Environmental Science & Technology

Editors' Choice
Open Access

pubs.acs.org/est

Critical Review

High-Resolution Mass Spectrometry for Human Exposomics: Expanding Chemical Space Coverage

Published as part of Environmental Science & Technology virtual special issue "The Exposome and Human Health".

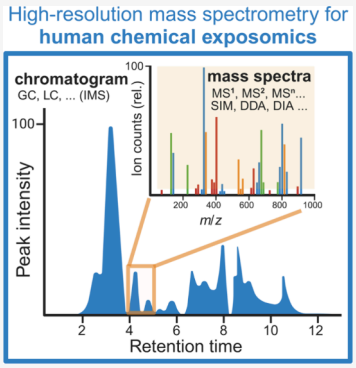
Yunjia Lai, Jeremy P. Koelmel, Douglas I. Walker, Elliott J. Price, Stefano Papazian, Katherine E. Manz, Delia Castilla-Fernández, John A. Bowden, Vladimir Nikiforov, Arthur David, Vincent Bessonneau, Bashar Amer, Suresh Seethapathy, Xin Hu, Elizabeth Z. Lin, Akrem Jbebli, Brooklynn R. McNeil, Dinesh Barupal, Marina Cerasa, Hongyu Xie, Vrinda Kalia, Renu Nandakumar, Randolph Singh, Zhenyu Tian, Peng Gao, Yujia Zhao, Jean Froment, Pawel Rostkowski, Saurabh Dubey, Kateřina Coufalíková, Hana Seličová, Helge Hecht, Sheng Liu, Hanisha H. Udhani, Sophie Restituito, Kam-Meng Tchou-Wong, Kun Lu, Jonathan W. Martin, Benedikt Warth, Krystal J. Godri Pollitt, Jana Klánová, Oliver Fiehn, Thomas O. Metz, Kurt D. Pennell, Dean P. Jones, and Gary W. Miller*

Cite This: *Environ. Sci. Technol.* 2024, 58, 12784−12822

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: In the modern "omics" era, measurement of the human exposome is a critical missing link between genetic drivers and disease outcomes. High-resolution mass spectrometry (HRMS), routinely used in proteomics and metabolomics, has emerged as a leading technology to broadly profile chemical exposure agents and related biomolecules for accurate mass measurement, high sensitivity, rapid data acquisition, and increased resolution of chemical space. Non-targeted approaches are increasingly accessible, supporting a shift from conventional hypothesis-driven, quantitation-centric targeted analyses toward data-driven, hypothesis-generating chemical exposome-wide profiling. However, HRMS-based exposomics encounters unique challenges. New analytical and computational infrastructures are needed to expand the analysis coverage through streamlined, scalable, and harmonized workflows and data pipelines that permit longitudinal chemical exposome tracking, retrospective validation, and multi-omics integration for meaningful health-oriented inferences. In this article, we survey the literature on state-of-the-art HRMS-based technologies, review current analytical workflows and informatic pipelines, and provide an up-to-date reference on exposomic approaches for chemists, toxicologists, epidemiologists, care providers, and stakeholders in health sciences and medicine. We propose efforts to benchmark fit-for-purpose platforms for expanding coverage of chemical space, including gas/liquid chromatography−HRMS (GC-HRMS and LC-HRMS), and discuss opportunities, challenges, and strategies to advance the burgeoning field of the exposome.



KEYWORDS: *exposome, toxicants, high-resolution mass spectrometry, chromatography, non-targeted analysis, environmental exposures, chemical space, metabolomics*

1. INTRODUCTION

The exposome encompasses non-genetic exposures and is the integrated compilation of all physical, chemical, biological, and psychosocial influences that impact biology, constituting a key determinant of health.[1−3] Due to anthropogenic impacts on both global and local scales, environmental pollution levels are on the rise, with numerous chemical stressors being dispersed into our surroundings including air, water, soil, and indoor environments.[4−6] These exposures occur in the context of non-chemical exposome components such as socio-cultural factors and lifestyles, which may modify effects and responses.[2] Humans face increasingly complex chemical exposures from both voluntary (e.g., cosmetics, pharmaceuticals and personal care products [PPCP])[7,8] and involuntary (e.g., inhalation of polluted air, food packaging)[9] sources, with potential toxic

Received: February 1, 2024
Revised: June 11, 2024
Accepted: June 12, 2024
Published: July 10, 2024

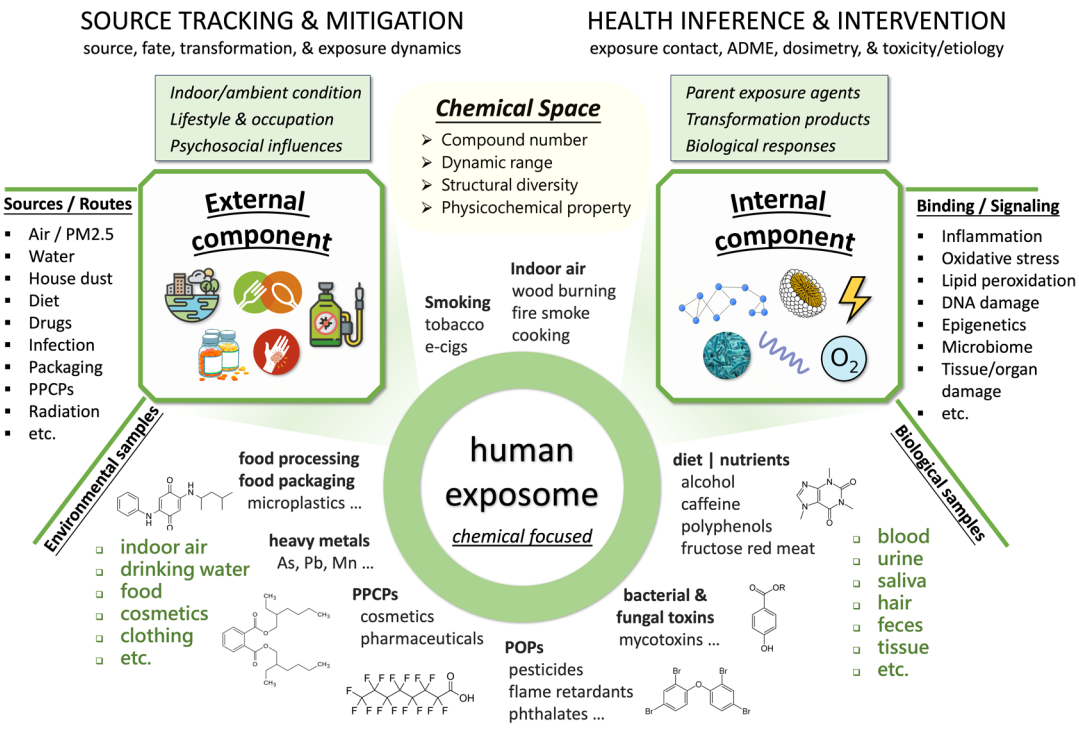


Figure 1. Schematic illustration of human chemical exposome consisting of both external and internal components, which embraces a vast chemical space by number, dynamic range, structural diversity, and physicochemical properties. The external component encompasses environmental chemicals humans are being exposed to and accumulate in the body, can have indoor, ambient, and occupational sources, and likely varies in individuals with distinct diet, drug, and lifestyle choices and psychosocial influences. The internal component is a dynamic reservoir of (*i*) parent exposure agents taken in, (*ii*) their biotransformation products, and (*iii*) endogenous biomolecules indicative of a toxicological and/or etiologic effect. **Abbreviations**: PPCP, pharmaceuticals and personal care products; ADME, absorption, distribution, metabolism, and excretion; e-cigs: E-cigarettes.

cocktail effects (e.g., additive, synergistic, antagonistic) likely arising from chemical mixtures that involve disparate dynamic ranges and modes of action (MOA).[10,11] Genomic studies have demonstrated that most chronic noncommunicable diseases are of a non-genetic origin,[6,12,13] and a recent exposome-wide association study (ExWAS) of aging and mortality in the UK Biobank (~500,000 participants) has further demonstrated that all-cause mortality is driven more by the exposome than the genome.[14]

In this article, we will focus on mapping the chemical component of the exposome to identify environmental drivers of disease, a key step toward exposomics—a transdisciplinary field aimed at enabling discovery-based analysis of the environmental factors that contribute to disease. This depends on our ability to detect, screen, and profile exposures to environmental chemicals and their transformation products in an unbiased and scalable manner.[15−17] The recent launching of large human studies and initiatives nationally (e.g., the NIH "*All of Us*")[18] and globally (e.g., EHEN, the European Human Exposome Network) provides unique opportunities for exposome research.[18,19] Through century-long development, mass spectrometry (MS)-based technologies stand out for identifying and quantifying molecules with high sensitivity, coverage, and a wide linear dynamic range.[20] Notably, the use of high-resolution mass spectrometry (HRMS) not only incentivizes a shift in biomonitoring of xenobiotics from targeted analyses (e.g., as undertaken in CDC's National Health and Nutrition Examination Survey, NHANES) toward non-targeted and mixture discovery,[21] but complements genome sequencing in biology and medicine for functional analyses. These functional capacities span proteomics, metabolomics, and now chemical exposomics, i.e., the omics-scale measurement and health-oriented inference of small-molecule (molecular weights ≤1000 Da) exposure agents, transformation products, and associated biomolecules through targeted and suspect approaches for expected and known compounds, and non-targeted analysis (NTA) for unexpected or unknown compounds.[2,22,23]

With analytical strengths and proof-of-principle evidence from metabolomics,[24] HRMS has emerged as an essential tool for chemical exposomics.[16,25,26] However, analytical challenges and limitations remain, largely due to the diverse chemical space encompassing a wide dynamic range of exogenous chemicals and their transformation products in the human body at substantially lower levels than endogenous biomolecules.[27] Standard practices adopted in metabolomics and related fields are not fully transferable to human chemical exposomics.[28,29] Although simultaneous analysis of exogenous chemicals, transformation products, and biomarker responses is possible through workflows optimized for metabolomics, analytical biases or gaps in chemical space coverage may occur in exposomics.[29] For instance, procedures may be needed to concentrate low-abundance exogenous analytes, distinguish/remove background contaminants (e.g., polyethylene glycol, phthalates), and counteract interferences from endogenous biomolecules (e.g., lyso- and phospholipids in blood plasma/serum).[30−34] On multiple levels, there are trade-offs between coverage and throughput, considering the sporadic occurrences, low abundance, structural diversity, and wide-ranging

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

physicochemical properties of environmental chemicals and their transformation products.[16]

Here, we survey the recent literature on HRMS-based analysis for insights into advancing human exposomics. We discuss existing techniques and practices through an analytical chemistry lens while focusing on pertinent topics from laboratory measurement (e.g., sampling, instrumentation, assay) to data analytics (e.g., feature detection, structural annotation). Balancing breadth with selective depth, we aim to identify new trends and prospects for chemical exposome research. We highlight a need to harmonize research efforts and benchmark emerging toolkits essential for expanding the analytical coverage of exposomics, such as gas chromatography (GC)-HRMS, liquid chromatography (LC)-HRMS, and ion mobility spectrometry (IMS). While this article focuses on organic molecules only, one should note that metals/metalloids constitute another critical exposome component commonly measured by inductively coupled plasma mass spectrometry (ICP-MS).[35] We discuss challenges, opportunities, and strategies for advancing HRMS-based exposomics, aiming to provide a primer reference for chemists, epidemiologists, toxicologists, care providers, and associated stakeholders in health sciences and beyond.

2. HUMAN EXPOSOME IN CHEMICAL SPACE

Understanding the chemical nature of the human exposome informs the rational design and implementation of chemical exposomics measurements and statistical analysis. The chemical space, referred to as the total collection of all possible molecules (theoretically or empirically) in a given context, represents a crucial concept for advancing biology and medicine.[36,37] The identification and prioritization of the diverse environmental chemicals, especially the persistent and bioaccumulative organic pollutants, have been attempted through models by physicochemical and/or toxicological properties and primarily for ambient environments.[38−40] Certain conceptual nuances of chemical space exist between traditional environmental modelers and modern non-targeted analysts in characterizing environmental chemicals.[40,41] Here, we define the human exposome in chemical space as the total collection of (*i*) chemical exposure agents humans are being exposed to and accumulate in the body, (*ii*) transformation products *in vivo*, and (*iii*) biomolecules indicative of a toxicological and/or etiologic effect in question (Figure 1).[38,39] For chemical exposome measurements, we refer to "analytical coverage" as the performance of a specific analytical workflow and the associated data pipeline to cover the chemical space in question by comprehensiveness, accuracy, and dynamic range.

Recent efforts to map possible organic chemical space for the human exposome have been based on curating literature, compound databases, and chemical inventories. Such attempts were conducted using multiple summary metrics, including compound number, class/use/source, dynamic range, lipophilicity, and inclusion of specific elements (e.g., halogens) and functional groups, with many derived from (blood) biomonitoring data targeting the internal component of the exposome.[27,42−46] To assess specific NTA workflows, the *ChemSpaceTool* was recently proposed as an integrated filtering framework, partitioning chemical space into (*i*) the detectable space, (*ii*) the identifiable space, and (*iii*) compound regions that are neither detectable nor identifiable using the select methods.[47] Substantial data curation efforts are being undertaken to make the search space of exposome compounds accessible and actionable, as exemplified by PubChemLite for Exposomics,[48] Exposome Explorer,[49] ChemMaps,[50] and the CompTox Chemicals Dashboard.[51]

2.1. Number of Compounds. Hundreds of millions of compounds and substances are being documented in public centralized chemical databases such as the Chemical Abstracts Services (CAS) (∼204 million items)[52] and PubChem (∼116 million compounds; 310 million substances),[53] raising the question: how vast and diverse is the chemical space for the human exposome? In 2016, the U.S. EPA launched CompTox Chemicals Dashboard, a highly integrated and curated hub of environmental chemicals that has cataloged ∼1.2 million searchable compounds and over 400 lists based on structure or category.[51] To assess chemicals in their commercial production and societal use, Wang and colleagues assembled the first global inventory of chemicals on the market into a catalog of over 350,000 compounds and mixtures, which unexpectedly tripled the number of previous listings.[42] Notably, due to corporate confidentiality, ∼120,000 substances remain inconclusively identified (unknowns or insufficiently described for a confident CAS# assignment), calling for internationally coordinated efforts among stakeholders (of research and regulation) to expand global inventories with transparency and accuracy.[42]

To make chemical exposomics feasible, PubChemLite for Exposomics is cataloging >360,000 candidate chemicals from PubChem's millions based on category (e.g., drugs, food additives, agrochemicals), toxicity, and disease relevance, to improve search space accessibility.[48] To gauge population exposure profiles and health effects, epidemiological research is essential. However, for most human cohort studies, only dozens up to a few hundred chemicals (and their biomarker metabolites) are analyzed.[42] Although certain prioritizations are necessary considering technical constraints, budget limitations, and disparate exposure patterns and health effects among individual chemicals, a harmonized prioritization framework is lacking. Strategies and tactics have only been recently discussed and attempted, highlighting a disconnect of understanding between what we need to measure and what we can measure.[21,54]

2.2. Dynamic Range. The dynamic range of exposome chemicals is vast, as revealed by MS-based analyses of blood compounds spanning up to 11 orders of magnitude.[27] Pollutants detected in blood were generally 1,000 times lower in abundance than compounds derived from food, drug, and endogenous origins, suggesting a need for more sensitive platforms.[27] Plasma and serum measurements for a variety of compounds determined by targeted LC-HRMS methods were compiled and categorized into 8 representative compound classes: the concentration profiles spanned 8 orders of magnitude, ranging from 10^{-2} (e.g., environmental pollutants) to 10^{6} ng/mL (e.g., lipids, nucleotides, food components).[16] Within compound class, molecules ranged over 7 orders of magnitude with certain classes reaching below the limits of detection (LOD) by LC-HRMS.[16] Together, these point to a challenge to profiling environmental pollutants which are often low-abundant and likely co-occur in structural congener mixtures (e.g., isomers, homologues, transformation products, isotopomers, etc.), thereby rendering a potential cocktail health effect difficult to discern.

Despite the analytical sensitivity and selectivity challenges, HRMS has demonstrated comparably good quantitative

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

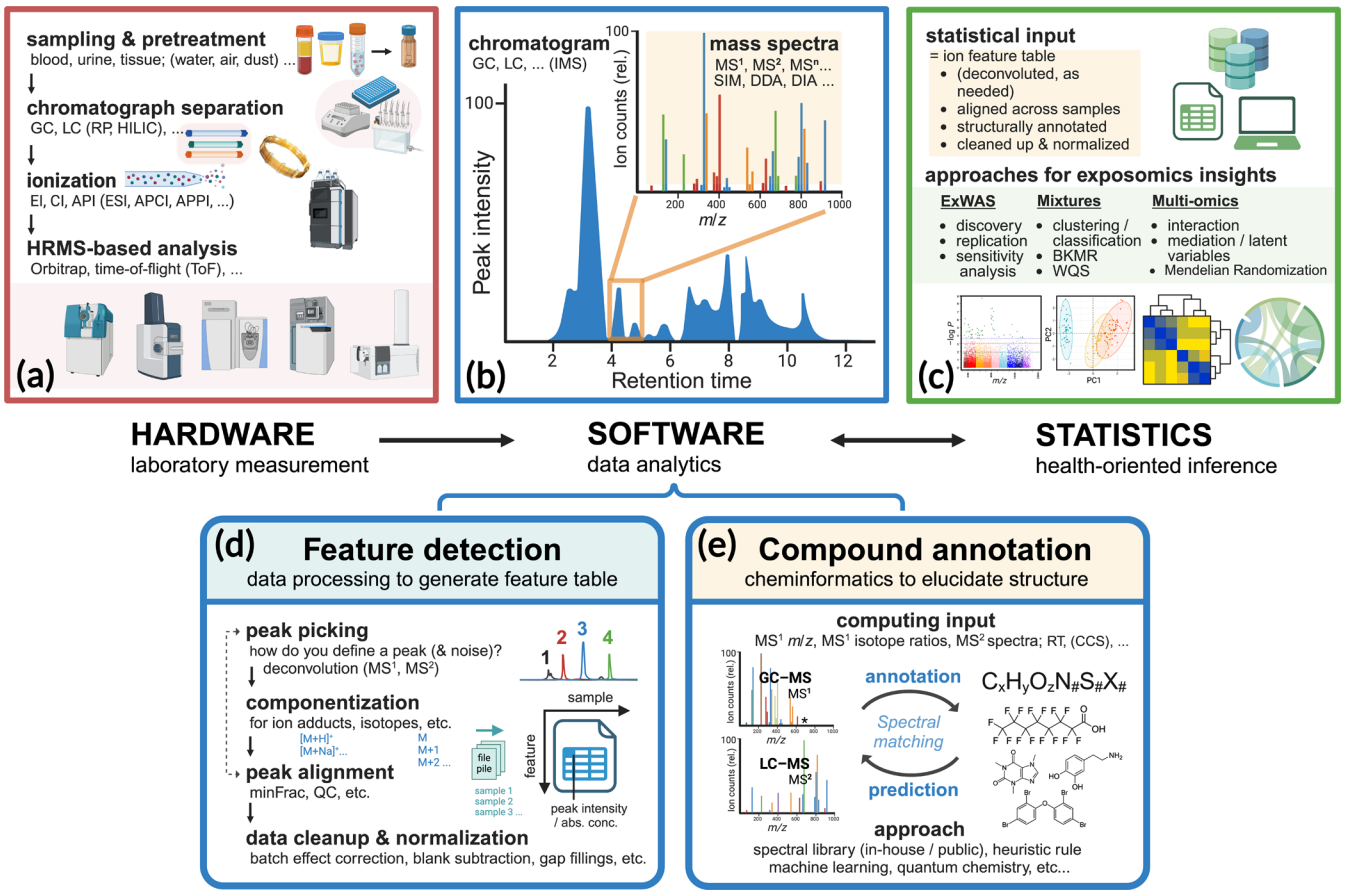


Figure 2. Critical steps for expanding the analytical coverage by HRMS-based exposomics. Modular components at the front end, from (**a**) laboratory measurement to (**b**) data analytics, are essential to generating quality feature tables for (**c**) advanced statistics successes at the later stages of analysis. To generate a feature table, the data analytics entails (**d**) feature detection and (**e**) compound annotation for best results, both of which have been critically reviewed in this article. The figure was generated using BioRender under a paid subscription. **Abbreviations**: GC, gas chromatography; LC, liquid chromatography; RP, reverse phase; HILIC, hydrophilic interaction chromatography; EI, electron ionization; CI, chemical ionization; API, atmospheric pressure ionization; ESI, electrospray ionization; APCI, atmospheric pressure chemical ionization; APPI, atmospheric pressure photoionization; IMS, ion mobility spectrometry; SIM, selective ion monitoring; DDA, data-dependent acquisition; DIA, data-independent acquisition; minFrac, minimum fraction (proportion of minimum samples where a peak has to be present in a group); QC, quality control; RT, retention time; CCS, collision cross section; ExWAS, exposome-wide association studies; PCA, principal component analysis; FA, factor analysis; NMF, non-negative matrix factorization; BKMR, Bayesian Kernel Machine Regression; WQS, Weighted Quantile Sum.

capability compared with sensitive targeted assays employed for decades using low-resolution MS (LRMS). In a recent study by Flasch et al. (2023), a triple quadrupole (QqQ) and an HRMS method (coupling to identical chromatography) were compared in their respective most frequently used acquisition modes (full-scan, also called "survey scan", for HRMS; multiple reaction monitoring for QqQ).[55] In the HRMS analyses, the median limit of quantitation (LOQ) was determined as 0.9 and 1.2 ng/mL in solvent and urine, respectively, while for the QqQ measurements, the median LOQ was 0.1 and 0.2 ng/mL in solvent and urine, respectively. In another work, the two approaches were compared for determining polyphenols in human urine, sera, and plasma, reaching a median LOD of 10−18 ng/mL for HRMS and 4.8−5.8 ng/mL for LRMS.[56] The high sensitivity of HRMS achieved sheerly via full-scan MS^1 suggested significant potential for HRMS to enable both targeted and non-targeted analyses within a single run.

2.3. Structural Diversity and Physicochemical Properties. On the molecular level, it is the chemical structure of the compound that governs its physicochemical properties and activity, which fundamentally determine exposure occurrences, biological function, health effects, and the associated strategies and approaches for chemical analysis, especially the extraction schemes at the front end.[2,57,58] For any given organic molecule, pertinent physicochemical properties include polarity (distribution of electrical charge across chemical atoms, bonds, functional groups, and the overall structure) and volatility (tendency to vaporize and partition in the air), among others.[57] These are collectively affected by structural characteristics, from the component elements (e.g., halogens and heteroatoms), molecular formula, molecular weight, and degree of saturation to the presence of specific substructures (e.g., fused rings, heterocyclic rings, prolonged aliphatic chains) and functional groups (e.g., amine, carboxylic group).

Faced with the sheer number and structural diversity of exogenous pollutants, environmental modelers and risk assessors have long leveraged the quantitative structure−activity relationships (QSAR) or similar models to conduct scalable predictions or read-across of environmental fate, transport, transformation (e.g., photo-, bio-), exposure dynamics, and ecological/human toxicity.[59,60] When modeling internal exposures and the associated functional effects in humans, more sophisticated cheminformatics or *in silico* New

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

Table 1. Thematic Summary of HRMS-Based Techniques by Modular Steps for Human Chemical Exposomics

Modular Steps	Modular Themes and Topics		Techniques, Characteristics, and Functionality	Practical Consideration	Notable References
Sample → Chemical extract mixtures	Sampling and Pretreatment	Sample choices	biological samples (e.g., blood, urine, saliva, tissue, feces); environmental samples (e.g., indoor air, drinking water, food packaging, cosmetics)	depends on (1) the goals of analysis, e.g., capturing internal/external components of the exposome, targeted vs non-targeted, and hypothesis-free vs effect-directed, (2) sample availability, (3) analytical resource restraints, (4) financial budgets, etc.	38,67−69,74,75,85
		Sampling modes	active vs passive sampling: the former extracts the whole sample, seeking the highest coverage and recovery possible, while the latter characterizes exposure *in situ* and/or bioavailable	sampling frequency, sensitivity, selectivity, recovery, and reproducibility	68,85,87,89,90
		Pretreatment: Extraction/Cleanup	dilute and shoot (DNS); solid−liquid or liquid−liquid extraction (SLE or LLE); solid-phase extraction (SPE); QuEChERS; accelerated solvent extraction (ASE); protein precipitation; chemical filters (e.g., OstroPlates, EMR-Lipid).	modes of analysis (targeted vs non-targeted); recovery; need to assess and curb matrix effect for quantitative analysis	25,67,69,111−114
		(Optional Pretreatment)	enzymatic digestion (e.g., sulfatase); effect-based approaches (e.g., ER protein pulldown); derivatization, to modify volatility/polarity to improve chromatographic separation and mass detection	necessary chemical modification to cater to downstream analysis, require a clear understanding of chemistry (reactions involved) and biology (e.g., knowledge of matrix; effect-based approach)	68,87,124
Chemical extract mixtures → Individual analytes separated	Pre-MS separation	Liquid chromatography (LC)	reverse phase column (RP, e.g., C_{18}, C_8, pentafluorophenyl [PFP]) for nonpolar to moderately polar species; HILIC (e.g., HILIC amide) for polar species	RP (C_{18}, C_8) is the most commonly used (>90% cases); RP with improved selectivity and peak capacity (e.g., PFP) is used for specific environmental chemicals; HILIC can be a plus when throughput is ensured through techniques such as dual chromatography and fast polarity switching	44,46,68,127
		Gas chromatography (GC)	for volatile species, nonpolar contaminants (e.g., halogenated POPs), and/or those of small MW (<300 Da) such as short-chain fatty acids (SCFAs)	a gold standard for many highly nonpolar and halogenated contaminants, fatty acids, can be applied to primary metabolite profiling, amenable to pyrolysis for microplastics analysis	44,46,68,127
		Ion mobility spectrometry (IMS)	(1) time-dispersive (e.g., DTIMS, TWIMS, SLIM); (2) space-dispersive (DMS); (3) ion trapping and select mobility release (e.g., TIMS)	offers added resolving power to chromatography without compromising throughput (*ms* time scale), but can be compound-specific; CCS as an orthogonal measure to improve compound identification; does not curb variability resulting from ionization and chromatography	126,128
		(Alternative pre-MS Modules/Modes)	(chromatography) adsorption, ion exchange, size exclusion, affinity chromatography, chiral chromatography	more selective, and can be used for novel chemical species	127
Analytes → Ions	Ionization	(for GC)	electron ionization (EI) chemical ionization (CI), etc.	EI is most widely used for GC: hard ionization, highly reproducible, too harsh with no $M^{+\bullet}$ to survive; CI is relatively softer with sensitivity varying between compound species. negative ionization can be sensitive for halogenated compounds but by far much less tested	127,161,293
		(for LC)	API: electrospray ionization (ESI), APCI, APPI, etc. matrix-assisted laser desorption/ionization (MALDI)	ESI: soft, most commonly used with wide chemical space coverage, high sensitivity (to subfemtomole), tolerates water, but needs to reduce artifacts/background and matrix effect; MALDI highly sensitive (to high attomole), for large molecules (e.g., proteins) and amenable for MS imaging	126,171,172,293
Ions → Spectral data	MS data acquisition	The overall strategy	(1) MS^1 full-scan vs MS^2 fragmentation (ion precursor-product pair), can be alternated to have both in data when cycle time allows; can be instrument-dependent (e.g., Q-Orbitrap vs Q-ToF); (2) MS^n is also possible with linear ion trap configured; (3) data format: profile mode vs centroid mode	full-scan MS^1 for profiling; MS^2 for specificity and structural elucidation. needs sufficient points/scan (preferably ≥10) over individual EICs to define and quantify a peak, challenging the characterization of low-abundant peaks	201,203
		MS^2 modes	(1) single/multiple reaction monitoring (SRM/MRM) or selected ion monitoring (SIM) or parallel reaction monitoring (PRM); (2) data-dependent acquisition (DDA, e.g., "Top N"); (3) data-independent acquisition (DIA), e.g., AIF, MS^E, SWATH, dia-PASEF	DDA generates MS^2 data—offers more than full-scan only and is more non-targeted than SRM/MRM/SIM/PRM, but could be biased; DIA is fully non-targeted, but deconvolution can be a challenge	204−206,211,215
	MS/MS fragmentation	MS^2 mechanism	(1) collision-induced dissociation (CID), with variants such as high-energy collisional dissociation (HCD); (2) electron-ion reaction-based dissociation (ExD), e.g., ECD, EAD, ETD, EIEIO; (3) proton-based activation, e.g., UVPD	CID/HCD is most commonly used for generic analysis; ExD is used when probing processes involving noncovalent interactions (e.g., PTM of proteins) UVPD helps resolve sterol species that challenge CID/HCD	159,199,218−222
Spectral data → Feature table (ready for statistics)	Data processing	Peak picking	algorithms to inspect MS spectra to define/characterize a peak, may entail tracking, binning, and clustering, e.g., centWave (XCMS), local maxima (MZmine), concavity/slicing (MS-DIAL), "moment" estimation (apLCMS); new algorithms such as "mass track" (asari) aligns peaks before picking, shunning potential provenance issues due to algorithmic drawbacks	need positive controls (e.g., QC pool samples) and negative controls (e.g., sample blanks or sample cleanup procedures) to assess algorithmic efficacy, i.e., proportions of false positive peaks and false negative peaks. could be challenging for NTA	295−298,300

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

Table 1. continued

Modular Steps	Modular Themes and Topics		Techniques, Characteristics, and Functionality	Practical Consideration	Notable References
		Componentization	(1) grouping specific spectral "fingerprints," e.g., isotopes, ion adducts, fragments; (2) deconvolution algorithms may be needed for GC-MS or LC-DIA-MS^2 data	crucial for qualitative and quantitative purposes; prone to errors in certain areas (e.g., ESI adduct distribution not yet well studied until this date); the accuracy for some can be evaluated using postprocessing tools such as MS-FLO	300,301
		Peak alignment	(1) algorithms to combine peak lists (from individual samples) into a master feature table (aligning all samples, yielding unique m/z, RT, peak intensity, etc.); (2) threshold filters to decide whether to include a peak; gap filling	duplicate peaks; false positive/negative peaks; the issue of sparsity in exposomics data: metabolomics settings (e.g., "80%" rule) are too stringent for exposomics; retrieval of isotopes information for compound identification	310,338
		Data cleanup/normalization	(1) blank subtraction to exclude background ions or contaminant signals; (2) normalization to correct for sample variability, batch effects, etc.	contamination or backgrounds are hard to dissect and subtract from exposure data; potential nonlinearity in NTA resulting in biased MS signal intensity ratios needs to be assessed	250,253,311−313
Ion feature → Chemical identity	Compound identification	MS spectral library	(1) in-house (RT, MS^1, MS^2, etc.) based on chemical standards; (2) public library based on chemical standards: open source (e.g., MassBank, METLIN) vs vendor-based (e.g., mzCloud); (3) *in silico* predicted library (e.g., CFM-ID, CSI-FingerID)	confidence level (in general): in-house library ≥ public library ≥ *in silico* generated library; the more resources available the better. regardless, experimental spectra are small compared to the vast chemical space of human exposome; GC-HRMS library and scoring options are lacking (with associated deconvolution to be benchmarked) compared to LC-HRMS. combining unit-mass GC library into HRMS application is needed	327−329
		Formula prediction and filtering	(1) can use accurate mass MS^1 m/z, isotopes, neutral loss, and MS^2 fragment m/z to retrieve formula from chemical database (e.g., PubChem) or conduct bottom-up formula construction (e.g., BUDDY); (2) can apply heuristic rules for filtering, e.g., Seven golden rules and SIRIUS to reduce search space for later structural analysis	becomes very effective (high accuracy) and computationally convenient in recent two decades, due to the rise of HRMS (enabling accurate-mass measurement), computing power, and integrated use of heuristic rules (e.g., Seven golden rules, SIRIUS)	342,367,376
		in silico approaches (chemical reaction-based)	experimental chemical reaction-based MS/MS, e.g., literature-reported (e.g., MS Frontier); hydrogen-rearrangement (HR) rules (MS-FINDER), etc.	generally, more confident than purely data-driven approaches (many with limited training data); some can be biased and unpredictable, e.g., HR rules may apply to ∼80% molecules at best as experimentally tested in select chemical space and mass spectra	322,358,359
		in silico approaches (data-driven)	machine learning-based prediction of MS/MS, e.g., MetFrag (Bayesian model); CFM-ID (probabilistic generative models for MS/MS fragmentation); GNPS (molecular networking); quantum chemistry calculation-based approaches	novel, efficient, but can lack ground truth data for systematic validation. largely not able to resolve structural nuances such as stereoisomerism and double-bond position in lipid PUFA. need to be integrated with experimental and ground truth data for accuracy	243,322,360,361
		MS spectral search: algorithms for matching and scoring	(1) similarity algorithms, e.g., probability-based matching (PBM); dot product, spectral entropy-based; (2) scoring, e.g., based on mass accuracy, bond dissociation energies (BDE), penalty of fragmentation linkages based as defined by similarity scores, etc.	weighted dot product has dominated mass spectral search for 30+ years; emerging approaches such as spectral entropy (with high computational efficiency) may expand rapidly	70,321,351,354,355
Annotated feature table → Statistical inference	Statistics and Epidemiology	(Dimensional reduction)	to deal with the high dimensionality in exposomics data sets where predictors (i.e., features) far outnumber observations (i.e., sample size). can leverage unsupervised (e.g., PCA, NMF) or supervised (e.g., PLS-DA) methods for dimensional reduction	approaches are emerging in exposomics data science with a lack of benchmarking. data can be hard to interpret	398
		Association and Interaction	(1) univariate statistics with multiple comparisons are sensitive in wide screening for associations, i.e., ExWAS; (2) multivariate statistics to reduce dimensionality, recognize patterns in data, and probe interactions	univariate statistics are considered sensitive and robust and still remain one of the most used methods in ExWAS/MWAS. usually performed first for exploratory analysis before delving into more sophisticated multivariate approaches	289,396,397
		Mixture modeling	integrates approaches to identifying potential mixture/cocktail effects (e.g., synergistic, antagonistic). may use wide-ranging univariate and multivariate methods for dimensionality reduction and pattern recognition, e.g., BKMR, WQS, random forest, etc.	the overall mixture effects are assessed first (e.g., by BKMR, WQS). if overall effects are observed, one may search for main contributors using penalized methods (e.g., elastic nets), BKMR, WQS. further interactions can be assessed	396,400−402
		Mediation and Causal Inferences	identifying causal factors or mediators: can use mediation analysis to assess indirect effect of exposures and reveal potential mediation pathways and mechanisms; can use Mendelian Randomization (MR) in combination with genomics data to determine if evidence of causality can be derived	for mediation analysis, challenges may involve confounding, reverse causation, measurement error, model specification, and missing data; for MR, mind the instrumental variable assumption, population stratification, pleiotropy, and weak instrument (SNPs weakly associated with exposure can bias MR estimates)	291,406,407

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

Approach Methods (NAM)[61] may come into play to enable high-throughput screening and prioritization of toxicants relating to human physiology and pathogenesis of disease, as represented by physiologically based kinetic (PBK) modeling,[62,63] *in vitro* to *in vivo* extrapolation (IVIVE),[64] and integrated approaches to testing and assessment (IATA) under the adverse outcome pathway (AOP) framework.[65]

The wide range of physicochemical properties (e.g., volatility, polarity) of compounds of the chemical exposome indicate a need for merging complementary strategies in sampling, extraction, separation, and ionization to reduce profiling biases and expand analytical coverages.[66] Unlike metabolomics, where exogenous chemicals are not necessarily included or treated as background contaminants in the analysis, exposomics prioritizes capturing these external exposure agents (or xenobiotics) alongside related biomarkers, with likely contaminant background issues harder to resolve. To achieve high coverage, both metabolomics and exposomics tend to minimize steps of sample preparation through balanced solvent choices and/or injection of whole samples/extracts to avoid analyte loss and degradation, although environmental chemicals often need a different pretreatment. In exposomics, this is true for the cases of ambient samples (e.g., air and water) but can be difficult for biological specimens. Additional concentration and cleanup steps may be needed in exposomics to capture low-abundance exogenous chemicals.[67] Likewise, for derivatization (if applicable) and analyte separation down the line, one notable division is chromatography choices, i.e., GC (more nonpolar, volatile, thermostable) vs LC (more polar, nonvolatile, thermally unstable), which have been discussed.[68]

2.4. HRMS-Based Chemical Characterization and Analytical Coverage. LC-HRMS has been predominantly used for exposomics due to the transferable analytical framework from metabolomics for small-molecule analysis. More recently, GC-HRMS equipped with high-resolution mass analyzers, specifically time-of-flight (ToF) and Orbitrap (high mass accuracy <1 ppm for 200 Da), has been developed for benchtop use.[69] Zhang and co-workers (2021) compiled 299 commonly monitored exogenous compounds and discovered that only half are relatively water-soluble and can be ionized under atmospheric pressure, hence amenable to LC-HRMS.[43] By surveying HRMS-based NTA of environmental and human samples, Manz and colleagues (2023) reported that only 16% of the studies used both LC-HRMS and GC-HRMS.[46] Coverage gaps of NTA between the two platforms were observed: in human samples (19 HRMS studies in total), LC-HRMS was able to detect phthalates (and their metabolites), per- and polyfluoroalkyl substances (PFAS), halogenated organics, and hair products ingredients, whereas GC-HRMS captured more volatile and nonpolar species including volatile organic compounds (VOCs), aldehydes, alkanes, alkenes, aromatics, and halogenated compounds.[46]

In another appraisal, a critical analysis of recent LC-HRMS-based NTA (2017−23) identified an alarmingly low chemical space coverage, with the number of confidently annotated compounds (Level 2 or higher, Schymanski Scale) in each sample accounting for roughly 5% of the detected features.[41,70] Such limitations in LC-HRMS NTA emphasize a need to address detection and annotation issues separately and use complementary techniques in the respective steps of extraction, chromatographic retention/separation, and mass spectral data acquisition.[41] More broadly, as HRMS instrumentation continues to advance, one should consider possible inter-dependencies between modular steps of the analytical workflow (Figure 2). While there is a need to compartmentalize and tackle detection and annotation separately, it is also crucial to identify key parameters in each that affect the performance of downstream modules and the overall coverage to enable wide profiling of chemical exposure agents, their transformation products *in vivo*, and alterations to endogenous biomolecular profiles indicative of a health effect (Figure 1). For instance, metabolites of exogenously derived chemicals are pivotal in assessing exposures. Various enzyme digestion methods, including the analysis of phase II sulfated metabolites through sulfatase treatment, are utilized for exogenous metabolite analysis.[71] The many technical aspects, from laboratory measurements to data analytics, are discussed at length in the following sections as they function as an integral companion to the burgeoning HRMS capacity.

