supporting this work with providing the ENTACT mixtures to develop new methods in structure elucidation for NTA and enhance our ability to detect new chemicals.

References

(1). Rappaport SM Implications of the Exposome for Exposure Science. Journal of Exposure Science & Environmental Epidemiology 2011 21:1 2010, 21 (1), 5–9. 10.1038/jes.2010.50.

(2). Teeguarden JG; Tan YM; Edwards SW; Leonard JA; Anderson KA; Corley RA; Kile ML; Simonich SM; Stone D; Tanguay RL; Waters KM; Harper SL; Williams DE Completing the Link between Exposure Science and Toxicology for Improved Environmental Health Decision Making: The Aggregate Exposure Pathway Framework. Environ Sci Technol 2016, 50 (9), 4579–4586. 10.1021/ACS.EST.5B05311/ASSET/IMAGES/LARGE/ES-2015-05311M_0003.JPEG. [PubMed: 26759916]

(3). Panagopoulos Abrahamsson D; Warner NA; Jantunen L; Jahnke A; Wong F; MacLeod M Investigating the Presence and Persistence of Volatile Methylsiloxanes in Arctic Sediments. Environ Sci Process Impacts 2020, 22 (4), 908–917. 10.1039/C9EM00455F. [PubMed: 32048673]

(4). Sühring R; Diamond ML; Bernstein S; Adams JK; Schuster JK; Fernie K; Elliott K; Stern G; Jantunen LM Organophosphate Esters in the Canadian Arctic Ocean. Environ Sci Technol 2021, 55 (1), 304–312. 10.1021/ACS.EST.0C04422/ASSET/IMAGES/LARGE/ES0C04422_0005.JPEG. [PubMed: 33305563]

(5). Chiaia-Hernandez AC; Keller A; Wächter D; Steinlin C; Camenzuli L; Hollender J; Krauss M Long-Term Persistence of Pesticides and TPs in Archived Agricultural Soil Samples and Comparison with Pesticide Application. Environ Sci Technol 2017, 51 (18), 10642–10651. 10.1021/ACS.EST.7B02529/SUPPL_FILE/ES7B02529_SI_001.PDF. [PubMed: 28829578]

(6). Kim KH; Kabir E; Jahan SA Exposure to Pesticides and the Associated Human Health Effects. Science of The Total Environment 2017, 575, 525–535. 10.1016/J.SCITOTENV.2016.09.009. [PubMed: 27614863]

(7). Segev O; Kushmaro A; Brenner A Environmental Impact of Flame Retardants (Persistence and Biodegradability). International Journal of Environmental Research and Public Health 2009, Vol. 6, Pages 478-491 2009, 6 (2), 478–491. 10.3390/IJERPH6020478.

(8). Chen D; Kannan K; Tan H; Zheng Z; Feng YL; Wu Y; Widelka M Bisphenol Analogues Other Than BPA: Environmental Occurrence, Human Exposure, and Toxicity - A Review. Environ Sci Technol 2016, 50 (11), 5438–5453. 10.1021/ACS.EST.5B05387/ASSET/IMAGES/LARGE/ES-2015-05387P_0004.JPEG. [PubMed: 27143250]

(9). Houde M; De Silva AO; Muir DCG; Letcher RJ Monitoring of Perfluorinated Compounds in Aquatic Biota: An Updated Review. Environ Sci Technol 2011, 45 (19), 7962–7973. 10.1021/ES104326W/SUPPL_FILE/ES104326W_SI_001.PDF. [PubMed: 21542574]

(10). Panagopoulos D; MacLeod M A Critical Assessment of the Environmental Fate of Linear and Cyclic Volatile Methylsiloxanes Using Multimedia Fugacity Models. Environ Sci Process Impacts 2018, 20 (1), 183–194. 10.1039/C7EM00524E. [PubMed: 29300410]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(11). Ma Y; Xie Z; Lohmann R; Mi W; Gao G Organophosphate Ester Flame Retardants and Plasticizers in Ocean Sediments from the North Pacific to the Arctic Ocean. Environ Sci Technol 2017, 51 (7), 3809–3815. 10.1021/ACS.EST.7B00755/ASSET/IMAGES/LARGE/ES-2017-00755G_0003.JPEG. [PubMed: 28245125]

(12). Nakada N; Kiri K; Shinohara H; Harada A; Kuroda K; Takizawa S; Takada H Evaluation of Pharmaceuticals and Personal Care Products as Water-Soluble Molecular Markers of Sewage. Environ Sci Technol 2008, 42 (17), 6347–6353. 10.1021/ES7030856/SUPPL_FILE/ES7030856-FILE002.PDF. [PubMed: 18800500]

(13). Cassani S; Gramatica P Identification of Potential PBT Behavior of Personal Care Products by Structural Approaches. Sustain Chem Pharm 2015, 1, 19–27. 10.1016/J.SCP.2015.10.002.

