Title

Large-Scale Implementation and Flaw Investigation of Human Serum Suspect Screening Analysis for Industrial Chemicals

Journal

Journal of The American Society for Mass Spectrometry, 32(9)

ISSN

1044-0305

Authors

Jiang, Ting
Wang, Miaomiao
Wang, Aolin
et al.

Publication Date

2021-09-01

DOI

10.1021/jasms.1c00135

Peer reviewed

ACS Partner Journal

Journal of the American Society for Mass Spectrometry

pubs.acs.org/jasms

Research Article

Large-Scale Implementation and Flaw Investigation of Human Serum Suspect Screening Analysis for Industrial Chemicals

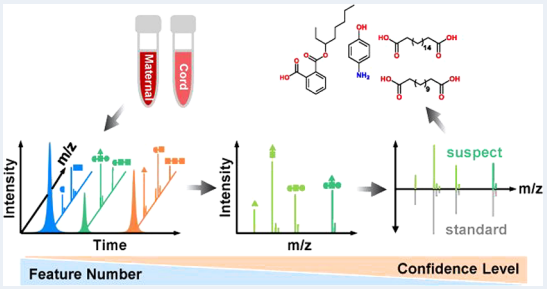
Ting Jiang, Miaomiao Wang, Aolin Wang, Dimitri Abrahamsson, Weixin Kuang, Rachel Morello-Frosch, June-Soo Park,* and Tracey J. Woodruff*

Cite This: *J. Am. Soc. Mass Spectrom.* 2021, 32, 2425−2435

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Non-targeted analysis (NTA), including both suspect screening analysis (SSA) and unknown compound analysis, has gained increasing popularity in various fields for its capability in identifying new compounds of interests. Current major challenges for NTA SSA are that (1) tremendous effort and resources are needed for large-scale identification and confirmation of suspect chemicals and (2) suspect chemicals generally show low matching rates during identification and confirmation processes. To narrow the gap between these challenges and smooth implementation of NTA SSA methodology in the biomonitoring field, we present a thorough SSA workflow for the large-scale screen, identification, and confirmation of industrial chemicals that may pose adverse health effects in pregnant women and newborns. The workflow was established in a study of 30 paired maternal and umbilical cord serum samples collected at delivery in the San Francisco Bay area. By analyzing LC-HRMS and MS/MS data, together with the assistance of a combination of resources including online MS/MS spectra libraries, online in silico fragmentation tools, and the EPA CompTox Chemicals Dashboard, we confirmed the identities of 17 chemicals, among which monoethylhexyl phthalate, 4-nitrophenol, tridecanedioic acid, and octadecanedioic acid are especially interesting due to possible toxicities and their high-volume use in industrial manufacturing. Similar to other previous studies in the SSA field, the suspect compounds show relatively low MS/MS identification (16%) and standard confirmation (8%) rates. Therefore, we also investigated origins of false positive features and unidentifiable suspected features, as well as technical obstacles encountered during the confirmation process, which would promote a better understanding of the flaw of low confirmation rate and encourage gaining more effective tools for tackling this issue in NTA SSA.



INTRODUCTION

Non-targeted analysis (NTA) is a fast-growing approach to uncover emerging environmental chemicals of concern and provide early warnings for industrial regulations and public-health improvement.[1−4] Its essential advantage over traditional targeted analysis lies in the capability to identify "known unknowns" (SSA, suspect screening analysis route) and "unknown unknowns" (unknown compound analysis route) rather than just focusing on "known knowns" (target analysis). For example, in recent NTA studies, new fluoroalkylether compounds from environmental and biological samples have been identified,[5] pesticide residues in food, food packaging, as well as those that end up in the human body have been screened,[6−8] and many more other chemicals can be tracked and monitored from various matrices to better understand their effects on human and environments.[9]

In NTA SSA studies, suspect features (compounds) are typically obtained by screening all the acquired raw features against a database that is composed of the chemicals of interest and are more likely to be valid with richer pieces of experimental evidence. According to the well-known and widely used confidence scale proposed by Schymanski et al.,[10] masses extracted from the raw TIC (total ion chromatograph) only have a level 5 identification confidence as *exact mass of interest*. The suspected features screened out from the database and assigned with isotopic patterns and formulas can be slightly improved in the confidence to level 4 as *unequivocal molecular formula*. Features with experimental evidence, such as MS/MS, or matched computationally predicted spectra, for possible structures but insufficient for further assignment have a level 3 identification confidence as *tentative candidates*. We categorize features that have their mass and MS/MS spectra matched with library spectra as level 2 identification confidence, *probable structures*, and those that have MS/MS spectra and retention time (RT) confirmed by analytical standards are qualified for the highest confidence of level 1 as *confirmed structures*.

Received: April 16, 2021
Revised: August 7, 2021
Accepted: August 9, 2021
Published: August 19, 2021

Some previous SSA studies only reported and discussed features with confidence levels up to level 4,[11] thus the information delivered by those studies was doubtful as the features were not confirmed. Many SSA studies limit their search for chemicals in certain chemical class or usage category (e.g., phthalates only,[12] pharmaceutical drugs only,[13] pesticides only[14]), and almost all of the limited existing SSA studies of human serum chose to apply very stringent prioritization criteria;[15−17] in these cases, they proceeded with only a small number of level 4 suspected features for further confirmation. Indeed, as a major challenge for NTA SSA, tremendous effort and resources are needed for identification and confirmation of suspected chemicals, especially in biomonitoring studies of serum samples, due to the large number of suspect chemicals typically presented and the complication brought by endogenous chemicals/metabolites.[17,18] It is also noteworthy that computationally predicted spectra produced by in silico fragmentation tools are typically used heavily for MS/MS identification in most previous NTA studies, yet many have shown that accuracies of predicted spectra are generally low.[19,20] Therefore, we recommend monitoring inaccurate predictions by moderate manual checking based on knowledge in organic chemistry, but it will further add burden to the amount of time and efforts needed for feature identification.

Another major challenge for NTA SSA is the low matching rates of suspect features (level 4) during identification (level 2) and confirmation (level 1).[2,21] For example, in a representative SSA study in wastewater conducted by Gago-Ferrero et al., with their original SSA method, out of 2524 level 5 features, 150 features were screened out as level 4. However, only 13 features reached the level 2 confidence, and only 7 were finally confirmed with the level 1 identification confidence.[2] These low MS/MS identification and analytical standard confirmation rates are also commonly observed in other SSA studies and are critical challenges to be addressed in this field.[22−24]

Despite the above challenges, the biomonitoring field would still benefit most from NTA SSA studies that (1) screen with a more inclusive database consisting of diverse categories of chemicals, (2) prioritize a larger number of interesting features, and (3) achieve higher feature confirmation rates. Therefore, in order to narrow the gap between the above NTA SSA challenges and beneficial implementation of NTA SSA methodology in the biomonitoring field, here we present a thorough SSA workflow for the large-scale screening, identification, and confirmation of industrial chemicals that may pose adverse health effects in pregnant women and newborns. Instead of focusing on environmental issues and exposomes of populations of different socioeconomic status as in our original study,[27] this work more comprehensively describes the methodological tools used during the MS/MS identification, origins of false positive features, and unidentifiable suspected features observed during the experiment, as well as technical obstacles encountered during the confirmation process with possible solutions proposed. Specifically, to focus research resources on the most interesting features relevant to this study, we prioritized level 4 features based on a tiered approach. During the feature identification using LC-HRMS/MS analysis for prioritized tier 1−3 features, we referred to the "ToxCast" and "source" indexes of U.S. EPA CompTox Chemicals Dashboard[25,26] to greatly minimize the efforts needed on the nonfingerprint type of fragmentation spectra. Finally with an 8% level 4 → level 1 confirmation rate, we confirmed 17 chemicals, including 4 possibly toxic industrial chemicals—monoethylhexyl phthalate, 4-nitrophenol, tridecanedioic acid, and octadecanedioic acid—by comparing to analytical standards. Most importantly, to promote a better understanding toward realistic analytical errors in SSA identification and confirmation, we also investigated typical reasons for identification failures and origins of false positives/negatives and discussed how to improve these rates for future studies. Thus, this work not only serves as one of the emerging studies on large-scale implementation of SSA in biomonitoring field to uncover environmental contaminant industrial chemicals of broad categories but also provides flaw investigation to assist the further advancement of SSA methodology in all related fields.

