



Tips and tricks for LC–MS-based metabolomics and lipidomics analysis

Stanislava Rakusanova, Tomas Cajka*

Institute of Physiology of the Czech Academy of Sciences, Videnska 1083, 14200, Prague, Czech Republic



ARTICLE INFO

Keywords:

Metabolomics
Lipidomics
Liquid chromatography
Mass spectrometry
Data processing
Quality control
Workflow
Troubleshooting
LC–MS

ABSTRACT

Metabolomics and lipidomics are rapidly growing fields, leading to novel discoveries and advancing the understanding of biological processes at the molecular level. However, designing a proper workflow and choosing from countless options can be challenging, especially for beginners in the field. To address this challenge, we provide a comprehensive overview of metabolomics and lipidomics tools and a step-by-step guide that includes “tips and tricks” based on current metabolomics and lipidomics analysis approaches. We include power analysis, sample collection and preparation, separation and detection of metabolites using primarily liquid chromatography–mass spectrometry (LC–MS), processing of raw instrumental files, quality control, statistical analysis, and data sharing. This guide offers practical insights applicable to diverse research areas, covering all the essential steps in metabolomic and lipidomic profiling.

1. Introduction

Metabolomics and lipidomics study low-molecular-weight compounds (typically <2000 Da) in biological matrices such as biofluids, tissues, or cells under different conditions. While metabolomics primarily focuses on water-soluble polar metabolites, such as sugars, amino acids, organic acids, and nucleotides, lipidomics aims to identify and quantify various lipid species [1]. There is a partial overlap between these two omics fields, and in some studies, polar metabolites and complex lipids are also referred to by the general term “metabolomics.”

Understanding how biological processes work can be crucial for biomarker discovery, clinical studies, metabolic phenotyping, physiology, or toxicology. To this end, various techniques for analyzing the metabolome and lipidome are available. These techniques include mass spectrometry (MS) [2,3] and nuclear magnetic resonance (NMR) [4,5]. MS methods are performed as liquid chromatography–MS (LC–MS), gas chromatography–MS (GC–MS), capillary electrophoresis–MS (CE–MS), direct infusion–MS, and also ion mobility–MS (IM–MS). Currently, LC–MS represents the most applied tool for analyzing polar and nonpolar metabolites and offers the highest coverage of the metabolome and lipidome compared to other techniques.

Metabolome and lipidome can be studied using untargeted and targeted methods. Untargeted methods analyze all the detectable metabolites in a sample and are mainly used for novel metabolite discovery and hypothesis-generating studies [6]. Conversely, targeted methods focus on analyzing defined metabolites for hypothesis-driven validation

[6,7]. While untargeted methods provide semiquantitative data (i.e., results are reported as peak areas or heights in arbitrary units), the targeted methods provide quantitative data reported in molar concentrations. In addition, untargeted and targeted methods can be combined, providing quantitative data for selected metabolites while reporting the other metabolites (including unknowns) in semiquantitative terms [8].

Since metabolomics and lipidomics are no longer emerging but rather well-established fields, dozens of untargeted and targeted analytical protocols based on MS or NMR, along with bioinformatics tools, are available. However, the availability of so many tools can be overwhelming for both beginners and experienced professionals in the field. Like other omics fields, metabolomics and lipidomics analyses are prone to biases. Thus, a community effort such as the recently established metabolomics Quality Assurance and Quality Control Consortium (mQACC, mqacc.org) may help standardize metabolomics and lipidomics analyses [9]. In this context, we present a comprehensive overview of metabolomics and lipidomics tools (Fig. 1), accompanied by a step-by-step guide that includes “tips and tricks” to avoid possible pitfalls and optimize workflows for generating reliable data for interpretation.

2. Power analysis and sample size

Determining the sample size (both total and per group) is crucial in study design. Insufficient sample size can result in multiple errors. Interestingly, even small differences with no actual significance might

* Corresponding author.

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

become statistically significant in a larger sample size. On the other hand, with a small number of samples, clinically important effects may appear statistically non-significant [10]. A small sample size can also cause a lack of precision. In contrast, a high sample size might lead to unnecessary waste of resources for minimal information gain. Additionally, ethical restrictions must be considered when determining an appropriate number of samples per group, especially concerning animals [11].

Calculating the minimal sample size is based on power analysis. Power analysis helps to determine the smallest number of samples per group, given a required significance level (usually $\alpha = 0.05$, and 0.1–0.2 for a pilot study), statistical power (usually 0.8), and effect size ($d = 0.8$, 0.5, 0.2 for large, medium, small effect size, respectively) [12]. Power analysis can be done before and after data analysis. However, sufficient sample size is required to obtain statistically validated data. Therefore, it is necessary to perform the power analysis before the beginning of the experiment [13]. For instance, G*Power software can provide effect size calculators and graphic options [14].

However, power analysis in untargeted metabolomics and lipidomics studies is challenging because the list of measured metabolites and the effect size are not known *a priori* [15]. Choosing a known metabolite to represent the entire metabolome/lipidome and conducting a univariate method to estimate the sample size can be used, although such statistics might oversimplify the metabolome/lipidome changes. It is often recommended that a pilot study is performed to obtain preliminary data before designing more extensive studies [16].

For metabolomics data, Billoir et al. [17] released an automated implementation of the Data-driven Sample size Determination (DSD) algorithm for MATLAB and GNU Octave, which enables the determination of optimized sample size in metabolic phenotyping studies. Their approach also uses analytical data from a small pilot cohort. On the contrary, Nyamundanda et al. [18] developed MetSizeR to estimate

sample size for metabolomics experiments even when experimental pilot data are unavailable. In 2020, Li et al. [19] introduced the online tool SSizer (idrlab.org/ssizer), enabling the assessment of sample sufficiency and determining the required number of samples for a user-input biological dataset.

2.1. Tips and tricks

- Since conducting power analysis for untargeted methods might be difficult, a general recommendation is used regarding the minimum samples per group to be analyzed. For cells under tightly controlled conditions, a minimum of 5 replicates is recommended [20], although some studies have used even fewer (3–4 replicates). In animal studies, factors such as age, housing, and diet are typically controlled; therefore, sample numbers can also be kept relatively low for ethical and practical reasons [20]. A minimum of 5–10 samples is recommended for biofluids and tissues per group (usually more for tissues due to their high heterogeneity) [21]. In human studies, at least 20–30 samples per class are advised [12,22]. However, the number of samples can range easily from hundreds [23] to even thousands [24] for reasonable statistics, leading to novel biomarker discoveries during clinical studies. Conversely, clinical studies may encounter limited subjects due to budget and recruiting constraints [15].
- The variation of animal and human metabolomes might be influenced (among others) by genetics, age, and sex. However, sex is still not always considered in metabolomics-based investigations, potentially limiting the comprehensive understanding of metabolic profiles and their implications. Therefore, in both animal and human studies, it is recommended to maintain a balanced ratio of male and female subjects to ensure a more accurate and representative exploration of metabolic dynamics [25].

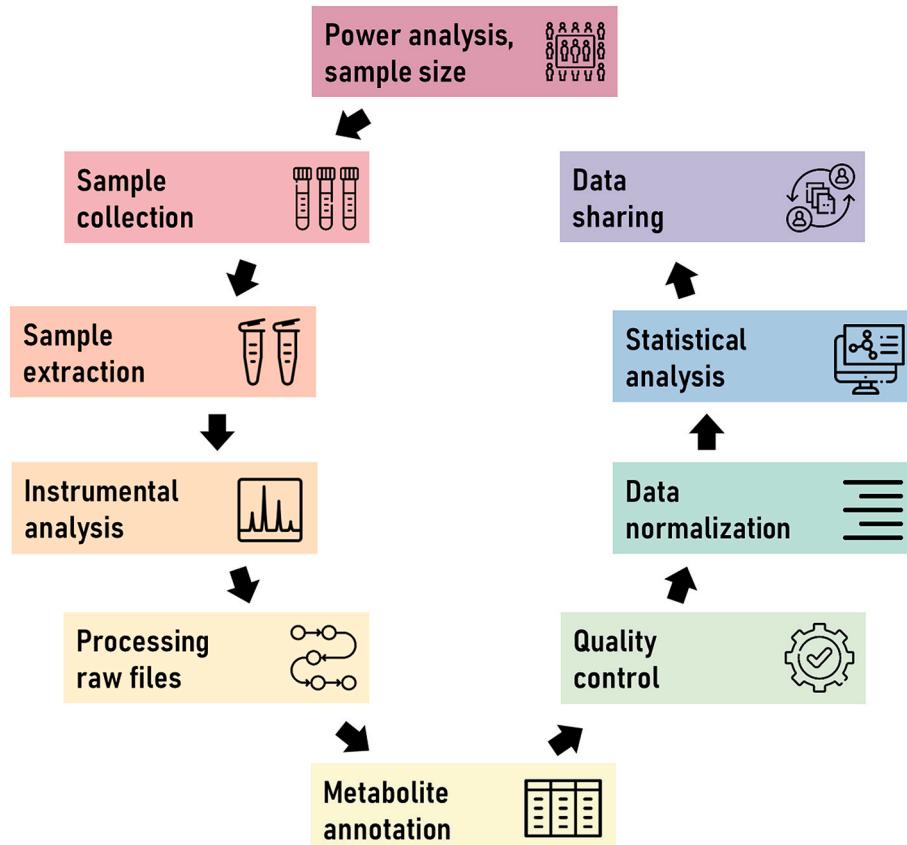


Fig. 1. Metabolomics and lipidomics workflow.

3. Sample collection

Sample collection methods differ depending on the sample type and analysis. The choice of biological material and the amount required for further processing steps are based on the experimental design and tested hypothesis. Incorrect sample collection or storage can lead to degradation of metabolites, high variability, or interferences with instrumentation [26]. In animal studies, specially trained personnel are recommended to use the time between collection and organ dissection as efficiently as possible. Proper training and experience are crucial while collecting multiple biofluids and tissues from all animals [27].

Quenching the metabolism of samples needs to be done as soon as possible, followed by storage of these samples at -80°C . Quenching should terminate all enzyme and chemical activities and avoid the perturbation of existing metabolite levels during harvesting [28]. This process should be done by liquid nitrogen, dry ice, or freeze-clamping. Before tissue collection, it is essential to determine which part of each organ will be sampled. Furthermore, for studies including plasma, the type of coagulant (e.g., ethylenediaminetetraacetic acid, citrate, heparin) must be specified in advance [29].

3.1. Tips and tricks

- Various sizes of plastic tubes are available to deliver biological samples for extraction. Remember that tissues undergo homogenization supported by grinding balls made of stainless steel, tungsten carbide, or zirconium oxide; thus, the use of 2 mL conical bottom tubes is needed for effective homogenization (i.e., 1.5 mL "V" bottom tubes are not helpful and grinding balls may be stuck during homogenization).
- There is an ongoing discussion about whether serum or plasma is the best for metabolomics and lipidomics experiments [30,31]. Remember that serum is prepared from whole blood, allowing the blood to clot by leaving it undisturbed at room temperature (usually 15–60 min, but this step is not always possible to keep under control and can vary based on the personnel's experience and training). On the other hand, plasma is prepared by collecting the whole blood into commercially available anticoagulant-treated tubes (in human studies). For animal studies (e.g., mouse, rat) collecting low volumes ($<100\text{ }\mu\text{L}$) of blood, the quantity of anticoagulant could be too high in tubes for human studies; thus, low-volume commercially anticoagulant-treated tubes or in-house tubes with an appropriate amount of anticoagulant are advised.
- Some metabolites are more prone to degradation or rearrangement, and some are less. Based on the considered analytical platform, remember that some metabolites are extremely unstable and require proper handling and usually a separate (targeted) sample preparation and instrumental platform; for instance, analysis of eicosanoids, endocannabinoids [32], dinucleotide redox cofactors (NADPH/NAD⁺, NADH/NAD⁺) [33], adenine nucleotides (AMP, ADP, and ATP), acyl-coenzymes A [34]. In addition, dry extracts are also prone to thermal reactions and should be stored at low temperatures (-80°C or -24°C) [35].

4. Sample extraction

Effective metabolite extraction is key to successful metabolomics and lipidomics studies. It helps to separate metabolites from undesired compounds and makes samples suitable for instrumental analysis [36]. Biological fluids such as urine can be analyzed directly without sample extraction. Often used steps in urine sample preparation include buffering, dilution, evaporation, or centrifugation. However, these steps may lead to metabolite losses, high salt concentration can lead to ionization suppression and altered adduct formation, and disrupt instrumental performance by forming nonvolatile residues. These complications can be reduced by adding an effective extraction step.

On the other hand, biofluids such as plasma or serum, which contain a wide range of interfering proteins, are not commonly analyzed directly, and an extraction step is added for protein removal [37]. Organic-solvent-based protein precipitation or liquid-liquid extraction (LLE) methods are often used during the extraction. These methods enable the extraction of a wide range of metabolites and simultaneously remove unwanted substances in biological samples, such as bulk proteins and salts [38]. However, in recent years, the all-in-one single extraction methods have been developed to isolate metabolites differing in their physicochemical properties, followed by fractionation. Each fraction is then analyzed under different separation conditions based on the polarity [39].

Single-phase extraction methods have been introduced using an isopropanol or butanol/methanol mixture to simultaneously extract complex lipids and polar metabolites from human plasma [40]. These methods were evaluated based on the metabolome and lipidome coverage, extraction efficiencies, effectiveness of protein precipitation, and reproducibility, and they were suitable for large-scale human population studies. Shortcomings of such methods might be found in ion suppression caused by co-eluting complex lipids in the chromatographic conditions during the analysis of polar metabolites.

Two-phase liquid extraction is commonly used in metabolomics and lipidomics, where one phase consists primarily of nonpolar metabolites (lipids), and the second phase holds mostly polar metabolites. This can be achieved using three solvent systems: methyl *tert*-butyl ether (MTBE)/methanol/water [41], chloroform/methanol/water [42], and dichloromethane/methanol/water [43].

Three-phase liquid extraction is possible using hexane, methyl acetate, acetonitrile, and water. The upper organic phase consists of neutral lipids such as triacylglycerols and cholesteryl esters, the middle organic phase is enriched in glycerophospholipids, and the bottom aqueous phase consists of polar metabolites and proteins [44]. Therefore, proteins need to be removed to analyze polar metabolites.

Along with the already mentioned LLE methods, solid-phase extraction (SPE) and solid-phase microextraction (SPME) methods are available for analyzing biological samples [45].

Besides biofluids, various tissues and cells are analyzed in metabolomics and lipidomics studies. While biofluids are convenient during their analysis, tissues and cells provide deeper insights into metabolomics, making them the focus of clinical research over the years [29].

