Environment International 156 (2021) 106630

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Full length article

Towards a comprehensive characterisation of the human internal chemical exposome: Challenges and perspectives

Arthur David [a,*], Jade Chaker [a], Elliott J. Price [b,c], Vincent Bessonneau [a], Andrew J. Chetwynd [d], Chiara M. Vitale [c], Jana Klánová [c], Douglas I. Walker [e], Jean-Philippe Antignac [f], Robert Barouki [g], Gary W. Miller [h]

[a] Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) – UMR_S 1085, F-35000 Rennes, France
[b] Faculty of Sports Studies, Masaryk University, Brno, Czech Republic
[c] RECETOX Centre, Masaryk University, Brno, Czech Republic
[d] School of Geography Earth and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
[e] Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, United States
[f] Oniris, INRAE, LABERCA, Nantes, France
[g] Unité UMR-S 1124 Inserm-Université Paris Descartes "Toxicologie Pharmacologie et Signalisation Cellulaire", Paris, France
[h] Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY, USA

ARTICLE INFO

Handling Editor: Chennai Guest Editor

Keywords:
Exposome
High-Resolution Mass Spectrometry
Internal chemical exposome
Non-targeted analysis
Suspect screening
EWAS

ABSTRACT

The holistic characterisation of the human internal chemical exposome using high-resolution mass spectrometry (HRMS) would be a step forward to investigate the environmental ætiology of chronic diseases with an unprecedented precision. HRMS-based methods are currently operational to reproducibly profile thousands of endogenous metabolites as well as externally-derived chemicals and their biotransformation products in a large number of biological samples from human cohorts. These approaches provide a solid ground for the discovery of unrecognised biomarkers of exposure and metabolic effects associated with many chronic diseases. Nevertheless, some limitations remain and have to be overcome so that chemical exposomics can provide unbiased detection of chemical exposures affecting disease susceptibility in epidemiological studies. Some of these limitations include (i) the lack of versatility of analytical techniques to capture the wide diversity of chemicals; (ii) the lack of analytical sensitivity that prevents the detection of exogenous (and endogenous) chemicals occurring at (ultra) trace levels from restricted sample amounts, and (iii) the lack of automation of the annotation/identification process. In this article, we discuss a number of technological and methodological limitations hindering applications of HRMS-based methods and propose initial steps to push towards a more comprehensive characterisation of the internal chemical exposome. We also discuss other challenges including the need for harmonisation and the difficulty inherent in assessing the dynamic nature of the internal chemical exposome, as well as the need for establishing a strong international collaboration, high level networking, and sustainable research infrastructure. A great amount of research, technological development and innovative bio-informatics tools are still needed to profile and characterise the "invisible" (not profiled), "hidden" (not detected) and "dark" (not annotated) components of the internal chemical exposome and concerted efforts across numerous research fields are paramount.

1. Introduction

In 2003, the Human Genome Project (HGP) was declared complete. Thirteen years, ~4 billons euros, and a solid international collaboration effort was necessary to sequence the ~3 billion base pairs and estimate that humans have between 20,000 and 25,000 genes (Consortium, 2004). The prospects of the HGP were an enormous step forward to understand "*at the chemical level, the role of genetic factors in a multitude of diseases, such as cancer, Alzheimer's disease, and schizophrenia, that diminish the individual lives of so many millions of people*" (Watson, 1990). Sequencing of the human genome and the development of more affordable high-throughput technologies paved the way for the

* Corresponding author.
E-mail address: arthur.david@ehesp.fr (A. David).

https://doi.org/10.1016/j.envint.2021.106630
Received 7 February 2021; Received in revised form 15 April 2021; Accepted 3 May 2021
Available online 15 May 2021

Definition Box
Definition and nomenclature box (in the context of this article)

GWAS: Genome-wide association studies; **EWAS**: Exposome-wide association studies

HRMS: High-resolution mass spectrometry; **ESI:** Electrospray ionisation; **EI+**: electron [impact] ionisation; **LC**: Liquid chromatography; **GC**: Gas chromatography; **IC**: Ion chromatography; **CE**: capillary electrophoresis.

Metabolite: substance involved in metabolic process [of an entity] (Price et al., 2021).

Metabolome: totality of metabolites [of an entity] (Price et al., 2021).

Metabolomics: the systematic and comprehensive study of metabolites (Price et al., 2021).

Exogenous chemicals: includes all chemical substances (naturally occurring or xenobiotics) from exogenous sources (e.g., food, air, water, soils) that accumulated in humans from intentional or unintentional exposure.

Internal chemical exposome: the totality of internal contact between exogenous chemicals including transformation products and an entity (Price et al., 2021).

***Invisible* internal chemical exposome/metabolome**: components of the internal chemical exposome/metabolome not profiled using a given analytical setup.

***Hidden* internal chemical exposome/metabolome:** components of the internal chemical exposome/metabolome not detected using a given analytical setup because of sensitivity issue.

***Dark* internal chemical exposome/metabolome:** components of the internal chemical exposome/metabolome detected using a given methodological setup but that remain uncharacterised (Da Silva et al., 2015).

Internal chemical exposomics: the systematic and comprehensive study of the internal chemical exposome (Price et al., 2021).

Exposome: All life-course environmental exposures (including lifestyle factors), from the prenatal period onwards (Wild, 2005).

Annotation: assigning a putative identity to *m/z* x Rt feature from HRMS datasets (Uppal et al., 2016). Different levels of confidence can be provided to an annotation. The highest level leading to the identification.

[Metabolic/Chemical] Biomarkers of exposure: exogenous chemical present as parent compounds or transformation products in human biological matrices.

[Metabolic] Biomarkers of effect: measurable change in metabolite profiles in an organism that, depending on the magnitude, can be recognised as associated with an established or possible health impairment.

implementation of genome-wide association studies (GWAS) aiming to use genetic mapping techniques to identify distinct genes affecting disease susceptibility (Bodmer and Bonilla, 2008). GWAS were first implemented to identify common variants (with frequencies ≥5%) influencing the incidence of common multifactorial diseases without prior assumptions and have uncovered previously unknown polymorphic variants associated with a variety of common chronic diseases (Bodmer and Bonilla, 2008). However, most of the variants identified so far in GWAS are associated with low odd ratios (i.e., 1.2–1.5) (Bodmer and Bonilla, 2008). Hence, practical applications of GWAS in terms of understanding the ætiology of a disease and in targeted prevention have been quite limited so far. The search for rare variants having frequencies between 0.1 and 1% in very large GWAS is anticipated to provide a substantial contribution to the study of multifactorial inheritance of common chronic diseases in the near future (Bodmer and Bonilla, 2008; Visscher et al., 2017). However, it has now become evident that the risk for an individual to develop chronic disease also depends on known and unknown environmental factors (Wray et al., 2007).

Environmental exposures, and their interactions with genetic factors, play an important role in many common chronic diseases. For instance, the health impact of environmental risk factors highlighted by the Global Burden of Disease project estimated the disease burden of 84 metabolic, environmental, occupational, and behavioural risk factors in 195 countries and territories, and found that these modifiable risks contribute to approximatively 60% of deaths worldwide (Gakidou et al., 2017; Vermeulen et al., 2020). Yet, compared to genetic factors, the accurate assessment of many environmental exposures in chronic disease epidemiology is underdeveloped (Wild, 2005). To address this, in 2005, C. Wild proposed the exposome as a complement to the genome. Defined as all life-course environmental exposures (non-genetic factors including lifestyle factors), from the prenatal period onwards (Wild, 2005), the exposome concept sought to draw attention to the need for methodological and technological developments in exposure assessment (Wild, 2005). Developing tools to reliably measure exposure history is a challenge because the exposome is highly variable and dynamic (Stingone et al., 2017). Like the HGP beforehand, tackling this challenge will require strong international collaboration.

High-resolution mass spectrometry (HRMS) chemical profiling methods have often been proposed as one of the most promising and operational approaches to tackle the complexity of chemical exposure assessment. These approaches enable the broad measurement of small molecules (generally 50–1200 Da) present in human biospecimen, encompassing endogenous metabolites as well as exogenous chemicals and their transformation products. Therefore, HRMS-based chemical profiling conceptually offers the possibility to identify both biomarkers of exposure (see Definition Box) and biomarkers of effect during the same set of analyses. This enables the identification of exposure signatures associated with disease susceptibility, and at the same time study of the underlying biological mechanisms involved in the development of chronic diseases (Bar et al., 2020; David et al., 2017; Orešič et al., 2020). Hence, the holistic characterisation of exogenous chemicals present as mixtures in human biological samples using HRMS platforms would be a step forward to investigate the environmental ætiology of chronic diseases with an unprecedented precision (Rappaport, 2011; Vermeulen et al., 2020; Wild, 2012).

Many proof of concept studies employing HRMS-based profiling to identify exogenous chemicals and their metabolic signatures in human biospecimen have been reported (Bonvallot et al., 2018; Haddad et al., 2019; Huhn et al., 2021; Jones, 2016; Niedzwiecki et al., 2019; Orešič et al., 2020; Uppal et al., 2016; Vineis et al., 2020). These studies

Box 1
Chemical exposomics and EWAS: where do we stand?

Context: Multifactorial diseases triggered by a combination of genetic and environmental factors are leading causes of mortality and morbidity worldwide. Exposure to chemical agents is a prevalent environmental factor. The comprehensive assessment of exposure, including chemical exposure is challenging and compared to genetic factors, the influence of environmental factors on health is understudied.

A step change in exposure assessment: Application of HRMS-based chemical exposomics in EWAS holds great promise to complement GWAS and identify without *a priori* unrecognised clusters of environmental chemical risk factors associated with multifactorial diseases.

Where do we stand: Hyphenated HRMS configurations can reproducibly profile thousands of endogenous metabolites and hundreds of exogenous chemicals plus their transformation products in human biological samples. Hence HRMS-based chemical profiling provides a solid ground for the discovery of new biomarkers of exposure and metabolic effects associated with many chronic diseases.

Current limitations: Among the hundreds of thousands of exogenous chemicals present on the market, only a small subset can currently be profiled, detected or identified using current HRMS-based methods, introducing a bias for the identification of relevant clusters of environmental chemical factors. Current limitations involve:

The lack of versatility of HRMS technologies limiting the **diversity of chemicals** which can be analysed.

The lack of analytical **sensitivity** which hinders the detection of low-abundant exogenous chemicals (and metabolites).

The lack of ease for accurate annotation meaning that the vast majority of information collected by HRMS-based methods remains, to date, "dark matter".

The **lack of harmonisation** of QA/QC, reporting procedures etc. that hinders concerted large-scale research efforts.

The lack of methodologies capable to assess the **dynamic aspect** of chemical exposure that limits the detection of non-persistent chemical agents and association with latent effects.

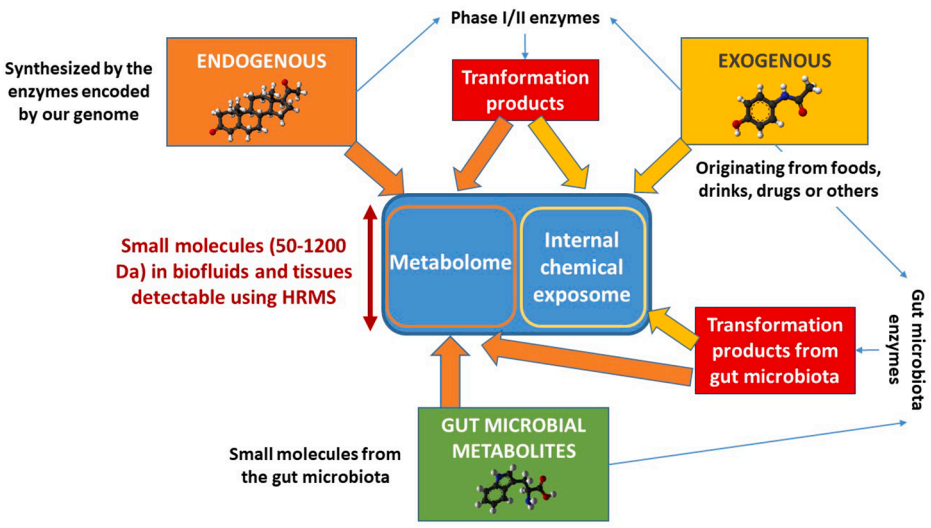

Fig 1. High-resolution mass spectrometry (HRMS) can theoretically profile during the same analysis all small molecules (50–1200 Da) present in human biofluids and tissues, encompassing endogenous metabolites, gut microbial metabolites as well as exogenous chemicals and their transformation products. The clear distinction between exposure and response allows for a better separation (and to avoid confusion) between measures of the endogenous metabolites and measures of exogenously-derived chemicals plus their biotransformation products.

supported the applicability of HRMS-based exposome-wide association studies (EWAS) and their potential to identify chemical risk factors associated with disease susceptibility in epidemiological and/or clinical studies. Now that we have evidence of the merits of HRMS-based methods to identify environmental chemical risk factors, it is critical to refine and further develop HRMS-based methodologies to identify, unbiasedly, the relevant clusters of chemical factors affecting disease susceptibility (see Box 1)).