(14). Wang Z; Walker GW; Muir DCG; Nagatani-Yoshida K Toward a Global Understanding of Chemical Pollution: A First Comprehensive Analysis of National and Regional Chemical Inventories. Environ Sci Technol 2020, 54 (5), 2575–2584. 10.1021/ACS.EST.9B06379/ASSET/IMAGES/LARGE/ES9B06379_0002.JPEG. [PubMed: 31968937]

(15). Persson L; Carney Almroth BM; Collins CD; Cornell S; de Wit CA; Diamond ML; Fantke P; Hassellöv M; MacLeod M; Ryberg MW; Søgaard Jørgensen P; Villarrubia-Gómez P; Wang Z; Hauschild MZ Outside the Safe Operating Space of the Planetary Boundary for Novel Entities. Environ Sci Technol 2022, 56 (3), 1510–1521. 10.1021/ACS.EST.1C04158/ASSET/IMAGES/LARGE/ES1C04158_0002.JPEG. [PubMed: 35038861]

(16). Rager JE; Bangma J; Carberry C; Chao A; Grossman J; Lu K; Manuck TA; Sobus JR; Szilagyi J; Fry RC Review of the Environmental Prenatal Exposome and Its Relationship to Maternal and Fetal Health. Reproductive Toxicology 2020, 98, 1–12. 10.1016/J.REPROTOX.2020.02.004. [PubMed: 32061676]

(17). Dzwilewski KLC; Schantz SL Prenatal Chemical Exposures and Child Language Development. J Commun Disord 2015, 57, 41–65. 10.1016/J.JCOMDIS.2015.07.002. [PubMed: 26255253]

(18). Padula AM; Monk C; Brennan PA; Borders A; Barrett ES; McEvoy CT; Foss S; Desai P; Alshawabkeh A; Wurth R; Salafia C; Fichorova R; Varshavsky J; Kress A; Woodruff TJ; Morello-Frosch R A Review of Maternal Prenatal Exposures to Environmental Chemicals and Psychosocial Stressors—Implications for Research on Perinatal Outcomes in the ECHO Program. Journal of Perinatology 2019 40:1 2019, 40 (1), 10–24. 10.1038/s41372-019-0510-y.

(19). Bommarito PA; Martin E; Fry RC Effects of Prenatal Exposure to Endocrine Disruptors and Toxic Metals on the Fetal Epigenome. 10.2217/epi-2016-0112 2017, 9 (3), 333–350. 10.2217/EPI-2016-0112.

(20). Antonelli MC; Pallarés ME; Ceccatelli S; Spulber S Long-Term Consequences of Prenatal Stress and Neurotoxicants Exposure on Neurodevelopment. Prog Neurobiol 2017, 155, 21–35. 10.1016/J.PNEUROBIO.2016.05.005. [PubMed: 27236051]

(21). Jedynak P; Maitre L; Guxens M; Gützkow KB; Julvez J; López-Vicente M; Sunyer J; Casas M; Chatzi L; Gražulevičienė R; Kampouri M; McEachan R; Mon-Williams M; Tamayo I; Thomsen C; Urquiza J; Vafeiadi M; Wright J; Basagaña X; Vrijheid M; Philippat C Prenatal Exposure to a Wide Range of Environmental Chemicals and Child Behaviour between 3 and 7 Years of Age – An Exposome-Based Approach in 5 European Cohorts. Science of The Total Environment 2021, 763, 144115. 10.1016/J.SCITOTENV.2020.144115. [PubMed: 33422710]

(22). Buck Louis GM; Yeung E; Sundaram R; Laughon SK; Zhang C The Exposome – Exciting Opportunities for Discoveries in Reproductive and Perinatal Epidemiology. Paediatr Perinat Epidemiol 2013, 27 (3), 229–236. 10.1111/PPE.12040. [PubMed: 23574410]

(23). Moschet C; Anumol T; Lew BM; Bennett DH; Young TM 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. 10.1021/ACS.EST.7B05767/ASSET/IMAGES/LARGE/ES-2017-05767T_0004.JPEG. [PubMed: 29437387]

(24). Petromelidou S; Margaritis D; Nannou C; Keramydas C; Lambropoulou DA HRMS Screening of Organophosphate Flame Retardants and Poly-/Perfluorinated Substances in Dust from Cars and Trucks: Occurrence and Human Exposure Implications. Science of The Total Environment 2022, 848, 157696. 10.1016/J.SCITOTENV.2022.157696. [PubMed: 35908702]

(25). Newton SR; McMahen RL; Sobus JR; Mansouri K; Williams AJ; McEachran AD; Strynar MJ Suspect Screening and Non-Targeted Analysis of Drinking Water Using Point-of-Use Filters.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Environmental Pollution 2018, 234, 297–306. 10.1016/J.ENVPOL.2017.11.033. [PubMed: 29182974]

(26). Park N; Choi Y; Kim D; Kim K; Jeon J Prioritization of Highly Exposable Pharmaceuticals via a Suspect/Non-Target Screening Approach: A Case Study for Yeongsan River, Korea. Science of The Total Environment 2018, 639, 570–579. 10.1016/J.SCITOTENV.2018.05.081. [PubMed: 29800850]

(27). Lowe CN; Phillips KA; Favela KA; Yau AY; Wambaugh JF; Sobus JR; Williams AJ; Pfirrman AJ; Isaacs KK Chemical Characterization of Recycled Consumer Products Using Suspect Screening Analysis. Environ Sci Technol 2021, 55 (16), 11375–11387. 10.1021/ACS.EST.1C01907/ASSET/IMAGES/LARGE/ES1C01907_0005.JPEG. [PubMed: 34347456]