4.1. Tips and tricks

- Different volumes of biofluids (e.g., plasma, serum, urine) and tissue amounts are needed for their analysis [46]. For untargeted methods, 10–30 μL of biofluids and 5–20 mg of tissues are sufficient. For targeted methods usually covering trace concentrations of some metabolites (e.g., eicosanoids, fatty acyl esters of hydroxy fatty acid), 200–1000 μL of biofluids and 50–200 mg of tissue samples are needed. Remember that it is always good to provide (or keep) backup samples if reanalysis is needed. For cells, usually at least 5×10^5 is needed. For adherent cells, a 6-well plate format is preferred. After completing the cell growth, the medium is removed, followed by washing with buffer (e.g., phosphate-buffered saline) to remove residues of the medium. After removing the washing buffer, the cells are either immediately extracted (with a cold organic solvent or mixture and scraping them) or frozen in liquid nitrogen and stored at -80°C before conducting the extraction. For suspension cells, centrifugation at low speed ($<1000 \times g$) is used first, followed by aspiration and discarding of the culture medium.
- Manual or electronic pipettes have inaccuracy, which depends on the range of a particular pipet and the pipetted volume [47]. For instance, for a 10–100 μL range pipet, the inaccuracy while pipetting 10 μL is usually $\pm 0.3\text{ }\mu\text{L}$ (i.e., 3 %), while for 100 μL $\pm 0.8\text{ }\mu\text{L}$ (i.e., 0.8 %) based on the vendor. However, for a 1–10 μL range pipet while pipetting 10 μL , the inaccuracy is lower $\pm 0.1\text{ }\mu\text{L}$ (i.e., 1 %);

thus, the pipetting volume should be considered when choosing the appropriate pipette for a specific task, taking into account the inaccuracy associated with the pipetted volume and the designated range. Although these errors fall within acceptable limits, regular maintenance, preferably annually, is advisable to ensure pipetting equipment's continued precision and reliability. For weighing the tissue samples, analytical balance (readability 0.1 mg) should be used instead of precision balance (readability 1 mg) since low amounts of tissues (5–20 mg) are often used. A dedicated person should also regularly calibrate the balance to ensure accuracy.

- Biofluid extraction can be performed using either glassware or plastic consumables. Glassware such as tubes, syringes, and pipettes is typically recommended to prevent background signals or the sorption of metabolites. However, when conducting large sample studies, using glassware consumables becomes impractical. In such cases, it is beneficial to evaluate background signals for various types of plastic tubes, typically made of polypropylene, and monitor potential leaking contaminants such as palmitate, stearate, phthalates, and butylated hydroxytoluene. It has been shown that using plastic consumables introduces a competing background signal of palmitate during fluxomics analysis [48]. On the contrary, a study comparing extraction in plastic tubes and glass vials using radioactive lipids found comparable recoveries [49].
- The choice of the extraction method also depends on the expected metabolite coverage. For instance, single-phase extraction can be sufficient for proper coverage of polar metabolites (e.g., extraction using methanol, acetonitrile, a mixture of isopropanol/acetonitrile/water, acetonitrile/methanol [50,51]) or complex lipids (e.g., extraction using isopropanol [52]). However, differences in the number of detected metabolites and method reproducibility should be anticipated [43]. On the other hand, for the simultaneous extraction of polar metabolites and complex lipids, bi-phase extraction using MTBE/methanol/water will ensure proper metabolome and lipidome coverage [53]. In addition, remember that each sample preparation may have a different efficiency in protein removal [54], which impacts the method's robustness, specifically the instrumental part (e.g., column clogging, retention time shifts, worsening separation of isomers).
- When employing two-phase liquid extraction methods, remember that in MTBE (ρ 0.74 g/cm³)/methanol (ρ 0.79 g/cm³)/water (ρ 1.00 g/cm³) protocol [41], the upper phase comprises the organic layer (MTBE and methanol), primarily containing complex lipids. In contrast, the lower phase consists of water (with a small portion of methanol), containing polar metabolites. Conversely, in chloroform (ρ 1.49 g/cm³)/methanol/water [42] and dichloromethane (ρ 1.33 g/cm³)/methanol/water protocols [43], the upper (polar) phase contains polar metabolites, and the bottom (organic) phase is enriched with complex lipids.
- While validation parameters such as selectivity, specificity, matrix effect, range, accuracy, precision, carry-over, dilution integrity, and stability are commonly employed for targeted methods, there are no established guidelines for validating the analytical aspect in an untargeted approach [55]. Instead, emphasis is placed on developing the extraction method with minimal steps, ensuring compatibility with the instrumental technique, and achieving comprehensive coverage of metabolites with high precision. These methods can be assessed by evaluating precision (intra-assay and inter-assay) to calculate the relative standard deviation (RSD) of molecular features (retention time–*m/z* pairs) or already annotated metabolites. Acceptable values are typically close to a 30 % RSD [56]. Furthermore, linearity can be confirmed by diluting the sample used during validation and employing more concentrated extracts to mimic potential ion suppression effects. Method blanks should also be incorporated to identify and address carryover during method development. Any features detected in blanks should be considered background signals [55].

- Using internal standards in the extraction and resuspension solvents ensures proper quality control (QC), such as that identical aliquots were collected from all extracts or that the autosampler injected the correct volume. Reanalysis is needed for samples where internal standards are completely missing or at low intensity compared to the rest of the study.

5. Instrumental analysis

LC–MS is the most popular platform for metabolomics and lipidomics, followed by GC–MS. GC–MS is used mainly to analyze volatile and primary metabolites after derivatization. CE–MS helps analyze polar charged metabolites. Recently, there has also been increased interest in IM–MS. In addition, analysis of polar and nonpolar metabolites can be done with NMR. However, the NMR method does not provide many annotated metabolites compared to MS [27].

Metabolomics and lipidomics detect many metabolites with high chemical diversity and complexity. LC–MS allows for separating and detecting isobars and isomers, reducing ion-suppression effects, and separating compounds according to their physicochemical properties [35]. No single method can cover the true breadth of a metabolome/lipidome. However, LC offers various stationary phases, column dimensions, mobile phase modifiers, and solvents [57]. Thus, different LC–MS separation modes enable the coverage of a wide range of metabolites.

Reversed-phase LC (RPLC) separates polar to mid-polar metabolites, and hydrophilic interaction chromatography (HILIC) separates highly polar metabolites. In addition to RPLC and HILIC, lipidomics also uses normal-phase LC (NPLC) and supercritical fluid chromatography (SFC) for sample separation [58,59]. In RPLC, C18 columns dominate, followed by C8 and C30. HILIC, on the other hand, employs more diverse stationary phase chemistries such as silica, aminopropylsilane, alkyl amide, and sulfobetaine groups [57]. Mobile phases containing water, acetonitrile, or methanol are used to analyze polar metabolites. In contrast, for RPLC-based lipidomics, stronger mobile phases with a high percentage of isopropanol are needed. The separation time usually ranges from 10 to 30 min. Fig. 2 shows the separation of very polar metabolites, demonstrating the benefits of HILIC configuration (high retention and separation of metabolites, Fig. 2A) compared to RPLC (elution of metabolites near the void volume, Fig. 2B).

Ultrahigh-performance LC (UHPLC) systems with sub-2 μ m particles significantly increased LC performance, such as improved speed, resolution, and sensitivity. Currently, LC–MS-based metabolomics and lipidomics studies apply short (50–150 mm) microbore columns (2.1 mm internal diameter, i.d.) with sub-2 μ m particles [7]. The following studies have described different applications [60,61]. Gray et al. [60] employed a 1 mm i.d. LC column rather than a <2.1 mm i.d. column, resulting in equivalent or superior performance in peak capacity, sensitivity, and robustness compared to conventional methodology. This miniaturized method required system optimization, such as reducing dispersion and ensuring appropriate connections to minimize band-broadening. This setup reduced solvent and sample consumption. Conversely, Schönberg et al. [61] employed a propylamine column (50 \times 2 mm i.d., 3 μ m propylamine particles) for the analysis of polar metabolites and a microLC column (100 \times 0.3 mm i.d., 1.8 μ m C18 particles) for the analysis of signaling lipids and retinoids. A sample was loaded with a short trap column (10 \times 1 mm i.d., 1.8 μ m C18 particles). This low-input metabolomics and lipidomics was applied to hematopoietic stem cells.

High-throughput LC–MS methods (<10 min) enable analyzing over a hundred samples daily. That can be achieved by modifying conventional UHPLC–MS methods, including using shorter columns, increasing column flow rate and temperature, and adjusting the LC gradient, ion source and MS settings (Fig. S1). High-throughput methods are popular in large-scale metabolomics and lipidomics. However, a lower annotation rate and chromatographic resolution compared to conventional

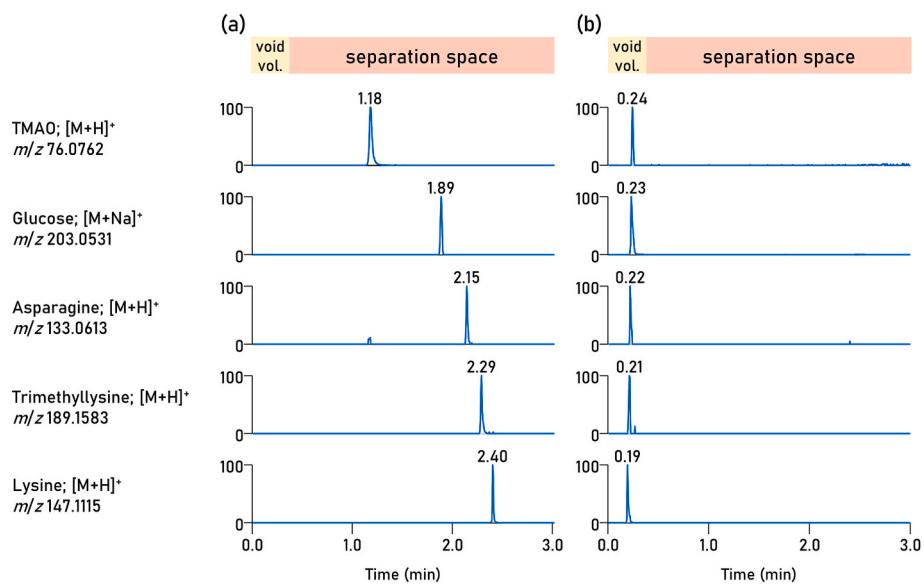


Fig. 2. Examples of separating polar metabolites in human serum extracts using fast HILIC-ESI(+)–MS and RPLC-ESI(+)–MS methods (4.4 min injection-to-injection time): trimethylamine *N*-oxide (TMAO) ($\text{XlogP} = 0.93$), glucose ($\text{XlogP} = 2.9$), asparagine ($\text{XlogP} = 4.3$), trimethyllysine ($\text{XlogP} = 6.2$), lysine ($\text{XlogP} = 3.2$). (a) The ACQUITY Premier BEH Amide column (50 mm \times 2.1 mm i.d.; 1.7 μm particle size) equipped with a VanGuard FIT cartridge (5 mm \times 2.1 mm i.d.; 1.7 μm particle size) (Waters, Milford, MA, USA) was utilized to separate polar metabolites based on the HILIC mechanism. (B) The ACQUITY Premier HSS T3 column (50 mm \times 2.1 mm i.d.; 1.8 μm particle size) equipped with a VanGuard FIT cartridge (5 mm \times 2.1 mm i.d.; 1.8 μm particle size) (Waters) was utilized to separate polar metabolites based on the RPLC mechanism. Details of separation conditions are provided in Ref. [35]. XlogP data for representative metabolites were taken from the HMDB (hmdb.ca).

methods should be anticipated [27].

After separating compounds, they undergo ionization in an ion source, forming charged species. Electrospray ionization (ESI) is the most frequently used method in LC–MS. It permits the ionization of small molecules (<2000 Da) as well as large molecules (peptides and proteins). ESI is considered a soft ionization technique, minimizing the fragmentation of the molecular ions. However, some metabolites can still be too fragile during ESI, resulting in various in-source fragments [62]. These fragments may also contribute to false metabolite annotations. Recent findings demonstrated that in-source fragments accounted for over 70 % of the observed peaks, suggesting that the spectra are largely influenced by fragment ions generated during ionization [63].

The presence of salts also impacts ESI performance; thus, the LC methods are limited to the use of only volatile mobile phase modifiers, with the most frequently used formic acid, acetic acid, ammonium hydroxide, ammonium formate, ammonium acetate, and ammonium bicarbonate [3]. In addition, ion suppression can occur because various analytes or other components of the analyzed matrix compete for ionization [64].

ESI can be operated in positive or negative ion mode, forming gas-phase cations or anions. Different ion forms can be formed based on the mobile phase modifiers, with protonated $[\text{M}+\text{H}]^+$ or deprotonated $[\text{M}-\text{H}]^-$ molecules the most commonly observed species. However, when ammonium salts (e.g., ammonium formate, ammonium acetate) are used as mobile phase modifiers, some metabolites tend to form $[\text{M}+\text{NH}_4]^+$ adducts (e.g., diacylglycerols, triacylglycerols, cholesteryl esters) in positive mode, or $[\text{M}+\text{HCOO}]^-/[\text{M}+\text{CH}_3\text{COO}]^-$ adducts in negative mode (e.g., phosphatidylcholines, ceramides, sphingomyelins) [59]. Furthermore, adducts such as $[\text{M}+\text{Na}]^+$, $[\text{M}+\text{K}]^+$, or $[\text{M}+\text{Cl}]^-$ are commonly observed due to leaking metals from glass bottles (mobile phase reservoirs) or their presence in analyzed extracts.

Different MS techniques are used for compound detection in targeted and untargeted metabolomics and lipidomics. These MS techniques can be divided into low-resolution MS and high-resolution MS instruments [7]. Low-resolution (tandem mass spectrometry) triple-quadrupole (QqQ) and quadrupole/linear ion trap (QLIT) are vastly used in targeted metabolomics and lipidomics. QqQ and QLIT generally operate in

a multiple reaction monitoring (MRM) mode. In principle, the precursor ion is isolated, followed by its fragmentation and subsequent detection of the generated fragment(s) (e.g., for TMAO, precursor ion m/z 76.2 \rightarrow collision energy (fragmentation) \rightarrow product ion m/z 58.2). These low-resolution mass analyzers enable metabolite identification and quantification, lowering the chance of false annotation; however, they can be used only for a defined number of metabolites and do not allow retrospective data analysis [65]. Contrarily, high-resolution mass analyzers used in untargeted approaches routinely operate in full mass spectra acquisition mode, which enables retrospective data mining [66].

High-resolution single-stage mass analyzers as time-of-flight (TOF) and orbital ion trap (Orbitrap) and hybrid analyzers as quadrupole/time-of-flight (QTOF) and quadrupole/orbital ion trap (Q/Orbitrap) are the main instruments used in the untargeted approach. High-resolution MS (HRMS) provides accurate mass data and improves the quality of the annotation of metabolites. The current mass resolving power for HRMS is usually 10,000–500,000 full width at half maximum (FWHM), which is much higher than the resolving power for low-resolution MS (~1000 FWHM) [67]. Additionally, LC–HRMS reduces the analysis time during metabolomics and lipidomics studies compared to LC–QqQ/QLIT analyzers [68]. It has been shown that a resolving power above 60,000 FWHM does not increase the number of molecular features [69]. However, higher resolving power can be beneficial during fluxomics analysis, where isotopologues must be resolved from other ions, including background ions. This capability minimizes the risk of spectral coelutions [70].

High-resolution tandem mass spectrometry (HR-MS/MS) is commonly used in metabolomics and lipidomics to increase confidence during metabolite annotation. HR-MS/MS experiments are performed in data-dependent acquisition (DDA) or data-independent acquisition (DIA) modes. In DDA mode, precursor ions are selected above a preset intensity threshold and using a narrow isolation window (0.4–4 m/z units) for further fragmentation; thus, product ions are then easily related to precursor ions. However, low-abundant precursor ions are often absent from MS/MS measurement [71]. DDA provides selective MS/MS spectra due to a narrow isolation window; however, DDA settings are more complicated than DIA settings, leading to a higher risk of

errors in method development and application [65].