MATERIALS AND METHODS

Study Samples. The study serum samples were collected from 30 pregnant women who enrolled in the Chemicals in Our Bodies 2 Study (CiOB2) when seeking prenatal and delivery care at the Zuckerberg San Francisco General Hospital and UCSF Mission Bay Medical Center between March 1, 2014 and March 31, 2016.[15,27]

In-House Industrial Chemicals Database.[27] Our suspect database consists of 3518 chemicals of different categories, including 369 environmental organic acids, 207 per- and polyfluoroalkyl substances, 44 flame retardants, and other industrial chemicals in U.S. EPA Chemical Data Reporting (CDR) 2016 database. These chemicals were compiled into the database using Agilent Mass Hunter Personal Compound Database and Library software (PCDL).

Sample Preparation and MS Instrumental Analysis. Serum samples together with blank samples (LCMS grade water) and spiked QC samples (Table S1) were extracted and prepared for instrumental analysis using the protein precipitation technique followed by centrifugation, concentration, and reconstitution (Figure S1). Ten microliters of each sample extract was injected sequentially into an Agilent 1290 UPLC interfaced with an iFunnel 6550 QTOF-MS system for TIC mass spectra acquisition in both negative (ESI−) and positive (ESI+) ionization mode in the 100−1000 m/z mass range. An Agilent Eclipse Plus C18 column (2.1 × 100 mm, 1.8 μm) was used with 5 mM ammonium acetate in water (0.1% methanol) as gradient A and 5 mM ammonium acetate in methanol with 10% water as gradient B. The gradient flow was set to be 0.3 mL/min. More detailed instrumental parameters are shown in Table S2. Two technical replicates were analyzed for each sample. Two blank samples and two quality control samples with two replicates were also analyzed together within one batch.

MS Data Analysis: Feature Extraction, Alignment, Cleaning and Screening (Figure 1). The raw MS TIC data files obtained were processed using Agilent MassHunter Profinder software (version B.08.00) with the raw molecular feature extraction (MFE) and target MFE algorithms to extract compound features recursively across the batch data files. In order to reduce duplicated binning on the same feature, for the initial raw MFE, we employed relaxed binning and alignment parameters of a 0.5 min RT window and a 50 ppm + 2 mDa mass window. If a stringent narrow window is used, there would be too many identical features being treated in different bins, which not only complicates the following data analysis but also distorts the detection frequency of the same features. The masses and retention times of the binned features obtained during raw MFE were then used to perform a recursive targeted MFE referred to as find by ion (FbI). During the subsequent target MFE step, a much narrower window of 5 ppm was applied to reduce false

https://doi.org/10.1021/jasms.1c00135
J. Am. Soc. Mass Spectrom. 2021, 32, 2425−2435

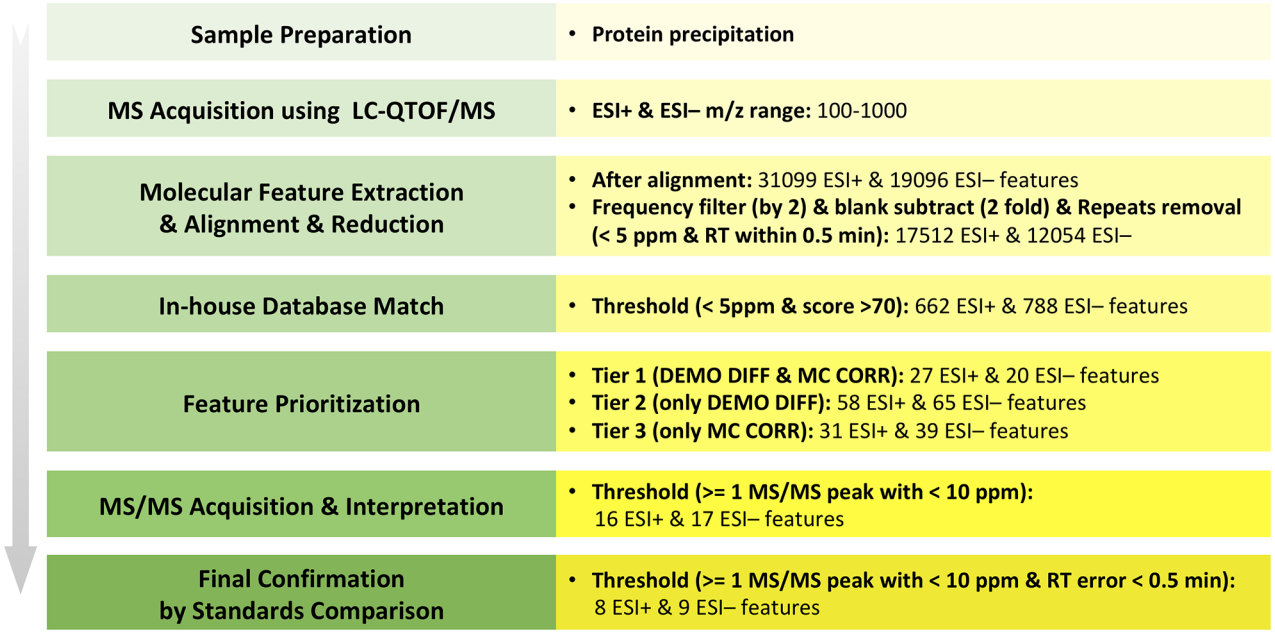


Figure 1. Workflow of the suspect screening analysis including the feature MS/MS identification and confirmation steps for the 30 pairs of pregnant women and umbilical cord serum (*tiers 1−3 also include the 100% detection frequency criterion as mentioned in the Materials and Methods section).

positive hits (Table S3). The extracted features were then aligned (RT correction window = 5% + 0.5 min, mass correction window = 10 ppm + 2 mDa, RT alignment window = 0.3 min) throughout the whole data set using the Agilent Mass Profiler Professional software (MPP, version 12.06.01). Further quality control filtering processes, such as (1) frequency filtering to filter out features that only appear once in the data set as technical replicates were analyzed, (2) blank subtraction to filter out features that do not have significantly higher concentrations in samples than in blanks (2-fold was used as the threshold), and (3) repeat removal to remove the repeated features that had not been successfully aligned together by the software (threshold: mass error <5 ppm, retention time difference <0.5 min), were applied. The resulting features were then screened against our in-house curated database for human exposure studies based on the spectral information (mass, isotopic abundance and patterns, as well as adduct ions) using MPP (mass tolerance window = 5 ppm + 0.01 mDa, score >70). Matched suspected chemicals were thus assigned a molecular formula and level 4 identification confidence.

