Trends in Analytical Chemistry 158 (2023) 116825

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

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

Toward building mass spectrometry-based metabolomics and lipidomics atlases for biological and clinical research

Stanislava Rakusanova [a], Oliver Fiehn [b], Tomas Cajka [a, *]

[a] Institute of Physiology of the Czech Academy of Sciences, Videnska 1083, 14220, Prague, Czech Republic
[b] West Coast Metabolomics Center, University of California, Davis, 451 Health Sciences Drive, Davis, CA, 95616, United States

ARTICLE INFO

Article history:
Received 12 August 2022
Received in revised form
1 November 2022
Accepted 8 November 2022
Available online 17 November 2022

Keywords:
Metabolomics
Lipidomics
Analytical methods
Mass spectrometry
Atlas
Database

ABSTRACT

Current metabolomics and lipidomics studies are limited in the number of examined matrices, the breadth and scope of methods, reporting the number of metabolites, and data sharing. Here, we discuss the concept of metabolomics and lipidomics atlases that characterize the quantitative distribution and relationships of metabolites in biological matrices and serve as a resource for future studies. Combined sample extraction is recommended to screen the metabolome and lipidome comprehensively. A multiplatform mass spectrometry-based approach with methods for each fraction should follow to separate and detect metabolites differing in their physicochemical properties. Since many known metabolites are detected through untargeted analysis, routine use of multiple internal standards for quantification is advised. This approach provides semiquantitative data and delivers molar concentrations for selected polar metabolites and lipids. An interactive web tool to query metabolites, generate statistical models, visualize data, and download the results should be developed to access generated data easily.

1. Introduction

Metabolomics and lipidomics are part of omics family research primarily focused on identifying and quantifying small molecules (<2000 Da) present in the metabolome and lipidome [1,2]. Compared to other omics areas such as genomics, transcriptomics, and proteomics with thousands of targets, the analysis of the metabolome and lipidome has been left behind due to the absence of reliable analytical platforms and appropriate bioinformatics tools [3]. However, over the last decade, it has been shown that chemical profiling by mass spectrometry-based metabolomics and lipidomics is valuable in characterizing different biological specimens [4,5].

Several comprehensive bioinformatics resources have been developed to map the metabolome and lipidome. These resources can be divided into:

(1) databases covering structures, physicochemical properties, and metadata of various compounds (e.g., PubChem (pubchem.ncbi.nlm.nih.gov), LIPID MAPS (lipidmaps.org)),
(2) databases for automated metabolite annotation (e.g., LipidBlast [6], MassBank of North America (massbank.us)), and
(3) databases with concentrations of particular metabolites in different biological matrices (e.g., Human Metabolome Database (hmdb.ca)).

The last ones are important for two reasons: First, reporting molar concentrations of particular metabolites is crucial to enable direct comparisons of the results between laboratories and studies. Second, absolute metabolite quantities immediately distinguish major from minor species, allowing biological interpretations of the results in the context of other analytes [7–9].

Historically, the Human Serum Metabolome Database was the first of this kind. Based on analyses from different instrumental platforms, it provides information on approximately 4200 endogenous metabolites in human serum [3]. The Human Metabolome Database (HMDB), introduced in 2007 [10], contains over 220 000 compound entries (both endogenous and exogenous) with molar concentrations, usually in biofluids (blood, urine) [11]. This huge number of compounds has been extracted from scientific literature, books, government documents, and other databases.

However, there is still a lack of sufficient data on the metabolome and lipidome characterizing different biofluids, tissues, or cells which can be easily accessible and reused at any time for

* Corresponding author.
E-mail address: tomas.cajka@fgu.cas.cz (T. Cajka).

https://doi.org/10.1016/j.trac.2022.116825

Abbreviations:

AU	arbitrary units
BAT	brown adipose tissue
CCS	collision cross section
CE	capillary electrophoresis
conc.	concentration
DDA	data-dependent acquisition
DIA	data-independent acquisition
EDTA	ethylenediaminetetraacetic acid
EI	electron ionization
FDR	false discovery rate
FWHM	full width at half maximum
GC	gas chromatography
HDX-MS	hydrogen/deuterium exchange mass spectrometry
HFD	high-fat diet
HILIC	hydrophilic interaction chromatography
HRMS	high-resolution mass spectrometry
i.d.	internal diameter
IE	iterative exclusion
IEX	ion-exchange chromatography
IM-MS	ion mobility mass spectrometry
IMS	imaging mass spectrometry
LC	liquid chromatography
LLE	liquid–liquid extraction
LSI	Lipidomics Standards Initiative
MoNA	MassBank of North America
mPFC	medial prefrontal cortex
MRM	multiple-reaction monitoring
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MSI	Metabolomics Standards Initiative
MTBE	methyl *tert*-butyl ether
NIST	National Institute of Standards and Technology
NPLC	normal-phase liquid chromatography
OGTT	oral glucose tolerance test
OPLS-DA	orthogonal partial least squares discriminant analysis
PC	phosphatidylcholine
PCA	principal component analysis
PGC	porous graphitic carbon
PLS-DA	partial least squares discriminant analysis
PPT	protein precipitation
QC	quality control
QLIT	quadrupole/linear ion trap
QqQ	triple quadrupole
QSRR	quantitative structure retention relationship
RPLC	reversed-phase liquid chromatography
SCN	suprachiasmatic nucleus
SFC	supercritical fluid chromatography
SPE	solid-phase extraction
SRM	standard reference material
SWATH	sequential window acquisition of all theoretical fragment-ion spectra
TIC	total ion chromatogram
TOF	time-of-flight
UHPLC	ultrahigh-performance liquid chromatography
VJC	vacuum jacketed column
WAT	white adipose tissue
ZT	Zeitgeber time

future studies. Indeed, a “call to action” has recently been issued, challenging researchers to create metabolome atlases of compound levels in organs and cells to compare individual studies against animal models and human population health data [12]. Current metabolomics and lipidomics studies are limited in many ways, though:

(1) *Examined matrices.* Metabolomics and lipidomics studies usually examine a single matrix related to the study design and the hypothesis tested. Biofluids such as plasma and serum have been in-depth characterized, while data on metabolite composition of different tissues is not so frequently reported. More than 42 000 metabolites and chemicals have been reported as detected in blood and collated in the Blood Exposome Database (bloodexposome.org) through text mining [13].
(2) *Breadth and scope of methods.* Historically, studies have focused either on analyzing polar metabolites (metabolomics) or simple and complex lipids (lipidomics) in biological samples. Thus, only fractions of the metabolites were typically captured. In addition, a multiplatform approach needed to cover polar metabolites and lipids comprehensively is still not a common approach because it requires developing and validating different instrumental platforms and metabolite annotation strategies.
(3) *Reported metabolites.* Reporting metabolomics and lipidomics data can range from the bare minimum, such as discussing statistically significant metabolites in a research paper, to comprehensive reports containing all annotated metabolites (including unknowns), their metadata, and signal intensities or concentrations in all analyzed samples. Unfortunately, some authors still use metabolite names to annotate chemicals instead of using authoritative identifiers such as InChI keys. This habit severely impedes comparability across datasets.
(4) *Data sharing.* Depositing raw instrumental files is still not common practice compared to other omics areas, which limits the transparency and reproducibility of scientific research. Thus, retrospective data mining at a different place is impossible, and validation data for new software programs are hard to assess.

As a solution, comprehensive metabolomics and lipidomics atlases can be built for hypothesis-testing and validation and to present (semi)-quantitative data for metabolites detected and annotated in biological samples using the multiplatform approach. We take this opportunity and provide critical steps (Fig. 1) toward building mass spectrometry-based metabolomics and lipidomics atlases for biological and clinical research. Although this review mainly covers liquid chromatography–mass spectrometry (LC–MS)-based methods, other techniques, including gas chromatography, capillary electrophoresis, direct infusion, ion mobility MS, imaging MS, and nuclear magnetic resonance (NMR), are also highlighted. The essential recommendations are summarized in Table 1 and discussed further in the following sections.

2. Sample collection and preparation

2.1. Sample collection – what and how to collect?

The kind and amount of biological material to be collected, aliquoted, stored, and prepared for instrumental analysis is based on

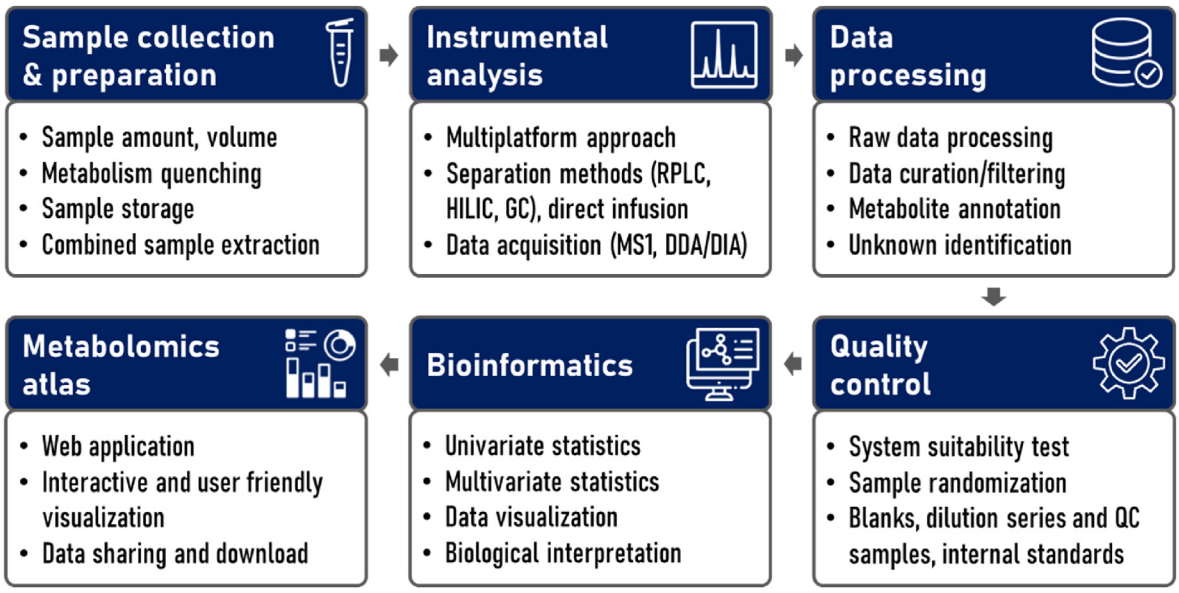

Fig. 1. Critical steps toward building mass spectrometry-based metabolomics and lipidomics atlases.

Table 1
Recommendations for building mass spectrometry-based metabolomics and lipidomics atlases for biological and clinical research.

Step	Recommendation
Sample collection and storage	• Define the section of each organ to be collected. • Use liquid nitrogen, dry ice, or freeze-clamping for quenching. • Keep consistent while using a particular anticoagulant (e.g., EDTA, citrate, heparin) for plasma. • Store samples at −80 °C or in liquid nitrogen.
Sample extraction	• Use combined all-in-one extraction methods to isolate a wide range of metabolites (from polar metabolites to nonpolar lipids) within a single extraction.
Sample separation	• Use multiple platforms to increase the number of annotated metabolites; lipidomics platforms (RPLC, HILIC, or direct infusion) may provide over 1000 unique lipid species, while fewer annotations are usually obtained for polar metabolites (RPLC, HILIC, GC, CE).
Method type	• Use a series of internal standards regardless the study is untargeted or targeted to control the workflow fully. • Add most of the internal standards during the extraction step to control the extraction step; use internal standards during the resuspension of dry extracts to monitor possible malfunction of the instrument.
Compound detection	• Acquire MS1 and MS/MS data for all samples simultaneously to increase the number of annotated metabolites (DDA is preferred to obtain clean MS/MS spectra). • Use iterative exclusion-MS to acquire MS/MS spectra for low abundant ions to increase the annotation rate.
Quality control	• Randomize the samples throughout the sequence with QC samples (pool of all samples) spread and regularly analyzed within the sequence. • To avoid detector saturation, use test injection (QC sample) to verify the injection volume and detector total ion chromatogram (TIC) response.
Data processing	• Use software that links MS1 data with MS/MS spectra for automated annotation. • "Trust, but verify" – use software that allows reviewing and curating annotated metabolites to remove false positives.
Metabolite annotation	• For data processing software, combine MS/MS spectral libraries from different sources to MSP format (e.g., NIST20, MoNA). • In-house MS/MS spectral libraries with acquired retention time improve annotation confidence.
Novel metabolite identification	• Ensure MS1 isotope pattern and MS/MS spectrum are obtained for *in-silico* software prediction. • Confirm novel metabolites by analyzing certified standards based on MS1, MS/MS, and retention time match.
Statistical analysis, data visualization	• Both univariate and multivariate analyses should be considered. • Web applications with interactive and user-friendly content visualization improve understanding of complex data sets.
Data sharing	• Provide instrumental files (blanks, QC, study samples) either in vendor format or open-access format (e.g., mzML, mzXML) via the data repository along with annotated metabolites and metadata.

the experimental design and tested hypothesis. In experiments with animals (e.g., mice and rats), trained personnel should be used to minimize the time between collecting or dissecting the organs. Proper training and personnel experience are paramount if multiple biofluids and tissues are collected for all animals. Liquid nitrogen, dry ice, or freeze-clamping should be used for quenching to stop the metabolism of samples as soon as possible, followed by storage of these samples at −80 °C [14]. For tissues, the section of each organ should be defined before its collection. Clinical samples usually involve collecting plasma or serum from the patients. The choice of a particular anticoagulant (e.g., EDTA, citrate, heparin) should be established beforehand and kept constant throughout the study [15]. A more detailed discussion regarding optimal pre-analytical handling of the most commonly used matrices, such as urine, blood, feces, tissues, and cells, can be found in recent review papers [14–19].

2.2. Sample extraction – how can it be streamlined to isolate a wide range of compounds?

The true breadth of a metabolome/lipidome cannot be captured by a single extraction method or instrumental platform because

biospecimens can contain metabolites spanning ~40 orders of magnitude on the predicted octanol/water partition coefficient scale and ranging from femtomolar to millimolar [4]. Such a concentration range exceeds the linear dynamic range of modern mass spectrometers by 10 000–100 000 fold [20]. Hence, in developing methods for metabolomics and lipidomic analyses, the main task is to cover metabolites using as few platforms as possible while maintaining the requisite precision, accuracy, and linear range for the metabolite classes detected by the chosen platforms.

Historically, sample preparation protocols for polar metabolites (aka metabolomics) and lipids (aka lipidomics) have been used separately. To this end, organic solvent-based protein precipitation or liquid–liquid extraction (LLE) methods are frequently used to extract a single sub-portion of the metabolome and to remove the bulk proteins and salts present in biological samples [21]. However, in the last few years, attempts have been made to isolate both subgroups of metabolites using an all-in-one single extraction method followed by fractionation and analyzing each fraction under different separation conditions [22,23]. Such an approach is helpful if the volume or amount of biological sample is limited, and different extraction methods for each platform would not be feasible. Table S1 provides examples of studies with combined extraction of polar metabolites and simple and complex lipids.

Single-step extraction using isopropanol or butanol/methanol (1:1) mixture has been evaluated and found to be best to isolate polar metabolites and lipids from plasma, followed by a separate analysis of each group of metabolites. Such a method benefits from high coverage and simplified operation, but ion suppression caused by co-eluting lipids was noted during polar metabolite analysis due to the high complexity of the extracts [24].

Reducing extract complexity is achieved during LLE when combined extraction of amphiphilic and lipophilic metabolites is conducted with solvents forming two phases: one mostly containing nonpolar metabolites (lipids) and the other mostly consisting of polar metabolites. These methods have been introduced with three solvent combinations: methyl *tert*-butyl ether (MTBE)/methanol/water [25], chloroform/methanol/water [26], and dichloromethane/methanol/water [27]. Overall, analyzing both phases under different instrumental conditions increases metabolome and lipidome coverage [4,5]. LLE has also been used for multi-omics studies to extract polar metabolites, lipids, and proteins [28,29].

Three-phase liquid extraction has been introduced using hexane, methyl acetate, acetonitrile, and water to reduce the sample complexity further. The upper organic phase is enriched in neutral lipids (triacylglycerols and cholesteryl esters), the middle organic phase contains the major glycerophospholipids, and the bottom phase contains polar metabolites a proteins [30]. Thus, analysis of polar metabolites is feasible after additional protein precipitation with an organic solvent.

Although this all-in-one extraction approach is highly interesting and progressively replaces separate extractions in metabolomics and lipidomics studies, there might be shortcomings to be explored more. For instance, acylcarnitines, composed of a polar carnitine head and a nonpolar carbon chain whose hydrophobicity increases with its chain length, require two LC–MS platforms. Specifically, acylcarnitines with carbon chain lengths from 2 to 10 partitioning to the polar phase are suited for hydrophilic interaction chromatography (HILIC). In contrast, acylcarnitines with carbon chain lengths above 12 are mainly extracted into the nonpolar phase of the MTBE/methanol extract and prefer reversed-phase liquid chromatography (RPLC) separation [31]. Furthermore, bi-phase extraction methods may become a bottleneck in large cohort studies because of the required manual operations. However, automated solvent-based protein precipitation using 96-well-plate formats recently became available [32,33] and promised a high-throughput, rapid, and cost-effective approach for the extraction of biofluids. Furthermore, these 96-well-plate formats are also available with various solid-phase extraction (SPE)-sorbents, which remove bulk phospholipids and isolate metabolites from biofluids in targeted methods [34].

Metabolomics and lipidomics studies largely rely on biofluids such as plasma, serum, and urine as a convenient specimen type for investigation, and literally, dozens of protocols are available for their analysis. However, biofluids comprise metabolites exported from different organs, complicating biochemical interpretation. Thus, metabolomics and lipidomics of tissues and cells should enable deeper insights into metabolism than biofluids [18]. Saoi et al. conducted a literature search and published reports comprising the liver, brain, heart, skeletal muscle, and kidney as the top five organs of mammals in pre-clinical animal studies [18]. Over the last few years, metabolomics and lipidomics of tissues have been rapidly expanding in clinical medicine and animal models. However, compared to protocols developed for biofluids, tissue samples require physical disruption using a homogenizer. First, the frozen tissue pieces (or lyophilized tissue powder) are placed into pre-cooled homogenization tubes containing ceramic beads. After the addition of cold extraction solvent, homogenization is performed [19]. During the method validation, the amount of tissue, the volumes of aliquots taken, the volumes of the resuspension solvent, and injection volumes should be optimized because different tissues may contain largely different concentrations of particular metabolites that need to be detected within the linear dynamic range of the detector. For adipose tissue, separate analyses are usually needed to cover high-abundant triacylglycerols and low-abundant complex lipids such as phospholipids. In the latter case, use of 80% methanol was found to resuspend low-abundant complex lipids such as phospholipids, diacylglycerols, fatty acids while keeping away the high-abundant triacylglycerols in the dry extract obtained during bi-phase extraction [35].

Hydro-organic mixtures are usually used for cell analysis (either as pellets or adherent cells) since they reduce water content, denature proteins, and provide an environment suitable for extracting and solubilizing a wide range of metabolites. Specifically, cold methanol/water (4:1) and acetonitrile/methanol/water (2:2:1) are commonly used [15]. However, to isolate highly nonpolar lipids such as triacylglycerols and cholesteryl esters, these solvent mixtures are insufficient for their extraction, and single or bi-phase extraction [25–27] with solvents with low polarity index (MTBE, chloroform, methanol) is needed.