Conversely, the DIA approach works with a wider precursor ion isolation window, allowing fragmentation of all precursor ions within the pre-set window. However, with such complex MS/MS spectra, it is hard to determine the relationship between multiple precursor ions and their fragments. Thus, DIA offers more comprehensive coverage for low-abundance precursor ions; however, the spectra quality is usually lower. Often used DIA methods include all-ion fragmentation modes (e.g., MS^E, Wide-band Fragmentation, All Ion Fragmentation, MS/MS^{ALL}), which fragments all the precursor ions in wider windows (e.g., 1000 m/z or more), and sequential window acquisition of all theoretical mass spectra (SWATH-MS/MS), which usually uses 20–50 m/z unit window [72]. With a wider isolation window, the risk of contamination increases. Several programs have been developed to deconvolute the raw MS/MS spectra (e.g., MS-DIAL, DecoMetDIA, DecoID) [27].

5.1. Tips and tricks

• A multiplatform LC-MS-based approach is commonly used during metabolomics and lipidomics analysis [73]. Remember that platforms might overlap, but some can provide unique detection capabilities for particular metabolites. For instance, LC-MS-based lipidomics in positive mode provide unique coverage for lipid subclasses such as diacylglycerols (DG), triacylglycerols (TG), cholesteryl esters (CE), and free cholesterol. In contrast, the negative mode detects free fatty acids (FA), fatty acid esters of hydroxy fatty acids (FAHFA), and cholesterol sulfate. Phospholipids such as (lyso)-phosphatidylcholines (LPC/PC), (lyso)-phosphatidylethanolamines (LPE/PE), (lyso)-phosphatidylinositols (LPI/PI), (lyso)-phosphatidylserines (LPS/PS), and (lyso)-phosphatidylglycerols (LPG/PG) can be detected in both platforms, but usually, one platform provides more annotations than the other (e.g., better coverage of LPC/PC in positive mode vs. LPE/PE in negative mode) [74].

- When using bi-phase extraction, two LC-MS platforms are typically required for acylcarnitines since they are partitioned into the polar phase (carbon chain lengths from 2 to 10) and the organic phase (carbon chain lengths above 12) [75]. HILIC is preferred for shorter (polar) acylcarnitines, while RPLC separation is optimal for longer (less polar) acylcarnitines.
- Polar metabolites are separated using RPLC or HILIC. However, HILIC is more efficient for very polar metabolites, while RPLC is for less polar metabolites. For instance, carnitine ($XlogP = -4.9$) can be detected using both RPLC and HILIC. However, in RPLC, the elution is close to a void volume where ion suppression can be expected (elution of other very polar metabolites, salts), while HILIC provides better retention and separation selectivity [76]. As a rule of thumb, an $XlogP > 0$ indicates a more hydrophobic analyte suitable for RPLC, while an $XlogP < 0$ suggests a more hydrophilic analyte for HILIC.
- RPLC-MS and HILIC-MS methods require different resuspension (or reconstitution) solvent compositions to be compatible with the initial conditions of the LC gradient. In RPLC-MS-based metabolomics, favoring a high percentage of water at the beginning of the gradient, the resuspension solvent should also contain a high percentage of water as well to avoid peak deterioration of early eluting metabolites, especially when high injection volumes are used (e.g., 5 μL) [77]. For HILIC-MS, typically featuring a high percentage of acetonitrile at the beginning of the gradient, a compromise is necessary due to the solubility of polar metabolites. This requires a high percentage of water in the resuspension solvent (i.e., 70–80 % acetonitrile) compared to initial conditions (e.g., 90–95 % acetonitrile) [78]. This compromise may negatively impact the peak shape of early eluting metabolites during the injection of high volumes.
- Vial caps can release polydimethylsiloxanes, especially during repeated injections from the same vial (Fig. 3A and B). In RPLC-MS-based lipidomic profiling, when using isopropanol as a strong mobile phase component, they elute as homologous series separated by a 74.0189 m/z unit, potentially impacting the intensity of endogenous

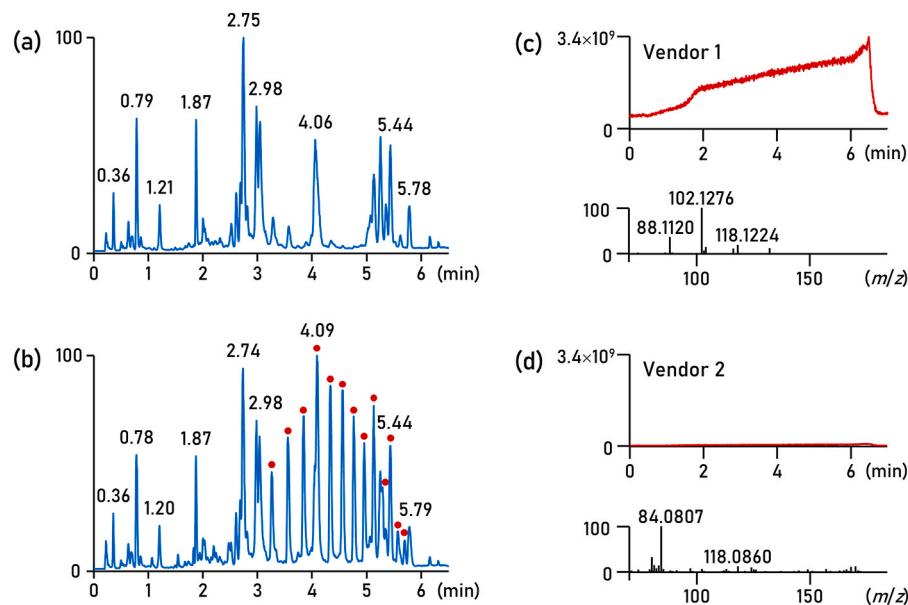


Fig. 3. RPLC-ESI(+)MS lipidomic profiling of human serum extracts (a) with and (b) without elution of polydimethylsiloxanes released from a vial cap (indicated by red circles ●). Polydimethylsiloxanes are eluted either as separate peaks or coeluting with the lipids. (c, d) Examples of RPLC-ESI(+)MS analysis with a mobile phase containing LC-MS-grade isopropanol from two vendors show the total ion chromatograms (m/z 70–1050) and MS1 spectra of mobile phase impurities. The ACQUITY Premier BEH C18 column (50 mm \times 2.1 mm i.d.; 1.7 μm particle size) equipped with a VanGuard FIT cartridge (5 mm \times 2.1 mm i.d.; 1.7 μm particle size) (Waters) was used for lipid separation with the mobile phase including (A) 60:40 acetonitrile/water with 10 mM ammonium formate and 0.1 % formic acid, and (B) 90:10:1 isopropanol/acetonitrile/water with 10 mM ammonium formate and 0.1 % formic acid. To improve the solubility of ammonium salts, they should first be dissolved in a small aliquot of water before being added to the 90:10 isopropanol/acetonitrile mixture. Fig. 3C and D reproduced (modified) with permission from Ref. [81]. Further details of separation conditions are provided in Refs. [73,81]. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

lipid species [79]. When multiple injections from a single vial are considered in the RPLC-MS method due to limited extract volume, vial caps from different vendors should be evaluated to minimize such contaminations. On the other hand, when using RPLC-MS and HILIC-MS with less strong mobile phases, polydimethylsiloxanes are usually not eluted; however, they tend to accumulate on the column. Thus, occasional rinsing of the column with isopropanol is helpful to remove this contamination.

- As a common rule, LC-MS-grade solvents and mobile-phase modifiers should be used during LC-MS analysis. These chemicals meet criteria such as a low mass noise level, minimal organic contamination, and minimal metal content [80]. However, the choice of such chemicals can be overwhelming because they are all labeled as “LC-MS-grade,” yet vendors may use different technologies to deliver the final products, resulting in variations in quality. Thus, it is recommended that solvent quality in LC-MS analysis be evaluated, especially when developing new methods. The choice of organic solvents can significantly impact the LC-MS system’s overall performance, affecting the ionization and detection of analytes [81]. Notably, a study evaluating isopropanol, a commonly used solvent, during RPLC-MS lipidomic profiling found that the most abundant mobile-phase impurities included the homologous series of amines, dioctyl phthalate, erucamide, Irgafos 168 oxide and its oxidized form, and Irganox 1076 [81]. Fig. 3C and D compare the use of isopropanol from two vendors, indicating significant differences in background contamination originating from mobile phases. Other common interferences and contaminants encountered during LC-MS are summarized in a review by Keller et al. [82].
- Mobile phase modifiers significantly impact metabolite retention, peak width and intensity, the ability to separate isomers, and the long-term stability of the retention times. For instance, a HILIC column based on a trifunctionally bonded amide phase, using 10 mM ammonium formate and 0.125 % formic acid, has outperformed other mobile phase modifiers [73]. Nevertheless, when evaluating a new column, it is advised to test mobile phase modifiers differing in composition and pH to optimize separation efficiency and selectivity

for a panel of expected groups of metabolites (e.g., amino acids, organic acids, biogenic amines, sugars) using a scoring system employing retention time, peak height intensity, peak width, and the ability to resolve isomers (e.g., leucine/isoleucine). Fig. 4 shows the impact of various mobile phase modifiers during HILIC-MS metabolomics analysis on the amino acid arginine in human serum extracts [73]. For RPLC-MS lipidomic profiling with a C18 column, 10 mM ammonium formate or 10 mM ammonium formate with 0.1 % formic acid in positive mode has been advised. On the other hand, using 10 mM ammonium acetate with 0.1 % acetic acid has been shown as a reasonable compromise regarding the signal intensity of the detected lipids and the stability of the retention times during long-term sequences (Fig. S2) [73].

- For targeted LC-MS methods operated in MRM mode, evaluating the sample preparation step or SPE cleanup is recommended by monitoring highly abundant phospholipids typically occurring in biological samples [83]. To this end, using additional MRM transitions for these lipids, precursor ion scan of m/z 184, or operating the mass spec in full scan mode helps with method optimization and avoiding carry-over of these lipids or achieving separation from target metabolites. For more in-depth characterization of potentially interfering lipids, including annotation using in-silico MS/MS spectra, it is recommended to use HRMS operated in MS/MS mode with the identical LC method [84].
- Current HRMS instruments for untargeted analysis usually have a linear dynamic range of about four orders of magnitude. Thus, it is necessary to determine the appropriate dilution of the extracts for injection and the typical intensity readout for the detector through an initial evaluation. To this end, serially diluted extracts are injected for a particular LC-MS platform ranging from very low to very high concentrated extracts. TOF instruments usually have an upper linearity of total ion chromatogram (TIC) limit of around 10^6 – 10^7 , while Orbitrap instruments reach around 10^9 . In TOF instruments, either saturation of the multichannel plate detector or reaching the analog-to-digital or time-to-digital converter limit leads to nonlinear signal behavior. In contrast, in Orbitrap instruments, highly abundant ions

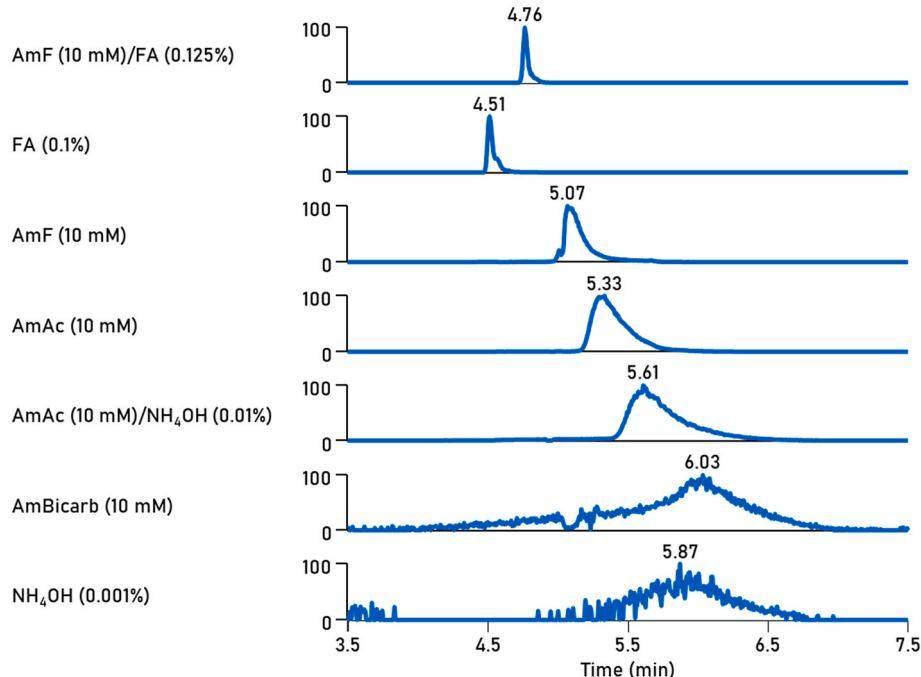


Fig. 4. Separation of arginine ($[M+H]^+$, m/z 175.1178) in human serum extracts using an ACQUITY UPLC BEH Amide column (50 mm \times 2.1 mm i.d.; 1.7 μ m particle size) coupled to an ACQUITY UPLC BEH Amide VanGuard pre-column (5 mm \times 2.1 mm i.d.; 1.7 μ m particle size) (Waters) under different mobile phase modifiers. Reproduced (modified) with permission from Ref. [73]. Legend: AmF, ammonium formate; FA, formic acid; AmF, ammonium formate; AmAc, ammonium acetate; NH₄OH, ammonium hydroxide; AmBicarb, ammonium bicarbonate.

quickly reach the automatic gain control (AGC) target, triggering automatic ion injection into the detector. This process can potentially lead to insufficient capacity to accumulate and detect the ions. Fig. 5, which illustrates examples of lipid extracts injected within (Fig. 5A) and outside (Fig. 5B) the linear dynamic range of the detector, highlights this phenomenon. While working within the linear dynamic range for quantitative and semiquantitative analysis is crucial, utilizing (reasonably) more concentrated samples can yield additional MS/MS spectra and improve their overall quality.

- High mass resolving power is crucial for removing potential interferences and separating isobaric compounds (e.g., compounds with the same nominal mass but different accurate masses). For example, in RPLC-MS-based lipidomics analysis, two lipids, PC P-34:1 ($[\text{M}+\text{H}]^+$, m/z 744.5902) and PE 36:2 ($[\text{M}+\text{H}]^+$, m/z 744.5538) are typically closely eluted [74]. A mass resolving power of at least 40,000 FWHM is required to achieve complete spectral separation of these two isobaric compounds. Therefore, low mass resolving power results in spectral co-elution, detecting these two lipids as a single mass peak with biased mass accuracy (Fig. S3).
- During DDA experiments, precursor ions can be isolated within a relatively narrow window, typically 0.4–1 m/z unit, resulting in highly selective isolation. However, this narrow window often reduces the intensity of MS/MS fragments due to the lower transmission of precursor ions. Conversely, employing a wider isolation window, such as 3–4 m/z units, enhances the sensitivity of detected fragments. However, it also increases the likelihood of potential interferences, as metabolites differing in mass by 2–3 m/z units can be co-isolated (Figs. S4 and S5). This results in mixed MS/MS spectra and a worse spectral match during library search.
- Since conventional DDA for acquiring MS/MS spectra is inefficient for low-abundance precursor ions, iterative exclusion-MS may help

increase the total number of MS/MS scans for metabolite annotation. Usually, the pool QC sample is injected, followed by excluding precursor ions for which MS/MS were acquired; thus, in sequential injections, low-abundant precursor ions are selected for MS/MS analysis, leading to a higher metabolite annotation rate [85].