(28). Sibiya I; Poma G; Cuykx M; Covaci A; Daso Adegbenro P; Okonkwo J Targeted and Non-Target Screening of Persistent Organic Pollutants and Organophosphorus Flame Retardants in Leachate and Sediment from Landfill Sites in Gauteng Province, South Africa. Science of The Total Environment 2019, 653, 1231–1239. 10.1016/J.SCITOTENV.2018.10.356. [PubMed: 30759563]

(29). Chao A; Grossman J; Carberry C; Lai Y; Williams AJ; Minucci JM; Thomas Purucker S; Szilagyi J; Lu K; Boggess K; Fry RC; Sobus JR; Rager JE Integrative Exposomic, Transcriptomic, Epigenomic Analyses of Human Placental Samples Links Understudied Chemicals to Preeclampsia. Environ Int 2022, 167, 107385. 10.1016/J.ENVINT.2022.107385. [PubMed: 35952468]

(30). Ng B; Quinete N; Gardinali PR Assessing Accuracy, Precision and Selectivity Using Quality Controls for Non-Targeted Analysis. Science of The Total Environment 2020, 713, 136568. 10.1016/J.SCITOTENV.2020.136568. [PubMed: 31955085]

(31). Ribbenstedt A; Ziarrusta H; Benskin JP Development, Characterization and Comparisons of Targeted and Non-Targeted Metabolomics Methods. PLoS One 2018, 13 (11), e0207082. 10.1371/JOURNAL.PONE.0207082. [PubMed: 30439966]

(32). García-Reyes JF; Hernando MD; Molina-Díaz A; Fernández-Alba AR Comprehensive Screening of Target, Non-Target and Unknown Pesticides in Food by LC-TOF-MS. TrAC Trends in Analytical Chemistry 2007, 26 (8), 828–841. 10.1016/J.TRAC.2007.06.006.

(33). Cervera MI; Portolés T; Pitarch E; Beltrán J; Hernández F Application of Gas Chromatography Time-of-Flight Mass Spectrometry for Target and Non-Target Analysis of Pesticide Residues in Fruits and Vegetables. J Chromatogr A 2012, 1244, 168–177. 10.1016/J.CHROMA.2012.04.063. [PubMed: 22608778]

(34). Schymanski EL; Singer HP; Slobodnik J; Ipolyi IM; Oswald P; Krauss M; Schulze T; Haglund P; Letzel T; Grosse S; Thomaidis NS; Bletsou A; Zwiener C; Ibáñez M; Portolés T; De Boer R; Reid MJ; Onghena M; Kunkel U; Schulz W; Guillon A; Noyon N; Leroy G; Bados P; Bogialli S; Stipaničev D; Rostkowski P; Hollender J Non-Target Screening with High-Resolution Mass Spectrometry: Critical Review Using a Collaborative Trial on Water Analysis. Anal Bioanal Chem 2015, 407 (21), 6237–6255. 10.1007/S00216-015-8681-7/FIGURES/4. [PubMed: 25976391]

(35). Panagopoulos Abrahamsson D; Wang A; Jiang T; Wang M; Siddharth A; Morello-Frosch R; Park JS; Sirota M; Woodruff TJ A Comprehensive Non-Targeted Analysis Study of the Prenatal Exposome. Environ Sci Technol 2021, 55 (15), 10542–10557. 10.1021/acs.est.1c01010. [PubMed: 34260856]

(36). Jiang T; Wang M; Wang A; Abrahamsson D; Kuang W; Morello-Frosch R; Park JS; Woodruff TJ Large-Scale Implementation and Flaw Investigation of Human Serum Suspect Screening Analysis for Industrial Chemicals. J Am Soc Mass Spectrom 2021, 32 (9), 2425–2435. 10.1021/JASMS.1C00135/SUPPL_FILE/JS1C00135_SI_002.XLSX. [PubMed: 34409840]

(37). Pellegrino RM; Di Veroli A; Valeri A; Goracci L; Cruciani G LC/MS Lipid Profiling from Human Serum: A New Method for Global Lipid Extraction. Anal Bioanal Chem 2014, 406 (30), 7937–7948. 10.1007/S00216-014-8255-0/FIGURES/6. [PubMed: 25381612]

(38). Reagen WK; Ellefson ME; Kannan K; Giesy JP Comparison of Extraction and Quantification Methods of Perfluorinated Compounds in Human Plasma, Serum, and Whole Blood. Anal Chim Acta 2008, 628 (2), 214–221. 10.1016/J.ACA.2008.09.029. [PubMed: 18929010]

(39). Ulrich EM; Sobus JR; Grulke CM; Richard AM; Newton SR; Strynar MJ; Mansouri K; Williams AJ EPA's Non-Targeted Analysis Collaborative Trial (ENTACT): Genesis, Design, and Initial

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Findings. Anal Bioanal Chem 2019, 411 (4), 853–866. 10.1007/S00216-018-1435-6/METRICS. [PubMed: 30519961]

(40). Lai Z; Tsugawa H; Wohlgemuth G; Mehta S; Mueller M; Zheng Y; Ogiwara A; Meissen J; Showalter M; Takeuchi K; Kind T; Beal P; Arita M; Fiehn O Identifying Metabolites by Integrating Metabolome Databases with Mass Spectrometry Cheminformatics. Nature Methods 2017 15:1 2017, 15 (1), 53–56. 10.1038/nmeth.4512.