2.3. Untargeted or targeted – or both?

In first-generation metabolomics (Metabolomics 1.0), untargeted and targeted methods are generally used separately, resulting in low sample throughput. Untargeted and targeted metabolomics methods can be characterized by data processing speed, the number of detected metabolites, and quantification reliability [4]. Here, reliability either refers to the accuracy of the absolute quantification, typically expressed in concentration units, or to the precision of semiquantitative assessments expressed in arbitrary (normalized) units.

In targeted metabolomics, structurally similar or identical (isotope-labeled) internal standards for compound quantification are used to achieve the best possible accuracy (Fig. 2). Untargeted metabolomics analyses can be accomplished without internal standards, but even here they tend to be used, albeit to check for method robustness (retention time shift, peak shape deterioration, and mass accuracy error) rather than for quantification [36,37]. Therefore, merging targeted and untargeted metabolomics

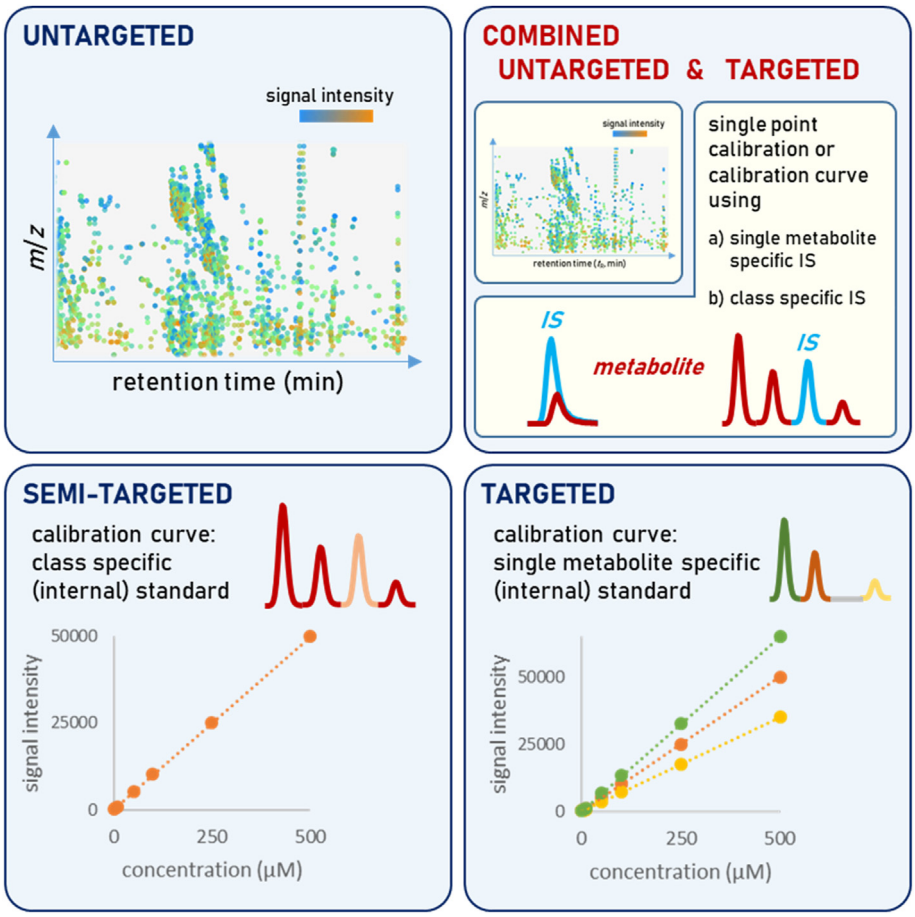

Fig. 2. Untargeted and targeted metabolomics and lipidomics methods.

workflows will require multipoint or one-point calibration using suitable internal standards added to samples during extraction, use of external calibration, or a combination of both (Fig. 2). Collectively, such methods open the gateway to second-generation metabolomics (Metabolomics 2.0) and can provide concentration results for particular metabolites and metabolites with similar physicochemical properties and untargeted signals [4]. In lipidomics, quantification strategies are usually based on adding at least one internal standard per lipid class followed by correction of type I isotopic effect for all lipid species of particular lipid class and type II overlap typically observed during direct infusion (shotgun) and HILIC methods [38]. Internal standards are usually added at appropriate concentrations during sample extraction for quantification. This can be achieved by adding a low volume of internal standard mixture to the sample prior to extraction, or internal standards can be present in the extraction solvents. When evaporation of an aliquot is needed, the resuspension solvent should also contain at least one internal standard to monitor possible instrument malfunctioning during the injection.

Recently, a few studies have shown the potential of combined untargeted and targeted analysis (Table 2). The combined methods can be conducted either within a single injection (e.g., MS1 for untargeted profiling and data-independent acquisition MS/MS for quantification of target metabolites [39]) or using separate platforms (e.g., RPLC for untargeted profiling and HILIC for targeted analysis of selected metabolites [40]). For instance, high-resolution mass spectrometry (HRMS) combined untargeted data collection with quantifying 55 targeted metabolites (amino acids, organic acids, purines, pyrimidines, and acylcarnitines) in human serum. This approach demonstrated improved quantification accuracy and precision for targeted metabolites and a high data acquisition speed that provided hypothesis-generating untargeted signals [39].

Several ring trials and multi-instrument comparisons [8,47–49] have shown that when a standardized approach is used, comparable results can be obtained for targeted metabolomics and lipidomics platforms. For instance, a recent ring trial with 14 laboratories provided quantitative data using the Biocrates AbsoluteIDQ p400HR kit with approximately 250 metabolites routinely quantified in plasma and serum samples. However, better performance (precision and accuracy) was observed for analytes measured using external calibration and LC with HRMS than those measured with stable-isotope dilution (single point) quantification by direct infusion-HRMS. Thus, multipoint calibration curves and chromatographic separation should be used to obtain the most accurate data when analytically and financially feasible [49].

3. Instrumental analysis

3.1. Sample separation – which chromatography to use?

Liquid chromatography–mass spectrometry (LC–MS) has become the most applied chromatography–MS tool for the analysis of both polar and nonpolar metabolites (~66% applications), followed by gas chromatography–mass spectrometry (GC–MS) used for the analysis of volatile metabolites and primary metabolites after derivatization (~30% applications) and capillary

Table 2
Examples of combined untargeted and targeted methods.

Matrix	Extraction	Untargeted platforms	Targeted platform, coverage (number of analytes)	Quantification	Ref.
Serum	Methanol	RPLC, metabolomics	RPLC, polar metabolites (55)	External standards	[39]
Plasma	Methanol	RPLC, metabolomics	HILIC, amino acids (12)	Internal standards	[40]
Plasma	Methanol, acetonitrile, acetone	RPLC, metabolomics	RPLC, amino acids and acylcarnitines (16)	Internal standards	[41]
Plasma	Methanol, MTBE, water	HILIC, metabolomics RPLC, lipidomics	HILIC, polar metabolites (59)	External and internal standards	[42]
Plasma	Methanol	HILIC, metabolomics	HILIC, amino acids (10)	External and internal standards	[43]
Plasma	Acetonitrile	RPLC × PGC, metabolomics	RPLC × PGC, polar metabolites (73)	External and internal standards	[44]
Cells	Methanol	IEX, metabolomics	IEX, polar metabolites (45)	External and internal standards	[45]
Plant	Methanol	RPLC, metabolomics	RPLC, secondary metabolites (12)	External standards	[46]

electrophoresis–mass spectrometry (CE–MS) for analysis of polar, charged metabolites (~4% applications) [7,50,51].

Coupling mass spectrometry with liquid chromatography provides several advantages, such as the separation and detection of isobars and isomers, reduced ion-suppression effects, and the possibility of separating compounds according to their physicochemical properties [52]. Luckily, various LC–MS modes, including different stationary phase chemistries and compositions of mobile-phase solvents and modifiers, are available to separate small molecules. In metabolomics, RPLC and HILIC are LC-based separation methods most frequently used [4]. Typically, RPLC is used to separate polar and medium-polar metabolites, and HILIC to separate highly-polar to medium-polar metabolites. In lipidomics, in addition to RPLC and HILIC, normal-phase LC (NPLC) and supercritical fluid chromatography (SFC) are also utilized for sample separation [53,54]. Through the fractionation of lipid classes (HILIC or NPLC) and follow-up analysis based on lipophilicity (RPLC), such complementary separation mechanisms enable increased lipidome coverage [55]. The separation time usually ranges between 10 and 30 min.

Concurrently, direct infusion methods are also available for complex lipid analysis [56] and a limited number of polar metabolites [57]. In addition, although overshadowed by MS in the number of annotated metabolites, nuclear magnetic resonance (NMR) spectroscopy offers highly reproducible and quantitative data over a wide dynamic range [58].

The compatibility of extraction solvents with initial conditions during LC–MS analysis must be considered, especially for bi-phase extraction protocols [25–27]. For instance, direct injection of MTBE/methanol or chloroform/methanol phase does not perform well for higher injection volumes causing deterioration of the peak shape of early-eluting lipids. Additionally, more attention should be paid during the data acquisition, and if reinjection is needed, this should be conducted immediately because evaporation of MTBE and chloroform occurs very quickly compared to other resuspension solvents (e.g., neat methanol) [4].

Analysis of the polar fraction is also challenging because it contains water-soluble proteins which were not precipitated using MTBE/methanol or chloroform/methanol mixtures [28]. Therefore, for the HILIC platform with the initial conditions favoring the high percentage of organic solvent, the aliquot of the polar fraction should be evaporated first. Then, a high percentage of organic solvent (e.g., acetonitrile/water, 4:1) in the resuspension solvent should be used, so the proteins are not dissolved anymore. On the other hand, for the RPLC platform with usually a high percentage of water at the beginning of the gradient, additional protein precipitation using an organic solvent is needed (e.g., the addition of acetonitrile/isopropanol mixture [59] or neat methanol [28] to polar extracts). After protein precipitation and evaporation of supernatant, metabolites are dissolvent in resuspension solvent with a high percentage of water (e.g., 5% methanol) [59].

The introduction, in 2004, of ultrahigh-performance LC (UHPLC) with sub-2 μm particles [60] led to significant improvements in chromatographic performance. Today, LC–MS-based metabolomics and lipidomics methods typically use short (50–150 mm) microbore columns (2.1 mm internal diameter, i.d.) with sub-2 μm particles [4]. However, using a 1 mm i.d. microbore column has been shown to provide similar or even better peak capacity and sensitivity than the frequently used 2.1 mm i.d. column. Even more interestingly, it seems that, while not requiring longer analysis time, the use of a 1 mm i.d. microbore column can lead to a 75% reduction in solvent consumption compared with the 2.1 mm i.d. variant [61]. However, some system modifications are necessary to reduce peak dispersion when using a 1 mm i.d. column with a conventional UHPLC system, such as reducing outlet tubing i.d., with a minimal length, and using a narrow bore stainless still spray capillary. Over the last few years, several LC systems suitable for micro- or even nano-flow have become available, enabling to work with 1 mm i.d. columns and low flow rates effectively. However, tubing with low internal diameters is prone to block, leading to high backpressure during analysis. In-line filters installed in the autosampler, at the entry to the trap column (optional), and at the entry to the analytical column help to minimize the problem. In addition, rigorous centrifugation of sample extracts and transfer of clean supernatant is critical [62].

For large-scale metabolomics and lipidomics studies, high-throughput LC–MS methods (<10 min) are attractive to many researchers because analysis of over a hundred samples per day is feasible. In conventional UHPLC systems, this can be achieved by modifications of conventional LC methods such as using short columns (e.g., 50 mm × 2.1 mm i.d.), increasing column flow rate and temperature, and modification of LC gradient. Since chromatographic peaks are usually narrower than in the original method, modification of MS settings is needed to obtain a sufficient number of data points per peak. Specifically, the number of data-dependent acquisition (DDA) scans is reduced during simultaneous MS1 and MS/MS acquisition. This can lead to a lower annotation rate, but automated iterative exclusion-MS may help to increase the total number of MS/MS scans for metabolite annotation [63]. High-throughput LC–MS methods may also suffer lower chromatographic resolution than conventional UHPLC methods. An attractive solution has been shown recently using vacuum jacketed columns (VJC) positioned at the source of the mass spectrometer. For a 5 min VJC lipidomics analysis of human plasma (Fig. 3A and B), the peak capacity increased by 66%, peak tailing reduced by up to 34%, and the number of lipids detected increased by 30% compared to conventional UHPLC [64].

Over the last few years, there has been an increased interest in novel technologies such as ion mobility mass spectrometry (IM-MS) for metabolomics and lipidomics applications. IM-MS allows collision cross-section (CCS) values to be measured for metabolites and lipids, increases the peak capacity and the signal-to-noise, and can be coupled to various fragmentation modes [65]. In addition,

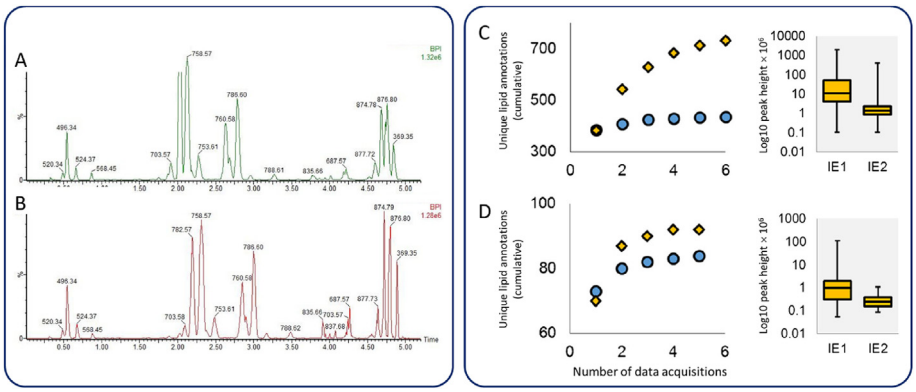

Fig. 3. (A + B) Comparison of chromatographic separation obtained from the LC−MS analysis in positive ion mode of the NIST SRM 1950 plasma. (A) Conventional method and (B) VJC method. For separation, an ACQUITY UPLC CSH C18 column was used (50 mm × 2.1 mm i.d., 1.7 μm particle size). Reproduced with permission from Ref. [64]. (C + D) Cumulative unique lipid annotations across multiple data acquisitions of plasma extract in (C) positive and (D) negative ion modes. Iterative exclusion (IE)-based DDA of top 5 precursor ions is compared with traditional DDA top 5 (without IE). Boxplots of log-transformed peak heights for unique lipid species annotated during (C) the first DDA top 5 acquisition (IE1) and (D) after applying an exclusion list using the IE algorithm (IE2) show that after applying IE in the second injection (IE2), peak heights of identified lipids were significantly lower than the initial injection. Reproduced (adapted) with permission from Ref. [63].

IM-MS can be combined with an ion source to perform direct infusion, coupled to online chromatography, or integrated into imaging mass spectrometry (IMS) [65].

During direct infusion IM-MS, molecules are separated according to their mobility times and *m*/*z* decreasing interferences before MS detection. For instance, a six-fold increase in peak capacity was shown for human blood metabolome when using direct infusion IM-MS compared to direct infusion alone [66]. Furthermore, IM-MS allows for overcoming the confounding factor of isobaric overlap of lipids during direct infusion analysis [67]. When IM-MS is coupled to liquid chromatography, improving the throughput of metabolome analysis using shorter columns and steeper gradients while increasing the peak capacity over conventional LC−MS has been observed. LC−IM-MS can provide CCS values as an orthogonal coordinate for compound identification in addition to retention time and *m*/*z* [68]. IM-MS is also helpful when integrated into imaging mass spectrometry (IMS) in spatial metabolomics and lipidomics. Compared to classical IMS, IM-IMS can separate discrete analyte classes and is often used to distinguish both isobaric and isomeric species [69].

3.2. Compound detection − how to acquire as much possible information from the sample?

Although various mass spectrometers can be used for the targeted analysis of metabolites, unit-resolution triple quadrupole (QqQ) and quadrupole/linear ion trap (QLIT) are the most frequently used for quantification in LC−MS. Both QqQ and QLIT are routinely operated in multiple-reaction monitoring (MRM) mode (1 precursor ion → 1−2 fragment ions) [1]. Conversely, high-resolution/accurate-mass mass spectrometers generally operate in full mass spectra acquisition mode in untargeted metabolomics and lipidomics. Such data acquisition enables the retrospective data mining of formerly unknown compounds, even years later. Specifically, single-stage instruments, such as time-of-flight (TOF) and orbital ion trap, and hybrid instruments, such as quadrupole/QTOF and quadrupole/orbital ion trap, dominate in untargeted metabolomics and lipidomics [1,53].

The latest high-resolution mass spectrometry-based instrumentation possesses several advantages, including fast acquisition speeds, better sensitivity, higher selectivity, and a wider linear dynamic range. Fast acquisition speeds enable MS1 data and multiple MS/MS spectra to be collected in very short cycle times (<1 s) [4,70]. With electronics and construction design improvements, these instruments can achieve very low detection limits. Current high-resolution instruments operate at a high mass resolving power of 10 000−450 000 full width at half maximum (FWHM); at such power, the more efficient spectral separation of isobaric compounds becomes feasible [71]. In addition, the current systems have a linear dynamic range of around four orders of magnitude, which enables the quantification of metabolites differing widely in concentration.

For MS/MS investigations, classic DDA fragmentation methods still lead the way in metabolomics and lipidomics. Typically, such an approach provides high-quality MS/MS spectra because the precursor ions are isolated using a narrow mass isolation window (0.7−4 Da). However, for complex mixtures, only high abundant precursor ions are selected because of the limited acquisition speed of a mass spectrometer. Therefore, critical rules should be applied to increase the number of acquired MS/MS spectra for metabolite annotation [72]. Among them, the optimal number of MS/MS scans per cycle, the proper threshold for selection of precursor ion selection, the use of an exclusion list, filters for precursor selection (e.g., isotope exclusion function, charge state), and performing DDA on all measured samples will ensure optimal performance. In addition, the use of an exclusion list with features (characterized by their *m*/*z* or *m*/*z* vs. retention time) that should never be selected for MS/MS can be based on blank analysis or analysis of actual samples with an automated iterative exclusion (IE), where selected precursors are excluded in sequential injections [63]. For example, for plasma extracts analyzed in positive and negative modes, applying IE and using six sequential injections increased the number of annotated lipid species by 69% and 18%, respectively, compared with the traditional DDA approach across six sequential injections (Fig. 3C and D) [63].