6. Processing raw files

Processing raw instrumental files usually includes feature detection, chromatogram building, deisotoping, peak alignment, and gap-filling [86,87]. Typical untargeted metabolomics and lipidomics analyses can produce hundreds of annotated metabolites and countless unknowns. A wide range of processing tools have been developed in recent years, including commercial vendor software (e.g., MarkerLynx, MarkerView, Mass Profiler Professional, MetaboScape, Compound Discoverer) or from independent developers (e.g., GeneData), open access software (e.g., XCMS, MZmine, MS-DIAL, MetAlign, IDEOM) or script platform (e.g., Matlab, R) [27]. Recently, it has been reported that the main differences among the data processing software programs were found in the number of false positive/negative peaks and gap-filling capability [88,89]. However, data quality is continuously improving through consistent software updates.

6.1. Tips and tricks

- “Trust, but verify” – software that allows reviewing and curating annotated metabolites should be used to remove false positives or annotations that appear multiple times due to increased baseline for particular metabolites.
- Using internal standards helps quickly assess the quality of generated data during data processing since they represent true positives in the

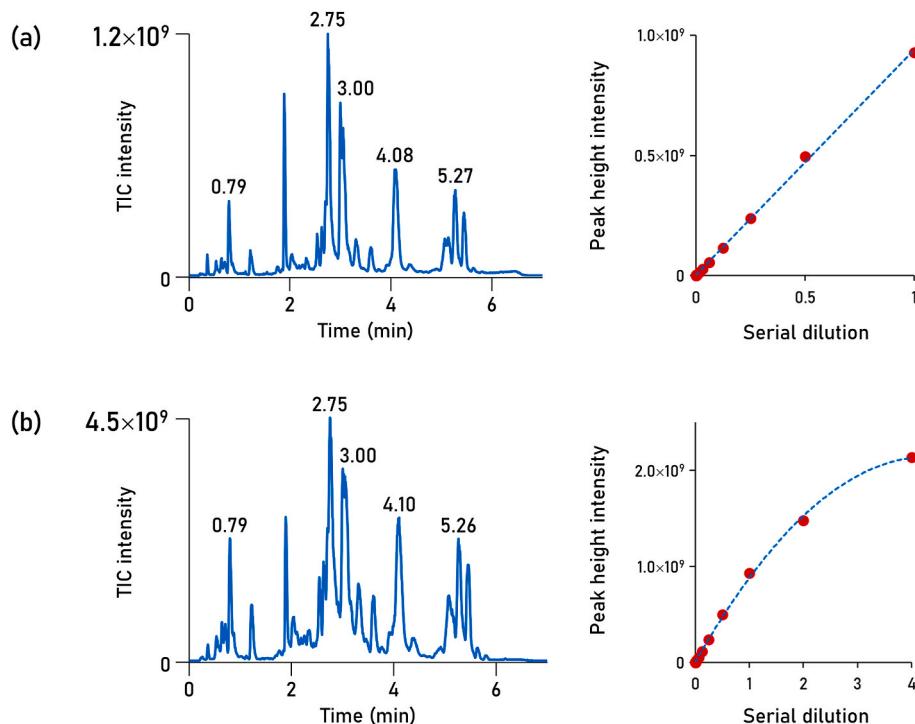


Fig. 5. Examples of TIC (m/z 200–1700) of lipid extracts from human serum injected (a) within the linear dynamic range (TIC 1.2×10^9) and (b) outside the linear dynamic range (TIC 4.5×10^9) on the Q Exactive Plus instrument. Working within the linear dynamic range provided a linear response ($R^2 > 0.99$) for lipids when using serial dilution samples of the lipid extracts. On the other hand, the linearity for highly abundant lipids (example shown for PC 34:2, $[\text{M}+\text{H}]^+$, m/z 758.5694) deteriorated when highly concentrated lipid extracts were injected. The ACQUITY Premier BEH C18 column (50 mm \times 2.1 mm i.d.; 1.7 μm particle size) equipped with a VanGuard FIT cartridge (5 mm \times 2.1 mm i.d.; 1.7 μm particle size) (Waters) was used for lipid separation with the mobile phase including (A) 60:40 acetonitrile/water with 10 mM ammonium formate and 0.1 % formic acid, and (B) 90:10:0.1 isopropanol/acetonitrile/water with 10 mM ammonium formate and 0.1 % formic acid. Further details of the RPLC-MS method can be found in Ref. [81].

sample. If they do not appear in the exported dataset, the cause should be investigated. This may include chromatographic or mass accuracy drift, wrong setting of data processing parameters, gross error during sample preparation, or LC malfunction.

- Parameters of data processing software, such as the minimal threshold (or signal-to-noise ratio, noise level, peak amplitude), should be optimized based on the MS instrument used; otherwise, excessively noisy data may appear in the exported dataset, thereby increasing data processing time. Furthermore, the smoothing level should be chosen judiciously to prevent the generation of artifact peaks from noise or the degradation of partially coeluted peaks detected afterward as a single peak.
- Spectral deconvolution (i.e., obtaining distinct peaks) of isomers is challenging during data processing, especially for partially co-eluted metabolites. When optimizing data processing settings, special attention should be given to parameters that may affect the ability to distinguish between isomers. Examples of these compounds include leucine/isoleucine, 3-hydroxybutyrate/3-hydroxisobutyrate/2-hydroxybutyrate, citrate/isocitrate, or sugar phosphates for polar metabolites. For lipids, examples of isomers may include phosphatidylcholines (e.g., PC 36:3-5, PC 38:4-5) or triacylglycerols (e.g., TG 54:5-7), each providing multiple isomers reported in plasma during RPLC-MS lipidomics analysis [90]. Additionally, lyso-forms of phospholipids are prone to isomerization, leading to 1-acyl-2-lyso-phospholipid and 1-lyso-2-acyl-phospholipid and detected as a double peak [91]. Fig. 6 shows examples of isomers that can be used to optimize data processing parameters.
- With new updates or versions of data processing software, their performance should be evaluated using validation data sets (instrumental files) for each platform, and data processing outcomes should be compared. This includes evaluating the total number of features, the efficiency of deconvolution of fully and partially separated isomers, and the impact on metabolite annotation, including annotation of new lipid subclasses if the new version provides such an update.

7. Metabolite annotation

Due to the structural complexity of metabolites, their annotation might be challenging [92]. Thousands of metabolite signals in a single sample can be obtained using untargeted metabolomics and lipidomics. However, only a small percentage of these signals is structurally known

[93].

The Metabolomics Standardization Initiative (MSI) explains how to report and perform metabolomics workflow using a community-based guideline. MSI proposed four confidence levels: (1) identified compounds, (2) putatively annotated compounds, (3) putatively characterized compound classes, and (4) unknown compounds [94]. However, multiple researchers have initiated revisions and possible modifications of the metabolite reporting standards to fit the current needs [95,96]. In 2020, the Lipidomics Standardization Initiative (LSI) was introduced, aiming to create guidelines for lipidomics workflows, including lipid annotation guidelines based on the hierarchical concept and application of shorthand notation [97].

The experimental MS/MS libraries play one of the key roles in metabolite annotation. Various public and commercial libraries are available. The most extensive spectral library, METLIN Gen2 (metlin.scripps.edu), contains over 900,000 molecular standards with MS/MS data generated in positive and negative ionization modes at multiple collision energies, collectively containing over 4 million tandem mass spectra [98]. The next largest library updated in 2023, the National Institute of Standards and Technology (NIST) MS/MS library (chemdata.nist.gov), includes over 2.3 million spectra from over 51 thousand standards.

One of the most comprehensive free access libraries is MassBank of North America (MoNA), which allows users to download and upload MS/MS spectra freely (massbank.us). MoNA contains over 2 million MS/MS spectra records, including experimental spectra for compounds such as natural products or endogenous metabolites [27]. Other public MS/MS spectral libraries are available, such as MassBank (massbank.jp), ReSpect (spectra.psc.riken.jp), RIKEN PlaSMA (plasma.riken.jp), mzCloud (mzcloud.org), GNPS (gnps.ucsd.edu), MSforID (msforid.com), or HMDB (hmdb.ca). As an example, Fig. 7 shows the visualization of processed LC-MS data in MS-DIAL software, including MS/MS library search and annotation.

In addition, the MS/MS spectral library can annotate simple and complex lipids (included in MS-DIAL and MZmine) or be available in a mass searchable format (MSP) for other data processing software (downloaded from MoNA). These in-silico libraries were created since many lipids break in an MS/MS experiment predictably, leading to fragmentation rules. These rules were then applied to lipid structures generated using in-silico methods to yield a comprehensive lipidomics library for compound annotations [99].

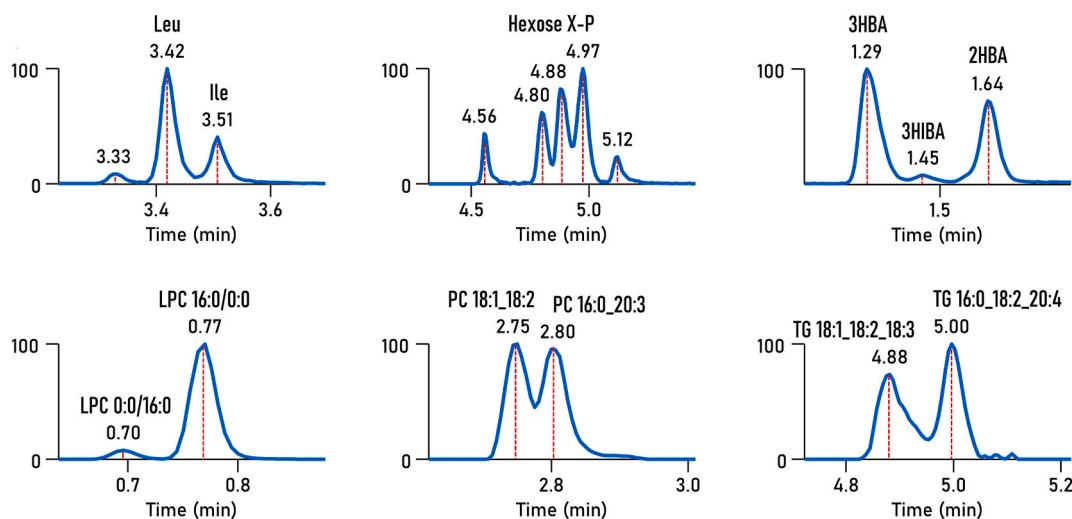


Fig. 6. Examples of isomers for optimization of data processing software: HILIC-MS metabolomic profiling: leucine and isoleucine ($[M+H]^+$, m/z 132.1019) in human serum extract; hexose X-phosphates ($[M-H]^-$, m/z 259.0224) in 3T3-L1 cell extract; RPLC-MS metabolomic profiling: 3-hydroxybutyrate, 3-hydroxisobutyrate, and 2-hydroxybutyrate ($[M-H]^-$, m/z 103.0400) in human serum extract; RPLC-MS lipidomic profiling: LPC 16:0 (LPC 0:0/16:0 and LPC 16:0/0:0, $[M+H]^+$, m/z 496.3398), PC 36:3 (PC 18:1_18:2 and PC 16:0_20:3, $[M+H]^+$, m/z 784.5851), TG 54:6 (TG 18:1_18:2_18:3 and TG 16:0_18:2_20:4, $[M+NH_4]^+$, m/z 896.7702) in human serum extract. The details of the LC-MS methods can be found in Ref. [73].



Fig. 7. Visualization of processed LC-MS data in MS-DIAL software: (1) list of processed LC-MS raw files; (2) alignment navigator; (3) peak filters; (4) peak spot viewer (spots colored based on the intensity of detected features/metabolites); (5) extracted ion chromatograms of aligned spot; (6) MS1 spectrum (full mass range of m/z 60–900); (7) details of annotated metabolite creatine in human plasma extract, showing that the difference between experimental (accurate) mass (m/z 132.07668) and calculated (exact) mass (m/z 132.0768) was as low as 0.12 mDa, passing through the MS1 accurate mass tolerance of 0.005 Da (5 mDa); (8) MS/MS spectrum acquire at stepped normalized collision energies of 20, 30, and 40 % using a Q Exactive Plus (upper panel) and reference (library) MS/MS spectrum acquire at stepped normalized collision energies of 20, 30, and 40 % using a Q Exactive HF available from MoNA (bottom panel), passing through the MS/MS accurate mass tolerance of 0.005 Da (5 mDa) as well. For separation, an ACQUITY UPLC BEH Amide column (50 mm \times 2.1 mm i.d.; 1.7 μ m particle size) coupled to an ACQUITY UPLC BEH Amide VanGuard pre-column (5 mm \times 2.1 mm i.d.; 1.7 μ m particle size) (Waters) was utilized [73].

Furthermore, most libraries are not specific for metabolites and include MS/MS spectra for other small molecules (e.g., food additives, pharmaceutical drugs, pesticides). Additionally, attention should be paid to the applied mass analyzers (e.g., QTOF or Orbitrap), collision energy, concentration and purity of chemical standards, and other experimental conditions because different libraries were acquired using different parameters [100]. Recently, Hoang et al. conducted cross-platform compassion (using QTOFs from Agilent, Bruker, SCIEX, Waters, and Orbitrap from Thermo) and tested collisional energies at 0, 10, 20, and 40 eV. They suggested that a collision energy of 20 eV provides optimal congruity and could significantly enhance uniformity and sensitivity, ultimately enhancing the accessibility and reproducibility of scientific data [101].

Another method that can be considered for compound annotation is IM-MS and LC-IM-MS. IM-MS provides an additional physicochemical property, CCS (collision cross section), which can be used for compound annotation and identification [102]. Coupling LC-MS with ion mobility separation (LC-IM-MS) improves the accuracy of metabolite annotation. It can also help with the process by improving peak capacity and resolving power, reducing the matrix effect, and increasing signal-to-noise ratios of metabolites [100]. Multiple experimental CCS databases for small molecules have been introduced [84,103,104].

It is estimated that only 10–20 % of molecular features detected during untargeted LC-MS-based metabolomics can be annotated according to mass spectra library matches [105]. In the next step, computational simulations predict the mass spectra from input structures, aiming to increase the annotation rate [106]. Various in-silico fragmentation software programs (Mass Frontier, CSI:FingerID, CFM-ID, MS-FINDER, MIDAS-G, and MetFrag) are used to determine these chemical structures [107]. Internal compound databases (e.g., HMDB, hmdb.ca; PubChem, pubchem.ncbi.nlm.nih.gov) provide putative chemical structures for “known unknowns” (i.e., existing but not yet

identified or characterized metabolites). For “unknown unknowns” (i.e., completely unknown and uncharacterized metabolites), in-silico tools for generating new metabolite structures (e.g., MINE, minedatabase.mcs.anl.gov; BioTransformer, biotransformer.ca) or network-based approaches (e.g., Global Natural Product Social Molecular Networking, gnps.ucsd.edu) are available to complement the metabolite coverage [108]. Furthermore, additional orthogonal filters can be based on retention time prediction [109,110] or hydrogen/deuterium exchange mass spectrometry (HDX-MS) [111,112].