3. HRMS: EXPERIMENTAL TECHNIQUES AND WORKFLOW

Quality mass spectral data acquisition is key to chemical exposomics success, comprising a complex, multi-step, and multifactorial process jointly coordinated by hardware and software (Figure 2**a-b**; Table 1).[72,73] In the past five years, research trends, limitations/feasibilities, and strategies have been discussed for HRMS-based exposomics, with in-depth reviews of specific techniques or assays. First, David and co-workers identified three methodological hurdles for exposomics: lack of technique versatility, sensitivity, and automated data annotation.[16] In another review, Vitale and colleagues surveyed laboratory-based components of HRMS-based exposomics and discussed benefits, costs, and strategies from sample pretreatment to instrumentation.[69] Critical questions remain to be addressed: How are samples selected and handled as a proxy readout of human chemical exposome? How can the analytes be properly separated, ionized, and detected by a mass spectrometer? What are the trade-offs and strategies to generate better (e.g., higher analytical coverage) and more health-relevant (i.e., effective study design for statistics) exposome data? In this section, we focus on experimental modules (instrumental setups and/or approaches) most integral to HRMS-centric workflows to address these questions (Figure 2a).[43]

3.1. Sample Matrices: Properties, Selection, and Sampling. Operationally, sample properties and handling directly define the chemical space to be covered in the analysis. Key aspects, including sample matrices, sampling/collection, transport, storage, and pretreatment, have been discussed.[67,69,74,75] For human exposomics, samples can be obtained in ambient/indoor environments or directly from humans and/or associated *in vitro/in vivo* models such as human induced pluripotent stem cell (iPSC)-differentiated cell culture and organ-on-chip systems.[43] Once specific goals of analysis are set,[76] important practical considerations come into play to balance relevance and convenience, sequentially for sample choices, sampling approaches, sampling frequency, and sample pretreatment. First, sample choices depend on the research question and technical feasibility. As of date, blood and urine are the two most commonly used matrices in biomonitoring (providing internal measures of the exposome), while air, dust, and water were frequently sampled for environmental monitoring (providing external measures of the exposome) (Figure 1).[44,77] Compared to tissue-type specimens which are sampled from organs or other bodily

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

compartments,[78] biofluids such as urine and blood[79] are generally less heterogeneous while offering broader chemical coverages.[80] While urine offers a timely, integrated snapshot of exposure profiles, blood is preferred in human cohort studies since it is health-indicative, accessible, and importantly, a circulating, uniform, and functional reservoir where environmental exposures and biological responses meet.[27,81] The reproducibility challenge remains; a recent meta-analysis of blood and urine exposome studies identified that both pharmacokinetics (mainly the half-life of elimination) and exposure patterns are key to reproducible exposomics results, although these could be compound-dependent and vary from case to case.[82]

To identify the functional exposome components from complex chemical mixtures, more focused strategies and approaches are needed for sample selection and preparation. These effect-based methods, such as effect-directed analysis (EDA) and toxicity identification evaluation (TIE), measure toxicity endpoints (*in vivo* or *in vitro*) to focus identification (and quantification) efforts on the compounds contributing to the observed toxicity.[83] EDA utilizes sample fractionation schemes and biological/toxicity assays designed for the problems formulated at the front end before delving into chemical analysis.[58,68] The target sample matrix (and extraction), test systems, and operational readouts of MOA and toxicity are determined first to guide the downstream chemical analysis, as implemented in recent successful cases.[84,85] Since molecular mechanisms of toxicity essentially boil down to ligand binding (to target receptors), specific and sensitive protein affinity-based assays can be used for selective extraction and screening, as demonstrated by bioassays based on endoplasmic reticulum (ER) protein pulldown.[86] For over two decades, EDA has been actively employed to enhance environmental monitoring (most commonly, water) and advance human exposomics.[87,88]

As opposed to the many active, invasive sampling methods aiming at a broad coverage (e.g., blood), it is equally important to devise passive sampling techniques to capture exposome components *in situ* (often considered the "bioavailable" fraction) in both environmental monitoring[89] and biomonitoring.[90] This is especially true when samples are not easily accessible, or intermittent, longitudinal, and noninvasive monitoring is desired, e.g., in vulnerable populations such as newborns, infants, young children, and pregnant women.[91] Polydimethylsiloxane (PDMS) or other silicone-based sorbent materials that are inert, nontoxic, and biocompatible play a crucial part in this.[92] Sorbents vary in properties and behaviors; some samplers enable accurate (time-integrated) measurement of ambient pollutants for the sampling durations, whereas others are designed to mimic how exposures reach and get absorbed/adsorbed by human individuals or organisms.[92−94] One notable example is solid-phase microextraction (SPME), which consists of hair-thin fibers precoated with high-purity silicone (e.g., PDMS) for passive (and usually non-depletive) sampling of organic pollutants (either gas-phase or in aqueous solution)[95] that has been increasingly applied *in vivo* as well.[90,96−98] Likewise, PDMS wristbands and new approaches such as Fresh Air wristbands and PDMS foam disks are likely to become more common in longitudinal NTA of personal, ambient, and indoor air contaminants for population exposure science, molecular epidemiology, and precision environmental health.[99−103] Meanwhile, novel, automated sample handling systems are being developed to link such minimal invasiveness (of microsampling) with scalability and throughput, as represented by dry blood spots,[104] urine stripes, and volumetrically accurate microsampling (VAMS) collection devices.[105−107]

3.2. Sample Pretreatment: Extraction, Cleanup, and Derivatization. Sample pretreatment procedures, including matrix normalization (e.g., creatinine for urine, hemoglobin for dried blood spots),[108] solvent extraction, and additional modification (e.g., enzymatic treatment, purification, derivatization), are most influential in determining the chemical space measured by MS-based assays.[69] Multiple strategies exist for exposomics to balance coverage, sensitivity, and consistency while countering interferences, whether it be targeted quantitation, NTA, or both. Partitioning or fractionation techniques, with wide-ranging solvent/sorbent choices, separate analytes into disparate portions/fractions by their physicochemical properties (e.g., lipophilicity, volatility, aqueous solubility, and pH) and/or elemental/functional specificity (e.g., halogenated, amines, glucuronidated).[67,69] These reduce co-eluting interferences and improve measurement specificity for known molecules but bear a risk of analyte loss and biased profiling for unknown chemicals, including many trace-level pollutants.[67] Thus, to achieve comprehensive chemical coverage, it is necessary to test and combine various extraction approaches. The cautions lie in the sampling−pretreatment design at the front end, knowledge/training of the operator who handles the samples (to avoid contamination and analyte loss), and effective validation for combinatory use.[109] Critical under-discussed considerations include how to select chemical standards as the appropriate proxy for appraising NTA, how to go beyond experimenting with a limited number of standards (from a few dozen to hundreds) as routinely adopted in current studies, and how to take *de novo* computational approaches to mining the non-targeted data directly for such prioritization and validation.[110]

Specific pretreatment techniques include protein precipitation (PPT), dilute and shoot (DNS), solid−liquid or liquid−liquid extraction (SLE or LLE), solid-phase extraction (SPE), dispersive solid-phase extraction (d-SPE, i.e., QuEChERS, short for "quick, easy, cheap, effective, rugged, and safe"), thermal desorption, and accelerated solvent extraction (ASE).[25,67,91,111−114] Purification may be needed, leveraging specific sorbents for phospholipid removal (PLR) and/or elimination of other interferences. Chemical filters to this end have proved effective recently, spanning OstroPlates, Phree, and Isolute PLD (96-well plates),[31,113,115,116] zirconia-based sorbents like HybridSPE (SPE cartridges),[114,117] and EMR-Lipid (d-SPE and SPE formats).[30,118] Complementary use of these techniques may expand exposomics coverages. For instance, a recent study reported only 43−54% of total ion features as overlapping between sample preparation approaches based on PPT and PLR plates, respectively, indicating the need for combining both methods.[31]

Consideration of the matrix effect remains a key factor for increasing the sensitivity and selectivity of analytical techniques. A matrix effect is any influence that the substrate (e.g., tissue, blood, water, or solvent) has on the analytical performance of a technique. This is typically characterized by ion suppression or ion enhancement that hampers detection and quantitative accuracy for given analytes of interest, likely due to the presence and concomitant ionization of coexisting molecules and/or overlapping signals of interferences in the matrix.[119,120] Strategically, the matrix interference may be

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

decreased by dilution, better cleanup (removal of interference), better chromatography (separation of interference from the analyte), and/or better detection (higher selectivity for the target analyte).[119,121]

The matrix effect hampers the detectability of low-abundance chemicals and analytical reproducibility in exposomics. Balancing its reduction with minimal analyte loss thus represents one primary goal in sample preparation. Generic approaches such as DNS are preferred due to minimal analyte losses incurred and have been applied successfully in detecting drugs, mycotoxins, and pesticides using LC-MS.[122] The DNS concept goes beyond liquid samples (e.g., urine, saliva) and applies to more complex matrices (e.g., blood, tissues) for which an extra SLE or LLE step is needed upfront to trigger off analyte transfer from matrix to liquid phase (before dilution).[122] For DNS, a dilution factor of 1:50 might be considered high with demonstrated benefits for certain matrices/compounds but can induce significant sensitivity loss in detecting other chemicals (e.g., pesticides) without mitigating matrix effects further.[77,123] DNS has also been applied to GC-MS analysis as the "dilute, evaporate, and shoot" approach, which is commonly used for the analysis of biological specimens like blood.[122] However, this fundamentally weakens the premise of minimal analyte loss. To resolve this, Hu and co-workers (2021) proposed a balanced and straightforward approach that selectively combines sample dilution, LLE, and QuEChERS-based cleanup (collectively termed "express liquid extraction," XLE) with high coverage and minimal recovery variability (for a range of matrices) in both targeted and NTA modes of GC-HRMS exposomics.[25]

In GC-MS analysis, a crucial consideration involves whether and how derivatization techniques are used—a process that modifies analyte structures for improved volatility, chromatographic separation, and/or detection.[124] Of note, GC-MS-based metabolomics often uses derivatization, such as trimethylsilylation to prevent the breakdown of carboxyl, hydroxyl, and amino groups of biomolecules.[125] For GC-MS-based exposomics, this has not been explored. How can specific and selective derivatization fit within an NTA framework? This may challenge the throughput since exposomics demands a balance of analyte recovery when covering both environmental chemicals (many are volatile, nonpolar, and bioaccumulative) and their transformation products/metabolites (likely more polar and fragile). Further, how can potential complications be avoided as resulting from the generation of partial or unwanted derivatives, artifacts, and multiple derivative products for the same species (e.g., multiple hydroxyl groups)? These are important questions to address in GC-based exposomics.

3.3. Pre-MS Separation: Chromatography and Ion Mobility Spectrometry. Efficient analyte separation before mass spectral detection is crucial for navigating the competing analytical goals of coverage and throughput. Compared to direct-injection modes, chromatographic separation delivers analytes into mass spectrometers slowly and steadily over time, minimizing ion suppression and source fouling. Correspondingly, chromatographic retention enables reliable peak integration for confident quantitation and offers an orthogonal metric for compound identification, with additional advantages arising from the time scale of separation: a chromatograph retains and resolves analytes in the seconds (*s*) scales, thus allowing incorporating in between the chromatographic step and a fast-scanning mass spectrometer (e.g., ToF, in microseconds, μs) additional separation modules such as ion mobility spectrometry (IMS, typically in milliseconds, *ms*).[126]

Chromatographic separation is achieved by moving a mobile phase (gas or liquid) that carries the analytes through a stationary phase-fixed system.[127] In partition chromatography (e.g., GC and LC), for example, analyte mixtures are vaporized/dissolved into mobile phases; chromatographic separations (i.e., differentiated retention on the stationary phase) are accomplished by gradient changes (alternatively, an isocratic setting) over the runtime (e.g., through changes in mobile phase polarity for LC and temperature for GC). Other chromatographic mechanisms include adsorption, ion exchange, size exclusion, as well as the more selective affinity chromatography (ligand reagents such as enzyme inhibitors or antibodies) and chiral chromatography (with stationary phase or mobile phase made chiral).[127] Non-chromatography techniques also exist, such as IMS[128] and capillary electrophoresis (CE).[129] This article focuses on partition chromatography (i.e., LC, GC) and IMS, considering their contribution to small molecule analysis as demonstrated in metabolomics and targeted environmental monitoring.

GC and LC are complementary in chemical space coverage of the exposome, as are their respective commonly implemented ionization techniques.[68] GC, paired to electron ionization (EI), captures relatively nonpolar (more bioaccumulative), volatile/semivolatile, and thermostable substances, including many hydrophobic organic chemicals (HOCs) like polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), organochlorine pesticides (OCPs), phthalates, and VOCs.[68] With versatile derivatizations, GC-MS has proved effective not only in metabolomics of primary metabolites including sugar, fatty acids, and amino acids,[130,131] but in targeting certain exogenous chemicals as well, such as UV-filters (additives in PPCPs) including benzophenone, pharmaceuticals, parabens, and phenols (e.g., bisphenols, alkyl- and halogenated phenols).[124,132] LC, coupled to electrospray ionization (ESI), widely applies to nonvolatile, relatively polar/hydrophilic, and thermally unstable chemicals, e.g., pesticides, prescription and illicit drugs, and mycotoxins.[133,134] LC-MS is noted for its speed, precision, and capability in the unequivocal detection of trace molecules *in vivo*, as demonstrated by the screening of doping or illicit drug use in equine or human athletics.[135,136] Most LC applications are based on reverse-phase LC (e.g., C_{18} and C_8). However, there are emerging alternatives like pentafluorophenyl (PFP) stationary phases and hydrophilic interaction chromatography (HILIC) for better retention of relatively polar species or unique isomeric selectivity.[68] Recent trends in chromatography include miniaturization (e.g., toward microLC/nanoLC), multidimensional (e.g., GC × GC), and parallel chromatography (featuring dual injector or fast polarity switching scans).[137,138] On a broader scale, these innovations offer added flexibility to balance efficacy, throughput, and cost-effectiveness in exposomics for many years to come.

In HRMS-based exposomics, LC-HRMS and, more recently, GC-HRMS are increasingly being used.[44] For biomonitoring, a recent review surveying 124 existing HRMS studies identified that 95 used LC-HRMS and 28 used GC-HRMS.[44] Within the LC/GC category, Orbitrap (n = 49) and ToF (n = 46) supported LC analysis equally, whereas for GC, magnetic sector GC-HRMS (n = 16) far outnumbered GC-ToF (n = 8) and GC-Orbitrap (n = 4) combined.[44] Note that sector GC-HRMS is designed for high sensitivity and selectivity with only

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

a borderline high mass resolution (e.g., 10,000 by full width at half-maximum, fwhm at m/z 322, as offered by the Thermo Scientific DFS model); in fact, all 16 sector GC-HRMS analyses were targeted quantitation of trace pollutants such as dioxin and illicit drugs.[44] NTA by GC-HRMS (Orbitrap or ToF), on the other hand, remains underexploited but has shown potential through successes in environmental monitoring (e.g., water, air, dust, and soil),[139] food safety assessment,[140] and biomonitoring.[25,141−143] Concerted efforts are warranted to bridge knowledge and technical gaps between these two platforms and, importantly, to expand the chemical space GC-HRMS covers. Technical specifics of feasibility and cost-effectiveness have been discussed for column choices, dimension/scale, MP modifiers, and operating temperature.[69]

Ion mobility spectrometry, or IMS, is an emerging technique incorporated into modern GC/LC-HRMS setups.[126,128] IMS separates gaseous ions by size, shape, and charge state through colliding with inert buffer gas under a guiding electric field. Similar to GC, IMS handles ions in the gaseous phase and can thus be readily coupled to MS.[126] IMS can be either (*i*) time-dispersive, such as drift-tube ion mobility spectrometry (DTIMS) (e.g., Agilent 6560 IM-Q-ToF) and traveling-wave ion mobility spectrometry (TWIMS) (e.g., Waters Synapt G2-*Si* and Structures for Lossless Ion Manipulation [SLIM] which provides high-resolution, lossless separation of ions at a low cost),[128] (*ii*) space-dispersive such as differential mobility spectrometry (DMS) (e.g., Sciex TripleTOF 5600+), or (*iii*) based on ion confinement (trapping) and selective mobility release such as trapped ion mobility spectrometry (TIMS) (e.g., Bruker timsTOF).

Three major advantages of incorporating IMS into HRMS instrumentation are (*i*) added orthogonal separation without compromising throughput, which benefits large-scale population studies,[126] (*ii*) complexity reduction in the MS^1 and MS/MS spectra (minimal chimeric spectra), and (*iii*) the use of collision cross section (CCS) as an additional property for compound identification. First, the data acquisition time frame in IMS separations (*ms*) is easily nested between that of chromatography (*s*) and mass analyzer (*μs* to low *ms*), offering respectively an orthogonal and semi-orthogonal separation for expanding the exposomic coverage.[126,144] In principle, IMS measurements such as drift time (in DTIMS) and CCS do not depend on mobile phases nor encounter cross-batch shifts much like chromatography does. In practice, however, reproducibility tests are still needed to counter unintended variations (e.g., due to improper implementation); aligning LC-IMS-MS data in this regard can be even more challenging than LC-MS. Further, CCS, a characteristic of the analyte's size and shape (given a specific counter gas), can be readily compared between laboratories and across instruments with proper and as-needed calibration.[145−147] *In silico* prediction of CCS (from structure) is more accessible than retention time; the latter depends on column (and possibly column batch), mobile phases, gradient, and many other physics/chemistry factors at play, inherently challenging harmonization and modeling across locations and instruments. Because of these, public libraries for experimental and predicted CCS have proliferated in recent years, covering compound classes amenable to both GC and LC.[146,148,149] The improvement in annotation rates and exposomics accuracy should not be understated when implementing CCS values.[150,151]

Over the past decade, IMS has advanced metabolomics, notably the subfield of lipidomics. IMS can distinguish lipid isobars/isomers and complex species that have been long challenging chromatography and mass spectrometry.[152] IMS also promotes the idea of "pan-omics," where biomolecules of different classes (e.g., nucleic acids, proteins, metabolites) are resolved within a single run without sophisticated fractionation upfront.[128,153] Besides enhanced peak capacity, IMS yields CCS data which can help uncover new environmental chemicals such as PFAS and other xenobiotics from the "dark metabolome."[154] However, it should be noted that the capability of resolving two adjacent peaks, as provided by these techniques, is generally the greatest for GC, followed by LC, and then IMS, of which the added separation may not be considered entirely orthogonal to chromatography. In resolving isomers, for example, lipids, although certain species have been reported to be distinguished on IMS (with distinct CCS) but not on chromatography (close affinity for stationary phase),[155,156] specific lipids of the same class (e.g., diacylglycerols and phosphatidylcholines) with varying acyl chains and double bond position may remain incompletely resolved on IMS but are baseline-separated on LC.[157] One IMS exception is SLIM, which has proven effective in resolving isotopologues (although depending on the specific separating path length).[158] For NTA, one should note that IMS separation occurs after ionization. When used alone, in-source ion suppression is not circumvented, limiting sensitivity while challenging (semi)quantitative analysis in IMS-assisted exposomics.

3.4. Ionization. Chromatography/IMS-differentiated analytes must be converted into charged gaseous species (in an ionization chamber) for sequential mass analysis and detection.[159] Ionization, therefore, is one of the most influential factors for compound coverage/selectivity, sensitivity, and annotation efficacy. Since the 1940s when EI was first introduced, mass spectrometry ionization methods have experienced phenomenal developments. These go beyond EI to relatively softer alternatives (e.g., chemical ionization, field ionization, photoionization) and the Nobel Prize-winning ESI and matrix-assisted laser desorption/ionization (MALDI).[160] The latter two techniques are considered "soft," allowing ionization of wide-ranging fragile molecular species, and have since revolutionized nearly every scientific discipline.

In modern MS analysis, EI has been a routine for GC-MS. Commonly applied at 70 eV, EI induces extensive fragmentation of the molecular ion ($M^{+\bullet}$).[161] The 70 eV EI spectrum has been reproducible, largely independent of specific GC-EI-Q-MS instruments in use, operators, and locations of analysis, promoting the idea of a "universal" reference spectral library.[161] However, challenges remain for broad GC-EI-HRMS profiling. For one thing, gas-phase formation of water adducts of EI cations, especially those highly labile species, was found prevalent in C-Trap compartments.[162] This presents an obstacle for trimethylsilyl (TMS) derivatives of fatty acids and native purines (e.g., alkaloid-like drugs), limiting the use of available unit-mass EI libraries.[162] For another, the hard EI inherently challenges NTA as molecular ions are not always present, and ion fragments are low-abundant with limited improvements observed thus far at lower EI energies (with possible platform- or source-specificity).[163,164] Given the low abundance of exogenous chemicals in biological samples, the presence of molecular ions is desirable, especially in coelution cases where ion detection of these can be masked by ion fragmentation of high-abundant molecules.[165] The emerging cold EI[166,167] may be one

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

solution; one study found that cold EI curbs in-source EI fragmentation and enhances annotation when applying a supersonic beam, reduced electron energy (to 18 eV), and lowered helium pressure, although these had only been tested on low-resolution instruments.[168]

Softer alternatives, e.g., chemical ionization (CI) that ionizes analytes through gentle proton transfer via chemical reagents, are increasingly used in GC-HRMS to generate intact molecular ions consistently, protonated or deprotonated.[142] This is necessary for quantitation and compound identification, given the trade-offs between ion source choices for GC-HRMS. Manz and co-workers (2023) reported that EI was always used and only occasionally complemented with CI (11% of all studies) in exposomics, highlighting a need to bridge such gaps for characterizing environmental molecules.[46] For sensitivity that can be compound-dependent, CI sensitivity irrespective of specific mechanisms (+/−) is generally an order of magnitude lower than EI, with the latter almost exclusively used in GC-MS quantitation.[169] For compound identification, EI produces more stable and consistent spectra than softer alternatives and thus makes spectral matching more confident and reproducible (as compared to CI and ESI). Still, the lack of molecular ions in EI presents a fundamental challenge to structural elucidation.

Capellades and colleagues (2021) explored different CI reagent gases (methane and isobutane) and compared them to EI for metabolomics by GC-HRMS.[170] The use of isobutane was discovered to prompt the $[M + H]^+$ isotopic envelope, facilitating the detection of isotopic enrichment in contrast to methane which induced unwanted $[M + H]^+$ fragmentation; no significant decline in sensitivity was observed for CI-isobutane.[170] Meanwhile, the authors determined that a low-energy EI (15 eV) still promoted greater fragmentation of $M^{+\bullet}$ than CI-isobutane.[170] Misra and Olivier (2020) compared EI-MS and CI-MS^2 (methane as reagent gas) in GC-Orbitrap HRMS.[142] Results showed that, of the spectra acquired for chemical standards of metabolites, roughly half (171 out of 330 GC-amenable compounds) were recorded by both EI-MS and CI-MS^2 (combining PCI and NCI modes).

Besides vacuum-assisted ionization techniques (e.g., EI and CI), atmospheric pressure ionization (API) has become increasingly popular for GC-MS, as represented by new GC-API-MS platforms incorporating plasma, laser, atmospheric pressure chemical ionization (APCI) or atmospheric pressure photoionization (APPI).[171,172] GC-API-MS instrumentation induces soft ionization that preserves the molecular or quasi-molecular ion and thus improves the detectability, selectivity, and precision over EI or CI sources.[171] Further, they can ionize a broader range of compounds than EI.[173] Thus, complementing EI with API for GC-HRMS may be conducive to enhancing NTA confidence, with technical specifics discussed elsewhere.[171,174] Recently, GC-APCI-IMS-ToF MS has shown the advantage of APCI (alongside IMS-CCS) in facilitating compound identification, especially for halogenated organics.[175]

For LC, soft API techniques are widely used for small molecules (e.g., ESI, APCI, and APPI). The ESI relies on solution chemistry, where analyte ions are believed to form in solution before in-chamber nebulization, desolvation, and ion evaporation.[176] LC-APCI requires that analytes turn gaseous for ionization to occur, passing LC eluents through a heated ceramic tubing to create a fine spray (i.e., nebulized and fully vaporized) and form protonated/deprotonated ions in contact with reagent/solvent vapor released from a corona discharge needle.[173] LC-APPI experiences the same nebulization as APCI but uses an ultraviolet lamp instead of a corona needle for ionization; additional mobile phase solvents/modifiers ("dopant") are usually added to assist with photoionization.[173] Interestingly, although LC-APCI and LC-APPI are better suited for low-polarity molecules (relative to LC-ESI), they are typically less sensitive than LC-ESI.[177,178] In addition, while soft ionization in LC more likely retains parent ions, such softness may result in less reproducible and, depending on the specific analytes, insufficiently fragmented spectra for qualitative and quantitative analyses.[159] To mitigate this downside, enhanced in-source fragmentation (EISA) techniques, among others, are being devised to improve compound annotation confidence and quantitative sensitivity (combined with tandem mass spectra when available) even in LRMS.[179−181]

The ionization mechanism remains to be elucidated for differing response factors of individual compounds with respect to its co-eluting species, that is, the matrix effect. Thus, ionization methods need to be tested or validated for the overall best coverage, sensitivity, and annotation confidence. Recent studies have imparted useful insights. Since the mid-2010s, the U.S. EPA has launched the non-targeted analysis collaborative trial (ENTACT) for systematic assessment of GC- and LC-MS on coverage and sensitivity using authentic chemical standards. One ENTACT study tested disparate ionization techniques, specifically comparing APCI (+/−) and ESI (+/−) for LC-HRMS analysis of 1,264 chemical standards (i.e., the ENTACT mixture).[182] Results showed that 1,116 were detected in at least one mode, while only 185 were detected in all four modes. Substructure enrichment analysis based on the ToxPrint sets[183] identified relatively hydrophilic substructures (e.g., alcohol moieties) as exclusively enriched in ESI data, whereas the more nonpolar naphthalene group clustered in APCI only. Relative to ESI, APCI data had less background with added chemical space coverage, suggesting that the two methods are complementary and together contribute to a broader coverage in LC-HRMS NTA.[182] Ring-trial studies as such are essential for benchmarking ionization methods for expanding the chemical exposome coverage.

3.5. Mass Analyzer: Mass Resolution, Sensitivity, and Scan Speed. The heart of any mass spectrometer is its mass analyzer, an essential modular component to separate, modify, and detect analytes by their mass-to-charge ratios (m/z).[159] Mass analyzers apply known electric and/or magnetic fields to the gaseous ions under an ultravacuum environment (1×10^{-3} to 1×10^{-10} Torr) to impart ions kinetic energy and momentum, respectively, and analyze the resultant motions of these ions being differentiated in time and/or space.[159] MS analyzers can either be low resolution or high resolution (10,000 or higher by fwhm), respectively conferring unit-mass measurement (e.g., quadrupole, linear ion trap, or LIT) and accurate-mass measurement as achieved by ToF, Orbitrap, or Fourier-transform ion cyclotron resonance (FT-ICR). HRMS basics and instrumentation are detailed elsewhere.[159,184,185] As the high-resolution accurate-mass (HRAM) capacity continues to expand and meet demands in small molecule analysis, under the umbrella term of "high-resolution," distinction can be drawn for HRMS between a borderline high mass resolution (10,000−50,000 fwhm, 3−10 ppm mass accuracy) and ultrahigh mass resolution (>50k fwhm, <3 ppm mass accuracy)

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

for deriving meaningful formula with minimal mass interferences.[186,187]

Characteristics for assessing analyzer performance include mass accuracy, mass resolving power ("resolution" refers to specific measurements), mass range, transmission, scan speed, and tandem mass capability.[188] In practical use, a trade-off between mass resolution, sensitivity, and scan speed will occur. On a given HRMS setup, scan speed affects mass accuracy, mass stability, and sizes of data files, and thus is key to acquiring good spectral data, qualitative and/or quantitative. High-speed scanning in full-scan mode is desired to derive meaningful quantitative integration and definition of co-eluting peaks, since slower scan speeds lead to data loss, resulting in less clean mass spectra and poorly "resolved" chromatographic peaks. In the actual sample analysis, more stringent cutoffs are encouraged when setting scans/peaks to counter matrix complexity.[189] Since each mass analyzer has its unique pros and cons, a hybrid configuration that enables mixed modes of analysis will provide a new solution. One example is the extended mass range conferred by Q-ToF: a quadrupole delivers a constant peak width across mass while its resolving power varies with it; in contrast, ToF maintains a constant resolving power almost independent of mass, but the peak width is mass-dependent.[190] Other proof of concept endorsing a hybrid use include QqQ (for MS/MS, or MS^2) and quadrupole linear ion trap (QqLIT) (for MS^n ion tree, n up to 10, theoretically),[191] both of which have advanced targeted sensitive analysis for decades.