Concurrently, data-independent acquisition (DIA) has improved coverage for low-abundance precursor ions [4]. However, when DIA is used, it uses a much wider precursor ion isolation window. For example, sequential window acquisition of all theoretical fragment-ion spectra (SWATH-MS/MS acquisition) typically uses narrow 20−50 *m*/*z* windows. On the other hand, for all-fragment-ion approaches (MS^E, All Ions MS/MS, MS^{ALL}, or All Ion Fragmentation) wider windows, such as 1000 *m*/*z* or even more, are used. However, because precursor-ion selection uses a wider isolation window than DDA, both approaches can contaminate the MS/MS spectra. Thus, the acquired MS/MS spectra must be deconvoluted to

reduce such contamination [4]. To this end, software programs (e.g., MS-DIAL [6], DecoMetDIA [73], DecoID [74]) have been introduced for precursor-peak spotting, followed by MS/MS-level deconvolution. A SWATH-MS/MS workflow reported a 30% increase in correctly identified metabolites, reaching over 1000 identified molecules [75]. In DIA with an all-fragment-ion approach, much higher spectral complexity is observed compared to DIA-SWATH. A correlation-based deconvolution (CorrDec) method that uses ion abundance correlations has been developed to improve the metabolite annotation rate [76]. Furthermore, DIA supports compound annotation using MS/MS library search but also quantification at the MS/MS level because, unlike classic DDA, it offers continuous MS/MS spectra acquisition for particular fragment ions. For quantification, modified MRMPROBS software developed for targeted metabolomics analysis can be used [77].

Overall, these massive improvements in MS methods enable to merge untargeted and targeted metabolomics workflows to detect vastly increased numbers of compounds. Furthermore, the simultaneous acquisition of MS/MS spectra for both high and low-abundant metabolites promises to significantly increase the rate of annotated compounds based on MS/MS library searching. In addition, the acquired MS/MS spectra of unknowns can be used for structural elucidation and, once more comprehensive MS/MS libraries become available, for retrospective data processing [78].

4. Data processing and metabolite annotation

4.1. Data processing – how to get a peak list from instrumental files?

LC–MS data processing pipelines include multiple stages such as filtering, feature detection, peak alignment, and gap-filling. The pipeline should also conduct deisotoping (i.e., removing ^{13}C isotopes of particular metabolites), annotation of adducts (e.g., $[M+NH_4]^+$ and $[M+Na]^+$), and reporting potential in-source fragments based on correlation analysis. For GC–MS, additional spectral deconvolution of multiple fragments generated during electron ionization (EI) and creating a pure EI spectrum for each compound is also required. Based on the software, the data set might be further filtered using method blank, a dilution series prepared from quality control (QC) sample, QC samples, and frequency occurrence in all samples or within particular classes. The final reports comprise mass-to-charge (m/z) values, retention times, and corresponding intensities for all detected peaks [79]. Several data processing tools are currently available [4], and many focus on processing MS1 data only. However, advanced software programs are needed because of the growing popularity of simultaneous MS/MS data acquisition during LC–MS-based metabolomics and lipidomics experiments. Such software should enable the automated processing of MS/MS signals, signal alignment with corresponding precursor ions from the MS1 data, bridging with MS/MS library search, and the optional export of MS/MS spectra in a proper format to other software programs or applications.

For processing metabolomics and lipidomics instrumental files, many software programs are available, including (1) commercial vendor software, (2) commercial software from independent developers, (3) free open-access software, and (4) free open-access script platform [4]. The most popular free open-access software include XCMS [80], XCMS Online [81], MZmine 3 [82], and MS-DIAL 4 [6]. Although some of them were developed more than a decade ago, there is still a consistent effort to improve their performance, such as reducing computational time, tuning graphical user interfaces for users, or addressing new challenges such as processing IM-MS data [83].

Differences between data processing algorithms may significantly impact the processing of the acquired instrumental files. In the few studies that have compared the results produced by different software tools, large differences were observed, particularly regarding the number of false positive/negative peak detections and gap-filling capability [33,84–86]. Therefore, more in-depth comparisons of the various data pre-processing tools are needed. Such comparisons should emphasize identifying the parameters that most impact the number and quality of the reported metabolites. However, these comparisons are valid only for a short period because software updates are used to improve the overall data processing quality. Therefore, it is advised to monitor the Metabolomics Society Forum (metabolomics-forum.com), where issues and fixes for different data processing software programs are listed. We also advise creating validation data sets (instrumental files) for different platforms and comparing data processing outcomes for each new version of particular software.

4.2. Metabolite annotation – how to get the most annotations?

During LC–MS-based metabolomics and lipidomics experiments, MS/MS spectral library search is the fastest way to annotate acquired MS/MS spectra correctly. However, confidence in the MS/MS-based annotation of chemical structures can be impacted by instrumental settings, data acquisition modes, scoring algorithms, and post-curation steps [87–89].

Both commercial and public MS/MS libraries are available [90]. The most comprehensive commercial library METLIN Gen2 (metlin.scripps.edu) includes over 860 000 molecular standards with MS/MS data generated in both positive and negative ionization modes at multiple collision energies and collectively contains over 4 million MS/MS spectra [91]. NIST20 MS/MS (chemdata.nist.gov) covers 1.3 million MS/MS spectra belonging to 31 thousand authentic standards. Unfortunately, several obstacles stand in the way of more efficient metabolite compound identification: (i) the high cost of commercial libraries and (ii) the reduced identification speed when using METLIN Gen2. Since data cannot be downloaded from it (for instance, in MSP format), MS/MS searches must be submitted as single or batch queries via an online browser reducing the overall speed of data processing and metabolite compound identification. Free access library such as MoNA – MassBank of North America (massbank.us) includes over 2 million mass spectra, including 205 thousand experimental MS/MS spectra for diverse compounds (e.g., natural products and endogenous metabolites) and more than 1.3 million *in-silico* MS/MS spectra for lipids. Fewer MS/MS spectra are covered in other public databases that do enable downloads, such as MassBank (massbank.jp), GNPS (gnps.ucsd.edu), RIKEN PlaSMA (plasma.riken.jp), ReSpect (spectra.psc.riken.jp), or HMDB (hmdb.ca). mzCloud (mzcloud.org) also permits MS/MS search as a single query using a web-based interface.

For processing metabolomics data with software such as MS-DIAL 4 [6] and MZmine 3 [82], a combined MS/MS library in MSP format can be generated from multiple sources (e.g., NIST20, MoNA, MS-DIAL curated library, in-house library) and used during a spectral library search. Furthermore, for annotation of simple and complex lipids, a computer-generated (*in-silico*) MS/MS database LipidBlast is included in MS-DIAL 4 [6] or available in MSP format (MoNA, massbank.us) for other data processing software.

During metabolite or lipid annotation, peak misidentification may occur. First, some metabolites can be present in the extracts as structurally similar isomers (e.g., hexose phosphates, citrate/isocitrate, leucine/isoleucine, valine/betaine), providing similar MS/MS spectra. Therefore, LC separation and knowledge about their particular retention times are needed to annotate such metabolites correctly [14]. In RPLC, the lipid species are separated according to the number of double bonds and the length of fatty acyl chains

(carbon number). The higher carbon number in fatty acyl chains corresponds to increased retention times, while an additional double bond has the opposite effect and leads to faster elution. Therefore, this retention behavior of lipids can be used to improve annotation or remove false positives [92]. Second, the formation of in-source degradation products is quite common during LC−MS analysis. For instance, glutamine and glutamate generate an in-source artifact by cyclization to pyroglutamate [93], and citrate can provide in-source fragments annotated as aconitate or itaconic acid [94]. In addition, phosphatidic acid artifacts may result from in-source fragmentation of phosphatidylserine counterparts, and dimethyl phosphatidylethanolamine artifacts are yielded from in-source dissociation of anion adducts of phosphatidylcholine counterparts [95]. Third, isomeric bis(monoacylglycero)phosphate (BMP) and phosphatidylglycerol (PG) species are another example where misidentification may occur. Both lipid classes can be detected in positive (as $[M+NH_4]^+$) and negative (as $[M-H]^-$) ion modes using LC−MS. Negative mode is usually more sensitive and thus preferred for detecting PG species. However, only the positive mode provides MS/MS spectra that reliably differentiate these two lipid classes, while the negative mode provides identical MS/MS spectra for both [6]. Thus, positive ion mode should be preferred for studies where BMP and PG are expected to occur. Alternatively, derivatization for LC−MS [96] and direct infusion methods [97] using trimethylsilyldiazomethane were described.

As a general guideline for annotation, Metabolomics Standards Initiative (MSI) has recommended four confidence levels ranging from unknown (Level 4) to putatively characterized compound classes (Level 3) to putatively annotated compounds (Level 2) to identified metabolites (Level 1) [98]. Level 2 can be achieved when an unknown spectrum (Level 4) matches MS1 spectrum and MS/MS from the spectral library. One should also differentiate “annotation” (Level 2−4 confidence) and “identification” (Level 1 confidence). In the latter case, matching an unknown compound's retention time, MS1 and MS/MS spectra with an authentic standard is required [99]. In 2014, Schymanski et al. proposed criteria for using LC−MS/MS-based metabolomics and exposomics, specifically in high-resolution mode [100]. In addition, the standard of metabolite reporting is currently under revision within the metabolomics community introducing six subclasses (A−F) for unambiguous metabolite identification, such as *cis*/*trans* configuration [5].

For lipid annotation, hierarchical concept and shorthand notation should be applied [101] based on the recent Lipidomics Standards Initiative (LSI) guidelines. A proper lipid annotation results from different degrees of structural resolution from MS/MS analysis. During typical lipidomic profiling, a structural resolution such as carbons and double bonds (e.g., PC 34:2) and fatty acyl constituents (e.g., PC 16:0_18:2) can be obtained. More in-depth characterization, such as positional isomers (e.g., PC 16:0/18:2), double bond position (e.g., PC 16:0/18:2(10,12), double bond *cis*/*trans* (e.g., PC 16:0/18/2(10*E*,12*Z*)), or stereochemistry (e.g., PC 16:0/18/2(10*E*,12*Z*)[*R*]) cannot be provided by traditional LC−MS/MS platforms [102].

IM-MS can also be considered for compound annotation since it provides a valuable physicochemical property, namely, CCS, which is orthogonal to retention time, accurate mass, and MS/MS spectra [65]. Therefore, direct matching accurate masses and CCS values of precursor ions with a spectral database would provide Level 2 confidence for metabolites detected using LC−MS. In addition, IM-MS CCS databases have recently become available for polar metabolites [103] and lipids [6,104] for confident metabolite annotation. However, a recent paper by Köfeler et al. [105] pointed out that reported lipidomics data with CCS values in a study conducted by Vasilopoulou et al. [106] would require additional experimental evidence for 46% of features reported as unique lipid species. Thus, quality control requirements should be implemented for interpreting and reporting IM-MS metabolomics and lipidomics data as for other platforms.

4.3. Improving metabolite identification − are we there yet?

LC−MS enables the reproducible detection of thousands of molecular features in biological samples within a single analysis [107]. However, it is estimated that only 20% of these features can be annotated based on mass spectra library matches; the other 80% of signals detected in untargeted metabolomics are unknown metabolites or exposome compounds, the “dark matter” of the metabolome [108]. Herein lies the most challenging and exciting aspect of the metabolome: the huge number of compounds yet to be identified. The origin and nature of these unknown chemicals are of major interest not only because some will be correlated with human disease but also because some may provide previously unimagined health benefits [109]. However, the missing information about these compounds severely hinders meaningful biochemical and pathway interpretation.

The simplest way to increase identification rates is by collecting information from the mass spectrometric analysis of certified standards and making that information publicly available in mass spectral repositories [110]. However, measuring the MS/MS spectra of all authentic metabolites and other exogenous compounds is unrealistic. Thus, the only feasible strategy is to simulate the fragmentation process computationally under low-energy collision-induced dissociation to obtain theoretical spectra of these input structures [111]. Furthermore, the combination of *in-silico* fragmentation software and MS library searching facilitates the identification of unknown compounds and increases the coverage of metabolites in spectral libraries [111].

For the identification of unknown metabolites and the calculation of their chemical structures, several *in-silico* fragmentation software programs already exist: Mass Frontier, CSI:FingerID, CFM-ID, MS-FINDER, AGMa+, MIDAS, and MetFrag [111]. Generally, these tools translate masses into molecular fragments through combinatorial structure generation. Some generate theoretical fragmentation spectra from a structure (e.g., CFM-ID, MS-FINDER, MetFrag); others compute a fingerprint from the experimental fragmentation spectrum and then match it against a structural database (CSI:-FingerID). In addition, a hybrid search combing conventional, direct peak matching with the logical equivalent of neutral-loss matching is valuable to find similar compounds with a structural difference confined to a single molecule region that does not substantially alter its fragmentation behavior [112].

Fig. 4 shows an example of using *in-silico* software MS-FINDER to identify *N*6,*N*6,*N*6-trimethyl-L-lysine in human plasma. Initially reported as an unknown signal, using MS1 isotopic pattern and MS/MS spectrum (Fig. 4A), the software provided a list of candidates (Fig. 4B), and the top ranking one was confirmed by analysis of authentic standard (Fig. 4C).

These tools have been proven to perform well and will play a crucial role in metabolite and exposome identification pipelines in the coming years [113]. Indeed, in the 2016 CASMI (Critical Assessment of Small Molecules Identification, casmi-contest.org) contest, *in-silico* fragmentation software programs returned the correct answer for positive ionization data in 74% of cases [113]. However, such a high compound identification rate is achieved because various internal compound databases and MS/MS spectra greatly boost compound identification beyond the results of the *in-silico* outputs alone. This means that these tools are currently helpful in identifying “known unknowns,” namely, compounds unknown to the investigator but known in the chemical literature, reference databases, or internet resources [114]. For reliable

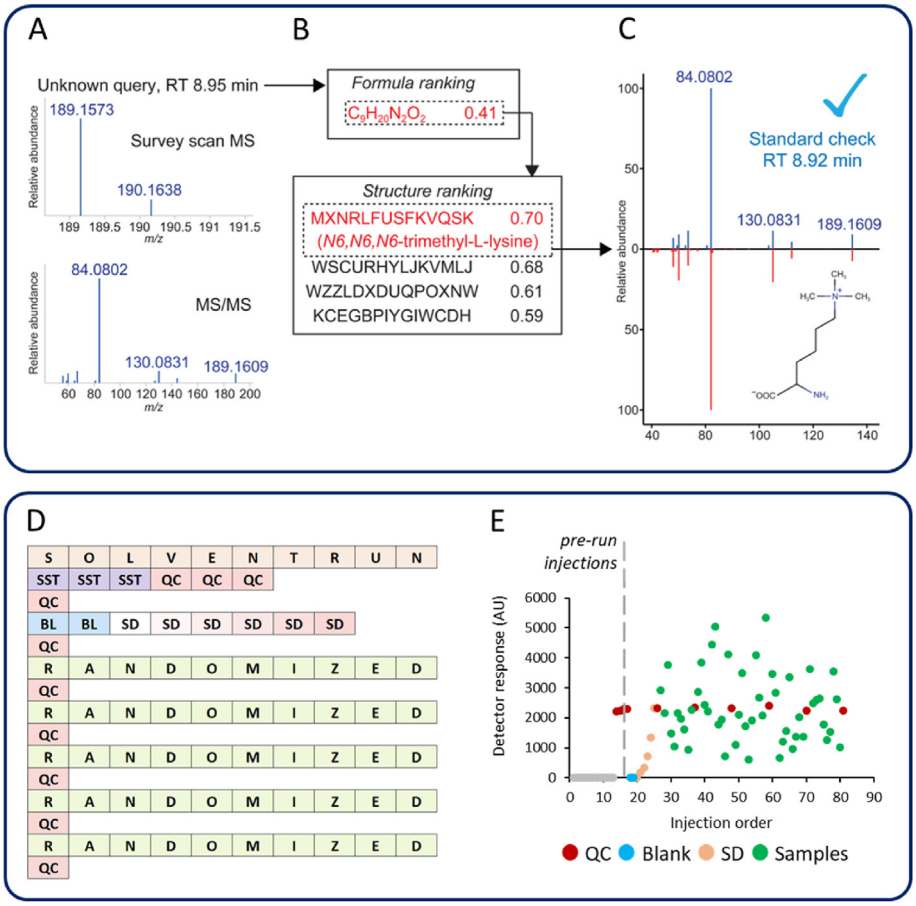

Fig. 4. (A–C) Identification of *N*6,*N*6,*N*6-trimethyl-L-lysine in human plasma with support of MS-FINDER software: (A) precursor MS1 and fragment ion MS/MS spectra, (B) formula prediction and structure search results for these spectra, sorted by the MS-FINDER score, and (C) the experimental MS/MS spectra from human plasma and purchased authentic standards are shown in blue and red, respectively. Reproduced (adapted) with permission from Ref. [111]. (D + E) Quality control in metabolomics and lipidomics studies: (D) a schematic example of a typical sequence for untargeted metabolomics and lipidomics, (E) an example of detector response for metabolite representative over the sequence (SST, system suitability test; QC, quality control sample; BL, blank; SD, a dilution series of QC sample).

identification of "unknown unknowns," much effort is still needed, though. In particular, a better understanding of the fragmentation mechanisms of small molecules is required if computational methods are to be improved to the point at which they become everyday tools for compound identification.

Additional orthogonal information can be crucial to determining the correct structure. For example, a frequently used technique in proteomics and structural elucidation of pharmaceutical impurities, yet not fully explored in metabolomics, represents hydrogen/deuterium exchange mass spectrometry (HDX-MS) [115,116]. Exchangeable hydrogen atoms (also called active, acidic, or labile) bound to heteroatoms such as oxygen, nitrogen, and sulfur are readily exchangeable with deuterium, whereas those bound to carbon are not exchanged. Thus, when metabolites are exposed to D_2O, labile hydrogens are replaced by deuterium, and their number can be obtained by determining the molecular mass before and after HDX using a mass spectrometer. Such information can be used as a useful filter to keep only a reduced number of potential candidates for structure elucidation [117]. For example, HDX-MS has been shown to identify *N*,*N*,*N*-trimethyl-L-alanyl-L-proline betaine [118] and (2R,3R)-2,3-dihydroxy-5-methylthio-4-pentenoic acid [119] in human plasma, and N-methyl lysophosphatidylethanolamines as abundant lipids in acidophilic mixed microbial communities [116].

Another orthogonal filter can be based on retention time prediction by quantitative structure retention relationship (QSRR) modeling, which relates the physicochemical properties of metabolites with their retention times under specific chromatographic conditions [120,121]. For instance, Bonini et al. used publicly available HILIC and RPLC data sets for model training in machine learning. The Keras algorithm outperformed other machine learning algorithms and yielded a mean absolute error of 0.78 min for HILIC and 0.57 min for RPLC. Furthermore, using mouse plasma samples as a test application, a 68% reduction in the number of candidate structures was observed when searching all isomers in MS-FINDER compound identification software [121].

5. Quality control

When performing metabolomics or lipidomics analysis, changes in instrument sensitivity may be observed over time. Typically, these are caused by retention-time shifts, degradation of the extracts, or contamination of the column and ion source by non-volatiles. To ensure that these factors do not confound metabolite signal intensity, samples must be analyzed in a random order, with sample classes randomly spread out across the whole run. Instrumental drift can be followed adequately over time by adding internal standards to samples and using quality control (QC) samples with a composition similar to the actual study samples [53,122]. In addition, these QC markers can be used for post-processing

normalization approaches. Such approaches include (i) the use of QC samples to fit a smoothed model for the intensity levels of certain features followed by the correction of all biological samples or (ii) the use of internal standards for normalizing particular metabolites and classes of metabolites. Several programs have been developed to handle unwanted variations in metabolomics and lipidomics data [83,123–125]. However, no consensus has yet been reached about which normalization strategy is the best.