7.1. Tips and tricks

- Typical metabolomics reports distinguish between Level 1 (matching based on retention time, MS1, and MS/MS spectrum) and Level 2 (matching based on MS1 and MS/MS spectrum). In Level 1, the term “identified” is used, whereas “annotated” is the appropriate terminology in the latter case [94]. Annotation based only on matching MS1 accurate mass (Level 3) may lead to many misannotations.
- As a common standard, the metabolite should be annotated with MS1 accurate mass and MS/MS spectra, which removes many false positive annotations. Using data processing software and spectral libraries, this is performed using settings such as narrow mass tolerance (e.g., ± 0.005 Da) for MS1 precursor ions and MS/MS fragments and calculating similarity scores (e.g., dot product, reversed dot product). On the other hand, even a combination of MS1 and MS/MS for metabolite annotation can be inefficient, especially for isomeric compounds such as hexose phosphates, citrate/isocitrate, or leucine/isoleucine with identical or very similar MS/MS spectra, thus, requiring retention time information for proper annotation [28].
- During LC-MS-based metabolomics analysis, thousands of molecular features are detected in biological samples per analysis. These

molecular features do not necessarily correspond to unique metabolites; they include isotopes, adducts, artifacts, and contaminants. For instance, Fig. 8A and B show an example of triacylglycerol TG 50:2 detected as multiple adducts ($[M+NH_4]^+$, $[M+Na]^+$, $[M+K]^+$, $[2M+NH_4]^+$, $[2M+Na]^+$), each with multiple isotopes, resulting in 21 molecular features. Furthermore, in-source fragmentation, which leads to diacylglycerol fragment ions (Fig. 8C), contributed eight additional molecular features. While adducts can be annotated based on typical mass differences, annotating in-source fragments can be more challenging. However, these fragments are typically chromatographically coeluted, exhibit the same peak shape, correlate in peak intensities across samples, and have similar statistical outcomes as molecular ions. They may also coincide with those obtained during MS/MS analysis (Fig. 8D). This underscores the importance of proper data processing and curation to yield a concise list of unique metabolites and relevant molecular features for statistical analysis. Furthermore, it has been shown that a systems-level annotation of a metabolomics dataset can reduce 25,000 molecular features to fewer than 1000 unique metabolites [113].

- Repeated annotations of some metabolites in LC-MS can be due to in-source artifact formation; for instance, glutamine and glutamic acid generate an in-source artifact by cyclizing to pyroglutamic acid (Fig. S6) [114]. Further, when the intensity of a particular metabolite is too high, the tailing peak may lead to additional annotations of the same metabolite (Fig. S7), which should be removed during data curation.
- Complex lipids are reported based on the structural resolution from MS1 and MS/MS analysis, such as carbons and double bonds (e.g., PC 36:2) and fatty acyl constituents (e.g., PC 18:0_18:2). The use of the underscore “_” means that there is certainty in the composition of the fatty acyl constituents, but not their placement on the glycerol backbone [115]. Lipid annotation, where the positional isomeric level of the fatty acyl chains (sn1 and sn2) is known, is indicated by a slash “/” (e.g., PC 18:0/18:2).
- Some lipids consist of a mixture of two or even more lipid isomers with the same number of carbon atoms and double bonds but differing in acyl chain lengths (e.g., TG 52:3; TG 16:1_18:1_18:1 and

TG 16:0_18:1_18:2; Fig. S8). Such coelutions unavoidably occur during LC-MS-based lipidomic profiling. In such instances, the software typically annotates the specific lipid with the acyl chain composition most similar to the in-silico MS/MS spectrum. However, other possible compositions may also be considered during data curation [99].

- Isomeric bis(monoacylglycerophosphate) (BMP) and phosphatidylglycerol (PG) species can be detected in positive (as $[M+NH_4]^+$) and negative (as $[M-H]^-$) ion modes using LC-MS. However, only the positive mode provides MS/MS spectra for reliable differentiation, while the negative mode provides identical MS/MS spectra for both lipid subclasses (Fig. S9) [84].
- Proposed guidelines on the expected fatty acyl chains in the major species of phospholipids, glycerolipids, and sphingolipids in mammalian samples should be checked to assess the likelihood of particular lipid species [116].
- During structure elucidation, the first step is properly annotating the ion form (adduct). Since multiple ion forms can be formed during ESI, it is suggested that ions with characteristic differences be explored to confirm particular adducts. In ESI(+), typically $[M+H]^+$, $[M+NH_4]^+$, $[M+Na]^+$, and $[M+K]^+$ species are formed. A mass difference of 21.9819 m/z indicates the presence of $[M+H]^+$ and $[M+Na]^+$, while a mass difference of 4.9554 m/z shows the presence of $[M+NH_4]^+$ and $[M+Na]^+$. In ESI(-), $[M-H]^-$, $[M+Cl]^-$, and $[M+HCOO]^-$ / $[M+CH_3COO]^-$ based on mobile phase modifiers can be observed. Thus, a mass difference of 35.9767 m/z indicates $[M-H]^-$ and $[M+Cl]^-$, while mass differences of 46.0055 and 60.0211 m/z are observed for $[M-H]^-$ and $[M+HCOO]^-$ / $[M+CH_3COO]^-$ ions, respectively. Additionally, it is useful to compare data acquired in both ionization modes. For instance, if a mass difference of 2.0146 m/z is observed, the molecule forms both $[M+H]^+$ and $[M-H]^-$ ions.
- The isotopic pattern of an unknown molecule with a suggested ion form should then be submitted to an in-silico software program to determine its elemental composition. In this step, the mass tolerance filter (e.g., 0.005 Da), isotopic ratio tolerance, selected elements, and selection of internal databases will significantly impact the number

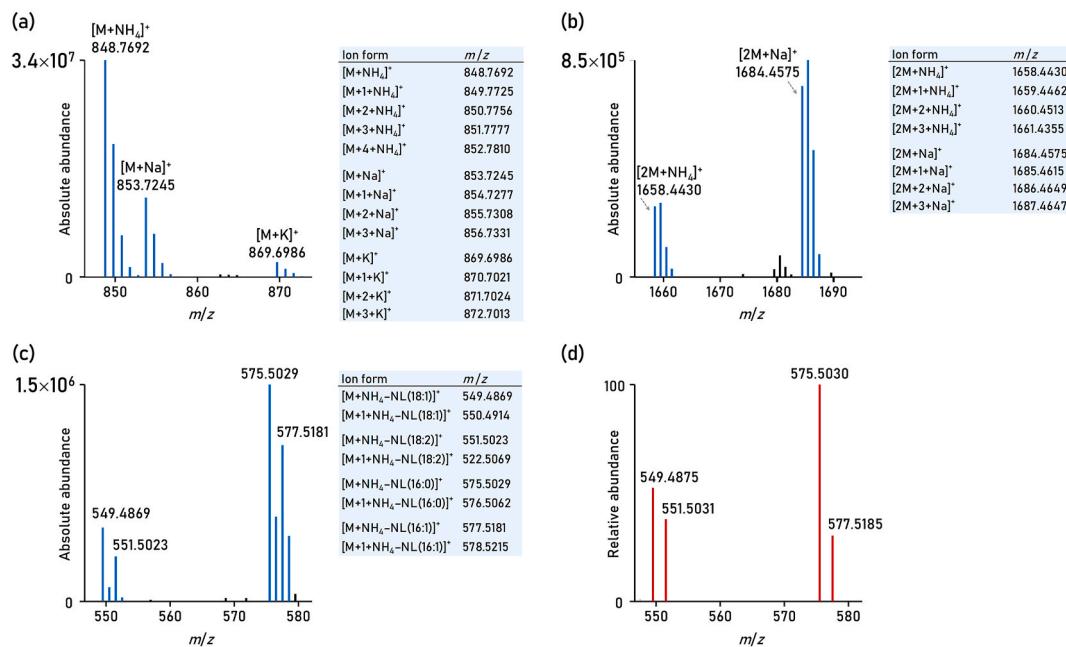


Fig. 8. Detection of TG 50:2 during RPLC-ESI(+) MS lipidomic profiling [81] of human serum extract: (a) intensity of $[M+NH_4]^+$, $[M+Na]^+$, $[M+K]^+$ adducts and their isotopes; (b) intensity of $[2M+NH_4]^+$, $[2M+Na]^+$ adducts and their isotopes; (c) intensity of in-source fragments of TG 50:2; (d) MS/MS spectrum (fragments) of TG 50:2 (precursor ions m/z 848.7692, retention time 5.24 min) leading to annotation of this TG as a mixture of TG 16:0_18:1_18:1 and TG 16:0_18:0_18:2 using MS-DIAL [84].

of possible formulae reported. Next, the MS/MS spectrum is used to generate a list of isomeric (scored) candidate structures from compound databases based on applied in-silico fragmentation rules (Fig. S10).

- A list of candidate structures can be reduced by applying HDX-MS. Exchangeable hydrogen atoms, bound to heteroatoms such as oxygen, nitrogen, and sulfur, readily exchange with deuterium, while those bound to carbon remain unaltered. When metabolites are exposed to deuterium-containing solvents (e.g., $\text{H}_2\text{O} \rightarrow \text{D}_2\text{O}$; methanol \rightarrow methanol-OD or methanol-D₄; isopropanol \rightarrow isopropanol-OD or isopropanol-D₈) and mobile phase modifiers (e.g., formic acid \rightarrow D₂-formic acid; ammonium formate \rightarrow D₅-ammonium formate), labile hydrogens within various functional groups ($-\text{NH}-$, $-\text{NH}_2$, $-\text{OH}$, $-\text{COOH}$, $-\text{SH}$) are substituted with deuterium [112]. The number of replaced hydrogens in the molecule can be determined by measuring the molecular mass before and after HDX (Fig. S11) using a mass spectrometer. As a final step, validation should be performed by analyzing authentic standards under the same LC-MS conditions and comparing retention times, MS1, and MS/MS spectra.

8. Quality control

The changes in the instrument's sensitivity are caused by the degradation of the extracts, contamination of ion source by nonvolatiles, or retention-time shifts, which can occur over time during the analytical run [7]. Therefore, the order of samples during the run should be randomized. Internal standards should also be added to the samples during analysis, which could help monitor chromatographic and mass accuracy drift and fluctuation of the signal intensity [9].

QC samples are generally employed to monitor the precision and stability of the method. It can also be used for signal correction and method standardization [117]. These samples should be identical biological samples, and their metabolic and sample matrix composition should be similar to the studied samples. These QC samples are injected regularly every 10–20 samples, depending on the analysis time and the number of biological samples [27].

QC samples are prepared as pooled aliquots from the study samples. The resulting mixture contains the mean concentration of all metabolites in these samples. However, this approach may not work for large-scale studies with over a thousand samples. In this case, the pooled QC sample may be replaced with a bulk “control” sample, a pool containing only a part of the study samples (preferably with samples covering all groups), or commercially available biofluids composed of multiple biological samples not present in the study [118]. Both approaches can also be combined [119]. A highly recommended commercial reference material is NIST Standard Reference Material 1950 – Metabolites in Frozen Human Plasma [83], used for human studies focusing on plasma or serum analysis [120].

The analysis should also include blanks and a serial dilution of QC samples. The no-injection blank runs mobile phase through the column without injection in the port of the instrument. At the beginning of the sequence, they aim to clean and condition the LC system, while within the sequence they provide information on carry-over of analytes or contaminants from previous injections. Reagent blanks are used to monitor potential contamination because manufacturing process modifications or supply chain changes can affect solvent quality without prior notice. Method blanks undergo the same steps during the analysis as the real samples but without a particular matrix. Both blanks and a serial dilution of QC samples help identify contamination in study samples and remove compounds with nonlinear behavior [9].

8.1. Tips and tricks

- Various measures such as (i) randomization of the samples within the sequence, (ii) regular injection of QC pool samples at the beginning, end, and between every ten actual samples for a specific matrix, (iii)

analysis of method blanks, (iv) analysis of serially diluted extracts prepared from QC sample (0, 1/16, 1/8, 1/4, 1/2, 1) (Fig. S12), and (v) control of the chromatographic peak shape, retention time, and intensity of internal standards are helpful to ensure QC [35].

- In practice, exported data sets can be filtered by removing metabolites with (i) a max sample peak height/blank peak height average <10 , (ii) an $R^2 < 0.8$ from a dilution series of QC sample, and (iii) a relative standard deviation (RSD) $>30\%$ from QC samples injected between 10 actual study samples [35]. Data can be further normalized using locally estimated scatterplot smoothing (LOESS) with QC samples injected between 10 actual study samples. In addition, the frequency (occurrence) cut-off (e.g., 50 %) of a particular metabolite within the entire data set or sample groups can be used.
- The total number of analyzed samples does not equal the number of injections per analytical platform. For instance, in a study with 100 study samples, the LC-MS sequence may include initial injections (e.g., solvent, QC samples) to ensure the stability of a particular platform (10 injections), system suitability tests (3 injections of a standard mixture or biological sample), method blanks (3 injections), a serial dilution of QC samples (6 injections), QC samples (12 injections), and study samples (100 injections). This means 34 more injections are needed to maintain proper QC (25 % of the total injections). Moreover, additional injections are required if the study is acquired in MS1 mode only, and MS/MS spectra are needed for metabolite annotations.
- Some lipids, such as (free) palmitic (FA 16:0) and stearic (FA 18:0) acids, are present when plastic tubes and tips are used [121]. Based on their signal intensity in method blanks and study samples, they can be removed during data filtration.
- Monitoring charts in untargeted metabolomics and lipidomics help visualize and track the performance and quality of analyses over time. Several types of monitoring charts can be implemented in the lab using analysis of QC samples at regular intervals to monitor the stability and consistency of each LC-MS platform. The extracted data from these injections can evaluate (i) the overall ion intensity (TIC) of the QC samples over time, (ii) retention times, peak heights and areas, and mass accuracy error of specific metabolites or internal standards across different runs, and (iii) the responses of blank samples and emerging new background contaminants. These parameters also help establish each LC-MS platform's “reference” status, allowing researchers to quickly identify and address any issues in the analytical process.

9. Data normalization

Normalization is crucial in removing undesired variations in the data while maintaining biological information. These variations can occur in ion intensities during the sequence and batches or sample-preparation steps. However, some unwanted biases in particular datasets might be challenging to remove. A wide range of strategies, from statistical methods to software tools, has been developed to address this issue. These normalization strategies are sample-based and data-based [122].

Sample-based strategies include intrinsic factors (e.g., ploidy/DNA, RNA, metabolites), cell-based approaches (e.g., count, volume), weight-based (e.g., fresh or dry weight), internal standards (e.g., isotopologues, stable-isotope labeled cells or extracts), QC-based (e.g., pooled QCs, standard reference material-based QCs), and others (e.g., specific gravity, osmolality). Data-based strategies include statistical methods (e.g., log, median, average) for batch effects (e.g., LOESS, ComBat) and analytical platform-based (e.g., total ion chromatogram) [123]. Multiple software tools have also been introduced, such as MetabR, MetaboDrift, NOREVA, and SERRF [124]. Fig. 9 shows an example of normalization using LOESS applied to two batches of study samples and accompanied QC samples.

However, choosing the appropriate normalization strategy can be difficult. Wu et al. [125] reviewed several strategies to help select a

proper normalization method. The review studied papers that worked with urine samples, cell extracts, and other biological samples (blood, sweat, and milk). Furthermore, Yang et al. [126] assessed 17 methods and their 110 possible combinations based on their performance. Various methods such as mean, cyclic LOESS, range scaling, and level scaling performed well under given criteria when combined with other normalizations.