Suspect Prioritization for MS/MS Analysis. We prioritized the suspect (database screened) chemicals based on three criteria that were the focus of the parent study:[27]

(1) Universal presence in general population: 100% detection frequency (DF), that is, DF ≥ 117 (total sample number = 30 pairs × duplicate injections = 120 samples), and feature peak areas rank among the top 50% of all features
(2) Significant demographic differences in peak areas of cord or maternal samples ($p < 0.05$, demographic details can be found in Table S4)
(3) Maternal and cord serum sample correlation:
   3a. half of cord samples have peak areas that are not lower than twice the median peak area of maternal samples
   3b. half of maternal samples have peak areas that are not lower than twice the median peak area of cord samples
   3c. the Spearman correlation between cord and maternal peak areas ≥0.5 ($p < 0.05$)

Based on these criteria, we obtained three tiers of prioritized chemicals. Features that met all three criteria were assigned to tier 1; features that met the common presence and demographic difference criteria were assigned to tier 2; features that met the common presence and maternal−cord correlation criteria were assigned to tier 3. Features in tiers 1−3 were prioritized for MS/MS spectra analysis for further identification (details of data prioritization strategies are reported elsewhere[27]).

Targeted MS/MS Spectra Acquisition (Table S5). The 30 pairs of maternal−cord serum samples were ranked according to peak areas of the prioritized features in descending order. For each feature, 10 μL extracts of the top ranked samples were used for individual injection into the same Agilent LC-QTOF-MS system used for MS acquisition to acquire the target MS/MS spectra. The targeted MS/MS acquisition method was built based on the list of target precursor ion masses and respective retention time ranges (feature retention time ±1 min) of the prioritized features. For each MS/MS spectra, different collision energies of 0, 10, 20, and 40 eV were applied.

MS/MS Spectra Interpretation. Agilent MassHunter Qualitative Analysis software (version B.08.00) was used to extract and review the MS/MS spectra of prioritized chemicals at different collision energies. For each chemical feature, online MS/MS libraries, mainly MassBank of North America (MoNA),[28] MassBank Europe,[29] HMDB,[30] and mzCloud[31] were first used to search for any existing MS/MS spectrum uploaded by other researchers to compare with the MS/MS spectra acquired. If existing MS/MS spectra were found and matched with acquired MS/MS spectra, the corresponding features were assigned with level 2 identification confidence. Otherwise, online in silico fragmentation tools, CFM-ID (competitive fragmentation modeling for metabolite identification)[32] and MetFrag,[33] were used to predict the MS/MS spectra and compare with the acquired ones. Empirical checking based on organic chemistry theories and chemical reactivity was

https://doi.org/10.1021/jasms.1c00135
J. Am. Soc. Mass Spectrom. 2021, 32, 2425−2435

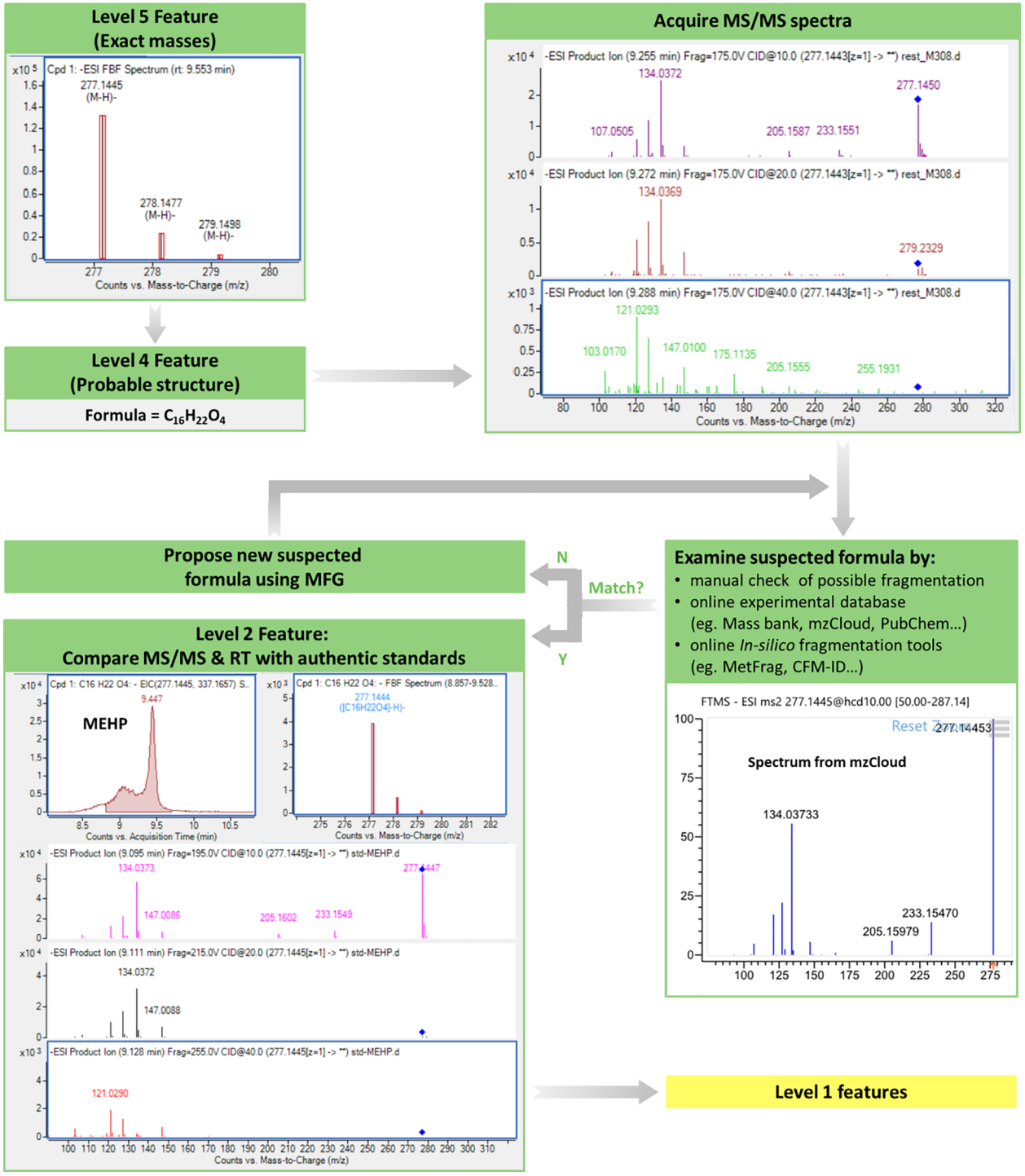


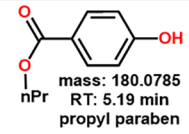
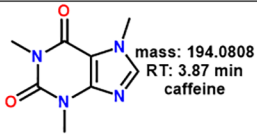
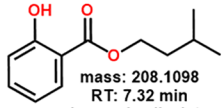
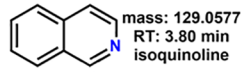
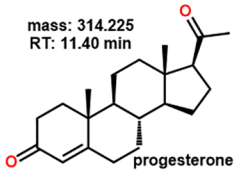
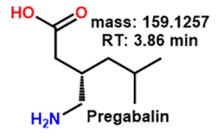
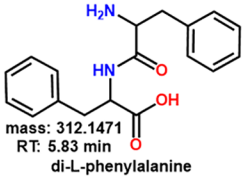
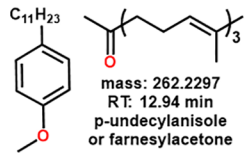
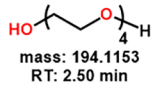
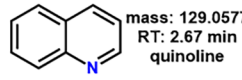
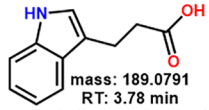
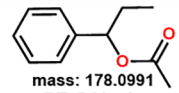
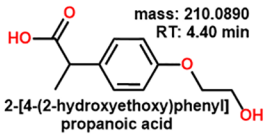
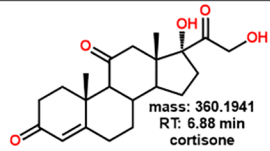
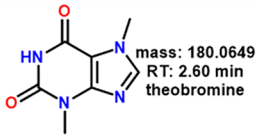
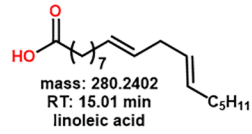
Figure 2. Flowchart that shows the processes of elevating the confidence levels of suspected features by examining MS/MS spectra (to level 2) and comparison with analytical standards spectra (to level 1) using MEHP (monoethylhexyl phthalate) as an example.

also performed during MS/MS interpretation, especially when comparing experimental MS/MS spectra with predicted ones from in silico fragmentation to monitor inaccurate predictions.