Modern HRMS systems use a hybrid configuration often through coupling a fast, selective, and low-resolution mass filter like quadrupole (Q) (for precursor selection, if needed) sequentially to a sensitive, accurate-mass detector (e.g., Orbitrap) to enable flexible data acquisition modes.[192,193] ToF and Orbitrap are the two most popular high-resolution mass analyzers, with technical nuances reviewed on the fundamental design and omics application.[185,194] By design, both analyzers impart considerable kinetic energy to prompt ion injection, sequentially followed by analysis of the ion motion in a gentle, non-electromagnetic space—ToF separates ions in a field-free drift region by their time-of-flight, while Orbitrap, the only new MS concept developed in the recent 30 years, traps ions in an electrostatic field (achieved by imposing high voltage) and determines m/z from its own resonant/oscillation frequency. The current Orbitrap HRMS, with the "high-field" design and Fourier-transform (FT) signal processing, offers an ultrahigh mass resolution that can approach FT-ICR MS.[185,195−197]

In principle, ToF produces essentially the same mass resolution over an entire mass range and across all scan speeds, whereas, for Orbitrap, the mass resolution is not only inversely proportional to scan speed but related to the specific m/z ($\propto$ sqrt$(1/(m/z)$, in one Orbitrap scan).[185] Compared to Orbitrap, the ToF analyzer scans fast and covers a wide mass range with no theoretical upper limit. However, ToF often encounters a limited (intra-scan) dynamic range for detecting trace-level compounds; possible causes include fast digitizers, the design of microchannel plate (MCP) detector, and a chemical background that commonly occurs when coupled to ESI or MALDI.[198] The interesting "chemical background" issue draws certain distinctions between ToF and FT/Orbitrap MS.[185] Granted that for ToF, it could be a factor limiting detection thresholds and dynamic ranges, FT analyzers (e.g., Orbitrap) are relatively free of such background, offering comparable or lower detection limits than ToF.[185] On the fundamental level, this is because ions need to remain intact within the FT analyzer for an extended period of time (e.g., many milliseconds), causing all stray or metastable ions either not FT-detected or to collectively form a broad, smooth background that can be readily subtracted from data.[185]

3.6. Mass Spectral Data Acquisition. Through innovative hybrid designs, HRMS enables versatile data acquisition modes to address wide-ranging analytical needs. Dating back to the 1960s, the conception of tandem mass spectrometry (MS/MS) first opened the door to selective, in-depth analysis of specific ions by the collision-induced dissociation (CID) mechanism.[199,200] Such setup of two spectrometers separated by a collision chamber (QqQ for MS^2 or QqLIT for MS^n), albeit unit-mass capacity only, offers unparalleled sensitivity and selectivity and has since served for decades as a powerhouse for targeted quantitative analysis of trace-level compounds in complex matrices.[120] In the HRMS era, the combinatory use of fast mass filters and high-resolution mass analyzers allows alternating full-scan and the scan of product ions (as resulting from fragmenting precursor ions), selective or nonselective, to yield quantitatively meaningful data with adequate scans/points (preferably $\geq$10) across an extracted ion chromatogram (EIC or XIC) even within a short cycle time interval. Such breakthrough is a prerequisite for enabling simultaneous exposome-wide profiling (quantitative) and high-coverage structural annotation (qualitative) with good throughput and scalability.

One most common data type for diagnostic confirmation or structural elucidation is tandem mass spectrometry, i.e., MS/MS (or MS^2), which consecutively implements m/z selection, fragmentation in a collision cell, and scanning of product ions, either preselected (e.g., selected/multiple reaction monitoring, SRM/MRM) or via an unbiased survey scan (e.g., parallel reaction monitoring, PRM).[201,202] With ion trap analyzers, the sophisticated ion tree approach (MS^n, n up to 10, theoretically) may come into play if more elaborate, in-depth structural analyses are intended.[203] For modern Q-HRMS instruments, significant advances have been made in sensitivity and acquisition speed, obtaining MS^2 spectra through data-dependent acquisition (DDA), data-independent acquisition (DIA), and beyond.[204] One driving force behind such advance is (shotgun/"bottom-up") proteomics, especially for DIA.[205,206] Proteomics data are acquired by alternating survey scans (of all precursor ions) with tandem mass scans that entail fragmenting select (peptide) precursor ions and scanning the resultant product ions.

The term "data-dependent" in DDA means that the MS selects specific ions over the others, typically the most intense/abundant ones (e.g., "top n"), for fragmentation and tandem mass analysis. While, DIA-MS^2 seeks to acquire complete and unbiased MS^2 data, fragmenting all ions possible (e.g., simultaneously or over sequential mass windows) regardless of ion abundance or structural characteristics.[204] For exposomics, it should be noted that DDA-MS^2 is inherently unsuitable for NTA, since it is biased toward the MS selection and lacks sample-to-sample reproducibility (e.g., owing to likely stochastic ion selection across samples/injections). In contrast, DIA-MS^2 aims to provide complete chemical coverage regardless of ion abundance or characteristics to serve NTA goals. Nonetheless, advances in both DDA and DIA have been primarily limited to the MS^2 level, leaving ion beam sampling for MS^1 scans inefficient. One emerging data

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

acquisition method, BoxCar,[207] boosts MS^1 sensitivity by filling multiple narrow m/z segments for a single scan, reaching 10-fold increases on quadrupole−Orbitrap MS in the mean ion injection time compared to a standard full-scan. This approach has been tried primarily in proteomics but recently demonstrated use for small molecule (amino acids) analysis,[208] thereby providing a potential boost to exposomics where co-eluting low-abundance analytes present a long-standing analytical challenge.

The DIA approaches are benchtop-accessible, from MS^E (Waters Q-ToF),[209] All Ions Fragmentation (AIF, Agilent Q-ToF),[210] SWATH ("sequential window acquisition of all theoretical mass spectra") (Sciex, TripleToF),[211] and DIA (Thermo, Q-Orbitrap),[212] to dia-PASEF (short for "parallel accumulation serial fragmentation") (Bruker, timsTOF).[213] DIA-MS^2 technologies continue to evolve and meet analytical demands across fields. For example, recent data showed DIA-MS^2 of the latest Orbitrap Astral (i.e., Asymmetric Track Lossless) MS model (Thermo) quantifies 5 times more peptides (per unit time) than the gold-standard Orbitrap MS, potentially furthering quantitative proteomics.[214] For small molecule analysis, full-scan, DDA, and DIA were compared; a trade-off was observed comparing MS^2 of DDA (better but fewer/biased spectra) vs of DIA (more spectra but with slightly lower quality), highlighting the potential of DIA-MS^2 to boost NTA.[215] The challenges lie in acquiring quality DIA-MS^2 data and deconvolution of these data—a demultiplexing algorithm to bridge precursor ions and fragment ions by *de novo* reconstruction of MS^2 for respective individual precursor ions.[216] Low-quality MS^2 with unreliable fragment intensities or chimeric peaks (due to missingness or artifacts) might be attributed to instrumental noise (overshading low-abundant analytes) and/or co-eluting ions (due to sample complexity).[216] Emerging solutions have been proposed, including a Bayesian approach that computes cumulative neutral losses to clean up DIA spectra *post hoc* with or without the time domain for fragment deconvolution.[217]

CID has been the dominant dissociation mechanism for ion fragmentation since the 1960s, colliding accelerated ions with a neutral inert gas (e.g., N_2, He, Ar) to induce bond cleavages in molecules, usually at sites of weakest bond energies and/or most convenient rearrangements.[159,199] On top of CID, high-energy collisional dissociation (HCD), now commonly employed in Orbitrap MS, applies high-energy electrons that allow even more extensive ion fragmentation.[218] Methods complementary to collision-activated dissociation include electron−ion reaction-based dissociation (ExD) and proton-based activation.[219] ExD methods induce radical-driven ion fragmentation at selective sites through various mechanisms, including electron-capture dissociation (ECD), electron-activated dissociation (EAD), electron transfer dissociation (ETD), and the emerging "electron impact excitation of ions from organics" (EIEIO).[220−222] As for proton-based activation and dissociation, ultraviolet photodissociation (UVPD) stands out, energizing ions via the absorption of high-energy photons.[221] While ECD and ETD have been widely applied in proteomics to preserve fragile moieties (e.g., post-translational modifications), UVPD has recently garnered interest for resolving co-eluting lipid/sterol isomers at various structural levels that have long challenged chromatography.[221] In parallel, computational data pipelines are being developed for spectral curation and reproducible annotation. One notable example is LibGen (2023), an automated pipeline to generate high-quality reference MS/MS spectral libraries for EAD-, UVPD-, and HCD-based HRMS using natural product standards for a showcase.[223] The LibGen pipeline corrects mass errors, denoises spectra through subformula assignments, and computes both spectral entropy and the explained intensity for quality control.[223]

The burgeoning Q-HRMS capabilities now allow flexible workflow development to tackle various analytical challenges in exposomics, from suspect screening to NTA of complex mixtures. Trends have shifted from the routine use of DDA-MS^2 (or dd-MS^2) to more DIA-MS^2-oriented applications over the past five years.[44,56] Recently, a Non-target Data Acquisition for Target Analysis (nDATA) approach has been developed using LC-HRMS.[224−226] The LC-nDATA workflow runs both full-scan (FS) and DIA-MS^2 scan and has succeeded in screening pesticides and their metabolites (n > 1,000) in foods and humans.[226] Specific DIA methods were evaluated on UHPLC-Q-Orbitrap MS, including variable DIA (vDIA) and multiplex DIA (mDIA).[227] Compared to vDIA (operationally FS-HRMS/vDIA-$HRMS^2$), mDIA (operationally FS-HRMS/mDIA-$HRMS^2$) proved comparably effective for high-throughput pesticide screening based on four measures for library search, namely MS^1 accurate mass, MS^1 isotopic pattern, chromatographic retention time, and characteristic MS^2 fragment ions. More DIA-MS^2-based biomonitoring should be conducted, given accumulative environmental monitoring successes for trace-level pollutants.[228] Notably, DIA-MS^2 has been rapidly expanding in LC-HRMS NTA but remains incipient for GC-HRMS, primarily due to the already extensive in-source fragmentation in EI. Nonetheless, DIA-MS^2 can be valuable for advancing GC-HRMS NTA when softer ion sources (e.g., CI) are benchmarked and applied.[44,46]

4. TOWARD MERGING TARGETED AND NON-TARGETED APPROACHES

For etiologic studies of the *Genome* × *Exposome* interplay, exposomics must emulate genome sequencing, requiring that both the laboratory measurement and data pipelines are streamlined and harmonized to accommodate the needs for scalability, accessibility, and throughput. Exposomics widely characterizes small molecules and has drawn inspiration in a certain way from MS-based metabolomics. Such cross-disciplinary fusion and differentiation, nonetheless, has led to confusion and a disconnect among fields/communities in conceptual understanding, formulation of glossaries/nomenclature, and workflow standardization. There is ambiguity in terminologies such as (*i*) target, suspect, non-targeted screening (identification focus) vs quantitation (absolute or relative), (*ii*) an exposure [event] and an exposure [agent/factor], and (*iii*) the definition of biomarker, specifically, to clarify a marker of exposure agents vs a marker of exposure-induced biological response. Some of the terminologies have been addressed in prior works.[15,229] This article will address terms related to analysis.

There has been notable confusion between targeted and non-targeted approaches in exposomics. On a granular level, such divergence is deeply rooted in the inherent dimensional challenges faced by exposomics and metabolomics—both handling up to millions of small molecules.[51,230] From a chemical analysis standpoint, this contrasts other upstream branches of the "omics" cascade. For example, genomics and proteomics sequence only a limited number of nucleotides (n = 4) or amino acids (n = 20), respectively, although there is a

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

unique spatial/conformational component to it (e.g., chromosome maps in genomics; post-translational modification in proteomics). Such wide gaps between analytical/budget limitations and a demand for high chemical space coverage result in trade-offs in exposomic practices and indicate a need for a tiered approach. Common chemicals with commercially available standards, not surprisingly, are favored in analysis over less familiar chemical spaces, which may be addressed through *de novo* approaches such as heuristic rules or machine learning.[48,51,230] The following sections discuss some pressing concepts/terms in HRMS-based chemical exposomics while stressing a need to merge targeted and non-targeted approaches.

4.1. Targeted, Suspect, Non-targeted Screening vs Quantitation. Chemical analysis aims to characterize molecules by means of qualitative ("What is the compound's chemical identity, i.e., elemental composition and structure?") and/or quantitative ("How abundant is it, or what is its concentration in the sample?") information. The advent of hybrid HRMS featuring rapidly evolving data acquisition techniques has opened up new opportunities for characterizing the chemical component of the exposome. In practical use, a tiered approach is commonly applied, from screening (e.g., targeted, suspect, or non-targeted) to quantitation (e.g., absolute or semiquantitative), which necessitates conceptual clarification.[229,231] Here, "semiquantitative" means using peak intensity (height or area) as a direct readout or relative to quantifiable compounds (e.g., spike-in internal standards) of ion/analyte abundances for data analysis rather than absolute concentration levels in the sample. The differentiation between targeted and non-targeted approaches lies in both data acquisition and data analysis (compound annotation and quantitation). Targeted data acquisition operates selective scanning of the MS^1 profile. The target ions are preselected and can be specific for known structures or substructural diagnostics. Non-targeted data acquisition, however, runs a nonselective MS^1 profiling, yielding data not limited to precursor ions defined *a priori*. Both approaches can provide qualitative (i.e., the presence of an analyte, for confirmatory or screening purposes) and/or quantitative information (absolute or semiquantitative). Compared to targeted counterparts, non-targeted approaches enable the detection and annotation of analytes not fully defined at the front end.[83]

The analytical design is context-specific and would depend on analysis goals, platform capability, sample availability, and budget limitations. For long-term exposomics success, high analytical coverage with reproducibility and throughput is crucial. This would depend on the harmonization, standardization, and merging of distinct analysis layers spanning workflow design, experimental implementation, data integration, and reporting results. Efforts to minimize or eliminate analytical variations (e.g., batch effects) and to facilitate such workflow merging are lacking but may benefit from sizable interlaboratory ring trials on national and international scales, as demonstrated in those for metabolomics, lipidomics, as well as ENTACT in recent years.[137,232,233]

The tiered analysis modes are tied to practical restrictions and ease of detection that favor certain chemicals over others for a given sample. For a first pass, targeted analysis (i.e., screening and/or quantitation) relies on authentic chemical standards or information on *known* compounds gathered under the same methods, generating confident data to validate platform efficacy while serving as a "positive control" before delving into NTA. Consensus for such confirmatory detection of known analytes in targeted screening and quantitation has been reached in criteria like the number of ions for confirmation, acceptable limits for quantitative precision, etc.; what remains undecided, however, is the acceptable detection rates per method.[234] NTA assessment is equivalent in this manner—due to a lack of ground truth, ring trials are required that compare known/spiked analytes in the sample.

NTA has been rapidly expanding and critically appraised in recent years. Notable endeavors include (*i*) the EPA's Non-Targeted Analysis Collaborative Trial (ENTACT),[235] (*ii*) the "Best Practices for Non-Targeted Analysis" (BP4NTA) working group that has created the Study Reporting Tool (SRT) for evaluation of reporting,[236,237] and (*iii*) NORMAN,[238] i.e., "network of reference laboratories, research centers and related organizations for monitoring of emerging substances" that has been continuously devising prioritization frameworks and resources for environmental monitoring over the past decade. The ENTACT comprises large-scale ring trials among ∼30 laboratories to evaluate cutting-edge NTA methods (largely HRMS-based), utilizing synthetic standards mixtures (of hundreds to ∼1,000 compounds) and multiple standardized media (e.g., human serum, house dust, and silicone bands). The first of its kind, ENTACT systematically compares GC-MS vs LC-MS techniques on coverage and sensitivity. While ENTACT strives to produce data testing on the performances of techniques, the BP4NTA working group focuses on generating rules and tools to improve NTA practices. Furthermore, the latest 2023 NORMAN guidance on suspect and non-targeted screening, originally aiming for a consensus on NTA in environmental monitoring, could be one valuable prototypic framework for human exposomics.[229] The guidance offers recommendations for critical steps in NTA, from sampling, sample preparation, and HRMS analysis to data evaluation/reporting.[229]

To further resolve ambiguity in terminologies, here we address the distinction between chemical exposomics approaches in a known−unknown quadrant chart, which can serve as a reference map for navigating analytical gaps (Figure 3a). In so doing, one steers through both hypothesis-driven and data-driven exposomics activities while compartmentalizing feature detection (i.e., data acquisition and processing) and annotation (i.e., diagnostic structural analysis).[41,47] The definition of "known/unknown" can be relative, subjective, and context-specific. For the first-place descriptive "known/unknown" on hypothesis/knowledge-based activities, what is known to Scientist A may not be known to Scientist B and vice versa. While for the second-place "known/unknown" on the actual analytical outcome, what is "known" to platform detection and common libraries (or suspect lists) may only represent a marginal fraction in more comprehensive databases (e.g., <1% of SciFinder). Nonetheless, a major division between targeted and non-targeted approaches lies in instruments and data acquisition modes. For example, one may use MRM/SRM on a low-resolution QqQ for quantitative work while operating DDA/DIA on an HRMS for non-targeted work. In some emerging cases, HRMS has proved to allow both arms of NTA and targeted analysis, employing non-selective MS^1 alongside selective monitoring of MS/MS diagnostic ions with the aid of internal chemical standards. Furthermore, the use of in-house library (authentic standards-based) is necessary for targeted analysis, whereas for NTA, first-principle or heuristic rules as well as *in silico* databases

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

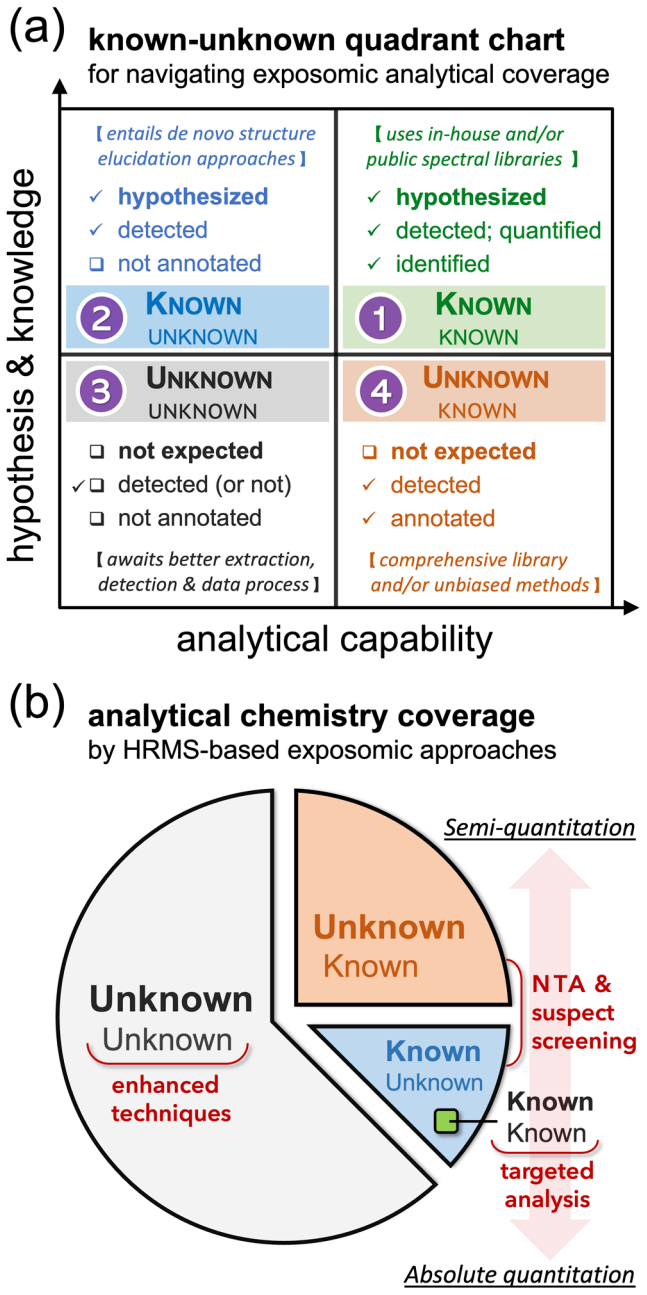


Figure 3. Conceptual navigation of analytical scenarios and approaches for expanding the chemical space coverage by HRMS-based exposomics. (**a**) The "known-unknown" quadrant chart as built from the Rumsfeld Matrix[328] as a framework to consider both the influences of hypothesis/knowledge-driven activities (i.e., expecting a feature to be detected and/or identified in a sample) (*y*-axis) and HRMS workflow capabilities (from feature detection to structural annotation) (*x*-axis) on the analytical coverage outcome. Within each quadrant, the definition of "known/unknown" can be relative, subjective, and context-specific; the first-place "known/unknown" term (in bold) describes the hypothesis/knowledge-driven activities (from unknown to known on the *y*-axis), while the second-place "known/unknown" term (not in bold) denotes the analytical outcome (from unknown to known on the *x*-axis). (**b**) Pie chart illustration of the analytical coverage by HRMS-based approaches including targeted analysis (screening/quantitation), suspect screening, and NTA. The double arrow does not indicate a quantitation continuum (i.e., absolute/semiquantitation is binary for individual compounds) across analysis modes. Rather, it illustrates the tendency or commonality for targeted and non-targeted approaches (or alternative analysis modes) to achieve different quantitative goals with affordable accuracy and sensitivity for target analytes/features.

based on patterns in chromatographic and mass spectral signals can be used for *de novo* annotation of chemicals.

Suspect screening holds a middle ground bridging targeted and non-targeted worlds and can occasionally be considered a subset of non-targeted procedures when NTA data are selectively analyzed.[83,229] Both nonselective MS^1 and user-defined MS^2 diagnostic ions (of known analytes/substructures) are used as opposed to *de novo* structural dereplication for NTA. Suspect screening can leverage community-based chemical databases and spectral libraries (with diagnostic information) to screen and annotate compounds, including spectral database screening, substructure-guided screening, derivatization-assisted screening, etc.[239] Since these libraries are "crowd-sourced" or developed by well-funded institutions/enterprises, they are substantially larger than in-house libraries while maintaining a level of confidence from chemical standards and curation, as represented by the NIST/WILEY GC Library 2023 (>2 million spectra) and mzCloud[240,241] (>12 million spectra). Suspect screening thus emerges as an avenue to expand analytical coverages (from targeted analyses) with generally higher annotation confidence than non-targeted approaches. The caveat, however, lies in addressing the many parametric differences in acquisition, instrumentation, workflows, and human error when applying public spectral libraries to local laboratory settings.

As shown in Figure 3a, *known−knowns* are arguably the most confident annotations, matching against in-house and/or validated libraries based on authentic chemical standards (targeted screening and/or quantitation), whereas *known-unknowns* (hypothesized, detected, but not annotated) and *unknown-knowns* (not expected, but detected and annotated) can be actively resolved through suspect screening and NTA, leaving *unknown−unknowns* to be revealed by enhanced techniques (Figure 3a,b). The overall annotation rates for given data sets are small, typically <5% combining *known−knowns* and *unknown-knowns* compared to other quadrants (Figure 3b).[41] For confidence and throughput, this status quo indicates a need for a tiered approach while encouraging cross-references and harmonization between quadrants to promote coverage expansion, result validation, and data/approach merging. New and integrated cheminformatics pipelines have been attempted for this, such as MetFrag for NTA under the integrated *patRoon* framework.[242,243] The developments in environmental monitoring have sparked enthusiasm toward merged approaches for human exposomics as well.[244]

4.2. Merging Targeted and Non-targeted Approaches: Vision, Strategies, and Feasibility. The idea of merging targeted and non-targeted approaches was first reviewed in 2016 for metabolomics (and the subfield of lipidomics) by Cajka and Fiehn, with methodological issues delineated for experimental workflow and data processing.[245] Targeted and non-targeted approaches have their unique pros and cons; selecting one over the other results in a trade-off between a low-accuracy overview of total molecular changes (i.e., non-targeted discovery) and a detailed yet limited snapshot of select compound subsets (i.e., targeted screen/quantitation). In some instances, researchers run both analyses via multiple injections of the same sample (on the same or a

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

different platform). These are typically implemented by a survey scan for non-targeted profiling (with HRMS) in the first injection to guide screen and/or quantitation of select targets implemented in a second injection, likely via MRM/SRM, single ion monitoring (SIM), PRM, DDA, or DIA.[246]

Merging approaches of exposomics entails balancing coverage, throughput, and capacities for quantitative and qualitative analyses, depending on the specific analysis goals to pursue (e.g., quantitation vs screening for targeted approaches). Such endeavors started in the mid-2010s[26] with a few more recent HRMS-based attempts.[25,111,247] However, there is a lack of community consensus; systematic workflow development and cross-laboratory harmonization are nascent for HRMS-based exposomics. Recent studies have started to provide clues. First, the use of authentic standards (ASDs) and stable isotope-labeling (SIL) internal standards (ISDs) appear to be instrumental in bridging the targeted and non-targeted worlds, either as a readout for quality control (QC), as a reference for quantitative purposes, or as tracers for characteristic structural moieties by isotope labeling or hydrogen−deuterium exchange,[248] to name a few. Jia and colleagues (2019) integrated targeted and non-targeted methods for exposomics of human urine by isotopically tagging exposure biomarkers with common functional groups (e.g., phenolic, hydroxyl, carboxyl, and primary amine) that can simultaneously be leveraged for quantitative purposes.[247] Likewise, the single-injection simultaneous quantitation and discovery (SQUAD) approach, developed for metabolomics, identifies (via ASDs) and quantifies (via calibration curves or one-point calibration using ASDs and ISDs) select compounds while allowing simultaneous data mining to look for broader molecular changes.[249] Second, the feasibility would largely depend on the specific instrumental capacity in use. For a proof of principle, high- and low-resolution MS (Orbitrap HRMS and QqQ, respectively) have demonstrated comparable sensitivity in both screening and quantitation of urinary exposome compounds, suggesting a synergy combining these acquisition modes for merged workflows.[55]

Nonlinearity remains an unresolved challenge in NTA toward workflow merging and risk assessments (e.g., quantitative dosimetry).[250] The quantitative accuracy and precision in NTA never match targeted assays due to a lack of surrogate/internal standards and certified samples.[83] Relative measures such as peak intensities and their ratios (i.e., fold changes) have been used for statistical analysis (e.g., when comparing data sets) and toxicity prioritization (e.g., the ToxPi framework).[251,252] However, pitfalls may occur due to coexisting interferents and matrix effect.[250,253] Recent studies have discovered significant quantitative biases using MS signal intensity ratios, particularly for ESI-MS platforms.[254] The nonlinearity may be partially addressed by the serially diluted QC-based calibration.[255] To broadly merge workflows and extend dynamic range, a chemometric classification of mixture components and/or methods based on *a priori* spectral information may also be helpful,[256] where multivariate statistical approaches (e.g., nonlinear regression models, machine learning) are used to address potential nonlinearity issues in NTA for background correction and more accurate quantitation.

HRMS instrumentation continues to evolve and make merged, streamlined exposomics workflows a reality. The latest Orbitrap and ToF MS models show promise, with examples of Orbitrap Tribrid and Q-Orbitrap Astral (Thermo Fisher Scientific),[214,257] ZenoTOF (Sciex),[258] and timsTOF (Bruker).[259] The Orbitrap Tribrid MS combines the advantages of quadrupole, Orbitrap, and LIT MS analyzers, operating two detectors (i.e., sensitive LIT and high-resolution Orbitrap) in parallel to accommodate both targeted and non-targeted analyses. Tribrid MS enables multiple MS^2 (e.g., HCD, CID, and UVPD) and MS^n approaches, allowing more in-depth and accurate unknown identification.[257] ZenoTOF MS leverages a Zeno trap pulsing to overcome common Q-ToF MS^2 duty deficiencies and (avowedly) achieves 5−20 times better sensitivity than older TripleTOF models. ZenoTOF offers both EAD and CID options for MS^2 fragmentation, enables SWATH-DIA and high-resolution MRM, and employs an ultrafast ToF scanner, with emerging chemical analyses demonstrating high capabilities.[258,260,261] The timsTOF achieves high sensitivity through ion confinement and selective mobility release (reaching 100% duty cycle) while incorporating IMS for orthogonal separation, specificity, and throughput. The CCS values and dia-PASEF data acquisition can increase confidence in compound annotation.[262] Through innovative design, all three kinds of instrumentation offer high mass resolution, high sensitivity, versatile fragmentation, and fast scan speeds that support both detection/discovery and quantitation capacities. The forthcoming years are undoubtedly filled with exciting new opportunities for boosting chemical exposomics, pushing HRMS limits in sensitivity, coverage, scalability, and analytical throughput.

4.3. Expanding Analytical Coverage to Uncover Causative Toxicants: Quo Vadis? What is the status quo of chemical space coverage of exposomics? Quo Vadis? In environmental monitoring (e.g., air, water, soil), HRMS-based analyses have been rapidly expanding over the past decade.[229,263] Platforms and workflows are being repurposed to target new and more pertinent chemical classes, as reflected in the Toxic Substances Control Act (TSCA) Inventory or more selective databases.[51,264] These new compounds and substances include many contaminants of emerging concern (CEC),[265] and are wide-ranging by structure, use, and toxicity, spanning PFAS,[266] phthalates,[267] pharmaceuticals,[268] UVCB substances (short for "unknown or variable composition, complex reaction products or of biological materials")[269] and microplastics,[270] among numerous others. Recent case studies to discover causative toxicants provide clues not only for profiling actual exposure occurrences and risk assessment but impart insights into the methodological design for HRMS-based exposomics altogether. The HRMS-led NTA capacity is broadly conducive to effect-based methods to identify individual drivers and modifiers of toxicity and disease.[38,87]

One seminal case study is the identification of 6PPD-quinone, a causative toxicant responsible for the massive acute mortality of coho salmon in seasons returning to spawn and a recurring event puzzling scientists for decades.[84] 6PPD-quinone is a major transformation product of 6PPD, or *N*-(1,3-dimethylbutyl)-*N*′-phenyl-*p*-phenylenediamine, a common antiozonant and antioxidant added to vehicle tires. The study was conducted under an EDA framework[68,271] which involved multistage chemical fractionations of tire rubber leachate mixtures to guide *in vivo* toxicity assays in serial disparate steps to identify the causative toxic component. The downstream HRMS-based NTA identified such components as 6PPD-quinone, which was confirmed by nuclear magnetic resonance (NMR).[84] With further follow-up NTA,[272,273] the

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

studies of 6PPD-quinone mark a methodological breakthrough in environmental sciences where HRMS analytical capacity expedites risk factor discovery and mitigation.

Quo Valis? Many CECs are synthesized to replace regulated chemicals. However, current chemical regulation frameworks, such as the Stockholm Convention, are more reactionary than precautionary.[274] First, substances are considered to be banned only when harm (human or ecological) is demonstrated through environmental, experimental, and epidemiologic observations, usually long after the first awareness of warning signs. Second, knowledge-based effect-directed methods leave numerous unknowns under-evaluated for exposure occurrences and health impacts. As for the replacement chemicals, although the structures of these substituents differ from the banned or legacy to-ban species, the core physicochemical properties likely stay and linger to fulfill similar functional performances (e.g., as flame retardants, coolants, and lubricants), often with comparable harmful effects. Furthermore, the new safe-by-design principles by the European Commission[275] may result in greater convergence of properties and activity among chemicals in the future, making the screening/prioritization more challenging with unforeseen adverse effects harder to dissect and elucidate.[83] All these will not change fundamentally unless market needs, industrial chain, and governmental policy experience synergistic changes, where green chemistry technologies may provide a solution.[276−278] To facilitate effective systems thinking[279,280] and innovative policy-making,[19] timely toxicant identification and health effect assessment at the front end are key.[281] The coming-of-age HRMS techniques and merged workflows will be crucial drivers in this, with demonstrated successes in suspect screening and NTA of ubiquitous complex mixtures such as PFAS[266,282,283] and plastics additives (using pyrolysis-GC-HRMS).[284−286] Studies are underway in targeting the internal component of the chemical exposome to uncover new chemical space that humans are critically exposed to.[46,287,288]

5. DATA ANALYTICS: FROM FEATURE DETECTION TO STRUCTURAL ANNOTATION

With the ever-evolving HRMS techniques, data generation rates and magnitude continue to expand. From hardware data acquisition to software data analytics, converting raw data files into aligned, annotated ion feature tables for statistical analysis (for interpretable health insights and to support chemical risk assessment) becomes attainable. Over the past decade, informatic tools and algorithms for small molecule analysis, specifically metabolomics, have been increasingly applied to accomplish big-data tasks in exposomics (Table 1). Still, key caveats and cautions remain under-discussed. In a given cohort of samples, how many ion peaks/features can be detected, properly integrated for quantitative purposes, and aligned across samples, specifically for chemical exposure agents? What are their chemical characteristics and structural identities? How can we leverage such information to conduct statistics for interpretable health-oriented insights? Here, we focus on feature detection (from raw data to aligned feature table, Figure 2d) and compound annotation (from ion feature to chemical structure, Figure 2e)—the two most pertinent data analytics steps in exposomics before one delves into statistical analyses[289−291] for health-oriented inferences (Figure 2c).

5.1. Feature Detection: Algorithms and Parameter Tuning. Feature detection, or data (pre)processing, is the first and foremost informatics step that converts raw HRMS data into a tabular, numerical format for data cleanup, statistics, and bioinformatic analysis (Figure 2d).[292] Such tabular datasets, typically derived from GC- or LC-HRMS, contain information of at least three dimensions, namely chromatographic retention time (RT), m/z (as well as the retrievable MS^1 isotopic pattern), and peak intensity (integrated area of MS ion signals or peak height) or more in case IMS CCS and/or other complementary techniques are not in use. The datasets are highly complex as the HRMS characterizes ion signals resulting from ionization and structural modulation of analytes that involve ion adducts, in-source fragments, isotopes, background, and contaminant ions.[293,294] Sophisticated informatics are thus needed to mine and clean up data before delving into statistical analysis.