9.1. Tips and tricks

- No sample-based normalization is usually needed for plasma and serum. On the other hand, normalization to creatinine or osmolality values is the most adopted strategy for urine with high variability in concentration and composition [127]. Fluorescence-based methods for DNA quantification can be used for cells [128]. For tissue mass (wet or dry), total protein concentration or DNA content can be used to recalculate the actual levels of metabolites in samples [122]. However, these approaches require additional analysis (e.g., DNA quantification, protein concentration, creatinine concentration).
- Data-based normalization includes TIC for all detected molecular features or mTIC with only genuine metabolites. Usually, there is a correlation between (m)TIC and cell numbers [129]; thus, this approach is helpful if sample-based normalization is impossible. However, the (m)TIC approach assumes that the decrease in the concentration of metabolites in one group is balanced with the increase in one group of metabolites. However, this assumption may not always be valid because a specific systematic error may affect some metabolites differently than others [124].

10. Statistical analysis

Many metabolomics and lipidomics studies aim to discover biomarker metabolites or understand the investigated biological processes. Various statistical methods have been introduced to identify key metabolites and molecular features. The *t*-test, a typically used univariate statistical method for two groups, provides a *p*-value corrected using Bonferroni correction [130] or the Benjamini–Hochberg [131] false discovery rate (FDR) control procedure to avoid a high number of false positives. For three or more groups, one-way analysis of variance (ANOVA) with one independent variable or two-way ANOVA with two independent variables (e.g., diet and sex) is employed, followed by post hoc tests [132] to determine specific groups responsible for significant differences detected by ANOVA. Multivariate methods mainly include principal component analysis, hierarchical cluster analysis, and partial least squares discriminant analysis (PLS-DA) [133]. Among various statistical tools and software, MetaboAnalyst (metaboanalyst.ca) is one of the most popular platforms for comprehensive metabolomics and lipidomics data analysis and interpretation [134].

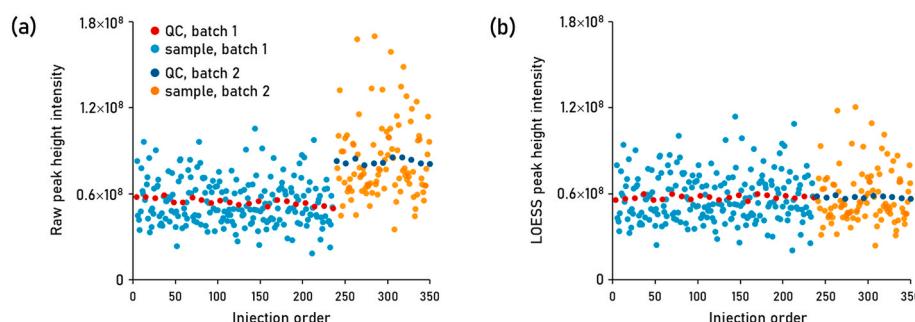


Fig. 9. Example of peak height intensities of acetylcarnitine (CAR 2:0) in human plasma acquired using HILIC-ESI(+)-MS [73] in QC and study samples over two batches. (a) Raw peak height intensities showing instrumental drift and batch effect; (b) peak height intensities after LOESS normalization and batch effect correction. Relative standard deviations (RSD) of 4.7 % and 2.7 % were observed for QC samples analyzed within batches 1 and 2, respectively, with an overall high RSD of 21.7 % for all QC samples due to batch effect. After LOESS normalization and batch effect correction, RSDs of 2.7 % and 1.6 % were obtained for QC samples within batches 1 and 2, respectively, with an overall RSD of 2.3 % for all QC samples.

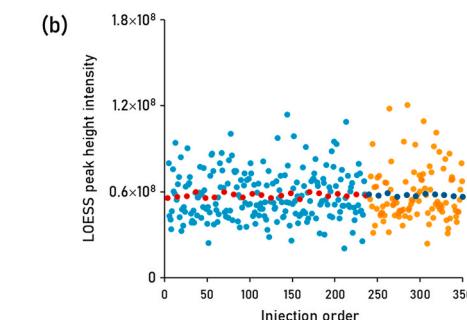
Integrating the results of an experiment with biological domain knowledge is necessary. For that purpose, metabolic network mapping can be performed with many software packages (e.g., MetaMapR, MetabNet, GNPS, MS2LDA). Metabolic pathway mapping can also be done with tools such as MetaboAnalyst or MetFlow [27]. For complex lipids, an ontology database and enrichment analysis (LION, lipidontology.com) and lipid over-representation analysis (LORA, lora.metabolomics.fgu.cas.cz) are available [135].

10.1. Tips and tricks

- When starting with statistical analysis, using test data available in MetaboAnalyst helps familiarize with the required data structure and available tools. In addition, the data for statistical analysis are available in some papers, usually in supporting materials, including the outcomes of the statistical analysis. Thus, these data sets can be used for practical hands-on exercises and to validate the acquired skills.
- Identifiers such as InChI Key, HMDB ID, or KEGG ID are helpful (and some are required) for advanced statistical methods such as enrichment and pathway analysis [134]. The Chemical Translation Service (cts.fiehnlab.ucdavis.edu) can translate different database identifiers or trivial names in metabolomics publications and databases to identifiers needed for bioinformatics tools [136]. Trivial names can also be converted to HMDB identifiers using MetaboAnalyst.
- A metabolite with a *p*-value ≤ 0.05 is typically used for further interpretation. For two-class comparison, fold-change can also be used as an additional measure (i.e., volcano plot). For instance, in MetaboAnalyst, a fold change of 2 is used as a default setting, although even fold changes of 1.3–1.5 have been shown to provide biologically meaningful sets rather than relying solely on *p*-values [137].
- When using PLS-DA, metabolites are sorted by their importance based on variable importance in projection (VIP) scores. A VIP score >1 is the typical rule for selecting relevant metabolites [138], although a higher value, such as 1.5 or even 2, can be considered a reasonable cut-off [139]. Consideration should also be given to PLS-DA cross-validation, specifically the model's predictive ability (Q^2) and the permutation test [140,141]. Good predictions will have a high Q^2 value, and $Q^2 > 0.5$ is considered very good in metabolomics studies [142]. If a low Q^2 value or even a negative Q^2 value is obtained, it indicates that the model lacks predictive ability or is overfitted, even if metabolites are reported with VIP >1 .

11. Data sharing

Data sharing and open data have become fundamental in science over the years. It is recommended that the data be shared in line with the



Findable, Accessible, Interoperable, and Reusable (FAIR) Guiding Principles for scientific data management and stewardship [143]. However, studies show that metabolomics and lipidomics still lack data sharing compared to other omics disciplines [144]. Dedicated data repositories, such as Metabolomics Workbench (metabolomicsworkbench.org), MetaboLights (ebi.ac.uk/metabolights), MassIVE (massive.ucsd.edu/ProteoSAFe/static/massive.jsp), and the general-purpose open repository Zenodo (zenodo.org), were developed to facilitate data sharing. Data should be uploaded in the original proprietary vendor format or converted to open access formats as mzML and mzXML using Proteo-Wizard software (proteowizard.sourceforge.io) [27].

Recently, the concept of metabolomics and lipidomics atlases has been introduced, aiming to provide open-access resources that characterize metabolites' quantitative distribution and relationships in different biological matrices, further highlighting the need for data reusability and sharing [145].

11.1. Tips and tricks

- When converting raw files to mzML and mzXML, remember that these files are usually much bigger than the original raw files. Thus, it is advised to apply reasonable filters such as centroiding (if profile mode was used during data acquisition) and cutting the noise to some extent.
- Raw LC-MS files from data repositories can also be helpful when adapting the protocols from other labs, permitting faster and more efficient method transfer by enabling direct comparison of newly acquired LC-MS records.
- Using authoritative identifiers, such as InChI keys or those from bioinformatics resources like the HMDB (hmdb.ca) and LIPID MAPS (lipidmaps.org), is advised to streamline the comparison of metabolites across studies.
- Details of lipidomic analyses can be summarized in a newly introduced minimal reporting checklist [146] and stored or shared in the supporting materials of papers or on Zenodo (zenodo.org). A reporting checklist is also available for different QC samples in untargeted metabolomics studies [9].

12. Conclusions

Recent advances in sample preparation, instrumentation, data processing, and bioinformatics have expanded the possibilities in metabolomics and lipidomics research. We have reviewed the steps involved in metabolomics and lipidomics experiments and compiled a guide for designing a tailored workflow based on some of the most popular techniques used in various studies. However, as with other omics fields, metabolomics and lipidomics analyses are prone to biases; therefore, tips and tricks are provided to avoid potential pitfalls. Given that the number of metabolomics and lipidomics studies has increased six-fold compared to a decade ago, maintaining QC is even more crucial for generating reliable data and meaningful biological interpretations. Providing sufficient information on the robustness and methodological details of novel metabolomics and lipidomics tools will be essential for faster and more efficient method transfer while minimizing potential pitfalls.

Future advancements in both omics fields are expected to merge untargeted (hypothesis-generating) and targeted (hypothesis-driven) workflows for comprehensive metabolite profiling, thanks to the availability of advanced mass spectrometers that enable faster data acquisition without compromising spectral quality. Additionally, spectral libraries are likely to expand further, leading to higher annotation rates during untargeted profiling. For processing raw LC-MS files, elucidating the structure of unknown metabolites, and data interpretation, advanced computational tools, including artificial intelligence and machine learning algorithms, may be crucial for automating and streamlining these processes. These improvements will accelerate and enhance

our understanding of complex biological systems. However, a chain is only as strong as its weakest link; thus, poorly acquired and processed LC-MS data can undermine the effectiveness of these advancements. Comprehensive training, rigorous standard operating procedures, and continuous QC measures will be essential to fully benefit from these technological improvements.

CRediT authorship contribution statement

Stanislava Rakusanova: Writing – review & editing, Writing – original draft. **Tomas Cajka:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

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 research was funded by the Czech Health Research Council (NU22-02-00161), the Czech Science Foundation (21-00477S), the Ministry of Education, Youth and Sport of the Czech Republic (LUAU24169), and the National Institute for Research of Metabolic and Cardiovascular Diseases (Programme EXCELES, ID Project No. LX22NPO5104) funded by the European Union—Next Generation EU. Examples of LC-MS chromatographic records (Vanquish UHPLC & Q Exactive Plus, Vanquish UHPLC & Orbitrap Exploris 480), processed data in MS-DIAL and MS-FINDER software, and LOESS normalization were provided by the Metabolomics Core Facility at the Institute of Physiology of the Czech Academy of Sciences (metabolomics.fgu.cas.cz).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.trac.2024.117940>.

References

1. R. Wang, B. Li, S.M. Lam, G. Shui, Integration of lipidomics and metabolomics for in-depth understanding of cellular mechanism and disease progression, *J. Genet. Genomics* 47 (2020) 69–83, <https://doi.org/10.1016/j.jgg.2019.11.009>.
2. O. Begou, H.G. Gika, I.D. Wilson, G. Theodoridis, Hyphenated MS-based targeted approaches in metabolomics, *Analyst* 142 (2017) 3079–3100, <https://doi.org/10.1039/C7AN00812K>.
3. G.A.N. Gowda, D. Djukovic, in: D. Raftery (Ed.), *Overview of Mass Spectrometry-Based Metabolomics: Opportunities and Challenges*, Springer New York, New York, NY, 2014, pp. 3–12.
4. J.T. Bjerrum, Y. Wang, F. Hao, M. Coskun, C. Ludwig, U. Günther, O.H. Nielsen, Metabolomics of human fecal extracts characterize ulcerative colitis, Crohn's disease and healthy individuals, *Metabolomics* 11 (2015) 122–133, <https://doi.org/10.1007/s11306-014-0677-3>.
5. Y. Lin, C. Ma, C. Liu, Z. Wang, J. Yang, X. Liu, Z. Shen, R. Wu, NMR-based fecal metabolomics fingerprinting as predictors of earlier diagnosis in patients with colorectal cancer, *Oncotarget* 7 (2016) 29454–29464, <https://doi.org/10.18632/oncotarget.8762>.
6. D.S. Wishart, Metabolomics for investigating physiological and pathophysiological processes, *Physiol. Rev.* 99 (2019) 1819–1875, <https://doi.org/10.1152/physrev.00035.2018>.
7. 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>.
8. B. Amer, R.R. Deshpande, S.S. Bird, Simultaneous quantitation and discovery (SQUAD) analysis: combining the best of targeted and untargeted mass

spectrometry-based metabolomics, *Metabolites* 13 (2023) 648, <https://doi.org/10.3390/metabo13050648>.

[9] 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, A. the metabolomics Quality, C. Quality Control, 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>.

[10] M.J. Gardner, D.G. Altman, Confidence intervals rather than P values: estimation rather than hypothesis testing, *Br. Med. J.* 292 (1986) 746, <https://doi.org/10.1136/2Fbmj.292.6522.746>.

[11] M. Vinaixa, S. Samino, I. Saez, J. Duran, J.J. Guinovart, O. Yanes, A guideline to univariate statistical analysis for LC/MS-based untargeted metabolomics-derived data, *Metabolites* 2 (2012) 775–795, <https://doi.org/10.3390/metabo2040775>.

[12] C.C. Serdar, M. Cihan, D. Yücel, M.A. Serdar, Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies, *Biochem. Med.* 31 (2021) 010502, <https://doi.org/10.11613/bm.2021.010502>.

[13] D. Dudzik, C. Barbas-Bernardos, A. García, C. Barbas, Quality assurance procedures for mass spectrometry untargeted metabolomics. A review, *J. Pharm. Biomed. Anal.* 147 (2018) 149–173, <https://doi.org/10.1016/j.jpba.2017.07.044>.

[14] F. Paul, E. Erdfelder, A. Buchner, A.G. Lang, Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses, *Behav. Res. Methods* 41 (2009) 1149–1160, <https://doi.org/10.3758/brm.41.4.1149>.

[15] X. Zhao, L. Niu, C. Clerici, R. Russo, M. Byrd, K.D.R. Setchell, Data analysis of MS-based clinical lipidomics studies with crossover design: a tutorial mini-review of statistical methods, *Clin. Mass Spectrom.* 13 (2019) 5–17, <https://doi.org/10.1016/j.clims.2019.05.002>.

[16] B.J. Blaise, G. Correia, A. Tin, J.H. Young, A.-C. Vergnaud, M. Lewis, J.T. M. Pearce, P. Elliott, J.K. Nicholson, E. Holmes, T.M.D. Ebbels, Power analysis and sample size determination in metabolic phenotyping, *Anal. Chem.* 88 (2016) 5179–5188, <https://doi.org/10.1021/acs.analchem.6b00188>.

[17] E. Billoir, V. Navratil, B.J. Blaise, Sample size calculation in metabolic phenotyping studies, *Brief. Bioinformation* 16 (2015) 813–819, <https://doi.org/10.1093/bib/bbu052>.

[18] G. Nyamundanda, I.C. Gormley, Y. Fan, W.M. Gallagher, L. Brennan, MetSizeR: selecting the optimal sample size for metabolomic studies using an analysis based approach, *BMC Bioinf.* 14 (2013) 338, <https://doi.org/10.1186/1471-2105-14-338>.

