Journal of Pharmaceutical and Biomedical Analysis 147 (2018) 149–173

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

Review

Quality assurance procedures for mass spectrometry untargeted metabolomics. a review

Danuta Dudzik[1], Cecilia Barbas-Bernardos[1], Antonia García, Coral Barbas*

Center for Metabolomics and Bioanalysis (CEMBIO), Faculty of Pharmacy, San Pablo CEU University, Boadilla del Monte, ES-28668, Madrid, Spain

ARTICLE INFO

Article history:
Available online 5 August 2017

Keywords:
Quality control
Unwanted variation
Bias
Standard operating procedures (SOPs)
Untargeted/non-targeted metabolomics
Uncertainty
Validation

ABSTRACT

Untargeted metabolomics, as a global approach, has already proven its great potential and capabilities for the investigation of health and disease, as well as the wide applicability for other research areas. Although great progress has been made on the feasibility of metabolomics experiments, there are still some challenges that should be faced and that includes all sources of fluctuations and bias affecting every step involved in multiplatform untargeted metabolomics studies. The identification and reduction of the main sources of unwanted variation regarding the pre-analytical, analytical and post-analytical phase of metabolomics experiments is essential to ensure high data quality. Nowadays, there is still a lack of information regarding harmonized guidelines for quality assurance as those available for targeted analysis.

In this review, sources of variations to be considered and minimized along with methodologies and strategies for monitoring and improvement the quality of the results are discussed. The given information is based on evidences from different groups among our own experiences and recommendations for each stage of the metabolomics workflow. The comprehensive overview with tools presented here might serve other researchers interested in monitoring, controlling and improving the reliability of their findings by implementation of good experimental quality practices in the untargeted metabolomics study.

Contents

Abbreviations: AIS, Multiple internal standard; BBMRI, Biomolecular resources research infrastructure; BPC, Base peak chromatogram; CAL samples, calibration samples; CAWG, Chemical analysis working group; CRM, Certified reference material; CV, Coefficient of variation; DBS, Dried blood spots; DIMS, Direct infusion mass spectrometry; DSD, Data-driven sample size determination; ECC, Extracted compound chromatogram; EDTA, Ethylenediaminetetraacetic acid; EIC, Extracted ion chromatogram; EuPA, EuPA biobank initiative; FBSC, Feature-based signal correction; FC, Fold change; FDR, False discovery rate; FFPE, Formalin-fixed paraffin-embedded; IS, Internal standard; LN_2, Liquid nitrogen; LOESS, Locally estimated scatterplot smoothing; MIIS, Matrix-induced ion suppression; MP, Mobile phase; MR, Mendelian randomization; MSI, Metabolomics Standards Initiative; MTBE, Methyl-*tert*-butyl ether; PQN, Probabilistic quotient normalization; Q, Qualification; QC-RLSC, Quality control-robust loess signal correction; QC-RSC, Quality control-robust spline correction; QC-SVRC, Quality control samples support vector regression; RM, Reference material; RT, Retention time; SDSD, Standard deviation step down; SIS, Single internal standard; SOP, Standard operating procedure; SST, System suitability test; UMS, Universal metabolome-standard; UPLC, Ultra high performance liquid chromatography.

* Corresponding author at: Center for Metabolomics and Bioanalysis (CEMBIO), Faculty of Pharmacy San Pablo CEU University, Campus Monteprincipe, Boadilla del Monte, 28668 Madrid, Spain.

E-mail addresses: danuta.dudzik@ceu.es (D. Dudzik), cec.barbas.ce@ceindo.ceu.es (C. Barbas-Bernardos), antogar@ceu.es (A. García), cbarbas@ceu.es (C. Barbas).

[1] Both authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jpba.2017.07.044
0731-7085/

1. Introduction

Metabolomics is the part of the – omics fields that focuses on the wide variety of low molecular weight metabolites present in biological samples (urine, blood, tissues etc.). Metabolites are present in the body either from endogenous sources, the different chemical reactions taking place in the organism, or from exogenous sources (diet, drugs, xenobiotics etc.). Therefore, metabolites have the advantage of showing directly the effect of gene expression, intrinsic metabolism, dietary habits, drug metabolism or even gut microbiota interactions. Indeed, metabolites include both endogenous and exogenous compounds belonging to different (bio)chemical classes (amino acids, fatty acids, lipids, alkaloids, acids, drugs metabolites, food metabolites etc.).

Nowadays the most common techniques to perform metabolite analysis are NMR and MS, the latter usually coupled to different separation techniques (LC, GC or CE) for the largest metabolite coverage with the highest sensitivity. Both can identify and quantify and both serve for targeted and untargeted metabolomics studies. The targeted approach is hypothesis-driven, while the untargeted one does not require a priori hypothesis. This review will focus only on untargeted MS metabolomics studies. Considering the analytical techniques coupled to MS, GC–MS is the oldest and the more robust coupling and moreover the most reproducible in relation to retention times (RT) and mass spectra, along with the accuracy and reliability of the peak assignment by comparison with spectral libraries (home-made and commercial ones). On the other hand, only a small set of compounds that can be made volatile can be analysed. It is generally considered a complementary technique to CE-MS and LC–MS, being the later the most universal considering the broad type of mechanisms, stationary phases and mobile phases, but with some drawbacks derived from signal suppression and the lack of mass spectral libraries. Meanwhile, CE-MS is a good choice for polar and charged metabolites in aqueous samples or samples with high saline content with minimal pretreatment and highest efficiency but with the drawbacks of lower sensitivity and migration time reproducibility.

As the aim of untargeted metabolomics is to measure as many metabolites as possible, including unknowns, to generate the metabolic fingerprint characteristic for a biological sample, the common way to work in untargeted metabolomics is by differential analysis. With that goal, fingerprints from different conditions (e.g., cases and controls) are overlaid and compared to select those signals with statistically significant differences. Afterwards, signals should be converted into identified compounds. This identification may result in the generation of new hypothesis, in the discovery of novel biochemical interactions (metabolic associations) or could generate potential novel biomarkers [1–5]. These approaches have also been used in drug development processes, food analysis [6–8] or more recently, in epidemiological studies [9–13], just to quote some examples.

This way of working based on differential analysis, leads to an urgent need to unify protocols to prevent variations and artifacts avoiding false discoveries describing differences without biological relevance and generating erroneous hypothesis [14]. Awareness should be raised about the need of standardization minimizing bias: standardize the conditions in which samples are collected, handled, pre-processed and analysed [15–17]; standardize the procedures to produce reliable and reproducible results in terms of data treatment and compound identification. Within this context the term "validation" appears, although there are different levels of validation: analytical, statistical and biological.

Analytical method validation implies a workflow with many challenging steps covering from sample collection to generating the analytical signal. The validation of analytical methods for biological samples in target analysis is clearly established [18,19], however researchers should agree how to prove that a method is suitable for its intended purpose in untargeted metabolomics.

Statistical validation is part of the following stage of the process where mathematical models should be validated [20] to avoid overfitting and discovery of markers by chance.

Biological validation is a complex process that could refer to test a hypothesis generated through metabolomics data in a proper biological system, but also in a second cohort, with more individuals per group [5,21,22].

Herein issues related to unwanted variation introduced in metabolomics studies are summarized together with tools and strategies available in order to apply quality assessment procedures in the analytical stage of metabolomics workflow. This review covers the current state of knowledge of quality assurance practices, that have been already proposed by some research groups and that are currently under a hot debate on metabolomics society [23–25]. In 2007 the Chemical Analysis Working Group (CAWG) as part of the Metabolomics Standards Initiative (MSI) proposed minimum

reporting standards related to metabolomics experiments including: sample preparation, experimental analysis, quality control, metabolite identification, and data pre-processing. That recommendations should be considered, developed and implemented to the general workflow of metabolomics based studies [26]. Data reporting in the literature and our own experience will be provided focused on general considerations for good laboratory practices in untargeted metabolomics studies.

2. Sources of variation in metabolomics workflow

In many cases, despite of all the efforts to design a robust study and perform appropriate statistical analysis, the results can result unsound. As illustrated in Fig. 1, variation can affect any stage of the metabolomics workflow, from study design, through sample handling, collection, pre-processing, measurement, to the final data treatment and interpretation. It may be caused by biological factors (e.g., gender, age, body mass index, diurnal cycle, smoking habits, etc.,), non-harmonized procedures, bias or it may be due to unexplainable random variation [27]. While appropriate statistical methods can reduce the effect of unwanted variation, they may not be able to eliminate it entirely [28]. However, understanding the causes of unwanted variation in a given metabolomics experiment and its removal can pose a challenging task. The overall unwanted variation factors that could alter the results from the metabolomics experiment are summarized in Fig. 1 and detailed below.

2.1. Pre-analytical

The term pre-analytical includes any process before sample measurement, such as (i) study design; (ii) sample handling; (iii) pre-processing and (iv) metabolite extraction. It has been proven that pre-analytical changes can have a major impact on the integrity and quality of samples and can significantly affect analytical results and decrease the reliability of the research outcomes [16,29]. Recently, metabolomics has moved toward defining Standard Operating Procedures (SOPs) for the pre-analytical phase [25,30,31], however complete standardization and guidelines for sample handling and pre-processing are not yet available.

2.1.1. Sample handling

Variation in sample handling-sample collection, processing and storage can significantly alter the molecular composition and overall quality of the samples. The major points that need to be considered are: (i) inactivation of ongoing metabolism; (ii) metabolite stability during sample pre-processing and (iii) sample integrity during storage [32]. All aspects of sample handling and pre-processing need to be carefully evaluated and described in experimental protocols before starting the study. Some guidelines have been already proposed [25,26,33] but still there are no official SOPs established.

- **Sample collection**

The choice of sample type and sampling method is an important consideration and should be directly related to the objectives of the study. Metabolomics aims to obtain an accurate snapshot of metabolites levels, that should prevent significant changes arising from enzymatic activity, degradation or originating from the environment. Typically blood [24,34] and urine [35] samples are collected but other biofluids (e.g., saliva [36], cerebrospinal fluid [37], amniotic fluid [2,38], milk [39] etc.), cell culture [40] or tissues [41–43] have also been used in metabolomics studies.

Plasma and serum: Both biofluids are frequently analyzed in metabolomics platforms. Different collecting procedures and the coagulation cascade influence concentrations of metabolites as those described by Yu et al. These group in a comprehensive study found a higher sensitivity of metabolites in serum but better reproducibility in plasma [44]. Collection of blood derived samples has been previously discussed in detail [16,17,45,46]. Yin et al. reviewed the effects of pre-analytical conditions on blood samples used in metabolomics studies [31] and they recommend, based on evidences, EDTA (ethylenediaminetetraacetic acid) as anticoagulant instead of heparin to diminish chemical noise in the mass spectra, avoid hemolyzed samples and place whole blood immediately in ice cold water. A systematic evaluation of different parameters for plasma and serum samples pre-processing has also been reported by Jobard et al. [30] and Tuck et al. [45] concluding that harmonized standard operation procedures (SOPs) should be implemented as small differences in SOPs between groups or within groups could yield potentially uninterpretable or biased results due to variations at multiple steps in the collection and handling process. More recently Hernandes et al. provided comprehensive overview of the most important factors for blood derived samples in aspect of metabolomics studies and keep them unaltered, that includes: fasting individuals, the choice of the same anticoagulant for all the samples in the study and how it influences in the matrix effect, why hemolyzed sample should be avoided, as well as how different clotting time and room temperature during sample processing can introduce variation in metabolite contents and blood plasma and serum stability, including considerations for Dried Blood Spots (DBS) samples [33].

Urine: Urine is a commonly used for metabolomics investigation due to non-invasive collection, the complex nature, and the ability to collect multiple samples over a period of time. However, it is important to determine appropriate sampling time (24 h collection *vs.* random or timed). 24 h sampling collection is preferred to eliminate large variability in metabolite profiles obtained in shorter collection periods and once fixed keep it for the whole study as diurnal cycle along with water intake are big sources of variation in urine metabolite content. The bacterial growing in urine is very rapid and preservatives (e.g., boric acid or sodium azide) have been used in clinical practice to prevent bacterial contamination that is known to impact on metabolite composition [47,48]. Although effective in preventing bacterial growth, preservatives are not recommended as those compounds could promote formation of chemical complexes in the samples or cause problems of sample handling consistency and overall sample quality [49]. Rox et al. suggested that boric acid could impact on the electrospray ionization and provoke chemical instability [47]. Instead, the samples should be rather collected on ice and frozen as soon as possible. In addition to bacterial contamination, Boomsma et al. reported that some labile compounds like catecholamines are easily oxidized to form quinones which is related to temperature and sample processing time [50]. Several studies as the one reported by Bando et al. investigated the influence of urine and plasma sample collection and handling procedures on GC–MS based metabolomics approach [51] and Rox et al. conducted the study with similar objectives with the urine samples analysed by LC–MS [47]. Both studies highlighted the fundamental meaning of the pooling period and the indispensable necessity to maintain low temperature during sample collection (4 °C or preferable collection on ice) that limits changes in metabolite content and is the most effective sampling condition. In addition, as 24 h collection could not be feasible in many studies, the factor of different dilutions should be considered as a source of variation for further normalization. Detailed description of all pre-analytical factors that should be considered prior to urine collection and preparation for metabolomics analysis could be found in the work of Fernández-Peralbo and Luque de Castro [49] or Khamis et al. [52].

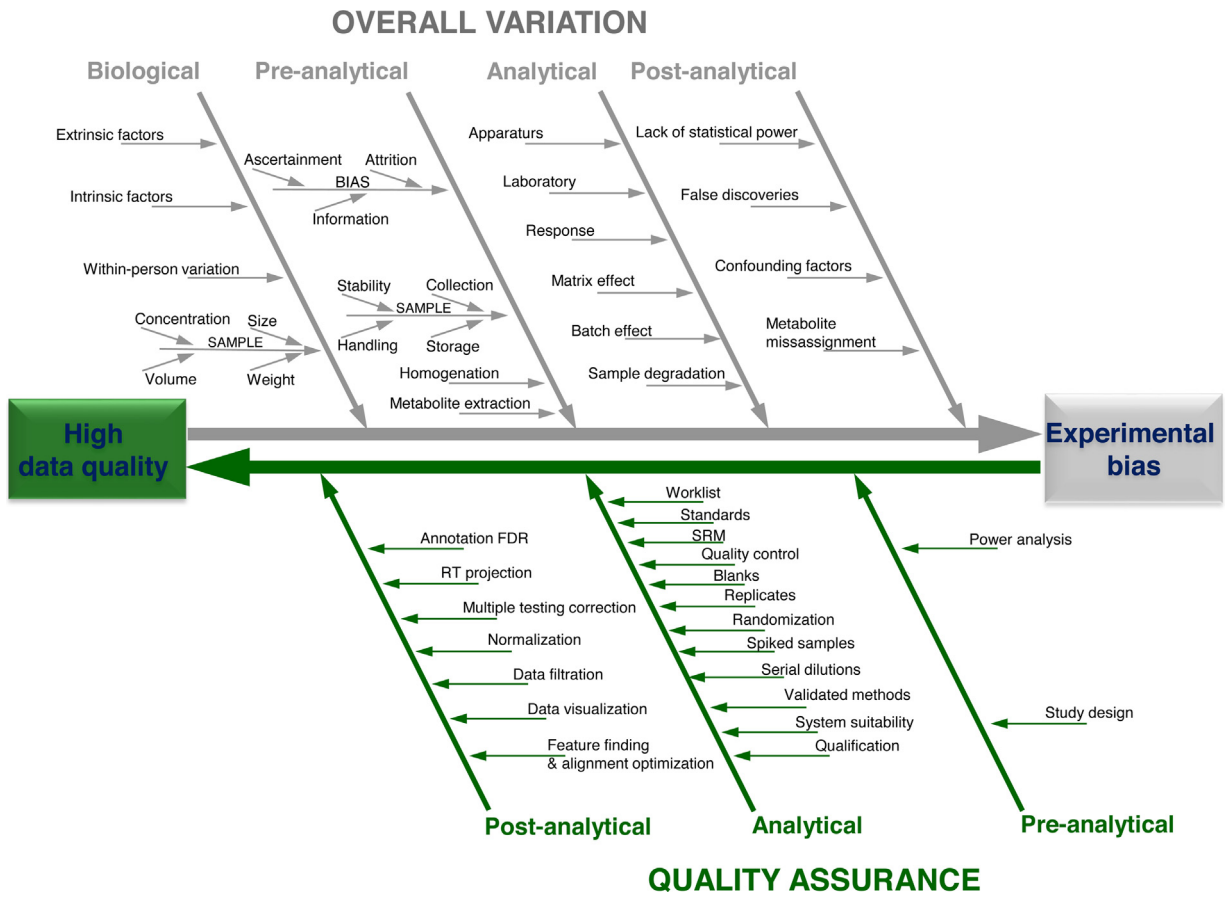

Fig. 1. Cause-effect diagram represents the overall variation and sources of experimental bias and variety of quality assurance procedures that guarantee high data quality of metabolomics study.

Tissues: Tissues are complex structures composed of heterogeneous mixtures of morphologically and functionally distinct cell types. Therefore, collection of representative and homogenous tissue sample requires critical considerations [43]. That point is challenging in general, but even more difficult in case of tumor sample collection due to its special heterogeneity [35]. In animal studies, whole tissue can be used, whereas human tissue sample is mainly harvested during biopsies or surgery. Tissue sampling should be performed as fast as possible and the sample should be immediately frozen to avoid metabolite changes caused by enzymatic reactions. Remaining blood should be removed before storage (e.g., rising the tissue with a buffer), as there is a high risk that the subsequent metabolic profile will be a combination of both blood and tissue [43,53]. Subsequent metabolism quenching to stop all enzymatic activities as quickly as possible, without the leakage of metabolites from the cell, to prevent metabolite degradation and *ex vivo* alteration of sample composition is typically performed by decreasing the temperature of the sample (usually in liquid nitrogen −195.8 °C) or denaturing the enzymes with acid or cold solvent addition [43,54]. When omitted, residual enzymatic activity can significantly change metabolite levels, resulting in poor correlation between the observed metabolome and the true metabolome at the time of sampling [55]. Notably, the healthy tissue samples obtained from post-mortem autopsy like human brain sample are apparently the only available source. Remarkable limitation that should be evaluated is the post-mortem time and induced changes in the brain metabolome. That has been recently studied by Gonzalez-Riano et al. The authors provide evidences based on LC–MS and GC–MS based metabolomics study in order to evaluate post-mortem changes in mice hippocampus [42]. One of the bottlenecks in human metabolomics research is the lack of repositories with a sufficient number of well-annotated frozen samples available for retrospective studies. This limitation could be overcome using formalin-fixed paraffin-embedded (FFPE) tissue sections acquired during diagnostic procedures. Although, FFPE based metabolomics studies have been conducted [56,57] special considerations should be taken regarding the metabolite stability and degree of reproducibility.

Cell and cell cultures. Recommendations concerning experimental design and revision of various methods for LC–MS based cell metabolomics are provided by Hounoum et al. [40]. The authors pointed that cell dependent methods, such as cellular quenching and extraction protocols, have been developed, but no universal methodology for sample pre-treatment has been developed yet. The use of cell lines poses several challenges like (i) variability of growth medium formulation and additives; (ii) differences in cell density which may modify the proliferation rates between cell lines; (iii) cultures of different ages (passage number); (iv) the quenching of cell cultures; and (v) the extraction procedure used [40,58]. For the complete analysis of a cell culture, it is important to measure both extracellular (footprint) and intracellular (fingerprint) metabolic profiles [58].

Cell quenching. Suspension cultures quenching is typically achieved by freezing or spraying into cold aqueous methanol solution (60%) or cold isotonic saline solutions, e.g., ice-cold aqueous NaCl solution [0.9% (w/v)], which better prevents cell membrane damage [40]. Quenching of adherent cells requires additionally the rinsing of cells before with a cold or warm saline buffer, such as phosphate-buffered saline (PBS) or aqueous NaCl solution [0.9% (w/v)]. It is also crucial to remove residual medium in order to

prevent its dissolution in the extraction solvent. That could cause suppression of ionization and impair analytical performance due to the contamination. Interestingly, Chen et al. described that the level of Matrix-Induced Ion Suppression (MIIS) is proportional to the concentration of sample matrix, and the authors proposed that phenomenon to estimate relative metabolite concentrations. It has been suggested that normalizing the metabolite concentration using cellular matrix concentration provided by the MIIS method could improve the data integrity when compared to the cell number approach [59].

When performing metabolomics on cell cultures, it is necessary to ensure cells are in metabolic steady state at the time of seeding and at harvesting. This can be achieved by seeding cells from log-phase cultures and ensuring the length of the experiment is short enough to avoid cells becoming in stationary phase before the time of harvesting and quenching of metabolism.

Cell harvesting. For suspension cultures, a cell pellet can be obtained by centrifugation or filtration [60]. By contrast, adherent cells must be detached from the surface of the culture vessel by one of two methods: trypsinization or scraping [61,62]. Bi et al. compared sample harvesting using trypsin/EDTA treatment and cell scraping in water. It was found that trypsin/EDTA treatment caused substantial metabolite leakage proving it inadequate for metabolomics studies. The results were consistent with the previous findings reported by Winder et al. [63] or with the study provided by Dettmer et al. [62]. The study of Bi et al. indicated that direct scraping cells consisting of an initial saline rinse to remove contaminating media residue, followed by liquid nitrogen quenching is the most suitable harvesting method for Panc-1 cells to obtain highest recovery of a diverse class of metabolites. Similar recommendations are given by Dettmer et al. that observed that direct scraping using methanol/water, is a method of choice to harvest and extract metabolites from adherently growing mammalian SW480 cells [62]. However, combination of all steps by using cold organic solvents that simultaneously quench metabolism, harvest cells and initiate metabolite extraction hamper the determination of cell count or protein concentration, that are often used to normalize resulting data. In all cases, caution should always be taken as differences in the number of cells harvested may result in differences in the cell exometabolome and endometabolome.

Extracellular metabolites. The cell culture medium must be collected rapidly, preferably at low temperature for minimizing the risk of cell damage. To separate culture media and cells before the analysis cold centrifugation or filtration for suspension cultures is performed or simple pipetting and centrifugation in case of adherent cells [40,58].