To process data, algorithms are programmed to inspect the spectral data to identify ion peaks (also called "m/z features," "ion features," or "features"), evaluate individual m/z over RT to construct extracted ion chromatograms (XIC), integrate that peak area (or record its heights), and align all peaks together across sample sets (of single or multiple batches). Key settings typically involve (*i*) m/z (MS^1 and/or MS^2) and RT ranges, (*ii*) peak picking methods that may entail tracking, binning, and clustering, e.g., centWave for XCMS,[295] concavity/slicing for MS-DIAL,[296] local maxima for MZmine,[297] and "moment" estimation for apLCMS,[298] and (*iii*) peak componentization to group various spectral "fingerprints" (e.g., isotopes, adducts, and ion fragments) originated from the same compound, and (*iv*) peak alignment parameters as a filter (e.g., by frequency, deisotoping, QC, blank, etc.) to decide whether an individual peak should be included in the final ion feature table.

Many tools and algorithms have been developed for nontargeted small-molecule data processing, especially for metabolomics (Table 1).[299] Yet, challenges in algorithmic design and parameter tuning remain, with technical nuances reviewed recently.[300] Both false positives and false negatives need to be addressed in characterizing individual peaks (in the individual samples) or features (aligned from multiple samples) for the overall best results, pointing to a trade-off between peak peaking and componentization.[301] Taking the slicing approach as an example where a "net" or slice space is projected against the spectral signals for m/z grouping and peak picking, the relationship between spectral peak width (i.e., m/z window) and slice width (toward defining a peak) represents an essential detail which highly depends on specific MS types in use.[295,296] If the peak width is larger than the slice width (typical in low-resolution MS), the signal from a single peak may "bleed" across multiple slices, leading to false positive peaks that were not MS-resolved. If the peak width is significantly smaller than the slice width, which occurs in ultrahigh HRMS analyzers such as FT MS analyzers (Orbitrap or ICR) or using precentroided data containing infinitely thin m/z peaks, a jagged peak shape may occur. Depending on the specific scan-to-scan instrumental precision, the signal from an analyte may fluctuate between adjacent slices over the chromatographic retention time, making an otherwise smooth peak shape appear jagged.[295,296]

The human chemical exposome is characterized by the vast chemical complexity, concentration at trace levels, and variability of exposure occurrences. Conventional data processing software faces peak picking and feature detection limitations due to low spectral signal intensity and poor peak shape. A universal one-for-all set of parameters could thus be

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

troublesome for NTA and exposomics in general which aims to accommodate both low- and high-abundant compounds. Certain algorithmic improvements provide a possible solution in the earlier days, such as the apLCMS/xMSanalyzer suite that optimizes for low-abundance peaks.[298,302] More recently, fast and self-optimizing processing workflows were released based on machine learning (e.g., SLAW)[303] as well as the *asari* approach which features the new "mass track" concept that shuns the provenance issue of peak picking (i.e., align first, pick peaks later).[304] For workflow optimization and harmonization, QC and appraisal tools are being developed to compare and validate peak-picking algorithms, including those for DIA strategies.[305] For systematic assessment, existing preprocessing software for LC-HRMS metabolomics was evaluated based on the FAIR (Findable, Accessible, Interoperable, and Reproducible) principles for research software (FAIR4RS); similar appraisals should be conducted for exposomics data processing.[292] On the GC-MS side, there have been multiple platforms, from the pioneering GC-MS BinBase,[306] which implements fully automated processing in the cloud for alignment and annotation, to the more recent automated data analysis pipeline for GC (ADAP-GC), Compound Discoverer-GC, and the global natural products social molecular networking (GNPS) GC workflow.[142,307,308] It should also be noted that, recently, harmonized quality assurance/quality control (QA/QC) guidelines (large collaborative efforts during the EU PARC Initiative) have been proposed for exposomics data preprocessing that allows community-wide access to the sensitivity of feature detection, reproducibility, integration accuracy, analytical precision and accuracy, and consistency.[309]

Metabolomics and exposomics loosely share a common archetypic framework for data processing but differ considerably in the technical specifics due to different analysis goals (i.e., metabolite-oriented vs exposure-oriented). The many assumptions applicable to metabolomics may not be suitable for exposomics. For example, the measure of total ion signals as a baseline is commonly applied to normalizing data in metabolomics, assuming that the whole metabolome is in a constant dynamic flux. Individual metabolites may be upregulated or downregulated, but the overall composite signal, after correcting for instrumental variation (e.g., batch effects), is considered constant. Likewise, for aligning samples of multiple groups, the "80% rule" is widely applied in metabolomics,[310] setting 80% in "*N*% detected in at least one group" for keeping features into final alignment under the assumption that metabolites are common (i.e., operationally 80%) at least in one group so that possible falsely picked peaks are removed. In contrast, exposome molecules are not always common, and exposomics may encounter unique challenges in (*i*) data sparsity, (*ii*) data normalization (e.g., for removal of batch effects), and (*iii*) statistical treatment (e.g., dimensional reduction).

Unlike canonical metabolites, chemical exposure agents can be random, erratic, and largely low-abundant, distributing sparsely among human subjects. Thus, metabolomics settings are usually too stringent for exposomics regarding thresholds of noise and percent feature presence. Meanwhile, more careful missing imputation approaches are needed to address the sparsity issue without losing the sensitivity to capture exposure patterns specific to a subgroup of individuals but largely uncommon across the whole cohort.[289] Furthermore, data normalization strategies to correct for batch effects in exposomics data, especially for large-scale cohorts, merit meticulous tests that may or may not involve pooled quality control (QC) samples[311] in the normalization algorithms.[312,313] In addition, statistical approaches linking exposure agents to their transformation products and metabolite markers are being developed. Notable developments include the "molecular gatekeeper" approach[314] and the Chemical Correlation Database (CCDB).[315] However, caution should be taken when relying solely on a single HRMS data file to find specific exposure−response links, as the assumption falsely disregards a lag of exposure event leading to an actual health effect.[15] Last, deconvolution poses a significant challenge for GC-EI-MS data, although strategies and approaches have been continuously evolving.[307,316] In compliance with the FAIR practices, a new MSHub/GNPS has recently been developed as a centralized hub/pipeline to perform automated deconvolution of ion fragments based on unsupervised non-negative matrix factorization (NMF) and molecular networking.[308] Likewise, ADAP-KDB, a spectral knowledge base that integrates the updated ADAP spectral deconvolution algorithms, has recently been developed to track and prioritize unknown GC-MS spectra from public data repositories.[307,317]

5.2. Compound Identification Basics: From Spectral Data to Structure. Compound annotation (structural assignment), or the more conclusive compound identification (structural assignment with affirmative validation by chemical standards or equivalent approaches), depends on measurements uniquely related to the chemical structure that is not shared by other molecules. In this molecular game of "guess who," multiple measurements are taken until only one molecule that fits within these measurements remains. In HRMS-based exposomics where MS is enhanced with analyte separation and tandem mass capacity, several measurements are commonly used for annotation. These include chromatographic RT (GC or LC), drift time and/or CCS (IMS, when applicable), accurate mass (MS^1), isotopic pattern (MS^1), ion adduct type and in-source fragments (MS^1), and ion fragmentation pattern (MS^2 or MS^n). In collective use, these measurements determine the confidence of an annotation. For example, in the Schymanski scale, a guideline for reporting annotation confidence that is widely accepted for LC-HRMS/MS studies, for a Level 1 (confirmed structure), the minimum requirements are RT, accurate mass, and a fragment(ation) match,[70] although in many cases RT and accurate mass alone are sufficient.[318−320] Likewise, in the Koelmel scale for GC-HRMS exposomics, an RT and fragment(ation) match are required.[321]

The annotation confidence depends on the quality of the reference mass spectral library in use, which in order from highest to lowest includes: (*i*) in-house standards acquired during the same experiment, (*ii*) in-house library from standards not acquired during the same experiment, (*iii*) community-shared library from standards under the same or similar experimental conditions, (*iv*) rule-based libraries based on standards, and (*v*) *in silico* or computer-predicted libraries.[70,321,322] Hence, on both Schymanski and Koelmel scales for confirming a structure to the highest Level 1, RTs and ion fragment(s) must be from an in-house library using reference chemical standards.[70] Community-based libraries, though not as confident as in-house counterparts (i.e., no RTs or CCS available), can significantly improve annotation coverage while maintaining a decent level of confidence.

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

Community-based libraries are best suited to measurements harmonized across laboratories, as far as methodologies do not differ between instruments tied to uncontrollable external variables. Harmonizable measurements across laboratories include the GC-EI and LC-ESI tandem mass spectrum, accurate mass, isotopic pattern, and CCS, although a significant proportion remains unshared.[145,323,324] In recent years, MS^2 coverage has been substantially expanded in public libraries, especially for LC (e.g., MassBank/MoNA,[325,326] METLIN,[327] mzCloud[240,241]), incorporating wide-ranging experimental parameters (e.g., collision energies, fragmentation types) to suit needs. Given the varying spectral sources and quality, expert knowledge may come in at a certain point to ensure proper library selection and matching for consistency by chemical space covered, analytical condition, organismal species, and sample matrix. One should note that even a high score in the spectral match against these MS^2 libraries is putative at best; one needs to further rely on authentic chemical standards or equivalent strategies (e.g., 2D NMR) to assign a full Level 1 identification.[70,321,328,329]

Given the essential roles of tandem mass spectra (i.e., MS^2 spectra, or MS^2) in expanding the analytical coverage, it is imperative to acquire quality MS^2 for as many analytes as possible. GC- and LC-derived spectral data entail separate approaches mainly due to the distinct nature of ionization processes. For GC-EI, in-source ion fragmentation occurs extensively due to hard EI, leaving no intact molecular ions for most compounds to be observed except for certain persistent species (e.g., PAHs and PCBs). For LC-ESI, the majority of analytes likely experience soft ionization with their intact parent structures preserved in the ESI ion adducts, with exceptions (e.g., of certain PFAS) which undergo fragmentation with little molecular ions. To obtain LC-ESI-MS^2, ions are selected (isolated) in a mass filter (e.g., a quadrupole with a ∼ 1-Da mass window) and put through a collision cell (e.g., a hexapole filled with N_2 or Ar) for fragmentation experiment(s). Since mass spectral acquisition takes time, MS^2 spectra may only get acquired on some analytes within a limited cycle time; only a portion of ions within the set mass window are selected for fragmentation and scanning.

Strategies to increase the MS^2 coverage include but are not limited to (*i*) fragmentation without precursor selection and (*ii*) reiterative analyses of the same sample. The first strategy involves DIA, thus requiring deconvolution to compute the precursor-fragment relationships and reconstruct pseudo-MS^2 spectra for individual precursors. Open-source computational tools, such as MS-DIAL[296] can perform spectral deconvolution on both LC-DIA MS^2 and GC-EI type of data with good results demonstrated for environmental samples (e.g., air pollution analysis by LC- and GC-HRMS).[330] Comparable algorithms and tools for spectral deconvolution include DEIMoS (for high dimensional LC data),[331] RAMClustR (for both LC-ESI-DIA and GC-EI),[332] and IonDecon (for LC-DIA).[333] However, it is acknowledged that these wide-scope non-targeted acquisition approaches may lead to inflated false positives; careful testing and procedural benchmarking to verify data quality and deconvolution efficacy should be performed.

The second strategy applies to samples of abundant amounts by continual reinjection of the same sample with iterative (ion) exclusion complementing previous runs until all ions of interest are fragmented; MS^2 can then be extracted by automated tools like IE-Omics[334] and PMDDA (short for "paired mass distance-dependent analysis").[335] More sophisticated methods may come into play for the iterative selection leveraging mass spectral patterns, blanks, and metadata, as exemplified by AcquireX[336] which automates background exclusion, component inclusion, iterative precursor exclusion, and deep scan to allow selective triggering of specific ion adduct types. These new approaches may further reduce the reinjection numbers needed to perform all requested fragmentation for an expanded MS^2 coverage.

For measurements that are more difficult to replicate accurately as compared to mass spectra, certain calculations can be done to correct for analytical variability or shifts. Take RT in GC, for example, which occasionally can be difficult to reproduce: even the same columns could differ considerably in RT for the same analyte across different batches, after trimming, or between manufacturers. This may be resolved through normalizing RTs to known peaks of reference standards across chromatographic runs such as alkanes and fatty acid methyl esters (FAMEs). Upon normalization, the resultant retention index (RI)[337] becomes more readily comparable across laboratories, although challenges remain for co-eluted peaks (e.g., PCB congeners) with likely misalignment issues.[338] In the case of LC,[339] RI can also be calculated and incorporated into public libraries. Still, caution should be taken to ensure a match of chromatographic solvents, gradient, and stationary phase, as the LC separation mode depends on all three. Regardless, to use the chromatographic retention to (optionally) complement the tandem mass spectral search, Schymanski Level 2 assignments (probable match) for LC-HRMS and Koelmel Level 2 (probable structure or close isomer) and 3 (tentative candidate) for GC-HRMS all rely on normalized/adjusted values for community-based libraries.[70,321] In GC-HRMS using community-based libraries for RI and EI spectral matching, there were slightly less than 25% false positives when also allowing for similar isomers to be considered matches, and over 50% false positives for exact matches (top-ranking hit).[321] This shows a substantial difference between using in-house libraries (from reference standards) and community-based libraries (with RIs) for matching; it is important to understand the difficulty of assigning correct annotations in HRMS-based exposomics studies.[321]

5.3. Cheminformatics for *De Novo* Structural Elucidation. Both community-based and in-house reference spectral libraries are limited to known chemicals, leaving the emerging unknowns (and their unwanted transformation byproducts) poorly characterized. Meanwhile, the biotic/human exposome remains largely unknown, indicating a need for NTA in exposomics to cast a wide net to identify the "unknown unknowns" (Figure 3). While the HRMS technologies continue to evolve and uncover new chemical space, on the computational side, one would need informatics algorithms for compound annotation and spectral library construction that focus on the intrinsic patterns of exposome spectral data structures between and/or within measurements or functional exposomics processes. For example, many anthropogenic chemicals, such as PFAS, hydrocarbons, and polymers, carry unique repeating units that differentiate them from naturally occurring compounds. The accurate mass patterns alone may suffice for finding these "chemical series," alternatively termed homologous series. Besides, atomic mass is normalized to ^{12}C (12.0000...); all mass defects come from atoms other than ^{12}C itself. By normalizing to CH_2, Kendrick

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

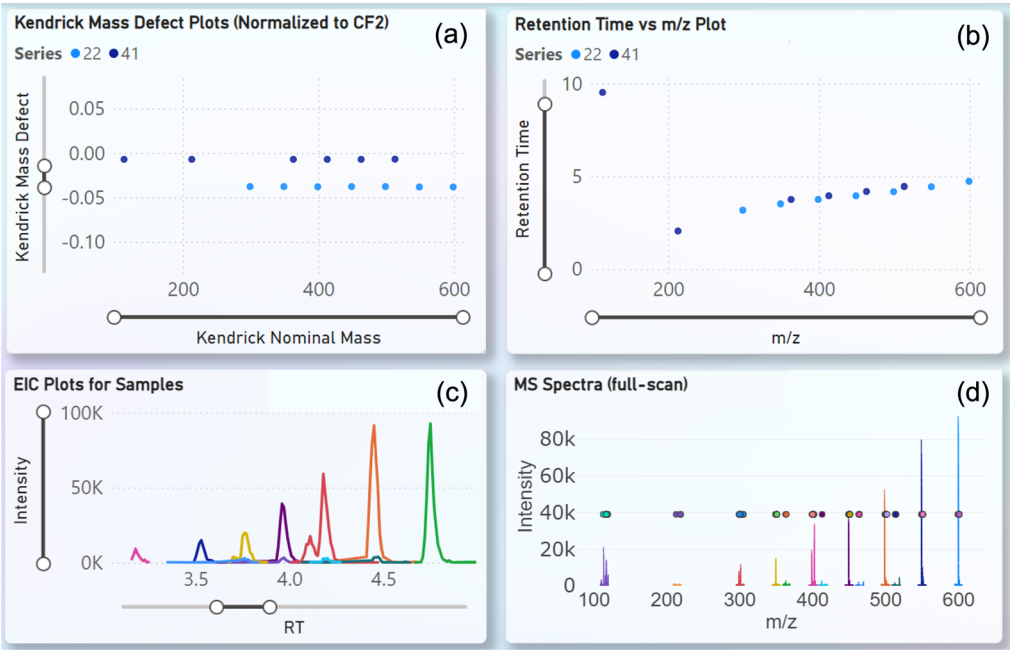


Figure 4. Patterns in the mass spectral measurements can be leveraged for *de novo* structural elucidation of exposome molecules, using FluoroMatch and PFAS for a showcase.[344,409] (**a**) The Kendrick plot reveals distinct mass defect patterns indicative of the CF_2 repeating units in the PFAS homologous series. (**b**) Similar patterns of RT and m/z space further confirm the homologous series orthogonally; homologues with differing repeat units will follow a different trend and can be readily removed. (**c**) The EIC plot view of the PFAS homologous series. (**d**) The full-scan MS spectral view of the PFAS homologous series.

realized that all non-CH_2-related additions to petroleum (unsaturation, N, S, aromaticity) could be inferred from mass defect—a technique now widely applied in chemical analysis of PFAS, pegylated polymers, lipids, and other chemical classes.[340]

Taking PFAS as an example, when normalizing to CF_2, one can make easy characterization and structural inferences, as all members of a class (e.g., all members containing a sulfonic acid functional group and a carbon−fluorine chain) will have the same mass defect and be 50 Da apart (CF_2 mass difference) (Figure 4a). Patterns in other measurements can further help confirm these homologous series. If simply differing in a repeating unit, these compounds should form a clear trend in RT vs m/z space. In contrast, other chemicals with the same accurate mass but different elemental composition or structure will not follow such trends and can be readily removed (Figure 4b). At the basic level, using these two techniques alone can substantially increase annotation rates for compounds bearing varied repeating units that are not covered in spectral libraries. At a more advanced level, MS^1 isotopic pattern could distinguish the presence of certain elements, such as Cl, S, and Br, which have unique isotopic patterns.[341,342] Using Kaufmann plots, isotopic pattern further helps determine structures by similarity in elemental composition and percentages. Specifically, the Kauffman plot maps mass defect over the estimated carbon number (from M+1 isotope) vs m/z to detect chemicals of specific classes (e.g., those highly fluorinated). Moving beyond MS^1, MS^2 spectra (and similarly, MS^n) are especially conducive to pattern analysis as ion fragments occur probabilistically based on the chemical structure (e.g., under CID, weakest bonds are typically the easiest breaks).[199,200]

To unravel unknown−unknowns and resolve known-unknowns, a wide range of strategies exist to infer from ion fragments without an experimental database match. These span ion fragment screening, matching against *in silico* fragmentation libraries, molecular networks, and formula prediction based on fragments and/or neutral loss.[322,343] First, fragment screening allows rule-based annotation, leveraging empirical substructure knowledge such as m/z 184 ion fragment for phosphocholine headgroup in lipidomics or $[SF_5]^-$ for PFAS which contains pentafluorosulfide anion/moieties.[344,345] Current mass spectral libraries mostly cover endogenous metabolites, pointing to a lack of reference spectral databases for exposome molecules (e.g., environmental pollutants or exogenous metabolites). There have been emerging *in silico* endeavors, as represented by the Blood Exposure Database,[346] a predictive model-based repository for trace organics of the human exposome. Furthermore, molecular networks constructed by spectral similarity referencing against the total spectrum (cosine) or by the number of matched spectral fragments, play a crucial role in non-targeted screening not only as a feature filter but facilitating compound identification.[343] By linking nodes (individual ion features) with similar ion fragments or neutral loss patterns, chemicals of similar structures or shared common motifs are grouped together and identified (e.g., PAHs may all group similarly in GC EI spectral networks).[330,343]

While expanding reference spectral resources for exposome compounds is crucial, spectral similarity algorithms and associated scoring/ranking represent another integral step in defining a match or annotation. The mathematical characterization of spectral similarity has been continually evolving since the 1970s when Probability-Based Matching (PBM)[347,348] and dot product (Finnigan/INCOS)[349] were first proposed. Further sophisticated algorithmic developments include the Hertz similarity index,[350] weighted dot product,[351] mass spectral tree search,[352] etc. For years, weighted dot product has remained the most widely adopted algorithm that

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

demonstrably outperformed PBM, Hertz similarity index, Euclidean distance, and absolute value distance by comparative testing.[351] In the recent five years, several new similarity algorithms have emerged that move beyond dot product similarity. Two notable examples include machine learning-enhanced search/ranking of structural analogs (e.g., MS2Query[353]) and spectral entropy-based search, which demonstrably outperforms dot product similarity for small-molecule annotations with a further boost in computational efficiency.[354,355]

5.4. Advancing Cheminformatics to Expand Exposomics Coverage: Strategies, Approaches, and Toolkits. From peak picking to chemical identity assignment, structural elucidation is key to bridging environmental exposure agents to biological/health effects with solid chemistry insights. The latest NORMAN guidelines (2023),[229] while intended for environmental monitoring needs, have listed multiple existing (non-vendor) tools and software for compound identification as being incorporated into a tiered strategic workflow for exposomic screening. The tiered workflow includes targeted screening based on in-house library match, MS/MS library search, suspect screening, and NTA (*de novo* structural inference).[229,328,329] The major difference between suspect screening and non-targeted screening in environmental monitoring, according to NORMAN 2023, lies in both prior knowledge about the contaminants (expected/suspect vs unexpected/unknown/non-targeted) and the goal of analysis. Suspect screening seeks to obtain a big picture of pollution through long-term monitoring of a large number of suspects at hand (e.g., NORMAN SusDat list) for modeling and regulation purposes. While non-targeted screening aims at identifying unknown chemicals causatively linked to adverse effects in question.[229] Nonetheless, both modes carry the discovery component highly applicable to human exposome analysis, with shared goals of tracking harmful environmental contaminants.

To expand the analytical coverage by human exposomics, the field of (computer-assisted) cheminformatics for HRMS-based structural elucidation is growing rapidly, covering aspects of matching algorithms, spectral prediction, formula and substructure assignments, networking, molecular weight predictions, orthogonal information, use of metadata, and artificial intelligence (AI) utility (even for single-stage GC-MS).[356] The *in silico* approaches for predictions between formula, structure, and spectrum can be classified into four categories:[322] (*i*) heuristic rules,[357] (*ii*) chemical reaction-based rules (e.g., MassFrontier,[358] MS-FINDER[306,359]), (*iii*) machine learning-based approaches (e.g., CFM-ID,[360] CSI-FingerID,[361] MetFrag[243]) and (*iv*) quantum chemistry modeling.[362,363] To illuminate exposome chemical space and health effects *in vivo*, new *in silico* platforms have been developed to capture biotransformation processes directly from spectral data. Recent examples include the Reactive compound Transformation Profiler (RTP) for probing reactive compound transformation products),[364] CyProduct for accurately predicting byproducts of human cytochrome P450 metabolism,[365] and BioTransformer (now ver. 3.0) for accurately predicting metabolic transformation products.[366]

The cheminformatic data pipeline can be prolonged, and both intermodular dependency and the need for streamlined and automated workflows motivate the development of toolsets—a suite (or cluster, collection) of tools functionally compatible with one another that collectively form synergistic advantages as supported by an "ecosystem" of active community participation to advance validation and research application. Many such developments stem from metabolomics, with the majority developed over the past 15 years. Seminal example suites include MS-DIAL/MS-FINDER (Fiehn),[296,359] SIRIUS/CSI-FingerID (Böcker),[361,367] XCMS/METLIN (Siuzdak),[327,368] apLCMS/xMSanalyzer/xMSannotator (Jones),[298,302,369] and TidyMass/metID (Snyder),[370,371] to name a few. Although it is often reasonable for individual users to stick with one suite of toolkits and exploit them fully for tackling a specific NTA problem at hand for best overall results, reproducibility, and reporting, it is equally essential to establish timely community consensus for standardization and optimization. To this end, the CASMI contest for "Critical Assessment of Small Molecule Identification," launched in 2012,[372] has continually provided valuable comparative insights to benefit the community.[373−375]

What do we learn from these toolkits and current annotation practices? How can we best benefit from the existing metabolomics innovations to advance human exposomics, particularly to boost compound annotation rates? Two key components in cheminformatics practices are chemical search space (i.e., candidate structures) and the *in silico* algorithm(s) for spectral prediction and matching. The search space consists of preselected databases or collections of chemical structures as input for *in silico* spectral prediction. For NTA, although it is non-targeted by intention which seeks an unbiased and comprehensive characterization, the task itself is daunting, hard to streamline and reproduce, and prone to errors and biases. Two strategies may help resolve this. One is to confine the search space only to relevant environmental occurrences and feasible analytical/computation power. For example, PubChem currently houses 116 million compounds, including synthetic molecules that humans are potentially exposed to. Using the entire list as input for NTA annotation can be unproductive and irreproducible. A lite version of PubChem, i.e., PubChemLite for Exposomics (<50,000 compounds),[48] has been devised by knowledge-based curation to make chemical exposomics NTA accessible. Another strategy to reduce the computational burden is to infer from data directly. One emerging tool is BUDDY,[376] a platform for molecular formula discovery via bottom-up MS/MS interrogation for global structural annotation. BUDDY decomposes the query MS^2 spectra into fragment-neutral loss pairs which allows for *de novo* discovery of new formula/subformula with high accuracy. This approach has been shown to reduce computational costs significantly compared to top-down algorithms such as SIRIUS,[367] which generates the entire potential candidate space using MS^1 data for downstream formula scoring, ranking, and filtering based on MS^1 and MS^2 data.

For spectral prediction and matching, one distinction of algorithms between exposomics and metabolomics is the need to capture spectral patterns uniquely linked to structures more prevalently observed in exogenous chemical exposure agents, such as the inclusion of halogens and fused rings. Unfortunately, current cheminformatics toolkits are lacking for annotating these exogenous chemicals, let alone the ones to be combined and streamlined for scalable use. The drawbacks of utilizing quantum chemistry-based approaches for *in silico* prediction of EI spectra for environmental chemicals have been recently noted.[377] As for LC-HRMS-based analysis, certain emerging tools show potential, including the MetFrag-based workflow that offers to merge varied modes of analysis.[48,243]

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

What's out there? What are humans exposed to? What and how can we measure?

CHEMICAL SPACE MAPPING FOR HUMAN EXPOSOME

1. **Chemical inventory:** (a) Timely updated chemical inventories of exposome compounds in market and actual societal use are lacking, nationally and globally; (b) Curation efforts for structural identities and metadata are not harmonized.
2. **Search space:** Available resources are lacking and largely not harmonized.
3. **Biomarkers *in vivo*:** Biomarkers of chemical exposure agents remain poorly cataloged.
4. **Prioritization:** Frameworks for long-term targeted tracking and suspect screening are lacking.

How to acquire quality HRMS data for chemical exposomics?

LABORATORY MEASUREMENT

1. **Reproducibility:** intra- and inter-variability of coverage, dynamic range, and quantitative accuracy.
2. **Streamlining:** (a) Benchmarking HRMS performance by modular steps; (b) Merging targeted and non-targeted components; (c) Data formats and merging; (d) Computation-aided automation.
3. **Harmonization:** (a) Establishing reference frameworks by authentic chemicals, certified reference materials (CRMs), and standardized protocol/result reporting; (b) Establishing communication mechanism for protocol exchange, workflow comparison, and community-based optimization.
4. **Scalability:** (a) Improving study design (sample size, metadata, etc.); (b) Optimizing *QA/QC* for monitoring, batch effect correction, and framework for data merging; (c) Constructing infrastructure for centralized storage and curation of disparate data streams for long-term tracking, retrospective validation, and multi-omics data integration.

How to translate HRMS data into feature tables for effective exposomics statistics?

DATA ANALYTICS

1. **Feature detection & alignment:** (a) Benchmarking algorithms for peak picking, from data conversion (e.g., "centroiding") and XIC peak definition/characterization to componentization (e.g., de-isotoping, adduct grouping) into ion features; (b) Benchmarking algorithms for peak alignment that address data sparsity, background noise, and contamination while applying filters based on detection frequency, blanks, and QC samples, etc. (c) Benchmarking algorithms for deconvolution of GC-HRMS and LC-HRMS (DIA) data.
2. **Data normalization and transformation:** (a) Normalizing sample matrix variability; (b) Calibrating NTA non-linearity; (c) Correcting for batch effects; (d) Benchmarking data transformation/scaling.
3. **Compound annotation:** (a) Expanding HRMS-based spectral libraries; (b) Benchmarking cheminformatic toolkits and pipelines; (c) Improving transparency and standardization in reporting annotation confidence; (d) Fostering efforts to merge, cross-reference, and synergize targeted screening, suspect screening, and NTA.

Figure 5. Select challenges in expanding the analytical coverage of human chemical exposome using HRMS-based approaches.

The functional modules span user-definable target screening, suspect screening (e.g., using NORMAN and Eawag-PPS suspect lists), and NTA which combines *in silico* fragmentor,[378] machine learning,[379] and the newly curated search space of PubChemLite for Exposomics.[48] Concerted cheminformatic efforts in both GC-HRMS and LC-HRMS applications are warranted in the forthcoming years to better cover xenobiotics and associated molecules to enable comprehensive, accurate, and reproducible human exposomics.

6. OUTSTANDING CHALLENGES, OPPORTUNITIES, AND FUTURE PROSPECTS

Current HRMS instrumentation, specifically hybrid Q-ToF and Q-Orbitrap MS, has gained increasing popularity for benchtop exposomics applications. High resolution and sensitivity are achieved, owing to the ever-evolving MS analyzers, faster electronics, enhanced ion optics, and improved detector technologies. The pre-MS steps, from sample extraction/fractionation to chromatographic (and/or ion

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

mobility) separation of analytes, are versatile and increasingly streamlined, as are the respective ionization techniques. Although the experimental advances show promise to tackle the all-encompassing exposome, a trade-off between budget, throughput, and chemical space coverage persists in practice and varies from case to case. As listed in Figure 5, certain challenges and opportunities remain for chemical space mapping and workflow development, both of which constitute the prerequisites for enabling population studies of the chemical exposome at much larger scales. In companion with Figure 5 and Table 1, this section selectively discusses some pertinent specifics, spanning from reference and standardization, benchmarking HRMS workflows, to data science and statistical approaches for meaningful exposome-based health inferences.

6.1. Standards, Reference Materials, and Data Formats. Despite the continuous efforts of NIST[328] and vendors (e.g., BioIVT, biocrates) to advance MS-based applications, the development and standardization of reference components remain a deeply unmet research need in HRMS-based exposomics. Notable types of reference span chemical standards, certified reference materials (CRM) (e.g., sample matrices), standardized protocols, analytical profiling kits, etc. These standards and CRMs are essential since they serve as benchmarks, baselines, and navigation points to bolster human chemical exposome studies by time, scale, and study design through workflow validation, quality assurance and calibration, platform comparison, and cross-laboratory harmonization.[380−382] Chemical standards of environmental chemicals can be procured either from vendors, research agencies (e.g., the EPA-housed *ToxCast* library), or synthesized through a fee-for-service mechanism. For reference materials, ideally, traceability, homogeneity, stability, and longevity are desired. Upon study design, the health relevance needs to be justified with matched matrices and system suitability. Common materials include blood and urine for human/biological samples and water and dust for environmental samples. However, challenges will be encountered if more heterogeneous and complex sample matrices are to be developed.

On the informatics and data analytics side, a few considerations ensue, spanning criteria of library matching, spectral data acquisition, use of metadata, and standardized reporting. First, although both Schymanski and Koelmel scales are released for actionable use to score annotation confidence, operationally, there exists a lack of consensus on the acceptable ranges for mass accuracy, RT drift, and CCS accuracy (if applicable) to define a "hit" by library matching. These may depend on platform, assay, or specific compounds and might change with the evolving technologies, making a community consensus difficult to reach. However, transparency in reporting these criteria should be encouraged at least to ensure reproducibility. Second, decisions should be made regarding the specifics of spectral data generation methods, from details like the collision energy to use (for MS^2 acquisition)[383] to the selection between profile mode and centroid mode.[300] Of note, profile mode data maintains the original entire continuous m/z signals but can result in large data size, whereas centroid spectral data record centroided data, i.e., discrete peaks through select sampling of the maximum intensity at a specific m/z (by a predefined window) and thus have much-reduced data sizes. Recovery of centroiding-induced information loss has been attempted to strike a balance between data size and information density (as well as authenticity).[384] Third, to abide by the FAIR principles,[385] centralized public depositories for exposomics data are needed, and raw data in various vendor formats are to be converted to standard data formats such as *.mzML[386] and *.mzXML.[387] Procedural details, use of metadata, and reporting of results can all be encapsulated into a standard format such as ISA-Tab[388] (for storing metadata) and mzTab[389] (metadata and procedures). Improvements as such together are integral to the effective construction of experimental and computational infrastructures.