[19] F. Li, Y. Zhou, X. Zhang, J. Tang, Q. Yang, Y. Zhang, Y. Luo, J. Hu, W. Xue, Y. Qiu, Q. He, B. Yang, F. Zhu, SSizer: determining the sample sufficiency for comparative biological study, *J. Mol. Biol.* 432 (2020) 3411–3421, <https://doi.org/10.1016/j.jmb.2020.01.027>.

[20] J. Ivanisevic, E.J. Want, From samples to insights into metabolism: uncovering biologically relevant information in LC-HRMS metabolomics data, *Metabolites* 9 (2019) 308, <https://doi.org/10.3390/metabo9120308>.

[21] W.N. Arifin, W.M. Zahiruddin, Sample size calculation in animal studies using resource equation approach, *Malays. J. Med. Sci.* 24 (2017) 101–105, <https://doi.org/10.21315/mjms2017.24.5.11>.

[22] R.V. Lenth, Some practical guidelines for effective sample size determination, *Am. Statistician* 55 (2001) 187–193, <https://doi.org/10.1198/000313001317098149>.

[23] I. Nemet, P.P. Saha, N. Gupta, W. Zhu, K.A. Romano, S.M. Skye, T. Cajka, M. L. Mohan, L. Li, Y. Wu, M. Funabashi, A.E. Ramer-Tait, S.V. Naga Prasad, O. Fiehn, F.E. Rey, W.H.W. Tang, M.A. Fischbach, J.A. DiDonato, S.L. Hazen, A cardiovascular disease-linked gut microbial metabolite acts via adrenergic receptors, *Cell* 180 (2020) 862–877.e822, <https://doi.org/10.1016/j.cell.2020.02.016>.

[24] D.H. Lee, Q. Jin, N. Shi, F. Wang, A.M. Bever, L. Liang, F.B. Hu, M. Song, O. A. Zelezniak, X. Zhang, A. Joshi, K. Wu, J.Y. Jeon, J.A. Meyerhardt, A.T. Chan, A. H. Eliassen, C. Clish, S.K. Clinton, E.L. Giovannucci, J. Li, F.K. Tabung, The metabolic potential of inflammatory and insulinemic dietary patterns and risk of type 2 diabetes, *Diabetologia* 67 (2024) 88–101, <https://doi.org/10.1007/s00125-023-06021-3>.

[25] M. Costanzo, M. Caterino, G. Sotgiu, M. Ruoppolo, F. Franconi, I. Campesi, Sex differences in the human metabolome, *Biol. Sex Differ.* 13 (2022) 30, <https://doi.org/10.1186/s13293-022-00440-4>.

[26] L. Smith, J. Villaret-Cazadumont, 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>.

[27] S. Rakusanova, O. Fiehn, T. Cajka, Toward building mass spectrometry-based metabolomics and lipidomics atlases for biological and clinical research, *Trends Anal. Chem.* 158 (2023) 116825, <https://doi.org/10.1016/j.trac.2022.116825>.

[28] S. Alseckh, 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. Perez de Souza, K. Saito, U. Sauer, F.C. Schroeder, S. Schuster, G. Siuzdak, A. Skirycz, L.W. Sumner, M.P. Snyder, H. Tang, T. Tohge, Y. Wang, W. Wen, S. Wu, G. 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>.

[29] J. Pezzatti, J. Boccard, S. Codesido, Y. Gagnebin, A. Joshi, D. Picard, V. González-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>.

[30] A. Vignoli, L. Tenori, C. Morsiani, P. Turano, M. Capri, C. Luchinat, Serum or plasma (and which plasma), that is the question, *J. Proteome Res.* 21 (2022) 1061–1072, <https://doi.org/10.1021/acs.jproteome.1c00935>.

[31] G.B. Reis, J.C. Rees, A.A. Ivanova, Z. Kuklenyik, N.M. Drew, J.L. Pirkle, J.R. Barr, Stability of lipids in plasma and serum: effects of temperature-related storage conditions on the human lipidome, *J. Mass Spectrom. Adv. Clin. Lab.* 22 (2021) 34–42, <https://doi.org/10.1016/j.jmsacl.2021.10.002>.

[32] K.R. Maddipati, S.L. Zhou, Stability and analysis of eicosanoids and docosanoids in tissue culture media, *Prostag. Other Lipid Mediat.* 94 (2011) 59–72, <https://doi.org/10.1016/j.prostaglandins.2011.01.003>.

[33] W. Lu, L. Wang, L. Chen, S. Hui, J.D. Rabinowitz, Extraction and quantitation of nicotinamide adenine dinucleotide redox cofactors, *Antioxidants Redox Signal.* 28 (2018) 167–179, <https://doi.org/10.1089/ars.2017.7014>.

[34] X. Fu, S. Deja, B. Kucejova, J.A.G. Duarte, J.G. McDonald, S.C. Burgess, Targeted determination of tissue energy status by LC-MS/MS, *Anal. Chem.* 91 (2019) 5881–5887, <https://doi.org/10.1021/acs.analchem.9b00217>.

[35] J. Hrlicko, L. Rudl Kulhava, M. Paucova, M. Novakova, O. Kuda, O. Fiehn, T. Cajka, Short-term stability of serum and liver extracts for untargeted metabolomics and lipidomics, *Antioxidants* 12 (2023) 986, <https://doi.org/10.3390/antiox12050986>.

[36] X. Du, J.J. Aristizabal-Henao, T.J. Garrett, M. Brochhausen, W.R. Hogan, D. J. Lemus, A checklist for reproducible computational analysis in clinical metabolomics research, *Metabolites* 12 (2022) 87, <https://doi.org/10.3390/metabo12010087>.

[37] B. Álvarez-Sánchez, F. Priego-Capote, M.D.L. Castro, Metabolomics analysis II. Preparation of biological samples prior to detection, *Trends Anal. Chem.* 29 (2010) 120–127, <https://doi.org/10.1016/j.trac.2009.12.004>.

[38] R.-J. Raterink, P.W. Lindenburg, R.J. Vreeken, R. Ramautar, T. Hankemeier, Recent developments in sample-pretreatment techniques for mass spectrometry-based metabolomics, *Trends Anal. Chem.* 61 (2014) 157–167, <https://doi.org/10.1016/j.trac.2014.06.003>.

[39] 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>.

[40] J. Medina, V. van der Velzen, T. Teav, Y. Guittot, 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>.

[41] V. Matyash, G. Liebisch, T.V. Kurzhalia, 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>.

[42] J. Folch, M. Lees, G.H. Sloane 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](https://doi.org/10.1016/S0021-9258(18)64849-5).

[43] 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/DANO1319F>.

[44] 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>.

[45] 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.jchroma.2015.07.048>.

[46] D. Vuckovic, in: H.J. Issaq, T.D. Veenstra (Eds.), Chapter 4 - Sample Preparation in Global Metabolomics of Biological Fluids and Tissues, Academic Press, Boston, 2020, pp. 53–83.

[47] X.L. Guan, D.P.S. Chang, Z.X. Mok, B. Lee, Assessing variations in manual pipetting: an under-investigated requirement of good laboratory practice, *J. Mass Spectrom. Adv. Clin. Lab.* 30 (2023) 25–29, <https://doi.org/10.1016/j.jmsacl.2023.09.001>.

[48] C.-H. Yao, G.-Y. Liu, K. Yang, R.W. Gross, G.J. Patti, Inaccurate quantitation of palmitate in metabolomics and isotope tracer studies due to plastics, *Metabolomics* 12 (2016) 143, <https://doi.org/10.1007/s11306-016-1081-y>.

[49] L. Löfgren, G.-B. Forsberg, M. Ståhlman, The BUME method: a new rapid and simple chloroform-free method for total lipid extraction of animal tissue, *Sci. Rep.* 6 (2016) 27688, <https://doi.org/10.1038/srep27688>.

[50] M.R. Showalter, E.B. Nonnecke, A.L. Linderholm, T. Cajka, M.R. Sa, B. Lönnerdal, N.J. Kenyon, O. Fiehn, Obesogenic diets alter metabolism in mice, *PLoS One* 13 (2018) e0190632, <https://doi.org/10.1371/journal.pone.0190632>.

[51] M. Lepoittevin, Q. Blancart-Remaury, T. Kerforne, L. Pellerin, T. Hauet, R. Thuillier, Comparison between 5 extraction methods in either plasma or serum to determine the optimal extraction and matrix combination for human metabolomics, *Cell. Mol. Biol. Lett.* 28 (2023) 43, <https://doi.org/10.1186/s11658-023-00452-x>.

[52] M. Jain, S. Ngoy, S.A. Sheth, R.A. Swanson, E.P. Rhee, R. Liao, C.B. Clish, V. K. Mootha, R. Nilsson, A systematic survey of lipids across mouse tissues, *Am. J. Physiol. Endocrinol. Metab.* 306 (2014) E854–E868, <https://doi.org/10.1152/ajpendo.00371.2013>.

[53] J. Sostare, R. Di Guida, J. Kirwan, K. Chalal, E. Palmer, W.B. Dunn, M.R. Viant, Comparison of modified Matyash method to conventional solvent systems for

polar metabolite and lipid extractions, *Anal. Chim. Acta* 1037 (2018) 301–315, <https://doi.org/10.1016/j.aca.2018.03.019>.

[54] M.H. Sarafian, M. Gaudin, M.R. Lewis, F.P. Martin, E. Holmes, J.K. Nicholson, M. E. Dumas, Objective set of criteria for optimization of sample preparation procedures for ultra-high throughput untargeted blood plasma lipid profiling by ultra performance liquid chromatography-mass spectrometry, *Anal. Chem.* 86 (2014) 5766–5774, <https://doi.org/10.1021/ac500317c>.

[55] S. Naz, M. Vallejo, A. García, C. Barbas, Method validation strategies involved in non-targeted metabolomics, *J. Chromatogr. A* 1353 (2014) 99–105, <https://doi.org/10.1016/j.chroma.2014.04.071>.

[56] C.T. Viswanathan, S. Bansal, B. Booth, A.J. DeStefano, M.J. Rose, J. Sailstad, V. P. Shah, J.P. Skelly, P.G. Swann, R. Weiner, Quantitative bioanalytical methods validation and implementation: best practices for chromatographic and ligand binding assays, *Pharm. Res. (N. Y.)* 24 (2007) 1962–1973, <https://doi.org/10.1007/s11095-007-9291-7>.

[57] E.M. Harrieler, F. Kretschmer, S. Böcker, M. Witting, Current state-of-the-art of separation methods used in LC-MS based metabolomics and lipidomics, *J. Chromatogr. B* 1188 (2022) 123069, <https://doi.org/10.1016/j.jchromb.2021.123069>.

[58] M. Lange, Z. 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>.

[59] T. Cajka, O. Fiehn, Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry, *Trends Anal. Chem.* 61 (2014) 192–206, <https://doi.org/10.1016/j.trac.2014.04.017>.

[60] 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>.

[61] K. Schönberger, M. Mitterer, J.M. Buescher, N. Cabezas-Wallscheid, Targeted LC-MS/MS-based metabolomics and lipidomics on limited hematopoietic stem cell numbers, *STAR Protocols* 3 (2022) 101408, <https://doi.org/10.1016/j.xpro.2022.101408>.

[62] X. Domingo-Almenara, J.R. Montenegro-Burke, C. Guijas, E.L.W. Majumder, H. P. Benton, G. Siuzdak, Autonomous METLIN-guided in-source fragment annotation for untargeted metabolomics, *Anal. Chem.* 91 (2019) 3246–3253, <https://doi.org/10.1021/acs.analchem.8b03126>.

[63] M. Giera, A. Aisporna, W. Uritboonthai, G. Siuzdak, The hidden impact of in-source fragmentation in metabolic and chemical mass spectrometry data interpretation, *Nat. Metab.* (2024), <https://doi.org/10.1038/s42255-024-01076-x>.

[64] C. Jaeger, J. Liseć, Towards unbiased evaluation of ionization performance in LC-HRMS metabolomics method development, *Metabolites* 12 (2022) 426, <https://doi.org/10.3390/metabo12050426>.

[65] E. Defossez, J. Bourquin, S. von Reuss, S. Rasmann, G. Glauser, Eight key rules for successful data-dependent acquisition in mass spectrometry-based metabolomics, *Mass Spectrom. Rev.* 42 (2023) 131–143, <https://doi.org/10.1002/mas.21715>.

[66] 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>.

[67] J. Rubert, M. Zachariasova, J. Hajslova, Advances in high-resolution mass spectrometry based on metabolomics studies for food – a review, *Food Addit. Contam.* 32 (2015) 1685–1708, <https://doi.org/10.1080/19440049.2015.1084539>.

[68] M.P. Hodson, G.J. Dear, J.L. Griffin, J.N. Haselden, An approach for the development and selection of chromatographic methods for high-throughput metabolomic screening of urine by ultra pressure LC-ESI-ToF-MS, *Metabolomics* 5 (2009) 166–182, <https://doi.org/10.1007/s11306-008-0135-1>.

[69] L. Najdekr, D. Friedecký, R. Tautenhahn, T. Pluskal, J. Wang, Y. Huang, T. Adam, Influence of mass resolving power in orbital ion-trap mass spectrometry-based metabolomics, *Anal. Chem.* 88 (2016) 11429–11435, <https://doi.org/10.1021/acs.analchem.6b02319>.

[70] M. Lopes, K. Brejchová, M. Riecan, M. Novakova, M. Rossmeisl, T. Cajka, O. Kuda, Metabolomics atlas of oral 13C-glucose tolerance test in mice, *Cell Rep.* 37 (2021) 109833, <https://doi.org/10.1016/j.celrep.2021.109833>.

[71] W.J. Nash, W.B. Dunn, From mass to metabolite in human untargeted metabolomics: recent advances in annotation of metabolites applying liquid chromatography-mass spectrometry data, *Trends Anal. Chem.* 120 (2019) 115324, <https://doi.org/10.1016/j.trac.2018.11.022>.

[72] J. Guo, T. Huan, Comparison of full-scan, data-dependent, and data-independent acquisition modes in liquid chromatography-mass spectrometry based untargeted metabolomics, *Anal. Chem.* 92 (2020) 8072–8080, <https://doi.org/10.1021/acs.analchem.9b05135>.

[73] T. Cajka, J. Hrlicko, L. Rudl Kulhava, M. Paucova, M. Novakova, O. Kuda, Optimization of mobile phase modifiers for fast LC-MS-based untargeted metabolomics and lipidomics, *Int. J. Mol. Sci.* 24 (2023) 1987, <https://doi.org/10.3390/ijms24031987>.

[74] T. Cajka, R. Davis, K. Austin, J. Newman, B. German, O. Feiñn, J. Smilowitz, Using a lipidomics approach for nutritional phenotyping in response to a test meal containing gamma-linolenic acid, *Metabolomics* 12 (2016) 127, <https://doi.org/10.1007/s11306-016-1075-9>.

[75] S. Chen, M. Hoene, J. Li, Y. Li, X. Zhao, H.U. Häring, E.D. Schleicher, C. Weigert, G. 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>.

[76] W. Lv, L. Guo, F. Zheng, Q. Wang, W. Wang, L. Cui, Y. Ouyang, X. Liu, E. Li, X. Shi, G. Xu, Alternate reversed-phase and hydrophilic interaction liquid chromatography coupled with mass spectrometry for broad coverage in metabolomics analysis, *J. Chromatogr. B* 1152 (2020) 122266, <https://doi.org/10.1016/j.jchromb.2020.122266>.