- **Sample storage and stability**

In biological matrices many analytes are degrading overtime and their stability should be known or investigated under various conditions. Therefore, considerations must be taken to the proper storage conditions in order to maintain sample quality until analysis are completed. Sample pre-aliquoting and storage at −80 °C is considered as the preferred and recommended procedure. To our best knowledge, the impact of long term −80 °C storage on stability of the entire metabolome still needs to be evaluated. Freeze-thaw cycles have to be avoided to minimize the potential for sample degradation as much as possible and in any case the groups to be compared should be treated under the same conditions.

Blood. The effects of blood derived sample collection, treatment, and storage conditions by comparing serum and plasma and by testing the impact of sample incubation time and temperature, sample storage time and temperature, and freeze-thaw cycles have been evaluated in many studies [16,64–66]. As reported by Jobard et al. there are no significant differences in stability of plasma and serum samples stored at least 3 months at −80 °C [30]. Hirayama et al. found that intra-individual changes in plasma and serum samples were significantly smaller than inter-individual changes, indicating that metabolomics changes due to freezing and thawing were smaller than inter-individual variations. Nevertheless, significant differences in serum samples were observed following 10 freeze-thaw cycles including increased concentration of amino acids like phenylalanine, glycine, arginine, asparagine, glutamine and other metabolites such as 5-oxoproline or hypoxanthine, whereas decreased level of 2-hydroxyvaleric acid, cystine and cysteineglutathione disulfide [64]. In other study, Anton et al. investigated the influence of up to four freeze-thaw cycles with assumption that this number may reflect the reality in average biobanking. No significant single metabolite concentration changes were seen although, concentration of some of the amino acids, namely glycine, methionine, phenylalanine, tryptophan and tyrosine slightly increased with rising numbers of freeze-thaw cycles [67]. As reported by Bhatnagar et al. freeze-thaw cycles induce conformational changes leading to aggregation or degradation of proteins [68]. That phenomena could explained increased levels of amino acids which reflects slightly ongoing metabolism. Few changes in serum lipid composition due to three freeze-thaw cycles have been also reported by Zivkovic et al. [69]. Regarding large-scale studies, where samples are collected and stored for long time period it is desirable to look for markers of sample degradation that will ensure the good sample quality before metabolomics study will be conducted, for example glutamate can result increased over large storages as glutamine decreased as described by Lorenzo et al. while comparing plasma samples from patients with bipolar disorders [70]. Trezzi et al. proposed a novel plasma quality marker called LacaScore based on the ratio of ascorbic acid to lactic acid which was found to be sensitive to the plasma pre-centrifugation conditions. However, this method could be only applied for EDTA plasma. The authors stated that in citrate plasma, ascorbic acid is stable over time and the method is also not applicable to heparinized plasma and serum in which the ascorbic acid degradation has shown to be slower than in EDTA plasma [71].

Urine. The stability of urine samples has been evaluated by Saude and Sykes. The samples were stored at room temperature (22 °C), in a refrigerator (4 °C), or in a deep-freeze (at −80 °C) for the 4-week duration of the study. The degree of the changes in metabolite concentrations was evaluated by NMR analysis. The study demonstrates that bacterial contamination of urine significantly alters the metabolic profile of urine over time and proper preparation and storage procedures must be followed to reduce these changes [48]. Urine stability has been also investigated by Gika et al. on LC–MS based study, which proved that urine is stable up to 6 months at −20 °C or below and up to nine freeze thaw cycles were without any apparent effect on the profile [72]. Generally, storage of human urine up to 6 months at −20 °C or below is suitable for metabolomics studies and the number of freeze–thaw cycles (up to 9) does not seem to affect sample integrity [49]. Nevertheless, for long-term stability of urine samples, it is recommended to avoid freeze-thaw cycles and rapidly freeze and store pre-aliquoted samples at −80 °C to minimize potential bacterial contamination and sample degradation.

Biorepositories. Special care must also be taken in case of already existing SOPs for their general applicability, to ensure they are suitable for sample collection for metabolomics approaches. Higher variation could be notable with samples from multicenter collections as usually they follow their own procedures and quality criteria. This leads to a large variation as quality, integrity and performance of individual samples can vary extensively. In consequence that provoke loos of statistical power and study significance. Malm et al. have highlighted the need for increased awareness regarding biobank samples collection, storage and uti-

lization. The authors give a global view of the field where best practice, the standardization aspects and conventional acceptances of biobanking are presented [73]. Up to date several initiatives that disseminate best practice for biobanking have been established such as Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) (www.bbmri-eric.eu), The European, Middle Eastern & African Society for Biopreservation & Biobanking (ESBB) (www.esbb.org) or EuPA Biobank Initiative (www.eupa-biobank.eu) [74]. Biospecimens like FFPE coming from different biobank repositories are also affected by high variations [75].

2.1.2. Metabolite extraction

Despite advances in automation and the use of robotic facilities, a large part of sample preparation still relies on human handling and it is considered one of the most error prone stages. Biological matrices are highly complex in nature, usually having high protein, lipid and salt content. Samples for MS based analysis require preparation that includes protein removal and metabolite extraction and/or derivatization in case of GC–MS analysis.

Plasma and serum. For plasma, serum and other liquid samples, the most common extraction method used is organic solvent-based protein precipitation followed by centrifugation [76]. This method allows the extraction of both hydrophilic and hydrophobic compounds, but the extend of recovery of different metabolite classes depends on the nature of solvent or solvent mixtures as well as solvent/sample ratios that have been evaluated in different studies [76–78]. Based on our own experience, the selected extraction methods should be adequate taking into consideration the metabolic coverage of the selected analytical platform.

Cells and tissues. Additionally, for cells or tissue, sample preparation requires cell wall disruption with freeze-thaw cycles or physical disruption like homogenization methods [43]. Two steps can be distinguished: cell/tissue homogenization and metabolite extraction. Naz et al. developed a comprehensive method for lung tissue metabolite extraction for untargeted multiplatform (LC–MS, GC–MS and CE-MS) fingerprinting. Three different combinations of homogenization solvents were tested in the study with different results: water with 5% formic acid, water:methanol (50:50), water:methanol (50:50) with 5% formic acid. The authors reported that the presence of formic acid in the homogenization solvent could promote the loss of volatile compounds during the drying step as they are in free acidic form. Finally, water:methanol (50:50) was chosen to give optimal performance, as water is necessary for good contact with the tissue and methanol act as a bridge for further extraction step [79]. The solvent mixture methanol:MTBE (methyl-*tert*-butyl ether) in the proportion 80:20 was chosen for metabolite extraction as Matyash et al. already demonstrated that such conditions provide better extraction while studying both polar and nonpolar compounds [80]. The same protocol has been successfully adopted for other tissue untargeted metabolomics studies [42,81].

Other authors described procedures for obtaining the aqueous and organic compounds from animal and human tissues using two consecutive extractions followed by a different untargeted UPLC–MS analysis method for each extract (hydrophilic interaction liquid chromatography (HILIC)) for the aqueous extract and reversed-phase chromatography for the organic one [43,82].

There are several works addressing the issue of method optimization for intracellular metabolites extraction [54,61,63,83–85]. In general, for efficient extraction of intracellular metabolites, the cell wall and membrane need to be permeabilized and in following step the metabolites are extracted by organic solvents, with or without water or other organic chlorinated solvents at different temperatures [40,61,86].

Urine. The advantage of urine is that it requires less complex sample pre-treatment due to the lower protein content. However, highly concentrated urine matrix can also interfere with the analysis and cause a significant GC–MS column and detector overload. To deplete urea in urine samples Fu et al. described simple method based on urease treatment to remove excessive urea from the samples [87].

As a general approach, it should be noted that over the metabolite extraction, metabolite losses may occur due to (i) co-precipitation with proteins; (ii) poor solubility in the extraction solvent or (iii) due to solvent saturation effects, that depend on the total metabolome content of the sample and should be checked before being applied to the study. For that reason different extraction methods have been developed for different biological samples [76,79,82,88,89].

The extraction of metabolites should be as comprehensive as possible while taking into consideration the following aspects: (i) the chemical stability of metabolites; (ii) recoveries of metabolites throughout the extraction and (iii) variability in metabolite concentrations during the extraction [14,76,86,89]. To summarize, sample preparation for metabolomics experiments should be (i) unselective; (ii) rapid with minimum number of steps; (iii) reproducible and (iv) incorporating metabolism quenching [55,89].

Chemical Derivatization

Prior any GC–MS analysis, volatile and thermally stable derivatives from the sample metabolites must be obtained. Silylation, acylation and alkylation are common types of derivatization but the first one is the most extended with biofluids and tissue [90]. The procedure also includes a previous step of oximation for protection of alpha keto groups with methoxamine hydrochloride/pyridine. Compounds with several acid protons can be converted into multiple derivatives with different retention times. The chemical derivatization must be performed under optimized and controlled conditions: reagent concentration, temperature and time in order to minimize randomly modified profiles and multi-peak and multi-origination phenomena. Amount of reagents should be optimized for the type of sample to guaranty total derivatization but not too much for preventing premature column damage along with injector & ionization source contamination. SOPs as well as online derivatization strategies are recommended.

2.2. Analytical

The analytical part refers to the steps following the sample treatment. Within the analytical method validation, different sources of variation can be considered, related to the sample itself, the instrument, and method parameters [91] that will affect the performance and final outcomes of the experiment. The most important sources of variations in the analytical phase include:

- **Apparatus:** Every apparatus used in an analytical chemistry laboratory must be under a quality assurance system and must be calibrated and verified for preventing variations. Even though the manufacturer may supply information about this parameter, routine maintenance should be always expected including: cleaning, verification and validation including daily used systems such as micro-pipettes, balances, heaters, freezers and refrigerators etc. as well as the equipments [92]. The last ones also need to be properly calibrated; (i) within HPLC systems, filters, flow rate, oven temperature should be controlled and calibrated every specific intervals of time; (ii) in GC, septum, injection liners, syringes, carrier gas pressure, and temperature in the oven too [93]; (iii) in CE, injection and rinsing pressure time and voltage. Mass spectrometers, will be tuned and calibrated, within a range of masses, prior to an analysis [24]. Along the experiment, the detector may suffer variations that affect the acquisition and mass accuracy, so normally reference masses are used along the whole analyti-

cal process allowing corrections of signal deviations. Not only the detector, but several pieces of the instruments should be checked.

- **Column/pre-column.** The selection of the column influences in the efficiency and selectivity of the separation (reverse phase columns, HILIC columns, particle and pore size) [94,95]. Characteristics of the column such as pH range tolerance, temperature resistance, pressure and flow rate tolerance, polarity should be taken into account [96]. Zelena et al. performed a comparison of identical columns coming from different batches in order to evaluate inter-batch column reproducibility, interesting approach for long-term studies where several columns may be needed [97]. Compatible pre-columns and online filters are recommended in order to avoid or reduce the risk of column clotting or blocking by particulate matter or macromolecules and to increase the column lifetime. Pressure curves should be saved in any analysis and compared to detect any change during the sequence.
- **Mobile phase (MP).** MP composition and stability in LC and background electrolyte (BGE) in CE could be a potential source of variation [98]. Amount of additives, pH, volatility, degradation and extraction of compounds from the containers should be considered [99]. The concentration of the buffer either in mobile phase and BGE may directly affect elution reproducibility and tailing peaks. Volatile buffers should always be used in mass spectrometry and select those with minimum response decay. Moreover, special containers should be selected with HILIC mechanisms for avoiding interferences from cations as sodium. The moment in which the mobile phase is going to be prepared needs to be pre-stablished (either the whole amount is prepared before the analysis or partially along the experiment), since this can also introduce unwanted variation. Differences in the reagent batches or instrument precision (pipettes, cylinders, volumetric flasks etc.) should be also considered and SOPs should be stablished.
- **System stability.** This term makes reference to how the equipment is performing during the experiment. Variance in system stability can occur due to e.g., (i) appearance of impurities or contaminations, that can manifest as random peaks; (ii) the presence of air bubbles in the system or salt precipitations, reflected in pressure stability or response drift; (iii) insufficient system conditioning or column failure, that will be reflected in RT shift.
- **Temperature.** It should be emphasized that, in whole metabolomics workflow the temperature is one of the most critical parameters that have to be strictly controlled. That point is specially significant in GC–MS analysis as high temperatures will enhance thermal degradation of the analytes and that can occur in several parts of the equipment: (i) injection port (usually 200–250 °C); (ii) column (temperature program 50–300 °C); (iii) mass-spectrometer ion source (> 200 °C); and (iv) transfer line (>200 °C) and these might lead to multi-peaks [90]. The use of SOPs, internal standards (IS) and quality control samples to monitor the process is highly recommended.

In case of analytical process, ambient laboratory temperature should remain constant (mobile phase can be altered with changes in temperature). The same restrictions must be applied to the samples before, during and after treatment. Moreover, sample vials are recommended to be located in a sample tray and the temperature in the controlled autosampler should be optimized and checked for the best stability and for investigation of precipitates on the bottom of the vial. Temperature influences a lot in the separation also in LC and CE, viscosity of the mobile phase or BGE and elution [100]. In relation to the detector, variations in the temperature of the ionization source may promote the formation of different adducts and that should be controlled and optimized for better results.

- **Dirtiness.** Background noise build up, systematic signal drop, pattern of mass peaks (in the case of LC series of mass peaks separated by 44 amu) are some of the indicators that the system becomes dirty [101]. Some sources of contamination are: compounds coming from plasticizers such as PEG, precipitation in the ion source of components from the matrix, in GC the volatility of derivatized compounds can decrease along the experiment forming aggregates in the liner along with septum bleeding or leak [93]. Pereira et al. propose cleaning the source after every 60 samples [102]. Based on our experience, QCs should be checked daily and the source should be cleaning whenever a response decay is detected.
- **Batch effect.** During large-scale experiments (but not exclusively), it is common to detect shifts in RT and variations in signal intensity. A batch is normally the set of samples that comprises a worklist analysed without stopping the equipment or without making big changes (MPs addition, source cleaning, recalibration, pre-column change or replacement). Variation between batches are very problematic in what data processing is concerned so the best experimental design is required in order to avoid later data treatment [27,103].
- **Matrix effect.** This phenomenon happens when there are molecules that co-elute altering the signal of the compound of interest [104,105]. Among the problems that this can cause, there are variations in ion suppression [106,107] and the corresponding abundances, low analyte detectability, inaccurate quantifications, changes in the total useful signal leading to decrease in the ruggedness of the methods and lower values for precision, accuracy and sensitivity [108]. Possible reasons for this effect to appear are MP and/or buffer additives, contaminations coming from materials or samples, inadequate flow rate, solvents/standards purity, dirt on source etc. Matrix effect can significantly affect analytical parameters like accuracy and precision, linearity and limits of detection and quantification [102]. Hence, it should be carefully investigated during method development and validation process either in LC–MS and GC–MS [109].
- **Carryover**. Mitulović et al. defined carryover as the process "when an analyte originating from a previously injected sample appears after the injection of buffer as a blank injection" [110]. From a chromatographic point of view there are two major sources of carryover, originated in the autosampler or in the column [111]. Accuracy and precision in LC–MS system can be seriously affected by this phenomenon therefore they should be investigated, reduced or completely eliminated during method development and validation. Nevertheless, it is advisable that the system is monitored routinely for carryover effect [111,112].

2.3. Post-analytical

Metabolomics experiments lead to highly complex data sets involving hundreds or even thousands of metabolic features or possible compounds. Considering this aspect, proper data handling is a fundamental step that has enormous impact on quality of metabolomics results. Raw data are typically pre-processed to provide structured data in an appropriate format for further data analysis. The main purpose of data pre-processing is to improve signal quality and reduce possible biases present in the raw data by (i) filtering background noise; (ii) peak peaking and deconvolution; (iii) aligning spectra across multiple samples; (iv) missing values imputation; and (v) data filtration [113,114].

This review is focused mostly on the pre-analytical and analytical aspects of metabolomics experiment, with the general idea that the quality that we put in these stages will result in the high quality of the gained data matrix. However, that aspect of post-analytical unwanted variation cannot be omitted completely, therefore, it will be shortly commented.

- **Data pre-processing challenges.** Metabolomics data analysis if one of the most time consuming steps in all metabolomics

workflow. It is generally expected to spend some days dedicated to sample analysis and then several weeks or months in data mining. That time could be even extended when high bias and unwanted variation have been introduced in the experiment. Post-analytical data pre-processing could also be prone to error. Recently, Cheng et al. discussed current challenges that are still facing data processing, treatment and results reporting [115]. The authors underline that data pre-processing and robust data analyses are more challenging for untargeted in comparison to targeted metabolomics. It has been recognized that existing software packages still have several limitations that must be considered. Special caution should be taken as one metabolite can be related to multiple features derived from different sources of noise, instead of truly biological background, resulting in a high number of 'false positive' identifications [116]. Many of these noise peaks can be attributed to (i) MS artefacts (e.g., adducts, dimmers, fragments, isotopes, clusters, charge states) and (ii) HPLC or peak-picking artefacts (e.g., split peaks, peak shoulders, baseline noise, and ghost peaks) and are often incorrectly identified as metabolites [117–120]. Artefacts called as 'ghost' peaks are characterized by almost the same mass and eluting in the same RT. Those peaks probably originate from the same molecule, but in such cases one mass is detected in half of the samples and the second similar mass in the other group of samples. The other major source of noise consists of background contamination, however, it can be filtered from the data by comparison with blank solvent samples, based on peak intensity and reproducibility. As discussed by Castillo et al. experimental settings with which data were recorded and the choice of the algorithms used for pre-processing may dramatically affect final results [114]. The properties and the quality of the data pre-processing algorithms is one of the most important factors that should considered, and that have been already evaluated in many studies [119,121–123]. Hence, the parameters chosen for data acquisition, data alignment (mass error, RT window) should be carefully evaluated based on the raw data and the specification of the analytical platform applied. Notably, in case of CE-MS, migration times and peak areas are not as reproducible as in GC–MS or LC–MS and peak alignment may fail [124]. Therefore, the outcome data matrix needs to be properly evaluated with respect to the possible alignment error. It should be remembered that a large number of aligned features may indicate failed peak alignment, due to erroneous settings of RT, mass window or abundance threshold. Such cases have to be carefully examined making the data pretreatment challenging and time consuming. Hence, it is advisable that substantial time should be first dedicated to verify quality of raw data before further data handling procedures.

3. Tools for quality control assessment

Herein, we provide examples and a short description of possible analytical strategies that are recommended in order to control experimental performance in metabolomics studies. A summary of the procedures is also presented in Table 1.

3.1. Pre-analytical

3.1.1. Study design

As it has been stated, any good scientific study starts with a rigorous experimental design [125,126] as poorly designed studies may give misleading results. According to this, a good experimental design means that the individuals are representative of the biological question to be answered and the samples to are compared in an unbiased way. Before stablishing the experimental design, information about the biological question to be answered, along with the type of samples and individuals will focus the way of obtaining a representative and homogeneous sample and select the appropriated techniques, data treatment and statistical analysis. Numerous metabolomics based studies have been designed to evaluate the impact of factors which may influence population homogeneity, especially in respect to (i) intrinsic factors: (e.g., age, [127–129], gender [128,130–132], genotype [133–135], health status [115,136,137] body composition [138,139], diurnal cycle [140,141] circadian rhythm [142–144], Body Mass Index [128,145,146], tissue turnover); (ii) extrinsic factors: diet [147–149], drugs [150], stress [151,152], physical activity [153,154], microbiota [155,156], smoking [157,158], alcohol intake [159], environmental factors, etc.; (iii) within-person variation; (iv) others: ethnicity [160,161], collection year or varying level of biofluid concentrations, number of cells per sample, sample measurements (e.g., volume, weight, amount of proteins). Therefore, collection of the data related to those factors can pave the way to identify confounding variables. All possible sources of variation should be considered and discussed in advance and the experiment should be designed in order to prevent them. Würtz et al. used Mendelian randomization (MR) to estimate causal effects of BMI on 82 metabolic measures in 12.664 adolescents and young adults from four population-based cohorts in Finland [162]. Briefly, Mendelian randomization is being extensively applied to estimate the long-term causal effects of various exposures on clinical and epidemiological outcomes using observational data [163]. This method employs phenotypic and genetic variants to remove bias due to confounding factors. The same approach has been recently applied by Borges et al. to define the metabolic profile associated with higher blood adiponectin concentration where *cis*-acting genetic variants in the vicinity of the adiponectin gene were used as instrumental variables [164]. The authors highlight the potential of MR in high-throughput metabolomics studies that could yield important insights to advance our understanding in the pathophysiology of common complex diseases. Well matched samples and appropriate methods applied for sample collection may substantially reduce the sources of bias that could affect sample processing, and further metabolome analysis.

Sample size and Power analysis

The choice of an appropriate sample size for a study is a crucial aspect of experimental design. It allows to have a high probability of detecting a difference of a given size if it exists but also to have a low probability of finding a significance when no real difference exists [28,165]. Studies in a laboratory environment with animal or cell culture models are well controlled and the impact of unwanted variation can be minimized or completely eliminated. Therefore, low sample size (6–15) can be used and that is sufficient to identify changes that are statistically robust [101,166]. However, caution should always be taken in complex studies applied to the general human population where observed variation is much higher.

Statistical power analysis should be performed before the beginning of the study. In order to obtain statistically validated data, it is mandatory to ensure that sufficient number of samples are acquired. When the number of variables greatly exceeds the number of samples it is always easy to find random multivariate correlations. Thereby, with an increased sample size, the standard error of an estimate will be reduced, leading to increased precision and study power [167]. Different approaches have been developed for sample size determination in metabolomics assays. Nyamundanda et al. proposed the MetSizeR method for determining sample size for metabolomics whilst controlling the false discovery rate (FDR). This approach employs permutation techniques to estimate sample size and accounts for correlation between metabolites and the effect size variability. The advantage of the developed method is its ability to determine sample size even when experimental pilot data are not available [168]. Billoir et al. released automated

Table 1
Implementation of tools and general considerations for QA procedures in untargeted metabolomics.