Fig. 4D shows a schematic example of a typical sequence for untargeted metabolomics and lipidomics and the detector response for metabolite representative over the sequence (Fig. 4E). The initial injections of solvents (e.g., resuspension solvent used for a particular platform) are performed to check the baseline. If no contamination from the mobile phase is visible, the system suitability test (SST) is performed. For SST, either a simple mixture of compounds or routinely used extracts of some biological material can be used. SST is performed to check the consistency of selected compounds' retention time, peak intensity, shape, and mass accuracy. In the next step, the equilibration of analytical platforms is conducted by multiple injections of QC (study) sample. QC samples are prepared as pooled aliquots from all study samples or their representatives. These QC samples are injected in regular order, typically every 10–20 biological samples, depending on the analysis time and the number of biological samples. Method blanks and a diluted series of QC samples (e.g., 0, 1/16, 1/8, 1/4, 1/2, 1) help filter processed data [126] based on the presence of metabolite/features in blank samples and remove compounds with non-linear behavior based the coefficient of determination, respectively.

Suppose QC samples are used for instrumental drift correction (e.g., local polynomial regression, LOESS [122], systematic error removal using random forest, SERRF [127]), only those metabolites/features detected in all QC samples will be used. Thus, a diluting effect should be considered for QC samples prepared from diverse samples (or the same matrix but with different phenotypes), leading to lower signal intensity or even vanishing some metabolites/features during detection. To address this issue, phenotypic QC samples from each experimental group might be used to highlight differences in signal intensities between particular groups [128]. QC samples are also valuable for modifying injection volume before initiation of the sequence to avoid detector saturation.

The use of NIST Standard Reference Material (SRM) 1950 – Metabolites in Frozen Human Plasma, a commercially available reference material, is highly recommended for human studies as an additional QC measure, specifically during plasma or serum analysis. This SRM has also been frequently used during ring trials [47,49,129,130] and has been qualitatively and quantitatively well characterized, including reference values for over 300 unique lipid species [130] and polar metabolites (e.g., glucose, amino acids, biogenic amines) [49]. Overview of well-characterized mixtures of authentic standards of complex and less well-characterized biosamples currently available for use as QC tools to help ensure metabolomic and lipidomic research has been published recently [131].

6. Bioinformatics

6.1. Data normalization – how to deal with unwanted variation?

Before statistical analysis, data normalization strategies should be considered for both untargeted and targeted methods. The first strategies are sample-based and include intrinsic factors (e.g., ploidy/DNA/RNA content, total thymine released by acid hydrolysis, total protein concentration, selected metabolite concentration), cell-based approaches (e.g., count, volume), weight-based (e.g., fresh weight, dry weight), internal standards (e.g., isotopologues, stable-isotope labeled cells/extracts), QC-based (pooled QCs, SRM-based QC) or others (e.g., specific gravity, osmolality) [132].

No sample-based normalization is usually needed for frequently used biofluids such as plasma and serum. On the other hand, urine exhibits significant variability in concentration and composition, with normalization to creatinine or osmolality values as the most adopted strategies [133]. For cell experiments, fluorescence-based methods for DNA quantification showed an excellent linear correlation between fluorescence intensities and cell numbers [134]. Although tissue mass (wet or dry), total protein concentration, or DNA content have been used to recalculate the actual levels of metabolites in samples, a recent study by Cuevas-Delgado et al. emphasized that data normalization cannot be treated in an automated manner and should be carried out according to the biological model [135]. Further, Modrzejewska et al. showed an example of normalizing L- and D-2-hydroxyglutarate levels in tissues as a potential error source when using tissue mass instead of total thymine [136].

The second strategies are data-based normalizations and include statistical methods (e.g., log, median, average), analytical platform-based methods (e.g., total ion chromatogram, TIC), or batch effect corrections (e.g., LOESS, ComBat). For example, TIC with all detected features or mTIC with only genuine metabolites for normalization has been used in many metabolomics and lipidomics studies. Furthermore, since correlations between TIC and cell numbers have been shown for cells [137], this approach is helpful if sample-based normalization is not possible (e.g., no information on cell numbers). In general, TIC/mTIC approach is based on the assumption of the self-averaging property. That is, the increase in the concentration of one group of metabolites is balanced by the decrease in the concentration of another group of metabolites in each sample. However, this assumption may not always be valid in metabolomics and lipidomics because a specific systematic error may affect some metabolites differently than others [127].

For data normalization, several software tools and resources are available [132]. Sample-based normalization methods generally allow for better data control, while data-based approaches are usually much easier to implement. Minimum reporting criteria during data normalization should be reported to avoid incorrect data pretreatment and misleading biological interpretation [135].

6.2. Statistical analysis, data visualization, and interpretation – how to get the most from the data set?

Generally, the identification of key features is the first step toward finding useful biomarkers or understanding the biological processes involved in the condition under investigation [138]. Various approaches have been developed for these tasks, most of which are based on classical univariate statistical methods (e.g., *t*-tests, analysis of variance). To avoid too many false positives, the *p*-values are typically corrected using either the rather conservative Bonferroni correction [139] or the Benjamini–Hochberg false discovery rate (FDR) control procedure [140]. However, such univariate approaches are often considered to be suboptimal because they ignore the correlations known to be present among the variables (i.e., peaks or metabolites) [141]. Multivariate methods simultaneously consider all variables and are generally more suitable for high-dimensional omics data analysis. They come in unsupervised and supervised versions: the most frequently used unsupervised multivariate methods are principal component analysis (PCA) and hierarchical clustering with heat map; the most frequently used supervised versions are partial least squares discriminant analysis (PLS-DA) and orthogonal partial least squares discriminant analysis (OPLS-DA)] [142]. While many commercial software programs and free R-scripts are available for univariate and multivariate methods,

MetaboAnalyst (metaboanalyst.ca) is currently the most popular application for metabolomics data analysis and interpretation. This comprehensive web server can handle most of the common metabolomics data types, platforms, and experiments.

Metabolomics experiments produce complex multi-dimensional biological, chemical, and analytical information. Metabolic network mapping is widely used to integrate the results of metabolomics experiments with biological domain knowledge. In addition to measuring compounds with identified biological roles, many measurements may also contain mass spectral or empirical information. Thus, to fully understand complex metabolomics data, it is necessary to integrate information about biochemical transformations, structural similarity, mass spectral similarity, and empirical correlations [143]. Various software packages (e.g., MetaMapR [143], MetabNet [144], GNPS [145], MS2LDA [146]) are available to achieve this integration. The results of such integration analysis can be presented as interactive visualizations or exported as high-quality graphics and numerical tables for import into common network analysis and visualization tools (e.g., Cytoscape [147]).

For metabolic pathway mapping, several tools have become available recently (e.g., MetaboAnalyst [148], MetFlow [149]). However, the biological interpretation of metabolome data can be hindered by the relatively low coverage of identified metabolites in pathway databases (e.g., KEGG, BioCyc). For example, while most metabolic reconstructions cover the majority of primary metabolites, the coverage of lipids and secondary metabolites is far lower. Recently, an ontology database and enrichment analysis (LION) with chemical and biophysical information specifically for lipid species has become available [150]. In addition, the chemical similarity enrichment analysis (ChemRICH) approach has been introduced as an alternative to biochemical pathway mapping. ChemRICH is based on chemical similarity rather than sparse biochemical knowledge annotations [151].

Recent reviews focused on new software tools, databases, and resources in metabolomics, including those for statistical analysis and visualization, show the trend in developing fast, computationally less intensive, robust, open-source, and user-friendly tools [83,125]. In addition, with emerging recent metabolomics and lipidomics atlases [152,153], a trend of providing an interactive web tool (or web application) is apparent. Such a tool allows for querying specific metabolites, generating statistical models, or data visualization to understand complex data sets better.

7. Data sharing and FAIR guiding principles

Metabolomics and lipidomics analyses generate instrumental files for all study samples and different accompanying samples such as blanks, a dilution series of QC samples, QC samples repeatedly injected during the assay, or calibration curve samples. Making instrumental files publicly available is the first step toward reproducible research. Sharing data with the publication is recommended, but only a small portion of data from metabolomics and lipidomics research outputs are publicly available [154].

In 2016, the FAIR Guiding Principles for scientific data management and stewardship were published [155] to provide guidelines to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets. To meet the requirements of FAIR compliance, metabolomics and lipidomics data must fulfill the following criteria:

(1) Findable: using a repository to upload the data,
(2) Accessible: using a repository that is accessible to the scientific community and publicly available,
(3) Interoperable: using harmonized format and standard units of uploaded data,
(4) Reusable: providing comparable data with long-term validity and independent of time.

Data repositories designed for metabolomics data are available to facilitate data sharing, such as MetaboLights (ebi.ac.uk/metabolights) [156], Metabolomics Workbench (metabolomicsworkbench.org) [157], and MassIVE (massive.ucsd.edu/ProteoSAFe/static/massive.jsp). These repositories also adhere to minimum reporting standards [98] to promote data reusability.

The instrumental files can be deposited in an original proprietary vendor format or converted to open-access formats such as mzML and mzXML using ProteoWizard software (proteowizard.sourceforge.io). Files in vendor format should be preferred because, in some cases, the size of mzML and mzXML files can increase dramatically after conversion. In addition, many data processing software (e.g., MS-DIAL 4 [6], MZmine 3 [82], or XCMS Online [81]) nowadays permit instrumental files in vendor format to be uploaded during processing directly.

8. Metabolomics and lipidomics atlases

The metabolomic and lipidomic profiles from biofluids or tissues provide a global snapshot of analyte concentrations in a particular biological sample at a specific physiological state, time, or intervention response.

Table 3 provides an overview of recent metabolomics and lipidomics atlases generated using the above technologies. Most studies used an LC−MS-based multiplatform approach [152,153,158−162], leading to hundreds or thousands of annotated metabolites. However, other techniques, such as GC−MS [152,158,159], CE−MS [163], or NMR [164], were valuable sources of polar metabolites. Specifically for NMR, its immediate quantitative output (even though for a limited number of metabolites) enables direct comparisons of the results between laboratories and studies which is not the case for untargeted metabolomics and lipidomics studies, where the results are usually reported in detector signal intensities (“arbitrary units”). In addition, IMS-based spatial profiling of tissues [165,166] has brought a new dimension to building atlases because it reveals the distribution of biomolecules, such as polar metabolites and lipids, and helps better understand the cellular profile of the biological system.

The confidence level (Level 1−3) has not always been available in all examined atlases but based on the experimental sections, most of the studies relied on using spectral libraries that permit either identification (Level 1) based on retention time, MS1 and MS/MS spectra of authentic standards, or annotation (Level 2−3) based on MS/MS spectra match. For future studies, it would be valuable to include unknown (Level 4) features (“dark matter” of the metabolome) in generated atlases, ideally with acquired MS1 and MS/MS spectra for follow-up annotation. These unknowns can be revised during updates when new versions of MS/MS spectral libraries are available.

Basic statistics, such as *t*-test, ANOVA, fold-change, and correlations, were used for statistical analysis, followed by advanced data analysis such as JTK_CYCLE (circadian rhythms), PCA, or PLS-DA. Data visualization was supported, for instance, using box plots, dendrograms, heatmaps, circos plots, fuzzy plots, Sankey diagrams, Venn diagrams, metabolic networks, association and forest plots. Although interactive web tools have not always been implemented to visualize and access complex data, the below-described recent metabolomics and lipidomics atlases can serve as examples for future studies.

Atlas of the aging mouse brain (mouse.atlas.metabolomics.us)

Table 3
Overview of metabolomics and lipidomics atlases.

Description	Design	Species	*n* subjects	Matrices	Sample preparation	Platforms	Number of reported metabolites	Web tool	Ref.
Atlas of aging mouse brain	Aging (week 3, 16, 59, 92) of 10 anatomical regions of the brain	Mouse (C57BL/6N)	64	Cerebral cortex, olfactory bulb, hippocampus, hypothalamus, basal ganglia, thalamus, midbrain, pons, medulla, cerebellum	LLE: methanol, MTBE, water	4 LC–MS GC–MS	1547 (AU)	mouse.atlas.metabolomics.us	[152]
Atlas of oral ^{13}C-glucose tolerance test in mice (GTTAtlas)	Mouse plasma and tissues during the oral [$^{13}C_6$]-glucose tolerance test	Mouse (C57BL/6J)	28	Plasma, liver, heart, pancreas, gastrocnemius muscle, soleus muscle, kidney, epididymal WAT, subcutaneous WAT, duodenum, jejunum, ileum, BAT	LLE: methanol, MTBE, water	5 LC–MS 6 LC–MS for WAT and BAT	1904 (AU)	gttatlas.metabolomics.fgu.cas.cz	[153]
Atlas of the mouse metabolome and lipidome (MetaboAtlas21)	Diet (chow vs. HFD)	Mouse (C57BL/6N)	18	Liver, pancreas, stomach, small intestine, large intestine, lungs, testicle, kidney, urine, bladder, feces, spleen, BAT, epididymal WAT, subcutaneous WAT, soleus muscle, gastrocnemius muscle, quadriceps muscle, plasma, heart, brain	LLE: methanol, MTBE, water	7 LC–MS 8 LC–MS for WAT and BAT	3000+ (AU)	metaboatlas21.metabolomics.fgu.cas.cz	[153]
Atlas of circadian metabolism	Diet (chow vs. HFD); circadian rhythms	Mouse (C57BL/6J)	130	Serum, liver, sperm, muscle, suprachiasmatic nucleus, medial prefrontal cortex, BAT, WAT	PPT: methanol	3–4 LC–MS GC–MS	540+ (AU)	circadiomics.ics.uci.edu	[158]
Atlas of postnatal mouse heart development	Postnatal development (day 1, 4, 9, 23)	Mouse (C57BL/6JO1aHsd)	92	Heart	PPT: methanol	LC–MS GC × GC–MS	151 (conc.)	–	[159]
Mouse multiple tissue metabolome database (MMMDB)	Tissue and plasm profiling	Mouse (C57BL/6J)	2	Cerebra, cerebella, thymus, spleen, lung, liver, kidney, heart, pancreas, testis, plasma	LLE: methanol, chloroform, water	CE–MS	219 (conc.)	mmmdb.iab.keio.ac.jp	[163]
Atlas of exercise metabolism	Acute exercise bout performed at different times of the day	Mouse (C57BL6/JBomTac)	24	Serum, BAT, eWAT, heart, hypothalamus, iWAT, liver, muscle	PPT: methanol	4 LC–MS	550–800 (AU)	circadiomics.ics.uci.edu	[161]
Atlas of plant metabolites	Building reference to investigate systemic and local effects of pathogen infection or environmental stress	Plant	6	Oil seed rapeseed, wheat seed, spike rachis, stem, rice roots, infected wheat seed	2,5-dihydroxybenzoic acid matrix for positive ion mode; 4-nitroaniline for negative ion mode	MALDI-IMS	90+ (AU)	–	[166]
Atlas of reference lipidome for human WAT (AdipoAtlas)	Lean and obese individuals	Human	86	WAT (abdominal visceral and subcutaneous)	LLE: chloroform, methanol, water followed by polarity-based LLE	3 LC–MS	1636 (AU), 737 (conc.)	–	[160]
Atlas of human plasma metabolite signatures of gut microbiome composition (GUTSY Atlas)	Associations between the gut microbiota and host plasma metabolome	Human	8583	Plasma	PPT: methanol	4 LC–MS	1052 (AU)	gutsyatlas.serve.scilifelab.se	[162]
Atlas of biomarkers for health and disease	Associations of biomarkers to prevalence, incidence, and mortality of over 700 common diseases	Human	118 461	Plasma	–	NMR	249 (conc.)	biomarker-atlas.nightingale.cloud	[164]
Atlas of lipids in the human caudate nucleus	Neurologically normal and Huntington's disease caudate nucleus	Human	23	Brain (caudate nucleus)	1,5-diaminonaphthalene matrix	MALDI-IMS	372 (AU)	–	[165]

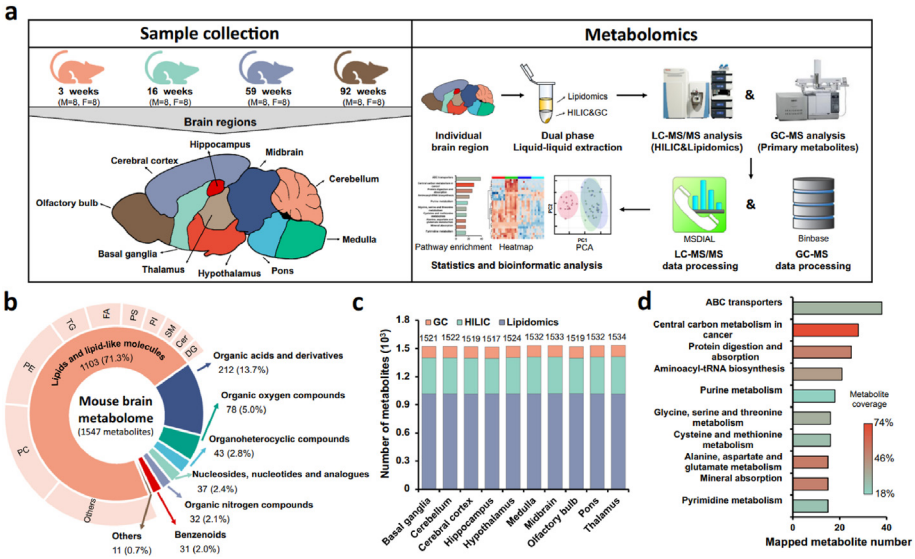

Fig. 5. Overview of the mouse brain atlas dataset. (a) Graphic illustration of the workflow to acquire aging mouse brain metabolome data. (b) Chemical composition of the mouse brain metabolome using ClassyFire categories to classify the metabolite diversity of all annotated metabolites across assays. (c) Number of annotated metabolites by metabolome assay and brain regions. (d) Mapping polar metabolites assayed by HILIC–MS and GC–MS to pathways. Top 10 mapped pathway-based sets are shown from a total of 118 pathways covered by Consensus PathDB. Reproduced with permission from Ref. [152].

[152]. This metabolome atlas provides the aging wildtype mouse (C57BL/6N) brain from 10 anatomical regions from adolescence to old age (Fig. 5). Multiplatform approach provided 1547 annotated metabolites. Almost all metabolites significantly differed between brain regions or age groups, but not by sex. Large changes in other metabolic pathways accompanied a shift in sphingolipid patterns related to myelin remodeling during aging. Functionally related brain regions (brain stem, cerebrum, and cerebellum) were also metabolically similar. However, in the cerebrum, metabolic correlations markedly weaken between adolescence and adulthood, whereas cross-region correlation patterns reflect decreased brain segregation at old age. The web application allows visualization of annotated metabolites in all 10 anatomical regions and developmental stages, highlighting statistical differences.

Atlas of oral ^{13}C-glucose tolerance test in mice – GTTAtlas (gttatlas.metabolomics.fgu.cas.cz) [153]. An analysis of $^{13}C_6$-glucose distribution was performed to map the metabolome and lipidome across 12 metabolically relevant mouse (C57BL/6J) organs and plasma, with integrated $^{13}C_6$-glucose-derived carbon tracing during oral glucose tolerance test (OGTT). Time profiles of water-soluble metabolites and lipids were acquired, and the global metabolite response was integrated into metabolic pathways. During the OGTT, glucose use was turned on with specific kinetics at the organ level, but fasting substrates like β-hydroxybutyrate were switched off in all organs simultaneously. Timeline profiling of ^{13}C-labeled fatty acids and triacylglycerols across tissues suggested that brown adipose tissue may contribute to the circulating fatty acid pool at maximal plasma glucose levels. The interactive web application is a unique resource for exploring whole-body glucose metabolism and time profiles of tissue and plasma metabolites during the OGTT.