(41). Schymanski EL; Jeon J; Gulde R; Fenner K; Ruff M; Singer HP; Hollender J Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. Environ Sci Technol 2014, 48 (4), 2097–2098. 10.1021/ES5002105/ASSET/IMAGES/LARGE/ES-2014-002105_0001.JPEG. [PubMed: 24476540]

(42). Wang A; Abrahamsson DP; Jiang T; Wang M; Morello-Frosch R; Park JS; Sirota M; Woodruff TJ Suspect Screening, Prioritization, and Confirmation of Environmental Chemicals in Maternal-Newborn Pairs from San Francisco. Environ Sci Technol 2021, 55 (8), 5037–5049. 10.1021/acs.est.0c05984. [PubMed: 33726493]

(43). CompTox Chemicals Dashboard. https://comptox.epa.gov/dashboard/ (accessed 2022-09-13).

(44). Human Metabolome Database: Browsing metabolites. https://hmdb.ca/metabolites?utf8=%E2%9C%93&quantified=1&detected=1&blood=1&filter=true (accessed 2022-09-13).

(45). Cody RB; Fouquet T Paper Spray and Kendrick Mass Defect Analysis of Block and Random Ethylene Oxide/Propylene Oxide Copolymers. Anal Chim Acta 2017, 989, 38–44. 10.1016/J.ACA.2017.08.005. [PubMed: 28915941]

(46). Bugsel B; Zwiener C LC-MS Screening of Poly- and Perfluoroalkyl Substances in Contaminated Soil by Kendrick Mass Analysis. Anal Bioanal Chem 2020, 412 (20), 4797–4805. 10.1007/S00216-019-02358-0/TABLES/4. [PubMed: 31919607]

(47). Müller WH; Verdin A; Kune C; Far J; De Pauw E; Malherbe C; Eppe G Dual-Polarity SALDI FT-ICR MS Imaging and Kendrick Mass Defect Data Filtering for Lipid Analysis. Anal Bioanal Chem 2021, 413 (10), 2821–2830. 10.1007/S00216-020-03020-W/FIGURES/6. [PubMed: 33125540]

(48). Kune C; McCann A; Raphaël LR; Arias AA; Tiquet M; Van Kruining D; Martinez PM; Ongena M; Eppe G; Quinton L; Far J; De Pauw E Rapid Visualization of Chemically Related Compounds Using Kendrick Mass Defect As a Filter in Mass Spectrometry Imaging. Anal Chem 2019, 91 (20), 13112–13118. 10.1021/ACS.ANALCHEM.9B03333/ASSET/IMAGES/LARGE/AC9B03333_0005.JPEG. [PubMed: 31509388]

(49). Verkh Y; Rozman M; Petrovic M A Non-Targeted High-Resolution Mass Spectrometry Data Analysis of Dissolved Organic Matter in Wastewater Treatment. Chemosphere 2018, 200, 397–404. 10.1016/J.CHEMOSPHERE.2018.02.095. [PubMed: 29499520]

(50). Distributed Structure-Searchable Toxicity (DSSTox) Database | US EPA. https://www.epa.gov/chemical-research/distributed-structure-searchable-toxicity-dsstox-database (accessed 2023-01-25).

(51). Barupel D; Fiehn O The Blood Exposome Database. Journal 2019.

(52). Borrel A; Rudel RA Cheminformatics Analysis of Chemicals That Increase Estrogen and Progesterone Synthesis for a Breast Cancer Hazard Assessment. Scientific Reports 2022 12:1 2022, 12 (1), 1–21. 10.1038/s41598-022-24889-w.

(53). Kumar N; Zhao HN; Awoyemi O; Kolodziej EP; Crago J Toxicity Testing of Effluent-Dominated Stream Using Predictive Molecular-Level Toxicity Signatures Based on High-Resolution Mass Spectrometry: A Case Study of the Lubbock Canyon Lake System. Environ Sci Technol 2021, 55 (5), 3070–3080. 10.1021/ACS.EST.0C05546/ASSET/IMAGES/LARGE/ES0C05546_0004.JPEG. [PubMed: 33600148]

(54). Hofmann AF Detoxification of Lithocholic Acid, A Toxic Bile Acid: Relevance to Drug Hepatotoxicity. 10.1081/DMR-200033475 2004, 36 (3–4), 703–722. 10.1081/DMR-200033475.

(55). CI-1044 - Chemical Details. https://comptox.epa.gov/dashboard/chemical/details/DTXSID5047291 (accessed 2023-01-26).