QA procedure	Controlled factor	General considerations[a]
Pre-analytical tools		
Study design	Sample handling	• It is essential to develop validated SOPs to ensure that samples across the study are handled and pre-processed in consistent manner.
	Sample collection	• Prospective *vs.* retrospective study.
	Sample pre-processing	• Population homogeneity: (i) intrinsic factors: (e.g., age, gender, genotype, health status, body composition, diurnal cycle, circadian rhythm resting metabolic rate, tissue turnover); (ii) extrinsic factors: diet, drugs, stress, physical activity, microbiota, smoking, alcohol intake, environmental factors, etc.); (iii) within-person variation; (iv) varying level of biofluids concentrations, cell size, sample measurements (e.g., volume, weight).
	Unwanted biological and analytical variation	• **Mendelian randomization** – determination of causal relationship between exposure and outcome of the interest
	Sample stability	• Provide representative samples and appropriate control group. Avoid poor quality samples e.g., hemolytic samples.
	Sample integrity	• Collect homogenous samples (especially for tissue samples).
		• Consider addition of additives (e.g., anticoagulants) or preservatives (e.g., sodium azide).
		• Stop metabolic turnover by inhibition of enzymatic activity (metabolism quenching) – LN2 or ice cold organic solvent.
		• Process samples under a controlled temperature (4 °C).
		• Minimize time for sample processing.
		• Prepare sample sub-aliquots.
		• Consider type of storage tubes.
		• Label samples in a clear way.
		• Storage condition (recommended temperature −80 °C, before and after extraction; be aware of light exposure and oxidation).
		• Avoid freeze-thaw cycles.
Power analysis	Sample size	• Minimum required sample size should be predetermined.
	Statistical power	• The more inherent variation controlled, the fewer sample is required to the study.
		• Power of the statistical test increases with increasing sample size. Large sample size has a greater ability than a small sample to detect an important effect if it exists.
		• The lower variability of the observations the higher statistical power detected.
		• The standard deviation for power calculation could be estimated from a pilot study or, if not available, from other published studies on the same subject.
		Adjust the sample size by recruiting more patients into the study at the outset (drop-out rate).
Analytical tools		
Qualification	System performance	• **Design Q** – setting the functional and performance specifications; user requirement specifications, operational specifications, vendor qualification.
		• **Installation Q** – performing and documenting installation; check as purchased; check proper installations of hardware and software.
		• **Operational Q** – testing the equipment to ensure specifications; performance testing.
		• **Performance Q** – testing the system performance according to the specifications and selected application; periodic calibration; preventive system maintenance
		• **Analytical instrument qualification:**
		• Level of qualification depends of the type of instrument and on application.
		• Group A – standard equipment with no measurement capability e.g., vortex mixers and centrifuges.
		• Group B – standard equipment and instruments providing measure values, e.g., balance, pH meters, oven, pipettes. This group also includes equipment controlling physical parameters, such temperature, pressure or flow.
		• Group C – instruments and computerized analytical systems, e.g., HPLC systems and mass spectrometers.

Table 1 (*Continued*)

QA procedure	Controlled factor	General considerations[a]
System suitability	System performance	• Verify that the system will perform in accordance with specified criterial for the chosen analytical procedure.
	Results quality	• Should be performed at the beginning of each analytical batch including tunning and mass calibration, together with chromatographic column and MS source cleaning. In case of GC–MS injector liner, septum or syringe cleaning or if necessary column cutting should be applied. In case of CE-MS system, that refers to the capillary cutting and ion source cleaning.
Analytical method	Linearity	• Linear response between signal and concentration within the broadest possible range of compound classes.
	Selectivity	• Prove the resolution with isobaric compounds.
	Sensitivity	• Calculation of the LOQ (accurately and precisely) when possible, if not, calculation of the LOD.
	Limit of detection	• Several measurements of the standards with a concentration in the calibration curve (inter and intraday precision).
	Limit of quantification	• Introduce intentionally variations to check if the method remains unaffected.
	Precision	
Serial dilutions	Proportionality	• For each analyte of interest, use a dilution factor to generate the calibration curve.
Spiked samples	Matrix effect	• Spike diluted QCs to check ion suppression, matrix effect and selectivity loss.
	Selectivity	
Randomization	Unwanted variation	• **Level of sample collection** – in order to obtain representative and homogenous samples; include participant enrollment, experimental group assignment.
	Unwanted correlations	• **Level of sample preparation and analysis** – to ensure that the analysis of the resulting data are not biased by sample preparation or analysis order.
		• **Constrained randomization** – dependent or matched samples. Applied to ensure that the bias related to instrumental drift will not be confounded with the biological variation of interest.
Replicates	Unwanted variation	• **Biological:** for quantities that vary due to external factors (e.g., blood glucose level which vary with food intake). It may be necessary to take several measurements over a period of time to get an accurate result.
		• **Analytical (technical):** for quantities that are difficult to measure accurately.
		Replicate sample – to acquire data regarding sample to sample processing consistency.
Blank samples	Signal-to-noise ratio	• Matrices that have no measurable amount of the analyte of interest.
	Contaminations	• **Collection blank** – sample that has been collected from the same source as the experimental sample.
	Interferences	• **Extraction blank** – prepared following exactly the same procedure as experimental samples.
	Artifact feature	• **Mobile phase blank** – consist only the mobile phase.
	Carry-over	• **Cleaning blank** – sample of strong solvent e.g., isopropanol introduced in order to clean the system.
	Matrix-effect	• Can be helpful to identify the artifacts.
		• Consider features with the main values in blank higher that 10% as non-relevant.
		• Essential for any QA procedure.
Quality Control	Efficiency of the sample preparation and metabolite extraction	• QCs should be representative of the qualitative and quantitative composition of the study samples.
	System stability and system performance	• Different type of QC sample (pooled, extraction, surrogate, commercially available), that should be considered before the study.
	Reliability of the system	• QC allows to correct signal drift within batches and between analytical batches.
	Reproducibility	• Should be used for%RSD calculation for both retention time (RSD < 1% as general rule) and metabolic features (LC–MS and CE-MS < 20%; GC–MS < 30%) that allow for estimation of the overall precision.
	Variability	• The results obtained from the QC samples should be monitored by plotting them on a control chart, check for systematic (trend, shift) and random error.
	Data quality	QC gives the highest level of QA.
	Method development	
	Data normalization	
Standards	Unwanted variation	• Known metabolites added to each biological sample before extraction.
		• IS cannot remove unwanted biological variation
		Variation captured by IS depends on its chemical properties and could include other source of variation e.g., arising from chromatographic system or ion suppression.

Table 1 (*Continued*)

QA procedure	Controlled factor	General considerations[a]
Worklist	System variability	• The length of the analytical run should be predetermined and based on the number and nature of samples as well as on analytical system.
	Unwanted correlations	• The samples should be randomized.
	Signal drift	• The mixture of standards (standard mix), blanks and QC samples should be included.
		• Large studies should be run in different analytical batches.
Post-analytical tools		
Data Visualization	Data variability	• A quick graphical summary of data.
	Signal changes	• Help to understand variation, evaluate analytical performance, trends, shift or outlying observations.
	Systematic or random error	• **Raw chromatogram** (EIC, BPC, ECC), pressure curve or current flow.
	Unwanted correlations	• **Control chart (Shewhart control chart)** for understanding process variability. Allow to overview of the analytical precision based on plotting the sum of metabolic feature intensities for every experimental and QC sample against run order.
	Potential outliers	• **PCA** – non-supervised multivariate method that allow for evaluating signal drift, sensitivity loss, variation in QC samples. The score plot can be used to identify the differences or similarities among the samples and identification of the outliers. The loadings plot can be used to identify the signals responsible for the grouping or separation among the samples.
Annotation	Metabolite identification	• **False discovery rate (FDR).** Developed to ensure annotation quality. Allows for reliable metabolite identification and minimize misinterpretation of metabolomics data.
	Quality of the results	• **Retention projection.** Strategy developed to calculate LC gradient retention across laboratories in order to enable to provide complete information for compound identification.
Normalization	Unwanted variation	• Method used to identify and remove sources of systematic variation between sample profiles due to the factors that are irrelevant with regard to biological processes.
	Sample preparation	• **Method-driven:**
	Metabolite extraction	• **Single internal standard (SIS).** The method assumes that every metabolite in a sample is subject to the same amount of unwanted variation. That is often found to be inadequate in removing unwanted variation.
	Signal drift	• **Multiple internal standards (AIS).** The AIS average is used for normalization.
	Within and between batch variation	• **Locally weighted scatter plot smoothing (LOESS).** Based on QC samples. It assumed that proportion of metabolic changes across biological samples is relatively small or that there is similar number of metabolites with increased and decreased signals across the peas intensity range. Can robustly estimate the unwanted batch variation but also possess the risk to spuriously missing metabolites.
		• **Data-driven:**
		• **Scaling factors** e.g., normalization by median or intensities; not applicable when the self-averaging property does not hold.
		• **Total intensity normalization.** Forces all samples to have equal total intensity; assumes that the total concentration of all metabolites in the sample does not vary across samples in a data set. The changes in the peak intensities of a few high concentration metabolites can notably compromise the normalization performance because of their substantial contribution to the total peak intensity.
		• **Median fold change (FC).** Adjust the median of log FC of peak intensities between samples to be approximately zero. Assumes that the metabolite peaks affected purely by dilution will exhibit the same FC between two sample profiles. The performance is robust against at least 50% peaks intensities exhibiting asymmetrical increase or decrease in response to biological factors.
		• **Quantile normalization.** Enforces all samples to have identical peak intensity distribution. Assumes that the distribution of peak intensity across a data set is nearly the same for all samples and that can be problematic with high intensity values. It assumed that there is a similar number of metabolites with increased and decreased signals across the peak intensity range.
		• **Probabilistic quotient normalization (PQN).** Transforms the metabolomics spectra according to an overall estimation on the most probable dilution.
		• **Variance-stabilizing transformation.** Assumes that the noise is multiplicative in nature and the standard deviation of peak intensity is proportional to its expected value.

[a] References provided in the text.

implementation of the Data-driven Sample size Determination (DSD) algorithm for MATLAB that allows the determination of an optimized sample size in metabolic phenotyping studies [169]. Recently, also Blaise et al. described power analysis and sample size determination in metabolic phenotyping. For effective power calculations, the authors recommend that pilot studies should be conducted to obtain good estimates of variable distributions and covariance structure [170]. A pilot study of 20 samples has been suggested as sufficient to perform robust power analysis. In case when such data are not available, it is possible to search for it through the scientific literature.

3.2. Analytical

3.2.1. Qualification

The recommendations for instrument qualification comes from the USP guidelines for analytical chemistry [171]. However, that concept could be also adapted for QA in metabolomics laboratories. The objective of any analytical measurement is to get consistent, reliable and accurate data. Therefore, proper performance of analytical instruments and computer systems is hugely important in achieving this goal. Establishment of appropriate qualification protocols requires assessment of many factors that are summarized in Table 1, whereas Fig. 3 illustrates the different components of metabolomics data quality.

3.2.2. System suitability

System suitability is fundamental part of the analytical procedure, because it verifies the adequate working of the instrument used to obtain the measurements. Bose et al. published a mini review in 2014 covering different parameters for system suitability [172], among them: peak symmetry, tailing factor, capacity factor, selectivity, resolution and theoretical plate count. The test should include a priori defined pass/fail criteria. These tests are usually performed using a mixture of chemically different standards which are individually evaluated for all the parameters to guarantee that the system's performance follows the initial expectations [173]. This analysis may also enable post-analytical evaluations [14]. A similar concept was introduced in 2006 by Sangster and co-workers who suggested the injection of QC samples along the analysis in order to evaluate the systems performance and detect possible analytical variations [174]. Even though the test mixtures normally contain low number of metabolites around 10 compounds is System Suitability Tests (SST) should be performed before and during the experiments avoiding the assumption that the behavior of the equipment will remain constant along the analysis. The main reason for analysing SST at different time intervals is based on "the concept that the equipment, electronics, analytical operations, and samples to be analysed constitute an integral system that can be evaluated as a whole" [175]. The Guidance for Bioanalytical Method Validation in the Industry could perfectly be applied [18]. It says that "if system suitability is assessed, a specific SOP should be used. Apparatus conditioning and instrument performance should be determined using spiked samples independent of the study calibrators, QCs, or study samples. Data should be maintained with the study records " [26].

Mass accuracy

Prior to any analysis with exact mass analyzers and when a detector tune is required, the equipment needs to be calibrated and reference masses are used. With exact mass instruments, at least two reference masses are used, one in the upper and one in the lower range of m/z and different pair for positive than for negative ionization. Every time the equipment stops, the source is cleaned or a new worklist starts, it should be recalibrated in order to make sure that all the masses along the analysis are equally calibrated both in intensity and in accurate mass.

3.2.3. Analytical method

For a full analytical measurement validation, several parameters need to be addressed, taking always into account that a method is validated for an intended use or application [92]. For all the parameters to be evaluated, a selection of standards should be used. The ideal scenario for an untargeted method validation would be to have metabolites spanning all the retention times along the techniques used, as well as enough standards to cover all the different biochemical classes, functional groups, polarities, molecular weights and ionization modes that could contain the sample. By doing this a method would be validated for nearly the whole metabolome ensuring the detection of the fingerprint [176].

Up to our knowledge, there has not been published any guideline or protocol about validation of untargeted metabolomics approaches. Nevertheless scientists are performing this type of studies adapting similar guidelines [18,19,91,177] and procedures published by different metabolomics groups [79,176,178,179]. Although movement is being seen around this field, e.g., FDA took steps forward on its edition published in 2013 with a section named "biomarkers" for targeted approaches, little is being done as far as methodology is concerned [18]. Under that title, FDA only describes what a biomarker is and what it serves for.

An analytical method validation for untargeted metabolomics should prove at least linearity and precision. The aim of untargeted metabolomics is the differential analysis of the broader possible set of compounds among two or more groups; therefore, recoveries could not be quantitative for all of them, however, as long as they are reproducible and the signal changes with the concentration of the analyte in the sample, the method can be used for the comparison.

- **Linearity:** refers to the relationship that exists between the instruments response and the real concentration of the analyte in the sample. In untargeted metabolomics studies it is unrealistic to talk about calibration curves of expected compounds, it would be commercially unattainable, difficult to handle and in most of cases there might not even exist, not to mention isotopically labelled standards. The calibration curve is obtained by plotting abundance versus added concentrations. If an IS is used, the ratio of the area from the analyte and the one from the IS versus the added concentration will be represented [61]. Guy et al. evaluated the usage of an IS for calculating linearity and observed that correlation coefficients obtained from the ratio of the analyte's area and the IS were worse than using the analyte area itself. In this scenario, where ISs are used, it might be preferable to use an endogenous isotopically labelled analyte which will also provide information on a possible matrix effect [180]. Venter et al. proposed using ten standard compounds, with increasing concentrations, to check if there was a linear response. They chose standards belonging to various compound classes which could be expected in their samples. For the concentration range, they used previously reported values from other groups with similar samples [181].

In cases where there is no literature related to the test sample, a pool of sample spiked with a standard can be used in order to estimate, in relation to the added amount, the concentration of analytes in the sample [176,182]. Once the average value of a metabolite is estimated, one approach could be designing the calibration curve from the 25 to the 200% of that mean value found in the preliminary analysis [176].

Methodology for preparing these samples can vary. One strategy may be diluting the samples and spiking standards after that,

with this, the ionic suppression decreases, as well as the matrix effect. However, this seems to be the best way for applying to all the compounds. Other options could be preparing the mixture of compounds in the appropriate solvent or diluting them in the mobile phase. Even though with this procedure calibration curves tend to a correlation coefficient close to one, the matrix effect is not observed. Requirements to validate linearity are very strict for quantification [18,19]; however, in untargeted metabolomics, the aim is estimating a fold change when comparing two or more groups and therefore, the validation should just prove that the signal increases proportional to the concentration of the analyte [177].

- **Selectivity:** A method must be able to distinguish between the different compounds present in the sample. This means, it should be able to resolve overlapping isobaric compounds. The way of assessing chromatographic resolution and the presence of interfering compounds could be the one Sandra et al. used and which was followed afterwards by Whiley et al. Two isobaric compounds were selected, with similar chemical structure, (e.g., PE18:1/18:1 and PC18:1/18:1) in order to prove the resolution of the method [183,184]. Differences in lipid concentrations as small as 25% were discriminated with a *P* value <0.01 and could even notice variances in the concentrations of 10%. Moreover, Whiley et al. following a similar protocol using PE(36:2) and PC(P-34:1), that differ in mass by 0.0364 Da, were able to separate them [183]. Validating selectivity is very uncommon in untargeted metabolomics and it could be very important, mainly when lipids are involved where different lipids co-elute and have very close or even equal accurate masses [185]. Another way of improving the selectivity of a method, as Siskos et al. suggests can be combining different analytical tools [185]. Whereas, Kuligowski et al. measured selectivity through the analysis of solvent blanks, non-spiked and spiked samples [182].
- **Sensitivity**: In bioanalytical assays the sensitivity of the method is considered to be the lowest concentration of a metabolite, measured accurately and precisely, in the same sample matrix allowing the detection with slight variations in metabolites concentrations [18]. This parameter is normally related to the limit of quantification (LOQ), which is calculated using the lowest level of the calibration curve measured with reliability. On the other hand, the LOQ is related to the signal to noise ratio (that normally is estimated being ten times higher than the signal to noise value) [181]. If the LOQ is outside the levels of the curve, means the analyte cannot be quantified with enough accuracy and precision so the limit of detection (LOD) is given. In untargeted metabolomics LOQ will be estimated only for a group of representative compounds, however it is interesting to explain why some compounds in a pathway could not be detected.
- **Precision**: evaluates the closeness of repeated measures of the same sample/analyte. This is a key parameter in untargeted metabolomics, when accuracy cannot be granted [97,181].

For single analyte experiments, FDA recommends that the signal detected in two-thirds of the QCs should be within the 15% of the QC mean though when the concentration is at or near the LOQ, they increase the tolerance limit towards 20%. For untargeted metabolomics where methods aim to detect the maximum number of compounds, Zelena et al. suggest an acceptance tolerance of 20% [97]. Normally expressed as CV, precision should be reported from one-day-run (intra-day precision known also as reproducibility or intermediate precision) and from different days-runs (inter-day precision) [182,186]. If the samples are prepared another day, by a second operator and with fresh reagents, then repeatability can be evaluated (precision of the method with time). These parameters are very important specially in large-scale studies where samples could be potentially analysed after several days of its preparation. Changes in metabolites with time should be known and controlled. Some authors suggest incorporating the evaluation of the inter-day precision as a routine procedure when dealing with analysis of different duration, this means, consider the length of each particular study and evaluate precision for that period of time [181,185].

- **Accuracy**: refers to the closeness of the experimental value with the one considered as "true value or accepted reference value" [91]. Accuracy and precision must remain constant even when the sample is diluted. Usually accuracy is calculated by the recovery of known amounts of a set of representative analytes. In any case, as previously commented, the recovery could not be close to 100% and even so the method could be acceptable for metabolomics.

3.2.4. Serial dilutions

Serial dilutions are prepared from the pool of samples with a dilution factor that is maintained. It is a useful tool to detect matrix effect since dilutions have a direct influence on ion suppression. Moreover, it gives information about the possible linear trend that a statistically significant feature may follow but this linear tendency of the QCs may not be reflection of what happens in the sample. Indeed, if an analyte appears to be statistically significant means it is possible to detect the increase/decrease of signal due to different concentrations [11]. Eliasson et al. used this tool to optimize data-processing parameter by a correlation between dilution series and integrated peaks. The authors suggested to utilize the concentration vector in the assessment of peak quality (the reliability index). All peaks that show a good linear correlation with the concentration vector are classified as reliable and could be assigned as potential biomarkers [187]. This series of diluted pool of samples enabled them also to assess linearity and to detect matrix effect [181,188].

3.2.5. Spiked samples

Method validations can be performed either in real samples or in solvents, being the first option the best approach since different behaviors can be observed with it and among them, matrix effects, ionic suppression, selectivity loss, etc. When talking about real samples, we refer to pooled QC samples (containing all the metabolites characteristic of the study). Into them, either individual or mixture of standards are added to check the behaviour of the analytes within the matrix, describe the profiles and even quantify possible compounds of interest within samples (in targeted studies) [102]. In the case of untargeted metabolomics, the selection of compounds should be done considering those with similar chemical properties to the metabolites expected in the samples. Solubility of these in the sample and compatibility with the system should be taken into account.

3.2.6. Randomization

Dunn et al. strongly recommended randomization strategy both for sample preparation and sample analysis order to ensure that no bias is introduced at this stage of the study [189]. González-Peña et al. applied sample randomization not only for metabolite extraction and corresponding analytical run, but also for homogenization, that offers good strategy for reducing the bias and to control unwanted variance [41]. As discussed by Want et al. randomization of a small sample set is a simple procedure, however, in case of cohort studies with large sample size it could possess several challenges [43]. An example of already existing approach for sample randomization is to use experimental design taking a priori information e.g., to obtain a balanced distribution between controls and cases in analytical batches [190].

Randomization in the analysis order. Samples must be randomized or appropriate block design created prior to analysis in order to ensure that there is no correlation between biological fac-

tors and analysis order. In case of samples prepared in different batches, the acquisition order should be randomized accordingly [189,191]. To enhance the overall quality of the data Jonsson et al. suggested constrained randomization procedure for mass spectrometry based metabolomics studies including dependent or matched samples e.g., controls or cases or the same subject before and after an intervention [190]. In this way systematic bias is reduced and the risk for confounding the biological variation of interest between the matched samples with the instrumental drift is minimized, that is rather difficult by a traditional randomization approach [190].

3.2.7. Replicates

Replicates generally yield more accurate and reliable summary statistics in experimental work and can be used to assess and isolate sources of variation in measurements and limit the effect of spurious variation on hypothesis testing and parameter estimation[192]. Such replicates include:

Biological replicates – refer to parallel measurements of biologically distinct samples that capture random biological variation, which may itself be a subject of study or a noise source [43,192]. To obtain meaningful results, multiple biological replicates must be analysed and subjected to appropriate statistical tests to ascertain the significant changes in metabolite concentrations between groups [193]. The typical number of replicates range from a minimum six for animal based studies where due to the controlled laboratory condition low variance is expected, to larger cohorts of human population, which is usually subjected to higher variation [101]. Limited number of biological replicates with high variance can yield poor statistical outcome and study failure.