Atlas of the mouse metabolome and lipidome – MetaboAtlas21 (metaboatlas21.metabolomics.fgu.cas.cz). This study presents a specific atlas of mouse metabolome and lipidome in systemic energy balance (chow diet) and under chronic nutrient stress (high-fat diet). Male mice (C57BL/6N) were fed a control (chow) diet for 2 months or a high-fat diet (HFD) for 2 and 10 months. Urine, plasma, feces, and 18 different tissues were collected from each animal for metabolomics and lipidomics analysis. These matrices covered digestive, excretory, respiratory, reproductive, endocrine, muscular, cardiovascular, and nervous systems. Over 3100 unique polar metabolites and simple and complex lipids were annotated. The web application allows the search of particular metabolites among all groups and samples and statistical analysis (*t*-test, fold change, Volcano plot, PCA, PLS-DA) to study the effect of the diet, the difference between matrices, or a combination of both.

Atlas of circadian metabolism (circadiomics.ics.uci.edu) [158]. Dyar et al. performed 24-hr metabolomics profiling of plasma and seven tissues of mice (C57BL/6J) and presented a temporal and spatial atlas of circadian metabolism under standard and HFD (Fig. 6). White adipose tissue, liver, brown adipose tissue, muscle, and serum were most impacted by HFD, with 40–60% of detected metabolites showing significant alterations. Conversely, brain and sperm metabolites appeared more resistant to diet-induced metabolite alterations. Subpathway enrichment analysis of significantly altered metabolites identified lipid classes altered by HFD, specifically long-chain fatty acids, polyunsaturated fatty acids, diacylglycerols, phospholipids, sphingolipids, glycerolipids, and lysolipids. In addition, polar metabolites such as arginine, proline, lysine, leucine, isoleucine, valine, and dipeptides were also highly impacted. This atlas constitutes a rich starting point for hypothesis generation and validation of metabolic networks across tissues and under different feeding conditions.

Atlas of postnatal mouse heart development [159]. Talman et al. conducted metabolomics profiling of mouse (C57BL/6JOlaHsd) ventricular tissue samples collected on postnatal days 1, 4, 9, and 23. In total, 151 metabolites with significant changes were identified. Differentially regulated metabolic pathways include branched-chain amino acid degradation, fatty acid metabolism, and the HMGCS (HMG-CoA [hydroxymethylglutaryl-coenzyme A] synthase)–mediated mevalonate pathway and ketogenesis. The metabolomics data were also combined with RNA sequencing and global proteomics. These integrated data of molecular changes associated with the loss of cardiac regeneration may open up new possibilities for developing regenerative therapies.

Mouse Multiple Tissue Metabolome Database – MMMDB (mmmdb.iab.keio.ac.jp) [163]. This atlas provides concentrations of over 200 polar metabolites in 10 different mouse (C57BL/6J) tissues and plasma.

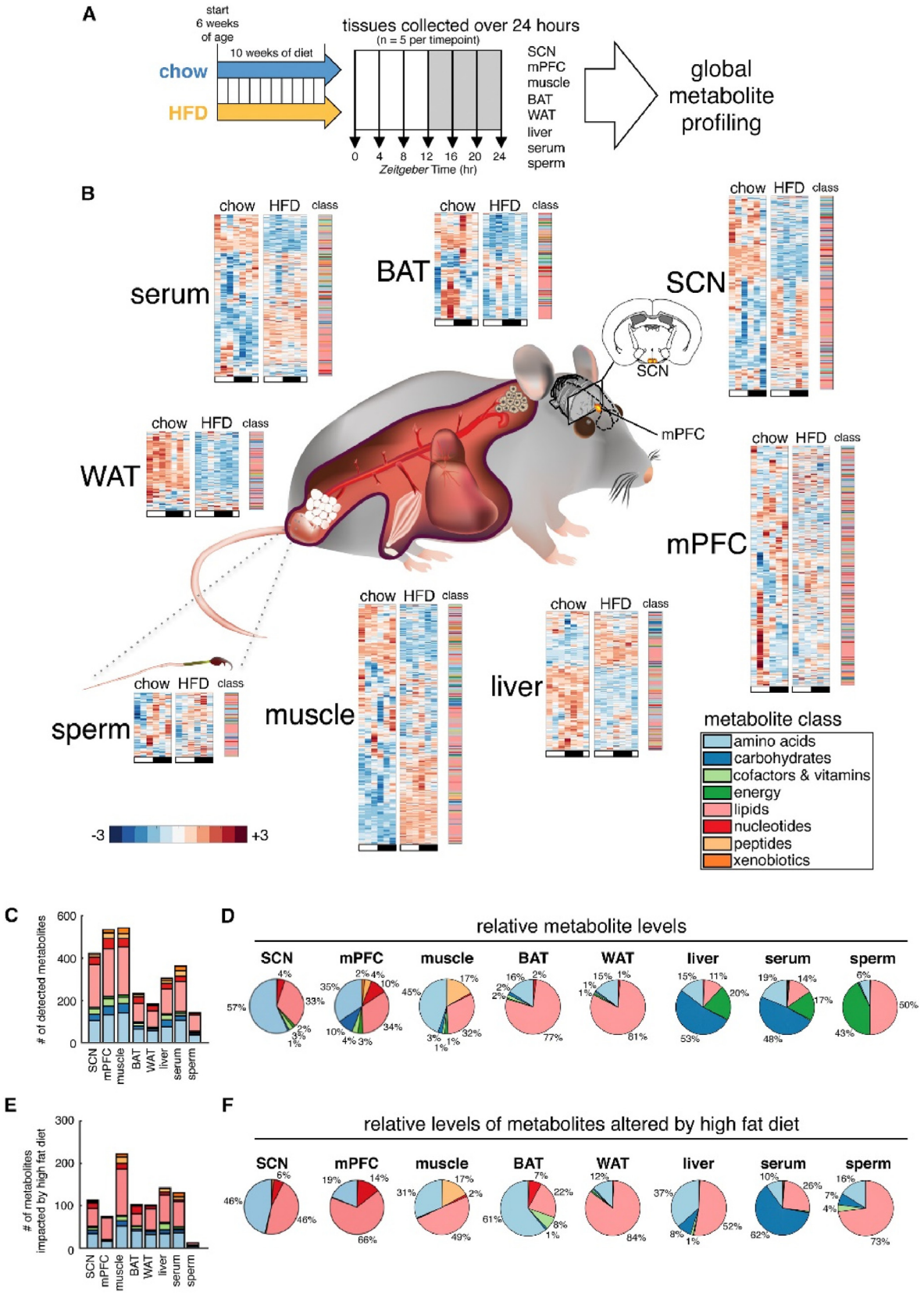

Fig. 6. Global metabolite profiling of mouse tissues over 24 h under chow and HFD. (a) Suprachiasmatic nucleus (SCN), medial prefrontal cortex (mPFC), gastrocnemius skeletal muscle, interscapular brown adipose tissue (BAT), epididymal white adipose tissue (WAT), liver, serum, and cauda epididymal sperm were collected every 4 h across the light/dark cycle from a single cohort of mice. (b) Tissue-specific metabolite heatmaps. Rows reflect normalized (*z*-score) metabolite abundance across the light/dark cycle (white bar, ZT0, 4, 8, and 24; black bar, ZT12, 16, and 20). Metabolite class is indicated at the right of each heatmap. (c) Counts and class of detected metabolites for each tissue. (d) Tissue metabolite class composition according to relative metabolite masses (sum of standardized abundances). (e) Counts and class of metabolites significantly impacted by diet (diet effect $p < 0.05$, linear regression model). (f) HFD-altered metabolite class composition for each tissue (relative metabolite masses affected by HFD). Reproduced with permission from Ref. [158].

Atlas of exercise metabolism (circadiomics.ics.uci.edu) [161]. This atlas includes global metabolomics profiling of multiple tissues and serum during exercise of male mice (C57BL6/JBomTac). Timed exercise rewires tissue-specific and systemic metabolism and differential tissue production and distribution of exerkines. 2-Hydroxybutyrate was also validated as a time-dependent exerkine at the systemic level. This resource provides physiological context about diurnal production and distribution of a wide range of signaling metabolites linked to sleep, memory, energy homeostasis, endurance capacity, and performance.

Atlas of plant metabolites [166]. This study aims toward a plant metabolite atlas using IMS of plant metabolites in multiple mono- and dicot species and major plant organs. Changes in metabolite patterns during plant development were investigated for germination of oilseed rape. More than 90 compounds allowed assignment to metabolic processes and indicated possible functions in plant tissues. In addition, the untargeted nature of IMS allowed the detection of marker compounds for the physiological status.

Atlas of reference lipidome for human white adipose tissue – AdipoAtlas [160]. In-depth lipidomics profiling of human subcutaneous and visceral white adipose tissue of lean and obese individuals was performed, leading to annotation and semi-absolute quantification of 1636 and 737 lipid molecular species, respectively. In addition, deep lipidomic profiling identified the main lipid (sub) classes undergoing depot-/phenotype-specific remodeling. This reference lipidome is a data-rich resource for developing WAT-specific high-throughput methods and a scaffold for systems medicine data integration.

Atlas of human plasma metabolite signatures of gut microbiome composition – GUTSY Atlas (gutsyatlas.serve.scilifelab.se) [162]. High-resolution deep metagenomic sequencing and mass spectrometry-based metabolite profiling were applied to analyze samples from 8583 individuals from a well-characterized population-based study. The authors found that the gut microbiota explains up to 58% of the variance of individual plasma metabolites. In addition, the potential of this resource was exemplified by presenting novel associations between dietary factors and oral medication with the gut microbiome and microbial species strongly associated with the uremic toxin *p*-cresol sulfate.

Atlas of biomarkers for health and disease (biomarker-atlas.nightingale.cloud) [164]. This comprehensive NMR-based atlas covers 249 lipoprotein lipids, fatty acids, and small molecules (out of which 17 represent amino acids, ketones, and glycolysis metabolites) quantified in 118 461 participants in the UK Biobank with associations of these biomarkers to prevalence, incidence, and mortality of over 700 common diseases. In addition, several biomarker associations, including susceptibility to infectious diseases and risk for the onset of various cancers, joint disorders, and mental health outcomes, indicate that abundant circulating lipids and metabolites are risk markers well beyond cardiometabolic diseases.

Atlas of lipids in the human caudate nucleus [165]. IMS was used to generate a comprehensive, unbiased, and spatially resolved lipidomic atlas of the caudate nucleus in human post-mortem tissue from neurologically normal and Huntington's disease subjects. Most lipid species were highly conserved in Huntington's disease subjects compared to age-matched controls. However, caudate nucleus specimens from Huntington's disease cases spanning a range of neuropathological grades showed a lower focal abundance of the neuroprotective docosahexaenoic and adrenic acids, several cardiolipins, the ganglioside GM1, and glycerophospholipids with long polyunsaturated fatty acyls.

9. Conclusions

Recent advanced in sample preparation, instrumental analysis, data processing, and bioinformatics have revolutionized metabolomics and lipidomics research. For building mass spectrometry-based metabolomics and lipidomics atlases, we recommend the following steps:

- All-in-one extraction with multiple internal standards should be preferred to isolate different fractions of the metabolome/lipidome during untargeted profiling combined with targeted analysis of selected polar metabolites and lipid classes. Such extraction reduces the need for extra samples and laboratory operations when separate extractions are conducted for each instrumental platform.
- A multiplatform MS-based approach with at least one method (LC–MS or direct infusion) for simple and complex lipids and one method (LC–MS or GC–MS) for polar metabolites is needed to cover metabolites differing in their physicochemical properties. In addition, advanced data processing software for high-resolution MS1 data (used for quantification or semi-quantification) linked with acquired MS/MS spectra (used for annotation) speeds up the workflow.
- Since complex data sets are generated during metabolomics and lipidomics analysis, we encourage presenting the data using an interactive web tool (or web application) where users can test different hypotheses, query annotated metabolites, and download the results. In addition, sharing raw instrumental files and reports with annotated metabolites (or even unknowns) will lead to reproducible research. These files can also be used while developing and validating new software tools for data processing.

Over the following years, novel technologies such as IM-MS will mature more and provide improved metabolite separation and annotation [65]. IMS to map the distribution of a wide range of compounds (as polar metabolites, lipids, and proteins) directly from thin tissue sections [167] and single, isolated or cultured cells [168] will significantly aid researchers in many fields; for instance, to examine subtle but important differences in cell types and their relationships in the brain.

The joint effort toward building multi-omics atlases (e.g., genomics, transcriptomics, proteomics, lipidomics, metabolomics, microbiomics) will generate a bigger picture and provide a deeper and more complete understanding of biological systems [169]. Importantly, simple, fast, and robust protocols for the combined extraction of polar metabolites, lipids, and proteins of the same sample have become available [28,29] and may help to reduce analysis cost and sample handling compared to separate analyses.

The current metabolomics and lipidomics atlases have focused mainly on animal models with analysis of different biofluids and tissues and human cohorts for which plasma and adipose tissues were collected. However, specifically for human cohorts, more comprehensive data will be needed to study biomarkers for health and disease. In addition, since metabolism is vital to every aspect of cell function [170], metabolomics and lipidomics atlases will also be needed for cells, including induced pluripotent stem cells, that have opened the field of regenerative medicine [171].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This study was supported by the Czech Health Research Council (grants no. NU20-01-00186, NU22-02-00161), Czech Science Foundation (20-21114S, 21-00477S), Ministry of Education, Youth and Sport of the Czech Republic (LTAUSA19124), and the project National Institute for Research of Metabolic and Cardiovascular Diseases (Programme EXCELES, ID Project No. LX22NPO5104) funded by the European Union – Next Generation EU.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.trac.2022.116825.