In this work, a MS/MS spectrum with at least two fragmentation peaks (mass error <10 ppm) congruent to the spectrum from the online MS/MS libraries was considered a match and the confidence level of the corresponding feature was elevated from level 4 to level 2. If there was only one fragmentation peak match between a MS/MS spectrum and the spectrum from online library, or if there was a match of at least one fragmentation peak between the MS/MS spectrum and in silico fragmentation prediction or empirical check, the confidence level of the corresponding feature was set as level 3. If the acquired MS/MS spectra of a suspected chemical was not a match, the confidence level of the corresponding feature was lowered to level 5. After new appropriate formulas and structures were proposed to these features by the molecular formula generation tool, the MS/MS spectra were examined again for the newly proposed structure, as shown in the flowchart (Figure 2). When proposing new structures, in order to benefit the most

https://doi.org/10.1021/jasms.1c00135
J. Am. Soc. Mass Spectrom. 2021, 32, 2425−2435

Table 1. Details of the 16 ESI+ Features That We Evaluated via MS/MS Identification and Standard Confirmation Steps for the 30 Pairs of Pregnant Women and Umbilic Cord Serum Samples

Suspected Features	Tier	Standard Availability	Confidence Level	Suspected Features	Tier	Standard Availability	Confidence Level
 mass: 180.0785 RT: 5.19 min propyl paraben	1	Y	3 ↓ Rejected	 mass: 194.0808 RT: 3.87 min caffeine	2	Y	2 ↓ 1
 mass: 208.1098 RT: 7.32 min Isoamyl salicylate	1	Y	3 ↓ Rejected	 mass: 129.0577 RT: 3.80 min isoquinoline	2	Y	2 ↓ Rejected
 mass: 314.225 RT: 11.40 min progesterone	1	Y	2 ↓ 1	 mass: 159.1257 RT: 3.86 min Pregabalin	2	N	2
 mass: 312.1471 RT: 5.83 min di-L-phenylalanine	1	Y	2 ↓ 1	 mass: 262.2297 RT: 12.94 min p-undecylanisole or farnesylacetone	2	Y for farnesyl-acetone; N for p-undecylanisole	3 (p-undecyl anisole)
 mass: 194.1153 RT: 2.50 min tetraethylene glycol	1	Y	2 ↓ 1	 mass: 129.0577 RT: 2.67 min quinoline	3	Y	2 ↓ Rejected
 mass: 189.0791 RT: 3.78 min 3-Indolepropionic acid	2	Y	2 ↓ 1	 mass: 178.0991 RT: 5.88 min 1-phenylpropyl acetate	3	N	3
 mass: 210.0890 RT: 4.40 min 2-[4-(2-hydroxyethoxy)phenyl] propanoic acid	2	N	3	 mass: 360.1941 RT: 6.88 min cortisone	3	Y	2 ↓ 1
 mass: 180.0649 RT: 2.60 min theobromine	2	Y	2 ↓ 1	 mass: 280.2402 RT: 15.01 min linoleic acid	3	Y	3 ↓ Rejected

from the collected MS/MS data, we did not just limit our search to the in-house database but instead looked for all possible isomers that were featured with a nonzero "ToxCast" index (toxicity forecaster, for reference, 48% chemicals of our in-house industrial chemical database are in the nonzero ToxCast collection) or a relatively high (>10) "sources" index (for reference, 93.7% chemicals of our in-house industrial chemical database are in the "sources > 10" collection) in U.S. EPA CompTox Chemicals Dashboard.[25,26] This way we were able to inclusively screen for a large number of possibly toxic chemicals with minimum efforts.

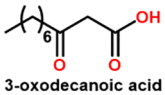
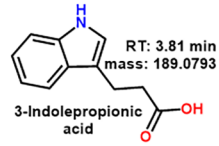
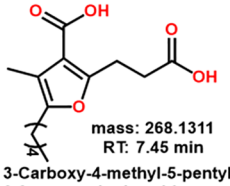
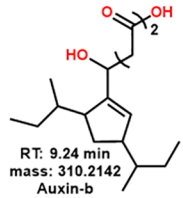
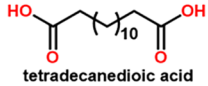
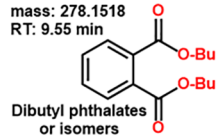
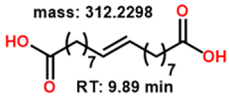
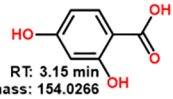
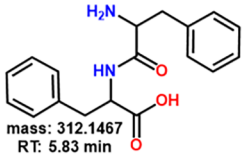
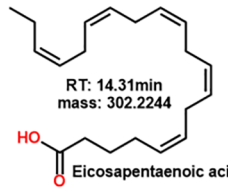
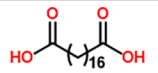
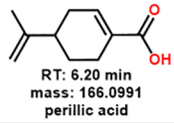
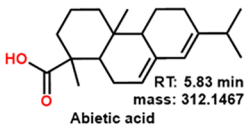
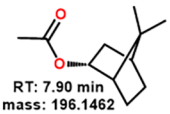
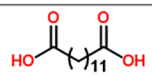
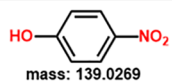
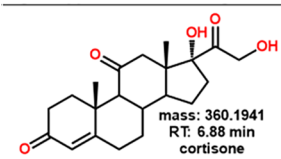
Feature Confirmation with Analytical Standards. Analytical standards for 25 level 2-3 features were purchased (see Table 1 and Table 2 for purchased standards) from Sigma-Aldrich and Thermo Fisher Scientific (standards for the rest of the eight features are not available for purchase, Table 1 and Table 2) and were dissolved in LCMS grade methanol/water mixture (ratio depending on the solubility of the chemicals) with concentrations on the nanogram to microgram/milliliter scale. A 10 μL solution of each standard was injected into the instrument for both the MS and MS/MS spectra. The spectra were then compared to those acquired for the suspected chemicals. The highest identification confidence, level 1, was assigned to matched suspected chemicals (Figure 2).