Technical replicate – repeated injections/measurements e.g., triplicate of the same biological sample performed to determine method reproducibility. Technical replicates could be introduced when measurements are difficult to perform. They could also be used to confirm whether an outlier sample is actually biologically different, or rather is related to system variability [43,116,189]. During data pre-processing the replicate measurements are summarized to average values, by determining the mean or median, to reduce the influence of noise in downstream data analysis. Although, this approach is very effective, it is also time consuming, that is great drawback in aspect of a high-throughput analysis. Recently, Bader et al. focused on the evaluation of technical replicates on the LC–MS profiles [194]. The authors strongly recommend to use at least two, or three technical replicates of each sample to enhance data quality. It was found that the analysis of multiple replicates provides the option of introducing stringent criteria for data filtering and allowed to minimize the false positive rate and to correct partially false negative findings occurring during the peak recognition [194]. In the other hand, Want et al. postulate that when using well validated method and QC samples there is no need for replicate analysis, as the QCs should provide clear evidence of analysis performance [43]. The exact number of replicates in metabolomics study depends on the variability between samples, variance between observed groups and the power of the test being performed [58].

Replicate sample – sample prepared from the same biological specimen, to acquire data regarding sample to sample processing consistency [195].

Ĉuperlović-Culf et al. discussed the need to use both biological and technical replicates in case of the cell culture applications for metabolomics studies. Cultures grown independently (in separate flasks) and subsequently treated with the same compound would represent biological replicates. Technical replicates would be different aliquots of cells harvested from the same culture flask after treatment or even multiple measurements of the same extracts. It is suggested to use at least three technical replicates to ensure accuracy of measurements to increase statistical accuracy of the results [58].

3.2.8. Blanks

The use of blanks is essential for any Quality Assurance (QA) procedure however this strategy sometimes seems to be overlooked. Blank samples are the matrices that have no measurable amount of the analyte of interest. All the conditions related to the blank preparation like sample pre-treatment, extraction, storage and analysis have to be the same as for experimental samples. The goal of including blank is to detect and minimize the carryover during method development (wash step). They also serve for identifying and removing artifacts from the data matrix, in order to identify features arising from contaminants and other sample components originating from e.g. tubes, solvent impurities, sample additives or preservatives [195–197]. At least one blank should be included at the beginning and the end of the batch. From the perspective of untargeted analysis it is not recommended to analyse blank samples during the analytical batch. The changes in the matrix composition may result in the partial deconditioning of the system (RT shift, changes in peak widths and chromatographic resolution) [191]. It is possible to use different types of blanks in metabolomics experiment.

Extraction blank or process blanks, that follow exactly the same procedures as experimental samples. Extraction blanks do not contain the biological specimen and are an important part of any analytical method development and validation process [195]. Such strategy has been applied for example by Kuligowski et al. for the LC-TOF-MS plasma analysis. From 4294, 4534 and 3815 features detected in 3 batches, 2600, 2745 and 2582 variables, respectively, left after blank subtraction. Features with the main values in blank higher that 10% of the mean value in samples were considered as non-relevant [196]. Blank filtering was also applied by Kirwan et al. and the peak was considered a contaminant if it had an intensity in the blank sample one-third or more of the intensity of the same peak in the biological samples. In that way it was allowed for high intensity peaks of biological origin to be retained [116]. Yao et al. used extraction blanks to quantify palmitate contaminations from plasticware [198]. It was found that plastic contamination signal was highly variable between consumables, even in the same lot. In such case it is challenging to remove accurately contamination by subtracting the background measured from a blank [198]. Extraction blank is prerequisite for GC–MS analysis to determine the ghost peaks originated from the derivatization process (reactants, presence of volatiles from the solvent, environment, etc.,) or the GC system e.g., column bleeding, contamination in the carrier gas or the inlet module [34].

Collection blank, means the sample that has been collected from the same source as the experimental sample, but are free of the biospecimen. Unfortunately, in spite of its undoubted benefits, such blank is not always available.

Mobile phase blank (system blank), consist in using only the mobile phase used for analytical run to identify impurities present in the solvent, cleanliness of glassware and equipment. Bader et al. as an example, used system blank as zero-injection and eliminated all features whose peak area and peak height did not substantially differ from the blank's (blank correction) to increase quality of untargeted LC–MS based screening [194].

Cleaning blank, is prepared with strong solvent e.g., isopropanol introduced to clean the system (column, injection port) in order to minimize carryover effect [199].

3.2.9. Quality control samples

Application of QA procedures for metabolomics studies are based on the recommendation of the FDA to define accuracy and precision of analytical methods in targeted studies [18]. One

of the most important aspects of QA procedures is the analysis of quality control samples (QCs). In 1997 Goldberger et al. described the use of QCs to control and verification of GC–MS analytical processes [200]. Subsequently, Sangster et al. [174] and van der Greef et al. [201] introduced the basic principles of QCs injections throughout metabolomics analysis. Nowadays, the concept of QCs has been already successfully implemented and widely used as the most relevant QA procedure for untargeted metabolomics experiments [174,189,202]. As we have shown elsewhere [2,4,5,34,39,41,42,81,203–205], it is also overall practice to use QC as a main tool for QA assessment in every metabolomics study conducted in our research center.

Among numerous requirements that must be fulfilled by QCs, the most important is that QC sample should be homogenous and representative of the qualitative and quantitative composition of the study samples [174,189]. To enhance the overall quality of the analysis there are several types of QC sample available that fit the purpose of metabolomics studies. There are advantages and limitations of each of these options mainly related to the closeness of QCs composition to study samples and the procedures for which technical variability will be finally evaluated.

Raw (pooled) QC. Sangster et al. proposed pragmatic use of the preparation QC samples that consists of small aliquots of each biological sample to be studied [174]. Sample prepared in such manner is representative of the composition of all samples in the study [174,189,202]. However, the preparation of a QCs is dependent on the sample type, availability and the size of the study [189]. Usually it is easy to prepare QC samples from biofluids like plasma or urine but for samples where limited sample volume is available or in case of tissue samples, preparation of pooled QCs could be challenging [189]. Similar situation occurs when we consider long-term experiments with several separate analytical batches applied. Running the large cohort samples in separated batches is recommended strategy due to the changes in signal response and the necessity of instrument re-calibration and cleaning [24]. Following Sangster's et al. recommendations for batches of 100 or so samples the QCs should represent a minimum of 10% of the total analysed samples or more if the batch size is smaller. Zelena et al. implement pooled QCs for the development of robust UPLC–MS based method for the long-term study of human serum samples [97]. For other purpose, Gika et al. used pooled QCs for determination within-day reproducibility [186] and day-to-day reproducibility of LC–MS metabolite profiling [206]. Though pooled QC provides many advantages, inconsistencies with experimental samples, due to the metabolite dilution could be also observed.

External (surrogate) QC. Taking into consideration that QCs should be identical for a whole study, external (surrogate) QC sample could be recommended. The surrogate QC should be matrix-matched with the study samples. The concept of surrogate QC sample is specially adapted to control the quality of the data in a large scale studies. The application of such QC sample and its utility in data treatment has been proposed by Dunn et al. [24]. In the HUSERMET project, sample preparation and data acquisition started before the whole sample set was completed [24]. In such case surrogate QC is an alternative that should be evaluated in the experimental design. Moreover, surrogate QC could be also used in cases when sample volume is limited or in case of especially important samples.

Commercially available QC. The commercially available QC samples will mimic the qualitative and quantitative composition, but with a lower accuracy than a pooled QC sample [24,205]. Such type of QCs could be also interesting in case of large cohort studies. Nevertheless, caution should always be taken when applying this option. In case of HUSERMET project it was reported that the metabolites composition and their relative concentrations were different between the commercially provided serum sample and the distribution of serum collected from the studied population [24]. Ideally, QC materials simulate the composition of patient samples as closely as possible in order to minimize matrix effects and correctly reflect the expected performance with patient samples. According to FDA recommendations information concerning the composition of the QC material should include: (i) base matrix characterization (e.g., plasma, serum, urine, etc.) and any known matrix effect that affect assay performance; (ii) source, e.g., human or animal species, synthetic or recombinant; (iii) information of added components such as stabilizers, preservative; (iv) specification of metabolite concentration level; (v) any other information pertinent to users concerning how the QC material is prepared; (vi) safety information [18].

Post-extraction QC. Dunn et al. discussed the use of post-extraction QC samples as an alternative for limited sample volume or in case of samples technically difficult to pool and sub-aliquot e.g., tissues or cells [189]. In such case a small aliquot from each of the extraction solutions is taken and combined [189]. Such QCs will provide only a measure of the variation associated with data acquisition and data preprocessing but will not include variation introduced during sample preparation.

Group specific QC. Prepared separately for each study group. Godzien et al. has proven that such QC offers the closest sample matrix to the samples and improves the statistical outcome, especially for biomarkers unique to one group [205].

QA strategy based on QC samples is a pragmatic solution, introduce several advantages from metabolomics perspective and could be applied to:

(i) Equilibrate or condition the system in order to achieve fully reproducible conditions [2,4,24,81,174,189]. One of the first studies that implemented QCs reported by Gika et al. showed that the first few injections on the system are not representative and should be discarded [82,186]. This study has proved that chromatographic system needs to be equilibrated to the samples (presumably by masking of active sites). Hence, the repeated injection (from 5 up to 10, depending on the analytical platform and sample type) of QCs that consist the matrix compatible with experimental samples are mandatory to achieve RT stability and fully reproducible results [186]. As suggested by Zhao et al. the frequencies of QC injection could also alter the correction factor calculated to further calibration method [13]. Zelena et al. suggested that for UPLC–MS serum analysis at least 10 conditioning injections of QCs are required before the commencement of the analytical run [97]. The data acquired from equilibration samples are removed from the dataset before further data pre-processing step [24].

(ii) Control of the performance of the analytical system. QC samples are analysed at the beginning and end of an analytical run as well as at regular intervals through the analysis in order to monitor stability and reproducibility of the analytical process [2,4,24,81]. That was evaluated by several experiments as reported by e.g., Sangster et al. [174], Gika et al. [186,206], Zelena et al. [97], Dunn et al. [189] among others. Gika et al. estimated the day-to-day variability of the LC–MS method by repeatitive analysis of a set of 60 urine samples during five different days and day-to-day reproducibility of the data was determined [206]. The results suggested that, 3 days continuous run is possible without the need for system maintenance. The variation as result of source contamination was notable over the 7-day of the study [206]. Nevertheless, Zelena et al. showed that the instrumental drift becomes intolerable after approximately 100 injections of the serum in UPLC–MS system. In this case the optimum batch length for studied application was consisted of 60 sample injections [97].

(iii) Correct small levels of drift in the measured signal over analysis (within batches) and between analytical batches [24].
(iv) Integrate the data from different analytical batches [24]. Long work lists or multiple matches with instrumental maintenance (cleaning, calibration) between batches exhibit risk of signal drift [207]. This can occur due to the changes in both chromatographic performance (e.g., from column aging) as well as signal intensity due to ion source contamination in case of LC–MS or injection device contamination of GC–MS system [207]. Additionally, the stability of GC–MS derivatized samples is also limited. This factors lead to poor repeatability and reproducibility and have a major negative impact on data quality [201]. Several authors have proposed the use of pooled QC samples, to correct instrumental drift within and between analytical batches [24,97,103,201,207]. In 2007 van der Greef developed a calibration procedure based on repeated analysis of a biological calibration samples (CAL samples) for removal of systematic batch-to-batch differences in response due to variation of extraction or derivatization procedures and LC–MS/GC–MS conditions. Proposed method was based on scaling of all metabolites with the optimal internal standard, previously selected due to statistical evaluation of CAL samples, determination of the temporal trends in CAL sample data per batch and the correction of these trends (curve fitting) [201]. Two years later van der Kloet et al. introduced a regression method that relies on pooled calibration samples and multiple internal standard strategy. QCs were used to select the best response scaling factor of internal standards for each feature. The procedure improved the coefficient of variation (CV) of the studied metabolites, however in case of increasing the number of features, large number of standards will be also required [208]. Zelena et al. used QCs for constructing the correction curve in order to correct sample to-sample signal drift [97]. In 2011 Dunn et al. applied Quality Control-Robust Loess Signal Correction (QC-RLSC) to correct for and minimize the impact of analytical variation. This algorithm incorporates locally estimated scatterplot smoothing function (LOESS) that is fitted to the QC data with respect to the order of injection. A correction curve for the complete data matrix is then interpolated, to which the total data set for that feature is normalized. Described method reduce the technical variation observed and allow to integrate data across different analytical experiments, maintaining the same level of biological variation [24]. Furthermore, Kamleh et al. tested the QC-based methods (feature-based signal correction, FBSC) against other current drift correction procedures e.g., total intensity normalization [207]. A remarkable improvement in reproducibility was observed with FBSC procedures compared to the pre-processed data and other correction procedures. The suggested protocol was found to be superior to currently used methods and showed a remarkable improvement in precision either within a block or between blocks of long LC–MS runs [207]. Kirwan et al. successfully applied method based on QC samples and robust cubic smoothing splines (Quality Control-Robust Spline Correction, QC-RSC) batch correction algorithm in a study designed to measure the validity and experimental reproducibility of a large, multi-batch (8 batches across 7 days) direct infusion mass spectrometry (DIMS) metabolomics study of mammalian cardiac tissue extracts. This method has shown to be useful for the evaluation and correction of both intra-and inter-batch instrumental stability. However, QC-RSC requires the analysis of replicates from several individuals, that allow the assessment of the correction performance [116]. Recently, Kuligowski et al. proposed intra-batch effect correction based on QC samples, support vector regression (QC-SVRC) and a radial basis function kernel. The authors evaluated and compared QC-SVRC algorithm with a recently developed QC-RSC method. The results show that QC-SVRC slightly outperformed QC-RSC and allows a straightforward fitting of the SVRC parameters to the instrument performance by using the ε-insensitive loss parameter [209]. Zhao et al. from the Guowang Xu Laboratory of Separation Science for Analytical Chemistry in China developed a novel correction strategy for large-scale and long-term studies [13]. The suggested method was applied to handle 1197 plant samples in nine batches analysed by two GC–MS instruments. Virtual QC of each sample was created by a linear fitting model of the feature intensities between two neighboring QCs to obtain a correction factor and remove the systematic bias. Results are very promising and showed that the approach outperforms the commonly used internal standard correction and total intensity signal correction methods.
(v) Calculation of metabolites precision [166,174,189,191]. For that purpose, several post data acquisition methods based on QC performance have been developed. Dunn et al. proposed QA procedure strategies for data filtering and signal correction based on QC samples [24]. It was suggested that data should be filtered based on at least 50% presence in QCs and exhibiting RSD lower than 20% in HPLC–MS data or 30% in GC–MS data as the sample preparation, metabolites extraction procedures (chemical derivatization) and samples injection can introduce higher variation. As previously mentioned FDA guidance for bioanalytical method validation in industry [18] for single analyte test tolerance limits are set, such that the measured response detected in two-thirds of QCs is within 15% of the QC mean. For compounds with concentrations at or near the LOQ a tolerance of 20% is acceptable. Whereby, the calculation of RSD for each metabolic feature for all the QCs across the analytical batch provides a quantitative measurement of precision. Moreover, features showing a comparable variation within the QC samples relative to the variability within the biological samples of interest in the study can also be considered unreliable and thus removed from subsequent statistical analysis [24,210].

Godzien et al. described QA+ procedure that offers an alternative to recover information that could be lost in filtering against surrogate QCs. That protocol takes into account, features absent in QCs but present in a high portion of samples from one of the experimental groups [205]. On this premise, other authors discuss several strategies for data filtering based on QC samples, e.g., by filtering by the presence of metabolic features in 50–80% of QCs samples [82,191]. Nevertheless, cut-off values are not straightforward and caution should be always taken with the abundance of some metabolites close to the limit of detection and subsequent dilution may result in intensities under a given threshold. For that reason some metabolic features could be lost during data preprocessing [14].

3.2.10. Standard reference materials

There are certain occasions where samples are difficult to obtaine or scarce and reference materials (RMs) are used in order to ensure the quality of the measurements. Certified reference materials (CRM) are RMs having a certificate of analysis for analytes, their concentration and confidence interval [211]. The best option is using the same type of sample but there are cases where reference materials do not exist and a matrix-matched QC surrogate is chosen [212–214]. During a validation study, CRM can be used for spiking the standards in them or as surrogate QCs allowing the follow-up of the experiments.

3.2.11. Other standards

The use of IS or test mixtures are approaches to control changes that may occur along the analysis (e.g., RT shifts, changes in peak

shape or in the detector response) improving system suitability and systems stability [97,173,215] as previously mentioned. Peng et al. used what they called universal metabolome-standard (UMS) method to test the reproducibility of their profiling method with LC–MS over a period of three months. An isotope labeled "standard" sample, prepared with the same type of sample (urine) but belonging to individuals that were not included in the initial study, "serve as a global internal standard that can be universally applied for metabolome profiling of any samples of the same type". After one year they verified and proved the method reproducibility and they propose advantages of UMS method over external calibrations [216]. Other authors as Koellensperger et al. [217] performed very recently internal standardization based on anion-exchange chromatography by addition of fully ^{13}C labeled extracts from *Pichia pastoris* in adherent cancer cells enabling absolute quantification of the primary metabolites and combining target and non-targeted metabolomics. With this strategy, quantitative analysis along with evaluation of response variation over the sequence, could be achieved without detrimental of global profiling evaluation.

3.2.12. Worklist

The length of the analytical run (worklist) is determined based on the number of samples in the study. In case of large studies it has been recommended to divide the experiment, and run samples in different analytical batches [218]. That strategy allows to decrease the previously described signal drift. The samples should be analysed as described in a random order. Previously the mixture of standards (standard mix) should be provided for the assessment of system suitability by examination of peak shape, RT stability and signal intensity. Blank samples have to be analysed at the beginning of the run, following injections of 5–10 QCs for system conditioning. Such injections can be used to assess system suitability before the experimental samples will be analysed. Subsequently QCs are intermittently analysed throughout the worklist (typically every 4–10 injection, depending of the length of analysis) [97,116,174,189]. However, choosing the optimal number of QCs is not straightforward and it depends on the type of samples, extraction procedures and the stability of the analytical system. According to van der Greef et al. when higher signal drift is observed, more QCs are required [201]. Optimal length of the batch is hugely important and should be also considered. As described above, long analytical run results in intolerable instrumental drift [24,97]. Therefore, for large studies it is recommended to perform analysis in separate analytical blocks. It was proved that an analytical batch with 100 injections is optimal for untargeted approaches [24,97,219]. Exemplary scheme of QC preparation by pooling aliquots of the study samples is represented in Fig. 2A, the analytical worklist is provided in Fig. 2 B and the superimposed QC chromatograms are depicted in Fig. 2C.

3.3. Post-analytical

3.3.1. Data visualization

The data visualization tools help to understand variation and could be used to evaluate analytical run quality and to check for systematic error and random error. Systematic error is evidenced by a change in the mean and may be demonstrated as a trend or a shift observed in QC samples (Fig. 2C). A trend indicates a gradual loss of reliability in the system and may be caused by e.g., gradual deterioration of the samples, aging of reagents, gradual deterioration of calibration. Shift observed in QCs represent a sudden change in the system performance and maybe caused by e.g., change in mobile phase composition, change of temperature, failure in the sampling system. Random error is defined as any deviation from an expected result, e.g., positive or negative deviation from the calculated mean for QC samples.

- **Chromatogram**

Inspection of raw data by the visualization of chromatograms and spectra allows the detection of outlying profiles that could occur due to several factors as e.g., acquisition problem, failed injection, incomplete extraction or derivatization [93]. That could also allow for quick overview of the data characteristic for the specified RT regions [114,220].

- **PCA**

Non-supervised method as principal component analysis (PCA) is a projection method used to checking for signal drift, sensitivity loss and variation in QC samples. Tightly clustering of QCs observed, indicate the precise analytical outcome. In most cases QCs prepared as a pool of all experimental samples, cluster in the middle of the scatter score plot, but it is also possible that QCs cluster moves in direction to one or other experimental group. That scenario is also possible and could be related to the higher amount of sample aliquots coming from one of the groups or stronger weight of variables in one group after dilution in the mixture (Fig. 2D).

- **Control chart (Shewhart control chart)**

Control charts are simple, robust tools for understanding process variability and is one of the main quality assurance tools. The control chart has a baseline and upper and lower limits. Measurements are plotted on the chart versus the acquisition order. Measurements that are outside the limits are consider as outlying results. The application of control charts is to control process stability. It enables fast and simple detection of abnormalities in the configuration of the marked points, and thus fast correction and confirmation of the reliability of the research. Data quality assessment include overview of the analytical precision based on plotting the sum of metabolic feature intensities for every experimental and QC sample against run order (Fig. 2E).