In this article, we emphasise technical and methodological issues that are hindering HRMS-based chemical profiling and propose initial steps to push towards a more comprehensive characterisation of the human internal chemical exposome. After providing a brief set of definitions for the terms we will use in the context of this article, we discuss (1) the need to combine complementary technologies to capture the large diversity of chemicals; (2) the need to enhance sensitivity in order to detect the low abundant species constituting the internal chemical exposome, even from limited amount of sample available for analysis; (3) the need to improve HRMS data annotation procedures as it is still one of the main bottlenecks of HRMS-based methods. We also discuss other challenges including the need for harmonisation and the difficulty to assess the dynamic aspect of the internal chemical exposome, as well as the need for establishing a strong international collaboration, networking, and sustainable research infrastructure in this field to tackle these different challenges.

2. A brief set of definitions

Recently, it has been proposed to reattribute the 'exposome' to the totality of environmental exposures (i.e. the totality of contact between external non-genetic factors and an individual) (Price et al., 2021). The clear distinction between exposure and response allows for a better separation (and to avoid confusion) between measures of the endogenous metabolites and measures of exogenously-derived chemicals plus their biotransformation products when profiling biospecimen (Fig. 1).

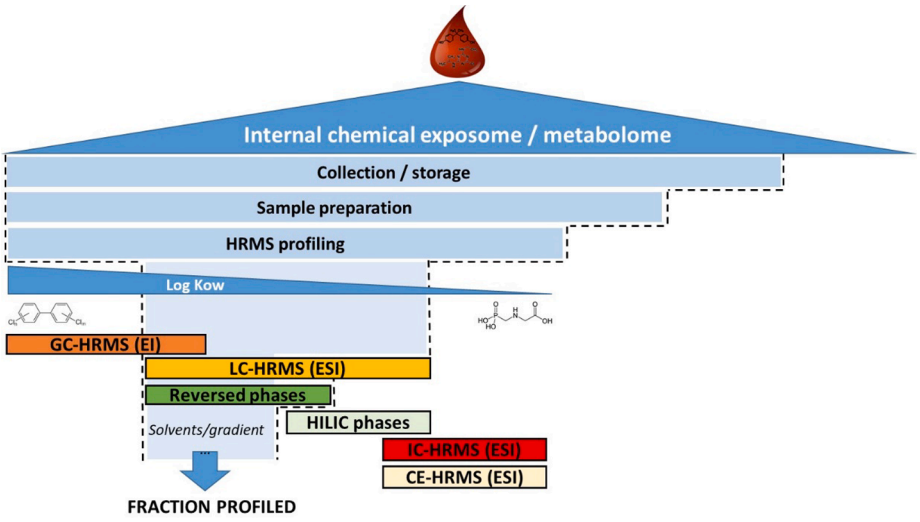

Fig. 2. Conceptual visualisation showing that only a small fraction of the internal exposome/metabolome can currently be profiled when a single LC-ESI-HRMS method is used. All the steps of the analytical workflow will impact on the final composition of the chemical exposome/metabolome. Specificities and overlaps of the different HRMS platforms are schematically represented. Log Kow = octanol/water partition coefficient; GC = gas chromatography; LC = liquid chromatography; IC = ion chromatography, CE = capillary electrophoresis, ESI = Electrospray ionisation.

Accordingly, the term internal chemical exposomics will be used here to refer to the systematic and comprehensive study of the internal chemical exposome, i.e. characterising contact between environmentally-derived chemicals plus their transformation products (50–1200 Da) and a cell, tissue, organ or organism. Metabolomics will refer to the study of the complete set of endogenous metabolites (50–1200 Da) that can be detected in cell, tissue, organ or organism. In this context, chemical exposomics include the different strategies that are currently used for exposure assessment, embracing conventional targeted methods and the increasingly emerging suspect and non-targeted screening approaches (Pourchet et al., 2020). From the strict instrumental point of view, the term HRMS-based chemical analysis/profiling will be used to refer to application of HRMS to both metabolomics and [internal] chemical exposomics, although a number of contextual, scientific and other technical elements distinguish those two domains.

To date, there are no comprehensive characterisations of the metabolome or the internal chemical exposome available for any model organism, tissue or cell (Viant et al., 2017). Therefore, many of the limitations we will discuss are applicable for both chemical exposomics and metabolomics. Nevertheless, providing separate terms will allow to better accommodate the need related to either exposure (e.g. need for acute analytical sensitivity, specific database) or response (e.g., use of biological pathway correlation for metabolite annotation) assessments.

3. The invisible, the hidden and the dark internal chemical exposome components

Great efforts are currently being made to improve the annotation/identification step (which continues to be a bottleneck of HRMS-based methods) as well as to advance the treatment of data generated to obtain exposure/response information. However, other steps in the general workflow, such as the generation of the chemical fingerprints, which includes many analytical steps (e.g., sample preparation, chromatographic separation, HRMS profiling), have yet not received the same attention (Chetwynd and David, 2018; David et al., 2014; David and Rostkowski, 2020). This is detrimental for chemical exposomics as it means that most of the exogenous chemicals, known to be present at trace levels (e.g., sub ng/ml in plasma/serum) compared to many endogenous metabolites (Rappaport et al., 2014), remain undetected with current methods. A great amount of research, technological development and innovative bio-informatics tools are therefore still needed to profile and characterise the "invisible" (not profiled), "hidden" (not detected) and "dark" (not annotated) components of the internal chemical exposome and concerted efforts across numerous research fields (analytical chemistry, bioinformatics, chemometrics) are paramount.

3.1. Capturing the large diversity of chemicals

It has been estimated that >100,000 chemicals are currently in use in the human population while only about 5,000 are estimated to be dispersed in the environment widely enough to pose a global threat to the human population (even though local specificities have to be accounted for) (Di Renzo et al., 2015; Vermeulen et al., 2020). Furthermore, the production of chemicals is predicted to increase by 3.4%/year until 2030 (Di Renzo et al., 2015) meaning that the realm of possibilities in terms of exogenous mixtures accumulating in human will also extend in the future.

These chemicals present in the environment have very diverse physical and chemical properties, for instance from high (e.g., glyphosate) to low (e.g., polychlorinated biphenyls) polarity. This high chemical diversity combined with the lack of versatility of the available technologies means that currently only a small fraction of the internal chemical exposome (and the metabolome) can be captured during a single analysis (Fig. 2). Hence, as opposed to genomics or transcriptomics that can rely on sequencing platforms to obtain comprehensive datasets, a combination of complementary analytical techniques for the sample preparation and HRMS platforms are needed to decipher the depth of the internal chemical exposome (Fig. 2).

This is challenging considering that application of chemical exposomics to epidemiological studies requires consecutive analysis of very large numbers of samples (up to several thousands) to achieve sufficient power to discover statistically significant associations at the population scale. To date, high-throughput HRMS-based methods using minimal sample preparation and/or very short chromatographic runs (or direct infusion) are often favoured so that a maximum number of samples can be analysed in a very short time (Chekmeneva et al., 2017; Gray et al., 2016; Nemkov et al., 2017). In addition to shortening the time of analysis, minimal sample preparation and short LC runs also contribute to reducing the costs. This strategy can be successful to study endogenous metabolites such as amino acids, specific lipid classes or other metabolites or exogenous chemicals present at sufficiently high concentrations. However, this strategy does not always fit well with chemical exposomics which aims to detect exogenous chemicals often present at trace levels in complex biological samples (Rappaport et al., 2014) and that would usually require more selective analytical strategies (e.g., solid phase extraction followed by MS/MS analyses) to be detected (Vorkamp et al., 2021). The challenge is therefore to use strategies that allow a sufficient decomplexification of samples prior

(sample preparation) or during (chromatographic separation) HRMS profiling. Some examples of technological advances or analytical strategies with potential to improve the coverage of the internal chemical exposome are given below.

3.1.1. Sample preparation

Sample preparation has often been overlooked for HRMS-based methods as it has often been argued that an ideal sample preparation method has to be simple, fast, and non-selective to ensure adequate depth of chemical coverage (David and Rostkowski, 2020; Vuckovic, 2012). However, it becomes more and more obvious that introduction of sample preparations with minimal level of selectivity such as solid-phase extraction (SPE) and/or protein removal and delipidation plates are beneficial to properly remove matrix components such as residual proteins and salts (David et al., 2014), and therefore reduce analytical interferences such as ion suppression and improve column reproducibility/prolongation (David et al., 2014; Michopoulos et al., 2009; Vuckovic, 2012). The implementation of minimally selective, e.g., online or offline SPE (David et al., 2014; David et al., 2017; Zhang et al., 2016), ultrafiltration or delipidation using commercially available lipid depletion plate (David et al., 2014; Tulipani et al., 2013), solid-phase microextraction (Bessonneau et al., 2017) or sequential sample preparation (Yang et al., 2013) to fractionate complex biological samples have already be used to improve the coverage the internal chemical exposome/metabolome.

Other sample preparation methods promising for higher sample throughput without compromising chemical coverage have been recently reviewed by Miggiels et al. (2019) for metabolomics applications. These methods include dispersive and non-dispersive liquid–liquid micro extraction (LLME) (Alexovič et al., 2017) which enable the fractionation of hydrophilic and hydrophobic chemicals, as well as electro-driven extractions (EE) which targets polar, charged analytes. EE is directly applicable for capillary electrophoresis (CE) but also now for hyphenation to LC, GC or direct injection thanks to recent developments increasing the versatility of EE (Miggiels et al., 2019; Oedit et al., 2016).

Most of the sample preparation methods mentioned here include miniaturisation and automation potential for high-throughput implementation, can improve pre-concentration factors and extract enrichment, and reduce extraction time while reducing reagent consumption (David et al., 2014; He et al., 2021; Miggiels et al., 2019). Notably, the notion that increasing selectivity of sample preparation such as SPE reduces the coverage due to greater compound losses is not necessarily true because lower recoveries can be compensated by the concentration of extracts or the removal of ion suppressing components (David et al., 2014). The miniaturization of the sample preparation process (i.e. use of limited amount of sorbent) is also beneficial when only limited amounts of precious biological samples are available.

Other strategies relying on the use of isotope labelling of urinary biomarkers with common functional groups (phenolic hydroxyl/carboxyl/primary amine) combined with a specific computational pipeline method for qualification and quantification have been proposed to improve the characterization of the urine chemical exposome (Jia et al., 2019).

It is nevertheless important to consider that any sample preparation selected for a specific study will impact the recoveries of the various component of the internal exposome. Hence, we discuss in the section *Harmonisation issues* the qualitative and quantitative parameters that could be considered to document the perimeter of the internal chemical exposome profiled and potentially validate a sample preparation method.

3.1.2. Chromatographic separation and ion mobility

Concurrently or alternatively, the use of dual or two dimensional (2D) chromatographic systems could be used to improve the coverage of the internal chemical exposome. In his review, Jones (2016) describes the use of dual chromatography setup in which one LC column is washed and re-equilibrated while the mass spectrometer is used to analyse a sample on another column. With this approach, the authors were able to use two types of columns to increase detection to more than 4000 chemical features (Soltow et al., 2013), and this was further increased to >10,000 by acquisition of an LTQ-Velos Orbitrap and dedicating use to HRMS (Jones, 2016).