6.2. Benchmarking HRMS Resources: From Workflow Specifics to Reference Spectral Libraries. To expand the analytical coverage by chemical exposomics, both GC-HRMS and LC-HRMS are needed. The former has been expanding rapidly compared to the latter which is more established.[356,390] To benchmark GC-HRMS, certain specifics are to be noted. First, for peak picking, most software was designed for LC-MS data; it is crucial to devise new algorithms with tunable parameters for addressing nuances of GC-MS data, such as sharper peak shapes (due to larger peak capacity) compared to LC-MS data. Second, because of extensive ion fragmentation (under EI), baseline estimation remains a hurdle for blank correction in GC to differentiate noises from low-abundant signals for removing unwanted interfering peaks. Third, for peak alignment, GC-MS data uses (linear) RIs, which can be advantageous for reproducible results compared to raw RT when significant RT shift occurs; universal RI development and application for LC, however, remains unfledged due to wide-ranging factors (e.g., LC mobile phase compositions) and the sheer number of LC-amenable molecules.[339] It should also be noted that GC-MS analyses are more prone to batch effect and matrix effect than LC-MS, thereby requiring more frequent tune/calibration, systematic evaluation, and better data cleanup and normalization strategies.[120,391,392]

For cheminformatics down the line, GC-HRMS annotation remains a bottleneck since most current GC-EI-MS spectral libraries are based on unit mass while having a high level of ion fragmentation. Thus, more than a list of m/z and RT/RI, spectral libraries that catalog HRMS-based fragments are required for identification purposes.[118] As GC-HRMS has started to expand only in the recent decade, most investigators have to rely on in-house libraries, which hardly cover thousands of compounds for suspect screening and NTA. One pressing issue is how to leverage the established low-resolution, unit-mass mass spectral libraries (e.g., NIST EI GC-MS library) to analyze the growing GC-HRMS data. Ideally, HRMS data are matched against accurate mass spectral libraries for compound identification. However, accurate-mass library generation is emerging; efforts are being undertaken to expand the number of entries to match that of the traditionally available unit-mass libraries. While public/commercial accurate-mass libraries remain largely not available, researchers using GC-HRMS first created libraries specific for hundreds of exposome molecules and released these to the public.[393] Alternatively, low-resolution spectral libraries can be converted into pseudo-high-resolution spectra as far as structures (and substructures) are unambiguously known or identified.[394] Nevertheless, unit-mass spectral libraries can still be useful. For example, one can use High-Resolution Factor (HRF)[395] to validate library hits (as implemented in Compound Discoverer-GC). If the library hit is correct, one should be able to explain every observed fragment using the subformula (or subset of the elemental formula) of that library

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

hit. The extent to which one can explain the fragments based on the elemental formula of the library hit is calculated as an HRF, with the process itself coined "high-resolution filtering."[395]

6.3. Data Science for Exposome-Based Health Inference: Approaches and Future Prospects. Long-term and large-scale human exposome research relies on computational infrastructures that can support the latest developments in chemical analysis. In turn, improvements in HRMS measurements, preprocessing algorithms, and annotation coverage will ultimately advance statistical analysis. Ideally, the experimental workflows and cheminformatics pipelines are streamlined and harmonized at the front end, permitting flexible statistical strategies to mine the exposomics datasets for meaningful health inference. Wide-ranging data science approaches can be employed to determine the health effects associated with exposures measured in a more comprehensive way, allowing for unbiased and effective assessment of drivers and modifiers of disease (Figure 2c). The challenges may originate from data complexity and sparsity, determining the combined effect of exposure in mixtures, omics data integration, and interpretability, which together motivate the development of next-generation data science for exposome research.[290,396]

The relationships between exposures and disease phenotypes are usually first assessed by univariate statistical approaches. Univariate analyses with multiple comparison corrections are considered sensitive and robust and still remain one of the most used methods in ExWAS/MWAS, taking each individual exposure as an independent variable to associate with the disease outcome.[397] To delve into effects and interactions, multivariate approaches are needed.[396] First, to overcome the challenge of high dimensionality in data where predictors (i.e., ion features or annotated exposures) far outnumber observations, dimensionality reduction approaches (e.g., principal component analysis [PCA], non-negative matrix factorization [NMF]) should be used to reduce noise and capture the essence of variability in data.[398] Such dimensional reduction facilitates feature extraction, classification, and visualization, as being utilized to explore the structure of one matrix (unsupervised methods) or estimate the relationship between exposures (one matrix) and an outcome (supervised methods). These are especially useful for association analysis (e.g., ExWAS), network analysis, and multi-omics integration to tease out exposure−disease links.[398,399]

Second, as chemical exposures commonly occur in complex mixtures, mixture modeling approaches can be utilized to identify potential synergistic or antagonistic effects of exposures and tease out individual toxic agents. Specific statistical analysis decisions would depend on the data structure and research question. Still, for a generic walkthrough, the overall effects of mixtures can be measured by Bayesian Kernel Machine Regression (BKMR),[400] Weighted Quantile Sum (WQS),[401] and Bayesian hierarchical models (when a hierarchical structure is suspected in data).[396,402] If overall effects are observed, one may go on to search for main contributors using penalized methods (e.g., elastic net, horseshoe regression), BKMR, WQS, and random forest. The interactions and nonlinearities can be further assessed using BKMR, random forest, etc. Third, machine learning (ML) approaches such as random forest, gradient boosting, support vector machine, and neural networks are advantageous in the search for features/biomarkers for classification and prediction of outcomes. In pattern recognition and model fitting, typically, ML conducts hyperparameter optimization and stacks all the results (i.e., the "ensemble" method) for best modeling/diagnostic performance.[403] How to balance such complexity/accuracy and model interpretability remains a challenge. Recently, an interpretable neural network (NN)-based framework was implemented for untargeted HRMS-based blood metabolomics data in a Parkinson's disease cohort.[404] This approach integrated model parameters (without the need for preselecting ion features), retrospective mining of key features contributing the most to an accurate model prediction (i.e., "interpreting" ML models), and benchmark testing of multiple ML methods for comparison and result validation, outperforming all other ML methods tested.[404]

Mediation analysis assesses the indirect effects of exposures through intermediate variables and can be used to reveal pathways and mechanisms of action.[405] To advance causal inference, one may take the Mendelian Randomization (MR) approach to the exposomics data (combined with genomics data) to establish causal relationships while overcoming confounding and reverse causation biases.[291,406] Once genome-wide association studies (GWAS) were conducted, MR considers SNPs associated with the outcome on individual exposome agents as instrumental variables to partition the cohort to determine if evidence of causality can be derived, with the assumption that variants are associated with exposures but not directly with confounders or outcomes except through the exposure.[291,406,407] Together, these emerging data science approaches show promise in molecular epidemiology and health sciences for exposure assessment, biomarker hunting, and causal inference, and will set the stage for mining the reams of newly generated exposomics data into interpretable and translational health insights.

7. IMPLICATIONS

The recent rise of HRMS and associated informatics approaches have ushered in an unprecedented new era to advance human exposome research. From hardware (for yielding better and more health-relevant data) to software (for translating such yielded data into meaningful information and insights into chemistry and biology), opportunities and challenges abound for expanding the analytical coverage. The recent launch of large-scale research programs and initiatives such as NIH "*All of Us*" and EHEN holds promise for improved exposomics study design (by statistical power, metadata, etc.) in the forthcoming years. Meanwhile, it highlights a need for an upgrade in the exposomic workflow toward expanded analytical coverage, paving ways for harmonized and scalable analysis with strategies, feasibilities, and prospects critically outlined in this article. It should be expected that longitudinal tracking, retrospective validation, and multi-omics analyses based on HRMS-based exposomics data, alongside novel mixture modeling and causal inference frameworks down the line, will benefit immensely from these analytical endeavors at the front end.[408]

Human exposomics, the transdisciplinary field that studies the exposome, is designed to enable discovery-based analysis of the environmental factors that contribute to disease. HRMS-aided chemical exposomics, at the leading edge with ever-increasing analytical comprehensiveness and accuracy, is now transforming precision medicine and precision environmental health. For example, with accumulating HRMS data and large-

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

scale cohorts in place, subjects are categorized and stratified based on the measured medication profiles and comorbidities (for pharmacology clinical trials and epidemiology), food questionnaires are augmented or replaced by hard data (for nutrition science), and environmental exposures are screened and monitored timely for effective prevention, intervention, and regulation (for environmental sciences). To illuminate the *Genome* × *Exposome* interplay, HRMS-based approaches at the forefront of exposomics help identify and quantify the non-genetic drivers of health and disease outcomes. We hope that the many technical issues and strategies reviewed in this article merit the attention of chemists (environmental, analytical, food, and pharmaceutical), toxicologists, epidemiologists, engineers, and physician-scientists as they pursue exposome-oriented research, as improvements in exposomics will drive improvements in human and environmental health.

AUTHOR INFORMATION

Corresponding Author

Gary W. Miller − *Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, New York 10032, United States;* orcid.org/0000-0001-8984-1284; Email: gm2815@cumc.columbia.edu

Authors

Yunjia Lai − *Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, New York 10032, United States;* orcid.org/0000-0002-1081-0897

Jeremy P. Koelmel − *Department of Environmental Health Sciences, Yale School of Public Health, New Haven, Connecticut 06520, United States*

Douglas I. Walker − *Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia 30322, United States*

Elliott J. Price − *RECETOX, Faculty of Science, Masaryk University, 611 37 Brno, Czech Republic;* orcid.org/0000-0001-5691-7000

Stefano Papazian − *Department of Environmental Science, Science for Life Laboratory, Stockholm University, SE-106 91 Stockholm, Sweden; National Facility for Exposomics, Metabolomics Platform, Science for Life Laboratory, Stockholm University, Solna 171 65, Sweden;* orcid.org/0000-0003-2538-8702

Katherine E. Manz − *Department of Environmental Health Sciences, School of Public Health, University of Michigan, Ann Arbor, Michigan 48109, United States*

Delia Castilla-Fernández − *Department of Food Chemistry and Toxicology, Faculty of Chemistry, University of Vienna, 1010 Vienna, Austria*

John A. Bowden − *Center for Environmental and Human Toxicology, Department of Physiological Sciences, College of Veterinary Medicine, University of Florida, Gainesville, Florida 32611, United States;* orcid.org/0000-0003-3135-4015

Vladimir Nikiforov − *NILU, Fram Centre, Tromsø NO-9296, Norway*

Arthur David − *Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) − UMR_S, 1085 Rennes, France;* orcid.org/0000-0001-8189-1826

Vincent Bessonneau − *Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) − UMR_S, 1085 Rennes, France*

Bashar Amer − *Thermo Fisher Scientific, San Jose, California 95134, United States*

Suresh Seethapathy − *Thermo Fisher Scientific, Somerset, New Jersey 08873, United States*

Xin Hu − *Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia 30322, United States*

Elizabeth Z. Lin − *Department of Environmental Health Sciences, Yale School of Public Health, New Haven, Connecticut 06520, United States*

Akrem Jbebli − *RECETOX, Faculty of Science, Masaryk University, 611 37 Brno, Czech Republic*

Brooklynn R. McNeil − *Biomarkers Core Laboratory, Irving Institute for Clinical and Translational Research, Columbia University Irving Medical Center, New York, New York 10032, United States*

Dinesh Barupal − *Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, New York 10029, United States;* orcid.org/0000-0002-9954-8628

Marina Cerasa − *Institute of Atmospheric Pollution Research, Italian National Research Council, 00015 Monterotondo, Rome, Italy*

Hongyu Xie − *Department of Environmental Science, Science for Life Laboratory, Stockholm University, SE-106 91 Stockholm, Sweden;* orcid.org/0000-0002-2422-0492

Vrinda Kalia − *Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, New York 10032, United States;* orcid.org/0000-0003-2244-6663

Renu Nandakumar − *Biomarkers Core Laboratory, Irving Institute for Clinical and Translational Research, Columbia University Irving Medical Center, New York, New York 10032, United States*

Randolph Singh − *Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, New York 10032, United States;* orcid.org/0000-0003-4500-3400

Zhenyu Tian − *Department of Chemistry and Chemical Biology, Northeastern University, Boston, Massachusetts 02115, United States;* orcid.org/0000-0002-7491-7028

Peng Gao − *Department of Environmental and Occupational Health, and Department of Civil and Environmental Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, United States; UPMC Hillman Cancer Center, Pittsburgh, Pennsylvania 15232, United States;* orcid.org/0000-0002-4311-584X

Yujia Zhao − *Institute for Risk Assessment Sciences, Utrecht University, Utrecht 3584CM, The Netherlands;* orcid.org/0000-0002-5275-0036

Jean Froment − *NILU, NO-2027 Kjeller, Norway;* orcid.org/0000-0003-4844-2919

Pawel Rostkowski − *NILU, NO-2027 Kjeller, Norway;* orcid.org/0000-0001-9778-4283

Saurabh Dubey − *Biomarkers Core Laboratory, Irving Institute for Clinical and Translational Research, Columbia University Irving Medical Center, New York, New York 10032, United States*

Kateřina Coufalíková − *RECETOX, Faculty of Science, Masaryk University, 611 37 Brno, Czech Republic*

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

Hana Seličová − *RECETOX, Faculty of Science, Masaryk University, 611 37 Brno, Czech Republic*
Helge Hecht − *RECETOX, Faculty of Science, Masaryk University, 611 37 Brno, Czech Republic;* orcid.org/0000-0001-6744-996X
Sheng Liu − *Department of Environmental Health Sciences, Yale School of Public Health, New Haven, Connecticut 06520, United States*
Hanisha H. Udhani − *Biomarkers Core Laboratory, Irving Institute for Clinical and Translational Research, Columbia University Irving Medical Center, New York, New York 10032, United States*
Sophie Restituito − *Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, New York 10032, United States*
Kam-Meng Tchou-Wong − *Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, New York 10032, United States*
Kun Lu − *Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, United States;* orcid.org/0000-0002-8125-2394
Jonathan W. Martin − *Department of Environmental Science, Science for Life Laboratory, Stockholm University, SE-106 91 Stockholm, Sweden; National Facility for Exposomics, Metabolomics Platform, Science for Life Laboratory, Stockholm University, Solna 171 65, Sweden;* orcid.org/0000-0001-6265-4294
Benedikt Warth − *Department of Food Chemistry and Toxicology, Faculty of Chemistry, University of Vienna, 1010 Vienna, Austria;* orcid.org/0000-0002-6104-0706
Krystal J. Godri Pollitt − *Department of Environmental Health Sciences, Yale School of Public Health, New Haven, Connecticut 06520, United States;* orcid.org/0000-0001-7332-2228
Jana Klánová − *RECETOX, Faculty of Science, Masaryk University, 611 37 Brno, Czech Republic*
Oliver Fiehn − *West Coast Metabolomics Center, University of California−Davis, Davis, California 95616, United States;* orcid.org/0000-0002-6261-8928
Thomas O. Metz − *Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99354, United States;* orcid.org/0000-0001-6049-3968
Kurt D. Pennell − *School of Engineering, Brown University, Providence, Rhode Island 02912, United States;* orcid.org/0000-0002-5788-6397
Dean P. Jones − *Department of Medicine, School of Medicine, Emory University, Atlanta, Georgia 30322, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.4c01156

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (NIH) through grant awards to G.W.M. (No. U2CES030163, National Institute of Environmental Health Sciences, NIEHS; R01AG067501, RF1AG066107, National Institute of Aging, NIA; UL1TR001873, National Center for Advancing Translational Sciences, NCATS), D.I.W. (R01ES032831, NIEHS), K.E.M. (K01ES035398, NIEHS), P.G. (R21ES036033, NIEHS), O.F. (R03OD034497, NIH), K.L. (P42ES031007, P30ES010126, NIEHS), and K.D.P. (R21ES034187, NIEHS). P.G. is a mentee of the NIH "Career MODE" Program through funding to G.W.M. (R25GM143298, National Institute of General Medical Sciences, NIGMS). K.L. acknowledges the support of the UNC Superfund Research Program and the UNC Center for Environmental Health and Susceptibility. The article content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. This work was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 857560 (CETOCOEN Excellence). This publication reflects only the author's view, and the European Commission is not responsible for any use that may be made of the information it contains. E.J.P., A.J., K.C., H.S., H.H. and J.K. acknowledge the research infrastructure RECETOX RI (LM2023069) and OP RDE (CZ.02.1.01/0.0/0.0/17_043/0009632). J.W.M. and H.X. acknowledge funding from the Swedish Research Council (VR, 2018-03409) and Formas (2018-02268). J.W.M. and S.P. acknowledge Swedish national funding to the National Facility for Exposomics (SciLifeLab). B.W. acknowledges that co-funding was provided by the European Union (ERC, EXPOMET 101043321 to B.W.). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. A.D. and V.B. acknowledge the research infrastructure France Exposome. B.A. is a Thermo Fisher employee. Views and opinions of this work are solely of the authors and do not represent any mass spectrometry manufacturers or associated commercial vendors. T.O.M. acknowledges support from the Pacific Northwest National Laboratory (PNNL) Laboratory Directed Research and Development Program, via the *m/q* Initiative. Battelle operates PNNL for the Department of Energy (DOE) under Contract DE-AC05-76RLO01830. The authors also acknowledge the GC-Orbitrap User Meeting mechanism led by K.J.D.P. (Yale University) and D.I.W. (Emory University) for the many stimulating ideas, critiques, and discussions that brought this work to fruition. The authors thank Drs. Hiroshi Tsugawa (TUAT, Japan) and Shuzhao Li (The Jackson Laboratory, CT, USA) for the consultation on databases and data preprocessing. The authors extend sincere gratitude to all the scientists, engineers, healthcare providers, and governmental/business stakeholders who have inspired, contributed, and offered support to the burgeoning field of the exposome.

■ GLOSSARY

Chemical space: the total collection of all possible molecules (theoretically or empirically) in a given context, with unique chemical structures, physicochemical properties, and functional activities.
Human exposome in chemical space: the total collection of (i) chemical exposure agents humans are being exposed to, (ii) transformation products *in vivo*, and (iii) biomolecules indicative of a toxicological and/or etiologic effect in question.
Analytical coverage: the performance of an analytical workflow and the associated data pipeline in covering the chemical space in question, by comprehensiveness, accuracy, and dynamic range.

Chemical exposomics: the omics-scale measurement of small-molecule exposure agents, transformation products, and associated biomolecules through targeted and/or suspect approaches for expected and known compounds, and non-targeted approaches for unexpected or unknown compounds.

High-resolution mass spectrometry (HRMS): an advanced analytical technique used to identify and quantify molecules based on their mass-to-charge ratio (m/z) and associated chemical transformation with high accuracy and precision. High resolution (10,000−50,000 fwhm, 3−10 ppm mass accuracy) and ultrahigh resolution (>50k fwhm, < 3 ppm mass accuracy) measurements are important for deriving meaningful formula with minimal mass interferences. HRMS can be flexibly coupled to GC, LC, IMS, or similar separation modules at the front end.

Exposome-Wide Association Study (ExWAS): statistical equivalent to Genome-Wide Association Study (GWAS) to interrogate environmental contributors to health and disease.

Non-targeted analysis (NTA): the analytical approach to measure a broad range of environmental exposures without limiting methods to prior knowledge of sample content.

Passive sampling: a sampling approach that accumulates target compounds over time, usually relying on the natural diffusion of compounds into the sampling medium (as opposed to grab sampling or active extraction).

Matrix effect: A matrix effect is any influence that the substrate (e.g., tissue, blood, water, or solvent) has on the analytical performance of a technique. This is typically characterized by ion suppression or ion enhancement that hampers detection and quantitative accuracy for given analytes of interest, likely due to the presence and concomitant ionization of coexisting molecules and/or overlapping signals of interferences in the matrix.

Spectral deconvolution (in data processing): a computational process that separates a complex spectrum of multiple co-eluting components (e.g., from GC-EI or LC-DIA) to generate a clean spectrum for each single component (i.e., putative compound).

Feature: an ion peak or analytical component with distinct m/z (MS^1) and retention time combination as algorithmically identified and componentized. Features are interchangeably referred to as ion feature or m/z feature and may or may not have associated fragmentation (MS^2) information.

Data preprocessing: a series of informatics steps converting raw HRMS data into a tabular, numerical format for follow-up data treatment (e.g., data normalization, cleanup), statistics, and informatic analysis.

Compound annotation: the process of assigning confidence chemical identities to ion features. It relies on multiple source evidence and orthogonal information, including retention time, MS^1 and MS^2 data, and spectral matching on available databases.

Heuristic rules: Simple, straightforward, and often empirical shortcuts for effective problem-solving (in HRMS-based exposomics for compound annotation).

Contaminants of Emerging Concerns (CECs): synthetic or naturally occurring compounds that have not been regulated but raise increasing concern due to (potentially) harmful effects on human and ecosystem health.

REFERENCES

(1) Wild, C. P. Complementing the genome with an "exposome": The outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidem Biomar* **2005**, *14* (8), 1847−1850.

(2) Vermeulen, R.; Schymanski, E. L.; Barabasi, A. L.; Miller, G. W. The exposome and health: Where chemistry meets biology. *Science* **2020**, *367* (6476), 392−396.

(3) Miller, G. W.; Jones, D. P. The nature of nurture: refining the definition of the exposome. *Toxicol. Sci.* **2014**, *137* (1), 1−2.

(4) Nriagu, J. O.; Pacyna, J. M. Quantitative Assessment of Worldwide Contamination of Air, Water and Soils by Trace-Metals. *Nature* **1988**, *333* (6169), 134−139.

(5) Rochman, C. M.; Browne, M. A.; Halpern, B. S.; Hentschel, B. T.; Hoh, E.; Karapanagioti, H. K.; Rios-Mendoza, L. M.; Takada, H.; Teh, S.; Thompson, R. C. Classify plastic waste as hazardous. *Nature* **2013**, *494* (7436), 169−171.

(6) Manisalidis, I.; Stavropoulou, E.; Stavropoulos, A.; Bezirtzoglou, E. Environmental and Health Impacts of Air Pollution: A Review. *Front. Public Health* **2020**, DOI: 10.3389/fpubh.2020.00014.

(7) Susmann, H. P.; Schaider, L. A.; Rodgers, K. M.; Rudel, R. A. Dietary Habits Related to Food Packaging and Population Exposure to PFASs. *Environ. Health Perspect* **2019**, *127* (10), No. 107003.

(8) Liu, S.; Hammond, S. K.; Rojas-Cheatham, A. Concentrations and potential health risks of metals in lip products. *Environ. Health Perspect* **2013**, *121* (6), 705−710.

(9) Ural, B. B.; Caron, D. P.; Dogra, P.; Wells, S. B.; Szabo, P. A.; Granot, T.; Senda, T.; Poon, M. M. L.; Lam, N.; Thapa, P.; et al. Inhaled particulate accumulation with age impairs immune function and architecture in human lung lymph nodes. *Nat. Med.* **2022**, *28* (12), 2622−2632.

(10) Hernández, A. F.; Gil, F.; Lacasaña, M. Toxicological interactions of pesticide mixtures: an update. *Arch. Toxicol.* **2017**, *91* (10), 3211−3223.

(11) Chi, Z. H.; Goodyer, C. G.; Hales, B. F.; Bayen, S. Characterization of different contaminants and current knowledge for defining chemical mixtures in human milk: A review. *Environ. Int.* **2023**, *171*, No. 107717.

(12) Polderman, T. J. C.; Benyamin, B.; de Leeuw, C. A.; Sullivan, P. F.; van Bochoven, A.; Visscher, P. M.; Posthuma, D. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **2015**, *47* (7), 702.

(13) Shen, J.; Liao, Y.; Hopper, J. L.; Goldberg, M.; Santella, R. M.; Terry, M. B. Dependence of cancer risk from environmental exposures on underlying genetic susceptibility: an illustration with polycyclic aromatic hydrocarbons and breast cancer. *Br. J. Cancer* **2017**, *116* (9), 1229−1233.

(14) Argentieri, M. A.; Amin, N.; Nevado-Holgado, A. J.; Sproviero, W.; Collister, J. A.; Keestra, S. M.; Doherty, A.; Hunter, D. J.; Alvergne, A.; Duijn, C. M. v. Integrating the environmental and genetic architectures of mortality and aging. *medRxiv*, May 16, 2023. DOI: 10.1101/2023.03.10.23286340.

(15) Price, E. J.; Vitale, C. M.; Miller, G. W.; David, A.; Barouki, R.; Audouze, K.; Walker, D. I.; Antignac, J.-P.; Coumoul, X.; Bessonneau, V.; et al. Merging the exposome into an integrated framework for "omics" sciences. *iScience* **2022**, *25* (3), No. 103976.

(16) David, A.; Chaker, J.; Price, E. J.; Bessonneau, V.; Chetwynd, A. J.; Vitale, C. M.; Klanova, J.; Walker, D. I.; Antignac, J. P.; Barouki, R.; et al. Towards a comprehensive characterisation of the human internal chemical exposome: Challenges and perspectives. *Environ. Int.* **2021**, *156*, No. 106630.

(17) Flasch, M.; Bueschl, C.; Del Favero, G.; Adam, G.; Schuhmacher, R.; Marko, D.; Warth, B. Elucidation of xenoestrogen metabolism by non-targeted, stable isotope-assisted mass spectrometry in breast cancer cells. *Environ. Int.* **2022**, *158*, No. 106940.

(18) The "All of Us" Research Program. *N. Engl. J. Med.* **2019**, *381* (7), 668−676. DOI: 10.1056/NEJMsr1809937.

(19) Hartung, T. A call for a Human Exposome Project. *ALTEX* **2023**, *40* (1), 4−33.

(20) Griffiths, J. A brief history of mass spectrometry. *Anal. Chem.* **2008**, *80* (15), 5678−5683.
(21) Sabbioni, G.; Berset, J.-D.; Day, B. W. Is It Realistic to Propose Determination of a Lifetime Internal Exposome? *Chem. Res. Toxicol.* **2020**, *33* (8), 2010−2021.
(22) Chait, B. T. Mass spectrometry in the postgenomic era. *Annu. Rev. Biochem.* **2011**, *80*, 239−246.
(23) Sun, J.; Fang, R.; Wang, H.; Xu, D.-X.; Yang, J.; Huang, X.; Cozzolino, D.; Fang, M.; Huang, Y. A review of environmental metabolism disrupting chemicals and effect biomarkers associating disease risks: Where exposomics meets metabolomics. *Environ. Int.* **2022**, *158*, No. 106941.
(24) Fiehn, O. Metabolomics − the link between genotypes and phenotypes. *Plant Molecular Biology* **2002**, *48* (1), 155−171.
(25) Hu, X.; Walker, D. I.; Liang, Y.; Smith, M. R.; Orr, M. L.; Juran, B. D.; Ma, C.; Uppal, K.; Koval, M.; Martin, G. S.; et al. A scalable workflow to characterize the human exposome. *Nat. Commun.* **2021**, *12* (1), 5575.
(26) Go, Y.-M.; Walker, D. I.; Liang, Y.; Uppal, K.; Soltow, Q. A.; Tran, V.; Strobel, F.; Quyyumi, A. A.; Ziegler, T. R.; Pennell, K. D.; et al. Reference Standardization for Mass Spectrometry and High-resolution Metabolomics Applications to Exposome Research. *Toxicol. Sci.* **2015**, *148* (2), 531−543.
(27) Rappaport, S. M.; Barupal, D. K.; Wishart, D.; Vineis, P.; Scalbert, A. The blood exposome and its role in discovering causes of disease. *Environ. Health Perspect* **2014**, *122* (8), 769−774.
(28) Chaker, J.; Gilles, E.; Léger, T.; Jégou, B.; David, A. From Metabolomics to HRMS-Based Exposomics: Adapting Peak Picking and Developing Scoring for MS1 Suspect Screening. *Anal. Chem.* **2021**, *93* (3), 1792−1800.
(29) Hyötyläinen, T. Analytical challenges in human exposome analysis with focus on environmental analysis combined with metabolomics. *J. Sep. Sci.* **2021**, *44* (8), 1769−1787.
(30) Sdougkou, K.; Xie, H.; Papazian, S.; Bonnefille, B.; Bergdahl, I. A.; Martin, J. W. Phospholipid Removal for Enhanced Chemical Exposomics in Human Plasma. *Environ. Sci. Technol.* **2023**, *57* (28), 10173−10184.
(31) Chaker, J.; Kristensen, D. M.; Halldorsson, T. I.; Olsen, S. F.; Monfort, C.; Chevrier, C.; Jégou, B.; David, A. Comprehensive Evaluation of Blood Plasma and Serum Sample Preparations for HRMS-Based Chemical Exposomics: Overlaps and Specificities. *Anal. Chem.* **2022**, *94* (2), 866−874.
(32) Parera, J.; Abalos, M.; Kärrman, A.; van Bavel, B.; Abad, E. 1.03 - Assessing and Controlling Sample Contamination. In *Comprehensive Sampling and Sample Preparation*; Pawliszyn, J., Ed.; Academic Press, 2012; pp 51−64.
(33) Psychogios, N.; Hau, D. D.; Peng, J.; Guo, A. C.; Mandal, R.; Bouatra, S.; Sinelnikov, I.; Krishnamurthy, R.; Eisner, R.; Gautam, B.; et al. The Human Serum Metabolome. *PLoS One* **2011**, *6* (2), No. e16957.
(34) David, A.; Abdul-Sada, A.; Lange, A.; Tyler, C. R.; Hill, E. M. A new approach for plasma (xeno)metabolomics based on solid-phase extraction and nanoflow liquid chromatography-nanoelectrospray ionisation mass spectrometry. *Journal of Chromatography A* **2014**, *1365*, 72−85.
(35) Sanchez, T. R.; Hu, X.; Zhao, J.; Tran, V.; Loiacono, N.; Go, Y.-M.; Goessler, W.; Cole, S.; Umans, J.; Jones, D. P.; et al. An atlas of metallome and metabolome interactions and associations with incident diabetes in the Strong Heart Family Study. *Environ. Int.* **2021**, *157*, No. 106810.
(36) Lipinski, C.; Hopkins, A. Navigating chemical space for biology and medicine. *Nature* **2004**, *432* (7019), 855−861.
(37) Stockwell, B. R. Exploring biology with small organic molecules. *Nature* **2004**, *432* (7019), 846−854.
(38) Ruan, T.; Li, P.; Wang, H.; Li, T.; Jiang, G. Identification and Prioritization of Environmental Organic Pollutants: From an Analytical and Toxicological Perspective. *Chem. Rev.* **2023**, *123* (17), 10584−10640.
(39) Gouin, T.; Mackay, D.; Webster, E.; Wania, F. Screening Chemicals for Persistence in the Environment. *Environ. Sci. Technol.* **2000**, *34* (5), 881−884.
(40) Reymond, J.-L. The Chemical Space Project. *Acc. Chem. Res.* **2015**, *48* (3), 722−730.
(41) Hulleman, T.; Turkina, V.; O'Brien, J. W.; Chojnacka, A.; Thomas, K. V.; Samanipour, S. Critical Assessment of the Chemical Space Covered by LC−HRMS Non-Targeted Analysis. *Environ. Sci. Technol.* **2023**, *57* (38), 14101−14112.
(42) Wang, Z.; Walker, G. W.; Muir, D. C. G.; 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.
(43) Zhang, P.; Carlsten, C.; Chaleckis, R.; Hanhineva, K.; Huang, M.; Isobe, T.; Koistinen, V. M.; Meister, I.; Papazian, S.; Sdougkou, K.; et al. Defining the Scope of Exposome Studies and Research Needs from a Multidisciplinary Perspective. *Environ. Sci. Technol. Lett.* **2021**, *8* (10), 839−852.
(44) Chen, Y. C.; Hsu, J. F.; Chang, C. W.; Li, S. W.; Yang, Y. C.; Chao, M. R.; Chen, H. C.; Liao, P. C. Connecting chemical exposome to human health using high-resolution mass spectrometry-based biomonitoring: Recent advances and future perspectives. *Mass Spectrom. Rev.* **2022**, No. e21805.
(45) Isaacs, K. K.; Egeghy, P.; Dionisio, K. L.; Phillips, K. A.; Zidek, A.; Ring, C.; Sobus, J. R.; Ulrich, E. M.; Wetmore, B. A.; Williams, A. J.; et al. The chemical landscape of high-throughput new approach methodologies for exposure. *J. Expo Sci. Environ. Epidemiol* **2022**, *32* (6), 820−832.
(46) Manz, K. E.; Feerick, A.; Braun, J. M.; Feng, Y. L.; Hall, A.; Koelmel, J.; Manzano, C.; Newton, S. R.; Pennell, K. D.; Place, B. J.; et al. Non-targeted analysis (NTA) and suspect screening analysis (SSA): a review of examining the chemical exposome. *J. Expo Sci. Environ. Epidemiol* **2023**, *33* (4), 524−536.
(47) Black, G.; Lowe, C.; Anumol, T.; Bade, J.; Favela, K.; Feng, Y.-L.; Knolhoff, A.; McEachran, A.; Nuñez, J.; Fisher, C.; et al. Exploring chemical space in non-targeted analysis: a proposed ChemSpace tool. *Anal. Bioanal. Chem.* **2023**, *415* (1), 35−44.
(48) Schymanski, E. L.; Kondic, T.; Neumann, S.; Thiessen, P. A.; Zhang, J.; Bolton, E. E. Empowering large chemical knowledge bases for exposomics: PubChemLite meets MetFrag. *J. Cheminform* **2021**, *13* (1), 19.
(49) Neveu, V.; Moussy, A.; Rouaix, H.; Wedekind, R.; Pon, A.; Knox, C.; Wishart, D. S.; Scalbert, A. Exposome-Explorer: a manually-curated database on biomarkers of exposure to dietary and environmental factors. *Nucleic Acids Res.* **2017**, *45* (D1), D979−D984.
(50) Borrel, A.; Conway, M.; Nolte, S. Z; Unnikrishnan, A.; Schmitt, C. P; Kleinstreuer, N. C ChemMaps.com v2.0: exploring the environmental chemical universe. *Nucleic Acids Res.* **2023**, *51* (W1), W78−W82.
(51) 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.; et al. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *J. Cheminform* **2017**, *9* (1), 61.
(52) Weisgerber, D. W. Chemical Abstracts Service Chemical Registry System: History, scope, and impacts. *Journal of the American Society for Information Science* **1997**, *48* (4), 349−360.
(53) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; et al. PubChem 2023 update. *Nucleic Acids Res.* **2023**, *51* (D1), D1373−D1380.
(54) Pellizzari, E. D.; Woodruff, T. J.; Boyles, R. R.; Kannan, K.; Beamer, P. I.; Buckley, J. P.; Wang, A.; Zhu, Y.; Bennett, D. H. Identifying and Prioritizing Chemicals with Uncertain Burden of Exposure: Opportunities for Biomonitoring and Health-Related Research. *Environ. Health Perspect* **2019**, *127* (12), No. 126001.
(55) Flasch, M.; Koellensperger, G.; Warth, B. Comparing the sensitivity of a low- and a high-resolution mass spectrometry approach for xenobiotic trace analysis: An exposome-type case study. *Anal. Chim. Acta* **2023**, *1279*, No. 341740.