3.3.2. Annotation

False discovery rate (FDR). The FDR of annotation quality has been already introduced in other 'omic' approaches [221]. For the purpose of reliable metabolite annotation Matsuda et al. developed a methodology for false discovery rates (FDR) evaluation for assessment of metabolome annotation quality. Based on the FDR analyses, the authors discussed several aspects of an elemental composition search, including setting a threshold, estimating FDR, and the types of elemental composition databases most reliable for metabolite annotation [222]. The authors estimated that false positive rates for the elemental composition search is relatively high, therefore caution should be kept in applying the search results for metabolite annotation. The authors postulate that evaluation of FDR in elemental composition search results is essential for improving the quality of elemental composition search and allow to minimize misinterpretation of metabolome data [222]. However, the strategy proposed by Matsuda et al. in 2009 has not found widespread use. Recently, another strategy has been proposed by Palmer et al. that reported pySM, a framework for FDR control of metabolite annotation at the level of the molecular sum formula, for high-mass-resolution (HR) imaging mass spectrometry. A comprehensive open-source bioinformatics framework for FDR-controlled metabolite annotation for HR imaging MS has been developed (https://github.com/alexandrovteam/pySM). As described by the authors, the pySm is based on the three following principles: database-driven annotation by screening for

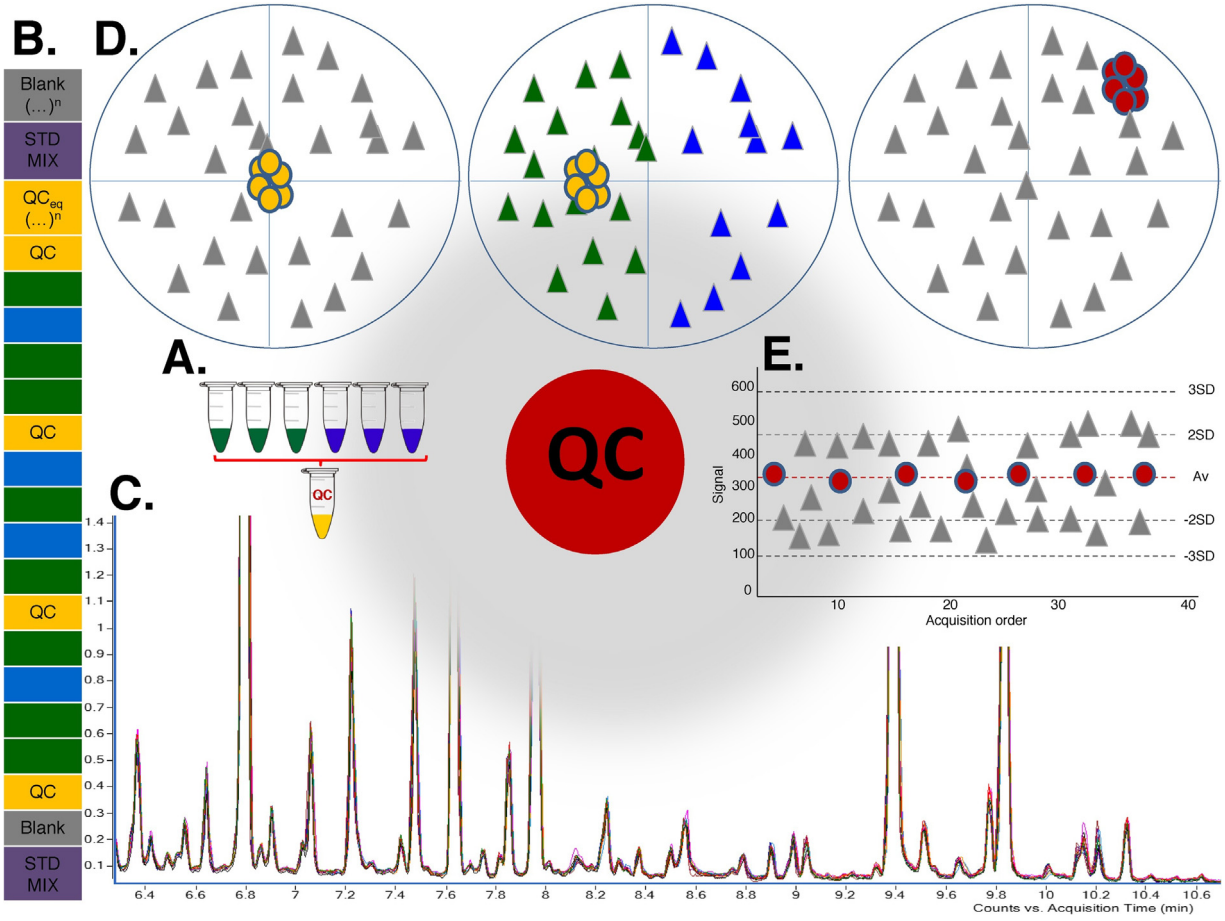

Fig. 2. Assessment of data quality based on QC samples. **2.A.** The QCs sample preparation. QC pooled from experimental samples, giving the best performance for QA procedure. **2.B.** Typical analytical batch for untargeted metabolomics experiment, started with blank injections, STD MIX, conditioning QC samples, then following QCs injection every equal intervals of experimental samples and at the end of the batch. **2.C.** First examination of raw chromatographic data (RT stability, peaks intensity). **2.D.** Series of multivariate PCA-X plots showing schematic prediction for the QCs. The tightly clustering of QCs could be observed, indicating the precise analytical outcome. QCs prepared by pooling equal volume of the balanced number of case-control samples will cluster at the center of the plot. When QCs consists the higher volume of one experimental group, the cluster will move toward to that group. In case of external material (e.g., surrogate QC), QCs could cluster in any area of the plot. **2.E.** Shewhart's chart enable the visual assessment of the changes in metabolic features responses over the acquisition order. The level of QCs and sample variance could be observed. (▲ ▲ ▲ experimental samples; ● ● QC samples).

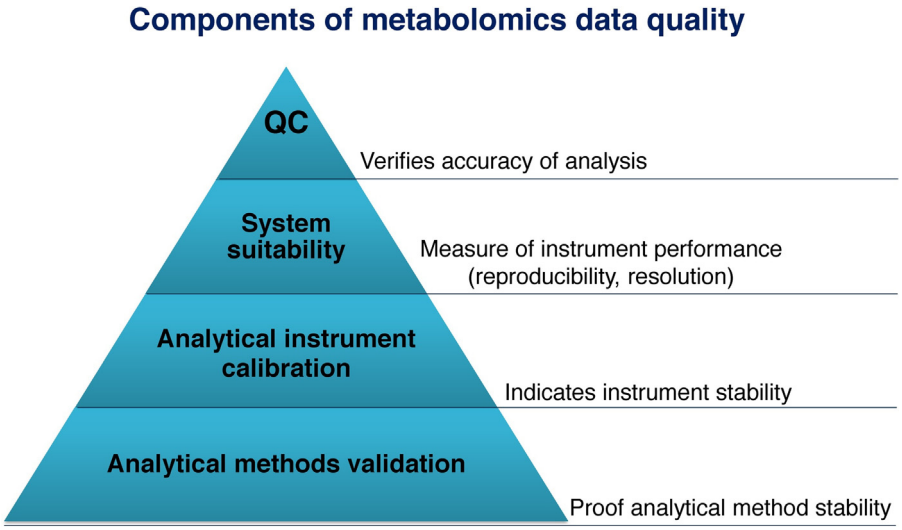

Fig. 3. The main factors required to obtain reliable and consistent data of metabolomics experiments. Adapted from USP 37 <Chapter 1058> [171].

metabolites with known sum formulas, an original metabolite-signal match (MSM) score combining spectral and spatial measures, and a novel target–decoy FDR-estimation approach with a decoy set generated via the use of implausible adducts [223].

Retention projection. Abate-Pella et al. developed the strategy aiming to calculate LC gradient retention across laboratories, where retention was reliably calculated under various multi-segment gradients and flow rates chosen independently. The methodology shows a great potential as a tool to enable LC retention information for compound identification and the open-access database is available at www.retentionprediction.org/hplc/database [224].

3.3.3. Statistical methods

Multiple testing corrections. Several methods have been already introduced to address multiple testing problem [225–228]. The simplest method, minimizing the family-wise error rate is the Bonferroni adjustment. However, this approach often produces overly conservative thresholds and for metabolomics specific data is rather too strict, especially when number of tests is large. Instead, False discovery rate control (FDR) has been proposed by Benjamini and Hochberg [225] or Benjamini–Yekutieili [226] and they are important tools to reduce type II error rates of statistical inference in feature selection [229]. In 2001 Storey et al. introduced complementary to the FDR, the pFDR, the positive false discovery rate and the q-value that gives each feature its own individual measure of significance has been introduced [227,228]. Whereas the p value is a measure of significance in terms of the false positive rate, the q value is a measure in terms of the FDR [230]. Storey at al. emphasized that the false positive rate and FDR are often mistakenly equated, but their difference is very important. The authors explain that a false positive rate of e.g., 5% means that on average 5% of the truly null features in the study will be called significant, whereas FDR of 5% means that among all features called significant, 5% of these are truly null on average [230]. Based on this assumption that the traditional FDR methods treat all features equally, which can cause substantial loss of statistical power to detect differentially expressed features. Chong et al. proposed to estimate a reliability index for MS based metabolomics data. The method amounts to a soft stratification of features based on their reliability levels. Each feature is compared to the null distribution derived from all the features with similar reliability level to obtain the lfdr values. The null density is computed based on permutation without changing the reliability indices [229]. Other methodology, standard deviation step down (SDSD) has been proposed by Wang et al. The author indicate that the method is more sensitive and appropriate than Bonferroni test for partially dependent data sets and the sensitivity and type I error rate of SDSD can be adjusted based on the degree of variable dependency [231].

Normalization. In untargeted metabolomics analysis, several factors related to unwanted experimental or biological variations, as presented in Fig. 1 may hamper the identification of differential metabolic features. Up to date several normalization methods have been widely applied for MS based metabolomics data. The commonly used strategies can be grouped into two categories: (1) method-driven normalization approaches based on IS or QCs and (2) data-driven normalization approaches, data scaling or transforming [20,232,233]. The choice of an appropriate normalization method should consider several aspects and it should also implicate basic statistical assumption [233,234] Comprehensive comparison of several data normalization methods has been already provided by Li at al. [20] or Ejigu et al. [232] among others and the most common used algorithms [233–240] are commented in Table 1 as single internal standard (SIS), multiple internal standards (AIS) and locally weighted scatter plot smoothing (LOESS) for method-driven normalization and normalization based on scaling factors, total intensity normalization, median fold change (FC), quantile normalization and probabilistic quotient normalization (PQN) for data-driven normalization.

4. Conclusions

All issues and queries related to pre-analytical, analytical and post-analytical sources of bias and uncertainty have to be gradually addressed in standard operating procedures based on continuous quality assurance, as an integral part of the successful untargeted metabolomics experiment. The factors associated with pre-analytical phase needs particular attention, because they are affected by high potential variations that would diminish the validity of the final results. That could be one of possible reasons why we cannot compare several studies or validate potential biomarker candidates. Special attention should be given to ensure consistency in sample collection within multicenter large cohort studies. Hence, it should be emphasized that high quality of biological samples is essential for the outcome of subsequent analysis.

Quality assurance seems to be still a 'dark side' of untargeted metabolomics and reliable strategies need to be correctly identified and routinely implemented in all steps of the metabolomics workflow. Several procedures that have been already developed to control analytical data quality could be and should be adapted for quality assurance procedures along with those especially suited and already implemented for untargeted metabolomics experiments. Standardization of the protocols will improve the overall quality and will contribute to the development of meaningful, accurate and precise management of untargeted studies.

Funding

This work was supported by the Spanish Ministry of Economy, Industry and Competitiveness (CTQ2014-55279-R). C. B.-B. thanks for the financial support provided by the Spanish Ministry of Education, Culture and Sport (FPU 15/03032).

References

[1] M. Frédérich, B. Pirotte, M. Fillet, P. de Tullio, Metabolomics as a challenging approach for medicinal chemistry and personalized medicine, J. Med. Chem. 59 (2016) 8649–8666, http://dx.doi.org/10.1021/acs.jmedchem.5b01335.
[2] D. Dudzik, R. Revello, C. Barbas, J.L. Bartha, LC-MS-based metabolomics identification of novel biomarkers of chorioamnionitis and its associated perinatal neurological damage, J. Proteome Res. 14 (2015) 1432–1444, http://dx.doi.org/10.1021/pr501087x.
[3] R. Bujak, E. Daghir-Wojtkowiak, R. Kaliszan, M.J. Markuszewski, PLS-based and regularization-based methods for the selection of relevant variables in non-targeted metabolomics data, Front. Mol. Biosci. 3 (2016) 35, http://dx.doi.org/10.3389/fmolb.2016.00035.
[4] D. Dudzik, M. Zorawski, M. Skotnicki, W. Zarzycki, G. Kozlowska, K. Bibik-Malinowska, M. Vallejo, A. García, C. Barbas, M.P. Ramos, Metabolic fingerprint of gestational diabetes mellitus, J. Proteom. 103 (2014) 57–71, http://dx.doi.org/10.1016/j.jprot.2014.03.025.
[5] A. Mastrangelo, M.I. Panadero, L.M. Perez, B.G. Galvez, A. Garcia, C. Barbas, F.J. Ruperez, New insight on obesity and adipose-derived stem cells using comprehensive metabolomics, Biochem. J. 473 (2016) 2187–2203, http://dx.doi.org/10.1042/BCJ20160241.
[6] I. Tzoulaki, T.M. Ebbels, A. Valdes, P. Elliott, J.P. Ioannidis, Design and analysis of metabolomics studies in epidemiologic research: a primer on −omic technologies, Am. J. Epidemiol. 180 (2014) 129–139, http://dx.doi.org/10.1093/aje/kwu143.
[7] M. Alewijn, H. van der Voet, S. van Ruth, Validation of multivariate classification methods using analytical fingerprints −concept and case study on organic feed for laying hens, J. Food Compos. Anal. 51 (2016) 15–23, http://dx.doi.org/10.1016/j.jfca.2016.06.003.
[8] J. Riedl, S. Esslinger, C. Fauhl-Hassek, Review of validation and reporting of non-targeted fingerprinting approaches for food authentication, Anal. Chim. Acta 885 (2015) 17–32, http://dx.doi.org/10.1016/j.aca.2015.06.003.
[9] R. Ramautar, R. Berger, J. van der Greef, T. Hankemeier, Human metabolomics: strategies to understand biology, Curr. Opin. Chem. Biol. 17 (2013) 841–846, http://dx.doi.org/10.1016/j.cbpa.2013.06.015.
[10] I. Garcia-Perez, J.M. Posma, R. Gibson, E.S. Chambers, T.H. Hansen, H. Vestergaard, T. Hansen, M. Beckmann, O. Pedersen, P. Elliott, J. Stamler, J.K. Nicholson, J. Draper, J.C. Mathers, E. Holmes, G. Frost, Objective assessment

of dietary patterns by use of metabolic phenotyping: a randomised, controlled, crossover trial, Lancet Diabetes Endocrinol. 5 (2017) 184–195, http://dx.doi.org/10.1016/S2213-8587(16)30419-3.
[11] E. Chekmeneva, G. Dos Santos Correia, Q. Chan, A. Wijeyesekera, A. Tin, J.H. Young, P. Elliott, J.K. Nicholson, E. Holmes, Optimization and large-scale application of direct infusion nanoelectrospray HRMS method for large-Scale urinary metabolic phenotyping in molecular epidemiology, J. Proteome Res. 16 (2017) 1646–1658, http://dx.doi.org/10.1021/acs.jproteome.6b01003.
[12] 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, http://dx.doi.org/10.1021/acs.analchem.6b01481.
[13] Y. Zhao, Z. Hao, C. Zhao, J. Zhao, J. Zhang, Y. Li, L. Li, X. Huang, X. Lin, Z. Zeng, X. Lu, G. Xu, A novel strategy for large-scale metabolomics study by calibrating gross and systematic errors in gas chromatography-mass spectrometry, Anal. Chem. 88 (2016) 2234–2242, http://dx.doi.org/10.1021/acs.analchem.5b03912.
[14] M.K.R. Engskog, J. Haglöf, T. Arvidsson, C. Pettersson, LC-MS based global metabolite profiling: the necessity of high data quality, Metabolomics 12 (2016) 114, http://dx.doi.org/10.1007/s11306-016-1058-x.
[15] J. Laparre, Z. Kaabia, M. Mooney, T. Buckley, M. Sherry, B. Le Bizec, G. Dervilly-Pinel, Impact of storage conditions on the urinary metabolomics fingerprint, Anal. Chim. Acta 951 (2017) 99–107, http://dx.doi.org/10.1016/j.aca.2016.11.055.
[16] B. Kamlage, S.G. Maldonado, B. Bethan, E. Peter, O. Schmitz, V. Liebenberg, P. Schatz, Quality markers addressing preanalytical variations of blood and plasma processing identified by broad and targeted metabolite profiling, Clin. Chem. 60 (2014) 399–412, http://dx.doi.org/10.1373/clinchem.2013.211979.
[17] P. Yin, A. Peter, H. Franken, X. Zhao, S.S. Neukamm, L. Rosenbaum, M. Lucio, A. Zell, H.U. Häring, G. Xu, R. Lehmann, Preanalytical aspects and sample quality assessment in metabolomics studies of human blood, Clin. Chem. 59 (2013) 833–845, http://dx.doi.org/10.1373/clinchem.2012.199257.
[18] FDA, Guidance for Industry: Bioanalytical Method Validation, 2013 http://www.labcompliance.de/documents/FDA/FDA-Others/Laboratory/f-507-bioanalytical-4252fnl.pdf.
[19] FDA, Guidance for Industry Analytical Procedures and Methods Validation for Drugs and Biologics, 2000.
[20] B. Li, J. Tang, Q. Yang, X. Cui, S. Li, S. Chen, Q. Cao, W. Xue, N. Chen, F. Zhu, Performance evaluation and online realization of data-driven normalization methods used in LC/MS based untargeted metabolomics analysis, Sci. Rep. 6 (2016) 38881, http://dx.doi.org/10.1038/srep38881.
[21] J.G. Cecatti, R.T. Souza, K. Sulek, M.L. Costa, L.C. Kenny, L.M. McCowan, R.C. Pacagnella, S.G. Villas-Boas, J. Mayrink, R. Passini, K.G. Franchini, P.N. Baker, Preterm SAMBA and SCOPE study groups, Use of metabolomics for the identification and validation of clinical biomarkers for preterm birth: preterm SAMBA, BMC Pregnancy Childbirth 16 (2016) 212, http://dx.doi.org/10.1186/s12884-016-1006-9.
[22] G. Martos-Moreno, A. Mastrangelo, V. Barrios, A. García, J.A. Chowen, F.J. Rupérez, C. Barbas, J. Argente, Metabolomics allow the discrimination of the pathophysiological relevance of hyperinsulinism in obese prepubertal children, Int. J. Obes. (2017), http://dx.doi.org/10.1038/ijo.2017.137.
[23] K. Ortmayr, V. Charwat, C. Kasper, S. Hann, G. Koellensperger, Uncertainty budgeting in fold change determination and implications for non-targeted metabolomics studies in model systems, Analyst 142 (2017) 80–90, http://dx.doi.org/10.1039/C6AN01342B.
[24] W.B. Dunn, D. Broadhurst, P. Begley, E. Zelena, S. Francis-McIntyre, N. Anderson, M. Brown, J.D. Knowles, A. Halsall, J.N. Haselden, A.W. Nicholls, I.D. Wilson, D.B. Kell, R. Goodacre, Human Serum Metabolome (HUSERMET) Consortium, 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, http://dx.doi.org/10.1038/nprot.2011.335.
[25] P. Bernini, I. Bertini, C. Luchinat, P. Nincheri, S. Staderini, P. Turano, Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks, J. Biomol. NMR 49 (2011) 231–243, http://dx.doi.org/10.1007/s10858-011-9489-1.
[26] 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 chemical analysis working group (CAWG) metabolomics standards initiative (MSI), Metabolomics 3 (2007) 211–221, http://dx.doi.org/10.1007/s11306-007-0082-2.
[27] A.M. De Livera, M. Sysi-Aho, L. Jacob, J.A. Gagnon-Bartsch, S. Castillo, J.A. Simpson, T.P. Speed, Statistical methods for handling unwanted variation in metabolomics data, Anal. Chem. 87 (2015) 3606–3615, http://dx.doi.org/10.1021/ac502439y.
[28] A. Petrie, C. Sabin, Medical Statistics at a Glance, Wiley-Blackwell, 2009.
[29] W.B. Dunn, D. Broadhurst, D.I. Ellis, M. Brown, A. Halsall, S. O'Hagan, I. Spasic, A. Tseng, D.B. Kell, A GC-TOF-MS study of the stability of serum and urine metabolomes during the UK Biobank sample collection and preparation protocols, Int. J. Epidemiol. 37 (2008) i23–i30, http://dx.doi.org/10.1093/ije/dym281.
[30] E. Jobard, O. Trédan, D. Postoly, F. André, A.L. Martin, B. Elena-Herrmann, S. Boyault, A systematic evaluation of blood serum and plasma pre-analytics for metabolomics cohort studies, Int. J. Mol. Sci. 17 (2016) 2035, http://dx.doi.org/10.3390/ijms17122035.
[31] P. Yin, R. Lehmann, G. Xu, Effects of pre-analytical processes on blood samples used in metabolomics studies, Anal. Bioanal. Chem. 407 (2015) 4879–4892, http://dx.doi.org/10.1007/s00216-015-8565-x.
[32] N.T. Holland, M.T. Smith, B. Eskenazi, M. Bastaki, Biological sample collection and processing for molecular epidemiological studies, Mutat. Res. 543 (2003) 217–234 http://www.ncbi.nlm.nih.gov/pubmed/12787814.
[33] V.V. Hernandes, C. Barbas, D. Dudzik, A review of blood sample handling and pre-processing for metabolomics studies, Electrophoresis (2017), http://dx.doi.org/10.1002/elps.201700086.
[34] D. Dudzik, M. Zorawski, M. Skotnicki, W. Zarzycki, A. García, S. Angulo, M.P. Lorenzo, C. Barbas, M.P. Ramos, GC–MS based Gestational Diabetes Mellitus longitudinal study: identification of 2-and 3-hydroxybutyrate as potential prognostic biomarkers, J. Pharm. Biomed. Anal. (2017), http://dx.doi.org/10.1016/j.jpba.2017.02.056.
[35] S. Bouatra, F. Aziat, R. Mandal, A.C. Guo, M.R. Wilson, C. Knox, T.C. Bjorndahl, R. Krishnamurthy, F. Saleem, P. Liu, Z.T. Dame, J. Poelzer, J. Huynh, F.S. Yallou, N. Psychogios, E. Dong, R. Bogumil, C. Roehring, D.S. Wishart, The human urine metabolome, PLoS One 8 (2013), http://dx.doi.org/10.1371/journal.pone.0073076.
[36] B. Álvarez-Sánchez, F. Priego-Capote, M.D. Luque de Castro, Study of sample preparation for metabolomic profiling of human saliva by liquid chromatography-time of flight/mass spectrometry, J. Chromatogr. A 1248 (2012) 178–181, http://dx.doi.org/10.1016/j.chroma.2012.05.029.
[37] D.S. Wishart, M.J. Lewis, J.A. Morrissey, M.D. Flegel, K. Jeroncic, Y. Xiong, D. Cheng, R. Eisner, B. Gautam, D. Tzur, S. Sawhney, F. Bamforth, R. Greiner, L. Li, The human cerebrospinal fluid metabolome, J. Chromatogr. B 871 (2008) 164–173, http://dx.doi.org/10.1016/j.jchromb.2008.05.001.
[38] G. Graça, B.J. Goodfellow, A.S. Barros, S. Diaz, I.F. Duarte, K. Spagou, K. Veselkov, E.J. Want, J.C. Lindon, I.M. Carreira, E. Galhano, C. Pita, A.M. Gil, UPLC-MS metabolic profiling of second trimester amniotic fluid and maternal urine and comparison with NMR spectral profiling for the identification of pregnancy disorder biomarkers, Mol. Biosyst. 8 (2012) 1243–1254, http://dx.doi.org/10.1039/c2mb05424h.
[39] A. Villaseñor, I. Garcia-Perez, A. Garcia, J.M. Posma, M. Fernández-López, A.J. Nicholas, N. Modi, E. Holmes, C. Barbas, Breast milk metabolome characterization in a single-phase extraction, multiplatform analytical approach, Anal. Chem. 86 (2014) 8245–8252, http://dx.doi.org/10.1021/ac501853d.
[40] B. Madji Hounoum, H. Blasco, P. Emond, S. Mavel, Liquid chromatography-high-resolution mass spectrometry-based cell metabolomics: experimental design, recommendations, and applications, TrAC – Trends Anal. Chem. 75 (2016) 118–128, http://dx.doi.org/10.1016/j.trac.2015.08.003.
[41] D. González-Peña, D. Dudzik, A. García, B. Ancos, C. Barbas, C. Sánchez-Moreno, Metabolomic fingerprinting in the comprehensive study of liver changes associated with onion supplementation in hypercholesterolemic wistar rats, Int. J. Mol. Sci. 18 (2017) 267, http://dx.doi.org/10.3390/ijms18020267.
[42] C. Gonzalez-Riano, S. Tapia-González, A. García, A. Muñoz, J. DeFelipe, C. Barbas, Metabolomics and neuroanatomical evaluation of post-mortem changes in the hippocampus, Brain Struct. Funct. (2017), http://dx.doi.org/10.1007/s00429-017-1375-5.
[43] E.J. Want, P. Masson, F. Michopoulos, I.D. Wilson, G. Theodoridis, R.S. Plumb, J. Shockcor, N. Loftus, E. Holmes, J.K. Nicholson, Global metabolic profiling of animal and human tissues via UPLC-MS, Nat. Protoc. 8 (2013) 17–32, http://dx.doi.org/10.1038/nprot.2012.135.
[44] Z. Yu, G. Kastenmüller, Y. He, P. Belcredi, G. Möller, C. Prehn, J. Mendes, S. Wahl, W. Roemisch-Margl, U. Ceglarek, A. Polonikov, N. Dahmen, H. Prokisch, L. Xie, Y. Li, H.E. Wichmann, A. Peters, F. Kronenberg, K. Suhre, J. Adamski, T. Illig, R. Wang-Sattler, Differences between human plasma and serum metabolite profiles, PLoS One. 6 (2011) e21230, http://dx.doi.org/10.1371/journal.pone.0021230.
[45] M.K. Tuck, D.W. Chan, D. Chia, A.K. Godwin, W.E. Grizzle, K.E. Krueger, W. Rom, M. Sanda, L. Sorbara, S. Stass, D.E. Brenner, Standard operating procedures for serum and plasma collection, J. Proteome Res. 8 (2010) 113–117, http://dx.doi.org/10.1021/pr800545q.Standard.
[46] J.R. Denery, A.A. Nunes, T.J. Dickerson, Characterization of differences between blood sample matrices in, Anal. Chem. 83 (2011) 1040–1047.
[47] A. Roux, E. Thévenot, F. Seguin, M.F. Olivier, C. Junot, Impact of collection conditions on the metabolite content of human urine samples as analyzed by liquid chromatography coupled to mass spectrometry and nuclear magnetic resonance spectroscopy, Metabolomics 11 (2014) 1095–1105, http://dx.doi.org/10.1007/s11306-014-0764-5.
[48] E.J. Saude, B.D. Sykes, Urine stability for metabolomic studies: effects of preparation and storage, Metabolomics 3 (2007) 19–27, http://dx.doi.org/10.1007/s11306-006-0042-2.
[49] M.A. Fernández-Peralbo, M.D. Luque de Castro, Preparation of urine samples prior to targeted or untargeted metabolomics mass-spectrometry analysis, TrAC – Trends Anal. Chem. 41 (2012) 75–85, http://dx.doi.org/10.1016/j.trac.2012.08.011.