The incorporation of ion mobility spectrometry (IMS) into current HRMS-based analytical methods is gaining more and more attention to improve the coverage, dynamic range and throughput of measurements (Metz et al., 2017). Many different IMS-based technologies have been developed so far and all have their specificities. Overall, the main advantage of IMS is to resolve isomers or isobars (which depends on the resolution of the technology used) that are difficult to distinguish using LC-MS alone. This can be extremely useful considering that many small molecules such as endogenous and exogenous chemicals have the same molecular formula and similar molecular structures but play very different roles in biological systems (Metz et al., 2017). LC-IMS-HRMS technologies produce higher volume of data that are more difficult to interpret, but the addition of another physico-chemical parameter (i.e. the collisional cross section) is also definitely something very valuable for the annotation process (Plante et al., 2019; Zheng et al., 2017).

3.1.3. Multi HRMS platform approaches

Concurrently, integration of multiple HRMS platform approaches is necessary to circumvent limitation related to the lack of versatility of a given technology. At present, the predominant HRMS-based platforms are liquid-chromatography equipped with an electrospray ionisation source and coupled with time-of-flight or Orbitrap HRMS analysers (LC-ESI-HRMS) (David and Rostkowski, 2020). These platforms have numerous advantages compared to others as they provide a soft ionisation process and a high dynamic range and versatility. LC-ESI-HRMS have the possibility to analyse efficiently mainly non-volatile compounds and their metabolites (as well as phase I biotransformation products from volatile chemicals). Different chromatographic set-ups (from reverse phases, HILIC or normal phases) are needed to capture the wide range of endogenous and exogenous chemicals detectable with LC-ESI-HRMS.

The use of gas chromatography (GC) coupled to HRMS is becoming increasingly more popular to profile human biological matrices in environmental health studies (Xin et al., 2020). In particular, GC-HRMS is very efficient to detect semi-volatile compounds such as persistent organic pollutants (POPs) (Valvi et al., 2020), many of which are recognised as carcinogens or endocrine disruptors and are not accessible with LC-HRMS. While these compounds are historic contaminants, sometimes restricted or banned for many years (e.g., organochlorine insecticides such as DDT), they still linger in the environment and are still present in our body because of their long half-lives. GC-HRMS does not suffer from the same degree of matrix effects as LC-ESI-HRMS. However, large spectral libraries are needed for the annotation process as the predominant ionisation process usually does not allow detection of the molecular ion (Stein, 2012). Derivatization of polar compounds (to reduce their polarity and improve thermal stability and volatility) can be done during sample preparation for GC-HRMS based methods to extend the versatility of the technique (David and Rostkowski, 2020) and many developments are currently being made to advance data treatment of GC–MS (Aksenov et al., 2020).

Another set of both endogenous and exogenous chemicals still remains, most of the time, out of our reach during HRMS-based profiling, even with the combination of GC-HRMS and LC-ESI-HRMS. The very polar chemicals (e.g. glyphosate and its metabolites) are usually difficult to analyse and usually require specific techniques to analyse them individually (e.g., derivatization techniques or ion pairing agents). The use of normal phase or HILIC columns with LC-ESI-HRMS platforms can help to detect these compounds but are sometimes difficult to implement in routine and are still more rarely utilised than reversed phase methods.

To improve the detection of highly polar components of the internal

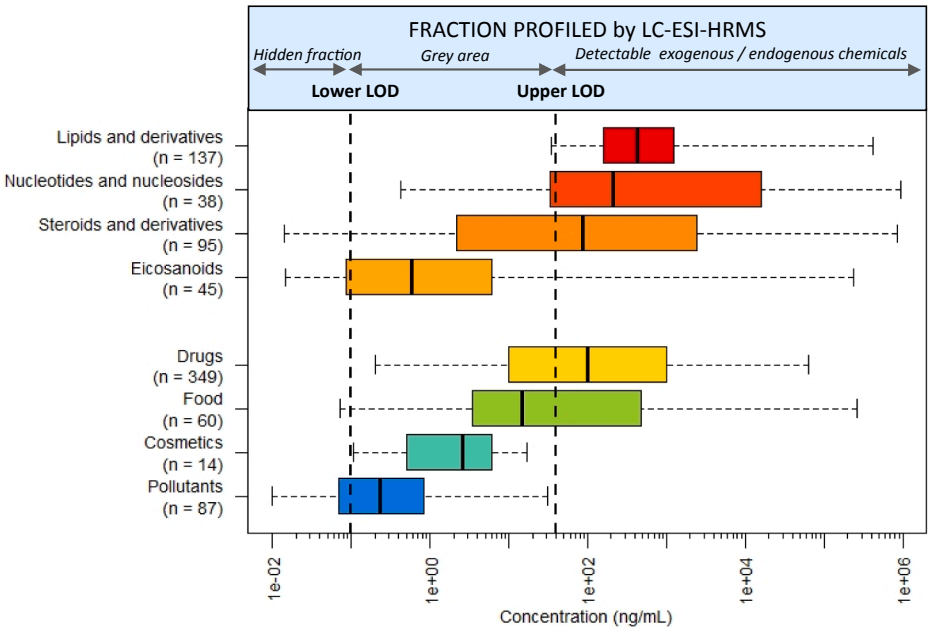

Fig. 3. Blood (plasma and serum) concentrations of different classes of endogenous and exogenous chemicals analysed using targeted MSMS methods and found in databases or from the scientific literature. Detection limits (lower and upper limits) measured for 30 exogenous chemicals (molecular weight from 133.15 to 424.53; logP from 0.07 to 5) spiked in either plasma or serum (from 0.01 to 50 ng/ml) and analysed in full scan using a method based on UHPLC-ESI-QTOF (X500R Sciex) are represented. Pollutants include insecticides, fungicides, herbicides and biocides.

chemical exposome and metabolome, CE-HRMS offers great promises as it provides efficient separation for highly charged and polar metabolites (Zhang and Ramautar, 2021). CE-HRMS has been growing in prominence in metabolomics in the past decade with the advent of sheathless platforms for improved sensitivity in addition to enhanced thermal regulation and improved capillary manufacturing allowing for highly reproducible migration times and peak areas (Drouin et al., 2020; Zhang and Ramautar, 2021). Due to the low flow rates employed of around 20 nL/min CE-HRMS benefits from the associated increases in ionisation efficiency and lack of chromatographic dilution allowing for highly sensitive analysis. The technique is particularly well suited for volume restricted samples (<5 μL) such as biopsies, cerebrospinal fluid or infant blood due to the small sample injection volumes required of a few nL per injection (Zhang et al., 2019). When higher sensitivity is required, sample pre-concentration can be applied by switching from hydrodynamic injections to variations on electrokinetic injections such as transient isotachophoresis or field amplified sample injection which have been extensively reviewed previously (Zhang and Ramautar, 2021). Combined, this makes CE-HRMS a valuable addition to the analytical toolkit moving forwards in the field of chemical exposomics, even though stability issues still prevent large-scale adaptation (Miggiels et al., 2019). Similarly, ion exchange chromatography (IC) coupled to HRMS has great, and yet unexplored, potential to profile the polar fraction of the chemical exposome since it has the ability to separate small charged non-volatile inorganic and organic compounds in diluted aqueous samples/extracts (Gallidabino et al., 2018). Direct injections of diluted samples such as urine make the entire procedure more rapid and straightforward (Gallidabino et al., 2018). Routine implementation of methods using IC or CE coupled to HRMS could definitely help to increase the coverage of polar components of the internal chemical exposome and the metabolome.

In the future, running samples on multiple HRMS platform approaches (using appropriate sample preparation methods depending on the technique used) will definitively contribute to enhance the coverage of the human chemical exposome by improving the overall separation selectivity of HRMS-based methods. However, it is also crucial to bear in mind that the coverage of the chemical space is also dependent on the analytical sensitivity of the technique so that low-abundant exogenous chemicals can also be detected. Presentation of HRMS systems that provide improved analytical sensitivity is covered in the *sensitivity issue* section.

3.1.4. *Towards the implementation of international research infrastructures*

To study human exposure to chemical factors for application in EWAS, we need to be prepared for a limitless combination of possibilities of mixtures of exogenous chemicals potentially involved in the diseases of interest due to personal history, local specificities, etc. To this aim, the development of large MS facilities (e.g., Hercules Exposome Research Center, Emory University, US, MRC-NIHR National Phenome Centre, UK) equipped with a wide range of HRMS-based instruments have the potential to reduce this bias. We also need to rely on the innovation potentials from smaller labs that are able to develop and propose new techniques allowing to profile specific families of endogenous/exogenous chemicals (e.g. polar chemicals) and their transformation product that stay, at present, invisible to us. Establishing a distributed European Research Infrastructure supporting a strong collaboration and harmonisation of relevant laboratories/analytical facilities in and outside Europe could offer a range of techniques and provide a sufficient capacity to profile as much as possible of the chemical exposome. In order to share biological samples (sometimes available in very low volumes) within the network, we need to develop complex workflows for sample preparation and use miniaturized analytical methods in order to reduce the volume/amount of samples needed for each technique.

3.2. *A sensitivity issue*

In addition to the invisible fraction of the internal chemical exposome, a large part of the internal chemical exposome remain hidden because of sensitivity issues. The poor detection of the components of the internal chemical exposome with current HRMS-based methods can be explained by the low concentrations of exogenous chemicals in human (e.g., sub ng/ml in plasma/serum) (Rappaport et al., 2014), poor ionization efficiency of ESI sources partially due to ion suppression or analytical interferences with highly abundant components of biological matrices. We discuss here the ion suppression phenomenon, which is one of the main issues of LC-ESI-HRMS (the most commonly used HRMS platform at present for chemical exposomics) and present some solutions that could be already implemented to improve the detection of the low-abundant components of the internal chemical exposome.

3.2.1. Ion suppression and ionisation efficiency of LC-ESI-HRMS

Ion suppression is one of the main factors that affects analytical performance, and detection limits for HRMS profiling based on LC-ESI-HRMS (Annesley, 2003; Antignac et al., 2005). The mechanisms of ion suppression in the case of LC-ESI-MS have been already extensively described in several reviews, e.g. (Annesley, 2003; Antignac et al., 2005; Trufelli et al., 2011). Briefly, the main cause of ion suppression is a change in the spray droplet solution properties (i.e., increase in viscosity and surface tension) caused by the presence of nonvolatile or less volatile solutes (Annesley, 2003; Antignac et al., 2005; Trufelli et al., 2011). These interfering or co-eluting non-volatile chemicals decrease the evaporation efficiency and therefore decrease the number of charged ions in the gas phase that ultimately reaches the detector (Annesley, 2003; Antignac et al., 2005). Furthermore, competition between analytes regarding the maximal ionization efficiency of the technique is another mechanism that could explain ion suppression (Antignac et al., 2005). This is especially true when low abundant exogenous chemicals are co-eluting with highly abundant metabolites (David et al., 2014). Consequences of ion suppression is a decrease of the analyte signal (or total suppression), or even a loss of mass accuracy which in turn can lead to the non-detection of many exogenous and endogenous chemicals present at low concentrations.

To illustrate to what extent current full scan HRMS-based methods are impacted by the sensitivity issue, we compiled concentrations in blood (plasma and serum) from a wide range of endogenous and exogenous chemicals mainly measured using targeted MS/MS methods and found in databases such as the Human Metabolome DataBase (HMDB) (Wishart et al., 2018) or from the scientific literature (Fig. 3). We focused mainly on chemicals detectable using LC-ESI-MS methods. We then experimentally measured limits of detection (LODs) for a mixture of 30 exogenous chemicals with a wide range of physical and chemical properties (molecular weight from 133.15 to 424.53; logP from 0.07 to 5) spiked at different concentrations (from 0.01 to 50 ng/ml) in either plasma or serum (prepared using methanolic protein precipitation). Samples were analysed in full scan mode using one of the latest generation of UHPLC-ESI-QTOF (X500R Sciex) with a one-hour long LC chromatographic separation using a reversed phase column (Waters Acquity UHPLC HSS-T3 column, 1.0 mm × 150 mm × 1.8 μm) and a flow rate of 100 μL/min to decrease co-elution at the maximum. The LODs were chemical-dependent and ranged from 0.1 to 40 ng/ml and can be seen in Fig. 3.

The results obtained here with the compilation of endogenous and exogenous chemicals are consistent with those previously reported on the blood chemical exposome (Rappaport et al., 2014). Of note is that mass concentrations were used to generate this figure (as opposed to molar concentrations for Rappaport et al. (2014)) and that we filtered for LC-ESI-MS detectability which eliminated a large share of compounds reported in Rappaport et al. (2014). We observed a dynamic range of mass concentrations (from 10^{-2} to 10^{6} ng/ml) through eight orders of magnitude. We observed that compared to concentrations of endogenous metabolites (median, first and third quartile concentrations of 170 [13; 1300] ng/ml), pharmaceuticals (100 [10;1000] ng/ml) and, to some extent, food chemicals (for which a large variability was observed, median, first and third quartile concentrations of 15 [4; 4400] ng/ml) were similar, while concentrations of exogenous chemicals such as pesticides and plasticizers were generally 400-700 times lower (concentrations of 0.23 [0.07; 0.82] ng/ml).