(56). Shen S; Shi H; Sun H; Huo S Mechanistic Study of the Oxidation of N-Phenyldiethanolamine by Bis(Hydrogen Periodato)Argentate(III) Complex Anion. Transition Metal Chemistry 2007, 32 (2), 167–171. 10.1007/S11243-006-0144-8/METRICS.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(57). Cui X; Yan D; Xiao D Synthesis and Characterization of Novel Polyamides Based on Tridecanedioic Acid: Nylons 3 13, 5 13, 6 13, 7 13, 9 13, 10 13, 11 13. E-Polymers 2004, 4 (1). 10.1515/EPOLY.2004.4.1.784/MACHINEREADABLECITATION/RIS.

(58). 3-Methyl-4,6-di-tert-butylphenol - Chemical Details. https://comptox.epa.gov/dashboard/chemical/details/DTXSID1041502 (accessed 2022-12-06).

(59). Krishna S; Berridge B; Kleinstreuer N High-Throughput Screening to Identify Chemical Cardiotoxic Potential. Chem Res Toxicol 2021, 34 (2), 566–583. 10.1021/ACS.CHEMRESTOX.0C00382/ASSET/IMAGES/LARGE/TX0C00382_0009.JPEG. [PubMed: 33346635]

(60). N-Cyclohexyl-2-benzothiazol-amine - Chemical Details. https://comptox.epa.gov/dashboard/chemical/details/DTXSID50891506 (accessed 2022-12-06).

(61). Chang D; Richardot WH; Miller EL; Dodder NG; Sedlak MD; Hoh E; Sutton R Framework for Nontargeted Investigation of Contaminants Released by Wildfires into Stormwater Runoff: Case Study in the Northern San Francisco Bay Area. Integr Environ Assess Manag 2021, 17 (6), 1179–1193. 10.1002/IEAM.4461. [PubMed: 34009690]

(62). Kawaguchi M; Ito R; Honda H; Endo N; Okanouchi N; Saito K; Seto Y; Nakazawa H Simultaneous Analysis of Benzophenone Sunscreen Compounds in Water Sample by Stir Bar Sorptive Extraction with in Situ Derivatization and Thermal Desorption–Gas Chromatography–Mass Spectrometry. J Chromatogr A 2008, 1200 (2), 260–263. 10.1016/J.CHROMA.2008.05.084. [PubMed: 18550077]

(63). Giannetti V; Mariani MB; Mannino P Monitoring of Contaminants in Recycled Paperboard for Food Contact Applications. J Sci Food Agric 2017, 97 (7), 2191–2198. 10.1002/JSFA.8028. [PubMed: 27611788]

(64). Gao X; Liu J; Zuo X; Feng X; Gao Y Recent Advances in Synthesis of Benzothiazole Compounds Related to Green Chemistry. Molecules 2020, Vol. 25, Page 1675 2020, 25 (7), 1675. 10.3390/MOLECULES25071675.

(65). Kong C; Liang W; Hu F; Xu X; Wang P; Jiang Y; Xing B Allelochemicals and Their Transformations in the Ageratum Conyzoides Intercropped Citrus Orchard Soils. Plant Soil 2004, 264 (1–2), 149–157. 10.1023/B:PLSO.0000047759.65133.FA/METRICS.

(66). 2,4-Bis[(4-dodecylphenyl)azo]resorcinol - Chemical Details. https://comptox.epa.gov/dashboard/chemical/details/DTXSID0052334 (accessed 2022-12-18).

(67). Butanedioic acid, 2-octyl-, 1,4-bis(tributylstannyl) ester - Chemical Details. https://comptox.epa.gov/dashboard/chemical/details/DTXSID00886866 (accessed 2022-12-18).

(68). Bonini P; Kind T; Tsugawa H; Barupal DK; Fiehn O Retip: Retention Time Prediction for Compound Annotation in Untargeted Metabolomics. Anal Chem 2020, 92 (11), 7515–7522. 10.1021/ACS.ANALCHEM.9B05765/SUPPL_FILE/AC9B05765_SI_002.PDF. [PubMed: 32390414]

(69). Hallberg I; Plassmann M; Olovsson M; Holte J; Damdimopoulou P; Sjunnesson YCB; Benskin JP; Persson S Suspect and Non-Target Screening of Ovarian Follicular Fluid and Serum – Identification of Anthropogenic Chemicals and Investigation of Their Association to Fertility. Environ Sci Process Impacts 2021, 23 (10), 1578–1588. 10.1039/D1EM00211B. [PubMed: 34581388]

(70). Dagnino S; Strynar MJ; McMahen RL; Lau CS; Ball C; Garantziotis S; Webster TF; McClean MD; Lindstrom AB Identification of Biomarkers of Exposure to FTOHs and PAPs in Humans Using a Targeted and Nontargeted Analysis Approach. Environ Sci Technol 2016, 50 (18), 10216–10225. 10.1021/ACS.EST.6B01170/ASSET/IMAGES/LARGE/ES-2016-011706_0006.JPEG. [PubMed: 27477586]

(71). Rotander A; Kärrman A; Toms LML; Kay M; Mueller JF; Gómez Ramos MJ Novel Fluorinated Surfactants Tentatively Identified in Firefighters Using Liquid Chromatography Quadrupole Time-of-Flight Tandem Mass Spectrometry and a Case-Control Approach. Environ Sci Technol 2015, 49 (4), 2434–2442. 10.1021/ES503653N/SUPPL_FILE/ES503653N_SI_001.PDF. [PubMed: 25611076]