References

[1] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass spectrometry-based metabolomics, Mass Spectrom. Rev. 26 (2007) 51–78. https://doi.org/10.1002/Mas.20108.
[2] H. Gallart-Ayala, T. Teav, J. Ivanisevic, Metabolomics meets lipidomics: assessing the small molecule component of metabolism, Bioessays 42 (2020), 2000052. https://doi.org/10.1002/bies.202000052.
[3] N. Psychogios, D.D. Hau, J. Peng, A.C. Guo, R. Mandal, S. Bouatra, I. Sinelnikov, R. Krishnamurthy, R. Eisner, B. Gautam, N. Young, J.G. Xia, C. Knox, E. Dong, P. Huang, Z. Hollander, T.L. Pedersen, S.R. Smith, F. Bamforth, R. Greiner, B. McManus, J.W. Newman, T. Goodfriend, D.S. Wishart, The human serum metabolome, PLoS One 6 (2011), e16957. https://doi.org/10.1371/journal.pone.0016957.
[4] T. Cajka, O. Fiehn, Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics, Anal. Chem. 88 (2016) 524–545. https://doi.org/10.1021/acs.analchem.5b04491.
[5] E. Rampler, Y. El Abiead, H. Schoeny, M. Rusz, F. Hildebrand, V. Fitz, G. Koellensperger, Recurrent topics in mass spectrometry-based metabolomics and lipidomics-standardization, coverage, and throughput, Anal. Chem. 93 (2021) 519–545. https://doi.org/10.1021/acs.analchem.0c04698.
[6] H. Tsugawa, K. Ikeda, M. Takahashi, A. Satoh, Y. Mori, H. Uchino, N. Okahashi, Y. Yamada, I. Tada, P. Bonini, Y. Higashi, Y. Okazaki, Z. Zhou, Z.-J. Zhu, J. Koelmel, T. Cajka, O. Fiehn, K. Saito, M. Arita, M. Arita, A lipidome atlas in MS-DIAL 4, Nat. Biotechnol. 38 (2020) 1159–1163. https://doi.org/10.1038/s41587-020-0531-2.
[7] T. Cajka, Towards merging targeted and untargeted analysis of the lipidome, metabolome, and exposome, LC-GC Eur. 32 (2019) 314–316.
[8] T. Cajka, J.T. Smilowitz, O. Fiehn, Validating quantitative untargeted lipidomics across nine liquid chromatography-high-resolution mass spectrometry platforms, Anal. Chem. 89 (2017) 12360–12368. https://doi.org/10.1021/acs.analchem.7b03404.
[9] K. Simons, How can omic science be improved? Proteomics 18 (2018), e1800039. https://doi.org/10.1002/Pmic.201800039.
[10] D.S. Wishart, D. Tzur, C. Knox, R. Eisner, A.C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M.A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D.D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G.E. Duggan, G.D. MacInnis, A.M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B.D. Sykes, H.J. Vogel, L. Querengesser, HMDB: the human metabolome database, Nucleic Acids Res. 35 (2007) D521–D526. https://doi.org/10.1093/nar/gkl923.
[11] D.S. Wishart, A.C. Guo, E. Oler, F. Wang, A. Anjum, H. Peters, R. Dizon, Z. Sayeeda, S.Y. Tian, B.L. Lee, M. Berjanskii, R. Mah, M. Yamamoto, J. Jovel, C. Torres-Calzada, M. Hiebert-Giesbrecht, V.W. Lui, D. Varshavi, D. Varshavi, D. Allen, D. Arndt, N. Khetarpal, A. Sivakumaran, K. Harford, S. Sanford, K. Yee, X. Cao, Z. Budinski, J. Liigand, L. Zhang, J.M. Zheng, R. Mandal, N. Karu, M. Dambrova, H.B. Schioth, R. Greiner, V. Gautam, HMDB 5.0: the human metabolome database for 2022, Nucleic Acids Res. 50 (2022) D622–D631. https://doi.org/10.1093/nar/gkab1062.
[12] P. Anikeeva, E. Boyden, C. Brangwynne, I.I. Cisse, O. Fiehn, P. Fromme, A.C. Gingras, C.S. Greene, E. Heard, S.W. Hell, E. Hillman, G.J. Jensen, R. Karchin, L.L. Kiessling, B.P. Kleinstiver, R. Knight, P. Kukura, M.A. Lancaster, N. Loman, L. Looger, E. Lundberg, Q.M. Luo, A. Miyawaki, E.W. Myers, G.P. Nolan, P. Picotti, W. Reik, M. Sauer, A.K. Shalek, J. Shendure, N. Slavov, A. Tanay, O. Troyanskaya, D. van Valen, H.W. Wang, C.Q. Yi, P. Yin, M. Zernicka-Goetz, X.W. Zhuang, Voices in methods development, Nat. Methods 16 (2019) 945–951. https://doi.org/10.1038/s41592-019-0585-6.
[13] D.K. Barupal, O. Fiehn, Generating the blood exposome database using a comprehensive text mining and database fusion approach, Environ. Health Perspect. 127 (2019), 97008. https://doi.org/10.1289/EHP4713.
[14] S. Alseekh, A. Aharoni, Y. Brotman, K. Contrepois, J. D'Auria, J. Ewald, J.C. Ewald, P.D. Fraser, P. Giavalisco, R.D. Hall, M. Heinemann, H. Link, J. Luo, S. Neumann, J. Nielsen, L.P. de Souza, K. Saito, U. Sauer, F.C. Schroeder, S. Schuster, G. Siuzdak, A. Skirycz, L.W. Sumner, M.P. Snyder, H.R. Tang, T. Tohge, Y.L. Wang, W.W. Wen, S. Wu, G.W. Xu, N. Zamboni, A.R. Fernie, Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices, Nat. Methods 18 (2021) 747–756. https://doi.org/10.1038/s41592-021-01197-1.
[15] J. Pezzatti, J. Boccard, S. Codesido, Y. Gagnebin, A. Joshi, D. Picard, V. Gonzalez-Ruiz, S. Rudaz, Implementation of liquid chromatography-high resolution mass spectrometry methods for untargeted metabolomic analyses of biological samples: a tutorial, Anal. Chim. Acta 1105 (2020) 28–44. https://doi.org/10.1016/j.aca.2019.12.062.
[16] L. Smith, J. Villaret-Cazadamont, S.P. Claus, C. Canlet, H. Guillou, N.J. Cabaton, S. Ellero-Simatos, Important considerations for sample collection in metabolomics studies with a special focus on applications to liver functions, Metabolites 10 (2020) 104. https://doi.org/10.3390/Metabo10030104.
[17] R. Lehmann, From bedside to bench-practical considerations to avoid pre-analytical pitfalls and assess sample quality for high-resolution metabolomics and lipidomics analyses of body fluids, Anal. Bioanal. Chem. 413 (2021) 5567–5585. https://doi.org/10.1007/s00216-021-03450-0.
[18] M. Saoi, P. Britz-McKibbin, New advances in tissue metabolomics: a review, Metabolites 11 (2021) 672. https://doi.org/10.3390/Metabo11100672.
[19] D. Vuckovic, Sample Preparation in Global Metabolomics of Biological Fluids and Tissues, 2020, pp. 53–83. https://doi.org/10.1016/B978-0-12-818607-7.00004-9.
[20] O. Fiehn, C.S. Bloszies, Using untargeted metabolomics for detecting exposome compounds, Curr. Opin. Toxicol. 8 (2018) 87–92. https://doi.org/10.1016/j.cotox.2018.03.002.
[21] R.J. Raterink, P.W. Lindenburg, R.J. Vreeken, R. Ramautar, T. Hankemeier, Recent developments in sample-pretreatment techniques for mass spectrometry-based metabolomics, TrAC-Trend Anal. Chem. 61 (2014) 157–167. https://doi.org/10.1016/j.trac.2014.06.003.
[22] L. Whiley, J. Godzien, F.J. Ruperez, C. Legido-Quigley, C. Barbas, In-vial dual extraction for direct LC-MS analysis of plasma for comprehensive and highly reproducible metabolic fingerprinting, Anal. Chem. 84 (2012) 5992–5999. https://doi.org/10.1021/Ac300716u.
[23] R.E. Patterson, A.J. Ducrocq, D.J. McDougall, T.J. Garrett, R.A. Yost, Comparison of blood plasma sample preparation methods for combined LC-MS lipidomics and metabolomics, J. Chromatogr. B 1002 (2015) 260–266. https://doi.org/10.1016/j.jchromb.2015.08.018.
[24] J. Medina, V. van der Velpen, T. Teav, Y. Guitton, H. Gallart-Ayala, J. Ivanisevic, Single-step extraction coupled with targeted HILIC-MS/MS approach for comprehensive analysis of human plasma lipidome and polar metabolome, Metabolites 10 (2020) 495. https://doi.org/10.3390/Metabo10120495.
[25] V. Matyash, G. Liebisch, T.V. Kurzchalia, A. Shevchenko, D. Schwudke, Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics, J. Lipid Res. 49 (2008) 1137–1146. https://doi.org/10.1194/jlr.D700041-JLR200.
[26] J. Folch, M. Lees, G.H.S. Stanley, A simple method for the isolation and purification of total lipides from animal tissues, J. Biol. Chem. 226 (1957) 497–509. https://doi.org/10.1016/S0021-9258(18)64849-5.
[27] A.D. Southam, L.D. Haglington, L. Najdekr, A. Jankevics, R.J.M. Weber, W.B. Dunn, Assessment of human plasma and urine sample preparation for reproducible and high-throughput UHPLC-MS clinical metabolic phenotyping, Analyst 145 (2020) 6511–6523. https://doi.org/10.1039/d0an01319f.
[28] C. Coman, F.A. Solari, A. Hentschel, A. Sickmann, R.P. Zahedi, R. Ahrends, Simultaneous metabolite, rotein, lipid extraction (SIMPLEX): a combinatorial multimolecular oOmics aproach for systems biology, Mol. Cell. Proteomics 15 (2016) 1453–1466. https://doi.org/10.1074/mcp.M115.053702.
[29] E.S. Nakayasu, C.D. Nicora, A.C. Sims, K.E. Burnum-Johnson, Y.M. Kim, J.E. Kyle, M.M. Matzke, A.K. Shukla, R.K. Chu, A.A. Schepmoes, J.M. Jacobs, R.S. Baric, B.J. Webb-Robertson, R.D. Smith, T.O. Metz, MPLEx: a robust and universal protocol for single-sample integrative proteomic, metabolomic, and lipidomic analyses, mSystems 1 (2016). https://doi.org/10.1128/mSystems.00043-16. e00043-00016.
[30] G. Vale, S.A. Martin, M.A. Mitsche, B.M. Thompson, K.M. Eckert, J.G. McDonald, Three-phase liquid extraction: a simple and fast method for lipidomic workflows, J. Lipid Res. 60 (2019) 694–706. https://doi.org/10.1194/jlr.D090795.
[31] S.L. Chen, M. Hoene, J. Li, Y.J. Li, X.J. Zhao, H.U. Haring, E.D. Schleicher, C. Weigert, G.W. Xu, R. Lehmann, Simultaneous extraction of metabolome and lipidome with methyl tert-butyl ether from a single small tissue sample for ultra-high performance liquid chromatography/mass spectrometry, J. Chromatogr. A 1298 (2013) 9–16. https://doi.org/10.1016/j.chroma.2013.05.019.
[32] K. Hanhineva, T. Barri, M. Kolehmainen, J. Pekkinen, J. Pihlajamaki, A. Vesterbacka, G. Solano-Aguilar, H. Mykkanen, L.O. Dragsted, J.F. Urban Jr., K. Poutanen, Comparative nontargeted profiling of metabolic changes in tissues and biofluids in high-fat diet-fed Ossabaw pig, J. Proteome Res. 12 (2013) 3980–3992. https://doi.org/10.1021/pr400257d.
[33] G. Gurdeniz, M. Kristensen, T. Skov, L.O. Dragsted, The effect of LC-MS data preprocessing methods on the selection of plasma biomarkers in fed vs. fasted rats, Metabolites 2 (2012) 77–99. https://doi.org/10.3390/metabo2010077.
[34] Y. Li, Z. Zhang, X. Liu, A. Li, Z. Hou, Y. Wang, Y. Zhang, A novel approach to the simultaneous extraction and non-targeted analysis of the small molecules metabolome and lipidome using 96-well solid phase extraction plates with column-switching technology, J. Chromatogr. A 1409 (2015) 277–281. https://doi.org/10.1016/j.chroma.2015.07.048.

[35] P. Janovska, V. Melenovsky, M. Svobodova, T. Havlenova, H. Kratochvilova, M. Haluzik, E. Hoskova, T. Pelikanova, J. Kautzner, L. Monzo, I. Jurcova, K. Adamcova, L. Lenkova, J. Buresova, M. Rossmeisl, O. Kuda, T. Cajka, J. Kopecky, Dysregulation of epicardial adipose tissue in cachexia due to heart failure: the role of natriuretic peptides and cardiolipin, J. Cachexia Sarcopenia Muscle 11 (2020) 1614–1627. https://doi.org/10.1002/jcsm.12631.
[36] Y.H. Yang, C. Cruickshank, M. Armstrong, S. Mahaffey, R. Reisdorph, N. Reisdorph, New sample preparation approach for mass spectrometry-based profiling of plasma results in improved coverage of metabolome, J. Chromatogr. A 1300 (2013) 217–226. https://doi.org/10.1016/j.chroma.2013.04.030.
[37] J.C. Garcia-Canaveras, M.T. Donato, J.V. Castell, A. Lahoz, A comprehensive untargeted metabonomic analysis of human steatotic liver tissue by RP and HILIC chromatography coupled to mass spectrometry reveals important metabolic alterations, J. Proteome Res. 10 (2011) 4825–4834. https://doi.org/10.1021/Pr200629p.
[38] H.C. Kofeler, R. Ahrends, E.S. Baker, K. Ekroos, X.L. Han, N. Hoffmann, M. Holcapek, M.R. Wenk, G. Liebisch, Recommendations for good practice in MS-based lipidomics, J. Lipid Res. 62 (2021), 100138. https://doi.org/10.1016/J.Jlr.2021.100138.
[39] I. Gertsman, J.A. Gangoiti, B.A. Barshop, Validation of a dual LC-HRMS platform for clinical metabolic diagnosis in serum, bridging quantitative analysis and untargeted metabolomics, Metabolomics 10 (2014) 312–323. https://doi.org/10.1007/s11306-013-0582-1.
[40] A.L. Taylor, D.E. Davis, S.G. Codreanu, F.E. Harrison, S.D. Sherrod, J.A. McLean, Targeted and untargeted mass spectrometry reveals the impact of high-fat diet on peripheral amino acid regulation in a mouse model of Alzheimer's disease, J. Proteome Res. 20 (2021) 4405–4414. https://doi.org/10.1021/acs.jproteome.1c00344.
[41] C. Roy, P.Y. Tremblay, J.F. Bienvenu, P. Ayotte, Quantitative analysis of amino acids and acylcarnitines combined with untargeted metabolomics using ultra-high performance liquid chromatography and quadrupole time-of-flight mass spectrometry, J. Chromatogr. B 1027 (2016) 40–49. https://doi.org/10.1016/j.jchromb.2016.05.006.
[42] M. Schwaiger, H. Schoeny, Y. El Abiead, G. Hermann, E. Rampler, G. Koellensperger, Merging metabolomics and lipidomics into one analytical run, Analyst 144 (2019) 220–229. https://doi.org/10.1039/c8an01219a.
[43] L. Zhang, W. Zheng, X. Li, S. Wang, M. Xiao, R. Xiao, D. Zhang, N. Ke, H. Cai, J. Cheng, X. Chen, M. Gong, A merged method for targeted analysis of amino acids and derivatives using parallel reaction monitoring combined with untargeted profiling by HILIC-Q-Orbitrap HRMS, J. Pharm. Biomed. Anal. 203 (2021), 114208. https://doi.org/10.1016/j.jpba.2021.114208.
[44] Y. Gao, Y.H. Chen, X.F. Yue, J.M. He, R.P. Zhang, J. Xu, Z. Zhou, Z.H. Wang, R. Zhang, Z. Abliz, Development of simultaneous targeted metabolite quantification and untargeted metabolomics strategy using dual-column liquid chromatography coupled with tandem mass spectrometry, Anal. Chim. Acta 1037 (2018) 369–379. https://doi.org/10.1016/j.aca.2018.08.042.
[45] M. Schwaiger, E. Rampler, G. Hermann, W. Miklos, W. Berger, G. Koellensperger, Anion-exchange chromatography coupled to high-resolution mass spectrometry: a powerful tool for merging targeted and non-targeted metabolomics, Anal. Chem. 89 (2017) 7667–7674. https://doi.org/10.1021/acs.analchem.7b01624.
[46] L. Liang, J. Xu, W.W. Zhou, E. Brand, H.B. Chen, Z.Z. Zhao, Integrating targeted and untargeted metabolomics to investigate the processing chemistry of Polygoni Multiflori Radix, Front. Pharmacol. 9 (2018) 934. https://doi.org/10.3389/Fphar.2018.00934.
[47] A.P. Siskos, P. Jain, W. Romisch-Margl, M. Bennet, D. Achaintre, Y. Asad, L. Marney, L. Richardson, A. Koulman, J.L. Griffin, F. Raynaud, A. Scalbert, J. Adamski, C. Prehn, H.C. Keun, Interlaboratory reproducibility of a targeted metabolomics platform for analysis of human serum and plasma, Anal. Chem. 89 (2017) 656–665. https://doi.org/10.1021/acs.analchem.6b02930.
[48] Y. Izumi, F. Matsuda, A. Hirayama, K. Ikeda, Y. Kita, K. Horie, D. Saigusa, K. Saito, Y. Sawada, H. Nakanishi, N. Okahashi, M. Takahashi, M. Nakao, K. Hata, Y. Hoshi, M. Morihara, K. Tanabe, T. Bamba, Y. Oda, Inter-laboratory comparison of metabolite measurements for metabolomics data integration, Metabolites 9 (2019) 257. https://doi.org/10.3390/Metabo9110257.
[49] J.W. Thompson, K.J. Adams, J. Adamski, Y. Asad, D. Borts, J.A. Bowden, G. Byram, V. Dang, W.B. Dunn, F. Fernandez, O. Fiehn, D.A. Gaul, A.F.R. Huhmer, A. Kalli, T. Koal, S. Koeniger, R. Mandal, F. Meier, F.J. Naser, D. O'Neil, A. Pal, G.J. Patti, P.T. Hai, C. Prehn, F.I. Raynaud, T. Shen, A.D. Southam, L. St John-Williams, K. Sulek, C.G. Vasilopoulou, M. Viant, C.L. Winder, D. Wishart, L. Zhang, J.M. Zheng, M.A. Moseley, International ring trial of a high resolution targeted metabolomics and lipidomics platform for serum and plasma analysis, Anal. Chem. 91 (2019) 14407–14416. https://doi.org/10.1021/acs.analchem.9b02908.
[50] J.L. Wolfender, G. Marti, A. Thomas, S. Bertrand, Current approaches and challenges for the metabolite profiling of complex natural extracts, J. Chromatogr. A 1382 (2015) 136–164. https://doi.org/10.1016/j.chroma.2014.10.091.
[51] G. Theodoridis, H.G. Gika, I.D. Wilson, Mass spectrometry-based holistic analytical approaches for metabolite profiling in systems biology studies, Mass Spectrom. Rev. 30 (2011) 884–906. https://doi.org/10.1002/mas.20306.
[52] S.J. Blanksby, T.W. Mitchell, Advances in mass spectrometry for lipidomics, Annu. Rev. Anal. Chem. 3 (2010) 433–465. https://doi.org/10.1146/annurev.anchem.111808.073705.
[53] T. Cajka, O. Fiehn, Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry, TrAC-Trend Anal. Chem. 61 (2014) 192–206. https://doi.org/10.1016/j.trac.2014.04.017.
[54] M. Lange, Z.X. Ni, A. Criscuolo, M. Fedorova, Liquid chromatography techniques in lipidomics research, Chromatographia 82 (2019) 77–100. https://doi.org/10.1007/s10337-018-3656-4.
[55] M. Lisa, E. Cifkova, M. Holcapek, Lipidomic profiling of biological tissues using off-line two-dimensional high-performance liquid chromatography mass spectrometry, J. Chromatogr. A 1218 (2011) 5146–5156. https://doi.org/10.1016/j.chroma.2011.05.081.
[56] S. Tumanov, J.J. Kamphorst, Recent advances in expanding the coverage of the lipidome, Curr. Opin. Biotechnol. 43 (2017) 127–133. https://doi.org/10.1016/j.copbio.2016.11.008.
[57] E. Chekmeneva, G.D. Correia, M. Gomez-Romero, J. Stamler, Q. Chan, P. Elliott, J.K. Nicholson, E. Holmes, Ultra-performance liquid chromatography high-resolution mass spectrometry and direct infusion-high-resolution mass spectrometry for combined exploratory and targeted metabolic profiling of human urine, J. Proteome Res. 17 (2018) 3492–3502. https://doi.org/10.1021/acs.jproteome.8b00413.
[58] J.L. Markley, R. Bruschweiler, A.S. Edison, H.R. Eghbalnia, R. Powers, D. Raftery, D.S. Wishart, The future of NMR-based metabolomics, Curr. Opin. Biotechnol. 43 (2017) 34–40. https://doi.org/10.1016/j.copbio.2016.08.001.
[59] A. Benova, M. Ferencakova, K. Bardova, J. Funda, J. Prochazka, F. Spoutil, T. Cajka, M. Dzubanova, T. Balcaen, G. Kerckhofs, W. Willekens, G.H. van Lenthe, G. Alquicer, A. Pecinova, T. Mracek, O. Horakova, M. Rossmeisl, J. Kopecky, M. Tencerova, Novel thiazolidinedione analog reduces a negative impact on bone and mesenchymal stem cell properties in obese mice compared to classical thiazolidinediones, Mol. Metabol. 65 (2022), 101598. https://doi.org/10.1016/j.molmet.2022.101598.
[60] S. Fekete, J. Schappler, J.L. Veuthey, D. Guillarme, Current and future trends in UHPLC, TrAC-Trend Anal. Chem. 63 (2014) 2–13. https://doi.org/10.1016/j.trac.2014.08.007.
[61] N. Gray, M.R. Lewis, R.S. Plumb, I.D. Wilson, J.K. Nicholson, High-throughput microbore UPLC-MS metabolic phenotyping of urine for large-scale epidemiology studies, J. Proteome Res. 14 (2015) 2714–2721. https://doi.org/10.1021/acs.jproteome.5b00203.
[62] K. Schonberger, M. Mitterer, J.M. Buescher, N. Cabezas-Wallscheid, Targeted LC-MS/MS-based metabolomics and lipidomics on limited hematopoietic stem cell numbers, STAR Protoc 3 (2022), 101408. https://doi.org/10.1016/j.xpro.2022.101408.
[63] J.P. Koelmel, N.M. Kroeger, E.L. Gill, C.Z. Ulmer, J.A. Bowden, R.E. Patterson, R.A. Yost, T.J. Garrett, Expanding lipidome coverage using LC-MS/MS data-dependent acquisition with automated exclusion list generation, J. Am. Soc. Mass Spectrom. 28 (2017) 908–917. https://doi.org/10.1007/s13361-017-1608-0.
[64] R.S. Plumb, G. Isaac, P.D. Rainville, J. Hill, L.A. Gethings, K.A. Johnson, J. Lauterbach, I.D. Wilson, High throughput UHPLC-MS-based lipidomics using vacuum jacketed columns, J. Proteome Res. 21 (2022) 691–701. https://doi.org/10.1021/acs.jproteome.1c00836.
[65] G. Paglia, A.J. Smith, G. Astarita, Ion mobility mass spectrometry in the omics era: challenges and opportunities for metabolomics and lipidomics, Mass Spectrom. Rev. 41 (2022) 722–765. https://doi.org/10.1002/mas.21686.
[66] P. Dwivedi, A.J. Schultz, H.H. Hill, Metabolic profiling of human blood by high-resolution ion mobility mass spectrometry (IM-MS), Int. J. Mass Spectrom. 298 (2010) 78–90. https://doi.org/10.1016/j.ijms.2010.02.007.
[67] B.K. Ubhi, Direct infusion-tandem mass spectrometry (DI-MS/MS) analysis of complex lipids in human plasma and serum using the Lipidyzer (TM) platform, Methods Mol. Biol. 1730 (2018) 227–236. https://doi.org/10.1007/978-1-4939-7592-1_15.
[68] G. Paglia, P. Angel, J.P. Williams, K. Richardson, H.J. Olivos, J.W. Thompson, L. Menikarachchi, S. Lai, C. Walsh, A. Moseley, R.S. Plumb, D.F. Grant, B.O. Palsson, J. Langridge, S. Geromanos, G. Astarita, Ion mobility-derived collision cross section as an additional measure for lipid fingerprinting and identification, Anal. Chem. 87 (2015) 1137–1144. https://doi.org/10.1021/ac503715v.
[69] E.S. Rivera, K.V. Djambazova, E.K. Neumann, R.M. Caprioli, J.M. Spraggins, Integrating ion mobility and imaging mass spectrometry for comprehensive analysis of biological tissues: a brief review and perspective, J. Mass Spectrom. 55 (2020) E4614. https://doi.org/10.1002/Jms.4614.
[70] D.P. Jones, Sequencing the exposome: a call to action, Toxicol Rep 3 (2016) 29–45. https://doi.org/10.1016/j.toxrep.2015.11.009.
[71] M. Holcapek, R. Jirasko, M. Lisa, Recent developments in liquid chromatography-mass spectrometry and related techniques, J. Chromatogr. A 1259 (2012) 3–15. https://doi.org/10.1016/j.chroma.2012.08.072.
[72] E. Defossez, J. Bourquin, S. Reuss, S. Rasmann, G. Glauser, Eight key rules for successful data-dependent acquisition in mass spectrometry-based metabolomics, Mass Spectrom. Rev. (2021). https://doi.org/10.1002/mas.21715. in press.
[73] Y.D. Yin, R.H. Wang, Y.P. Cai, Z.Z. Wang, Z.J. Zhu, DecoMetDIA: deconvolution of multiplexed MS/MS spectra for metabolite identification in SWATH-MS-based untargeted metabolomics, Anal. Chem. 91 (2019) 11897–11904. https://doi.org/10.1021/acs.analchem.9b02655.
[74] E. Stancliffe, M. Schwaiger-Haber, M. Sindelar, G.J. Patti, DecoID improves