By superposing LODs on the ranges of concentrations compiled for classes of endogenous and exogenous chemicals, we highlight that (1) HRMS-based methods are still not sensitive enough to detect important exogenous chemicals accumulated in blood with known toxicity potential, (2) fractionation of complex biological samples might be necessary to account for the dynamic ranges of concentrations observed (the dynamic range of most detectors is ~6 orders of magnitude), and (3) the lack of sensitivity is also very relevant for some metabolites involved in important signalling pathways, such as steroids and eicosanoids (Funk, 2001; Harizi et al., 2008). The need for an improved analytical sensitivity therefore appears mandatory for the more comprehensive detection of exogenous chemicals in complex biological matrices. This future technical progress toward better sensitivity will also benefit scientists working in the field of metabolomics (Gallart-Ayala et al., 2020).

3.2.2. Increasing analytical sensitivity using micro or nanoflow-nanoESI-HRMS

While the performance and the capacity to deal with matrix effects of HRMS technologies have considerably improved, additional progress could be made to lower detection limits of HRMS-based methods. One of the most obvious would be the implementation of nanoflow (or microflow) LC- nano ESI platforms (Chetwynd and David, 2018; Chetwynd et al., 2014; David et al., 2014; David et al., 2017), using flow rates on the nL/min scale (10–1000 nL/min) with ESI emitters' internal diameters (i.d.) ranging between 10 and 50 μm (Chervet et al., 1996; Wickremsinhe et al., 2006). The increased sensitivity of the nLC-nESI can be attributed to the use of nESI emitters producing droplets that are 100–1000 fold smaller than the typical droplets emitted from conventional ESI emitters (Wilm and Mann, 1996). The generation of significantly smaller plume droplets with lower volumes considerably increases the rate of desolvation, resulting in up to 500 times more ions being formed and entering the mass analyser (Karas et al., 2000; Marginean et al., 2008; Marginean et al., 2014; Wilm and Mann, 1996). In proteomics, the use of nanoflow (or microflow) is widespread and plays an important role to ensure a comprehensive coverage of the proteome. Automated nanoflow two-dimensional reversed-phase LC system has even been implemented to enable a wider coverage of the proteome and phosphopeptides (Dou et al., 2019). However, the uptake of nanoflowLC-nanoESI for small molecules analyses is still very slow, which this could be explained by the following technical limitations: (1) sample preparation requires extra care to reduce clogging and extend the column and emitter lifetimes and (2) these methods have longer chromatographic run times and are therefore less compatible with high throughput and large scale analysis than conventional LC-ESI platforms (Chetwynd and David, 2018; David et al., 2014). Hence, the implementation of microflow (e.g., using 0.3 mm i.d. columns, ESI emitters i.d. of 50 μm and flow rates ranging from 1 to 50 μL/min) could be an attractive alternative to keep the benefits of low flow without significantly having all the above-mentioned analytical limitations (Geller et al., 2021). Furthermore, it can be highlighted that, similarly to CE-HRMS, micro or nanoflow-nanoESI platforms required lower injection volumes (usually below 1 μL) which is advantageous for volume restricted samples.

3.2.3. Ultra-fractionation of complex biological samples

To date, it is not known how far (sensitivity-wise) and how wide we need to go in order to profile the entire "chemical space" characterising complex biological samples. This in-depth examination could be done by performing an ultra-fractionation of biological samples using a combination of different SPEs/semi-preparative HPLC systems and HRMS platforms, finally leading to complementary profiles of each analysed sample. Effect direct analyses (EDA) (Brack, 2003) using a range of various bioassays (e.g., *in vitro* cell lines) with different endpoints (e.g., estrogenicity, anti-androgenicity…) could be also used in order to guide HRMS profiling and isolate fractions showing presence of relevant biologically active chemicals still missing. This would involve a high level of networking and could help to close the gap between environmental and human health communities, which are often relying on similar tools to identify environmental contaminants. Providing quantitative data alongside non-targeted investigations would allow to provide layers of concentrations (from sub ng/ml(g) to hundreds of μg/ml(g) up to mg/ml (g)) for the different exposures encountered and indicate which analytical techniques would be the most appropriate in the future to monitor the different chemical exposures.

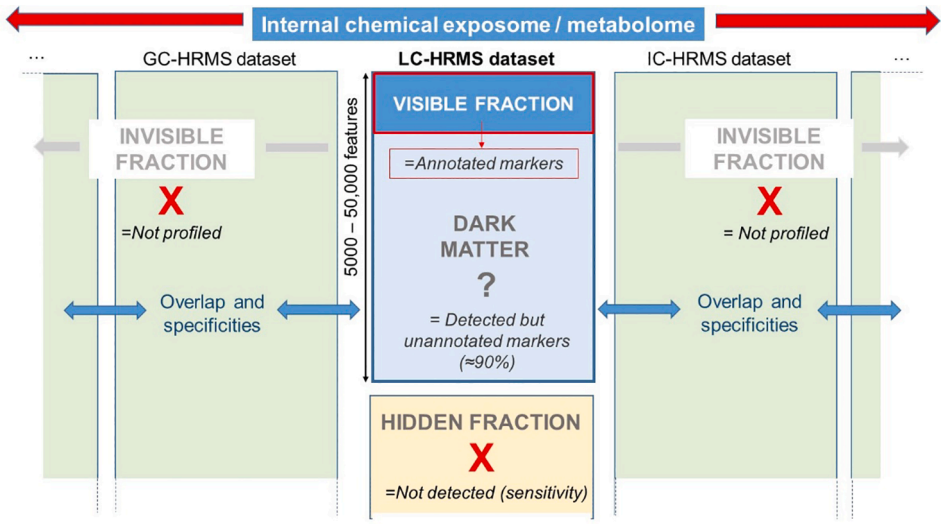

Fig. 4. Example of the conceptual visualisation of the visible fraction of the internal chemical exposome / metabolome when biofluids or tissues are analysed only using a LC-ESI-HRMS method. It is widely acknowledged that only a small portion of the 5000–50,000 features generated using HRMS-based methods can be annotated. An unknown share is not detected because of sensitivity issue (hidden fraction) while other chemicals are not detected using th LC-ESI-HRMS platforms (invisible fraction) because of their physical and chemical properties.

3.2.4. *Finding a balance between high-throughput and analytical sensitivity*

Interestingly, most of the strategies previously described to improve the coverage of the internal chemical exposome will be also relevant to improve the sensitivity issue. These two lines of work are interconnected and should be developed concomitantly to feed each other. Some solutions are already available but are usually not considered because most of them are deemed too time consuming compared to conventional HRMS methods. Scaling-up EWAS is fundamental but a compromise has to be found in order to go deeper into the "sequencing" of the internal chemical exposome. Improving analytical sensitivity will also help to ensure that bioinformatics tools designed to process LC-HRMS data can disentangle the less abundant chemicals' signals from the noise and therefore help reducing the rates of false negative that can be observed during the peak picking process (Chaker et al., 2021).

3.3. *The annotation issue*

Despite the fact that current HRMS-based methods cannot profile and detect the vast expanse of the metabolome and the internal chemical exposome, these methods are already operational to produce very large datasets including >10,000 features (represented as a Rt × *m/z* together with its abundance measured as area/intensity). These complex HRMS datasets are a gold mine of information and it is therefore important to develop automated workflows to improve the annotation of these features.

3.3.1. *A tedious and incomplete process*

Identification of features is a tedious process involving a final step to confirm the structure of the putative feature by matching exact mass, retention time (Rt), isotopic pattern and MS/MS spectra with those from a reference standard (Schymanski et al., 2014). It is usually admitted that only a small percentage (usually below 10%) of all features generated in HRMS can be identified. This percentage is varying from one experiment to another and depends on the time invested, but in the end, it is a known fact that the vast majority of features remain, for now, unannotated or unknown. This means that the vast majority of information collected by HRMS-based methods remains, to date, "dark matter," i.e. chemical features/signatures that remain unannotated (da Silva et al., 2015) (Fig. 4).

Annotation is even more complex for exogenous chemicals considering that online libraries such as HMDB (Wishart et al., 2018) or KEGG (Kanehisa, 2002) were initially designed for the annotation of the metabolome (Wishart et al., 2007). Several databases such as CompTox Chemicals Dashboard (Williams et al., 2017), which include environmentally and toxicologically-relevant compounds, or the Blood Exposome Database (Barupal and Fiehn, 2019), which includes >60,000 chemicals, are now available for exposomics research and many large databases such as PubChem include information relevant for the annotation of the internal chemical exposome. An extended and open access database dedicated to exposure markers has been also developed within the European Human Biomonitoring for Europe (HBM4EU) initiative, that encompass more than 70,000 parent substances and more than 300,000 known or modelled metabolites (Meijer et al., 2021). Developing tools to automatise the annotation of complex HRMS datasets is another priority to move the field of chemical exposomics forward.

3.3.2. *Developing MS/MS libraries and in silico fragmentation software*

The *m/z* of a supposed molecular ion generated in HRMS with an accuracy of 5 ppm can still produce 1000 of candidates when searching large databases such as KEGG or PubChem (Kind and Fiehn, 2006). It is therefore necessary to connect additional predictors such as the experimental MS/MS data of the candidate with those of chemical structures found in public MS databases such as HMDB (Wishart et al., 2018) or Metlin (Guijas et al., 2018; Xue et al., 2020) in order to narrow down the search to a limited numbers of serious candidates. But this approach is hampered by the fact that MS/MS data are available for only a very limited number of chemicals and have been created largely with commercially available compounds (Dührkop et al., 2020). For instance, there are about 100 million molecules in PubChem, but only hundreds of thousands of MS/MS spectra representing about <50,000 molecules that are accessible for non-targeted experiments (da Silva et al., 2015). More recently, Oberacher et al. (2020) explored representative percent coverage of measured MS/MS spectra in selected major environmental suspect databases of interest in the context of human biomonitoring, and confirmed the current very large gap between the number of potential substances of interest (up to hundreds of thousands) and measured spectra (0.57–3.6% of the total chemicals have spectral information available). In this line of work, great efforts are being made within the HBM4EU initiative to provide harmonization guidelines for the acquisition and processing of MS/MS data (Oberacher et al., 2020).

Besides acquisition of experimental MS/MS data, there is no doubt that computational methods such as machine learning and molecular network approaches will play a pivotal role to help automate this annotation process. In that respect, tools such as CSI (compound structure identification):FingerID (Dührkop et al., 2015) are among the most promising approaches to aid in the annotation of chemical features. CSI: FingerID uses fragmentation trees that best explains the fragmentation spectrum of an unknown molecule. Then the fragmentation tree is used

to predict the molecular structure fingerprint of the unknown compound using machine learning (Dührkop et al., 2015). Other *in silico* fragmentation software based on fragmentation trees (e.g. MetFrag (Ruttkies et al., 2016) or MAGMa (Ridder et al., 2012)) are now commonly utilised for computer assisted annotation of unknown features. However, manual curation (e.g., match with Rt and MS/MS from a standard, orthogonal method, additional information regarding the use/biological function of the metabolite or exogenous chemicals) is still needed to confirm identification or to increase the level of confidence for an annotation (Schymanski et al., 2014).

3.3.3. Challenges related to the acquisition of reliable and informative MS/MS data

The development MS/MS libraries (and *in silico* predicted MS/MS spectra) is a prerequisite to advance the efficiency (as well as the automation) of the annotation process. However, the choice of the acquisition strategies used to provide MS/MS data for a large numbers of samples is crucial to ensure that reproducible MS/MS patterns are obtained but also to ensure that MS/MS acquisition method is triggered for low-abundant components of the chemical exposome.