[50] F. Boomsma, G. Alberts, L. van Eijk, A.J. Man in 't Veld, M.A. Schalekamp, Optimal collection and storage conditions for catecholamine measurements in human plasma and urine, Clin. Chem. 39 (1993) 2503–2508.
[51] K. Bando, R. Kawahara, T. Kunimatsu, J. Sakai, J. Kimura, H. Funabashi, T. Seki, T. Bamba, E. Fukusaki, Influences of biofluid sample collection and handling procedures on GC–MS based metabolomic studies, J. Biosci. Bioeng. 110 (2010) 491–499, http://dx.doi.org/10.1016/j.jbiosc.2010.04.010.
[52] M.M. Khamis, D.J. Adamko, A. El-Aneed, Mass spectrometric based approaches in urine metabolomics and biomarker discovery, Mass Spectrom. Rev. 36 (2017) 115–134, http://dx.doi.org/10.1002/mas.21455.
[53] P. Masson, A.C. Alves, T.M.D. Ebbels, J.K. Nicholson, E.J. Want, Optimization and evaluation of metabolite extraction protocols for untargeted metabolic profiling of liver samples by UPLC-MS, Anal. Chem. 82 (2010) 7779–7786, http://dx.doi.org/10.1021/ac101722e.
[54] R. El Rammouz, F. Létisse, S. Durand, J.C. Portais, Z.W. Moussa, X. Fernandez, Analysis of skeletal muscle metabolome: evaluation of extraction methods for targeted metabolite quantification using liquid chromatography tandem mass spectrometry, Anal. Biochem. 398 (2010) 169–177, http://dx.doi.org/10.1016/j.ab.2009.12.006.
[55] D. Vuckovic, Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry, Anal. Bioanal. Chem. 403 (2012) 1523–1548, http://dx.doi.org/10.1007/s00216-012-6039-y.
[56] S. Cacciatore, G. Zadra, C. Bango, K.L. Penney, S. Tyekucheva, O. Yanes, M. Loda, Metabolic profiling in formalin-fixed and paraffin embedded prostate cancer tissues, Mol. Cancer Res. 15 (2017) 439–447, http://dx.doi.org/10.1158/1541-7786.MCR-16-0262.
[57] A. Wojakowska, M. Chekan, Ł. Marczak, K. Polanski, D. Lange, M. Pietrowska, P. Widlak, Detection of metabolites discriminating subtypes of thyroid cancer: molecular profiling of FFPE samples using the GC/MS approach, Mol. Cell. Endocrinol. 417 (2015) 149–157, http://dx.doi.org/10.1016/j.mce.2015.09.021.
[58] M. Čuperlović-Culf, D.A. Barnett, A.S. Culf, I. Chute, Cell culture metabolomics: applications and future directions, Drug Discov. Today 15 (2010) 610–621, http://dx.doi.org/10.1016/j.drudis.2010.06.012.
[59] G.Y. Chen, H.W. Liao, I.L. Tsai, Y.J. Tseng, C.H. Kuo, Using the matrix-induced ion suppression method for concentration normalization in cellular metabolomics studies, Anal. Chem. 87 (2015) 9731–9739, http://dx.doi.org/10.1021/acs.analchem.5b01869.
[60] B. Madji Hounoum, H. Blasco, P. Emond, S. Mavel, Liquid chromatography-high-resolution mass spectrometry-based cell metabolomics: experimental design, recommendations, and applications, TrAC – Trends Anal. Chem. 75 (2016) 118–128, http://dx.doi.org/10.1016/j.trac.2015.08.003.
[61] H. Bi, K.W. Krausz, S.K. Manna, F. Li, C.H. Johnson, F.J. Gonzalez, Optimization of harvesting, extraction, and analytical protocols for UPLC-ESI–MS-based metabolomic analysis of adherent mammalian cancer cells, Anal. Bioanal. Chem. 405 (2013) 5279–5289, http://dx.doi.org/10.1007/s00216-013-6927-9.
[62] K. Dettmer, N. Nürnberger, H. Kaspar, M.A. Gruber, M.F. Almstetter, P.J. Oefner, Metabolite extraction from adherently growing mammalian cells for metabolomics studies: optimization of harvesting and extraction protocols, Anal. Bioanal. Chem. 399 (2011) 1127–1139, http://dx.doi.org/10.1007/s00216-010-4425-x.
[63] C.L. Winder, W.B. Dunn, S. Schuler, D. Broadhurst, R. Jarvis, G.M. Stephens, R. Goodacre, Global metabolic profiling of escherichia coli cultures: an evaluation of methods for quenching and extraction of intracellular metabolites, Anal. Chem. 80 (2008) 2939–2948, http://dx.doi.org/10.1021/ac7023409.
[64] A. Hirayama, M. Sugimoto, A. Suzuki, Y. Hatakeyama, A. Enomoto, S. Harada, T. Soga, M. Tomita, T. Takebayashi, Effects of processing and storage conditions on charged metabolomic profiles in blood, Electrophoresis 36 (2015) 2148–2155, http://dx.doi.org/10.1002/elps.201400600.
[65] P. Yin, R. Lehmann, G. Xu, Effects of pre-analytical processes on blood samples used in metabolomics studies, Anal. Bioanal. Chem. 407 (2015) 4879–4892, http://dx.doi.org/10.1007/s00216-015-8565-x.
[66] M. Breier, S. Wahl, C. Prehn, M. Fugmann, U. Ferrari, M. Weise, F. Banning, J. Seissler, H. Grallert, J. Adamski, A. Lechner, Targeted metabolomics identifies reliable and stable metabolites in human serum and plasma samples, PLoS One 9 (2014) 1–11, http://dx.doi.org/10.1371/journal.pone.0089728.
[67] G. Anton, R. Wilson, Z.H. Yu, C. Prehn, S. Zukunft, J. Adamski, M. Heier, C. Meisinger, W. Römisch-Margl, R. Wang-Sattler, K. Hveem, B. Wolfenbuttel, A. Peters, G. Kastenmüller, M. Waldenberger, Pre-analytical sample quality: metabolite ratios as an intrinsic marker for prolonged room temperature exposure of serum samples, PLoS One. 10 (2015) 1–19, http://dx.doi.org/10.1371/journal.pone.0121495.
[68] B.S. Bhatnagar, R.H. Bogner, M.J. Pikal, Protein stability during freezing: separation of stresses and mechanisms of protein stabilization, Pharm. Dev. Technol. 12 (2007) 505–523, http://dx.doi.org/10.1080/10837450701481157.
[69] A.M. Zivkovic, M.M. Wiest, U.T. Nguyen, R. Davis, S.M. Watkins, J.B. German, Effects of sample handling and storage on quantitative lipid analysis in human serum, Metabolomics 5 (2009) 507–516, http://dx.doi.org/10.1007/s11306-009-0174-2.
[70] M.P. Lorenzo, A. Villaseñor, A. Ramamoorthy, A. Garcia, Optimization and validation of a capillary electrophoresis laser-induced fluorescence method for amino acids determination in human plasma: application to bipolar disorder study, Electrophoresis 34 (2013) 1701–1709, http://dx.doi.org/10.1002/elps.201200632.
[71] J.P. Trezzi, A. Bulla, C. Bellora, M. Rose, P. Lescuyer, M. Kiehntopf, K. Hiller, F. Betsou, LacaScore: a novel plasma sample quality control tool based on ascorbic acid and lactic acid levels, Metabolomics 12 (2016) 96, http://dx.doi.org/10.1007/s11306-016-1038-1.
[72] H.G. Gika, G.A. Theodoridis, I.D. Wilson, Liquid chromatography and ultra-performance liquid chromatography-mass spectrometry fingerprinting of human urine. Sample stability under different handling and storage conditions for metabonomics studies, J. Chromatogr. A 1189 (2008) 314–322, http://dx.doi.org/10.1016/j.chroma.2007.10.066.
[73] J. Malm, T.E. Fehniger, P. Danmyr, A. Végvári, C. Welinder, H. Lindberg, R. Appelqvist, K. Sjödin, E. Wieslander, T. Laurell, S. Hober, F.S. Berven, D. Fenyö, X. Wang, P.E. Andrén, G. Edula, E. Carlsohn, M. Fuentes, C.L. Nilsson, M. Dahlbäck, M. Rezeli, D. Erlinge, G. Marko-Varga, Developments in biobanking workflow standardization providing sample integrity and stability, J. Proteom. 95 (2013) 38–45, http://dx.doi.org/10.1016/j.jprot.2013.06.035.
[74] Å.M. Wheelock, L. Paulson, J.E. Litton, EuPA biobank initiative group, the EuPA biobank initiative: meeting the future challenges of biobanking in proteomics & systems medicine, J. Proteom. 127 (2015) 414–416, http://dx.doi.org/10.1016/j.jprot.2015.07.014.
[75] D. Simeon-Dubach, S.M. Zeisberger, S.P. Hoerstrup, Quality assurance in biobanking for pre-Clinical research, Transfus. Med. Hemother. 43 (2016) 353–357, http://dx.doi.org/10.1159/000448254.
[76] T. Cajka, O. Fiehn, Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics, Anal. Chem. 88 (2016) 524–545, http://dx.doi.org/10.1021/acs.analchem.5b04491.
[77] S.J. Bruce, I. Tavazzi, V. Parisod, S. Rezzi, S. Kochhar, P.A. Guy, Investigation of human blood plasma sample preparation for performing metabolomics using ultrahigh performance liquid chromatography/mass spectrometry, Anal. Chem. 81 (2009) 3285–3296, http://dx.doi.org/10.1021/ac8024569.
[78] E.J. Want, G. O'Maille, C.A. Smith, T.R. Brandon, W. Uritboonthai, C. Qin, S.A. Trauger, G. Siuzdak, Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry, Anal. Chem. 78 (2006) 743–752, http://dx.doi.org/10.1021/ac051312t.
[79] S. Naz, A. García, C. Barbas, Multiplatform analytical methodology for metabolic fingerprinting of lung tissue, Anal. Chem. 85 (2013) 10941–10948, http://dx.doi.org/10.1021/ac402411n.
[80] V. Matyash, G. Liebisch, T.V. Kurzchalia, A. Shevchenko, D. Schwudke, Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics, J. Lpid Res. 49 (2008) 1137–1146, http://dx.doi.org/10.1194/jlr.D700041-JLR200.
[81] D. González-Peña, D. Dudzik, C. Colina-Coca, B. de Ancos, A. García, C. Barbas, C. Sánchez-Moreno, Evaluation of onion as a functional ingredient in the prevention of metabolic impairments associated to diet-induced hypercholesterolaemia using a multiplatform approach based on LC-MS, CE-MS and GC–MS, J. Funct. Foods 19 (2015) 363–375.
[82] P.A. Vorkas, G. Isaac, M.A. Anwar, A.H. Davies, E.J. Want, J.K. Nicholson, E. Holmes, Untargeted UPLC-MS profiling pipeline to expand tissue metabolome coverage: application to cardiovascular disease, Anal. Chem. 87 (2015) 4184–4193, http://dx.doi.org/10.1021/ac503775m.
[83] A.B. Canelas, A. ten Pierick, C. Ras, R.M. Seifar, J.C. van Dam, W.M. van Gulik, J.J. Heijnen, Quantitative evaluation of intracellular metabolite extraction techniques for yeast metabolomics, Anal. Chem. 81 (2009) 7379–7389, http://dx.doi.org/10.1021/ac900999t.
[84] M. Faijes, A.E. Mars, E.J. Smid, Comparison of quenching and extraction methodologies for metabolome analysis of Lactobacillus plantarum, Microb. Cell Fact. 6 (2007) 27, http://dx.doi.org/10.1186/1475-2859-6-27.
[85] S.G. Villas-Bôas, J. Højer-Pedersen, M. Akesson, J. Smedsgaard, J. Nielsen, Global metabolite analysis of yeast: evaluation of sample preparation methods, Yeast 22 (2005) 1155–1169, http://dx.doi.org/10.1002/yea.1308.
[86] X. Duportet, R.B.M. Aggio, S. Carneiro, S.G. Villas-Bôas, The biological interpretation of metabolomic data can be misled by the extraction method used, Metabolomics 8 (2012) 410–421, http://dx.doi.org/10.1007/s11306-011-0324-1.
[87] X. Fu, M. Iga, M. Kimura, S. Yamaguchi, Simplified screening for organic acidemia using GC/MS and dried urine filter paper: a study on neonatal mass screening, Early Hum. Dev. 58 (2000) 41–55 http://www.ncbi.nlm.nih.gov/pubmed/10785335.
[88] Z.G. Gong, J. Hu, X. Wu, Y.J. Xu, The recent developments in sample preparation for mass spectrometry-based metabolomics, Crit. Rev. Anal. Chem. (2017) 1–7, http://dx.doi.org/10.1080/10408347.2017.1289836.
[89] R.J. Raterink, P.W. Lindenburg, R.J. Vreeken, R. Ramautar, T. Hankemeier, Recent developments in sample-pretreatment techniques for mass spectrometry-based metabolomics, TrAC – Trends Anal. Chem. 61 (2014) 157–167, http://dx.doi.org/10.1016/j.trac.2014.06.003.
[90] F. Xu, L. Zou, C.N. Ong, Experiment-originated variations, and multi-peak and multi-origination phenomena in derivatization-based GC–MS metabolomics, TrAC Trends Anal. Chem. 29 (2010) 269–280, http://dx.doi.org/10.1016/j.trac.2009.12.007.
[91] ICH, Validation of Analytical Procedures: Text and Methodology Q2(R1), 1994 (2005) 17. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q2_R1/Step4/Q2_R1__Guideline.pdf.
[92] Laboratory and Scientific Section, United Nations, Guidance for the Validation of Analytical Methodology and Calibration of Equipment used for