identification rates in metabolomics through database-assisted MS/MS deconvolution, Nat. Methods 18 (2021) 779–787. https://doi.org/10.1038/s41592-021-01195-3.
[75] H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. VanderGheynst, O. Fiehn, M. Arita, MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis, Nat. Methods 12 (2015) 523–526. https://doi.org/10.1038/nmeth.3393.
[76] I. Tada, R. Chaleckis, H. Tsugawa, I. Meister, P. Zhang, N. Lazarinis, B. Dahlen, C.E. Wheelock, M. Arita, Correlation-based deconvolution (CorrDec) to generate high-quality MS2 spectra from data-independent acquisition in multisample studies, Anal. Chem. 92 (2020) 11310–11317. https://doi.org/10.1021/acs.analchem.0c01980.
[77] H. Tsugawa, M. Arita, M. Kanazawa, A. Ogiwara, T. Bamba, E. Fukusaki, MRMPROBS: a data assessment and metabolite identification tool for large-scale multiple reaction monitoring based widely targeted metabolomics, Anal. Chem. 85 (2013) 5191–5199. https://doi.org/10.1021/ac400515s.
[78] D.B. Kell, S.G. Oliver, The metabolome 18 years on: a concept comes of age, Metabolomics 12 (2016) 148. https://doi.org/10.1007/s11306-016-1108-4.
[79] J. Boccard, J.L. Veuthey, S. Rudaz, Knowledge discovery in metabolomics: an overview of MS data handling, J. Separ. Sci. 33 (2010) 290–304. https://doi.org/10.1002/jssc.200900609.
[80] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, XCMS: processing mass spectrometry data for metabolite profiling using Nonlinear peak alignment, matching, and identification, Anal. Chem. 78 (2006) 779–787. https://doi.org/10.1021/ac051437y.
[81] R. Tautenhahn, G.J. Patti, D. Rinehart, G. Siuzdak, XCMS Online: a web-based platform to process untargeted metabolomic data, Anal. Chem. 84 (2012) 5035–5039. https://doi.org/10.1021/ac300698c.
[82] T. Pluskal, S. Castillo, A. Villar-Briones, M. Oresic, MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, BMC Bioinf. 11 (2010) 395. https://doi.org/10.1186/1471-2105-11-395.
[83] B.B. Misra, New software tools, databases, and resources in metabolomics: updates from 2020, Metabolomics 17 (2021) 49. https://doi.org/10.1007/S11306-021-01796-1.
[84] J.B. Coble, C.G. Fraga, Comparative evaluation of preprocessing freeware on chromatography/mass spectrometry data for signature discovery, J. Chromatogr. A 1358 (2014) 155–164. https://doi.org/10.1016/j.chroma.2014.06.100.
[85] Z.C. Li, Y. Lu, Y.F. Guo, H.J. Cao, Q.H. Wang, W.Q. Shui, Comprehensive evaluation of untargeted metabolomics data processing software in feature detection, quantification and discriminating marker selection, Anal. Chim. Acta 1029 (2018) 50–57. https://doi.org/10.1016/j.aca.2018.05.001.
[86] O.D. Myers, S.J. Sumner, S.Z. Li, S. Barnes, X.X. Du, Detailed investigation and comparison of the XCMS and MZmine 2 chromatogram construction and chromatographic peak detection methods for preprocessing mass spectrometry metabolomics data, Anal. Chem. 89 (2017) 8689–8695. https://doi.org/10.1021/acs.analchem.7b01069.
[87] S. Stein, Mass spectral reference libraries: an ever-expanding resource for chemical identification, Anal. Chem. 84 (2012) 7274–7282. https://doi.org/10.1021/ac301205z.
[88] M. Vinaixa, E.L. Schymanski, S. Neumann, M. Navarro, R.M. Salek, O. Yanes, Mass spectral databases for LC/MS- and GC/MS-based metabolomics: state of the field and future prospects, TrAC-Trend Anal. Chem. 78 (2016) 23–35. https://doi.org/10.1016/j.trac.2015.09.005.
[89] Y.Y. Li, T. Kind, J. Folz, A. Vaniya, S.S. Mehta, O. Fiehn, Spectral entropy outperforms MS/MS dot product similarity for small-molecule compound identification, Nat. Methods 18 (2021) 1524–1531. https://doi.org/10.1038/s41592-021-01331-z.
[90] T. Kind, H. Tsugawa, T. Cajka, Y. Ma, Z.J. Lai, S.S. Mehta, G. Wohlgemuth, D.K. Barupal, M.R. Showalter, M. Arita, O. Fiehn, Identification of small molecules using accurate mass MS/MS search, Mass Spectrom. Rev. 37 (2018) 513–532. https://doi.org/10.1002/mas.21535.
[91] J.C. Xue, C. Guijas, H.P. Benton, B. Warth, G. Siuzdak, METLIN MS2 molecular standards database: a broad chemical and biological resource, Nat. Methods 17 (2020) 953–954. https://doi.org/10.1038/s41592-020-0942-5.
[92] Z. Vankova, O. Peterka, M. Chocholouskova, D. Wolrab, R. Jirasko, M. Holcapek, Retention dependences support highly confident identification of lipid species in human plasma by reversed-phase UHPLC/MS, Anal. Bioanal. Chem. 414 (2022) 319–331. https://doi.org/10.1007/s00216-021-03492-4.
[93] P. Purwaha, L.P. Silva, D.H. Hawke, J.N. Weinstein, P.L. Lorenzi, An artifact in LC-MS/MS measurement of glutamine and glutamic acid: in-source cyclization to pyroglutamic acid, Anal. Chem. 86 (2014) 5633–5637. https://doi.org/10.1021/ac501451v.
[94] W.Y. Lu, X.Y. Su, M.S. Klein, I.A. Lewis, O. Fiehn, J.D. Rabinowitz, Metabolite measurement: pitfalls to avoid and practices to follow, Annu. Rev. Biochem. 86 (2017) 277–304. https://doi.org/10.1146/annurev-biochem-061516-044952.
[95] C.F. Hu, W.Q. Luo, J. Xu, X.L. Han, Recognition and avoidance of ion source-generated artifacts in lipidomics analysis, Mass Spectrom. Rev. 41 (2022) 15–31. https://doi.org/10.1002/mas.21659.
[96] X.Y. Wang, M.V. Schmitt, L.N. Xu, Y.P. Jiao, L.J. Guo, P. Lienau, A. Reichel, X.H. Liu, Quantitative molecular tissue atlas of bis(monoacylglycero)phosphate and phosphatidylglycerol membrane lipids in rodent organs generated by methylation assisted high resolution mass spectrometry, Anal. Chim. Acta 1084 (2019) 60–70. https://doi.org/10.1016/j.aca.2019.07.060.
[97] M. Wang, J.P. Palavicini, A. Csereszny, X.L. Han, Strategy for quantitative analysis of isomeric bis(monoacylglycero)phosphate and phosphatidylglycerol species by shotgun lipidomics after one-step methylation, Anal. Chem. 89 (2017) 8490–8495. https://doi.org/10.1021/acs.analchem.7b02058.
[98] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W. Fan, O. Fiehn, R. Goodacre, J.L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A.N. Lane, J.C. Lindon, P. Marriott, A.W. Nicholls, M.D. Reily, J.J. Thaden, M.R. Viant, Proposed minimum reporting standards for chemical analysis chemical analysis working group (CAWG) metabolomics standards initiative (MSI), Metabolomics 3 (2007) 211–221. https://doi.org/10.1007/s11306-007-0082-2.
[99] N.A. Reisdorph, S. Walmsley, R. Reisdorph, A perspective and framework for developing sample type specific databases for LC/MS-based clinical metabolomics, Metabolites 10 (2020) 8. https://doi.org/10.3390/Metabo10010008.
[100] E.L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H.P. Singer, J. Hollender, Identifying small molecules via high resolution mass spectrometry: communicating confidence, Environ. Sci. Technol. 48 (2014) 2097–2098. https://doi.org/10.1021/es5002105.
[101] G. Liebisch, E. Fahy, J. Aoki, E.A. Dennis, T. Durand, C.S. Ejsing, M. Fedorova, I. Feussner, W.J. Griffiths, H. Kofeler, A.H. Merrill, R.C. Murphy, V.B. O'Donnell, O. Oskolkova, S. Subramaniam, M.J.O. Wakelam, F. Spener, Update on LIPID MAPS classification, nomenclature, and shorthand notation for MS-derived lipid structures, J. Lipid Res. 61 (2020) 1539–1555. https://doi.org/10.1194/jlr.S120001025.
[102] J.P. Koelmel, C.Z. Ulmer, C.M. Jones, R.A. Yost, J.A. Bowden, Common cases of improper lipid annotation using high resolution tandem mass spectrometry data and corresponding limitations in biological interpretation, BBA-Mol. Cell Biol. L. 1862 (2017). https://doi.org/10.1016/j.bbalip.2017.06.013, 1024-1024.
[103] Z.W. Zhou, M.D. Luo, X. Chen, Y.D. Yin, X. Xiong, R.H. Wang, Z.J. Zhu, Ion mobility collision cross-section atlas for known and unknown metabolite annotation in untargeted metabolomics, Nat. Commun. 11 (2020) 4334. https://doi.org/10.1038/S41467-020-18171-8.
[104] K.L. Leaptrot, J.C. May, J.N. Dodds, J.A. McLean, Ion mobility conformational lipid atlas for high confidence lipidomics, Nat. Commun. 10 (2019) 985. https://doi.org/10.1038/s41467-019-08897-5.
[105] H.C. Kofeler, T.O. Eichmann, R. Ahrends, J.A. Bowden, N. Danne-Rasche, E.A. Dennis, M. Fedorova, W.J. Griffiths, X.L. Han, J. Hartler, M. Holcapek, R. Jirasko, J.P. Koelmel, C.S. Ejsing, G. Liebisch, Z. Ni, V.B. O'Donnell, O. Quehenberger, D. Schwudke, A. Shevchenko, M.J.O. Wakelam, M.R. Wenk, D. Wolrab, K. Ekroos, Quality control requirements for the correct annotation of lipidomics data, Nat. Commun. 12 (2021) 4771. https://doi.org/10.1038/S41467-021-24984-Y.
[106] C.G. Vasilopoulou, K. Sulek, A.D. Brunner, N.S. Meitei, U. Schweiger-Hufnagel, S.W. Meyer, A. Barsch, M. Mann, F. Meier, Trapped ion mobility spectrometry and PASEF enable in-depth lipidomics from minimal sample amounts, Nat. Commun. 11 (2020) 331. https://doi.org/10.1038/s41467-019-14044-x.
[107] K. Uppal, D.I. Walker, K. Liu, S. Li, Y.M. Go, D.P. Jones, Computational metabolomics: a framework for the million metabolome, Chem. Res. Toxicol. 29 (2016) 1956–1975. https://doi.org/10.1021/acs.chemrestox.6b00179.
[108] R.R. da Silva, P.C. Dorrestein, R.A. Quinn, Illuminating the dark matter in metabolomics, P. Natl. Acad. Sci. USA 112 (2015) 12549–12550. https://doi.org/10.1073/pnas.1516878112.
[109] K.K. Dennis, D.P. Jones, The exposome: a new frontier for education, Am. Biol. Teach. 78 (2016) 542–548. https://doi.org/10.1525/abt.2016.78.7.542.
[110] T. Kind, O. Fiehn, Advances in structure elucidation of small molecules using mass spectrometry, Bioanal. Rev. 2 (2010) 23–60. https://doi.org/10.1007/s12566-010-0015-9.
[111] H. Tsugawa, T. Kind, R. Nakabayashi, D. Yukihira, W. Tanaka, T. Cajka, K. Saito, O. Fiehn, M. Arita, Hydrogen rearrangement rules: computational MS/MS fragmentation and structure elucidation using MS-FINDER software, Anal. Chem. 88 (2016) 7946–7958. https://doi.org/10.1021/acs.analchem.6b00770.
[112] B.T. Cooper, X.J. Yan, Y. Simon-Manso, D.V. Tchekhovskoi, Y.A. Mirokhin, S.E. Stein, Hybrid search: a method for identifying metabolites absent from tandem mass spectrometry libraries, Anal. Chem. 91 (2019) 13924–13932. https://doi.org/10.1021/acs.analchem.9b03415.
[113] S. Bocker, Searching molecular structure databases using tandem MS data: are we there yet? Curr. Opin. Chem. Biol. 36 (2017) 1–6. https://doi.org/10.1016/j.cbpa.2016.12.010.
[114] J.L. Little, A.J. Williams, A. Pshenichnov, V. Tkachenko, Identification of "known unknowns" utilizing accurate mass data and ChemSpider, J. Am. Soc. Mass Spectrom. 23 (2012) 179–185. https://doi.org/10.1007/s13361-011-0265-y.
[115] D.Q. Liu, L.M. Wu, M.J. Sun, P.A. MacGregor, On-line H/D exchange LC-MS strategy for structural elucidation of pharmaceutical impurities, J. Pharmaceut. Biomed. 44 (2007) 320–329. https://doi.org/10.1016/j.jpba.2007.01.019.
[116] C.R. Fischer, P. Wilmes, B.P. Bowen, T.R. Northen, J.F. Banfield, Deuterium-exchange metabolomics identifies N-methyl lyso phosphatidylethanolamines as abundant lipids in acidophilic mixed microbial communities, Metabolomics 8 (2012) 566–578. https://doi.org/10.1007/s11306-011-0344-x.