In recent years, advances in data multi-event acquisition offered different solutions for MS/MS acquisition such as methods data-dependent (DDA) and data-independent acquisition (DIA) (Fenaille et al., 2017). In DDA, an MS/MS event can be triggered based upon the most intense peaks observed in the latest MS1 survey scan (Davies et al., 2021). A benefit of DDA is that the MS2 spectra can be associated to a specific precursor (*m/z* isolation window typically of the order of 1 Da); the disadvantages being that limited number of ions can be fragmented within a single injection and the stochastic nature of fragmentation (Barbier Saint Hilaire et al., 2020; Davies et al., 2021). DIA operates in a less-selective manner in the way that valuable MS/MS data are obtained on all ions. SWATH (Sequential Window Acquisition of all Theoretical fragment-ion spectra) or SWATH-like methods are the most used DIA acquisition workflow (Gillet et al., 2012). It involves the selection of wide and consecutive *m/z* isolation windows (typically between 10 and 50 Da) to cover the whole mass range; then all ions within these windows are fragmented simultaneously (Barbier Saint Hilaire et al., 2020). Other DIA acquisition workflows include all ion fragmentation (AIF) and MS^E. Data resulting from DIA require more processing, in particular to link a specific precursor to its fragments (Barbier Saint Hilaire et al., 2020; Davies et al., 2021).

Several studies have evaluated DDA and DIA approaches for metabolomics/chemical exposomics applications and these reports demonstrate the value and complementarity of DDA and SWATH-type DIA approaches, e.g., (Barbier Saint Hilaire et al., 2020; Guo and Huan, 2020). However, additional studies would be needed to improve MS/MS acquisition methods and optimize parameters allowing to trigger MS/MS for low-abundant exogenous chemicals in various complex human biological matrices (e.g., blood, urine) as well as to assess the impact of selected analytical parameters (sample preparation and extended chromatographic separation using low flow LC) on the acquisition of reliable MS/MS data for the internal components of the chemical exposome.

3.3.4. Automation of suspect screening workflow and developing robust MS1 annotation

Another approach used to speed-up the annotation process that has gained in popularity is called the suspect screening strategy. The first step consists in building a library of chemicals (exogenous and/or metabolites and their respective metabolites) suspected to be present and/or active in toxicity assays. The number of exogenous and endogenous chemicals that can be screened is theoretically unlimited and can be adapted to the focus of the study (Moschet et al., 2013). However, parameters other than the *m/z* must be added in the suspect screening workflow (e.g., experimental or predicted retention time, isotopic patterns, MS/MS data) to decrease dramatically the rate of false positives and therefore limit the number of putative annotations that need manual curation. Several tools have been developed for the prediction of retention time such as PredRet, (Stanstrup et al., 2015) the RTI platform (Aalizadeh et al., 2019), Retip (Bonini et al., 2020) and simple linear regression using log Kow or log P (Bade et al., 2015). Recent research has shown that Rt predictions have great potential to improve scoring for MS1 suspect screening to prioritize pre-annotated features and reduce false positives (Chaker et al., 2021; Hu et al., 2018). Besides *m/z* and Rt, adding parameters linked to isotopes can help to filter exogenous chemicals having distinguishable isotopic patterns (e.g., compounds with halogenated atoms such as Cl or Br) (Chaker et al., 2021). In addition to MS1 descriptors, it is now possible, with the development of *in silico* fragmentation software, to include MS/MS data for a wide range of chemicals present in the database (Chao et al., 2020). In order to provide annotation scoring, intermediate annotation scores for each predictor (e.g., *m/z*, Rt) can be developed as well as a composite score, via automated workflow (Chaker et al., 2021; Hu et al., 2018).

It is worth mentioning that in the absence of reliable MS/MS data for a given feature (e.g. because of very low MS1 signals in the sample), which is relatively common for exogenous chemicals, the combination of MS1 predictors based on *m/z*, Rt prediction, isotope ratios, and the presence of metabolites originating from the same parent compounds could already provide a solid ground for the annotation process.

3.3.5. Implementing biotransformation metabolites recognition in annotation tools

Another difficulty that hampers the detection of components of the internal chemical exposome lies into the fact that many of the databases lack information about the potential structures of chemicals undergoing biological transformation (e.g., phase I/II metabolites). Many exogenous chemicals (e.g., plasticizers, pesticides, cosmetics) will go through extensive liver (and to some extent kidney and intestinal) biotransformation rapidly after ingestion. Computational tools such as Meteor Nexus (Marchant et al., 2008) or BioTransformer (Djoumbou-Feunang et al., 2019) can be used to predict some of the potential phase I/II metabolites. However, many enzymatic reactions involved in the metabolism of non-persistent chemicals remain unknown, in particular for the gut microbial metabolism (David et al., 2021). Therefore, more experimental database providing accurate liver, kidney or gastro intestinal biotransformation processes (including for instance the starting reactant, the reaction product and the name or type of the enzyme catalysing the biotransformation) would be very valuable to help for the annotation of complex HRMS datasets by identifying new enzymatic pathways that can be used as pattern recognition for a vast array of chemicals.

3.3.6. Producing high quality annotated datasets for machine learning

Despite the fact that many tools are already under development to help with automation of the annotation process, there is no doubt that we will have to go through a step of manual annotation using expert knowledge to confirm the annotation provided by these tools. This step of manual annotation is fundamental because the same chemicals present in 1000s of biological samples such as urine or blood can show subtle changes regarding the qualitative parameters used for its annotation (e.g., absence/presence of fragments, variability of ratio of specific fragments, drift in Rt, variability of isotopologues ratio, etc…) that can be explained by the complexity of the matrix and the inter-individual variations. Defining the acceptable variability for these qualitative parameters used for the annotation can only be done as first step through manual expertise by checking individually spectra and chromatograms. The more we will provide robust annotated datasets confirmed by HRMS experts, the more we will provide robust training datasets for machine learning. Manual annotation/curation will therefore play a central role by providing reliable datasets that will allow to decipher the chemical exposome using machine learning tools.

4. Harmonization issues

Harmonizing QA/QC procedures will provide confidence in the reliability and robustness of the data produced, and allow to characterise, to some extent, the portion of the chemical exposome that can be profiled using a specific methodology (which in return could help to validate or invalidate putative annotations) (Monteiro Bastos da Silva et al., 2017).

4.1. Developing harmonised quality assurance/quality criteria for method validation

So far, parameters used to validate sample preparation method for metabolomics/chemical exposomics include mostly qualitative parameters (e.g., the number of features generated -blanks excluded- or the main classes of exogenous and endogenous chemicals detected). Adapting quantitative parameters commonly used to validate targeted multi-residue methods such as recovery experiments, detection/quantification limits, repeatability/reproducibility using a set of exogenous and endogenous chemicals carefully chosen (e.g., with a wide range of K_{ow}) (David et al., 2014; Schulze et al., 2020) would help to document the perimeter of the internal chemical exposome profiled. Harmonisation could be done regarding sets of exogenous and endogenous chemicals to be tested or the choice of common reference materials for these quantitative parameters (Caballero-Casero et al., 2021). Validation methods should also be extended to bioinformatics tools; parameter selection of bio-informatics software for data pre-processing tools should be assessed to ensure (1) optimization of peak picking to limit false negatives; (2) poor data integration; and (3) optimized computing time (Chaker et al., 2021). Similarly, mixtures of exogenous and endogenous chemicals spiked at low and high levels in the biological matrices under study could be used to assess the performance of this step.

4.2. Implementing standard procedures for batch effect removal

Standard procedures for the data pre-processing/treatment component of HRMS-based profiling are still needed to correct for intra- or inter-batches instrumental drift (e.g., loss of analytical sensitivity, drift in Rt) (Liu et al., 2020). Batch effects removal is especially important for large scale applications where the analyses can run from several consecutive days up to several weeks. Composite quality control (QC) samples reflecting the average exogenous and endogenous chemical concentrations from all study samples run in all batches can be used to this aim and also assess the analytical performance for all the compounds by calculating the relative standard deviation (population standard deviation divided by the population mean) in these pooled samples (van der Kloet et al., 2009). Workflows using pooled QC samples and multiple internal standard strategy have already been developed (Liu et al., 2020; van der Kloet et al., 2009). Recently, several metrics such as retention time drift, number of compounds detected, missing values, and MS reproducibility were proposed to assess data quality for LC–MS-based global metabolomics with the goal of providing an efficient and improved ability to evaluate the data quality (Zhang et al., 2020). Reference standardization protocol can also be used to assess concentrations of individual chemical in unknown samples by comparison to a concurrently analysed, pooled reference sample with known chemical concentrations (Go et al., 2015).

Adopting a common scaling and normalisation workflow is also important to compare datasets coming from different HRMS instruments or orthogonal methods (or to integrate data from positive and negative ionization modes). Several strategies have been developed to combine heterogeneous datasets, such as statistical heterospectroscopy (for MS and NMR datasets) which works through the analysis of the intrinsic covariance between signal intensities in the same and related molecules measured by different techniques across cohorts of samples (Crockford et al., 2006). Other methods based on multiple kernel learning and OPLS-DA has been developed to offer an efficient tool for the fusion of multiblock Omics data obtained from multiple sources (Boccard and Rutledge, 2013).

4.3. Strategies to provide quantitative data from HRMS-based methods

Another limitation hindering the application of HRMS-based methods in epidemiological studies is the lack of efficient methods available to provide accurate quantitative data (which is a key parameters to test statistical associations with health outcomes). One of the main obstacle arises from the fact that in LC-ESI platforms (the most currently used platform for HRMS), different compounds ionize to a very different extent (and the ionization will also depend on the system used (e.g., ESI vs nanoESI), the mobile phase or the biological matrices), meaning that the analytical response (e.g. the abundance) can greatly vary for compounds present at the same concentrations (Kruve, 2020).

Kruve (Kruve, 2020) recently reviewed a number of different strategies developed to obtain quantitatively meaningful results without standard substances. It involves approaches using (1) peak areas directly (if analyses of several samples have been carried out using the same methodology) or in combination with statistical data treatment, e.g., (Plassmann et al., 2018); (2) isotope dilution and its alternatives, e.g., (Mashego et al., 2004); (3) radiolabelling, (4) structurally similar compounds for quantitation, e.g., (Dahal et al., 2011) and (5) quantitation based on the predicted ionization efficiencies, e.g., (Oss et al., 2010). According to Kruve (2020), the highest accuracy is possessed by the isotopically labelled standards and radiochromatography, but these strategies are application dependent. In the case of chemical exposomics, quantification based on peak areas or on the predicted ionization efficiencies would seem more appropriate.

Another strategy recently used to provide quantitative data for a large numbers of exogenous chemicals and metabolites relies on the use of multianalyte targeted metabolomics/exposomics analyses (González-Domínguez et al., 2020; Preindl et al., 2019). Hence, González-Domínguez et al. (2020) recently developed a methodology based on scheduled multiple reaction monitoring (sMRM) mode which enables the simultaneous quantitative investigation of more than 1000 metabolites and exogenously-derived chemicals in total. Although, these methodologies relying on targeted methods involve pre-established lists of metabolites and exogenous chemicals, they can be seen as complementary to discovery-based HRMS methods and could be used as a second step to validate the results obtained after the initial HRMS-based exploratory investigation.

4.4. Providing annotation guidelines to harmonise the reporting of results

Confidence levels from 1 to 5 based on information generated during the analytical process (*m/z*, retention time, isotopic pattern, MS/MS data) have already been proposed for reporting annotation by Schymanski et al. (2014). Some of these annotation levels could now be updated to include for instance the match with a predicted Rt or the presence of additional biotransformation products to better characterise MS1 annotation. The reporting of qualitative criteria should include all the relevant information and provide a confidence level based on these recommendations or the ones from other initiatives (e.g. Metabolomics Standard Initiative (Sumner et al., 2007)). Other criteria can be added (e.g., specific use, production volume, toxicity data…) to provide even more context to a new annotation. Other criteria such as the molecule identifier (e.g., InChI, SMILES) still need to be adopted and require additional work to provide a harmonised reporting template.