(72). Abrahamsson D; Siddharth A; Young TM; Sirota M; Park JS; Martin JW; Woodruff TJ In Silico Structure Predictions for Non-Targeted Analysis: From Physicochemical Properties to Molecular

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Structures. J Am Soc Mass Spectrom 2022, 33 (7), 1134–1147. 10.1021/JASMS.1C00386/ASSET/IMAGES/LARGE/JS1C00386_0010.JPEG. [PubMed: 35649165]

(73). Wagner WE; Gold SC Legal Obstacles to Toxic Chemical Research. Science (1979) 2022, 375 (6577), 138–141. 10.1126/SCIENCE.ABL4383/SUPPL_FILE/SCIENCE.ABL4383_SM.PDF.

(74). Wagner WE; Gold SC Legal Obstacles to Toxic Chemical Research. Science (1979) 2022, 375 (6577), 138–141. 10.1126/SCIENCE.ABL4383/SUPPL_FILE/SCIENCE.ABL4383_SM.PDF.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Highlights

- We applied NTA to discover underreported chemicals in the prenatal exposome.
- ENTACT mixtures and Kendrick Mass Defect were added to enhance the NTA Workflow.
- Serum from 296 maternal/cord pairs were analyzed via LC-QTOF-MS.
- We detected 45 chemicals in the cohort that have been underreported.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

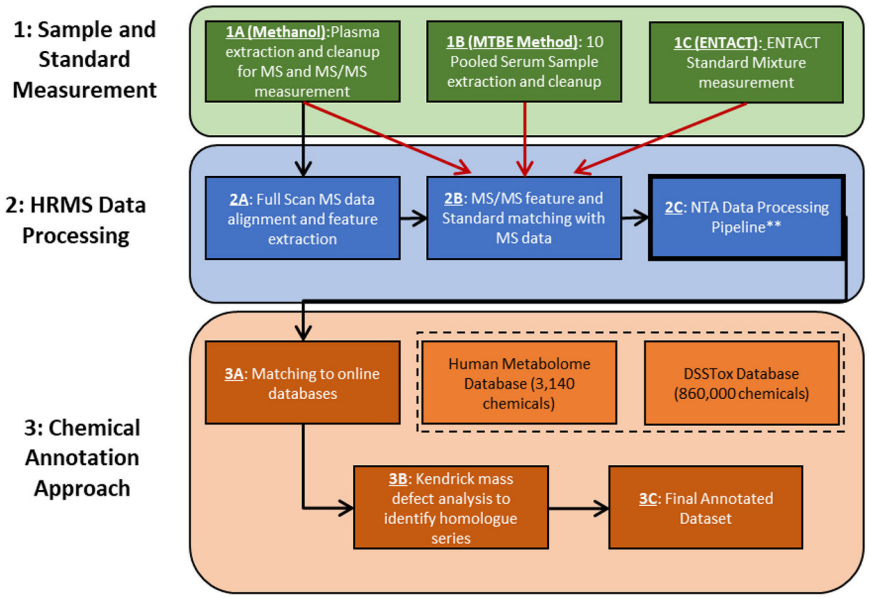

Figure 1:

Overall NTA Workflow schematic for extracting and annotating chemical features: Sample and standard measurement, data processing, and chemical annotation approach. ** Non-Targeted Analysis Data Process is described in Figure S1.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