[117] T. De Vijlder, D. Valkenborg, F. Lemiere, E.P. Romijn, K. Laukens, F. Cuyckens, A tutorial in small molecule identification via electrospray ionization-mass spectrometry: the practical art of structural elucidation, Mass Spectrom. Rev. 37 (2018) 607–629. https://doi.org/10.1002/mas.21551.
[118] Q.B. Zhang, L.A. Ford, A.M. Evans, D.R. Toal, Structure elucidation of metabolite x17299 by interpretation of mass spectrometric data, Metabolomics 13 (2017) 92. https://doi.org/10.1007/s11306-017-1231-x.
[119] Q.B. Zhang, L.A. Ford, A.M. Evans, D.R. Toal, Identification of an endogenous organosulfur metabolite by interpretation of mass spectrometric data, Org. Lett. 20 (2018) 2100–2103. https://doi.org/10.1021/acs.orglett.8b00664.
[120] M. Witting, S. Bocker, Current status of retention time prediction in metabolite identification, J. Separ. Sci. 43 (2020) 1746–1754. https://doi.org/10.1002/jssc.202000060.
[121] P. Bonini, T. Kind, H. Tsugawa, D.K. Barupal, O. Fiehn, Retip: retention time prediction for compound annotation in untargeted metabolomics, Anal. Chem. 92 (2020) 7515–7522. https://doi.org/10.1021/acs.analchem.9b05765.
[122] C W.B. Dunn, D. Broadhurst, P. Begley, E. Zelena, S. Francis-McIntyre, N. Anderson, M. Brown, J.D. Knowles, A. Halsall, J.N. Haselden, A.W. Nicholls, I.D. Wilson, D.B. Kell, R. Goodacre, Human Serum Metabolome, Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry, Nat. Protoc. 6 (2011) 1060–1083. https://doi.org/10.1038/nprot.2011.335.
[123] F. Fernandez-Albert, R. Llorach, M. Garcia-Aloy, A. Ziyatdinov, C. Andres-Lacueva, A. Perera, Intensity drift removal in LC/MS metabolomics by common variance compensation, Bioinformatics 30 (2014) 2899–2905. https://doi.org/10.1093/bioinformatics/btu423.
[124] A.M. De Livera, M. Sysi-Aho, L. Jacob, J.A. Gagnon-Bartsch, S. Castillo, J.A. Simpson, T.P. Speed, Statistical methods for handling unwanted variation in metabolomics data, Anal. Chem. 87 (2015) 3606–3615. https://doi.org/10.1021/ac502439y.
[125] K. O'Shea, B.B. Misra, Software tools, databases and resources in metabolomics: updates from 2018 to 2019, Metabolomics 16 (2020) 36. https://doi.org/10.1007/S11306-020-01657-3.
[126] T. Matsuo, H. Tsugawa, H. Miyagawa, E. Fukusaki, Integrated strategy for unknown EI-MS identification using quality control calibration curve, multivariate analysis, EI-MS spectral database, and retention index prediction, Anal. Chem. 89 (2017) 6766–6773. https://doi.org/10.1021/acs.analchem.7b01010.
[127] S.L. Fan, T. Kind, T. Cajka, S.L. Hazen, W.H.W. Tang, R. Kaddurah-Daouk, M.R. Irvin, D.K. Arnett, D.K. Barupal, O. Fiehn, Systematic error removal using random forest for normalizing large-scale untargeted lipidomics data, Anal. Chem. 91 (2019) 3590–3596. https://doi.org/10.1021/acs.analchem.8b05592.
[128] J.A. Kirwan, H. Gika, R.D. Beger, D. Bearden, W.B. Dunn, R. Goodacre, G. Theodoridis, M. Witting, L.R. Yu, I.D. Wilson, Quality assurance and quality control reporting in untargeted metabolic phenotyping: mQACC recommendations for analytical quality management, Metabolomics 18 (2022) 70. https://doi.org/10.1007/S11306-022-01926-3.
[129] M. Ghorasaini, Y. Mohammed, J. Adamski, L. Bettcher, J.A. Bowden, M. Cabruja, K. Contrepois, M. Ellenberger, B. Gajera, M. Haid, D. Hornburg, C. Hunter, C.M. Jones, T. Klein, O. Mayboroda, M. Mirzaian, R. Moaddel, L. Ferrucci, J. Lovett, K. Nazir, M. Pearson, B.K. Ubhi, D. Raftery, F. Riols, R. Sayers, E.J.G. Sijbrands, M.P. Snyder, B.L. Su, V. Velagapudi, K.J. Williams, Y.B. de Rijke, M. Giera, Cross-laboratory standardization of preclinical lipidomics using differential mobility spectrometry and multiple reaction monitoring, Anal. Chem. 93 (2021) 16369–16378. https://doi.org/10.1021/acs.analchem.1c02826.
[130] J.A. Bowden, A. Heckert, C.Z. Ulmer, C.M. Jones, J.P. Koelmel, L. Abdullah, L. Ahonen, Y. Alnouti, A.M. Armando, J.M. Asara, T. Bamba, J.R. Barr, J. Bergquist, C.H. Borchers, J. Brandsma, S.B. Breitkopf, T. Cajka, A. Cazenave-Gassiot, A. Checa, M.A. Cinel, R.A. Colas, S. Cremers, E.A. Dennis, J.E. Evans, A. Fauland, O. Fiehn, M.S. Gardner, T.J. Garrett, K.H. Gotlinger, J. Han, Y.Y. Huang, A.H.P. Neo, T. Hyotylainen, Y. Izumi, H.F. Jiang, H.L. Jiang, J. Jiang, M. Kachman, R. Kiyonami, K. Klavins, C. Klose, H.C. Kofeler, J. Kolmert, T. Koal, G. Koster, Z. Kuklenyik, I.J. Kurland, M. Leadley, K. Lin, K.R. Maddipati, D. McDougall, P.J. Meikle, N.A. Mellett, C. Monnin, M.A. Moseley, R. Nandakumar, M. Oresic, R. Patterson, D. Peake, J.S. Pierce, M. Post, A.D. Postle, R. Pugh, Y.P. Qiu, O. Quehenberger, P. Ramrup, J. Rees, B. Rembiesa, D. Reynaud, M.R. Roth, S. Sales, K. Schuhmann, M.L. Schwartzman, C.N. Serhan, A. Shevchenko, S.E. Somerville, L.S. John-Williams, M.A. Surma, H. Takeda, R. Thakare, J.W. Thompson, F. Torta, A. Triebl, M. Trotzmuller, S.J.K. Ubhayasekera, D. Vuckovic, J.M. Weir, R. Welti, M.R. Wenk, C.E. Wheelock, L.B. Yao, M. Yuan, X.Q.H. Zhao, S.L. Zhou, Harmonizing lipidomics: NIST interlaboratory comparison exercise for lipidomics using SRM 1950-Metabolites in Frozen Human Plasma, J. Lipid Res. 58 (2017) 2275–2288. https://doi.org/10.1194/jlr.M079012.
[131] K.A. Lippa, J.J. Aristizabal-Henao, R.D. Beger, J.A. Bowden, C. Broeckling, C. Beecher, W. Clay Davis, W.B. Dunn, R. Flores, R. Goodacre, G.J. Gouveia, A.C. Harms, T. Hartung, C.M. Jones, M.R. Lewis, I. Ntai, A.J. Percy, D. Raftery, T.B. Schock, J. Sun, G. Theodoridis, F. Tayyari, F. Torta, C.Z. Ulmer, I. Wilson, B.K. Ubhi, Reference materials for MS-based untargeted metabolomics and lipidomics: a review by the metabolomics quality assurance and quality control consortium (mQACC), Metabolomics 18 (2022) 24. https://doi.org/10.1007/s11306-021-01848-6.
[132] B.B. Misra, Data normalization strategies in metabolomics: current challenges, approaches, and tools, Eur. J. Mass Spectrom. 26 (2020) 165–174. https://doi.org/10.1177/1469066720918446.
[133] M.M. Khamis, T. Holt, H. Awad, A. El-Aneed, D.J. Adamko, Comparative analysis of creatinine and osmolality as urine normalization strategies in targeted metabolomics for the differential diagnosis of asthma and COPD, Metabolomics 14 (2018) 115. https://doi.org/10.1007/S11306-018-1418-9.
[134] C. Muschet, G. Moller, C. Prehn, M.H. de Angelis, J. Adamski, J. Tokarz, Removing the bottlenecks of cell culture metabolomics: fast normalization procedure, correlation of metabolites to cell number, and impact of the cell harvesting method, Metabolomics 12 (2016) 151. https://doi.org/10.1007/S11306-016-1104-8.
[135] P. Cuevas-Delgado, D. Dudzik, V. Miguel, S. Lamas, C. Barbas, Data-dependent normalization strategies for untargeted metabolomics-a case study, Anal. Bioanal. Chem. 412 (2020) 6391–6405. https://doi.org/10.1007/s00216-020-02594-9.
[136] M. Modrzejewska, M. Gawronski, D. Gackowski, Normalization of metabolic data to total thymine content and its application to determination of 2-hydroxyglutarate, Anal. Biochem. 618 (2021), 114129. https://doi.org/10.1016/J.Ab.2021.114129.
[137] A. Hutschenreuther, A. Kiontke, G. Birkenmeier, C. Birkemeyer, Comparison of extraction conditions and normalization approaches for cellular metabolomics of adherent growing cells with GC-MS, Anal. Methods 4 (2012) 1953–1963. https://doi.org/10.1039/c2ay25046b.
[138] C.H. Johnson, J. Ivanisevic, G. Siuzdak, Metabolomics: beyond biomarkers and towards mechanisms, Nat. Rev. Mol. Cell Biol. 17 (2016) 451–459. https://doi.org/10.1038/nrm.2016.25.
[139] P. Sedgwick, Multiple significance tests: the Bonferroni correction, Br. Med. J. 344 (2012) e509. https://doi.org/10.1136/bmj.e509.
[140] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate - a practical and powerful approach to multiple testing, J. Roy. Stat. Soc. B Met. 57 (1995) 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
[141] J. Xia, D.S. Wishart, Using MetaboAnalyst 3.0 for comprehensive metabolomics data analysis, Curr. Protoc. Bioinformatics 55 (2016). https://doi.org/10.1002/cpbi.11, 14.10.11-14.10.91.
[142] S. Ren, A.A. Hinzman, E.L. Kang, R.D. Szczesniak, L.J. Lu, Computational and statistical analysis of metabolomics data, Metabolomics 11 (2015) 1492–1513. https://doi.org/10.1007/s11306-015-0823-6.
[143] D. Grapov, K. Wanichthanarak, O. Fiehn, MetaMapR: pathway independent metabolomic network analysis incorporating unknowns, Bioinformatics 31 (2015) 2757–2760. https://doi.org/10.1093/bioinformatics/btv194.
[144] K. Uppal, Q.A. Soltow, D.E. Promislow, L.M. Wachtman, A.A. Quyyumi, D.P. Jones, MetabNet: an R package for metabolic association analysis of high-resolution metabolomics data, Front. Bioeng. Biotechnol. 3 (2015) 87. https://doi.org/10.3389/fbioe.2015.00087.
[145] M.X. Wang, J.J. Carver, V.V. Phelan, L.M. Sanchez, N. Garg, Y. Peng, D.D. Nguyen, J. Watrous, C.A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A.V. Melnik, M.J. Meehan, W.T. Liu, M. Criisemann, P.D. Boudreau, E. Esquenazi, M. Sandoval-Calderon, R.D. Kersten, L.A. Pace, R.A. Quinn, K.R. Duncan, C.C. Hsu, D.J. Floros, R.G. Gavilan, K. Kleigrewe, T. Northen, R.J. Dutton, D. Parrot, E.E. Carlson, B. Aigle, C.F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B.T. Murphy, L. Gerwick, C.C. Liaw, Y.L. Yang, H.U. Humpf, M. Maansson, R.A. Keyzers, A.C. Sims, A.R. Johnson, A.M. Sidebottom, B.E. Sedio, A. Klitgaard, C.B. Larson, C.A. Boya, D. Torres-Mendoza, D.J. Gonzalez, D.B. Silva, L.M. Marques, D.P. Demarque, E. Pociute, E.C. O'Neill, E. Briand, E.J.N. Helfrich, E.A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J.J. Kharbush, Y. Zeng, J.A. Vorholt, K.L. Kurita, P. Charusanti, K.L. McPhail, K.F. Nielsen, L. Vuong, M. Elfeki, M.F. Traxler, N. Engene, N. Koyama, O.B. Vining, R. Baric, R.R. Silva, S.J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P.G. Williams, J.Q. Dai, R. Neupane, J. Gurr, A.M.C. Rodriguez, A. Lamsa, C. Zhang, K. Dorrestein, B.M. Duggan, J. Almaliti, P.M. Allard, P. Phapale, L.F. Nothias, T. Alexandrovr, M. Litaudon, J.L. Wolfender, J.E. Kyle, T.O. Metz, T. Peryea, D.T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Muller, K.M. Waters, W.Y. Shi, X.T. Liu, L.X. Zhang, R. Knight, P.R. Jensen, B.O. Palsson, K. Pogliano, R.G. Linington, M. Gutierrez, N.P. Lopes, W.H. Gerwick, B.S. Moore, P.C. Dorrestein, N. Bandeira, Sharing and community curation of mass spectrometry data with global natural products social molecular networking, Nat. Biotechnol. 34 (2016) 828–837. https://doi.org/10.1038/nbt.3597.
[146] J.J.J. van der Hooft, J. Wandy, M.P. Barrett, K.E.V. Burgess, S. Rogers, Topic modeling for untargeted substructure exploration in metabolomics, P. Natl. Acad. Sci. USA 113 (2016) 13738–13743. https://doi.org/10.1073/pnas.1608041113.
[147] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (2003) 2498–2504. https://doi.org/10.1101/gr.1239303.
[148] Z.Q. Pang, J. Chong, G.Y. Zhou, D.A.D. Morais, L. Chang, M. Barrette, C. Gauthier, P.E. Jacques, S.Z. Li, J.G. Xia, MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights, Nucleic Acids Res. 49 (2021) W388–W396. https://doi.org/10.1093/nar/gkab382.
[149] X.T. Shen, Z.J. Zhu, MetFlow: an interactive and integrated workflow for metabolomics data cleaning and differential metabolite discovery, Bioinformatics 35 (2019) 2870–2872. https://doi.org/10.1093/bioinformatics/

bty1066.
[150] M.R. Molenaar, A. Jeucken, T.A. Wassenaar, C.H.A. van de Lest, J.F. Brouwers, J.B. Helms, LION/web: a web-based ontology enrichment tool for lipidomic data analysis, GigaScience 8 (2019) giz061. https://doi.org/10.1093/gigascience/giz061.
[151] D.K. Barupal, O. Fiehn, Chemical Similarity Enrichment Analysis (ChemRICH) as alternative to biochemical pathway mapping for metabolomic datasets, Sci. Rep. 7 (2017). https://doi.org/10.1038/S41598-017-15231-W.
[152] J. Ding, J. Ji, Z. Rabow, T. Shen, J. Folz, C.R. Brydges, S.L. Fan, X.C. Lu, S. Mehta, M.R. Showalter, Y. Zhang, R. Araiza, L.R. Bower, K.C.K. Lloyd, O. Fiehn, A metabolome atlas of the aging mouse brain, Nat. Commun. 12 (2021). https://doi.org/10.1038/S41467-021-26310-Y.
[153] M. Lopes, K. Brejchova, M. Riecan, M. Novakova, M. Rossmeisl, T. Cajka, O. Kuda, Metabolomics atlas of oral 13C-glucose tolerance test in mice, Cell Rep. 37 (2021). https://doi.org/10.1016/J.Celrep.2021.109833.
[154] X.S. Du, J.J. Aristizabal-Henao, T.J. Garrett, M. Brochhausen, W.R. Hogan, D.J. Lemas, A Checklist for reproducible computational analysis in clinical metabolomics research, Metabolites 12 (2022). https://doi.org/10.3390/Metabo12010087.
[155] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.W. Boiten, L.B.D. Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, Comment: the FAIR Guiding Principles for scientific data management and stewardship, Sci. Data 3 (2016), 160018. https://doi.org/10.1038/Sdata.2016.18.
[156] K. Haug, K. Cochrane, V.C. Nainala, M. Williams, J.K. Chang, K.V. Jayaseelan, C. O'Donovan, MetaboLights: a resource evolving in response to the needs of its scientific community, Nucleic Acids Res. 48 (2020) D440–D444. https://doi.org/10.1093/nar/gkz1019.
[157] M. Sud, E. Fahy, D. Cotter, K. Azam, I. Vadivelu, C. Burant, A. Edison, O. Fiehn, R. Higashi, K.S. Nair, S. Sumner, S. Subramaniam, Metabolomics Workbench: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools, Nucleic Acids Res. 44 (2016) D463–D470. https://doi.org/10.1093/nar/gkv1042.
[158] K.A. Dyar, D. Lutter, A. Artati, N.J. Ceglia, Y. Liu, D. Armenta, M. Jastroch, S. Schneider, S. de Mateo, M. Cervantes, S. Abbondante, P. Tognini, R. Orozco-Solis, K. Kinouchi, C. Wang, R. Swerdloff, S. Nadeef, S. Masri, P. Magistretti, V. Orlando, E. Borrelli, N.H. Uhlenhaut, P. Baldi, J. Adamski, M.H. Tschop, K. Eckel-Mahan, P. Sassone-Corsi, Atlas of circadian metabolism reveals system-wide coordination and communication between clocks, Cell 174 (2018) 1571–1585. https://doi.org/10.1016/j.cell.2018.08.042.
[159] V. Talman, J. Teppo, P. Poho, P. Movahedi, A. Vaikkinen, S.T. Karhu, K. Trost, T. Suvitaival, J. Heikkonen, T. Pahikkala, T. Kotiaho, R. Kostiainen, M. Varjosalo, H. Ruskoaho, Molecular atlas of postnatal mouse heart development, J. Am. Heart Assoc. 7 (2018), e010378. https://doi.org/10.1161/JAHA.118.010378.
[160] M. Lange, G. Angelidou, Z.X. Ni, A. Criscuolo, J. Schiller, M. Bluher, M. Fedorova, AdipoAtlas: a reference lipidome for human white adipose tissue, Cell Rep. Med. 2 (2021), 100407. https://doi.org/10.1016/J.Xcrm.2021.100407.
[161] S. Sato, K.A. Dyar, J.T. Treebak, S.L. Jepsen, A.M. Ehrlich, S.P. Ashcroft, K. Trost, T. Kunzke, V.M. Prade, L. Small, A.L. Basse, M. Schonke, S. Chen, M. Samad, P. Baldi, R. Barres, A. Walch, T. Moritz, J.J. Holst, D. Lutter, J.R. Zierath, P. Sassone-Corsi, Atlas of exercise metabolism reveals time-dependent signatures of metabolic homeostasis, Cell Metabol. 34 (2022) 329–345. https://doi.org/10.1016/j.cmet.2021.12.016.
[162] K.F. Dekkers, S. Sayols-Baixeras, G. Baldanzi, C. Nowak, U. Hammar, D. Nguyen, G. Varotsis, L. Brunkwall, N. Nielsen, A.C. Eklund, J. Bak Holm, H.B. Nielsen, F. Ottosson, Y.T. Lin, S. Ahmad, L. Lind, J. Sundstrom, G. Engstrom, J.G. Smith, J. Arnlov, M. Orho-Melander, T. Fall, An online atlas of human plasma metabolite signatures of gut microbiome composition, Nat. Commun. 13 (2022) 5370. https://doi.org/10.1038/s41467-022-33050-0.
[163] M. Sugimoto, S. Ikeda, K. Niigata, M. Tomita, H. Sato, T. Soga, MMMDB: mouse multiple tissue metabolome database, Nucleic Acids Res. 40 (2012) D809–D814. https://doi.org/10.1093/nar/gkr1170.
[164] H. Julkunen, A. Cichońska, M. Tiainen, H. Koskela, K. Nybo, V. Mäkelä, J. Nokso-Koivisto, K. Kristiansson, M. Perola, V. Salomaa, P. Jousilahti, A. Lundqvist, A.J. Kangas, P. Soininen, J.C. Barrett, P. Würtz, Atlas of Plasma Nuclear Magnetic Resonance Biomarkers for Health and Disease in 118,461 Individuals from the UK Biobank, 2022. https://doi.org/10.1101/2022.06.13.22276332.
[165] M. Hunter, N.J. Demarais, R.L.M. Faull, A.C. Grey, M.A. Curtis, An imaging mass spectrometry atlas of lipids in the human neurologically normal and Huntington's disease caudate nucleus, J. Neurochem. 157 (2021) 2158–2172. https://doi.org/10.1111/jnc.15325.
[166] D.R. Bhandari, Q. Wang, W. Friedt, B. Spengler, S. Gottwald, A. Rompp, High resolution mass spectrometry imaging of plant tissues: towards a plant metabolite atlas, Analyst 140 (2015) 7696–7709. https://doi.org/10.1039/c5an01065a.
[167] G.Q. Wang, B. Heijs, S. Kostidis, A. Mahfouz, R.G.J. Rietjens, R. Bijkerk, A. Koudijs, L.A.K. Van der Pluijm, C.W. Van den Berg, S.J. Dumas, P. Carmeliet, M. Giera, B.M. Van den Berg, T.J. Rabelink, Analyzing cell-type-specific dynamics of metabolism in kidney repair, Nat. Metab. 4 (2022) 1109. https://doi.org/10.1038/s42255-022-00615-8.
[168] L. Capolupo, I. Khven, A.R. Lederer, L. Mazzeo, G. Glousker, S. Ho, F. Russo, J.P. Montoya, D.R. Bhandari, A.P. Bowman, S.R. Ellis, R. Guiet, O. Burri, J. Detzner, J. Muthing, K. Homicsko, F. Kuonen, M. Gilliet, B. Spengler, R.M.A. Heeren, G.P. Dotto, G. La Manno, G. D'Angelo, Sphingolipids control dermal fibroblast heterogeneity, Science 376 (2022) 262. https://doi.org/10.1126/science.abh1623.
[169] Y. Hasin, M. Seldin, A. Lusis, Multi-omics approaches to disease, Genome Biol. 18 (2017) 83. https://doi.org/10.1186/s13059-017-1215-1.
[170] A.D. Panopoulos, O. Yanes, S. Ruiz, Y.S. Kida, D. Diep, R. Tautenhahn, A. Herrerias, E.M. Batchelder, N. Plongthongkum, M. Lutz, W.T. Berggren, K. Zhang, R.M. Evans, G. Siuzdak, J.C.I. Belmonte, The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming, Cell Res. 22 (2012) 168–177. https://doi.org/10.1038/cr.2011.177.
[171] S.L. Leibel, I. Tseu, A.S. Zhou, A. Hodges, J. Yin, C. Bilodeau, O. Goltsis, M. Post, Metabolomic profiling of human pluripotent stem cell differentiation into lung progenitors, iScience 25 (2022), 103797. https://doi.org/10.1016/J.Isci.2022.103797.