RESULTS AND DISCUSSION

Feature Extraction and Prioritization for MS/MS Identification. Figure 1 shows the workflow of the SSA, including feature MS/MS identification and confirmation steps. From the acquired raw TIC of the 30 pairs of maternal−cord serum samples, molecular feature extraction and alignment

https://doi.org/10.1021/jasms.1c00135
J. Am. Soc. Mass Spectrom. 2021, 32, 2425−2435

Table 2. Details of the 17 ESI− Features That We Evaluated via MS/MS Identification and Standard Confirmation Steps for the 30 Pairs of Pregnant Women and Umbilic Cord Serum Samples

Suspected Features	Tier	Standard Availability	Confidence Level	Suspected Features	Tier	Standard Availability	Confidence Level
 3-oxodecanoic acid mass: 186.1255 RT: 6.71	1	N	2	 RT: 3.81 min mass: 189.0793 3-Indolepropionic acid	2	Y	2 ↓ 1
 mass: 268.1311 RT: 7.45 min 3-Carboxy-4-methyl-5-pentyl-2-furanpropionic acid	1	N	2	 RT: 9.24 min mass: 310.2142 Auxin-b	2	N	3
 tetradecanedioic acid mass: 258.1825 RT: 6.60 min	1	Y	2 ↓ 1	 mass: 278.1518 RT: 9.55 min Dibutyl phthalates or isomers	2	Y	2 ↓ 1 (MEHP)
 mass: 312.2298 RT: 9.89 min (9Z)-9-octadecenedioic acid	1	N	2	 RT: 3.15 min mass: 154.0266 2,4-Dihydroxybenzoic acid	2	Y	2 ↓ Rejected
 mass: 312.1467 RT: 5.83 min di-L-phenylalanine	1	Y	2 ↓ 1	 RT: 14.31min mass: 302.2244 Eicosapentaenoic acid	3	Y	2 ↓ 1
 Octadecanedioic acid mass: 314.2456 RT: 10.90 min	1	Y	2 ↓ 1	 RT: 6.20 min mass: 166.0991 perillic acid	3	Y	2 ↓ Rejected
 RT: 5.83 min mass: 312.1467 Abietic acid	1	Y	3 ↓ Rejected	 RT: 7.90 min mass: 196.1462 bornyl acetate	3	Y	3 ↓ Rejected
 tridecanedioic acid mass: 244.1674 RT: 6.01 min	2	Y	2 ↓ 1	 mass: 139.0269 RT: 4.9 min 4-Nitrophenol	3	Y	2 ↓ 1
 mass: 360.1941 RT: 6.88 min cortisone	2	Y	2 ↓ 1				

using Agilent Profinder and MPP software yield a total of 31099 ESI+ features and 19096 ESI− features. Considering that technical replicates of sample extracts were used to collect the spectra, a frequency filter was applied to remove 287 ESI+ features and 546 ESI− features that only appear once throughout the whole data set, as they were likely to be artifacts. Moreover, 10638 (34%) ESI+ features and 5710 (30%) ESI− features were removed as repeated features that were likely to result from incomplete binning/alignment rooted in the software algorithm. Furthermore, 2662 (9%) ESI+ features and 784 (4%) ESI− features had peak areas (<2-fold) similar to those in the blank samples and were subtracted as they are likely to be introduced during the experimental process, i.e., sample preparation and injection. This data cleaning and reduction process helps remove a large number of background and artifact features, and as a result, 17512 (56%) ESI+ and 12054 (63%)

https://doi.org/10.1021/jasms.1c00135
J. Am. Soc. Mass Spectrom. 2021, 32, 2425−2435

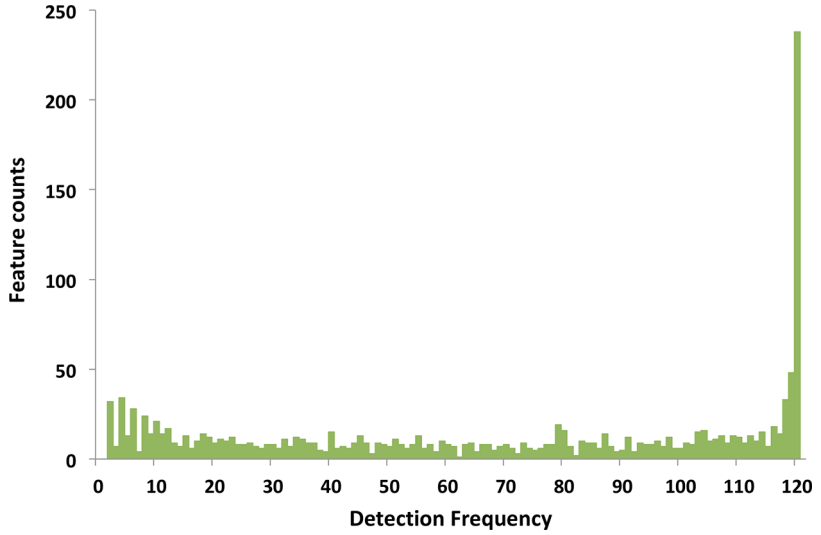


Figure 3. Histogram of DF (detection frequencies) of all level 4 suspected chemicals in the 30 pairs of maternal and cord samples. *DF is a statistical value commonly used in the biomonitoring field to characterize the extent of exposure to a specific chemical by the investigated population.

ESI− level 5 features remained for further SSA and unknown identification analysis. For SSA, screening against our in-house chemical database yields 662 ESI+ and 788 ESI− hits (level 4 features), consisting only ∼5% of the total cleaned-up features. The majority of the remaining 16850 ESI+ and 11266 ESI− unknown features are suspected to be likely related to endogenous and metabolite compounds, as well as other exogenous chemicals that are not covered by our in-house chemical database. Instrumental background, in-source fragmentation, and improper peak grouping in the feature extraction process may also compose a portion of the unknown features. These unknown features can be analyzed via the NTA unknown identification route and can be screened against new databases of interest for additional SSA analysis. This capability of retrospective analysis without the need of reinjecting samples is another advantage of the NTA technique.

Figure 3 shows the histogram of DF of all level 4 suspected chemicals in the 30 pairs of maternal and cord serums (a total of 120 samples when counting technical replicates). Feature counts are much higher at the high detection frequency end of the diagram, consistent with the matrix similarities, as well as the expectation that the general population is exposed to similar chemicals. There are also low-count (count number <40) features with fewer DF (e.g., DF between 3 and 116). Features with low DF are likely to be contributed by individual differences and unique experiences, which can include differences in diet, personal care product use, and residential location. Industrial chemicals exposed to the general population are of special interest our study; therefore, universal presence in general population (the threshold of DF = 100% was in this work) was set as the top criterion that all prior features needed to meet. A total of 328 features had 100% DF (DF ≥ 117 out of the 120 samples; also see the Materials and Methods section) in the experimental maternal−cord pairs (Figure 3). The median peak areas across all 120 samples for these 328 features also rank the top 50th percentile when sorting the median peak areas of all features in descending order. Among these 328 features, 27 ESI+ and 20 ESI− features had both demographic differences and maternal−cord correlations and were assigned as tier 1 features; 58 ESI+ and 65 ESI− features had maternal−cord correlations but no demographic differences and were assigned as tier 2 features; 31 ESI+ and 39 ESI− features had only demographic differences without strong maternal−cord correlations and were assigned as tier 3 features. Based on these criteria, we prioritized a total of 240 features, 116 ESI+ and 124 ESI−, in tiers 1−3 for further MS/MS identification (the full list is shown in Table S6).

MS/MS Identification and Level 2 and 3 Features. We found a total of 10 ESI+ and 14 ESI− features identified with level 2 confidence, plus 6 ESI+ and 3 ESI− features with level 3 confidence from the three tiers of prioritized chemicals (Table 1 and Table 2). Therefore, the level 2 and 3 identification rate is 16/116 = 13.8% for ESI+ suspected features and 17/124 = 13.7% for ESI− suspected features.

In this study, there were several factors that lowered the rate of level 2 and 3 identification:

(1) Fragmentation patterns of ∼38% of the features did not match to the suspect structure and only correspond to the loss of −COOH, −OH, or $-CH_3$, etc., which is not specific fingerprint type evidence and thus limits our ability to assign the features to specific chemical structures. Further, searching all of the possible structures on EPA CompTox Chemicals Dashboard[25,26] did not yield any structure associated with a nonzero "ToxCast" index or a high (>10) "sources" index. It is also possible that these features were endogenous compounds and were false positively identified as level 4 hits during database matching. For features with specific isotopic patterns, such as those that contain halogens, they are more likely to be true positive exogenous compounds and exhibit more distinct fragmentation signature of halogen loss.