[77] A. Lindahl, S. Sääf, J. Lehtiö, A. Nordström, Tuning metabolome coverage in reversed phase LC-MS metabolomics of MeOH extracted samples using the reconstitution solvent composition, *Anal. Chem.* 89 (2017) 7356–7364, <https://doi.org/10.1021/acs.analchem.7b00475>.

[78] M. Kaczmarek, N. Zhang, L. Buzhansky, S. Gilead, E. Gazit, Optimization strategies for mass spectrometry-based untargeted metabolomics analysis of small polar molecules in human plasma, *Metabolites* 13 (2023) 923, <https://doi.org/10.3390/metabo13080923>.

[79] P.I. Benke, B. Burla, K. Ekorroos, M.R. Wenk, F. Torta, Impact of ion suppression by sample cap liners in lipidomics, *Anal. Chim. Acta* 1137 (2020) 136–142, <https://doi.org/10.1016/j.aca.2020.09.055>.

[80] A. Lenk, Solvents: an overlooked ally for liquid chromatography-mass spectrometry, *Column* 14 (2018) 19–21.

[81] T. Cajka, J. Hrlicko, L. Rudl Kulhava, M. Paucova, M. Novakova, O. Fiehn, O. Kuda, Exploring the impact of organic solvent quality and unusual adduct formation during LC-MS-based lipidomic profiling, *Metabolites* 13 (2023) 966, <https://doi.org/10.3390/metabo13090966>.

[82] B.O. Keller, J. Sui, A.B. Young, R.M. Whittal, Interferences and contaminants encountered in modern mass spectrometry, *Anal. Chim. Acta* 627 (2008) 71–81, <https://doi.org/10.1016/j.aca.2008.04.043>.

[83] 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. Cremer, E.A. Dennis, J.E. Evans, A. Fauland, O. Fiehn, M.S. Gardner, T.J. Garrett, K.H. Gotlinger, J. Han, Y. Huang, A.H. Neo, T. Hyötyläinen, Y. Izumi, H. Jiang, H. Jiang, J. Jiang, M. Kachman, R. Kiyonami, K. Klavins, C. Klose, H.C. Köfeler, 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. Qiu, O. Quehenberger, P. Ramrups, 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. St John-Williams, M.A. Surma, H. Takeda, R. Thakare, J. W. Thompson, F. Torta, A. Triebel, M. Trötzmüller, S.J.K. Ubhayasekera, D. Vuckovic, J.M. Weir, R. Welti, M.R. Wenk, C.E. Wheelock, L. Yao, M. Yuan, X. H. Zhao, S. 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>.

[84] 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>.

[85] 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>.

[86] 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>.

[87] M. Katajamaa, M. Oresic, Data processing for mass spectrometry-based metabolomics, *J. Chromatogr. A* 1158 (2007) 318–328, <https://doi.org/10.1016/j.chroma.2007.04.021>.

[88] Z. Li, Y. Lu, Y. Guo, H. Cao, Q. Wang, W. 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>.

[89] O.D. Myers, S.J. Sumner, S. Li, S. Barnes, 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>.

[90] 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/10.1012/acs.analchem.7b03404>.

[91] M. Croset, N. Brossard, A. Polette, M. Lagarde, Characterization of plasma unsaturated lysophosphatidylcholines in human and rat, *Biochem. J.* 345 (1999) 61–67, <https://doi.org/10.1042/bj3450061>.

[92] R. Chaleckis, I. Meister, P. Zhang, C.E. Wheelock, Challenges, progress and promises of metabolite annotation for LC-MS-based metabolomics, *Curr. Opin. Biotechnol.* 55 (2019) 44–50, <https://doi.org/10.1016/j.copbio.2018.07.010>.

[93] I. Blaženović, T. Kind, J. Ji, O. Fiehn, Software tools and approaches for compound identification of LC-MS/MS data in metabolomics, *Metabolites* 8 (2018) 31, <https://doi.org/10.3390/metabo8020031>.

[94] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W. M. 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, *Metabolomics* 3 (2007) 211–221, <https://doi.org/10.1007/s11306-007-0082-2>.

[95] E. Rampler, Y.E. 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>.

[96] 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, vol. 48, 2014, pp. 2097–2098, <https://doi.org/10.1021/es5002105>.

[97] G. Liebisch, E. Fahy, J. Aoki, E.A. Dennis, T. Durand, C.S. Ejsing, M. Fedorova, I. Feussner, W.J. Griffiths, H. Köfeler, A.H. Merrill Jr., 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>.

[98] J. 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>.

[99] T. Cajka, O. Fiehn, LC-MS-based lipidomics and automated identification of lipids using the LipidBlast in-silico MS/MS library, *Methods Mol. Biol.* 1609 (2017) 149–170, https://doi.org/10.1007/978-1-4939-6996-8_14.

[100] Y. Cai, Z. Zhou, Z.-J. Zhu, Advanced analytical and informatic strategies for metabolite annotation in untargeted metabolomics, *Trends Anal. Chem.* 158 (2023) 116903, <https://doi.org/10.1016/j.trac.2022.116903>.

[101] C. Hoang, W. Uritboonthai, L. Hoang, E.M. Billings, A. Aisporna, F.A. Nia, R.J. E. Derk, J.R. Williamson, M. Giera, G. Siuzdak, Tandem mass spectrometry across platforms, *Anal. Chem.* (2024), <https://doi.org/10.1021/acs.analchem.3c05576>.

[102] 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>.

[103] 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>.

[104] E.S. Baker, C. Hoang, W. Uritboonthai, H.M. Heyman, B. Pratt, M. MacCoss, B. MacLean, R. Plumb, A. Aisporna, G. Siuzdak, METLIN-CCS: an ion mobility spectrometry collision cross section database, *Nat. Methods* 20 (2023) 1836–1837, <https://doi.org/10.1038/s41592-023-02078-5>.

[105] N.F. de Jonge, K. Mildau, D. Meijer, J.J.R. Louwen, C. Bueschl, F. Huber, J.J. J. van der Hooft, Good practices and recommendations for using and benchmarking computational metabolomics metabolite annotation tools, *Metabolomics* 18 (2022) 103, <https://doi.org/10.1007/s11306-022-01963-y>.

[106] 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>.

[107] C.A. Kretzler, G.G. Thallinger, A map of mass spectrometry-based *in silico* fragmentation prediction and compound identification in metabolomics, *Brief. Bioinform.* 22 (2021) 1–25, <https://doi.org/10.1093/bib/bbab073>.

[108] Z. Zhou, M. Luo, H. Zhang, Y. Yin, Y. Cai, Z.-J. Zhu, Metabolite annotation from knowns to unknowns through knowledge-guided multi-layer metabolic networking, *Nat. Commun.* 13 (2022) 6656, <https://doi.org/10.1038/s41467-022-34537-6>.

[109] M. Witting, S. Böcker, Current status of retention time prediction in metabolite identification, *J. Separ. Sci.* 43 (2020) 1746–1754, <https://doi.org/10.1002/jssc.202000060>.

[110] 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>.

[111] A. Damont, A. Legrand, C. Cao, F. Fenaille, J.-C. Tabet, Hydrogen/deuterium exchange mass spectrometry in the world of small molecules, *Mass Spectrom. Rev.* 42 (2023) 1300–1331, <https://doi.org/10.1002/mas.21765>.

[112] T. Cajka, J. Hrlicko, S. Rakusanova, K. Brejchova, M. Novakova, L. Rudl Kulhava, V. Hola, M. Paucova, O. Fiehn, O. Kuda, Hydrophilic interaction liquid chromatography–hydrogen/deuterium exchange–mass spectrometry (HILIC-HDX-MS) for untargeted metabolomics, *Int. J. Mol. Sci.* 25 (2024) 2899, <https://doi.org/10.3390/ijms25052899>.

[113] N.G. Mahieu, G.J. Patti, Systems-level annotation of a metabolomics data set reduces 25 000 features to fewer than 1000 unique metabolites, *Anal. Chem.* 89 (2017) 10397–10406, <https://doi.org/10.1021/acs.analchem.7b02380>.

[114] 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>.

[115] 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, *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* 1862 (2017) 766–770, <https://doi.org/10.1016/j.bbapli.2017.02.016>.

[116] T. Skotland, K. Ekoos, J. McDonald, R. Ahrends, G. Liebisch, K. Sandvig, Pitfalls in lipid mass spectrometry of mammalian samples — a brief guide for biologists, *Nat. Rev. Mol. Cell Biol.* (2024), <https://doi.org/10.1038/s41580-024-00758-4>.

[117] O. Begou, H.G. Gika, G.A. Theodoridis, I.D. Wilson, Quality control and validation issues in LC-MS metabolomics, *Methods Mol. Biol.* 1738 (2018) 15–26, https://doi.org/10.1007/978-1-4939-7643-0_2.

[118] 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, C. 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>.

[119] M.R. Lewis, J.T.M. Pearce, K. Spagou, M. Green, A.C. Dona, A.H.Y. Yuen, M. David, D.J. Berry, K. Chappell, V. Horneffer-van der Sluis, R. Shaw, S. Lovestone, P. Elliott, J. Shockcor, J.C. Lindon, O. Cloarec, Z. Takats, E. Holmes, J.K. Nicholson, Development and application of ultra-performance liquid chromatography-TOF MS for precision large scale urinary metabolic phenotyping, *Anal. Chem.* 88 (2016) 9004–9013, <https://doi.org/10.1021/acs.analchem.6b01481>.

[120] 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. Hümer, A. Kalli, T. Koal, S. Koeniger, R. Mandal, F. Meier, F.J. Naser, D. O'Neil, A. Pal, G. J. Patti, H. Pham-Tuan, 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. 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>.

[121] A. Olivieri, O.S. Degenhardt, G.R. McDonald, D. Narang, I.M. Paulsen, J. L. Kozuska, A. Holt, On the disruption of biochemical and biological assays by chemicals leaching from disposable laboratory plasticware, *Can. J. Physiol. Pharmacol.* 90 (2012) 697–703, <https://doi.org/10.1139/y2012-049>.

[122] 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>.

[123] 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>.

[124] S. 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>.

[125] Y. Wu, L. Li, Sample normalization methods in quantitative metabolomics, *J. Chromatogr. A* 1430 (2016) 80–95, <https://doi.org/10.1016/j.chroma.2015.12.007>.

[126] Q. Yang, J. Hong, Y. Li, W. Xue, S. Li, H. Yang, F. Zhu, A novel bioinformatics approach to identify the consistently well-performing normalization strategy for current metabolomic studies, *Briefings Bioinf.* 21 (2020) 2142–2152, <https://doi.org/10.1093/bib/bbz137>.

[127] 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>.

[128] C. Muschet, G. Möller, 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>.

[129] 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>.

[130] P. Sedgwick, Multiple significance tests: the Bonferroni correction, *BMJ* 344 (2012) e509, <https://doi.org/10.1136/bmj.e509>.

[131] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Stat. Soc. Ser. B* 57 (1995) 289–300, <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.

[132] M.L. McHugh, Multiple comparison analysis testing in ANOVA, *Biochem. Med.* 21 (2011) 203–209, <https://doi.org/10.11613/BM.2011.029>.

[133] S. Ren, A.A. Hinman, 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>.

[134] Z. Pang, J. Chong, G. Zhou, D.A. de Lima Morais, L. Chang, M. Barrette, C. Gauthier, P.-É. Jacques, S. Li, J. 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>.

[135] M. Vondrackova, D. Kopczynski, N. Hoffmann, O. Kuda, LORA, Lipid Over-Representation Analysis based on structural information, *Anal. Chem.* 95 (2023) 12600–12604, <https://doi.org/10.1021/acs.analchem.3c02039>.

[136] G. Wohlgemuth, P.K. Haldiya, E. Willighagen, T. Kind, O. Fiehn, The Chemical Translation Service—a web-based tool to improve standardization of metabolomics reports, *Bioinformatics* 26 (2010) 2647–2648, <https://doi.org/10.1093/bioinformatics/btq476>.

[137] D.J. McCarthy, G.K. Smyth, Testing significance relative to a fold-change threshold is a TREAT, *Bioinformatics* 25 (2009) 765–771, <https://doi.org/10.1093/bioinformatics/btp053>.

[138] N. Akarachantachote, S. Chadham, K. Saithanu, Cutoff threshold of variable importance in projection for variable selection, *Int. J. Pure Appl. Math.* 94 (2014), <https://doi.org/10.12732/ijppm.v94i3.2>.

[139] D. Pieragostino, M. D'Alessandro, M. di Iorio, C. Rossi, M. Zucchini, A. Urbani, C. Di Ilio, A. Lugarasi, P. Sacchetta, P. Del Boccio, An integrated metabolomics approach for the research of new cerebrospinal fluid biomarkers of multiple

sclerosis, Mol. Biosyst. 11 (2015) 1563–1572, <https://doi.org/10.1039/C4MB00700J>.

[140] J.A. Westerhuis, H.C.J. Hoefsloot, S. Smit, D.J. Vis, A.K. Smilde, E.J.J. van Velzen, J.P.M. van Duijnhoven, F.A. van Dorsten, Assessment of PLSDA cross validation, Metabolomics 4 (2008) 81–89, <https://doi.org/10.1007/s11306-007-0099-6>.

[141] M.N. Triba, L. Le Moyec, R. Amathieu, C. Goossens, N. Bouchemal, P. Nahon, D. N. Rutledge, P. Savarin, PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters, Mol. Biosyst. 11 (2015) 13–19, <https://doi.org/10.1039/C4MB00414K>.

[142] E. Nemutlu, S. Zhang, A. Gupta, N.O. Juranic, S.I. Macura, A. Terzic, A. Jahangir, P. Dzeja, Dynamic phosphometabolomic profiling of human tissues and transgenic models by 18O-assisted ³¹P NMR and mass spectrometry, Physiol. Genom. 44 (2012) 386–402, <https://doi.org/10.1152/physiolgenomics.00152.2011>.

[143] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva 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 Schalk, 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, The FAIR guiding principles for scientific data management and stewardship, Sci. Data 3 (2016) 160018, <https://doi.org/10.1038/sdata.2016.18>.

[144] R.A. Spicer, C. Steinbeck, A lost opportunity for science: journals promote data sharing in metabolomics but do not enforce it, Metabolomics 14 (2017) 16, <https://doi.org/10.1007/s11306-017-1309-5>.

[145] M. Witting, (Re-)use and (re-)analysis of publicly available metabolomics data, Proteomics 23 (2023) 2300032, <https://doi.org/10.1002/pmic.202300032>.

[146] J.G. McDonald, C.S. Ejsing, D. Kopczynski, M. Holcapek, J. Aoki, M. Arita, M. Arita, E.S. Baker, J. Bertrand-Michel, J.A. Bowden, B. Brügger, S.R. Ellis, M. Fedorova, W.J. Griffiths, X. Han, J. Hartler, N. Hoffmann, J.P. Koelmel, H. C. Köfeler, T.W. Mitchell, V.B. O'Donnell, D. Saigusa, D. Schwudke, A. Shevchenko, C.Z. Ulmer, M.R. Wenk, M. Witting, D. Wolrab, Y. Xia, R. Ahrends, G. Liebisch, K. Ekoos, Introducing the lipidomics minimal reporting checklist, Nat. Metab. 4 (2022) 1086–1088, <https://doi.org/10.1038/s42255-022-00628-3>.