To help for the harmonisation and standardisation of data processing, one possibility will be to rely in the future on online workflows (e.g., XCMS online (Tautenhahn et al., 2012)), galaxy workflows (e.g. Workflow4Metabolomics (Giacomoni et al., 2015)) or Cloud Research Environment (e.g., PhenoMeNal (Peters et al., 2019)) which are promising

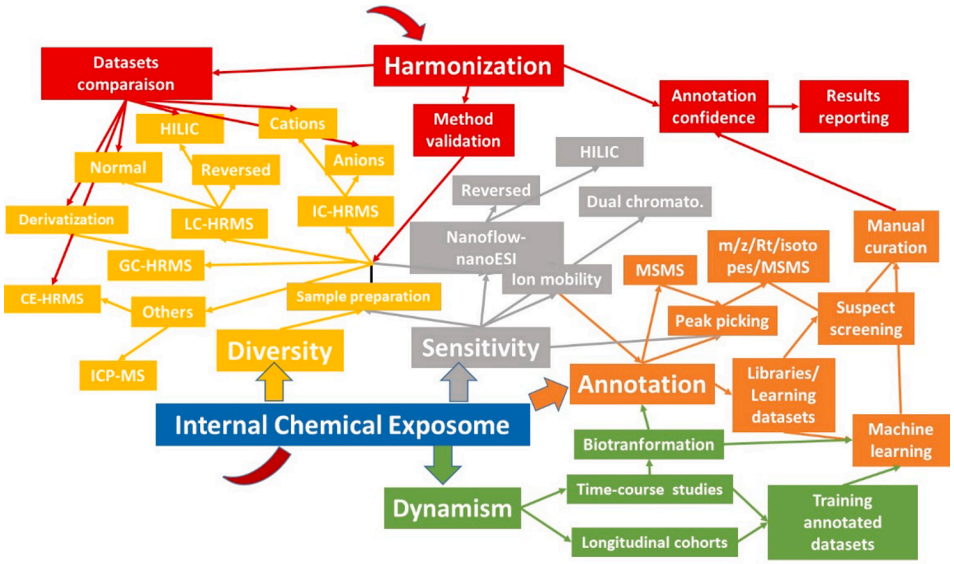

Fig. 5. Summarised mind mapping showing important lines of work (i.e., capturing the chemical diversity, improving analytical sensitivity, annotation and harmonisation issues and assessing the dynamic aspect of exposures) needed to improve the comprehensive characterisation of the internal chemical exposome and some of their connexions. These lines of work and the subtasks provided here are not exhaustive. Some items included in the sensitivity and diversity lines of work could be inverted since the coverage of the internal chemical exposome is dependent on the overall separation selectivity and the ability to detect low-abundant chemicals.

approaches, not only to standardize the process but also for the automation of some of these steps. It is also worth noting that bioinformatic workflows are now being developed specifically for non-targeted analyses (e.g., patRoon (Helmus et al., 2021)).

In the end, working on harmonisation issues is also necessary to generate a workflow for the registration and make raw data and metadata available and reusable under FAIR (findable, accessible, interoperable and reusable) principles. Enhancing the FAIRness of open resources can mutually enhance several resources for whole community benefit (Schymanski et al., 2020). Illustration of this includes the recent development of PubChemLite (Schymanski et al., 2020), a subset of the PubChem database that can be integrated into HRMS-based workflows. PubChemLite is using benchmarking datasets from earlier publications to show how experimental knowledge and existing datasets can be used to detect and fill all gaps in compound DB to progressively improve large resources such as PubChem (Schymanski et al., 2020).

5. Assessing the dynamic aspect of the internal chemical exposome

The initial foundation for characterisation of the internal chemical exposome is to identify comprehensively the types of agents (e.g., pesticides and biocides, personal care products, pharmaceuticals, food chemicals) humans are exposed to. It is also necessary to push towards characterising in more detail the magnitude, contact point/surface, and time period of exposure. From this point of view, it must be emphasised that not even qualitative experimental characterisations based on targeted methods are complete.

Vineis et al. (2020) recently suggested that one of the most urgent needs is the setting up of a new generation of cohort studies with improved and repeated biosample collection, improved questionnaire data, and the deployment of the enhanced exposure assessment methodologies. The collection of various biological samples on a very short time scale (e.g., several times a day for urine and once a week for blood) over a long period of time (months or up to one year) could be a strong asset to improve our knowledge of chemical toxicokinetics, including biotransformation processes (e.g., (David et al., 2021)), which would assist with annotation of complex HRMS datasets. These cohorts could be initially performed on a limited number of person to focus mainly on measuring more precisely the dynamics of exposure (Schüssler-Fiorenza Rose et al., 2019). The knowledge we would gain by characterising the chemical exposome of even few individuals over a limited period of time could then be translated to selected human biological samples from existing cohorts. Lifespan exposure modelling approaches are also emerging that will contribute to shift from punctual human biomonitoring measurement(s) to a more representative internal exposure estimate (Pruvost-Couvreur et al., 2020a,b).

6. Conclusion and perspectives

Several European (e.g., The Human Early-Life Exposome (HELIX) (Vrijheid et al., 2014), EXPOsOMICS (Vineis et al., 2017)), and international (e.g., HERCULES (Niedzwieck and Miller, 2019), NIH Human Health Exposure Analysis Resource (HHEAR) laboratory network) initiatives have already been deployed to perform large-scale exposomics studies (Huhn et al., 2021) and develop research infrastructures aiming to expand analytical capabilities, to provide support for data analysis, and to train future researchers in exposomics. Recently, the European Human Exposome Network, a cluster of nine new Horizon 2020 Exposome projects representing 106 M€ funding from the EU over 5 years, has been launched to improve technologies and methodologies for exposome research and to better understand how environmental factors affect our health. All these collaborative projects involving large international partnerships have already provided and will keep providing very important landmarks for the successful implementation of chemical exposomics based on HRMS (as well as other technologies -e.g., sensors- or advanced statistical models to treat big data) in epidemiological studies.

It is nevertheless important that we improve the versatility of HRMS chemical profiling methods and incorporate complementary techniques to explore the ‘invisible’; enhance their analytical sensitivity to uncover the ‘hidden’, and continue to develop automated and reliable annotation process to illuminate the ‘dark’ matter of the internal chemical exposome. Implementing harmonised quality control criteria will be crucial to provide confidence in the reliability and robustness of the data produced and adopting sample collection strategies which allow the assessment of the dynamic aspect of environmental exposures paramount. Importantly, all these lines of work are inter-connected and should be developed concomitantly to feed each other (see Fig. 5 for examples of connections between these lines of work). Setting up a distributed European Research Infrastructure supporting a strong collaboration and harmonisation of relevant laboratories/analytical facilities in and outside Europe will be necessary to develop this global view of all actions needed to decipher the internal chemical exposome.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: 'Dr. Miller receives royalties for his books The Exposome: A Primer and The Exposome: A New Paradigm for the Environment and Health. He has no other conflicts of interest. The other authors declare they have no actual or potential competing financial interests.'.

Acknowledgments

This article was supported by the MoU signed between Inserm and the Mailman School of Public Health of Columbia University on Nov. 12 2019. JC was funded by the Réseau Doctoral en Santé Publique. E.J.P. acknowledges support from the Czech Operational Programme Research, Development and Education – Project Postdoc@MUNI (No. CZ.02.2.69/ 0.0/0.0/16_027/0008360) and Czech Operational Programme Research, Development and Education – Project MSCA-fellow4@MUNI (No. CZ.02.2.69/0.0/0.0/20_079/0017045).