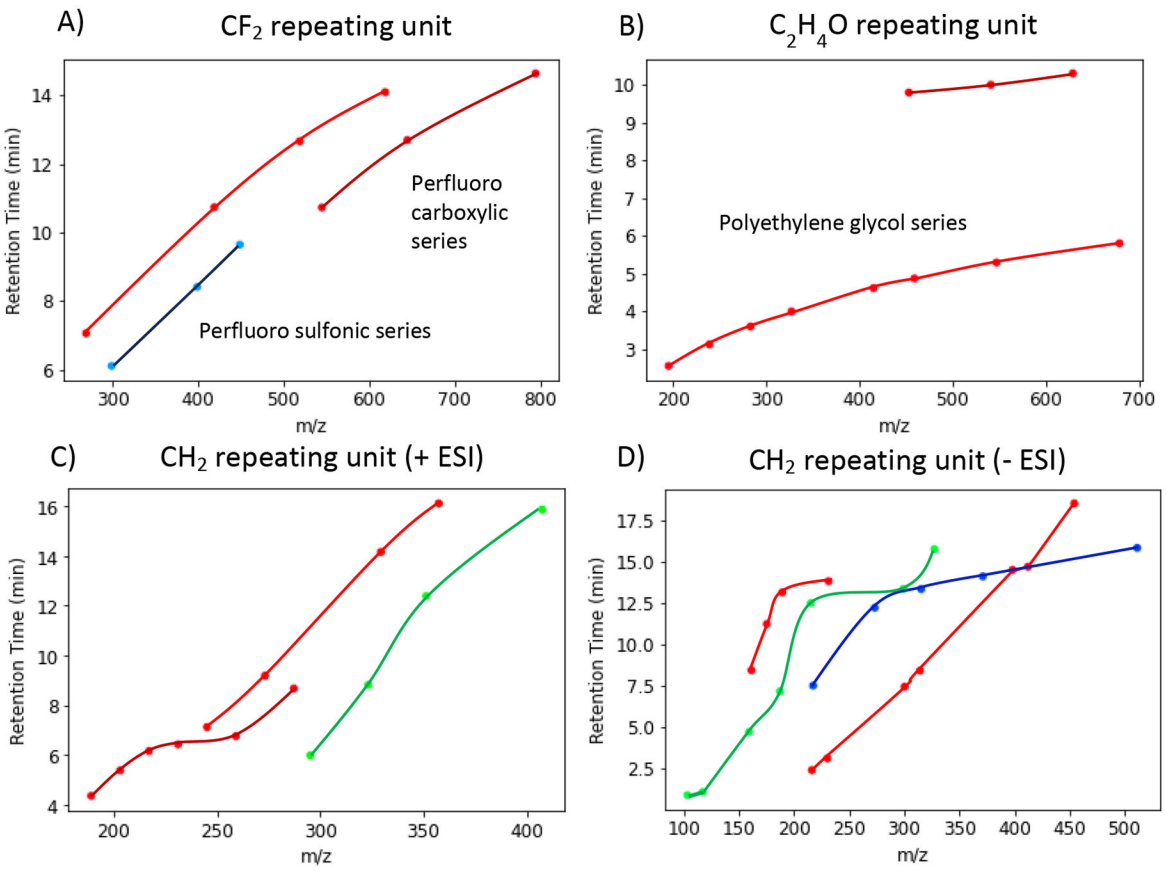

Figure 2:

Feature plot of homologous series candidates as a function of m/z and RT for A) CF_2, B) C_2H_4O, C) CH_2 (positive), and D) CH_2 (negative). Features with connected lines are the identified homologous series. Dashed lines in D) show 2 potential routes of the homologous series.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

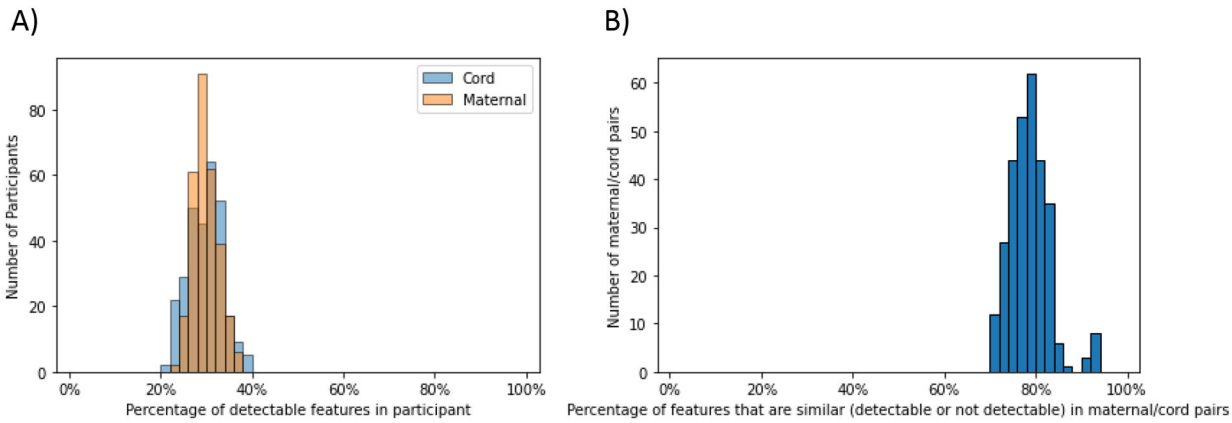

Figure 3:

A) Percentage of total features that are detectable in serum samples and B) the percentage of feature detectability in maternal/cord pairs (e.g., if feature is detectable in maternal and cord pair or not, then feature = 1, if feature is found in either maternal or cord, but not the other, then feature = 0).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1:

Level Annotation breakdown for Total features, unique compounds, exogenous chemicals, and endogenous chemicals. Level 4 annotations only contain molecular formula and level 5 does not contain any assigned compound, so Level 4 and Level 5 classification are not included in Unique Compounds, Exogenous chemicals, and Endogenous Chemicals.

Level Annotation Confidence	Total Features	Unique Compounds	Exogenous Chemicals	Endogenous Chemicals
Level 1	24	21	16	5
Level 2	90	88	67	21
Level 3	7,607	3,977	3,210	767
Level 4	15,424	N/A	N/A	N/A
Level 5	9,052	N/A	N/A	N/A
TOTAL	**32,197**	**4,086**	**3,293**	**793**

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2:

Summary of select under-reported exogenous chemicals detected in our pregnancy cohort (level 1 or 2 confidence). The rest of the chemicals identified as 'Natural Product' is provided in SI.