(2) Signal/noise ratios (S/N) of ∼28% of acquired MS/MS spectra are too low for interpretation, probably due to low precursor (parent mass) intensities or the difficulties of fragmenting some inert precursors and/or matrix interferences.

(3) Precursor ions for ∼10% features cannot be found during MS/MS experiments, possibly due to the wrong grouping of mass peaks by the molecular feature extraction algorithm during feature extraction process. They could also be interferences from instrument parts, such as analytical columns, sources, and detectors, which only appeared during MS injections and were absent in MS/MS injections, as they were acquired at different times.

https://doi.org/10.1021/jasms.1c00135
J. Am. Soc. Mass Spectrom. 2021, 32, 2425−2435

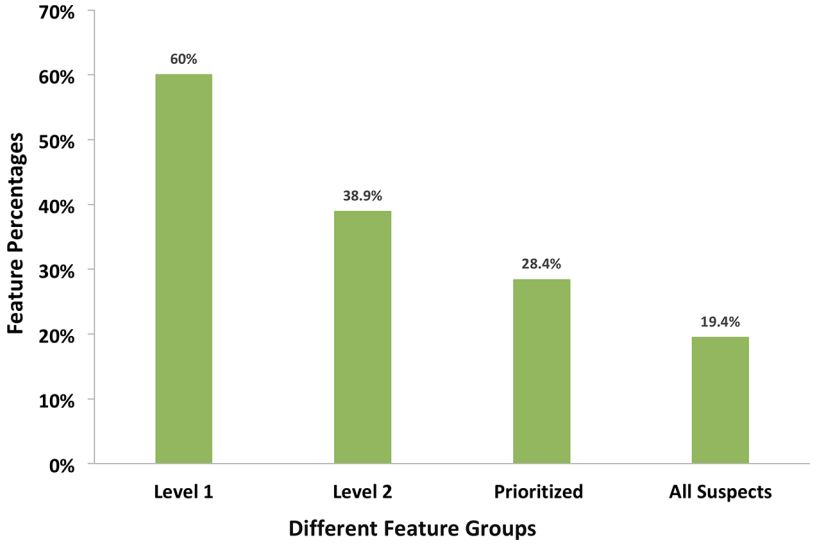


Figure 4. Percentages of features that can be found in both ESI+ and ESI− modes in level 1 features, level 2 features, prioritized features, and all suspect features.

(4) Masses of about ∼3% features are low, and their fragments (product ions) are not in the detection range (m/z = 100−1000).

Analytical Standards Comparison and Level 1 Features. Among 16 ESI+ and 17 ESI− level 2 and 3 features, analytical standards of 4 ESI+ and 4 ESI− features (Table 1 and Table 2) were not commercially available at the time of the experiment, thus we only conducted confirmation on 12 ESI+ and 13 ESI− features. By comparison, 8 ESI+ and 9 ESI− level 2 features were matched with their analytical standards and were confirmed to be level 1 features (Table 1 and Table 2). All audited level 3 (3 ESI+ and 2 ESI−) features failed to match with their analytical standards, suggesting that it may be more efficient to focus on only confirming level 2 features. The remaining four level 3 features (3 ESI+ and 1 ESI−) cannot be examined due to the unavailability of analytical standards and thus remained at level 3. For the unmatched ones, either the retention times (e.g., quinoline) and/or the MS/MS spectra (e.g., abietic acid) are different from those of the standard compounds, or the standard compounds (e.g., bornyl acetate) do not ionize much in the respective ionization modes. Hence the level 2 and 3 → level 1 confirmation rate is 8/16 = 50% for the ESI+ level 2 and 3 features and 9/17 = 52.9% for ESI− level 2 and 3 features, and if we only consider level 2 features, the level 2 → level 1 confirmation rates are considerable higher with 8/10 = 80% in ESI+ mode and 9/14 = 64.3% in ESI− mode.

False Positives Found by MS/MS Identification and Analytical Standards Comparison. Aided by the MS/MS identification and standard confirmation process, we found about 45% of false positive (FP) hits which were corrected or dropped. Most of these hits had been assigned to level 4 identification confidence, and we identified them as being false positive at the MS/MS identification step. In many cases of these FP hits, the acquired MS/MS spectra turned out to match with endogenous chemicals. For example, the MS/MS spectra collected for $C_{18}H_{20}N_2O_3$ at 5.8 min, albeit suspected to be *N*-(2-ethoxyphenyl)-*N*′-(2-ethylphenyl)-ethanediamide, instead matched with di-L-phenylalanine, a metabolite of essential amino acid phenylalanine (Figure S3.7 and S3.8), and the MS/MS spectra collected for $C_{30}H_{56}O_4$ at 15.8 min, suspected to be ditridecyl (2*Z*)-2-butenedioate, were likely to match with a fatty acid. In other cases, the acquired MS/MS spectra match with different high-use industrial chemicals. For example, those collected for $C_{18}H_{30}O$ at 12.9 min, suspected to be 4-*s*-butyl-2,6-di-*tert*-butylphenol, 2,4,6-tris(*tert*-butyl)phenol, or 4-dodecylphenol, match better with *p*-undecylanisole and farnesylacetone (used as flavoring agent) which share the same formula but are not in our in-house database used for screening (Figure S4.1 and S4.2). Moreover, some features are in-source fragmentation products of other features: for example, the MS/MS spectra collected for $C_8H_4O_3$ at 12.63 min match with the suspected structure, phthalic anhydride; however, the RT is too late for this formula, and it is suspected to be an in-source fragmentation product of another suspect feature, dibutyl phthalate, which also has the matched RT (Figure S5).

There were some FP features even after the MS/MS identification step (level 2 FP) and were not discovered until the analytical standard comparison step. As discussed above, either the retention times and/or the MS/MS spectra of these level 2 FP hits are different from those of the standard compounds (e.g., stereoisomers of the standards), or the standard compounds (e.g., bornyl acetate) could not be detected in the ionization modes in which the level 2 FP hits were detected.

We also noticed that in one rare special case, a level 2 FP might pass the analytical standard comparison step depending on the experimental setting employed. For example, the MS/MS spectra collected for feature C_9H_7N at 3.8 min matched with the suspected structure quinoline or isoquinoline. If quinoline/isoquinoline and another analytical standard, 3-indolepropionic acid, were prepared in one mixed solution for injection, which can be a common practice for standard MS spectra acquisition, this C_9H_7N at 3.8 min feature would be coincidently and wrongly identified as a level 1 true positive as quinoline/isoquinoline because that RT of 3-indolepropionic acid is 3.8 min. In fact, RT = 6.5 and 6.2 min are observed for the isoquinoline and quinoline analytical standards, respectively, when injected alone, and the C_9H_7N at 3.8 min feature is the in-source fragmentation product of the level 1 feature $C_{11}H_{11}NO_2$ at 3.8 min, which is confirmed to be 3-indolepropionic acid (Figure S2).

https://doi.org/10.1021/jasms.1c00135
J. Am. Soc. Mass Spectrom. 2021, 32, 2425−2435

Generally, features that are detected in both ESI+ and ESI− modes are more likely to be real features and identified with higher identification confidence. As shown in Figure 4, the percentage of features that are detected in both ESI modes are highest in the group of 18 level 1 features and is lowest in the group of all suspect chemicals. We use values in % instead of in raw numbers in the figure because the numbers of level 1 features, level 2 features, prioritized features, and all suspect features vary drastically. Values in % better convey the idea that features that can be detected in both ESI+ and ESI− modes are more likely to be real features and identified with higher identification confidence. Therefore, appearances in both ESI+ and ESI− modes can be used as a critical factor for feature prioritization in future studies. However, it is also worth noting that these features also belong to a special class of chemicals which can be ionized into both cations and anions.