Testing of Illicit Drugs in Seized Materials and Biological Specimens, n.d. https://www.unodc.org/documents/scientific/validation_E.pdf.
[93] A. Mastrangelo, A. Ferrarini, F. Rey-Stolle, A. García, C. Barbas, From sample treatment to biomarker discovery: a tutorial for untargeted metabolomics based on GC-(EI)-Q-MS, Anal. Chim. Acta 900 (2015) 21–35, http://dx.doi.org/10.1016/j.aca.2015.10.001.
[94] C.R. Evans, A. Karnovsky, M.A. Kovach, T.J. Standiford, C.F. Burant, K.A. Stringer, Untargeted LC-MS metabolomics of bronchoalveolar lavage fluid differentiates acute respiratory distress syndrome from health, J. Proteome Res. 13 (2014) 640–649, http://dx.doi.org/10.1021/pr4007624.
[95] K.M. Sas, A. Karnovsky, G. Michailidis, S. Pennathur, Metabolomics and diabetes: analytical and computational approaches, Diabetes 64 (2015) 718–732, http://dx.doi.org/10.2337/db14-0509.
[96] K.A. Stringer, R.T. McKay, A. Karnovsky, B. Quémerais, P. Lacy, Metabolomics and its application to acute lung diseases, Front. Immunol. 7 (2016), http://dx.doi.org/10.3389/fimmu.2016.00044.
[97] E. Zelena, W.B. Dunn, D. Broadhurst, S. Francis-McIntyre, K.M. Carroll, P. Begley, S. O'Hagan, J.D. Knowles, A. Halsall, H. Consortium, I.D. Wilson, D.B. Kell, Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum, Anal. Chem. 81 (2009) 1357–1364, http://dx.doi.org/10.1021/ac8019366.
[98] Rahul P. Chilbule, Rajendra B. Kakde, Experimental design based optimization of HPLC method for determination of blonanserin in in-vitro human serum sample, forced degraded sample and pharmaceutical formulations, Int. J. Pharm. Sci. Res. 19 (2015) 646–657, http://dx.doi.org/10.13040/IJPSR.0975-8232.8(2).646-57.
[99] L.R. Snyder, J.J. Joseph, J. Kirkland, J.L. Glajch, Practical HPLC Method Development, Wiley, 2012, http://dx.doi.org/10.1002/9781118592014.
[100] J. Pu, X. Zhao, L. Xiao, H. Zhao, Development and validation of a HILIC-ELSD method for simultaneous analysis of non-substituted and acetylated xylo-oligosaccharides, J. Pharm. Biomed. Anal. 139 (2017) 232–237, http://dx.doi.org/10.1016/j.jpba.2017.03.007.
[101] M. Lämmerhofer, W. Weckwerth, Metabolomics in Practice: Successful Strategies to Generate and Analyze Metabolic Data, Wiley-VCH, 2013.
[102] H. Pereira, J.F. Martin, C. Joly, J.L. Sébédio, E. Pujos-Guillot, Development and validation of a UPLC/MS method for a nutritional metabolomic study of human plasma, Metabolomics 6 (2010) 207–218, http://dx.doi.org/10.1007/s11306-009-0188-9.
[103] M. Rusilowicz, M. Dickinson, A. Charlton, S. O'Keefe, J. Wilson, A batch correction method for liquid chromatography-mass spectrometry data that does not depend on quality control samples, Metabolomics 12 (2016) 1–11, http://dx.doi.org/10.1007/s11306-016-0972-2.
[104] E. Chambers, D.M. Wagrowski-Diehl, Z. Lu, J.R. Mazzeo, Systematic and comprehensive strategy for reducing matrix effects in LC/MS/MS analyses, J. Chromatogr. B 852 (2007) 22–34, http://dx.doi.org/10.1016/j.jchromb.2006.12.030.
[105] J.C. Van De Steene, W.E. Lambert, Comparison of matrix effects in HPLC-MS/MS and UPLC-MS/MS analysis of nine basic pharmaceuticals in surface waters, J. Am. Soc. Mass Spectrom. 19 (2008) 713–718, http://dx.doi.org/10.1016/j.jasms.2008.01.013.
[106] A.J. Girón, K. Deventer, K. Roels, P. Van Eenoo, Development and validation of an open screening method for diuretics, stimulants and selected compounds in human urine by UHPLC-HRMS for doping control, Anal. Chim. Acta 721 (2012) 137–146, http://dx.doi.org/10.1016/j.aca.2012.02.002.
[107] D.N. Heller, Ruggedness testing of quantitative atmospheric pressure ionization mass spectrometry methods: the effect of co-injected matrix on matrix effects, Rapid Commun. Mass Spectrom. 21 (2007) 644–652, http://dx.doi.org/10.1002/rcm.2882.
[108] P.J. Taylor, Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography-electrospray-tandem mass spectrometry, Clin. Biochem. 38 (2005) 328–334, http://dx.doi.org/10.1016/j.clinbiochem.2004.11.007.
[109] S.G. Villas-Bôas, K.F. Smart, S. Sivakumaran, G.A. Lane, Alkylation or silylation for analysis of amino and non-amino organic acids by GC–MS? Metabolites 1 (2011) 3–20, http://dx.doi.org/10.3390/metabo1010003.
[110] G. Mitulović, C. Stingl, I. Steinmacher, O. Hudecz, J.R.A. Hutchins, J.M. Peters, K. Mechtler, Preventing carryover of peptides and proteins in nano LC-MS separations, Anal. Chem. 81 (2009) 5955–5960, http://dx.doi.org/10.1021/ac900696m.
[111] N.C. Hughes, E.Y.K. Wong, J. Fan, N. Bajaj, Determination of carryover and contamination for mass spectrometry-based chromatographic assays, AAPS J. 9 (2007) E353–E360, http://dx.doi.org/10.1208/aapsj0903042.
[112] W. Zeng, D.G. Musson, A.L. Fisher, A.Q. Wang, A new approach for evaluating carryover and its influence on quantitation in high-performance liquid chromatography and tandem mass spectrometry assay, Rapid Commun. Mass Spectrom. 20 (2006) 635–640, http://dx.doi.org/10.1002/rcm.2353.
[113] A. Alonso, S. Marsal, A. Julià, Analytical methods in untargeted metabolomics: state of the artin 2015, Front. Bioeng. Biotechnol. 3 (2015) 23, http://dx.doi.org/10.3389/fbioe.2015.00023.
[114] S. Castillo, P. Gopalacharyulu, L. Yetukuri, M. Orešič, Algorithms and tools for the preprocessing of LC-MS metabolomics data, Chemom. Intell. Lab. Syst. 108 (2011) 23–32, http://dx.doi.org/10.1016/j.chemolab.2011.03.010.
[115] S. Cheng, S.H. Shah, E.J. Corwin, O. Fiehn, R.L. Fitzgerald, R.E. Gerszten, T. Illig, E.P. Rhee, P.R. Srinivas, T.J. Wang, M. Jain, American Heart Association Council on Functional Genomics and Translational Biology; Council on Cardiovascular and Stroke Nursing; Council on Clinical Cardiology; and Stroke Council, Potential Impact and Study Considerations of Metabolomics in Cardiovascular Health and Disease: A Scientific Statement From the American Heart Association, Circ. Cardiovasc. Genet. 10 (2017) e000032, http://dx.doi.org/10.1161/HCG.0000000000000032.
[116] J.A. Kirwan, R.J.M. Weber, D.I. Broadhurst, M.R. Viant, Direct infusion mass spectrometry metabolomics dataset: a benchmark for data processing and quality control, Sci. Data 1 (2014) 140012, http://dx.doi.org/10.1038/sdata.2014.12.
[117] D.J. Creek, A. Jankevics, K.E.V. Burgess, R. Breitling, M.P. Barrett, IDEOM: an Excel interface for analysis of LC-MS-based metabolomics data, Bioinformatics 28 (2012) 1048–1049, http://dx.doi.org/10.1093/bioinformatics/bts069.
[118] R. Scheltema, S. Decuypere, J. Dujardin, D. Watson, R. Jansen, R. Breitling, Simple data-reduction method for high-resolution LC-MS data in metabolomics, Bioanalysis 1 (2009) 1551–1557, http://dx.doi.org/10.4155/bio.09.146.
[119] R. Tautenhahn, C. Bottcher, S. Neumann, Highly sensitive feature detection for high resolution LC/MS, BMC Bioinf. 9 (2008) 504, http://dx.doi.org/10.1186/1471-2105-9-504.
[120] H. Treutler, S. Neumann, Prediction, detection, and validation of isotope clusters in mass spectrometry data, Metabolites 6 (2016) 37, http://dx.doi.org/10.3390/metabo6040037.
[121] E. Lange, R. Tautenhahn, S. Neumann, C. Gröpl, Critical assessment of alignment procedures for LC-MS proteomics and metabolomics measurements, BMC Bioinf. 9 (2008) 375, http://dx.doi.org/10.1186/1471-2105-9-375.
[122] Y. Koh, K.K. Pasikanti, C.W. Yap, E.C.Y. Chan, Comparative evaluation of software for retention time alignment of gas chromatography/time-of-flight mass spectrometry-based metabonomic data, J. Chromatogr. A 1217 (2010) 8308–8316, http://dx.doi.org/10.1016/j.chroma.2010.10.101.
[123] J.D. Watrous, M. Henglin, B. Claggett, K. Lehmann, M.G. Larson, S. Cheng, M. Jain, Visualization, quantification and alignment of spectral drift in population scale untargeted metabolomics data, Anal. Chem. 89 (2017) 1399–1404, http://dx.doi.org/10.1021/acs.analchem.6b04337.
[124] A. Garcia, S. Naz, C. Barbas, Metabolite fingerprinting by capillary electrophoresis-mass spectrometry, Methods Mol. Biol. (2014) 107–123, http://dx.doi.org/10.1007/978-1-4939-1258-2_8.
[125] Douglas C. Montgomery (Ed.), Design and Analysis of Experiments, 8th edition, Wiley, 2012 http://eu.wiley.com/WileyCDA/WileyTitle/productCd-EHEP002024.html.
[126] R. Goodacre, D. Broadhurst, A.K. Smilde, B.S. Kristal, J.D. Baker, R. Beger, C. Bessant, S. Connor, G. Capuani, A. Craig, T. Ebbels, D.B. Kell, C. Manetti, J. Newton, G. Paternostro, R. Somorjai, M. Sjöström, J. Trygg, F. Wulfert, Proposed minimum reporting standards for data analysis in metabolomics, Metabolomics 3 (2007) 231–241, http://dx.doi.org/10.1007/s11306-007-0081-3.
[127] R. Chaleckis, I. Murakami, J. Takada, H. Kondoh, M. Yanagida, Individual variability in human blood metabolites identifies age-related differences, Proc. Natl. Acad. Sci. U. S. A. 113 (2016) 4252–4259, http://dx.doi.org/10.1073/pnas.1603023113.
[128] E.A. Thévenot, A. Roux, Y. Xu, E. Ezan, C. Junot, Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses, J. Proteome Res. 14 (2015) 3322–3335, http://dx.doi.org/10.1021/acs.jproteome.5b00354.
[129] Z. Yu, G. Zhai, P. Singmann, Y. He, T. Xu, C. Prehn, W. Römisch-Margl, E. Lattka, C. Gieger, N. Soranzo, J. Heinrich, M. Standl, E. Thiering, K. Mittelstraß, H.-E. Wichmann, A. Peters, K. Suhre, Y. Li, J. Adamski, T.D. Spector, T. Illig, R. Wang-Sattler, Human serum metabolic profiles are age dependent, Aging Cell. 11 (2012) 960–967, http://dx.doi.org/10.1111/j.1474-9726.2012.00865.x.
[130] A. Mastrangelo, G. Martos-Moreno, A. García, V. Barrios, F. Rupérez, J. Chowen, C. Barbas, J. Argente, Insulin resistance in prepubertal obese children correlates with sex-dependent early onset metabolomic alterations, Int. J. Obes. 40 (2016) 1494–1502, http://dx.doi.org/10.1038/ijo.2016.92.
[131] M.K. Townsend, C.B. Clish, P. Kraft, C. Wu, A.L. Souza, A.A. Deik, S.S. Tworoger, B.M. Wolpin, Reproducibility of metabolomic profiles among men and women in 2 large cohort studies, Clin. Chem. 59 (2013) 1657–1667, http://dx.doi.org/10.1373/clinchem.2012.199133.
[133] L.K. Reed, C.F. Baer, A.S. Edison, Considerations when choosing a genetic model organism for metabolomics studies, Curr. Opin. Chem. Biol. 36 (2017) 7–14, http://dx.doi.org/10.1016/j.cbpa.2016.12.005.
[134] O. Fiehn, Metabolomics-the link between genotypes and phenotypes, Plant Mol. Biol. 48 (2002) 155–171 http://www.ncbi.nlm.nih.gov/pubmed/11860207.
[135] V.G. Marrachelli, P. Rentero, M.L. Mansego, J.M. Morales, I. Galan, M. Pardo-Tendero, F. Martinez, J.C. Martin-Escudero, L. Briongos, F.J. Chaves, J. Redon, D. Monleon, Genomic and metabolomic profile associated to clustering of cardio-metabolic risk factors, PLoS One. 11 (2016) e0160656, http://dx.doi.org/10.1371/journal.pone.0160656.
[136] J.L. Sébédio, Metabolomics, nutrition, and potential biomarkers of food quality, intake, and health status, Adv. Food Nutr. Res. 82 (2017) 83–116, http://dx.doi.org/10.1016/bs.afnr.2017.01.001.
[137] J.B. German, B.D. Hammock, S.M. Watkins, Metabolomics: building on a century of biochemistry to guide human health, Metabolomics 1 (2005) 3–9, http://dx.doi.org/10.1007/s11306-005-1102-8.

[138] A. Rietman, T.L. Stanley, C. Clish, V. Mootha, M. Mensink, S.K. Grinspoon, H. Makimura, Associations between plasma branched-chain amino acids, β-aminoisobutyric acid and body composition, J. Nutr. Sci. 5 (2016) e6, http://dx.doi.org/10.1017/jns.2015.37.
[139] J. Brignardello, E. Holmes, I. Garcia-Perez, Metabolic phenotyping of diet and dietary intake, Adv. Food Nutr. Res. 81 (2017) 231–270, http://dx.doi.org/10.1016/bs.afnr.2016.12.002.
[140] G.F. Giskeødegård, S.K. Davies, V.L. Revell, H. Keun, D.J. Skene, Diurnal rhythms in the human urine metabolome during sleep and total sleep deprivation, Sci. Rep. 5 (2015) 14843, http://dx.doi.org/10.1038/srep14843.
[141] J.J. Gooley, E.C. Chua, Diurnal regulation of lipid metabolism and applications of circadian lipidomics, J. Genet. Genomics. 41 (2014) 231–250, http://dx.doi.org/10.1016/j.jgg.2014.04.001.
[142] S.K. Davies, J.E. Ang, V.L. Revell, B. Holmes, A. Mann, F.P. Robertson, N. Cui, B. Middleton, K. Ackermann, M. Kayser, A.E. Thumser, F.I. Raynaud, D.J. Skene, Effect of sleep deprivation on the human metabolome, Proc. Natl. Acad. Sci. 111 (2014) 10761–10766, http://dx.doi.org/10.1073/pnas.1402663111.
[143] J.E. Ang, V. Revell, A. Mann, S. Mäntele, D.T. Otway, J.D. Johnston, A.E. Thumser, D.J. Skene, F. Raynaud, Identification of human plasma metabolites exhibiting time-of-day variation using an untargeted liquid chromatography-mass spectrometry metabolomic approach, Chronobiol. Int. 29 (2012) 868–881, http://dx.doi.org/10.3109/07420528.2012.699122.
[144] R. Dallmann, A.U. Viola, L. Tarokh, C. Cajochen, S.A. Brown, The human circadian metabolome, Proc. Natl. Acad. Sci. 109 (2012) 2625–2629, http://dx.doi.org/10.1073/pnas.1114410109.
[145] C. Hellmuth, K.L. Lindsay, O. Uhl, C. Buss, P.D. Wadhwa, B. Koletzko, S. Entringer, Association of maternal prepregnancy BMI with metabolomic profile across gestation, Int. J. Obes. 41 (2017) 159–169, http://dx.doi.org/10.1038/ijo.2016.153.
[146] S.C. Moore, C.E. Matthews, J.N. Sampson, R.Z. Stolzenberg-Solomon, W. Zheng, Q. Cai, Y.T. Tan, W.H. Chow, B.T. Ji, D.K. Liu, Q. Xiao, S.M. Boca, M.F. Leitzmann, G. Yang, Y.B. Xiang, R. Sinha, X.O. Shu, A.J. Cross, Human metabolic correlates of body mass index, Metabolomics 10 (2014) 259–269, http://dx.doi.org/10.1007/s11306-013-0574-1.
[147] R. Kleemann, L. Verschuren, M.J. van Erk, Y. Nikolsky, N.H. Cnubben, E.R. Verheij, A.K. Smilde, H.F. Hendriks, S. Zadelaar, G.J. Smith, V. Kaznacheev, T. Nikolskaya, A. Melnikov, E. Hurt-Camejo, J. van der Greef, B. van Ommen, T. Kooistra, Atherosclerosis and liver inflammation induced by increased dietary cholesterol intake: a combined transcriptomics and metabolomics analysis, Genome Biol. 8 (2007) R200, http://dx.doi.org/10.1186/gb-2007-8-9-r200.
[148] H. Gibbons, E. Carr, B.A. McNulty, A.P. Nugent, J. Walton, A. Flynn, M.J. Gibney, L. Brennan, Metabolomic based identification of clusters that reflect dietary patterns, Mol. Nutr. Food Res. (2017) 1601050, http://dx.doi.org/10.1002/mnfr.201601050.
[149] M.C. Playdon, J.N. Sampson, A.J. Cross, R. Sinha, K.A. Guertin, K.A. Moy, N. Rothman, M.L. Irwin, S.T. Mayne, R. Stolzenberg-Solomon, S.C. Moore, Comparing metabolite profiles of habitual diet in serum and urine, Am. J. Clin. Nutr. 104 (2016) 776–789, http://dx.doi.org/10.3945/ajcn.116.135301.
[150] T.A. Clayton, J.C. Lindon, O. Cloarec, H. Antti, C. Charuel, G. Hanton, J.P. Provost, J.L. Le Net, D. Baker, R.J. Walley, J.R. Everett, J.K. Nicholson, Pharmaco-metabonomic phenotyping and personalized drug treatment, Nature 440 (2006) 1073–1077, http://dx.doi.org/10.1038/nature04648.
[151] Y. Ni, M. Su, J. Lin, X. Wang, Y. Qiu, A. Zhao, T. Chen, W. Jia, Metabolic profiling reveals disorder of amino acid metabolism in four brain regions from a rat model of chronic unpredictable mild stress, FEBS Lett. 582 (2008) 2627–2636, http://dx.doi.org/10.1016/j.febslet.2008.06.040.
[152] S. Prabakaran, J.E. Swatton, M.M. Ryan, S.J. Huffaker, J.T. Huang, J.L. Griffin, M. Wayland, T. Freeman, F. Dudbridge, K.S. Lilley, N.A. Karp, S. Hester, D. Tkachev, M.L. Mimmack, R.H. Yolken, M.J. Webster, E.F. Torrey, S. Bahn, Mitochondrial dysfunction in Schizophrenia: evidence for compromised brain metabolism and oxidative stress, Mol. Psychiatry 9 (2004) 684–697, http://dx.doi.org/10.1038/sj.mp.4001511, 643.
[153] O. Deda, H. Gika, T. Panagoulis, I. Taitzoglou, N. Raikos, G. Theodoridis, Impact of exercise on fecal and cecal metabolome over aging: a longitudinal study in rats, Bioanalysis 9 (2017) 21–36, http://dx.doi.org/10.4155/bio-2016-0222.
[154] A. Siopi, O. Deda, V. Manou, S. Kellis, I. Kosmidis, D. Komninou, N. Raikos, K. Christoulas, G. Theodoridis, V. Mougios, Effects of different exercise modes on the urinary metabolic fingerprint of men with and without metabolic syndrome, Metabolites 7 (2017) 5, http://dx.doi.org/10.3390/metabo7010005.
[155] K.B. Martinez, V. Leone, E.B. Chang, Microbial metabolites in health and disease: navigating the unknown in search of function, J. Biol. Chem. 292 (2017) 8553–8559, http://dx.doi.org/10.1074/jbc.R116.752899.
[156] P. Vernocchi, F. Del Chierico, L. Putignani, Gut microbiota profiling: metabolomics based approach to unravel compounds affecting human health, Front. Microbiol. 7 (2016) 1144, http://dx.doi.org/10.3389/fmicb.2016.01144.
[157] R. Wang-Sattler, Y. Yu, K. Mittelstrass, E. Lattka, E. Altmaier, C. Gieger, K.H. Ladwig, N. Dahmen, K.M. Weinberger, P. Hao, L. Liu, Y. Li, H.-E. Wichmann, J. Adamski, K. Suhre, T. Illig, Metabolic profiling reveals distinct variations linked to nicotine consumption in humans-first results from the KORA study, PLoS One 3 (2008) e3863, http://dx.doi.org/10.1371/journal.pone.0003863.
[158] U.E. Rolle-Kampczyk, J. Krumsiek, W. Otto, S.W. Röder, T. Kohajda, M. Borte, F. Theis, I. Lehmann, M. von Bergen, Metabolomics reveals effects of maternal smoking on endogenous metabolites from lipid metabolism in cord blood of newborns, Metabolomics 12 (2016) 76, http://dx.doi.org/10.1007/s11306-016-0983-z.
[159] P. Würtz, S. Cook, Q. Wang, M. Tiainen, T. Tynkkynen, A.J. Kangas, P. Soininen, J. Laitinen, J. Viikari, M. Kähönen, T. Lehtimäki, M. Perola, S. Blankenberg, T. Zeller, S. Männistö, V. Salomaa, M.R. Järvelin, O.T. Raitakari, M. Ala-Korpela, D.A. Leon, Metabolic profiling of alcohol consumption in 9778 young adults, Int. J. Epidemiol. 45 (2016) 1493–1506, http://dx.doi.org/10.1093/ije/dyw175.
[160] M.C. Walsh, G.A. McLoughlin, H.M. Roche, J.F. Ferguson, C.A. Drevon, W.H.M. Saris, J.A. Lovegrove, U. Risérus, J. López-Miranda, C. Defoort, B. Kieć-Wilk, L. Brennan, M.J. Gibney, Impact of geographical region on urinary metabolomic and plasma fatty acid profiles in subjects with the metabolic syndrome across Europe: the LIPGENE study, Br. J. Nutr. 111 (2014) 424–431, http://dx.doi.org/10.1017/S0007114513002602.
[161] K.J. Nahon, M.R. Boon, L.E.H. Bakker, C. Prehn, J. Adamski, I.M. Jazet, K.W. van Dijk, P.C.N. Rensen, D.O. Mook-Kanamori, Physiological changes due to mild cooling in healthy lean males of white Caucasian and South Asian descent: a metabolomics study, Arch. Biochem. Biophys. 589 (2016) 152–157, http://dx.doi.org/10.1016/j.abb.2015.09.001.
[162] P. Würtz, Q. Wang, A.J. Kangas, R.C. Richmond, J. Skarp, M. Tiainen, T. Tynkkynen, P. Soininen, A.S. Havulinna, M. Kaakinen, J.S. Viikari, M.J. Savolainen, M. Kähönen, T. Lehtimäki, S. Männistö, S. Blankenberg, T. Zeller, J. Laitinen, A. Pouta, P. Mäntyselkä, M. Vanhala, P. Elliott, K.H. Pietiläinen, S. Ripatti, V. Salomaa, O.T. Raitakari, M.R. Järvelin, G.D. Smith, M. Ala-Korpela, Metabolic signatures of adiposity in young adults: mendelian randomization analysis and effects of weight change, PLoS Med. 11 (2014) e1001765, http://dx.doi.org/10.1371/journal.pmed.1001765.
[163] S. Burgess, R.M. Daniel, A.S. Butterworth, S.G. Thompson, EPIC-InterAct Consortium, Network Mendelian randomization: using genetic variants as instrumental variables to investigate mediation in causal pathways, Int. J. Epidemiol. 44 (2015) 484–495, http://dx.doi.org/10.1093/ije/dyu176.
[164] M.C. Borges, A.J. Barros, D.L.S. Ferreira, J.P. Casas, B.L. Horta, M. Kivimaki, M. Kumari, U. Menon, T.R. Gaunt, Y. Ben-Shlomo, D.F. Freitas, I.O. Oliveira, A. Gentry-Maharaj, E. Fourkala, D.A. Lawlor, A.D. Hingorani, Metabolic profiling of adiponectin levels in adults: mendelian randomization analysis, bioRxiv (2017), http://dx.doi.org/10.1101/126789.
[165] S. Barnes, H.P. Benton, K. Casazza, S.J. Cooper, X. Cui, X. Du, J. Engler, J.H. Kabarowski, S. Li, W. Pathmasiri, J.K. Prasain, M.B. Renfrow, H.K. Tiwari, Training in metabolomics research. I. Designing the experiment, collecting and extracting samples and generating metabolomics data, J. Mass Spectrom. 51 (461) (2016) 475, http://dx.doi.org/10.1002/jms.3782.
[166] W.B. Dunn, D.I. Broadhurst, H.J. Atherton, R. Goodacre, J.L. Griffin, Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy, Chem. Soc. Rev. 40 (2011) 387–426, http://dx.doi.org/10.1039/b906712b.
[167] M. Krzywinski, N. Altman, Points of significance: power and sample size, Nat. Methods. 10 (2013) 1139–1140, http://dx.doi.org/10.1038/nmeth.2738.
[168] G. Nyamundanda, I. 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, http://dx.doi.org/10.1186/1471-2105-14-338.
[169] E. Billoir, V. Navratil, B.J. Blaise, Sample size calculation in metabolic phenotyping studies, Brief. Bioinform. 16 (2014) 813–819, http://dx.doi.org/10.1093/bib/bbu052.
[170] 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, http://dx.doi.org/10.1021/acs.analchem.6b00188.
[171] USP 37 Chapter 1058, < 1058 > Analytical Instrument Qualification, 2012 747–751.
[172] A. Bose, HPLC calibration process parameters in terms of system suitability test, Austin Chromatogr. 1 (2014) 4, http://dx.doi.org/10.1080/01483919108049375.
[173] E.J. Want, I.D. Wilson, H. Gika, G. Theodoridis, R.S. Plumb, J. Shockcor, E. Holmes, J.K. Nicholson, Global metabolic profiling procedures for urine using UPLC −MS, Nat. Protoc. 5 (2010) 1005–1018, http://dx.doi.org/10.1038/nprot.2010.50.
[174] T. Sangster, H. Major, R. Plumb, A.J. Wilson, I.D. Wilson, A pragmatic and readily implemented quality control strategy for HPLC-MS and GC–MS-based metabonomic analysis, Analyst 131 (2006) 1075–1078, http://dx.doi.org/10.1039/b604498k.
[175] M. Dong, R. Paul, L. Gershanov, Getting the peaks perfect: System suitability for HPLC, Todayís Chem. Work. (n.d.). https://pubs.acs.org/subscribe/archive/tcaw/10/i09/html/09dong.html.
[176] S. Naz, A. Garcia, M. Rusak, C. Barbas, Method development and validation for rat serum fingerprinting with CE-MS: Application to ventilator-induced-lung-injury study, Anal. Bioanal. Chem. 405 (2013) 4849–4858, http://dx.doi.org/10.1007/s00216-013-6882-5.
[177] EMA, Guideline on Bioanalytical Method Validation, vol. 44, 2012, pp. 1–23, EMEA/CHMP/EWP/192217/2009.
[178] A. Kruve, R. Rebane, K. Kipper, M.L. Oldekop, H. Evard, K. Herodes, P. Ravio, I. Leito, Tutorial review on validation of liquid chromatography-mass spectrometry methods: part II, Anal. Chim. Acta 870 (2015) 8–28, http://dx.doi.org/10.1016/j.aca.2015.02.016.