(56) Oesterle, I.; Pristner, M.; Berger, S.; Wang, M.; Verri Hernandes, V.; Rompel, A.; Warth, B. Exposomic Biomonitoring of Polyphenols by Non-Targeted Analysis and Suspect Screening. *Anal. Chem.* **2023**, *95* (28), 10686−10694.
(57) Schwarzenbach, R. P.; Gschwend, P. M.; Imboden, D. M. *Environmental Organic Chemistry*; Wiley, 2017.
(58) Escher, B. I.; Stapleton, H. M.; Schymanski, E. L. Tracking complex mixtures of chemicals in our changing environment. *Science* **2020**, *367* (6476), 388−392.
(59) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; et al. QSAR Modeling: Where Have You Been? Where Are You Going To? *J. Med. Chem.* **2014**, *57* (12), 4977−5010.
(60) Rand-Weaver, M.; Margiotta-Casaluci, L.; Patel, A.; Panter, G. H.; Owen, S. F.; Sumpter, J. P. The Read-Across Hypothesis and Environmental Risk Assessment of Pharmaceuticals. *Environ. Sci. Technol.* **2013**, *47* (20), 11384−11395.
(61) Schmeisser, S.; Miccoli, A.; von Bergen, M.; Berggren, E.; Braeuning, A.; Busch, W.; Desaintes, C.; Gourmelon, A.; Grafström, R.; Harrill, J.; et al. New approach methodologies in human regulatory toxicology − Not if, but how and when! *Environ. Int.* **2023**, *178*, No. 108082.
(62) Sager, J. E.; Yu, J.; Ragueneau-Majlessi, I.; Isoherranen, N. Physiologically Based Pharmacokinetic (PBPK) Modeling and Simulation Approaches: A Systematic Review of Published Models, Applications, and Model Verification. *Drug Metab. Dispos.* **2015**, *43* (11), 1823−1837.
(63) Wang, X.; Zhao, X.; Shi, D.; Dong, Z.; Zhang, X.; Liang, W.; Liu, L.; Wang, X.; Wu, F. Integrating Physiologically Based Pharmacokinetic Modeling-Based Forward Dosimetry and in Vitro Bioassays to Improve the Risk Assessment of Organophosphate Esters on Human Health. *Environ. Sci. Technol.* **2023**, *57* (4), 1764−1775.
(64) Müller, F. A.; Stamou, M.; Englert, F. H.; Frenzel, O.; Diedrich, S.; Suter-Dick, L.; Wambaugh, J. F.; Sturla, S. J. In vitro to in vivo extrapolation and high-content imaging for simultaneous characterization of chemically induced liver steatosis and markers of hepatotoxicity. *Arch. Toxicol.* **2023**, *97* (6), 1701−1721.
(65) Tollefsen, K. E.; Scholz, S.; Cronin, M. T.; Edwards, S. W.; de Knecht, J.; Crofton, K.; Garcia-Reyero, N.; Hartung, T.; Worth, A.; Patlewicz, G. Applying Adverse Outcome Pathways (AOPs) to support Integrated Approaches to Testing and Assessment (IATA). *Regul. Toxicol. Pharmacol.* **2014**, *70* (3), 629−640.
(66) González-Domínguez, R.; Jáuregui, O.; Queipo-Ortuño, M. I.; Andrés-Lacueva, C. Characterization of the Human Exposome by a Comprehensive and Quantitative Large-Scale Multianalyte Metabolomics Platform. *Anal. Chem.* **2020**, *92* (20), 13767−13775.
(67) Gu, Y.; Peach, J. T.; Warth, B. Sample preparation strategies for mass spectrometry analysis in human exposome research: Current status and future perspectives. *TrAC Trends in Analytical Chemistry* **2023**, *166*, No. 117151.
(68) Brack, W.; Ait-Aissa, S.; Burgess, R. M.; Busch, W.; Creusot, N.; Di Paolo, C.; Escher, B. I.; Mark Hewitt, L.; Hilscherova, K.; Hollender, J.; et al. Effect-directed analysis supporting monitoring of aquatic environments–An in-depth overview. *Sci. Total Environ.* **2016**, *544*, 1073−1118.
(69) Vitale, C. M.; Price, E. J.; Miller, G. W.; David, A.; Antignac, J.-P.; Barouki, R.; Klánová, J. Analytical strategies for chemical exposomics: exploring limits and feasibility. *Exposome* **2021**, DOI: 10.1093/exposome/osab003.
(70) 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.
(71) Fitzgerald, C. C. J.; Hedman, R.; Uduwela, D. R.; Paszerbovics, B.; Carroll, A. J.; Neeman, T.; Cawley, A.; Brooker, L.; McLeod, M. D. Profiling Urinary Sulfate Metabolites With Mass Spectrometry. *Front. Mol. Biosci.* **2022**, DOI: 10.3389/fmolb.2022.829511.
(72) Xue, J.; Lai, Y.; Liu, C. W.; Ru, H. Towards Mass Spectrometry-Based Chemical Exposome: Current Approaches, Challenges, and Future Directions. *Toxics* **2019**, *7* (3), 41.
(73) Manrai, A. K.; Cui, Y.; Bushel, P. R.; Hall, M.; Karakitsios, S.; Mattingly, C. J.; Ritchie, M.; Schmitt, C.; Sarigiannis, D. A.; Thomas, D. C.; et al. Informatics and Data Analytics to Support Exposome-Based Discovery for Public Health. *Annu. Rev. Public Health* **2017**, *38*, 279−294.
(74) 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.
(75) Hajeb, P.; Zhu, L.; Bossi, R.; Vorkamp, K. Sample preparation techniques for suspect and non-target screening of emerging contaminants. *Chemosphere* **2022**, *287*, No. 132306.
(76) Lioy, P. J.; Rappaport, S. M. Exposure science and the exposome: an opportunity for coherence in the environmental health sciences. *Environ. Health Perspect* **2011**, *119* (11), A466−467.
(77) Musatadi, M.; Andrés-Maguregi, A.; De Angelis, F.; Prieto, A.; Anakabe, E.; Olivares, M.; Etxebarria, N.; Zuloaga, O. The role of sample preparation in suspect and non-target screening for exposome analysis using human urine. *Chemosphere* **2023**, *339*, No. 139690.
(78) Damaraju, S.; Driga, A.; Cook, L.; Calder, K.; Graham, K.; Dabbs, K.; Steed, H.; Berendt, R.; Mackey, J. R.; Cass, C. E. 3.02 - Considerations on Dealing with Tissues and Cell Samples (Include Tissue Banking). In *Comprehensive Sampling and Sample Preparation*; Pawliszyn, J., Ed.; Academic Press, 2012; pp 21−31.
(79) Guthrie, J. W. 3.01 - General Considerations when Dealing with Biological Fluid Samples. *In Comprehensive Sampling and Sample Preparation*; Pawliszyn, J., Ed.; Academic Press, 2012; pp 1−19.
(80) Zhou, W.; Yang, S.; Wang, P. G. Matrix effects and application of matrix effect factor. *Bioanalysis* **2017**, *9* (23), 1839−1844.
(81) Barupal, D. K.; Fiehn, O. Generating the Blood Exposome Database Using a Comprehensive Text Mining and Database Fusion Approach. *Environ. Health Persp.* **2019**, *127* (9), No. 097008.
(82) Goerdten, J.; Yuan, L.; Huybrechts, I.; Neveu, V.; Nöthlings, U.; Ahrens, W.; Scalbert, A.; Floegel, A. Reproducibility of the Blood and Urine Exposome: A Systematic Literature Review and Meta-Analysis. *Cancer Epidemiology, Biomarkers & Prevention* **2022**, *31* (9), 1683−1692.
(83) Tkalec, Ž.; Antignac, J.-P.; Bandow, N.; Béen, F. M.; Belova, L.; Bessems, J.; Le Bizec, B.; Brack, W.; Cano-Sancho, G.; Chaker, J.; et al. Innovative analytical methodologies for characterizing chemical exposure with a view to next-generation risk assessment. *Environ. Int.* **2024**, *186*, No. 108585.
(84) Tian, Z.; Zhao, H.; Peter, K. T.; Gonzalez, M.; Wetzel, J.; Wu, C.; Hu, X.; Prat, J.; Mudrock, E.; Hettinger, R.; et al. A ubiquitous tire rubber-derived chemical induces acute mortality in coho salmon. *Science* **2021**, *371* (6525), 185−189.
(85) Vinggaard, A. M.; Bonefeld-Jørgensen, E. C.; Jensen, T. K.; Fernandez, M. F.; Rosenmai, A. K.; Taxvig, C.; Rodriguez-Carrillo, A.; Wielsøe, M.; Long, M.; Olea, N.; et al. Receptor-based in vitro activities to assess human exposure to chemical mixtures and related health impacts. *Environ. Int.* **2021**, *146*, No. 106191.
(86) Chung, M. K.; Rappaport, S. M.; Wheelock, C. E.; Nguyen, V. K.; Meer, T. P. v. d.; Miller, G. W.; Vermeulen, R.; Patel, C. J. Utilizing a Biology-Driven Approach to Map the Exposome in Health and Disease: An Essential Investment to Drive the Next Generation of Environmental Discovery. *Environ. Health Persp.* **2021**, *129* (8), No. 085001.
(87) Tian, Z.; McMinn, M. H.; Fang, M. Effect-directed analysis and beyond: how to find causal environmental toxicants. *Exposome* **2023**, DOI: 10.1093/exposome/osad002.
(88) Simon, E.; Lamoree, M. H.; Hamers, T.; de Boer, J. Challenges in effect-directed analysis with a focus on biological samples. *TrAC Trends in Analytical Chemistry* **2015**, *67*, 179−191.
(89) Semple, K. T.; Doick, K. J.; Jones, K. C.; Burauel, P.; Craven, A.; Harms, H. Defining bioavailability and bioaccessibility of

contaminated soil and sediment is complicated. *Environ. Sci. Technol.* **2004**, *38* (12), 228A−231A.
(90) Yu, M.; Roszkowska, A.; Pawliszyn, J. In Vivo Solid-Phase Microextraction and Applications in Environmental Sciences. *ACS Environmental Au* **2022**, *2* (1), 30−41.
(91) Lin, E. Z.; Nichols, A.; Zhou, Y.; Koelmel, J. P.; Godri Pollitt, K. J. Characterizing the external exposome using passive samplers—comparative assessment of chemical exposures using different wearable form factors. *Journal of Exposure Science & Environmental Epidemiology* **2023**, *33* (4), 558−565.
(92) Rusina, T. P.; Smedes, F.; Klanova, J.; Booij, K.; Holoubek, I. Polymer selection for passive sampling: A comparison of critical properties. *Chemosphere* **2007**, *68* (7), 1344−1351.
(93) Adams, R. G.; Lohmann, R.; Fernandez, L. A.; MacFarlane, J. K.; Gschwend, P. M. Polyethylene Devices: Passive Samplers for Measuring Dissolved Hydrophobic Organic Compounds in Aquatic Environments. *Environ. Sci. Technol.* **2007**, *41* (4), 1317−1323.
(94) Bohlin, P.; Jones, K. C.; Strandberg, B. Occupational and indoor air exposure to persistent organic pollutants: A review of passive sampling techniques and needs. *Journal of Environmental Monitoring* **2007**, *9* (6), 501−509.
(95) Souza-Silva, É. A.; Jiang, R.; Rodríguez-Lafuente, A.; Gionfriddo, E.; Pawliszyn, J. A critical review of the state of the art of solid-phase microextraction of complex matrices I. Environmental analysis. *TrAC Trends in Analytical Chemistry* **2015**, *71*, 224−235.
(96) Bessonneau, V.; Ings, J.; McMaster, M.; Smith, R.; Bragg, L.; Servos, M.; Pawliszyn, J. In vivo microsampling to capture the elusive exposome. *Sci. Rep.* **2017**, *7* (1), No. 44038.
(97) Kleeblatt, J.; Schubert, J. K.; Zimmermann, R. Detection of Gaseous Compounds by Needle Trap Sampling and Direct Thermal-Desorption Photoionization Mass Spectrometry: Concept and Demonstrative Application to Breath Gas Analysis. *Anal. Chem.* **2015**, *87* (3), 1773−1781.
(98) Bessonneau, V.; Boyaci, E.; Maciazek-Jurczyk, M.; Pawliszyn, J. In vivo solid phase microextraction sampling of human saliva for non-invasive and on-site monitoring. *Anal. Chim. Acta* **2015**, *856*, 35−45.
(99) Papazian, S.; Fornaroli, C.; Bonnefille, B.; Pesquet, E.; Xie, H.; Martin, J. W. Silicone Foam for Passive Sampling and Nontarget Analysis of Air. *Environ. Sci. Tech Let* **2023**, *10*, 989.
(100) Koelmel, J. P.; Lin, E. Z.; Guo, P.; Zhou, J.; He, J.; Chen, A.; Gao, Y.; Deng, F.; Dong, H.; Liu, Y.; et al. Exploring the external exposome using wearable passive samplers - The China BAPE study. *Environ. Pollut.* **2021**, *270*, No. 116228.
(101) Okeme, J. O.; Koelmel, J. P.; Johnson, E.; Lin, E. Z.; Gao, D.; Pollitt, K. J. G. Wearable Passive Samplers for Assessing Environmental Exposure to Organic Chemicals: Current Approaches and Future Directions. *Current Environmental Health Reports* **2023**, *10* (2), 84−98.
(102) Samon, S. M.; Hammel, S. C.; Stapleton, H. M.; Anderson, K. A. Silicone wristbands as personal passive sampling devices: Current knowledge, recommendations for use, and future directions. *Environ. Int.* **2022**, *169*, No. 107339.
(103) O'Connell, S. G.; Kincl, L. D.; Anderson, K. A. Silicone Wristbands as Personal Passive Samplers. *Environ. Sci. Technol.* **2014**, *48* (6), 3327−3335.
(104) Jobst, K. J.; Arora, A.; Pollitt, K. G.; Sled, J. G. Dried blood spots for the identification of bioaccumulating organic compounds: Current challenges and future perspectives. *Current Opinion in Environmental Science & Health* **2020**, *15*, 66−73.
(105) Kocur, A.; Pawinski, T. Volumetric Absorptive Microsampling in Therapeutic Drug Monitoring of Immunosuppressive Drugs-From Sampling and Analytical Issues to Clinical Application. *Int. J. Mol. Sci.* **2023**, *24* (1), 681.
(106) Nemkov, T.; Yoshida, T.; Nikulina, M.; D'Alessandro, A. High-Throughput Metabolomics Platform for the Rapid Data-Driven Development of Novel Additive Solutions for Blood Storage. *Front Physiol* **2022**, *13*, No. 833242.
(107) Jacobson, T. A.; Bae, Y.; Kler, J. S.; Iyer, R.; Zhang, R.; Montgomery, N. D.; Nunes, D.; Pleil, J. D.; Funk, W. E. Advancing Global Health Surveillance of Mycotoxin Exposures using Minimally Invasive Sampling Techniques: A State-of-the-Science Review. *Environ. Sci. Technol.* **2024**, *58* (8), 3580−3594.
(108) Jain, A.; Morris, M.; Lin, E. Z.; Khan, S. A.; Ma, X.; Deziel, N. C.; Godri Pollitt, K. J.; Johnson, C. H. Hemoglobin normalization outperforms other methods for standardizing dried blood spot metabolomics: A comparative study. *Science of The Total Environment* **2023**, *854*, No. 158716.
(109) Berndt, A. E. Sampling Methods. *Journal of Human Lactation* **2020**, *36* (2), 224−226.
(110) Fisher, C. M.; Peter, K. T.; Newton, S. R.; Schaub, A. J.; Sobus, J. R. Approaches for assessing performance of high-resolution mass spectrometry−based non-targeted analysis methods. *Anal. Bioanal. Chem.* **2022**, *414* (22), 6455−6471.
(111) Manz, K. E.; Yamada, K.; Scheidl, L.; La Merrill, M. A.; Lind, L.; Pennell, K. D. Targeted and Nontargeted Detection and Characterization of Trace Organic Chemicals in Human Serum and Plasma Using QuEChERS Extraction. *Toxicol. Sci.* **2021**, *185* (1), 77−88.
(112) Tulipani, S.; Llorach, R.; Urpi-Sarda, M.; Andres-Lacueva, C. Comparative Analysis of Sample Preparation Methods To Handle the Complexity of the Blood Fluid Metabolome: When Less Is More. *Anal. Chem.* **2013**, *85* (1), 341−348.
(113) Al-Salhi, R.; Monfort, C.; Bonvallot, N.; David, A. Analytical strategies to profile the internal chemical exposome and the metabolome of human placenta. *Anal. Chim. Acta* **2022**, *1219*, No. 339983.
(114) Mertens, H.; Noll, B.; Schwerdtle, T.; Abraham, K.; Monien, B. H. Less is more: a methodological assessment of extraction techniques for per- and polyfluoroalkyl substances (PFAS) analysis in mammalian tissues. *Anal. Bioanal. Chem.* **2023**, *415* (24), 5925−5938.
(115) De Baere, S.; Ochieng, P. E.; Kemboi, D. C.; Scippo, M.-L.; Okoth, S.; Lindahl, J. F.; Gathumbi, J. K.; Antonissen, G.; Croubels, S. Development of High-Throughput Sample Preparation Procedures for the Quantitative Determination of Aflatoxins in Biological Matrices of Chickens and Cattle Using UHPLC-MS/MS. *Toxins* **2023**, *15* (1), 37.
(116) Zeleke, G.; De Baere, S.; Suleman, S.; Devreese, M. Development and Validation of a Reliable UHPLC-MS/MS Method for Simultaneous Quantification of Macrocyclic Lactones in Bovine Plasma. *Molecules* **2022**, *27* (3), 998.
(117) Tanoue, R.; Nozaki, K.; Nomiyama, K.; Kunisue, T.; Tanabe, S. Rapid analysis of 65 pharmaceuticals and 7 personal care products in plasma and whole-body tissue samples of fish using acidic extraction, zirconia-coated silica cleanup, and liquid chromatography-tandem mass spectrometry. *Journal of Chromatography A* **2020**, *1631*, No. 461586.
(118) Pourchet, M.; Narduzzi, L.; Jean, A.; Guiffard, I.; Bichon, E.; Cariou, R.; Guitton, Y.; Hutinet, S.; Vlaanderen, J.; Meijer, J.; et al. Non-targeted screening methodology to characterise human internal chemical exposure: Application to halogenated compounds in human milk. *Talanta* **2021**, *225*, No. 121979.
(119) Taylor, P. J. Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography−electrospray−tandem mass spectrometry. *Clinical Biochemistry* **2005**, *38* (4), 328−334.
(120) Matuszewski, B. K.; Constanzer, M. L.; Chavez-Eng, C. M. Strategies for the Assessment of Matrix Effect in Quantitative Bioanalytical Methods Based on HPLC−MS/MS. *Anal. Chem.* **2003**, *75* (13), 3019−3030.
(121) Stahnke, H.; Kittlaus, S.; Kempe, G.; Alder, L. Reduction of Matrix Effects in Liquid Chromatography−Electrospray Ionization−Mass Spectrometry by Dilution of the Sample Extracts: How Much Dilution is Needed? *Anal. Chem.* **2012**, *84* (3), 1474−1482.
(122) Greer, B.; Chevallier, O.; Quinn, B.; Botana, L. M.; Elliott, C. T. Redefining dilute and shoot: The evolution of the technique and its application in the analysis of foods and biological matrices by liquid chromatography mass spectrometry. *TrAC Trends in Analytical Chemistry* **2021**, *141*, No. 116284.

(123) Alcántara-Durán, J.; Moreno-González, D.; Beneito-Cambra, M.; García-Reyes, J. F. Dilute-and-shoot coupled to nanoflow liquid chromatography high resolution mass spectrometry for the determination of drugs of abuse and sport drugs in human urine. *Talanta* **2018**, *182*, 218−224.
(124) Atapattu, S. N.; Rosenfeld, J. M. Analytical derivatizations in environmental analysis. *Journal of Chromatography A* **2022**, *1678*, No. 463348.
(125) Lai, Z.; Fiehn, O. Mass spectral fragmentation of trimethylsilylated small molecules. *Mass Spectrom. Rev.* **2018**, *37* (3), 245−257.
(126) Metz, T. O.; Baker, E. S.; Schymanski, E. L.; Renslow, R. S.; Thomas, D. G.; Causon, T. J.; Webb, I. K.; Hann, S.; Smith, R. D.; Teeguarden, J. G. Integrating ion mobility spectrometry into mass spectrometry-based exposome measurements: what can it add and how far can it go? *Bioanalysis* **2017**, *9* (1), 81−98.
(127) Skoog, D. A.; West, D. M.; Holler, F. J.; Crouch, S. R. *Fundamentals of Analytical Chemistry*; Cengage Learning, 2013.
(128) May, J. C.; McLean, J. A. Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation. *Anal. Chem.* **2015**, *87* (3), 1422−1436.
(129) Shanmuganathan, M.; Kroezen, Z.; Gill, B.; Azab, S.; de Souza, R. J.; Teo, K. K.; Atkinson, S.; Subbarao, P.; Desai, D.; Anand, S. S.; et al. The maternal serum metabolome by multisegment injection-capillary electrophoresis-mass spectrometry: a high-throughput platform and standardized data workflow for large-scale epidemiological studies. *Nat. Protoc.* **2021**, *16* (4), 1966−1994.
(130) Fiehn, O. Metabolomics by Gas Chromatography−Mass Spectrometry: Combined Targeted and Untargeted Profiling. *Curr. Protoc. Mol. Biol.* **2016**, *114* (1), 30.34.31−30.34.32.
(131) Fiehn, O.; Kopka, J.; Dörmann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **2000**, *18* (11), 1157−1161.
(132) Gago-Ferrero, P.; Díaz-Cruz, M. S.; Barceló, D. An overview of UV-absorbing compounds (organic UV filters) in aquatic biota. *Anal. Bioanal. Chem.* **2012**, *404* (9), 2597−2610.
(133) Mol, H. G. J.; Plaza-Bolaños, P.; Zomer, P.; de Rijk, T. C.; Stolker, A. A. M.; Mulder, P. P. J. Toward a Generic Extraction Method for Simultaneous Determination of Pesticides, Mycotoxins, Plant Toxins, and Veterinary Drugs in Feed and Food Matrixes. *Anal. Chem.* **2008**, *80* (24), 9450−9459.
(134) Kanakaki, C.; Traka, T.; Thomaidis, N. S. Development and validation of multi-analyte methods for the determination of migrating substances from plastic food contact materials by GC-EI-QqQ-MS and GC-APCI-QTOF-MS. *Front. Sustain. Food Syst.* **2023**, DOI: 10.3389/fsufs.2023.1159002.
(135) Wong, J. K. Y.; Choi, T. L. S.; Kwok, K. Y.; Lei, E. N. Y.; Wan, T. S. M. Doping control analysis of 121 prohibited substances in equine hair by liquid chromatography−tandem mass spectrometry. *J. Pharm. Biomed. Anal.* **2018**, *158*, 189−203.
(136) Thevis, M.; Thomas, A.; Pop, V.; Schänzer, W. Ultrahigh pressure liquid chromatography-(tandem) mass spectrometry in human sports drug testing: possibilities and limitations. *J. Chromatogr A* **2013**, *1292*, 38−50.
(137) Rampler, E.; Abiead, Y. E.; Schoeny, H.; Rusz, M.; Hildebrand, F.; Fitz, V.; Koellensperger, G. Recurrent Topics in Mass Spectrometry-Based Metabolomics and Lipidomics—Standardization, Coverage, and Throughput. *Anal. Chem.* **2021**, *93* (1), 519−545.
(138) Flasch, M.; Fitz, V.; Rampler, E.; Ezekiel, C. N.; Koellensperger, G.; Warth, B. Integrated Exposomics/Metabolomics for Rapid Exposure and Effect Analyses. *JACS Au* **2022**, *2* (11), 2548−2560.
(139) González-Gaya, B.; Lopez-Herguedas, N.; Bilbao, D.; Mijangos, L.; Iker, A. M.; Etxebarria, N.; Irazola, M.; Prieto, A.; Olivares, M.; Zuloaga, O. Suspect and non-target screening: the last frontier in environmental analysis. *Analytical Methods* **2021**, *13* (16), 1876−1904.
(140) Dürig, W.; Lindblad, S.; Golovko, O.; Gkotsis, G.; Aalizadeh, R.; Nika, M.-C.; Thomaidis, N.; Alygizakis, N. A.; Plassmann, M.; Haglund, P.; et al. What is in the fish? Collaborative trial in suspect and non-target screening of organic micropollutants using LC- and GC-HRMS. *Environ. Int.* **2023**, *181*, No. 108288.
(141) Manz, K. E.; Dodson, R. E.; Liu, Y.; Scheidl, L.; Burks, S.; Dunn, F.; Gairola, R.; Lee, N. F.; Walker, E. D.; Pennell, K. D.; et al. Effects of Corsi-Rosenthal boxes on indoor air contaminants: non-targeted analysis using high resolution mass spectrometry. *Journal of Exposure Science & Environmental Epidemiology* **2023**, *33* (4), 537−547.
(142) Misra, B. B.; Olivier, M. High Resolution GC-Orbitrap-MS Metabolomics Using Both Electron Ionization and Chemical Ionization for Analysis of Human Plasma. *J. Proteome Res.* **2020**, *19* (7), 2717−2731.
(143) Brack, W.; Hollender, J.; de Alda, M. L.; Müller, C.; Schulze, T.; Schymanski, E.; Slobodnik, J.; Krauss, M. High-resolution mass spectrometry to complement monitoring and track emerging chemicals and pollution trends in European water resources. *Environmental Sciences Europe* **2019**, *31* (1), 62.
(144) Morris, C. B.; Poland, J. C.; May, J. C.; McLean, J. A. Fundamentals of Ion Mobility-Mass Spectrometry for the Analysis of Biomolecules. In *Ion Mobility-Mass Spectrometry: Methods and Protocols*; Paglia, G., Astarita, G., Eds.; Springer US, 2020; pp 1−31.
(145) Zhou, Z.; Luo, M.; Chen, X.; Yin, Y.; Xiong, X.; Wang, R.; Zhu, Z.-J. Ion mobility collision cross-section atlas for known and unknown metabolite annotation in untargeted metabolomics. *Nat. Commun.* **2020**, *11* (1), 4334.
(146) Baker, E. S.; Hoang, C.; Uritboonthai, W.; Heyman, H. M.; Pratt, B.; MacCoss, M.; MacLean, B.; Plumb, R.; Aisporna, A.; Siuzdak, G. METLIN-CCS: an ion mobility spectrometry collision cross section database. *Nat. Methods* **2023**, *20*, 1836.
(147) Feuerstein, M. L.; Hernández-Mesa, M.; Kiehne, A.; Le Bizec, B.; Hann, S.; Dervilly, G.; Causon, T. Comparability of Steroid Collision Cross Sections Using Three Different IM-HRMS Technologies: An Interplatform Study. *J. Am. Soc. Mass Spectrom.* **2022**, *33* (10), 1951−1959.
(148) Celma, A.; Sancho, J. V.; Schymanski, E. L.; Fabregat-Safont, D.; Ibáñez, M.; Goshawk, J.; Barknowitz, G.; Hernández, F.; Bijlsma, L. Improving Target and Suspect Screening High-Resolution Mass Spectrometry Workflows in Environmental Analysis by Ion Mobility Separation. *Environ. Sci. Technol.* **2020**, *54* (23), 15120−15131.
(149) Izquierdo-Sandoval, D.; Fabregat-Safont, D.; Lacalle-Bergeron, L.; Sancho, J. V.; Hernández, F.; Portoles, T. Benefits of Ion Mobility Separation in GC-APCI-HRMS Screening: From the Construction of a CCS Library to the Application to Real-World Samples. *Anal. Chem.* **2022**, *94* (25), 9040−9047.
(150) Mairinger, T.; Causon, T. J.; Hann, S. The potential of ion mobility−mass spectrometry for non-targeted metabolomics. *Curr. Opin. Chem. Biol.* **2018**, *42*, 9−15.
(151) Gonzalez de Vega, R.; Cameron, A.; Clases, D.; Dodgen, T. M.; Doble, P. A.; Bishop, D. P. Simultaneous targeted and non-targeted analysis of per- and polyfluoroalkyl substances in environmental samples by liquid chromatography-ion mobility-quadrupole time of flight-mass spectrometry and mass defect analysis. *J. Chromatogr A* **2021**, *1653*, No. 462423.
(152) Tu, J.; Zhou, Z.; Li, T.; Zhu, Z.-J. The emerging role of ion mobility-mass spectrometry in lipidomics to facilitate lipid separation and identification. *TrAC Trends in Analytical Chemistry* **2019**, *116*, 332−339.
(153) Giles, K.; Ujma, J.; Wildgoose, J.; Pringle, S.; Richardson, K.; Langridge, D.; Green, M. A Cyclic Ion Mobility-Mass Spectrometry System. *Anal. Chem.* **2019**, *91* (13), 8564−8573.
(154) Foster, M.; Rainey, M.; Watson, C.; Dodds, J. N.; Kirkwood, K. I.; Fernández, F. M.; Baker, E. S. Uncovering PFAS and Other Xenobiotics in the Dark Metabolome Using Ion Mobility Spectrometry, Mass Defect Analysis, and Machine Learning. *Environ. Sci. Technol.* **2022**, *56* (12), 9133−9143.