Chemical name	DTXSID	Source(s)	Confidence level	Chemical formula	Monoisotopic mass	RT (min)	Number of Features	Cord samples detected (%)	Maternal sample detected (%)
Benzothiazolone	DTXSID6061315	Food additive	2	C_7H_5NOS	151.01	6.08	1	48	74
2-Hydroxybenzophenone	DTXSID5047879	Food Packaging and Sunscreens	2	$C_{13}H_{10}O_2$	198.07	7.42	1	3	8
N-Cyclohexyl-2-benzothiazol-amine	DTXSID50891506	Industry	2	$C_{13}H_{16}N_2S$	232.1	11.72	1	1	1
N-Ethyl-4-menthane-3-carboxamide	DTXSID5047039	Food additive	1	$C_{13}H_{25}NO$	211.19	10.33	1	2	6
4,4'-Sulfonylbis[2-(prop-2-en-1-yl)phenol]	DTXSID9047598	Industry	1	$C_{18}H_{18}O_4S$	330.09	9.01	1	21	20
CI-1044	DTXSID5047291	Anti-inflammation agent	1	$C_{23}H_{19}N_5O_2$	397.16	5.47	1	18	38
3-Methyl-4,6-di-tert-butylphenol	DTXSID1041502	Hydrocarbon-based products	2	$C_{15}H_{24}O$	220.18	14.68	1	50	45
tert-Butyric acid	DTXSID0061450	Oral Care Products	2	$C_6H_{12}O_2$	116.08	3.45	1	100	97
Valine betaine	*	Dietary Supplement	2	$C_8H_{17}NO_2$	159.13	3.99	1	34	41
Chrysanthemic Acid	DTXSID30871866	Insecticide	2	$C_{10}H_{16}O_2$	168.12	7.01	1	100	100
(1s,5s)-1,5-dimethyl-3,7-diazabicyclo[3.3.1]nonan-9-one	DTXSID30354301	Organic Compound Synthesis	2	$C_9H_{16}N_2O$	168.12	13.23	1	54	91
2-(1H-indol-3-yl)propanoic acid	*	Organic Compound Synthesis	2	$C_{11}H_{11}NO_2$	189.08	3.78	1	99	99
1H-Indene-3-carboxylic acid, 3a,4,5,6,7,7a-hexahydro-3a,7,7-trimethyl-	DTXSID5022483	Industry	2	$C_{13}H_{20}O_2$	208.15	8.75	1	10	29
Cyclo(Leu-Pro)	DTXSID40218965	Food Additive	2	$C_{11}H_{18}N_2O_2$	210.14	4.81	1	100	100
tetradec-5-ynoic acid	DTXSID1047152	Organic Compound Synthesis	2	$C_{14}H_{24}O_2$	224.18	12.05	1	100	100
Hydroxypestalotin	DTXSID40242028	Fungus Metabolite	2	$C_{11}H_{18}O_5$	230.11	3.52	1	17	16

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Chemical name	DTXSID	Source(s)	Confidence level	Chemical formula	Monoisotopic mass	RT (min)	Number of Features	Cord samples detected (%)	Maternal sample detected (%)
(4S,5Z,6S)-4-(2-methoxy-2-oxoethyl)-5-[2-[(E)-3-phenylprop-2-enoyl]oxyethylidene]-6-[(2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxy-4H-pyran-3-carboxylic acid	DTXSID4041654	Organic Compound Synthesis	2	$C_{15}H_{22}O_2$	234.16	11.96	1	100	98
(1R,7R)-7-ethenyl-1,4a,7-trimethyl-3,4,4b,5,6,9,10,10a-octahydro-2H-phenanthrene-1-carboxylic acid	*	Bile Acid	2	$C_{20}H_{30}O_2$	302.22	13.15	1	100	100
Pyrenophorol	DTXSID30448991	Fungi	2	$C_{16}H_{24}O_6$	312.15	5.91	1	100	100
(4aR,5aS,9R)-9-ethynyl-9a,11b-dimethylhexadecahydrocyclopenta[1,2]phenanthro[8a,9-b]oxirene-3,9-diol	*	Unknown	2	$C_{21}H_{30}O_3$	330.22	8.48	1	100	83
Lithocholic acid	DTXSID70859365	Secondary Bile Acid	1	$C_{24}H_{40}O_3$	376.30	13.74	1	48	68
Azithromycin (Zithromax)	*	Pharmaceutical	2	$C_{38}H_{72}N_2O_{12}$	748.51	9.67	1	3	2
N-Phenyldiethanolamine	DTXSID5021962	hair coloring agent, resin	1	$C_{10}H_{15}NO_2$	181.11	5.6-6.6	2	2-100	5-98
6,7-dimethoxy-2,2-dimethyl-2h-1-benzopyran	DTXSID9060942	insecticide	2	$C_{13}H_{16}O_3$	220.11	10.70	1	2	9
Tridecanedioic acid	DTXSID9021683	nylon production	1	$C_{13}H_{24}O_4$	244.17	6.03	1	71	97
9-stearolic acid	DTXSID90198647	Fatty Acid	2	$C_{18}H_{32}O_2$	318.19	4.89	1	67	38
Tetrahydro-3-pentyl-2H-pyran-4-yl acetate	DTXSID7044556	Cosmetics	2	$C_{12}H_{22}O_3$	214.15	8.55	1	>99	100

* - Chemicals did not contain DTXSID information because they were not registered to the dashboard.