[179] M.M. Moein, A. El Beqqali, M. Abdel-Rehim, Bioanalytical method development and validation: critical concepts and strategies, J. Chromatogr. B 1043 (2017) 3–11, http://dx.doi.org/10.1016/j.jchromb.2016.09.028.
[180] P.A. Guy, I. Tavazzi, S.J. Bruce, Z. Ramadan, S. Kochhar, Global metabolic profiling analysis on human urine by UPLC-TOFMS: issues and method validation in nutritional metabolomics, J. Chromatogr. B 871 (2008) 253–260, http://dx.doi.org/10.1016/j.jchromb.2008.04.034.
[181] L. Venter, P.J. van Rensburg, D.T. Loots, A. Vosloo, J.Z. Lindeque, Untargeted metabolite profiling of abalone using gas chromatography mass spectrometry, Food Anal. Methods. 9 (2016) 1254–1261, http://dx.doi.org/10.1007/s12161-015-0285-5.
[182] J. Kuligowski, I. Torres-Cuevas, G. Quintás, D. Rook, J.B. Van Goudoever, E. Cubells, M. Asensi, I. Lliso, A. Nuñez, M. Vento, J. Escobar, Assessment of oxidative damage to proteins and DNA in urine of newborn infants by a validated UPLC-MS/MS approach, PLoS One 9 (2014), http://dx.doi.org/10.1371/journal.pone.0093703.
[183] L. Whiley, J. Godzien, F.J. Ruperez, C. Legido-Quigley, C. Barbas, In-vial dual extraction for direct LC-MS analysis of plasma for comprehensive and highly reproducible metabolic fingerprinting, Anal. Chem. 84 (2012) 5992–5999.
[184] K. Sandra, S. Pereira Ados, G. Vanhoenacker, F. David, P. Sandra, Comprehensive blood plasma lipidomics by liquid chromatography/quadrupole time-of-flight mass spectrometry, J. Chromatogr. A 1217 (2010) 4087–4099, http://dx.doi.org/10.1016/j.chroma.2010.02.039.
[185] A.P. Siskos, P. Jain, W. Römisch-Margl, M. Bennett, D. Achaintre, Y. Asad, L. Marney, L. Richardson, A. Koulman, J.L. Griffin, F. Raynaud, A. Scalbert, J. Adamski, C. Prehn, H.C. Keun, Interlaboratory reproducibility of a targeted metabolomics platform for analysis of human serum and plasma, Anal. Chem. 89 (2017) 656–665, http://dx.doi.org/10.1021/acs.analchem.6b02930.
[186] H.G. Gika, G.A. Theodoridis, J.E. Wingate, I.D. Wilson, Within-day reproducibility of an HPLC-MS-based method for metabonomic analysis: application to human urine, J. Proteome Res. 6 (2007) 3291–3303, http://dx.doi.org/10.1021/pr070183p.
[187] M. Eliasson, S. Rännar, R. Madsen, M.A. Donten, E. Marsden-Edwards, T. Moritz, J.P. Shockcor, E. Johansson, J. Trygg, Strategy for optimizing LC-MS data processing in metabolomics: a design of experiments approach, Anal. Chem. 84 (2012) 6869–6876, http://dx.doi.org/10.1021/ac301482k.
[188] I. Gertsman, J.A. Gangoiti, B.A. Barshop, Validation of a dual LC-HRMS platform for clinical metabolic diagnosis in serum, bridging quantitative analysis and untargeted metabolomics, Metabolomics 10 (2014) 312–323, http://dx.doi.org/10.1007/s11306-013-0582-1.
[189] W.B. Dunn, I.D. Wilson, A.W. Nicholls, D. Broadhurst, The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans, Bioanalysis 4 (2012) 2249–2264, http://dx.doi.org/10.4155/bio.12.204.
[190] P. Jonsson, A. Wuolikainen, E. Thysell, E. Chorell, P. Stattin, P. Wikström, H. Antti, Constrained randomization and multivariate effect projections improve information extraction and biomarker pattern discovery in metabolomics studies involving dependent samples, Metabolomics 11 (2015) 1667–1678, http://dx.doi.org/10.1007/s11306-015-0818-3.
[191] H.G. Gika, C. Zisi, G. Theodoridis, I.D. Wilson, Protocol for quality control in metabolic profiling of biological fluids by U(H)PLC-MS, J. Chromatogr. B 1008 (2016) 15–25, http://dx.doi.org/10.1016/j.jchromb.2015.10.045.
[192] P. Blainey, M. Krzywinski, N. Altman, Points of significance: replication, Nat. Methods 11 (2014) 879–880, http://dx.doi.org/10.1038/nmeth.3091.
[193] L.D. Roberts, A.L. Souza, R.E. Gerszten, C.B. Clish, Targeted metabolomics, Curr. Protoc. Mol. Biol. 30 (2012) 1–24, http://dx.doi.org/10.1002/0471142727.mb3002s98.
[194] T. Bader, W. Schulz, K. Kümmerer, R. Winzenbacher, General strategies to increase the repeatability in non-target screening by liquid chromatography-high resolution mass spectrometry, Anal. Chim. Acta 935 (2016) 173–186, http://dx.doi.org/10.1016/j.aca.2016.06.030.
[195] W.B. Dunn, D.I. Broadhurst, A. Edison, C. Guillou, M.R. Viant, D.W. Bearden, R.D. Beger, Quality assurance and quality control processes: summary of a metabolomics community questionnaire, Metabolomics 13 (2017) 50, http://dx.doi.org/10.1007/s11306-017-1188-9.
[196] J. Kuligowski, D. Pérez-Guaita, I. Lliso, J. Escobar, Z. León, L. Gombau, R. Solberg, O.D. Saugstad, M. Vento, G. Quintás, Detection of batch effects in liquid chromatography-mass spectrometry metabolomic data using guided principal component analysis, Talanta 130 (2014) 442–448, http://dx.doi.org/10.1016/j.talanta.2014.07.031.
[197] O. Yanes, R. Tautenhahn, G.J. Patti, G. Siuzdak, Expanding coverage of the metabolome for global metabolite profiling, Anal. Chem. 83 (2011) 2152–2161, http://dx.doi.org/10.1021/ac102981k.
[198] 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, http://dx.doi.org/10.1007/s11306-016-1081-y.
[199] S. Beisken, M. Eiden, R.M. Salek, Getting the right answers: understanding metabolomics challenges, Expert Rev. Mol. Diagn. 15 (2015) 97–109, http://dx.doi.org/10.1586/14737159.2015.974562.
[200] B.A. Goldberger, M.A. Huestis, D.G. Wilkins, Commonly practiced quality control and quality assurance procedures for gas chromatography-mass spectrometry analysis in forensic urine drug-testing laboratories, Forensic Sci. Rev. 9 (1997) 59–80 http://www.ncbi.nlm.nih.gov/pubmed/26269941.
[201] J. van der Greef, S. Martin, P. Juhasz, A. Adourian, T. Plasterer, E.R. Verheij, R.N. McBurney, The art and practice of systems biology in medicine: mapping patterns of relationships, J. Proteome Res. 6 (2007) 1540–1559, http://dx.doi.org/10.1021/pr0606530.
[202] X. Liu, P. Zheng, X. Zhao, Y. Zhang, C. Hu, J. Li, J. Zhao, J. Zhou, P. Xie, G. Xu, Discovery and validation of plasma biomarkers for major depressive disorder classification based on liquid chromatography-mass spectrometry, J. Proteome Res. 14 (2015) 2322–2330, http://dx.doi.org/10.1021/acs.jproteome.5b00144.
[203] R. Bujak, J. Mateo, I. Blanco, J.L. Izquierdo-García, D. Dudzik, M.J. Markuszewski, V.I. Peinado, M. Laclaustra, J.A. Barberá, C. Barbas, J. Ruiz-Cabello, New biochemical insights into the mechanisms of pulmonary arterial hypertension in humans, PLoS One 11 (2016) e0160505, http://dx.doi.org/10.1371/journal.pone.0160505.
[204] C. Barbas-Bernardos, E.G. Armitage, A. García, S. Mérida, A. Navea, F. Bosch-Morell, C. Barbas, Looking into aqueous humor through metabolomics spectacles – exploring its metabolic characteristics in relation to myopia, J. Pharm. Biomed. Anal. 127 (2016) 18–25, http://dx.doi.org/10.1016/j.jpba.2016.03.032.
[205] J. Godzien, V. Alonso-Herranz, C. Barbas, G. Armitage, E.G. Armitage, Controlling the quality of metabolomics data: new strategies to get the best out of the QC sample, Metabolomics (2015) 518–528, http://dx.doi.org/10.1007/s11306-014-0712-4.
[206] H.G. Gika, G.A. Theodoridis, M. Earll, I.D. Wilson, A QC approach to the determination of day-to-day reproducibility and robustness of LC-MS methods for global metabolite profiling in metabonomics/metabolomics, Bioanalysis 4 (2012) 2239–2247, http://dx.doi.org/10.4155/bio.12.212.
[207] M.A. Kamleh, T.M.D. Ebbels, K. Spagou, P. Masson, E.J. Want, Optimizing the use of quality control samples for signal drift correction in large-scale urine metabolic profiling studies, Anal. Chem. 84 (2012) 2670–2677, http://dx.doi.org/10.1021/ac202733q.
[208] F.M. van der Kloet, I. Bobeldijk, E.R. Verheij, R.H. Jellema, Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping, J. Proteome Res. 8 (2009) 5132–5141, http://dx.doi.org/10.1021/pr900499r.
[209] J. Kuligowski, Á. Sánchez-Illana, D. Sanjuán-Herráez, M. Vento, G. Quintás, Intra-batch effect correction in liquid chromatography-mass spectrometry using quality control samples and support vector regression (QC-SVRC), Analyst 140 (2015) 7810–7817, http://dx.doi.org/10.1039/C5AN01638J.
[210] B. Zhou, J.F. Xiao, L. Tuli, H.W. Ressom, LC-MS-based metabolomics, Mol. Biosyst. 8 (2012) 470–481, http://dx.doi.org/10.1039/c1mb05350g.
[211] A. Kiełbasa, R. Gadzała-Kopciuch, B. Buszewski, Reference materials: significance, general requirements, and demand, Crit. Rev. Anal. Chem. 46 (2016) 224–235, http://dx.doi.org/10.1080/10408347.2015.1045120.
[212] A. Leinenbach, J. Pannee, T. Dülffer, A. Huber, T. Bittner, U. Andreasson, J. Gobom, H. Zetterberg, U. Kobold, E. Portelius, K. Blennow, IFCC Scientific Division Working Group on CSF proteins, Mass spectrometry-based candidate reference measurement procedure for quantification of amyloid-β in cerebrospinal fluid, Clin. Chem. 60 (Jul (7)) (2014) 987–994, http://dx.doi.org/10.1373/clinchem.2013.220392.
[213] K.W. Phinney, G. Ballihaut, M. Bedner, B.S. Benford, J.E. Camara, S.J. Christopher, W.C. Davis, N.G. Dodder, G. Eppe, B.E. Lang, S.E. Long, M.S. Lowenthal, E.A. McGaw, K.E. Murphy, B.C. Nelson, J.L. Prendergast, J.L. Reiner, C.A. Rimmer, L.C. Sander, M.M. Schantz, K.E. Sharpless, L.T. Sniegoski, S.S. Tai, J.B. Thomas, T.W. Vetter, M.J. Welch, S.A. Wise, L.J. Wood, W.F. Guthrie, C.R. Hagwood, S.D. Leigh, J.H. Yen, N.F. Zhang, M. Chaudhary-Webb, H. Chen, Z. Fazili, D.J. LaVoie, L.F. McCoy, S.S. Momin, N. Paladugula, E.C. Pendergrast, C.M. Pfeiffer, C.D. Powers, D. Rabinowitz, M.E. Rybak, R.L. Schleicher, B.M.H. Toombs, M. Xu, M. Zhang, A.L. Castle, Development of a Standard Reference Material for metabolomics research, Anal. Chem. 85 (2013) 11732–11738, http://dx.doi.org/10.1021/ac402689t.
[214] A.M.H. van der Veen, T.P.J. Linsinger, A. Lamberty, J. Pauwels, Uncertainty calculations in the certification of reference materials 3. Stability study, Accred Qual Assur. 6 (2001) 257–263 http://files.instrument.com.cn/bbs/upfile/200851735314.pdf.
[215] H.G. Gika, E. Macpherson, G.A. Theodoridis, I.D. Wilson, Evaluation of the repeatability of ultra-performance liquid chromatography-TOF-MS for global metabolic profiling of human urine samples, J. Chromatogr. B 871 (2008) 299–305, http://dx.doi.org/10.1016/j.jchromb.2008.05.048.
[216] J. Peng, Y.T. Chen, C.L. Chen, L. Li, Development of a universal metabolome-standard method for long-term LC-MS metabolome profiling and its application for bladder cancer urine-metabolite-biomarker discovery, Anal. Chem. 86 (2014) 6540–6547, http://dx.doi.org/10.1021/ac5011684.
[217] M. Schwaiger, E. Rampler, G. Hermann, W. Miklos, W. Berger, G. Koellensperger, Anion-exchange chromatography coupled to high resolution mass spectrometry: a powerful tool for merging targeted and non-targeted metabolomics, Anal. Chem. (2017), http://dx.doi.org/10.1021/acs.analchem.7b01624.
[218] W.B. Dunn, W. Lin, D. Broadhurst, P. Begley, M. Brown, E. Zelena, A.A. Vaughan, A. Halsall, N. Harding, J.D. Knowles, S. Francis-McIntyre, A. Tseng, D.I. Ellis, S. OíHagan, G. Aarons, B. Benjamin, S. Chew-Graham, C. Moseley, P. Potter, C.L. Winder, C. Potts, P. Thornton, C. McWhirter, M. Zubair, M. Pan, A. Burns, J.K. Cruickshank, G.C. Jayson, N. Purandare, F.C.W. Wu, J.D. Finn, J.N. Haselden, A.W. Nicholls, I.D. Wilson, R. Goodacre, D.B. Kell, Molecular phenotyping of a UK population: defining the human serum metabolome,

Metabolomics 11 (2014) 9–26, http://dx.doi.org/10.1007/s11306-014-0707-1.
[219] P. Begley, S. Francis-McIntyre, W.B. Dunn, D.I. Broadhurst, A. Halsall, A. Tseng, J. Knowles, R. Goodacre, D.B. Kell, D.B. Kell, Development and performance of a gas chromatography-time-of-flight mass spectrometry analysis for large-scale nontargeted metabolomic studies of human serum, Anal. Chem. 81 (2009) 7038–7046, http://dx.doi.org/10.1021/ac9011599.
[220] L. Yetukuri, M. Katajamaa, G. Medina-Gomez, T. Seppanen-Laakso, A. Vidal-Puig, M. Oresic, Bioinformatics strategies for lipidomics analysis: characterization of obesity related hepatic steatosis, BMC Syst. Biol. 1 (2007) 12, http://dx.doi.org/10.1186/1752-0509-1-12.
[221] L. Käll, J.D. Storey, M.J. MacCoss, W.S. Noble, Assigning significance to peptides identified by tandem mass spectrometry using decoy databases, J. Proteome Res. 7 (2008) 29–34, http://dx.doi.org/10.1021/pr700600n.
[222] F. Matsuda, Y. Shinbo, A. Oikawa, M.Y. Hirai, O. Fiehn, S. Kanaya, K. Saito, Assessment of metabolome annotation quality: a method for evaluating the false discovery rate of elemental composition searches, PLoS One 4 (2009) e7490, http://dx.doi.org/10.1371/journal.pone.0007490.
[223] A. Palmer, P. Phapale, I. Chernyavsky, R. Lavigne, D. Fay, A. Tarasov, V. Kovalev, J. Fuchser, S. Nikolenko, C. Pineau, M. Becker, T. Alexandrov, FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry, Nat. Methods 14 (2017) 57–60, http://dx.doi.org/10.1038/nmeth.4072.
[224] D. Abate-Pella, D.M. Freund, Y. Ma, Y. Simón-Manso, J. Hollender, C.D. Broeckling, D.V. Huhman, O.V. Krokhin, D.R. Stoll, A.D. Hegeman, T. Kind, O. Fiehn, E.L. Schymanski, J.E. Prenni, L.W. Sumner, P.G. Boswell, Retention projection enables accurate calculation of liquid chromatographic retention times across labs and methods, J. Chromatogr. A 2015 (1412) 43–51, http://dx.doi.org/10.1016/j.chroma.2015.07.108.
[225] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. 57 (1995) 289–300.
[226] Y. Benjamini, D. Yekutieli, False discovery rate- adjusted multiple confidence intervals for selected parameters, J. Am. Stat. Assoc. 100 (2005) 71–81, http://dx.doi.org/10.1198/016214504000001907.
[227] J.D. Storey, The positive false discovery rate: a Bayesian interpretation and the q −value, Ann. Stat. 31 (2003) 2013–2035, http://dx.doi.org/10.1214/aos/1074290335.
[228] J.D. Storey, A direct approach to false discovery rates, J. R. Stat. Soc. 64 (2002) 479–498.
[229] E.Y. Chong, Y. Huang, H. Wu, N. Ghasemzadeh, K. Uppal, A.A. Quyyumi, D.P. Jones, T. Yu, Local false discovery rate estimation using feature reliability in LC/MS metabolomics data, Sci. Rep. 5 (2015) 17221, http://dx.doi.org/10.1038/srep17221.
[230] J.D. Storey, R. Tibshirani, Statistical significance for genomewide studies, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 9440–9445, http://dx.doi.org/10.1073/pnas.1530509100.
[231] B. Wang, Z. Shi, G.F. Weber, M.A. Kennedy, Introduction of a new critical p value correction method for statistical significance analysis of metabonomics data, Anal. Bioanal. Chem. 405 (2013) 8419–8429, http://dx.doi.org/10.1007/s00216-013-7284-4.
[232] B.A. Ejigu, D. Valkenborg, G. Baggerman, M. Vanaerschot, E. Witters, J.C. Dujardin, T. Burzykowski, M. Berg, Evaluation of normalization methods to pave the way towards large-scale LC-MS-based metabolomics profiling experiments, Omi. A J. Integr. Biol. 17 (2013) 473–485, http://dx.doi.org/10.1089/omi.2013.0010.
[233] A.M. De Livera, D.A. Dias, D. De Souza, T. Rupasinghe, J. Pyke, D. Tull, U. Roessner, M. McConville, T.P. Speed, Normalizing and integrating metabolomics data, Anal. Chem. 84 (2012) 10768–10776, http://dx.doi.org/10.1021/ac302748b.
[234] A.M. De Livera, M. Sysi-Aho, L. Jacob, J.A. Gagnon-Bartsch, S. Castillo, J.A. Simpson, T.P. Speed, Statistical methods for handling unwanted variation in metabolomics data, Anal. Chem. 87 (2015) 3606–3615, http://dx.doi.org/10.1021/ac502439y.
[235] R. a van den Berg, H.C.J. Hoefsloot, J. a Westerhuis, A.K. Smilde, M.J. van der Werf, Centering, scaling, and transformations: improving the biological information content of metabolomics data, BMC Genomics 7 (2006) 142, http://dx.doi.org/10.1186/1471-2164-7-142.
[236] M. Sysi-Aho, M. Katajamaa, L. Yetukuri, M. Oresic, Normalization method for metabolomics data using optimal selection of multiple internal standards, BMC Bioinf. 8 (2007) 93, http://dx.doi.org/10.1186/1471-2105-8-93.
[237] M. Berg, M. Vanaerschot, A. Jankevics, B. Cuypers, R. Breitling, J.C. Dujardin, LC-MC metabolomics from study design to data-analysis -using a versatile pathogen as a test case, Comput. Struct. Biotechnol. J. 4 (2013) 1–8, http://dx.doi.org/10.5936/csbj.201301002.
[238] Y. Wu, L. Li, Sample normalization methods in quantitative metabolomics, J. Chromatogr. A 2016 (1430) 80–95, http://dx.doi.org/10.1016/j.chroma.2015.12.007.
[239] M. Katajamaa, M. Orešič, Processing methods for differential analysis of LC/MS profile data, BMC Bioinf. 6 (2005) 179, http://dx.doi.org/10.1186/1471-2105-6-179.
[240] B.M. Warrack, S. Hnatyshyn, K.-H. Ott, M.D. Reily, M. Sanders, H. Zhang, D.M. Drexler, Normalization strategies for metabonomic analysis of urine samples, J. Chromatogr. B 877 (2009) 547–552, http://dx.doi.org/10.1016/j.jchromb.2009.01.007.