(155) Chouinard, C. D.; Cruzeiro, V. W. D.; Roitberg, A. E.; Yost, R. A. Experimental and Theoretical Investigation of Sodiated Multimers of Steroid Epimers with Ion Mobility-Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (2), 323−331.
(156) Kyle, J. E.; Zhang, X.; Weitz, K. K.; Monroe, M. E.; Ibrahim, Y. M.; Moore, R. J.; Cha, J.; Sun, X.; Lovelace, E. S.; Wagoner, J.; et al. Uncovering biologically significant lipid isomers with liquid chromatography, ion mobility spectrometry and mass spectrometry. *Analyst* **2016**, *141* (5), 1649−1659.
(157) Jeanne Dit Fouque, K.; Ramirez, C. E.; Lewis, R. L.; Koelmel, J. P.; Garrett, T. J.; Yost, R. A.; Fernandez-Lima, F. Effective Liquid Chromatography−Trapped Ion Mobility Spectrometry−Mass Spectrometry Separation of Isomeric Lipid Species. *Anal. Chem.* **2019**, *91* (8), 5021−5027.
(158) Wojcik, R.; Nagy, G.; Attah, I. K.; Webb, I. K.; Garimella, S. V. B.; Weitz, K. K.; Hollerbach, A.; Monroe, M. E.; Ligare, M. R.; Nielson, F. F.; et al. SLIM Ultrahigh Resolution Ion Mobility Spectrometry Separations of Isotopologues and Isotopomers Reveal Mobility Shifts due to Mass Distribution Changes. *Anal. Chem.* **2019**, *91* (18), 11952−11962.
(159) Watson, J. T.; Sparkman, O. D. *Introduction to Mass Spectrometry: Instrumentation, Applications, and Strategies for Data Interpretation*; Wiley, 2007.
(160) Finkelstein, J. Development of ionization methods. *Nat. Methods* **2015**, *12* (1), 6−7.
(161) Beale, D. J.; Pinu, F. R.; Kouremenos, K. A.; Poojary, M. M.; Narayana, V. K.; Boughton, B. A.; Kanojia, K.; Dayalan, S.; Jones, O. A. H.; Dias, D. A. Review of recent developments in GC−MS approaches to metabolomics-based research. *Metabolomics* **2018**, *14* (11), 152.
(162) Baumeister, T. U. H.; Ueberschaar, N.; Pohnert, G. Gas-Phase Chemistry in the GC Orbitrap Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (4), 573−580.
(163) Hong, S.; Kang, W.; Su, Y.; Guo, Y. Analysis of Trace-Level Volatile Compounds in Fresh Turf Crop (Lolium perenne L.) by Gas Chromatography Quadrupole Time-of-Flight Mass Spectrometry. *Chin. J. Chem.* **2013**, *31* (10), 1329−1335.
(164) Omer, E.; Bichon, E.; Hutinet, S.; Royer, A. L.; Monteau, F.; Germon, H.; Hill, P.; Remaud, G.; Dervilly-Pinel, G.; Cariou, R.; et al. Toward the characterisation of non-intentionally added substances migrating from polyester-polyurethane lacquers by comprehensive gas chromatography-mass spectrometry technologies. *J. Chromatogr A* **2019**, *1601*, 327−334.
(165) Vogel, P.; Lazarou, C.; Gazeli, O.; Brandt, S.; Franzke, J.; Moreno-González, D. Study of Controlled Atmosphere Flexible Microtube Plasma Soft Ionization Mass Spectrometry for Detection of Volatile Organic Compounds as Potential Biomarkers in Saliva for Cancer. *Anal. Chem.* **2020**, *92* (14), 9722−9729.
(166) Alon, T.; Amirav, A. How enhanced molecular ions in Cold EI improve compound identification by the NIST library. *Rapid Commun. Mass Spectrom.* **2015**, *29* (23), 2287−2292.
(167) Margolin Eren, K. J.; Elkabets, O.; Amirav, A. A comparison of electron ionization mass spectra obtained at 70 eV, low electron energies, and with cold EI and their NIST library identification probabilities. *Journal of Mass Spectrometry* **2020**, *55* (12), No. e4646.
(168) Tsizin, S.; Bokka, R.; Keshet, U.; Alon, T.; Fialkov, A. B.; Tal, N.; Amirav, A. Comparison of electrospray LC−MS, LC−MS with Cold EI and GC−MS with Cold EI for sample identification. *Int. J. Mass Spectrom.* **2017**, *422*, 119−125.
(169) Dougherty, R. C. Positive and Negative Chemical Ionization Mass Spectrometry. In *Mass Spectrometry in Environmental Sciences*; Karasek, F. W., Hutzinger, O., Safe, S., Eds.; Springer US, 1985; pp 77−91.
(170) Capellades, J.; Junza, A.; Samino, S.; Brunner, J. S.; Schabbauer, G.; Vinaixa, M.; Yanes, O. Exploring the Use of Gas Chromatography Coupled to Chemical Ionization Mass Spectrometry (GC-CI-MS) for Stable Isotope Labeling in Metabolomics. *Anal. Chem.* **2021**, *93* (3), 1242−1248.
(171) Ayala-Cabrera, J. F.; Montero, L.; Meckelmann, S. W.; Uteschil, F.; Schmitz, O. J. Review on atmospheric pressure ionization sources for gas chromatography-mass spectrometry. Part I: Current ion source developments and improvements in ionization strategies. *Anal. Chim. Acta* **2023**, *1238*, No. 340353.
(172) Niu, Y.; Liu, J.; Yang, R.; Zhang, J.; Shao, B. Atmospheric pressure chemical ionization source as an advantageous technique for gas chromatography-tandem mass spectrometry. *TrAC Trends in Analytical Chemistry* **2020**, *132*, No. 116053.
(173) de Hoffmann, E. *Mass Spectrometry: Principles and Applications*; John Wiley & Sons, 2007.
(174) Li, X.; Dorman, F. L.; Helm, P. A.; Kleywegt, S.; Simpson, A.; Simpson, M. J.; Jobst, K. J. Nontargeted Screening Using Gas Chromatography−Atmospheric Pressure Ionization Mass Spectrometry: Recent Trends and Emerging Potential. *Molecules* **2021**, *26* (22), 6911.
(175) Singh, R. R.; Aminot, Y.; Héas-Moisan, K.; Preud'homme, H.; Munschy, C. Cracked and shucked: GC-APCI-IMS-HRMS facilitates identification of unknown halogenated organic chemicals in French marine bivalves. *Environ. Int.* **2023**, *178*, No. 108094.
(176) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85* (1), 2−9.
(177) De O. Silva, R.; De Menezes, M. G.G.; De Castro, R. C.; De A. Nobre, C.; Milhome, M. A.L.; Do Nascimento, R. F. Efficiency of ESI and APCI ionization sources in LC-MS/MS systems for analysis of 22 pesticide residues in food matrix. *Food Chem.* **2019**, *297*, No. 124934.
(178) Martínez-Villalba, A.; Moyano, E.; Galceran, M. T. Ultra-high performance liquid chromatography−atmospheric pressure chemical ionization−tandem mass spectrometry for the analysis of benzimidazole compounds in milk samples. *Journal of Chromatography A* **2013**, *1313*, 119−131.
(179) Xue, J.; Domingo-Almenara, X.; Guijas, C.; Palermo, A.; Rinschen, M. M.; Isbell, J.; Benton, H. P.; Siuzdak, G. Enhanced in-Source Fragmentation Annotation Enables Novel Data Independent Acquisition and Autonomous METLIN Molecular Identification. *Anal. Chem.* **2020**, *92* (8), 6051−6059.
(180) Bernardo-Bermejo, S.; Xue, J.; Hoang, L.; Billings, E.; Webb, B.; Honders, M. W.; Venneker, S.; Heijs, B.; Castro-Puyana, M.; Marina, M. L.; et al. Quantitative multiple fragment monitoring with enhanced in-source fragmentation/annotation mass spectrometry. *Nat. Protoc.* **2023**, *18* (4), 1296−1315.
(181) Xue, J.; Zhu, J.; Hu, L.; Yang, J.; Lv, Y.; Zhao, F.; Liu, Y.; Zhang, T.; Cai, Y.; Fang, M. EISA-EXPOSOME: One Highly Sensitive and Autonomous Exposomic Platform with Enhanced in-Source Fragmentation/Annotation. *Anal. Chem.* **2023**, *95*, 17228.
(182) Singh, R. R.; Chao, A.; Phillips, K. A.; Xia, X. R.; Shea, D.; Sobus, J. R.; Schymanski, E. L.; Ulrich, E. M. Expanded coverage of non-targeted LC-HRMS using atmospheric pressure chemical ionization: a case study with ENTACT mixtures. *Anal Bioanal Chem.* **2020**, *412* (20), 4931−4939.
(183) Richard, A. M.; Huang, R.; Waidyanatha, S.; Shinn, P.; Collins, B. J.; Thillainadarajah, I.; Grulke, C. M.; Williams, A. J.; Lougee, R. R.; Judson, R. S.; et al. The Tox21 10K Compound Library: Collaborative Chemistry Advancing Toxicology. *Chem. Res. Toxicol.* **2021**, *34* (2), 189−216.
(184) Xian, F.; Hendrickson, C. L.; Marshall, A. G. High Resolution Mass Spectrometry. *Anal. Chem.* **2012**, *84* (2), 708−719.
(185) Zubarev, R. A.; Makarov, A. Orbitrap Mass Spectrometry. *Anal. Chem.* **2013**, *85* (11), 5288−5296.
(186) Kind, T.; Fiehn, O. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics* **2006**, *7* (1), 234.
(187) Balogh, M. P. Debating resolution and mass accuracy. *LC-GC N. Am.* **2004**, *22* (2), 118.
(188) Siuzdak, G. *The Expanding Role of Mass Spectrometry in Biotechnology*; MCC Press, 2006.

https://doi.org/10.1021/acs.est.4c01156
Environ. Sci. Technol. 2024, 58, 12784−12822

(189) Ni, Y.; Su, M.; Qiu, Y.; Jia, W.; Du, X. ADAP-GC 3.0: Improved Peak Detection and Deconvolution of Co-eluting Metabolites from GC/TOF-MS Data for Metabolomics Studies. *Anal. Chem.* **2016**, *88* (17), 8802−8811.
(190) Worsfold, P.; Townshend, A.; Poole, C. F.; Miró, M. *Encyclopedia of Analytical Science*; Elsevier Science, 2019.
(191) Llorca, M.; Farré, M.; Picó, Y.; Barceló, D. Study of the performance of three LC-MS/MS platforms for analysis of perfluorinated compounds. *Anal Bioanal Chem.* **2010**, *398* (3), 1145−1159.
(192) Michalski, A.; Damoc, E.; Hauschild, J. P.; Lange, O.; Wieghaus, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol. Cell Proteomics* **2011**, *10* (9), No. M111.011015.
(193) Scheltema, R. A.; Hauschild, J. P.; Lange, O.; Hornburg, D.; Denisov, E.; Damoc, E.; Kuehn, A.; Makarov, A.; Mann, M. The Q Exactive HF, a Benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. *Mol. Cell Proteomics* **2014**, *13* (12), 3698−3708.
(194) Eichhorn, P.; Pérez, S.; Barceló, D. Chapter 5 - Time-of-Flight Mass Spectrometry Versus Orbitrap-Based Mass Spectrometry for the Screening and Identification of Drugs and Metabolites: Is There a Winner? In *Comprehensive Analytical Chemistry*; Fernandez-Alba, A. R., Ed.; Elsevier, 2012; Vol. *58*, pp 217−272.
(195) Makarov, A.; Denisov, E.; Lange, O. Performance Evaluation of a High-field Orbitrap Mass Analyzer. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (8), 1391−1396.
(196) Eliuk, S.; Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu. Rev. Anal Chem. (Palo Alto Calif)* **2015**, *8*, 61−80.
(197) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. The Orbitrap: a new mass spectrometer. *J. Mass Spectrom* **2005**, *40* (4), 430−443.
(198) Kaufmann, A.; Walker, S. Comparison of linear intrascan and interscan dynamic ranges of Orbitrap and ion-mobility time-of-flight mass spectrometers. *Rapid Commun. Mass Spectrom.* **2017**, *31* (22), 1915−1926.
(199) Johnson, R. Forming fragments. *Nat. Methods* **2015**, *12* (1), 14−14.
(200) Cooks, R. G. Special feature: Historical. Collision-induced dissociation: Readings and commentary. *Journal of Mass Spectrometry* **1995**, *30* (9), 1215−1221.
(201) de Hoffmann, E. Tandem mass spectrometry: A primer. *Journal of Mass Spectrometry* **1996**, *31* (2), 129−137.
(202) Bourmaud, A.; Gallien, S.; Domon, B. Parallel reaction monitoring using quadrupole-Orbitrap mass spectrometer: Principle and applications. *Proteomics* **2016**, *16* (15−16), 2146−2159.
(203) Vaniya, A.; Fiehn, O. Using fragmentation trees and mass spectral trees for identifying unknown compounds in metabolomics. *TrAC Trends in Analytical Chemistry* **2015**, *69*, 52−61.
(204) Li, J.; Smith, L. S.; Zhu, H.-J. Data-independent acquisition (DIA): An emerging proteomics technology for analysis of drug-metabolizing enzymes and transporters. *Drug Discovery Today: Technologies* **2021**, *39*, 49−56.
(205) Krasny, L.; Huang, P. H. Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology. *Molecular Omics* **2021**, *17* (1), 29−42.
(206) Zhang, F.; Ge, W.; Ruan, G.; Cai, X.; Guo, T. Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *Proteomics* **2020**, *20* (17−18), No. e1900276.
(207) Meier, F.; Geyer, P. E.; Virreira Winter, S.; Cox, J.; Mann, M. BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nat. Methods* **2018**, *15* (6), 440−448.
(208) Wu, J.; Wang, H.; Zhao, X.; Qiu, H.; Li, N. High sensitivity and high-confidence compound identification with a flexible BoxCar acquisition method. *J. Pharm. Biomed. Anal.* **2022**, *219*, No. 114973.
(209) Liu, K.; Song, Y.; Liu, Y.; Peng, M.; Li, H.; Li, X.; Feng, B.; Xu, P.; Su, D. An integrated strategy using UPLC−QTOF-MSE and UPLC−QTOF-MRM (enhanced target) for pharmacokinetics study of wine processed Schisandra Chinensis fructus in rats. *J. Pharm. Biomed. Anal.* **2017**, *139*, 165−178.
(210) Castro, G.; Ramil, M.; Cela, R.; Rodríguez, I. Identification and determination of emerging pollutants in sewage sludge driven by UPLC-QTOF-MS data mining. *Science of The Total Environment* **2021**, *778*, No. 146256.
(211) Collins, B. C.; Hunter, C. L.; Liu, Y.; Schilling, B.; Rosenberger, G.; Bader, S. L.; Chan, D. W.; Gibson, B. W.; Gingras, A.-C.; Held, J. M.; et al. Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nat. Commun.* **2017**, *8* (1), 291.
(212) Koopmans, F.; Ho, J. T. C.; Smit, A. B.; Li, K. W. Comparative Analyses of Data Independent Acquisition Mass Spectrometric Approaches: DIA, WiSIM-DIA, and Untargeted DIA. *PROTEOMICS* **2018**, *18* (1), No. 1700304.
(213) Meier, F.; Brunner, A.-D.; Frank, M.; Ha, A.; Bludau, I.; Voytik, E.; Kaspar-Schoenefeld, S.; Lubeck, M.; Raether, O.; Bache, N.; et al. diaPASEF: parallel accumulation−serial fragmentation combined with data-independent acquisition. *Nat. Methods* **2020**, *17* (12), 1229−1236.
(214) Heil, L. R.; Damoc, E.; Arrey, T. N.; Pashkova, A.; Denisov, E.; Petzoldt, J.; Peterson, A. C.; Hsu, C.; Searle, B. C.; Shulman, N.; et al. Evaluating the Performance of the Astral Mass Analyzer for Quantitative Proteomics Using Data-Independent Acquisition. *J. Proteome Res.* **2023**, *22* (10), 3290−3300.
(215) Guo, J.; Huan, T. Comparison of Full-Scan, Data-Dependent, and Data-Independent Acquisition Modes in Liquid Chromatography−Mass Spectrometry Based Untargeted Metabolomics. *Anal. Chem.* **2020**, *92* (12), 8072−8080.
(216) Nikolskiy, I.; Mahieu, N. G.; Chen, Y., Jr.; Tautenhahn, R.; Patti, G. J. An Untargeted Metabolomic Workflow to Improve Structural Characterization of Metabolites. *Anal. Chem.* **2013**, *85* (16), 7713−7719.
(217) van Herwerden, D.; O'Brien, J. W.; Lege, S.; Pirok, B. W. J.; Thomas, K. V.; Samanipour, S. Cumulative Neutral Loss Model for Fragment Deconvolution in Electrospray Ionization High-Resolution Mass Spectrometry Data. *Anal. Chem.* **2023**, *95* (33), 12247−12255.
(218) Pejchinovski, M.; Klein, J.; Ramírez-Torres, A.; Bitsika, V.; Mermelekas, G.; Vlahou, A.; Mullen, W.; Mischak, H.; Jankowski, V. Comparison of higher energy collisional dissociation and collision-induced dissociation MS/MS sequencing methods for identification of naturally occurring peptides in human urine. *Proteomics Clin Appl.* **2015**, *9* (5−6), 531−542.
(219) Peters-Clarke, T.; Coon, J.; Riley, N. *Instrumentation at the Leading Edge of Proteomics*; American Chemical Society (ACS), 2023.
(220) Chen, X.; Wang, Z.; Wong, Y.-L. E.; Wu, R.; Zhang, F.; Chan, T.-W. D. Electron-ion reaction-based dissociation: A powerful ion activation method for the elucidation of natural product structures. *Mass Spectrom. Rev.* **2018**, *37* (6), 793−810.
(221) Brodbelt, J. S.; Morrison, L. J.; Santos, I. Ultraviolet Photodissociation Mass Spectrometry for Analysis of Biological Molecules. *Chem. Rev.* **2020**, *120* (7), 3328−3380.
(222) Baba, T.; Campbell, J. L.; Le Blanc, J. C. Y.; Baker, P. R. S. In-depth sphingomyelin characterization using electron impact excitation of ions from organics and mass spectrometry[S]. *J. Lipid Res.* **2016**, *57* (5), 858−867.
(223) Kong, F.; Keshet, U.; Shen, T.; Rodriguez, E.; Fiehn, O. LibGen: Generating High Quality Spectral Libraries of Natural Products for EAD-, UVPD-, and HCD-High Resolution Mass Spectrometers. *Anal. Chem.* **2023**, *95* (46), 16810−16818.
(224) Wang, J.; Chow, W.; Wong, J. W.; Chang, J. Applications of nDATA for screening, quantitation, and identification of pesticide residues in fruits and vegetables using UHPLC/ESI Q-Orbitrap all ion fragmentation and data independent acquisition. *J. Mass Spectrom* **2021**, *56* (9), No. e4783.

(225) Wang, J.; Chow, W.; Wong, J. W.; Leung, D.; Chang, J.; Li, M. Non-target data acquisition for target analysis (nDATA) of 845 pesticide residues in fruits and vegetables using UHPLC/ESI Q-Orbitrap. *Anal Bioanal Chem.* **2019**, *411* (7), 1421−1431.
(226) Wong; Wang, J. W.; Chang, J.; Chow, J. S.; Carlson, W.; Rajski, R.; Fernández-Alba, A. R.; Self, R.; Cooke, W. K.; Lock, C. M.; et al. Multilaboratory Collaborative Study of a Nontarget Data Acquisition for Target Analysis (nDATA) Workflow Using Liquid Chromatography-High-Resolution Accurate Mass Spectrometry for Pesticide Screening in Fruits and Vegetables. *J. Agric. Food Chem.* **2021**, *69* (44), 13200−13216.
(227) Wong, J. W.; Wang, J.; Chow, W.; Carlson, R.; Jia, Z.; Zhang, K.; Hayward, D. G.; Chang, J. S. Perspectives on Liquid Chromatography-High-Resolution Mass Spectrometry for Pesticide Screening in Foods. *J. Agric. Food Chem.* **2018**, *66* (37), 9573−9581.
(228) Manjarrés, D. P.; Montemurro, N.; Pérez, S. Development of a USE/d-SPE and targeted DIA-Orbitrap-MS acquisition methodology for the analysis of wastewater-derived organic pollutants in fish tissues and body fluids. *MethodsX* **2022**, *9*, No. 101705.
(229) Hollender, J.; Schymanski, E. L.; Ahrens, L.; Alygizakis, N.; Béen, F.; Bijlsma, L.; Brunner, A. M.; Celma, A.; Fildier, A.; Fu, Q.; et al. NORMAN guidance on suspect and non-target screening in environmental monitoring. *Environmental Sciences Europe* **2023**, *35* (1), 75.
(230) Uppal, K.; Walker, D. I.; Liu, K.; Li, S.; Go, Y.-M.; Jones, D. P. Computational Metabolomics: A Framework for the Million Metabolome. *Chem. Res. Toxicol.* **2016**, *29* (12), 1956−1975.
(231) Liu, X.; Locasale, J. W. Metabolomics: A Primer. *Trends Biochem. Sci.* **2017**, *42* (4), 274−284.
(232) Thompson, J. W.; Adams, K. J.; Adamski, J.; Asad, Y.; Borts, D.; Bowden, J. A.; Byram, G.; Dang, V.; Dunn, W. B.; Fernandez, F.; et al. International Ring Trial of a High Resolution Targeted Metabolomics and Lipidomics Platform for Serum and Plasma Analysis. *Anal. Chem.* **2019**, *91* (22), 14407−14416.
(233) Bowden, J. A.; Heckert, A.; Ulmer, C. Z.; Jones, C. M.; Koelmel, J. P.; Abdullah, L.; Ahonen, L.; Alnouti, Y.; Armando, A. M.; Asara, J. M.; et al. Harmonizing lipidomics: NIST interlaboratory comparison exercise for lipidomics using SRM 1950-Metabolites in Frozen Human Plasma. *J. Lipid Res.* **2017**, *58* (12), 2275−2288.
(234) Gago-Ferrero, P.; Bletsou, A. A.; Damalas, D. E.; Aalizadeh, R.; Alygizakis, N. A.; Singer, H. P.; Hollender, J.; Thomaidis, N. S. Wide-scope target screening of > 2000 emerging contaminants in wastewater samples with UPLC-Q-ToF-HRMS/MS and smart evaluation of its performance through the validation of 195 selected representative analytes. *J. Hazard Mater.* **2020**, *387*, No. 121712.
(235) Ulrich, E. M.; Sobus, J. R.; Grulke, C. M.; Richard, A. M.; Newton, S. R.; Strynar, M. J.; Mansouri, K.; Williams, A. J. EPA's non-targeted analysis collaborative trial (ENTACT): genesis, design, and initial findings. *Anal. Bioanal. Chem.* **2019**, *411* (4), 853−866.
(236) Place, B. J.; Ulrich, E. M.; Challis, J. K.; Chao, A.; Du, B.; Favela, K.; Feng, Y.-L.; Fisher, C. M.; Gardinali, P.; Hood, A.; et al. An Introduction to the Benchmarking and Publications for Non-Targeted Analysis Working Group. *Anal. Chem.* **2021**, *93* (49), 16289−16296.
(237) Peter, K. T.; Phillips, A. L.; Knolhoff, A. M.; Gardinali, P. R.; Manzano, C. A.; Miller, K. E.; Pristner, M.; Sabourin, L.; Sumarah, M. W.; Warth, B.; et al. Nontargeted Analysis Study Reporting Tool: A Framework to Improve Research Transparency and Reproducibility. *Anal. Chem.* **2021**, *93* (41), 13870−13879.
(238) Mohammed Taha, H.; Aalizadeh, R.; Alygizakis, N.; Antignac, J.-P.; Arp, H. P. H.; Bade, R.; Baker, N.; Belova, L.; Bijlsma, L.; Bolton, E. E.; et al. The NORMAN Suspect List Exchange (NORMAN-SLE): facilitating European and worldwide collaboration on suspect screening in high resolution mass spectrometry. *Environmental Sciences Europe* **2022**, *34* (1), 104.
(239) Shi, C.; Yang, J.; You, Z.; Zhang, Z.; Fang, M. Suspect screening analysis by tandem mass spectra from metabolomics to exposomics. *TrAC Trends in Analytical Chemistry* **2024**, *175*, No. 117699.
(240) Mistrik, R. *mzCLOUD: A spectral tree library for the Identification of "unknown unknowns"*, 2018; Vol. *255*.
(241) Wang, J.; P, D. A.; Mistrik, R.; Huang, Y.; Araujo, G. D. *A platform to identify endogenous metabolites using a novel high performance orbitrap MS and the mzCloud library*, 2013. https://apps.thermoscientific.com/media/cmd/ASMS-TNG-Roadshow/TNG/resouces/870_PN_ASMS13_MP%20045_JWang.pdf.
(242) Helmus, R.; ter Laak, T. L.; van Wezel, A. P.; de Voogt, P.; Schymanski, E. L. patRoon: open source software platform for environmental mass spectrometry based non-target screening. *Journal of Cheminformatics* **2021**, *13* (1), 1.
(243) Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *Journal of Cheminformatics* **2016**, *8* (1), 3.
(244) Talavera Andújar, B.; Aurich, D.; Aho, V. T. E.; Singh, R. R.; Cheng, T.; Zaslavsky, L.; Bolton, E. E.; Mollenhauer, B.; Wilmes, P.; Schymanski, E. L. Studying the Parkinson's disease metabolome and exposome in biological samples through different analytical and cheminformatics approaches: a pilot study. *Anal. Bioanal. Chem.* **2022**, *414* (25), 7399−7419.
(245) Cajka, T.; Fiehn, O. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Anal. Chem.* **2016**, *88* (1), 524−545.
(246) Comte, B.; Monnerie, S.; Brandolini-Bunlon, M.; Canlet, C.; Castelli, F.; Chu-Van, E.; Colsch, B.; Fenaille, F.; Joly, C.; Jourdan, F.; et al. Multiplatform metabolomics for an integrative exploration of metabolic syndrome in older men. *eBioMedicine* **2021**, *69*, No. 103440.
(247) Jia, S.; Xu, T.; Huan, T.; Chong, M.; Liu, M.; Fang, W.; Fang, M. Chemical Isotope Labeling Exposome (CIL-EXPOSOME): One High-Throughput Platform for Human Urinary Global Exposome Characterization. *Environ. Sci. Technol.* **2019**, *53* (9), 5445−5453.
(248) James, E. I.; Murphree, T. A.; Vorauer, C.; Engen, J. R.; Guttman, M. Advances in Hydrogen/Deuterium Exchange Mass Spectrometry and the Pursuit of Challenging Biological Systems. *Chem. Rev.* **2022**, *122* (8), 7562−7623.
(249) Amer, B.; Deshpande, R. R.; Bird, S. S. Simultaneous Quantitation and Discovery (SQUAD) Analysis: Combining the Best of Targeted and Untargeted Mass Spectrometry-Based Metabolomics. *Metabolites* **2023**, *13* (5), 648.
(250) McCord, J. P.; Groff, L. C.; Sobus, J. R. Quantitative non-targeted analysis: Bridging the gap between contaminant discovery and risk characterization. *Environ. Int.* **2022**, *158*, No. 107011.
(251) Rager, J. E.; Strynar, M. J.; Liang, S.; McMahen, R. L.; Richard, A. M.; Grulke, C. M.; Wambaugh, J. F.; Isaacs, K. K.; Judson, R.; Williams, A. J.; et al. Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring. *Environ. Int.* **2016**, *88*, 269−280.
(252) 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.
(253) Guo, J.; Yu, H.; Xing, S.; Huan, T. Addressing big data challenges in mass spectrometry-based metabolomics. *Chem. Commun.* **2022**, *58* (72), 9979−9990.
(254) Yu, H.; Huan, T. Patterned Signal Ratio Biases in Mass Spectrometry-Based Quantitative Metabolomics. *Anal. Chem.* **2021**, *93* (4), 2254−2262.
(255) Yu, H.; Xing, S.; Nierves, L.; Lange, P. F.; Huan, T. Fold-Change Compression: An Unexplored But Correctable Quantitative Bias Caused by Nonlinear Electrospray Ionization Responses in Untargeted Metabolomics. *Anal. Chem.* **2020**, *92* (10), 7011−7019.
(256) Liang, Y.-Z.; Kvalheim, O. M.; Manne, R. White, grey and black multicomponent systems: A classification of mixture problems and methods for their quantitative analysis. *Chemometrics and Intelligent Laboratory Systems* **1993**, *18* (3), 235−250.
(257) He, Y.; Brademan, D. R.; Hutchins, P. D.; Overmyer, K. A.; Coon, J. J. Maximizing MS/MS Acquisition for Lipidomics Using

Capillary Separation and Orbitrap Tribrid Mass Spectrometer. *Anal. Chem.* **2022**, *94* (7), 3394−3399.
(258) Karasawa, K.; Duchoslav, E.; Baba, T. Fast Electron Detachment Dissociation of Oligonucleotides in Electron-Nitrogen Plasma Stored in Magneto Radio-Frequency Ion Traps. *Anal. Chem.* **2022**, *94* (44), 15510−15517.
(259) Di Poto, C.; Tian, X.; Peng, X.; Heyman, H. M.; Szesny, M.; Hess, S.; Cazares, L. H. Metabolomic Profiling of Human Urine Samples Using LC-TIMS-QTOF Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2021**, *32* (8), 2072−2080.
(260) Wang, Z.; Mülleder, M.; Batruch, I.; Chelur, A.; Textoris-Taube, K.; Schwecke, T.; Hartl, J.; Causon, J.; Castro-Perez, J.; Demichev, V.; et al. High-throughput proteomics of nanogram-scale samples with Zeno SWATH MS. *eLife* **2022**, *11*, No. e83947.
(261) Hernandes, V. V.; Warth, B. Uniting targeted (Zeno MRM-HR) and untargeted (SWATH) LC-MS in a single run for sensitive high-resolution exposomics. *ChemRxiv* **2024**, DOI: 10.26434/chemrxiv-2024-s6sw6.
(262) Wolf, C.; Behrens, A.; Brungs, C.; Mende, E. D.; Lenz, M.; Piechutta, P. C.; Roblick, C.; Karst, U. Mobility-resolved broadband dissociation and parallel reaction monitoring for laser desorption/ionization-mass spectrometry - Tattoo pigment identification supported by trapped ion mobility spectrometry. *Anal. Chim. Acta* **2023**, *1242*, No. 340796.
(263) 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.
(264) Meijer, J.; Lamoree, M.; Hamers, T.; Antignac, J.-P.; Hutinet, S.; Debrauwer, L.; Covaci, A.; Huber, C.; Krauss, M.; Walker, D. I.; et al. An annotation database for chemicals of emerging concern in exposome research. *Environ. Int.* **2021**, *152*, No. 106511.
(265) Sauvé, S.; Desrosiers, M. A review of what is an emerging contaminant. *Chemistry Central Journal* **2014**, *8* (1), 15.
(266) Liu, Y.; D'Agostino, L. A.; Qu, G.; Jiang, G.; Martin, J. W. High-resolution mass spectrometry (HRMS) methods for nontarget discovery and characterization of poly- and per-fluoroalkyl substances (PFASs) in environmental and human samples. *TrAC Trends in Analytical Chemistry* **2019**, *121*, No. 115420.
(267) Hsu, J. Y.; Shih, C. L.; Liao, P. C. Exposure Marker Discovery of Phthalates Using Mass Spectrometry. *Mass Spectrom.* **2017**, *6*, S0062.
(268) Singer, H. P.; Wössner, A. E.; McArdell, C. S.; Fenner, K. Rapid Screening for Exposure to "Non-Target" Pharmaceuticals from Wastewater Effluents by Combining HRMS-Based Suspect Screening and Exposure Modeling. *Environ. Sci. Technol.* **2016**, *50* (13), 6698−6707.
(269) Lai, A.; Clark, A. M.; Escher, B. I.; Fernandez, M.; McEwen, L. R.; Tian, Z.; Wang, Z.; Schymanski, E. L. The Next Frontier of Environmental Unknowns: Substances of Unknown or Variable Composition, Complex Reaction Products, or Biological Materials (UVCBs). *Environ. Sci. Technol.* **2022**, *56* (12), 7448−7466.
(270) Vega-Herrera, A.; Llorca, M.; Savva, K.; León, V. M.; Abad, E.; Farré, M. Screening and Quantification of Micro(Nano)Plastics and Plastic Additives in the Seawater of Mar Menor Lagoon. *Front. Marine Sci.* **2021**, DOI: 10.3389/fmars.2021.697424.
(271) Muschket, M.; Di Paolo, C.; Tindall, A. J.; Touak, G.; Phan, A.; Krauss, M.; Kirchner, K.; Seiler, T.-B.; Hollert, H.; Brack, W. Identification of Unknown Antiandrogenic Compounds in Surface Waters by Effect-Directed Analysis (EDA) Using a Parallel Fractionation Approach. *Environ. Sci. Technol.* **2018**, *52* (1), 288−297.
(272) Zhao, H. N.; Hu, X.; Tian, Z.; Gonzalez, M.; Rideout, C. A.; Peter, K. T.; Dodd, M. C.; Kolodziej, E. P. Transformation Products of Tire Rubber Antioxidant 6PPD in Heterogeneous Gas-Phase Ozonation: Identification and Environmental Occurrence. *Environ. Sci. Technol.* **2023**, *57* (14), 5621−5632.
(273) Hu, X.; Zhao, H. N.; Tian, Z.; Peter, K. T.; Dodd, M. C.; Kolodziej, E. P. Chemical characteristics, leaching, and stability of the ubiquitous tire rubber-derived toxicant 6PPD-quinone. *Environ. Sci. Process Impacts* **2023**, *25* (5), 901−911.
(274) Arp, H. P. H.; Aurich, D.; Schymanski, E. L.; Sims, K.; Hale, S. E. Avoiding the Next Silent Spring: Our Chemical Past, Present, and Future. *Environ. Sci. Technol.* **2023**, *57* (16), 6355−6359.
(275) *Safe and Sustainable by Design Chemicals and Materials—A European Assessment Framework*; Publications Office of the European Union, 2022. DOI: 10.2777/86120.
(276) Rogers, L.; Jensen, K. F. Continuous manufacturing − the Green Chemistry promise? *Green Chem.* **2019**, *21* (13), 3481−3498.
(277) Zhu, B.; Wang, D.; Wei, N. Enzyme discovery and engineering for sustainable plastic recycling. *Trends Biotechnol.* **2022**, *40* (1), 22−37.
(278) Mulvihill, M. J.; Beach, E. S.; Zimmerman, J. B.; Anastas, P. T. Green Chemistry and Green Engineering: A Framework for Sustainable Technology Development. *Annual Review of Environment and Resources* **2011**, *36* (1), 271−293.
(279) Meadows, D. H.; Wright, D. *Thinking in Systems: A Primer*; Chelsea Green Publishing, 2008.
(280) McAlister, M. M.; Zhang, Q.; Annis, J.; Schweitzer, R. W.; Guidotti, S.; Mihelcic, J. R. Systems Thinking for Effective Interventions in Global Environmental Health. *Environ. Sci. Technol.* **2022**, *56* (2), 732−738.
(281) Wang, Z.; DeWitt, J. C.; Higgins, C. P.; Cousins, I. T. A Never-Ending Story of Per- and Polyfluoroalkyl Substances (PFASs)? *Environ. Sci. Technol.* **2017**, *51* (5), 2508−2518.
(282) Casey, J. S.; Jackson, S. R.; Ryan, J.; Newton, S. R. The use of gas chromatography − high resolution mass spectrometry for suspect screening and non-targeted analysis of per- and polyfluoroalkyl substances. *Journal of Chromatography A* **2023**, *1693*, No. 463884.
(283) Koelmel, J. P.; Lin, E. Z.; DeLay, K.; Williams, A. J.; Zhou, Y.; Bornman, R.; Obida, M.; Chevrier, J.; Godri Pollitt, K. J. Assessing the External Exposome Using Wearable Passive Samplers and High-Resolution Mass Spectrometry among South African Children Participating in the VHEMBE Study. *Environ. Sci. Technol.* **2022**, *56* (4), 2191−2203.
(284) Akoueson, F.; Chbib, C.; Brémard, A.; Monchy, S.; Paul-Pont, I.; Doyen, P.; Dehaut, A.; Duflos, G. Identification of plastic additives: Py/TD-GC-HRMS method development and application on food containers. *Journal of Analytical and Applied Pyrolysis* **2022**, *168*, No. 105745.
(285) Hankett, J. M.; Holtz, J. L.; Walker-Franklin, I.; Shaffer, K.; Jourdan, J.; Batiste, D. C.; Garcia, J. M.; Kaczan, C.; Wohlleben, W.; Ferguson, L. Matrix Matters: novel insights for the extraction, preparation, and quantitation of microplastics in a freshwater mesocosm study. *Microplastics and Nanoplastics* **2023**, *3* (1), 13.
(286) Fischer, M.; Scholz-Böttcher, B. M. Simultaneous Trace Identification and Quantification of Common Types of Microplastics in Environmental Samples by Pyrolysis-Gas Chromatography−Mass Spectrometry. *Environ. Sci. Technol.* **2017**, *51* (9), 5052−5060.
(287) Zheng, T.; Kelsey, K.; Zhu, C.; Pennell, K. D.; Yao, Q.; Manz, K. E.; Zheng, Y. F.; Braun, J. M.; Liu, Y.; Papandonatos, G.; et al. Adverse birth outcomes related to concentrations of per- and polyfluoroalkyl substances (PFAS) in maternal blood collected from pregnant women in 1960−1966. *Environ. Res.* **2024**, *241*, No. 117010.
(288) Panagopoulos Abrahamsson, D.; Wang, A.; Jiang, T.; Wang, M.; Siddharth, A.; Morello-Frosch, R.; Park, J.-S.; Sirota, M.; Woodruff, T. J. A Comprehensive Non-targeted Analysis Study of the Prenatal Exposome. *Environ. Sci. Technol.* **2021**, *55* (15), 10542−10557.
(289) Warembourg, C.; Anguita-Ruiz, A.; Siroux, V.; Slama, R.; Vrijheid, M.; Richiardi, L.; Basagaña, X. Statistical Approaches to Study Exposome-Health Associations in the Context of Repeated Exposure Data: A Simulation Study. *Environ. Sci. Technol.* **2023**, *57* (43), 16232−16243.
(290) Gibson, E. A.; Goldsmith, J.; Kioumourtzoglou, M.-A. Complex Mixtures, Complex Analyses: an Emphasis on Interpretable Results. *Current Environmental Health Reports* **2019**, *6* (2), 53−61.