Conclusion and Future Perspectives. We analyzed 30 pairs of maternal and umbilical cord serum samples collected in the San Francisco Bay area for a large-scale implementation, especially in terms of the broadness of industrial chemical screening database used and the large number of suspect features prioritized for further confirmation, of NTA SSA methodology in the biomonitoring field. From a total of 17512 ESI+ and 12054 ESI− features extracted and cleaned from these samples, 662 ESI+ and 788 ESI− were matched with suspected chemicals out of the in-house constructed industrial chemical database. Among them, 116 ESI+ and 124 ESI− features were prioritized for further identification and confirmation. By analyzing the targeted MS/MS experimental data of these prioritized features, 16 ESI+ and 17 ESI− features were tentatively identified with a level 2 and 3 confidence. After the MS and MS/MS of these level 2 and 3 features were compared with those of the purchased analytical standards, eight structures for the ESI+ mode and nine structures for the ESI− mode are confirmed with level 1 confidence.

Among the confirmed level 1 features, monoethylhexyl phthalate (MEHP), 4-nitrophenol, tridecanedioic acid, and octadecanedioic acid are especially interesting for further exploration in a future exposure study. MEHP is a common plasticizer metabolite; 4-nitrophenol is found in diesel exhaust particles and is also used for manufacture of pharmaceutical drugs and fungicides; the last two are abnormal fatty acids that appear in high levels in patients with Zellweger syndrome[34] and Reye's syndrome,[35,36] respectively, but an excess amount in healthy people suggests that the exposure from these two compounds may be from industrial chemicals because they are also widely used in manufacturing plastics.

The advantage of the utilized tiered approach for feature prioritization is that we are able to focus first on the most interesting features at the time and can revisit the features in lower tiers (e.g., tier 4, tier 5, etc.) later given additional time and resources. For example, if we set tier 4 to be fluorinated chemicals regardless of the DF, demographic difference, or maternal−cord correlations, we could identify and confirm another nine poly- and perfluoroalkyl substances as additional level 1 features (Table S7). We can also set tier 5 as features with $80 \leq DF < 100\%$ in follow-up studies by setting DF = 100% in this study: (1) our approach can miss interesting chemicals to which the population is exposed universally but were not detected at a high frequency due to reasons such as individual differences of the subjects from whom the serum samples were collected, the specific type of LC-MS instrument used in this experiment, sample extraction efficiencies, matrix effects, limit of detection, etc.; (2) endogenous chemicals, and background or matrix interference existing in all samples that are also higher than blank levels, are more likely to be picked up. Moreover, chemicals of common exposure instead of universal exposure are also meaningful for biomonitoring or regulation, and some chemicals of interests may be only distributed in limited population and associated with occupational exposure, age, gender, race/ethnicity etc.

Despite the interference of a serious number of endogenous compounds in human serum (possibly up to 38% of total level 4 features, vide supra) and low-quality MS/MS spectra (∼28% with low S/N, ∼10% with no precursor ion found, ∼3% out of detection range, vide supra), we were able to achieve a level 2 and 3 identification rate of ∼14% and confirm the identify of ∼50% of these level 2 and 3 features with level 1 confidence. Nonetheless, the general low rates observed in all SSA studies is still a long-standing challenge and calls for both intralab improvement and interorganizational collaboration of researchers worldwide for improvements in many aspects. Based on our analysis of origins of false positive features and unidentifiable suspected features, as well as obstacles encountered during the confirmation process, we believe the following aspects are possible directions for future advancement: (1) improve the sample extraction process for more efficient removal of lipids and other endogenous compounds, and make sample extracts more concentrated so that low concentration features can be detected with better resolved isotopic patterns; (2) refine the feature extraction algorithm to better group and bin the peaks detected in MS scan to decrease duplicated or artifact features (like isotopic peaks of high intensity wrongly identified as new features); (3) find more suitable "blanks" for blank subtraction process, because serum blanks cannot be used since the exposome of the blank serum cannot be determined, using procedure DI water sample as blanks as in this study, only a small fraction of features can be subtracted, whereas many endogenous/metabolite features from the complicated serum matrix unrelated to exposome are retained; (4) improve suspect screening database to better reflect environmental chemical exposure and highlight those with high production volume or that are more frequently seen in commonly used commercial products, as these are more likely to be true positives; (5) enhance the automatic feature prioritization strategies, such as incorporating chemical structure/retention time predicting tools, to pick out interesting true positive features more efficiently; (6) expand existing online MS/MS libraries to cover more industrial chemicals, which can be greatly benefited from collaborations from analytical research laboratories around the world; (7) improve instrument sensitivities and resolution, especially in MS^2 acquisition; (8) incorporate machine learning techniques to improve automated online spectra matching and fragmentation pattern recognition.

As for SSA studies in the biomonitoring field with biological samples specifically, apart from the above aspects, we cannot ignore the fact that a lot of exogenous chemicals that enter the body will be metabolized by the liver, kidneys, etc. into different unrecognized structures and be removed from the body, which may also be a critical, even the core, reason for the overall low identification and confirmation rates in works like this. In this respect, studying the metabolites of industrial chemicals and building MS and MS/MS databases for these metabolites will be of great benefit to better identify industrial chemicals in the body.

https://doi.org/10.1021/jasms.1c00135
J. Am. Soc. Mass Spectrom. 2021, 32, 2425−2435

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jasms.1c00135.

Figures S1−S5 and Tables S1−S7 (PDF)

In-house chemical database used in this study in an excel spreadsheet (XLSX)

■ AUTHOR INFORMATION

Corresponding Authors

June-Soo Park − *California Department of Toxic Substances Control, California Environmental Protection Agency, Berkeley, California 94710, United States; Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco, California 94158, United States*; Email: June-Soo.Park@dtsc.ca.gov

Tracey J. Woodruff − *Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco, California 94158, United States*; Email: Tracey.Woodruff@ucsf.edu

Authors

Ting Jiang − *California Department of Toxic Substances Control, California Environmental Protection Agency, Berkeley, California 94710, United States; Public Health Institute, Oakland, California 94607, United States*; orcid.org/0000-0002-4408-7852

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

Aolin Wang − *Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco, California 94158, United States*

Dimitri Abrahamsson − *Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco, California 94158, United States*

Weixin Kuang − *California Department of Toxic Substances Control, California Environmental Protection Agency, Berkeley, California 94710, United States; Public Health Institute, Oakland, California 94607, United States*

Rachel Morello-Frosch − *School of Public Health and Department of Environmental Science, Policy and Management, University of California, Berkeley, California 94720-3114, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jasms.1c00135

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Hollender, J.; Schymanski, E. L.; Singer, H. P.; Ferguson, P. L. Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? *Environ. Sci. Technol.* **2017**, *51* (20), 11505−11512.

(2) Gago-Ferrero, P.; Schymanski, E. L.; Bletsou, A. A.; Aalizadeh, R.; Hollender, J.; Thomaidis, N. S. Extended Suspect and Non-Target Strategies to Characterize Emerging Polar Organic Contaminants in Raw Wastewater with LC-HRMS/MS. *Environ. Sci. Technol.* **2015**, *49* (20), 12333−12341.