References

Aalizadeh, R., Nika, M.C., Thomaidis, N.S., 2019. Development and application of retention time prediction models in the suspect and non-target screening of emerging contaminants. J. Hazard. Mater. 363, 277–285.
Aksenov, A.A., Laponogov, I., Zhang, Z., Doran, S.L.F., Belluomo, I., Veselkov, D., et al., 2020. Auto-deconvolution and molecular networking of gas chromatography–mass spectrometry data. Nat. Biotechnol. 39, 169–173.
Alexovič, M., Horstkotte, B., Šrámková, I., Solich, P., Sabo, J., 2017. Automation of dispersive liquid–liquid microextraction and related techniques. Approaches based on flow, batch, flow-batch and in-syringe modes. TrAC, Trends Anal. Chem. 86, 39–55.
Annesley, T.M., 2003. Ion suppression in mass spectrometry. Clin. Chem. 49, 1041–1044.
Antignac, J.P., de Wasch, K., Monteau, F., De Brabander, H., Andre, F., Le Bizec, B., 2005. The ion suppression phenomenon in liquid chromatography–mass spectrometry and its consequences in the field of residue analysis. Anal. Chim. Acta 529, 129–136.
Bade, R., Bijlsma, L., Sancho, J.V., Hernández, F., 2015. Critical evaluation of a simple retention time predictor based on LogKow as a complementary tool in the identification of emerging contaminants in water. Talanta 139, 143–149.
Bar, N., Korem, T., Weissbrod, O., Zeevi, D., Rothschild, D., Leviatan, S., et al., 2020. A reference map of potential determinants for the human serum metabolome. Nature 588, 135–140.
Barbier Saint Hilaire, P., Rousseau, K., Seyer, A., Dechaumet, S., Damont, A., Junot, C., Fenaille, F., 2020. Comparative evaluation of data dependent and data independent acquisition workflows implemented on an orbitrap fusion for untargeted metabolomics. Metabolites 10, 158.
Barupal, D.K., Fiehn, O., 2019. Generating the blood exposome database using a comprehensive text mining and database fusion approach. Environ. Health Perspect. 127, 97008.
Bessonneau, V., Ings, J., McMaster, M., Smith, R., Bragg, L., Servos, M., Pawliszyn, J., 2017. In vivo microsampling to capture the elusive exposome. Sci. Rep. 7, 44038.
Boccard, J., Rutledge, D.N., 2013. A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion. Anal. Chim. Acta 769, 30–39.
Bodmer, W., Bonilla, C., 2008. Common and rare variants in multifactorial susceptibility to common diseases. Nat. Genet. 40, 695–701.
Bonini, P., Kind, T., Tsugawa, H., Barupal, D.K., Fiehn, O., 2020. Retip: Retention time prediction for compound annotation in untargeted metabolomics. Anal. Chem. 92, 7515–7522.
Bonvallot, N., David, A., Chalmel, F., Chevrier, C., Cordier, S., Cravedi, J.-P., Zalko, D., 2018. Metabolomics as a powerful tool to decipher the biological effects of environmental contaminants in humans. Curr. Opin. Toxicol. 8, 48–56.
Brack, W., 2003. Effect-directed analysis: a promising tool for the identification of organic toxicants in complex mixtures? Anal. Bioanal. Chem. 377, 397–407.
Caballero-Casero, N., Belova, L., Vervliet, P., Antignac, J.-P., Castaño, A., Debrauwer, L., et al., 2021. Towards harmonized criteria in Quality Assurance and Quality Control of suspect and non-target LC-HRMS analytical workflows for screening of emerging contaminants in human biomonitoring. TrAC, Trends Anal. Chem. 116201.
Chaker, J., Gilles, E., Léger, T., Jégou, B., David, A., 2021. From metabolomics to HRMS-based exposomics: Adapting peak picking and developing scoring for MS1 suspect screening. Anal. Chem. 93, 1792–1800.
Chao, A., Al-Ghoul, H., McEachran, A.D., Balabin, I., Transue, T., Cathey, T., et al., 2020. In silico MS/MS spectra for identifying unknowns: a critical examination using CFM-ID algorithms and ENTACT mixture samples. Anal. Bioanal. Chem. 412, 1303–1315.
Chekmeneva, E., Dos Santos Correia, G., Chan, Q., Wijeyesekera, A., Tin, A., Young, J.H., Elliott, P., Nicholson, J.K., Holmes, E., 2017. Optimization and application of direct infusion nanoelectrospray HRMS method for large-scale urinary metabolic phenotyping in molecular epidemiology. J. Proteome Res. 16, 1646–1658.
Chervet, J.P., Ursem, M., Salzmann, J.P., 1996. Instrumental requirements for nanoscale liquid chromatography. Anal. Chem. 68, 1507–1512.
Chetwynd, A.J., David, A., 2018. A review of nanoscale LC-ESI for metabolomics and its potential to enhance the metabolome coverage. Talanta 182, 380–390.
Chetwynd, A.J., David, A., Hill, E.M., Abdul-Sada, A., 2014. Evaluation of analytical performance and reliability of direct nanoLC-nanoESI-high resolution mass spectrometry for profiling the (xeno)metabolome. J. Mass Spectrom. 49, 1063–1069.
Consortium, I.H.G.S., 2004. Finishing the euchromatic sequence of the human genome. Nature 431, 931–945.
Crockford, D.J., Holmes, E., Lindon, J.C., Plumb, R.S., Zirah, S., Bruce, S.J., Rainville, P., Stumpf, C.L., Nicholson, J.K., 2006. Statistical Heterospectroscopy, an Approach to the Integrated Analysis of NMR and UPLC-MS Data Sets: Application in Metabonomic Toxicology Studies. Anal. Chem. 78, 363–371.
da Silva, R.R., Dorrestein, P.C., Quinn, R.A., 2015. Illuminating the dark matter in metabolomics. Proc. Natl. Acad. Sci. 112, 12549–12550.
Dahal, U.P., Jones, J.P., Davis, J.A., Rock, D.A., 2011. Small Molecule Quantification by Liquid Chromatography-Mass Spectrometry for Metabolites of Drugs and Drug Candidates. Drug Metab. Dispos. 39, 2355–2360.
David, A., Abdul-Sada, A., Lange, A., Tyler, C.R., Hill, E.M., 2014. A new approach for plasma (xeno)metabolomics based on solid-phase extraction and nanoflow liquid chromatography-nanoelectrospray ionisation mass spectrometry. J. Chromatogr. A 1365, 72–85.
David, A., Chaker, J., Léger, T., Al-Salhi, R., Dalgaard, M.D., Styrishave, B., et al., 2021. Acetaminophen metabolism revisited using non-targeted analyses: Implications for human biomonitoring. Environ. Int. 149, 106388.
David, A., Lange, A., Abdul-Sada, A., Tyler, C.R., Hill, E.M., 2017. Disruption of the Prostaglandin Metabolome and Characterization of the Pharmaceutical Exposome in Fish Exposed to Wastewater Treatment Works Effluent As Revealed by Nanoflow-Nanospray Mass Spectrometry-Based Metabolomics. Environ. Sci. Technol. 51, 616–624.
David, A., Rostkowski, P., 2020. Chapter 2 – Analytical techniques in metabolomics. In: Álvarez-Muñoz D., Farré M. (Eds.), Environmental Metabolomics. Elsevier.
Davies, V., Wandy, J., Weidt, S., van der Hooft, J.J.J., Miller, A., Daly, R., Rogers, S., 2021. Rapid Development of Improved Data-Dependent Acquisition Strategies. Anal. Chem. 93, 5676–5683.
Di Renzo, G.C., Conry, J.A., Blake, J., DeFrancesco, M.S., DeNicola, N., Martin Jr., J.N., et al., 2015. International Federation of Gynecology and Obstetrics opinion on reproductive health impacts of exposure to toxic environmental chemicals. Int. J. Gynaecol. Obstet. 131, 219–225.
Djoumbou-Feunang, Y., Fiamoncini, J., Gil-de-la-Fuente, A., Greiner, R., Manach, C., Wishart, D.S., 2019. BioTransformer: a comprehensive computational tool for small molecule metabolism prediction and metabolite identification. J. Cheminform. 11, 2.
Dou, M., Tsai, C.-F., Piehowski, P.D., Wang, Y., Fillmore, T.L., Zhao, R., et al., 2019. Automated Nanoflow Two-Dimensional Reversed-Phase Liquid Chromatography System Enables In-Depth Proteome and Phosphoproteome Profiling of Nanoscale Samples. Anal. Chem. 91, 9707–9715.
Drouin, N., van Mever, M., Zhang, W., Tobolkina, E., Ferre, S., Servais, A.-C., et al., 2020. Capillary Electrophoresis-Mass Spectrometry at Trial by Metabo-Ring: Effective Electrophoretic Mobility for Reproducible and Robust Compound Annotation. Anal. Chem. 92, 14103–14112.
Dührkop, K., Nothias, L.-F., Fleischauer, M., Reher, R., Ludwig, M., Hoffmann, M.A., et al., 2020. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. Nat. Biotechnol. 39, 462–471.
Dührkop, K., Shen, H., Meusel, M., Rousu, J., Böcker, S., 2015. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. Proc. Natl. Acad. Sci. U. S. A. 112, 12580–12585.
Fenaille, F., Barbier Saint-Hilaire, P., Rousseau, K., Junot, C., 2017. Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: Where do we stand? J. Chromatogr. A 1526, 1–12.
Funk, C.D., 2001. Prostaglandins and leukotrienes: advances in eicosanoid biology. Science 294, 1871–1875.
Gakidou, E., Afshin, A., Abajobir, A.A., Abate, K.H., Abbafati, C., Abbas, K.M., et al., 2017. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet 390, 1345–1422.
Gallart-Ayala, H., Teav, T., Ivanisevic, J., 2020. Metabolomics meets lipidomics: Assessing the small molecule component of metabolism. BioEssays 42, 2000052.
Gallidabino, M.D., Hamdan, L., Murphy, B., Barron, L.P., 2018. Suspect screening of halogenated carboxylic acids in drinking water using ion exchange chromatography - high resolution (Orbitrap) mass spectrometry (IC-HRMS). Talanta 178, 57–68.
Geller, S., Lieberman, H., Kloss, A., Ivanov, A.R., 2021. A systematic approach to development of analytical scale and microflow-based liquid chromatography coupled to mass spectrometry metabolomics methods to support drug discovery and development. J. Chromatogr. A 1642, 462047.
Giacomoni, F., Le Corguillé, G., Monsoor, M., Landi, M., Pericard, P., Pétéra, M., et al., 2015. Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. Bioinformatics 31, 1493–1495.
Gillet, L.C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., Aebersold, R., 2012. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis*. Mol. Cell. Proteomics 11 (O111), 016717.
Go, Y.M., Walker, D.I., Liang, Y., Uppal, K., Soltow, Q.A., Tran, V., et al., 2015. Reference Standardization for Mass Spectrometry and High-resolution Metabolomics Applications to Exposome Research. Toxicol. Sci. 148, 531–543.

González-Domínguez, R., Jáuregui, O., Queipo-Ortuño, M.I., Andrés-Lacueva, C., 2020. Characterization of the Human Exposome by a Comprehensive and Quantitative Large-Scale Multianalyte Metabolomics Platform. Anal. Chem. 92, 13767–13775.
Gray, N., Adesina-Georgiadis, K., Chekmeneva, E., Plumb, R.S., Wilson, I.D., Nicholson, J.K., 2016. Development of a Rapid Microbore Metabolic Profiling Ultraperformance Liquid Chromatography-Mass Spectrometry Approach for High-Throughput Phenotyping Studies. Anal. Chem. 88, 5742–5751.
Guijas, C., Montenegro-Burke, J.R., Domingo-Almenara, X., Palermo, A., Warth, B., Hermann, G., et al., 2018. METLIN: A Technology Platform for Identifying Knowns and Unknowns. Anal. Chem. 90, 3156–3164.
Guo, J., Huan, T., 2020. Evaluation of significant features discovered from different data acquisition modes in mass spectrometry-based untargeted metabolomics. Anal. Chim. Acta 1137, 37–46.
Haddad, N., Andrianou, X.D., Makris, K.C., 2019. A Scoping Review on the Characteristics of Human Exposome Studies. Curr. Pollut. Rep. 5, 378–393.
Harizi, H., Corcuff, J.B., Gualde, N., 2008. Arachidonic-acid-derived eicosanoids: roles in biology and immunopathology. Trends Mol. Med. 14, 461–469.
He, Y., Miggiels, P., Wouters, B., Drouin, N., Guled, F., Hankemeier, T., Lindenburg, P. W., 2021. A high-throughput, ultrafast, and online three-phase electro-extraction method for analysis of trace level pharmaceuticals. Anal. Chim. Acta 1149, 338204.
Helmus, R., ter Laak, T.L., van Wezel, A.P., de Voogt, P., Schymanski, E.L., 2021. patRoon: open source software platform for environmental mass spectrometry based non-target screening. J. Cheminf. 13, 1.
Hu, M., Müller, E., Schymanski, E.L., Ruttkies, C., Schulze, T., Brack, W., Krauss, M., 2018. Performance of combined fragmentation and retention prediction for the identification of organic micropollutants by LC-HRMS. Anal. Bioanal. Chem. 410, 1931–1941.
Huhn, S., Escher, B.I., Krauss, M., Scholz, S., Hackermüller, J., Altenburger, R., 2021. Unravelling the chemical exposome in cohort studies: routes explored and steps to become comprehensive. Environ. Sci. Eur. 33, 17.
Jia, S., Xu, T., Huan, T., Chong, M., Liu, M., Fang, W., Fang, M., 2019. Chemical Isotope Labeling Exposome (CIL-EXPOSOME): One High-Throughput Platform for Human Urinary Global Exposome Characterization. Environ. Sci. Technol. 53, 5445–5453.
Jones, D.P., 2016. Sequencing the exposome: A call to action. Toxicol. Rep. 3, 29–45.
Kanehisa, M., 2002. The KEGG database. Novartis Found Symp 247, 91–101.
Karas, M., Bahr, U., Dulcks, T., 2000. Nano-electrospray ionization mass spectrometry: addressing analytical problems beyond routine. Fresenius J. Anal. Chem. 366, 669–676.
Kind, T., Fiehn, O., 2006. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. BMC Bioinf. 7, 234.
Kruve, A., 2020. Strategies for Drawing Quantitative Conclusions from Nontargeted Liquid Chromatography–High-Resolution Mass Spectrometry Analysis. Anal. Chem. 92, 4691–4699.
Liu, Q., Walker, D., Uppal, K., Liu, Z., Ma, C., Tran, V., Li, S., Jones, D.P., Yu, T., 2020. Addressing the batch effect issue for LC/MS metabolomics data in data preprocessing. Sci. Rep. 10, 13856.
Marchant, C.A., Briggs, K.A., Long, A., 2008. In silico tools for sharing data and knowledge on toxicity and metabolism: derek for windows, meteor, and vitic. Toxicol. Mech. Methods 18, 177–187.
Marginean, I., Kelly, R.T., Prior, D.C., LaMarche, B.L., Tang, K., Smith, R.D., 2008. Analytical characterization of the electrospray ion source in the nanoflow regime. Anal. Chem. 80, 6573–6579.
Marginean, I., Tang, K., Smith, R.D., Kelly, R.T., 2014. Picoelectrospray ionization mass spectrometry using narrow-bore chemically etched emitters. J. Am. Soc. Mass Spectrom. 25, 30–36.
Mashego, M.R., Wu, L., Van Dam, J.C., Ras, C., Vinke, J.L., Van Winden, W.A., Van Gulik, W.M., Heijnen, J.J., 2004. MIRCLE: mass isotopomer ratio analysis of U-13C-labeled extracts. A new method for accurate quantification of changes in concentrations of intracellular metabolites. Biotechnol. Bioeng. 85, 620–628.
Meijer, J., Lamoree, M., Hamers, T., Antignac, J.-P., Hutinet, S., Debrauwer, L., et al., 2021. An annotation database for chemicals of emerging concern in exposome research. Environ. Int. 152, 106511.
Metz, T.O., Baker, E.S., Schymanski, E.L., Renslow, R.S., Thomas, D.G., Causon, T.J., et al., 2017. Integrating ion mobility spectrometry into mass spectrometry-based exposome measurements: what can it add and how far can it go? Bioanalysis 9, 81–98.
Michopoulos, F., Lai, L., Gika, H., Theodoridis, G., Wilson, I., 2009. UPLC-MS-based analysis of human plasma for metabonomics using solvent precipitation or solid phase extraction. J. Proteome Res. 8, 2114–2121.
Miggiels, P., Wouters, B., van Westen, G.J.P., Dubbelman, A.-C., Hankemeier, T., 2019. Novel technologies for metabolomics: More for less. TrAC, Trends Anal. Chem. 120, 115323.
Monteiro Bastos da Silva, J., Chaker, J., Martail, A., Costa Moreira, J., David, A., Le Bot, B., 2021. Improving exposure assessment using non-targeted and suspect screening: The ISO/IEC 17025: 2017 Quality standard as a guideline. J. Xenobiot. 11, 1–15.
Moschet, C., Piazzoli, A., Singer, H., Hollender, J., 2013. Alleviating the reference standard dilemma using a systematic exact mass suspect screening approach with liquid chromatography-high resolution mass spectrometry. Anal. Chem. 85, 10312–10320.
Nemkov, T., Hansen, K.C., D'Alessandro, A., 2017. A three-minute method for high-throughput quantitative metabolomics and quantitative tracing experiments of central carbon and nitrogen pathways. Rapid Commun. Mass Spectrom. 31, 663–673.
Niedzwieck, M.M., Miller, G.W., 2019. HERCULES: An Academic Center to Support Exposome Research. In: Dagnino, S., Macherone, A. (Eds.), Unraveling the Exposome. Springer, Cham. https://doiorg/101007/978-3-319-89321-1_13.
Niedzwiecki, M.M., Walker, D.I., Vermeulen, R., Chadeau-Hyam, M., Jones, D.P., Miller, G.W., 2019. The Exposome: Molecules to Populations. Annu. Rev. Pharmacol. Toxicol. 59, 107–127.
Oberacher, H., Sasse, M., Antignac, J.-P., Guitton, Y., Debrauwer, L., Jamin, E.L., et al., 2020. A European proposal for quality control and quality assurance of tandem mass spectral libraries. Environ. Sci. Eur. 32, 43.
Oedit, A., Ramautar, R., Hankemeier, T., Lindenburg, P.W., 2016. Electroextraction and electromembrane extraction: Advances in hyphenation to analytical techniques. Electrophoresis 37, 1170–1186.
Orešič, M., McGlinchey, A., Wheelock, C.E., Hyötyläinen, T., 2020. Metabolic Signatures of the Exposome—Quantifying the Impact of Exposure to Environmental Chemicals on Human Health. Metabolites 10, 454.
Oss, M., Kruve, A., Herodes, K., Leito, I., 2010. Electrospray Ionization Efficiency Scale of Organic Compounds. Anal. Chem. 82, 2865–2872.
Peters, K., Bradbury, J., Bergmann, S., Capuccini, M., Cascante, M., de Atauri, P., et al., 2019. PhenoMeNal: processing and analysis of metabolomics data in the cloud. GigaScience 8, giy149.
Plante, P.-L., Francovic-Fontaine, É., May, J.C., McLean, J.A., Baker, E.S., Laviolette, F., Marchand, M., Corbeil, J., 2019. Predicting Ion Mobility Collision Cross-Sections Using a Deep Neural Network: DeepCCS. Anal. Chem. 91, 5191–5199.
Plassmann, M.M., Fischer, S., Benskin, J.P., 2018. Nontarget Time Trend Screening in Human Blood. Environ. Sci. Technol. Lett. 5, 335–340.
Pourchet, M., Debrauwer, L., Klanova, J., Price, E.J., Covaci, A., Caballero-Casero, N., et al., 2020. Suspect and non-targeted screening of chemicals of emerging concern for human biomonitoring, environmental health studies and support to risk assessment: From promises to challenges and harmonisation issues. Environ. Int. 139, 105545.
Preindl, K., Braun, D., Aichinger, G., Sieri, S., Fang, M., Marko, D., Warth, B., 2019. A Generic Liquid Chromatography−Tandem Mass Spectrometry Exposome Method for the Determination of Xenoestrogens in Biological Matrices. Anal. Chem. 91, 11334–11342.
Price, E.J., Vitale, C.M., Miller, G.W., Barouki, R., Audouze, K., David, A., et al., 2021. Merging the exposome in an integrated framework for "omic". Sciences. https://doi.org/10.5281/zenodo.4475649.
Pruvost-Couvreur, M., Le Bizec, B., Béchaux, C., Rivière, G., 2020a. Dietary risk assessment methodology: how to deal with changes through life. Food Additiv. Contamin.: Part A 37, 705–722.
Pruvost-Couvreur, M., Le Bizec, B., Béchaux, C., Rivière, G., 2020b. A method to assess lifetime dietary risk: Example of cadmium exposure. Food Chem. Toxicol. 137, 111130.
Rappaport, S.M., 2011. Implications of the exposome for exposure science. J. Expo. Sci. Environ. Epidemiol. 21, 5–9.
Rappaport, S.M., Barupal, D.K., Wishart, D., Vineis, P., Scalbert, A., 2014. The blood exposome and its role in discovering causes of disease. Environ. Health Perspect. 122, 769–774.
Ridder, L., van der Hooft, J.J., Verhoeven, S., de Vos, R.C., van Schaik, R., Vervoort, J., 2012. Substructure-based annotation of high-resolution multistage MS(n) spectral trees. Rapid Commun. Mass Spectrom. 26, 2461–2471.
Ruttkies, C., Schymanski, E.L., Wolf, S., Hollender, J., Neumann, S., 2016. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. J. Cheminform. 8.
Schulze, B., Jeon, Y., Kaserzon, S., Heffernan, A.L., Dewapriya, P., O'Brien, J., et al., 2020. An assessment of quality assurance/quality control efforts in high resolution mass spectrometry non-target workflows for analysis of environmental samples. TrAC, Trends Anal. Chem. 133, 116063.
Schüssler-Fiorenza Rose, S.M., Contrepois, K., Moneghetti, K.J., Zhou, W., Mishra, T., Mataraso, S., et al., 2019. A longitudinal big data approach for precision health. Nat. Med. 25, 792–804.
Schymanski, E., Kondic, T., Neumann, S., Thiessen, P.A., Zhang, J., Bolton, E., 2020. Empowering Large Chemical Knowledge Bases for Exposomics: Pubchemlite Meets Metfrag. J. Cheminform. 13, 19.
Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ. Sci. Technol. 48, 2097–2098.
Soltow, Q.A., Strobel, F.H., Mansfield, K.G., Wachtman, L., Park, Y., Jones, D.P., 2013. High-performance metabolic profiling with dual chromatography-Fourier-transform mass spectrometry (DC-FTMS) for study of the exposome. Metabolomics 9, S132–S143.
Stanstrup, J., Neumann, S., Vrhovšek, U., 2015. PredRet: Prediction of Retention Time by Direct Mapping between Multiple Chromatographic Systems. Anal. Chem. 87, 9421–9428.
Stein, S., 2012. Mass spectral reference libraries: an ever-expanding resource for chemical identification. Anal. Chem. 84, 7274–7282.
Stingone, J.A., Buck Louis, G.M., Nakayama, S.F., Vermeulen, R.C., Kwok, R.K., Cui, Y., Balshaw, D.M., Teitelbaum, S.L., 2017. Toward Greater Implementation of the Exposome Research Paradigm within Environmental Epidemiology. Annu. Rev. Public Health 38, 315–327.
Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., et al., 2007. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics 3, 211–221.
Tautenhahn, R., Patti, G.J., Rinehart, D., Siuzdak, G., 2012. XCMS Online: a web-based platform to process untargeted metabolomic data. Anal. Chem. 84, 5035–5039.