(291) Avery, C. L.; Howard, A. G.; Ballou, A. F.; Buchanan, V. L.; Collins, J. M.; Downie, C. G.; Engel, S. M.; Graff, M.; Highland, H. M.; Lee, M. P.; et al. Strengthening Causal Inference in Exposomics Research: Application of Genetic Data and Methods. *Environ. Health Persp.* **2022**, *130* (5), No. 055001.
(292) Du, X.; Dastmalchi, F.; Ye, H.; Garrett, T. J.; Diller, M. A.; Liu, M.; Hogan, W. R.; Brochhausen, M.; Lemas, D. J. Evaluating LC-HRMS metabolomics data processing software using FAIR principles for research software. *Metabolomics* **2023**, *19* (2), 11.
(293) Siuzdak, G. *Activity Metabolomics and Mass Spectrometry*; MCC Press: San Diego, USA, 2024. DOI: 10.63025/LCUW3037.
(294) Mahieu, N. G.; Patti, G. J. Systems-Level Annotation of a Metabolomics Data Set Reduces 25 000 Features to Fewer than 1000 Unique Metabolites. *Anal. Chem.* **2017**, *89* (19), 10397−10406.
(295) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **2006**, *78* (3), 779−787.
(296) Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* **2015**, *12* (6), 523−526.
(297) Schmid, R.; Heuckeroth, S.; Korf, A.; Smirnov, A.; Myers, O.; Dyrlund, T. S.; Bushuiev, R.; Murray, K. J.; Hoffmann, N.; Lu, M.; et al. Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nat. Biotechnol.* **2023**, *41* (4), 447−449.
(298) Yu, T.; Park, Y.; Johnson, J. M.; Jones, D. P. apLCMS—adaptive processing of high-resolution LC/MS data. *Bioinformatics* **2009**, *25* (15), 1930−1936.
(299) Misra, B. B. New software tools, databases, and resources in metabolomics: updates from 2020. *Metabolomics* **2021**, *17* (5), 49.
(300) Renner, G.; Reuschenbach, M. Critical review on data processing algorithms in non-target screening: challenges and opportunities to improve result comparability. *Anal. Bioanal. Chem.* **2023**, *415* (18), 4111−4123.
(301) DeFelice, B. C.; Mehta, S. S.; Samra, S.; Čajka, T.; Wancewicz, B.; Fahrmann, J. F.; Fiehn, O. Mass Spectral Feature List Optimizer (MS-FLO): A Tool To Minimize False Positive Peak Reports in Untargeted Liquid Chromatography−Mass Spectroscopy (LC-MS) Data Processing. *Anal. Chem.* **2017**, *89* (6), 3250−3255.
(302) Uppal, K.; Soltow, Q. A.; Strobel, F. H.; Pittard, W. S.; Gernert, K. M.; Yu, T.; Jones, D. P. xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. *BMC Bioinformatics* **2013**, *14* (1), 15.
(303) Delabriere, A.; Warmer, P.; Brennsteiner, V.; Zamboni, N. SLAW: A Scalable and Self-Optimizing Processing Workflow for Untargeted LC-MS. *Anal. Chem.* **2021**, *93* (45), 15024−15032.
(304) Li, S.; Siddiqa, A.; Thapa, M.; Chi, Y.; Zheng, S. Trackable and scalable LC-MS metabolomics data processing using asari. *Nat. Commun.* **2023**, *14* (1), 4113.
(305) Sing, J. C.; Charkow, J.; AlHigaylan, M.; Horecka, I.; Xu, L.; Röst, H. L. MassDash: A Web-Based Dashboard for Data-Independent Acquisition Mass Spectrometry Visualization. *J. Proteome Res.* **2024**, *23*, 2306.
(306) Lai, Z.; Tsugawa, H.; Wohlgemuth, G.; Mehta, S.; Mueller, M.; Zheng, Y.; Ogiwara, A.; Meissen, J.; Showalter, M.; Takeuchi, K.; et al. Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nat. Methods* **2018**, *15* (1), 53−56.
(307) Smirnov, A.; Qiu, Y.; Jia, W.; Walker, D. I.; Jones, D. P.; Du, X. ADAP-GC 4.0: Application of Clustering-Assisted Multivariate Curve Resolution to Spectral Deconvolution of Gas Chromatography−Mass Spectrometry Metabolomics Data. *Anal. Chem.* **2019**, *91* (14), 9069−9077.
(308) Aksenov, A. A.; Laponogov, I.; Zhang, Z.; Doran, S. L. F.; Belluomo, I.; Veselkov, D.; Bittremieux, W.; Nothias, L. F.; Nothias-Esposito, M.; Maloney, K. N.; et al. Auto-deconvolution and molecular networking of gas chromatography-mass spectrometry data. *Nat. Biotechnol.* **2021**, *39* (2), 169−173.
(309) Lennon, S.; Chaker, J.; Price, E. J.; Hollender, J.; Huber, C.; Schulze, T.; Ahrens, L.; Béen, F.; Creusot, N.; Debrauwer, L.; et al. Harmonized quality assurance/quality control provisions to assess completeness and robustness of MS1 data preprocessing for LC-HRMS-based suspect screening and non-targeted analysis. *TrAC Trends in Analytical Chemistry* **2024**, *174*, No. 117674.
(310) Yang, J.; Zhao, X.; Lu, X.; Lin, X.; Xu, G. A data preprocessing strategy for metabolomics to reduce the mask effect in data analysis. *Front Mol. Biosci* **2015**, *2*, 4.
(311) Stancliffe, E.; Schwaiger-Haber, M.; Sindelar, M.; Murphy, M. J.; Soerensen, M.; Patti, G. J. An Untargeted Metabolomics Workflow that Scales to Thousands of Samples for Population-Based Studies. *Anal. Chem.* **2022**, *94* (50), 17370−17378.
(312) Fan, S.; Kind, T.; Cajka, T.; Hazen, S. L.; Tang, W. H. W.; Kaddurah-Daouk, R.; Irvin, M. R.; Arnett, D. K.; Barupal, D. K.; Fiehn, O. Systematic Error Removal Using Random Forest for Normalizing Large-Scale Untargeted Lipidomics Data. *Anal. Chem.* **2019**, *91* (5), 3590−3596.
(313) Rong, Z.; Tan, Q.; Cao, L.; Zhang, L.; Deng, K.; Huang, Y.; Zhu, Z.-J.; Li, Z.; Li, K. NormAE: Deep Adversarial Learning Model to Remove Batch Effects in Liquid Chromatography Mass Spectrometry-Based Metabolomics Data. *Anal. Chem.* **2020**, *92* (7), 5082−5090.
(314) Yu, M.; Teitelbaum, S. L.; Dolios, G.; Dang, L. T.; Tu, P.; Wolff, M. S.; Petrick, L. M. Molecular Gatekeeper Discovery: Workflow for Linking Multiple Exposure Biomarkers to Metabolomics. *Environ. Sci. Technol.* **2022**, *56* (10), 6162−6171.
(315) Barupal, D. K.; Mahajan, P.; Fakouri-Baygi, S.; Wright, R. O.; Arora, M.; Teitelbaum, S. L. CCDB: A database for exploring inter-chemical correlations in metabolomics and exposomics datasets. *Environ. Int.* **2022**, *164*, No. 107240.
(316) Du, X.; Zeisel, S. H. SPECTRAL DECONVOLUTION FOR GAS CHROMATOGRAPHY MASS SPECTROMETRY-BASED METABOLOMICS: CURRENT STATUS AND FUTURE PERSPECTIVES. *Computational and Structural Biotechnology Journal* **2013**, *4* (5), No. e201301013.
(317) Smirnov, A.; Liao, Y.; Fahy, E.; Subramaniam, S.; Du, X. ADAP-KDB: A Spectral Knowledgebase for Tracking and Prioritizing Unknown GC−MS Spectra in the NIH's Metabolomics Data Repository. *Anal. Chem.* **2021**, *93* (36), 12213−12220.
(318) Ardente, A. J.; Garrett, T. J.; Wells, R. S.; Walsh, M.; Smith, C. R.; Colee, J.; Hill, R. C. A Targeted Metabolomics Assay to Measure Eight Purines in the Diet of Common Bottlenose Dolphins, Tursiops truncatus. *J. Chromatogr. Sep. Technol.* **2016**, DOI: 10.4172/2157-7064.1000334.
(319) Griffiths, W. J.; Koal, T.; Wang, Y.; Kohl, M.; Enot, D. P.; Deigner, H.-P. Targeted Metabolomics for Biomarker Discovery. *Angew. Chem., Int. Ed.* **2010**, *49* (32), 5426−5445.
(320) Yazd, H. S.; Rubio, V. Y.; Chamberlain, C. A.; Yost, R. A.; Garrett, T. J. Metabolomic and lipidomic characterization of an X-chromosome deletion disorder in neural progenitor cells by UHPLC-HRMS. *Journal of Mass Spectrometry and Advances in the Clinical Lab* **2021**, *20*, 11−24.
(321) Koelmel, J. P.; Xie, H.; Price, E. J.; Lin, E. Z.; Manz, K. E.; Stelben, P.; Paige, M. K.; Papazian, S.; Okeme, J.; Jones, D. P. An actionable annotation scoring framework for gas chromatography-high-resolution mass spectrometry. *Exposome* **2022**, DOI: 10.1093/exposome/osac007.
(322) Blaženović, I.; Kind, T.; Ji, J.; Fiehn, O. Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. *Metabolites* **2018**, *8* (2), 31.
(323) Zhang, H.; Luo, M.; Wang, H.; Ren, F.; Yin, Y.; Zhu, Z.-J. AllCCS2: Curation of Ion Mobility Collision Cross-Section Atlas for Small Molecules Using Comprehensive Molecular Representations. *Anal. Chem.* **2023**, *95* (37), 13913−13921.
(324) Simón-Manso, Y.; Lowenthal, M. S.; Kilpatrick, L. E.; Sampson, M. L.; Telu, K. H.; Rudnick, P. A.; Mallard, W. G.;

Bearden, D. W.; Schock, T. B.; Tchekhovskoi, D. V.; et al. Metabolite Profiling of a NIST Standard Reference Material for Human Plasma (SRM 1950): GC-MS, LC-MS, NMR, and Clinical Laboratory Analyses, Libraries, and Web-Based Resources. *Anal. Chem.* **2013**, *85* (24), 11725−11731.
(325) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; et al. MassBank: a public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry* **2010**, *45* (7), 703−714.
(326) Wohlgemuth, G.; Mehta, S. S.; Mejia, R. F.; Neumann, S.; Pedrosa, D.; Pluskal, T.; Schymanski, E. L.; Willighagen, E. L.; Wilson, M.; Wishart, D. S.; et al. SPLASH, a hashed identifier for mass spectra. *Nat. Biotechnol.* **2016**, *34* (11), 1099−1101.
(327) Guijas, C.; Montenegro-Burke, J. R.; Domingo-Almenara, X.; Palermo, A.; Warth, B.; Hermann, G.; Koellensperger, G.; Huan, T.; Uritboonthai, W.; Aisporna, A. E.; et al. METLIN: A Technology Platform for Identifying Knowns and Unknowns. *Anal. Chem.* **2018**, *90* (5), 3156−3164.
(328) Stein, S. Mass Spectral Reference Libraries: An Ever-Expanding Resource for Chemical Identification. *Anal. Chem.* **2012**, *84* (17), 7274−7282.
(329) Kind, T.; Tsugawa, H.; Cajka, T.; Ma, Y.; Lai, Z.; Mehta, S. S.; Wohlgemuth, G.; Barupal, D. K.; Showalter, M. R.; Arita, M.; et al. Identification of small molecules using accurate mass MS/MS search. *Mass Spectrom. Rev.* **2018**, *37* (4), 513−532.
(330) Papazian, S.; D'Agostino, L. A.; Sadiktsis, I.; Froment, J.; Bonnefille, B.; Sdougkou, K.; Xie, H.; Athanassiadis, I.; Budhavant, K.; Dasari, S.; et al. Nontarget mass spectrometry and in silico molecular characterization of air pollution from the Indian subcontinent. *Communications Earth & Environment* **2022**, *3* (1), 35.
(331) Colby, S. M.; Chang, C. H.; Bade, J. L.; Nunez, J. R.; Blumer, M. R.; Orton, D. J.; Bloodsworth, K. J.; Nakayasu, E. S.; Smith, R. D.; Ibrahim, Y. M.; et al. DEIMoS: An Open-Source Tool for Processing High-Dimensional Mass Spectrometry Data. *Anal. Chem.* **2022**, *94* (16), 6130−6138.
(332) Broeckling, C. D.; Afsar, F. A.; Neumann, S.; Ben-Hur, A.; Prenni, J. E. RAMClust: A Novel Feature Clustering Method Enables Spectral-Matching-Based Annotation for Metabolomics Data. *Anal. Chem.* **2014**, *86* (14), 6812−6817.
(333) Koelmel, J. P.; Kummer, M.; Chevallier, O.; Hindle, R.; Hunt, K.; Camacho, C. G.; Abril, N.; Gill, E. L.; Beecher, C. W. W.; Garrett, T. J.; et al. Expanding Per- and Polyfluoroalkyl Substances Coverage in Nontargeted Analysis Using Data-Independent Analysis and IonDecon. *J. Am. Soc. Mass Spectrom.* **2023**, *34* (11), 2525−2537.
(334) Koelmel, J. P.; Kroeger, N. M.; Gill, E. L.; Ulmer, C. Z.; Bowden, J. A.; Patterson, R. E.; Yost, R. A.; Garrett, T. J. Expanding Lipidome Coverage Using LC-MS/MS Data-Dependent Acquisition with Automated Exclusion List Generation. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (5), 908−917.
(335) Yu, M.; Dolios, G.; Petrick, L. Reproducible untargeted metabolomics workflow for exhaustive MS2 data acquisition of MS1 features. *Journal of Cheminformatics* **2022**, *14* (1), 6.
(336) Cooper, B.; Yang, R. An assessment of AcquireX and Compound Discoverer software 3.3 for non-targeted metabolomics. *Sci. Rep.* **2024**, *14* (1), 4841.
(337) Kováts, E. Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone. *Helv. Chim. Acta* **1958**, *41* (7), 1915−1932.
(338) Li, M.; Wang, X. R. Peak alignment of gas chromatography−mass spectrometry data with deep learning. *Journal of Chromatography A* **2019**, *1604*, No. 460476.
(339) Rigano, F.; Arigò, A.; Oteri, M.; La Tella, R.; Dugo, P.; Mondello, L. The retention index approach in liquid chromatography: An historical review and recent advances. *Journal of Chromatography A* **2021**, *1640*, No. 461963.
(340) Hughey, C. A.; Hendrickson, C. L.; Rodgers, R. P.; Marshall, A. G.; Qian, K. Kendrick Mass Defect Spectrum: A Compact Visual Analysis for Ultrahigh-Resolution Broadband Mass Spectra. *Anal. Chem.* **2001**, *73* (19), 4676−4681.
(341) Zhang, X.; Di Lorenzo, R. A.; Helm, P. A.; Reiner, E. J.; Howard, P. H.; Muir, D. C. G.; Sled, J. G.; Jobst, K. J. Compositional space: A guide for environmental chemists on the identification of persistent and bioaccumulative organics using mass spectrometry. *Environ. Int.* **2019**, *132*, No. 104808.
(342) Kind, T.; Fiehn, O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* **2007**, *8* (1), 105.
(343) Aron, A. T.; Gentry, E. C.; McPhail, K. L.; Nothias, L.-F.; Nothias-Esposito, M.; Bouslimani, A.; Petras, D.; Gauglitz, J. M.; Sikora, N.; Vargas, F.; et al. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nat. Protoc.* **2020**, *15* (6), 1954−1991.
(344) Koelmel, J. P.; Paige, M. K.; Aristizabal-Henao, J. J.; Robey, N. M.; Nason, S. L.; Stelben, P. J.; Li, Y.; Kroeger, N. M.; Napolitano, M. P.; Savvaides, T.; et al. Toward Comprehensive Per- and Polyfluoroalkyl Substances Annotation Using FluoroMatch Software and Intelligent High-Resolution Tandem Mass Spectrometry Acquisition. *Anal. Chem.* **2020**, *92* (16), 11186−11194.
(345) Koelmel, J. P.; Kroeger, N. M.; Ulmer, C. Z.; Bowden, J. A.; Patterson, R. E.; Cochran, J. A.; Beecher, C. W. W.; Garrett, T. J.; Yost, R. A. LipidMatch: an automated workflow for rule-based lipid identification using untargeted high-resolution tandem mass spectrometry data. *BMC Bioinformatics* **2017**, *18* (1), 331.
(346) Zhao, F.; Li, L.; Lin, P.; Chen, Y.; Xing, S.; Du, H.; Wang, Z.; Yang, J.; Huan, T.; Long, C.; et al. HExpPredict: In Vivo Exposure Prediction of Human Blood Exposome Using a Random Forest Model and Its Application in Chemical Risk Prioritization. *Environ. Health Persp.* **2023**, *131* (3), No. 037009.
(347) Pesyna, G. M.; Venkataraghavan, R.; Dayringer, H. E.; McLafferty, F. W. Probability based matching system using a large collection of reference mass spectra. *Anal. Chem.* **1976**, *48* (9), 1362−1368.
(348) McLafferty, F. W.; Hertel, R.; Villwock, R. Probability based matching of mass spectra. Rapid identification of specific compounds in mixtures. *Journal of Mass Spectrometry* **1974**, *9*, 690−702.
(349) Sokolow, S.; K, J.; Gustafson, P. *The Finnigan Library Search Program, Finnigan Application Report 2, San Jose, CA*; 1978. https://littlemsandsailing.com/wp-content/uploads/2015/03/finigan-incos-app-note-fixed.pdf.
(350) Hertz, H. S.; Hites, R. A.; Biemann, K. Identification of mass spectra by computer-searching a file of known spectra. *Anal. Chem.* **1971**, *43* (6), 681−691.
(351) Stein, S. E.; Scott, D. R. Optimization and testing of mass spectral library search algorithms for compound identification. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (9), 859−866.
(352) Sheldon, M. T.; Mistrik, R.; Croley, T. R. Determination of Ion Structures in Structurally Related Compounds Using Precursor Ion Fingerprinting. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (3), 370−376.
(353) de Jonge, N. F.; Louwen, J. J. R.; Chekmeneva, E.; Camuzeaux, S.; Vermeir, F. J.; Jansen, R. S.; Huber, F.; van der Hooft, J. J. J. MS2Query: reliable and scalable MS2 mass spectra-based analogue search. *Nat. Commun.* **2023**, *14* (1), 1752.
(354) Li, Y.; Kind, T.; Folz, J.; Vaniya, A.; Mehta, S. S.; Fiehn, O. Spectral entropy outperforms MS/MS dot product similarity for small-molecule compound identification. *Nat. Methods* **2021**, *18* (12), 1524−1531.
(355) Li, Y.; Fiehn, O. Flash entropy search to query all mass spectral libraries in real time. *Nat. Methods* **2023**, *20* (10), 1475−1478.
(356) Cai, Y.; Zhou, Z.; Zhu, Z.-J. Advanced analytical and informatic strategies for metabolite annotation in untargeted metabolomics. *TrAC Trends in Analytical Chemistry* **2023**, *158*, No. 116903.
(357) Kind, T.; Liu, K.-H.; Lee, D. Y.; DeFelice, B.; Meissen, J. K.; Fiehn, O. LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nat. Methods* **2013**, *10* (8), 755−758.

(358) Zhou, J.; Weber, R. J.; Allwood, J. W.; Mistrik, R.; Zhu, Z.; Ji, Z.; Chen, S.; Dunn, W. B.; He, S.; Viant, M. R. HAMMER: automated operation of mass frontier to construct in silico mass spectral fragmentation libraries. *Bioinformatics* **2014**, *30* (4), 581−583.
(359) Tsugawa, H.; Kind, T.; Nakabayashi, R.; Yukihira, D.; Tanaka, W.; Cajka, T.; Saito, K.; Fiehn, O.; Arita, M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. *Anal. Chem.* **2016**, *88* (16), 7946−7958.
(360) Wang, F.; Liigand, J.; Tian, S.; Arndt, D.; Greiner, R.; Wishart, D. S. CFM-ID 4.0: More Accurate ESI-MS/MS Spectral Prediction and Compound Identification. *Anal. Chem.* **2021**, *93* (34), 11692−11700.
(361) Duhrkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Bocker, S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (41), 12580−12585.
(362) Borges, R. M.; Colby, S. M.; Das, S.; Edison, A. S.; Fiehn, O.; Kind, T.; Lee, J.; Merrill, A. T.; Merz, K. M., Jr; Metz, T. O.; et al. Quantum Chemistry Calculations for Metabolomics. *Chem. Rev.* **2021**, *121* (10), 5633−5670.
(363) Wang, S.; Kind, T.; Bremer, P. L.; Tantillo, D. J.; Fiehn, O. Quantum Chemical Prediction of Electron Ionization Mass Spectra of Trimethylsilylated Metabolites. *Anal. Chem.* **2022**, *94* (3), 1559−1566.
(364) Liu, M.; Jiang, J.; Zheng, J.; Huan, T.; Gao, B.; Fei, X.; Wang, Y.; Fang, M. RTP: One Effective Platform to Probe Reactive Compound Transformation Products and Its Applications for a Reactive Plasticizer BADGE. *Environ. Sci. Technol.* **2021**, *55* (23), 16034−16043.
(365) Tian, S.; Cao, X.; Greiner, R.; Li, C.; Guo, A.; Wishart, D. S. CyProduct: A Software Tool for Accurately Predicting the Byproducts of Human Cytochrome P450 Metabolism. *J. Chem. Inf. Model.* **2021**, *61* (6), 3128−3140.
(366) Wishart, D. S; Tian, S.; Allen, D.; Oler, E.; Peters, H.; Lui, V. W; Gautam, V.; Djoumbou-Feunang, Y.; Greiner, R.; Metz, T. O BioTransformer 3.0—a web server for accurately predicting metabolic transformation products. *Nucleic Acids Res.* **2022**, *50* (W1), W115−W123.
(367) Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A. A.; Melnik, A. V.; Meusel, M.; Dorrestein, P. C.; Rousu, J.; Böcker, S. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* **2019**, *16* (4), 299−302.
(368) Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G. XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Anal. Chem.* **2012**, *84* (11), 5035−5039.
(369) Uppal, K.; Walker, D. I.; Jones, D. P. xMSannotator: An R Package for Network-Based Annotation of High-Resolution Metabolomics Data. *Anal. Chem.* **2017**, *89* (2), 1063−1067.
(370) Shen, X.; Yan, H.; Wang, C.; Gao, P.; Johnson, C. H.; Snyder, M. P. TidyMass an object-oriented reproducible analysis framework for LC−MS data. *Nat. Commun.* **2022**, *13* (1), 4365.
(371) Shen, X.; Wu, S.; Liang, L.; Chen, S.; Contrepois, K.; Zhu, Z.-J.; Snyder, M. metID: an R package for automatable compound annotation for LC−MS-based data. *Bioinformatics* **2022**, *38* (2), 568−569.
(372) Schymanski, E. L.; Neumann, S. The Critical Assessment of Small Molecule Identification (CASMI): Challenges and Solutions. *Metabolites* **2013**, *3* (3), 517−538.
(373) McEachran, A. D.; Chao, A.; Al-Ghoul, H.; Lowe, C.; Grulke, C.; Sobus, J. R.; Williams, A. J. Revisiting Five Years of CASMI Contests with EPA Identification Tools. *Metabolites* **2020**, *10* (6), 260.
(374) Bertrand, S.; Guitton, Y.; Roullier, C. Successes and pitfalls in automated dereplication strategy using liquid chromatography coupled to mass spectrometry data: A CASMI 2016 experience. *Phytochemistry Letters* **2017**, *21*, 297−305.
(375) Blaženović, I.; Kind, T.; Torbašinović, H.; Obrenović, S.; Mehta, S. S.; Tsugawa, H.; Wermuth, T.; Schauer, N.; Jahn, M.; Biedendieck, R.; et al. Comprehensive comparison of in silico MS/MS fragmentation tools of the CASMI contest: database boosting is needed to achieve 93% accuracy. *Journal of Cheminformatics* **2017**, *9* (1), 32.
(376) Xing, S.; Shen, S.; Xu, B.; Li, X.; Huan, T. BUDDY: molecular formula discovery via bottom-up MS/MS interrogation. *Nat. Methods* **2023**, *20* (6), 881−890.
(377) Hecht, H.; Rojas, W. Y.; Ahmad, Z.; Křenek, A.; Klánová, J.; EJ, P. Quantum chemistry based prediction of electron ionization mass spectra for environmental chemicals. *ChemRxiv* **2024**, DOI: 10.26434/chemrxiv-2024-2ngwq-v2.
(378) Wolf, S.; Schmidt, S.; Müller-Hannemann, M.; Neumann, S. In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinformatics* **2010**, *11* (1), 148.
(379) Ruttkies, C.; Neumann, S.; Posch, S. Improving MetFrag with statistical learning of fragment annotations. *BMC Bioinformatics* **2019**, *20* (1), 376.
(380) Phinney, K. W.; Ballihaut, G.; Bedner, M.; Benford, B. S.; Camara, J. E.; Christopher, S. J.; Davis, W. C.; Dodder, N. G.; Eppe, G.; Lang, B. E.; et al. Development of a Standard Reference Material for Metabolomics Research. *Anal. Chem.* **2013**, *85* (24), 11732−11738.
(381) Hampel, D.; Shahab-Ferdows, S.; Hossain, M.; Islam, M. M.; Ahmed, T.; Allen, L. H. Validation and Application of Biocrates AbsoluteIDQ(®) p180 Targeted Metabolomics Kit Using Human Milk. *Nutrients* **2019**, *11* (8), 1733.
(382) Lippa, K. A.; Aristizabal-Henao, J. J.; Beger, R. D.; Bowden, J. A.; Broeckling, C.; Beecher, C.; Clay Davis, W.; Dunn, W. B.; Flores, R.; Goodacre, R.; et al. Reference materials for MS-based untargeted metabolomics and lipidomics: a review by the metabolomics quality assurance and quality control consortium (mQACC). *Metabolomics* **2022**, *18* (4), 24.
(383) Hoang, C.; Uritboonthai, W.; Hoang, L.; Billings, E. M.; Aisporna, A.; Nia, F. A.; Derks, R. J. E.; Williamson, J. R.; Giera, M.; Siuzdak, G. Tandem Mass Spectrometry across Platforms. *Anal. Chem.* **2024**, *96* (14), 5478−5488.
(384) Samanipour, S.; Choi, P.; O'Brien, J. W.; Pirok, B. W. J.; Reid, M. J.; Thomas, K. V. From Centroided to Profile Mode: Machine Learning for Prediction of Peak Width in HRMS Data. *Anal. Chem.* **2021**, *93* (49), 16562−16570.
(385) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **2016**, *3* (1), No. 160018.
(386) Martens, L.; Chambers, M.; Sturm, M.; Kessner, D.; Levander, F.; Shofstahl, J.; Tang, W. H.; Römpp, A.; Neumann, S.; Pizarro, A. D.; et al. mzML—a Community Standard for Mass Spectrometry Data*. *Mol. Cell. Proteomics* **2011**, *10* (1), No. R110.000133.
(387) Lin, S. M.; Zhu, L.; Winter, A. Q.; Sasinowski, M.; Kibbe, W. A. What is mzXML good for? *Expert Review of Proteomics* **2005**, *2* (6), 839−845.
(388) Larralde, M.; Lawson, T. N.; Weber, R. J. M.; Moreno, P.; Haug, K.; Rocca-Serra, P.; Viant, M. R.; Steinbeck, C.; Salek, R. M. mzML2ISA and nmrML2ISA: generating enriched ISA-Tab metadata files from metabolomics XML data. *Bioinformatics* **2017**, *33* (16), 2598−2600.
(389) Hoffmann, N.; Rein, J.; Sachsenberg, T.; Hartler, J.; Haug, K.; Mayer, G.; Alka, O.; Dayalan, S.; Pearce, J. T. M.; Rocca-Serra, P.; et al. mzTab-M: A Data Standard for Sharing Quantitative Results in Mass Spectrometry Metabolomics. *Anal. Chem.* **2019**, *91* (5), 3302−3310.
(390) Misra, B. B. Advances in high resolution GC-MS technology: a focus on the application of GC-Orbitrap-MS in metabolomics and exposomics for FAIR practices. *Analytical Methods* **2021**, *13* (20), 2265−2282.

(391) Uclés, S.; Lozano, A.; Sosa, A.; Parrilla Vázquez, P.; Valverde, A.; Fernández-Alba, A. R. Matrix interference evaluation employing GC and LC coupled to triple quadrupole tandem mass spectrometry. *Talanta* **2017**, *174*, 72−81.
(392) Raposo, F.; Barceló, D. Challenges and strategies of matrix effects using chromatography-mass spectrometry: An overview from research versus regulatory viewpoints. *TrAC Trends in Analytical Chemistry* **2021**, *134*, No. 116068.
(393) Price, E. J.; Palát, J.; Coufaliková, K.; Kukučka, P.; Codling, G.; Vitale, C. M.; Koudelka, Š.; Klánová, J. Open, High-Resolution EI + Spectral Library of Anthropogenic Compounds. *Front. Public Health* **2021**, DOI: 10.3389/fpubh.2021.622558.
(394) *RECETOX High-resolution filtering workflow*; Zenodo, 2024. DOI: 10.5281/zenodo.11212089.
(395) Kwiecien, N. W.; Bailey, D. J.; Rush, M. J. P.; Cole, J. S.; Ulbrich, A.; Hebert, A. S.; Westphall, M. S.; Coon, J. J. High-Resolution Filtering for Improved Small Molecule Identification via GC/MS. *Anal. Chem.* **2015**, *87* (16), 8328−8335.
(396) Maitre, L.; Guimbaud, J.-B.; Warembourg, C.; Güil-Oumrait, N.; Petrone, P. M.; Chadeau-Hyam, M.; Vrijheid, M.; Basagaña, X.; Gonzalez, J. R. State-of-the-art methods for exposure-health studies: Results from the exposome data challenge event. *Environ. Int.* **2022**, *168*, No. 107422.
(397) Tzoulaki, I.; Ebbels, T. M. D.; Valdes, A.; Elliott, P.; Ioannidis, J. P. A. Design and Analysis of Metabolomics Studies in Epidemiologic Research: A Primer on -Omic Technologies. *American Journal of Epidemiology* **2014**, *180* (2), 129−139.
(398) Kalia, V.; Walker, D. I.; Krasnodemski, K. M.; Jones, D. P.; Miller, G. W.; Kioumourtzoglou, M.-A. Unsupervised dimensionality reduction for exposome research. *Current Opinion in Environmental Science & Health* **2020**, *15*, 32−38.
(399) Momeni, Z.; Hassanzadeh, E.; Saniee Abadeh, M.; Bellazzi, R. A survey on single and multi omics data mining methods in cancer data classification. *Journal of Biomedical Informatics* **2020**, *107*, No. 103466.
(400) Bobb, J. F.; Valeri, L.; Claus Henn, B.; Christiani, D. C.; Wright, R. O.; Mazumdar, M.; Godleski, J. J.; Coull, B. A. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics* **2015**, *16* (3), 493−508.
(401) Carrico, C.; Gennings, C.; Wheeler, D. C.; Factor-Litvak, P. Characterization of Weighted Quantile Sum Regression for Highly Correlated Data in a Risk Analysis Setting. *Journal of Agricultural, Biological, and Environmental Statistics* **2015**, *20* (1), 100−120.
(402) Gibson, E. A.; Nunez, Y.; Abuawad, A.; Zota, A. R.; Renzetti, S.; Devick, K. L.; Gennings, C.; Goldsmith, J.; Coull, B. A.; Kioumourtzoglou, M.-A. An overview of methods to address distinct research questions on environmental mixtures: an application to persistent organic pollutants and leukocyte telomere length. *Environmental Health* **2019**, *18* (1), 76.
(403) Sarker, I. H. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science* **2021**, *2* (6), 420.
(404) Zhang, J. D.; Xue, C.; Kolachalama, V. B.; Donald, W. A. Interpretable Machine Learning on Metabolomics Data Reveals Biomarkers for Parkinson's Disease. *ACS Central Science* **2023**, *9* (5), 1035−1045.
(405) MacKinnon, D. P.; Fairchild, A. J.; Fritz, M. S. Mediation Analysis. *Annual Review of Psychology* **2007**, *58* (1), 593−614.
(406) Davies, N. M.; Holmes, M. V.; Smith, G. D. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* **2018**, *362*, k601.
(407) Sanderson, E.; Glymour, M. M.; Holmes, M. V.; Kang, H.; Morrison, J.; Munafò, M. R.; Palmer, T.; Schooling, C. M.; Wallace, C.; Zhao, Q.; et al. Mendelian randomization. *Nature Reviews Methods Primers* **2022**, *2* (1), 6.
(408) Gao, P.; Shen, X. T.; Zhang, X. Y.; Jiang, C.; Zhang, S.; Zhou, X.; Rose, S. M. S. F.; Snyder, M. Precision environmental health monitoring by longitudinal exposome and multi-omics profiling. *Genome Res.* **2022**, *32* (6), 1199−1214.
(409) Koelmel, J. P.; Stelben, P.; McDonough, C. A.; Dukes, D. A.; Aristizabal-Henao, J. J.; Nason, S. L.; Li, Y.; Sternberg, S.; Lin, E.; Beckmann, M.; et al. FluoroMatch 2.0—making automated and comprehensive non-targeted PFAS annotation a reality. *Anal. Bioanal. Chem.* **2022**, *414* (3), 1201−1215.