(3) Grigoriadou, A.; Schwarzbauer, J. Non-target Screening of Organic Contaminants in Sediments from the Industrial Coastal Area of Kavala City (NE Greece). *Water, Air, Soil Pollut.* **2011**, *214* (1), 623−643.

(4) Krauss, M.; Singer, H.; Hollender, J. LC−high resolution MS in environmental analysis: from target screening to the identification of unknowns. *Anal. Bioanal. Chem.* **2010**, *397* (3), 943−951.

(5) Munoz, G.; Liu, J.; Vo Duy, S.; Sauvé, S. Analysis of F-53B, Gen-X, ADONA, and emerging fluoroalkylether substances in environmental and biomonitoring samples: A review. *Trends Environ. Anal. Chem.* **2019**, *23*, e00066.

(6) Roca, M.; Leon, N.; Pastor, A.; Yusà, V. Comprehensive analytical strategy for biomonitoring of pesticides in urine by liquid chromatography−orbitrap high resolution mass spectrometry. *Journal of Chromatography A* **2014**, *1374*, 66−76.

(7) Picó, Y.; El-Sheikh, M. A.; Alfarhan, A. H.; Barceló, D. Target vs non-target analysis to determine pesticide residues in fruits from Saudi Arabia and influence in potential risk associated with exposure. *Food Chem. Toxicol.* **2018**, *111*, 53−63.

(8) Guo, Z.; Zhu, Z.; Huang, S.; Wang, J. Non-targeted screening of pesticides for food analysis using liquid chromatography high-resolution mass spectrometry-a review. *Food Addit. Contam., Part A* **2020**, *37* (7), 1180−1201.

(9) Escher, B. I.; Stapleton, H. M.; Schymanski, E. L. Tracking complex mixtures of chemicals in our changing environment. *Science* **2020**, *367* (6476), 388.

(10) 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.

(11) Yukioka, S.; Tanaka, S.; Suzuki, Y.; Echigo, S.; Kärrman, A.; Fujii, S. A profile analysis with suspect screening of per- and polyfluoroalkyl substances (PFASs) in firefighting foam impacted waters in Okinawa, Japan. *Water Res.* **2020**, *184*, 116207.

(12) Yao, C.; Wang, J.; Chan, P.; Feng, Y.-L. A novel non-targeted screening method for urinary exposure biomarker discovery of phthalates using liquid chromatography-mass spectrometry. *Anal. Methods* **2018**, *10* (9), 959−967.

(13) Baz-Lomba, J. A.; Reid, M. J.; Thomas, K. V. Target and suspect screening of psychoactive substances in sewage-based samples by UHPLC-QTOF. *Anal. Chim. Acta* **2016**, *914*, 81−90.

(14) Bauer, A.; Luetjohann, J.; Rohn, S.; Jantzen, E.; Kuballa, J. Development of a Suspect Screening Strategy for Pesticide Metabolites in Fruit and Vegetables by UPLC-Q-Tof-MS. *Food Analytical Methods* **2018**, *11* (6), 1591−1607.

(15) 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), 077009−077009.

(16) 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.

(17) 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.

(18) 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.

(19) Blaženović, I.; Kind, T.; Torbašinović, H.; Obrenović, S.; Mehta, S. S.; Tsugawa, H.; Wermuth, T.; Schauer, N.; Jahn, M.; Biedendieck, R.; Jahn, D.; Fiehn, O. Comprehensive comparison of in silico MS/MS fragmentation tools of the CASMI contest: database boosting is needed to achieve 93% accuracy. *J. Cheminf.* **2017**, *9* (1), 32.

(20) Nguyen, D. H.; Nguyen, C. H.; Mamitsuka, H. Recent advances and prospects of computational methods for metabolite identification: a review with emphasis on machine learning approaches. *Briefings Bioinf.* **2019**, *20* (6), 2028−2043.
(21) Hug, C.; Ulrich, N.; Schulze, T.; Brack, W.; Krauss, M. Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening. *Environ. Pollut.* **2014**, *184*, 25−32.
(22) 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.
(23) Knolhoff, A. M.; Croley, T. R. Non-targeted screening approaches for contaminants and adulterants in food using liquid chromatography hyphenated to high resolution mass spectrometry. *Journal of Chromatography A* **2016**, *1428*, 86−96.
(24) Guo, Z.; Huang, S.; Wang, J.; Feng, Y.-L. Recent advances in non-targeted screening analysis using liquid chromatography - high resolution mass spectrometry to explore new biomarkers for human exposure. *Talanta* **2020**, *219*, 121339.
(25) Richard, A. M.; Judson, R. S.; Houck, K. A.; Grulke, C. M.; Volarath, P.; Thillainadarajah, I.; Yang, C.; Rathman, J.; Martin, M. T.; Wambaugh, J. F.; Knudsen, T. B.; Kancherla, J.; Mansouri, K.; Patlewicz, G.; Williams, A. J.; Little, S. B.; Crofton, K. M.; Thomas, R. S. ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology. *Chem. Res. Toxicol.* **2016**, *29* (8), 1225−1251.
(26) Williams, A. J.; Grulke, C. M.; Edwards, J.; McEachran, A. D.; Mansouri, K.; Baker, N. C.; Patlewicz, G.; Shah, I.; Wambaugh, J. F.; Judson, R. S.; Richard, A. M. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *J. Cheminf.* **2017**, *9* (1), 61.
(27) Wang, A.; Abrahamsson, D. P.; Jiang, T.; Wang, M.; Morello-Frosch, R.; Park, J.-S.; Sirota, M.; Woodruff, T. J. Woodruff Suspect screening, prioritization and confirmation of environmental chemicals in maternal-newborn pairs from San Francisco. *Environ. Sci. Technol.* **2021**, *55*, 5037−5049.
(28) MassBank of North America (MoNA), a metadata-centric, auto-curating repository designed for efficient storage and querying of mass spectral records. Available online: https://mona.fiehnlab.ucdavis.edu (accessed on 2021-01-18).
(29) MassBank Europe, a high quality mass spectral database. Available online: https://massbank.eu/MassBank/Index (accessed on 2021-01-18).
(30) 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.
(31) mzCloud, a state of the art mass spectral database that assists analysts in identifying compounds in areas such as life sciences, metabolomics, pharmaceutical research, toxicology, forensic investigations, environmental analysis, food control and various industrial applications. Available online: https://www.mzcloud.org/ (accessed on 2021-01-18).
(32) Djoumbou-Feunang, Y.; Pon, A.; Karu, N.; Zheng, J.; Li, C.; Arndt, D.; Gautam, M.; Allen, F.; Wishart, D. S. CFM-ID 3.0: Significantly Improved ESI-MS/MS Prediction and Compound Identification. *Metabolites* **2019**, *9* (4), 72.
(33) Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *J. Cheminf.* **2016**, *8* (1), 3.
(34) Rocchiccioli, F.; Aubourg, P.; Bougneres, P. F. Medium- and Long-Chain Dicarboxylic Aciduria in Patients with Zellweger Syndrome and Neonatal Adrenoleukodystrophy. *Pediatr. Res.* **1986**, *20* (1), 62−66.
(35) Tonsgard, J. H. Serum dicarboxylic acids in patients with Reye syndrome. *J. Pediatr.* **1986**, *109* (3), 440−445.
(36) Ng, K. J.; Andresen, B. D.; Hilty, M. D.; Bianchine, J. R. Identification of long chain dicarboxylic acids in the serum of two patients with Reye's syndrome. *J. Chromatogr., Biomed. Appl.* **1983**, *276*, 1−10.

https://doi.org/10.1021/jasms.1c00135
J. Am. Soc. Mass Spectrom. 2021, 32, 2425−2435