Trufelli, H., Palma, P., Famiglini, G., Cappiello, A., 2011. An overview of matrix effects in liquid chromatography-mass spectrometry. Mass Spectrom. Rev. 30, 491–509.
Tulipani, S., Llorach, R., Urpi-Sarda, M., Andres-Lacueva, C., 2013. Comparative Analysis of Sample Preparation Methods To Handle the Complexity of the Blood Fluid Metabolome: When Less Is More. Anal. Chem. 85, 341–348.
Uppal, K., Walker, D.I., Liu, K., Li, S., Go, Y.M., Jones, D.P., 2016. Computational Metabolomics: A Framework for the Million Metabolome. Chem. Res. Toxicol. 29, 1956–1975.
Valvi, D., Walker, D.I., Inge, T., Bartell, S.M., Jenkins, T., Helmrath, M., et al., 2020. Environmental chemical burden in metabolic tissues and systemic biological pathways in adolescent bariatric surgery patients: A pilot untargeted metabolomic approach. Environ. Int. 143, 105957.
van der Kloet, F.M., Bobeldijk, I., Verheij, E.R., Jellema, R.H., 2009. Analytical Error Reduction Using Single Point Calibration for Accurate and Precise Metabolomic Phenotyping. J. Proteome Res. 8, 5132–5141.
Vermeulen, R., Schymanski, E.L., Barabasi, A.L., Miller, G.W., 2020. The exposome and health: Where chemistry meets biology. Science 367, 392–396.
Viant, M.R., Kurland, I.J., Jones, M.R., Dunn, W.B., 2017. How close are we to complete annotation of metabolomes? Curr. Opin. Chem. Biol. 36, 64–69.
Vineis, P., Chadeau-Hyam, M., Gmuender, H., Gulliver, J., Herceg, Z., Kleinjans, J., et al., 2017. The exposome in practice: Design of the EXPOsOMICS project. Int. J. Hyg. Environ. Health 220, 142–151.
Vineis, P., Robinson, O., Chadeau-Hyam, M., Dehghan, A., Mudway, I., Dagnino, S., 2020. What is new in the exposome? Environ. Int. 143, 105887.
Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., Yang, J., 2017. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am. J. Hum. Genet. 101, 5–22.
Vorkamp, K., Castaño, A., Antignac, J.-P., Boada, L.D., Cequier, E., Covaci, A., et al., 2021. Biomarkers, matrices and analytical methods targeting human exposure to chemicals selected for a European human biomonitoring initiative. Environ. Int. 146, 106082.
Vrijheid, M., Slama, R., Robinson, O., Chatzi, L., Coen, M., Hazel, P.v.d., et al., 2014. The Human Early-Life Exposome (HELIX): Project Rationale and Design. Environ. Health Perspect. 122, 535–544.
Vuckovic, D., 2012. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. Anal. Bioanal. Chem. 403, 1523–1548.
Watson, J., 1990. The human genome project: past, present, and future. Science 248, 44–49.
Wickremsinhe, E.R., Singh, G., Ackermann, B.L., Gillespie, T.A., Chaudhary, A.K., 2006. A review of nanoelectrospray ionization applications for drug metabolism and pharmacokinetics. Curr. Drug Metab. 7, 913–928.
Wild, C.P., 2005. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol. Biomarkers Prev. 14, 1847–1850.
Wild, C.P., 2012. The exposome: from concept to utility. Int. J. Epidemiol. 41, 24–32.
Williams, A.J., Grulke, C.M., Edwards, J., McEachran, A.D., Mansouri, K., Baker, N.C., et al., 2017. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. J. Cheminf. 9, 61.
Wilm, M., Mann, M., 1996. Analytical properties of the nanoelectrospray ion source. Anal. Chem. 68, 1–8.
Wishart, D.S., Feunang, Y.D., Marcu, A., Guo, A.C., Liang, K., Vazquez-Fresno, R., et al., 2018. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res. 46, D608–D617.
Wishart, D.S., Tzur, D., Knox, C., Eisner, R., Guo, A.C., Young, N., et al., 2007. HMDB: the Human Metabolome Database. Nucleic Acids Res. 35, D521–D526.
Wray, N.R., Goddard, M.E., Visscher, P.M., 2007. Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res. 17, 1520–1528.
Xin, H., Douglas, W., YongLiang, L., Matthew, S., Michael, O., Brian, J., et al., 2020. A scalable workflow for the human exposome. Res. Sq. https://doi.org/10.21203/rs.3.rs-83843/v1.
Xue, J., Guijas, C., Benton, H.P., Warth, B., Siuzdak, G., 2020. METLIN MS2 molecular standards database: a broad chemical and biological resource. Nat. Methods 17, 953–954.
Yang, Y.H., Cruickshank, C., Armstrong, M., Mahaffey, S., Reisdorph, R., Reisdorph, N., 2013. New sample preparation approach for mass spectrometry-based profiling of plasma results in improved coverage of metabolome. J. Chromatogr. A 1300, 217–226.
Zhang, W., Guled, F., Hankemeier, T., Ramautar, R., 2019. Utility of sheathless capillary electrophoresis-mass spectrometry for metabolic profiling of limited sample amounts. J. Chromatogr. B 1105, 10–14.
Zhang, W., Ramautar, R., 2021. CE-MS for metabolomics: Developments and applications in the period 2018–2020. Electrophoresis 42, 381–401.
Zhang, X., Dong, J., Raftery, D., 2020. Five Easy Metrics of Data Quality for LC-MS-Based Global Metabolomics. Anal. Chem. 92, 12925–12933.
Zhang, X., Romm, M., Zheng, X., Zink, E.M., Kim, Y.-M., Burnum-Johnson, K.E., et al., 2016. SPE-IMS-MS: An automated platform for sub-sixty second surveillance of endogenous metabolites and xenobiotics in biofluids. Clin. Mass Spectromet. 2, 1–10.
Zheng, X., Aly, N.A., Zhou, Y., Dupuis, K.T., Bilbao, A., Paurus, Vanessa L., et al., 2017. A structural examination and collision cross section database for over 500 metabolites and xenobiotics using drift tube ion mobility spectrometry. Chem. Sci. 8, 7724–7736.