Metabolomics (2022) 18:24
https://doi.org/10.1007/s11306-021-01848-6

REVIEW ARTICLE

Reference materials for MS-based untargeted metabolomics and lipidomics: a review by the metabolomics quality assurance and quality control consortium (mQACC)

Katrice A. Lippa[1] · Juan J. Aristizabal-Henao[2,3] · Richard D. Beger[4] · John A. Bowden[2] · Corey Broeckling[5] · Chris Beecher[6] · W. Clay Davis[7] · Warwick B. Dunn[8] · Roberto Flores[9] · Royston Goodacre[10] · Gonçalo J. Gouveia[11] · Amy C. Harms[12] · Thomas Hartung[13] · Christina M. Jones[1] · Matthew R. Lewis[14] · Ioanna Ntai[15] · Andrew J. Percy[16] · Dan Raftery[17] · Tracey B. Schock[7] · Jinchun Sun[4] · Georgios Theodoridis[18] · Fariba Tayyari[19] · Federico Torta[20] · Candice Z. Ulmer[21] · Ian Wilson[22] · Baljit K. Ubhi[23]

Received: 29 June 2021 / Accepted: 7 October 2021
This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2022

Abstract

Introduction The metabolomics quality assurance and quality control consortium (mQACC) is enabling the identification, development, prioritization, and promotion of suitable reference materials (RMs) to be used in quality assurance (QA) and quality control (QC) for untargeted metabolomics research.

Objectives This review aims to highlight current RMs, and methodologies used within untargeted metabolomics and lipidomics communities to ensure standardization of results obtained from data analysis, interpretation and cross-study, and cross-laboratory comparisons. The essence of the aims is also applicable to other 'omics areas that generate high dimensional data.

Results The potential for game-changing biochemical discoveries through mass spectrometry-based (MS) untargeted metabolomics and lipidomics are predicated on the evolution of more confident qualitative (and eventually quantitative) results from research laboratories. RMs are thus critical QC tools to be able to assure standardization, comparability, repeatability and reproducibility for untargeted data analysis, interpretation, to compare data within and across studies and across multiple laboratories. Standard operating procedures (SOPs) that promote, describe and exemplify the use of RMs will also improve QC for the metabolomics and lipidomics communities.

Conclusions The application of RMs described in this review may significantly improve data quality to support metabolomics and lipidomics research. The continued development and deployment of new RMs, together with interlaboratory studies and educational outreach and training, will further promote sound QA practices in the community.

Keywords Reference materials · Certified reference materials · Internal standards · Untargeted analysis · Mass spectrometry · Metabolomics · Lipidomics · Metabolomics quality assurance and quality control consortium (mQACC)

1 Introduction

The metabolomics Quality Assurance and Quality Control Consortium (mQACC) was established in 2018 to build a collaborative effort among relevant stakeholders from academia, industry, and governmental organizations to address key quality assurance (QA) and quality control (QC) issues in untargeted metabolomics (Beger et al., 2019). As part of its mission, the mQACC is engaging the metabolomics community to identify and to prioritize key reference materials (RMs) to be used in QA/QC for untargeted metabolomics research. RMs are artifact-based measurement standards that have been characterized for a known composition of specific physical, chemical or biological properties. They are often described by their function (e.g., calibration, quality control, method validation) and range in design from matrix-based materials from natural (e.g., biological) sources to "matrix-free" standards, such as pure substances or standard solutions and mixtures. The focus of untargeted metabolomics research is to detect and identify hundreds of metabolites and minimize sources of variance (biological *versus* technical) to

✉ Baljit K. Ubhi
drbkubhi@gmail.com

Extended author information available on the last page of the article

Published online: 09 April 2022

identify differential metabolomics patterns of interest with an eventual goal to quantify select metabolites of biological interest. Thus, the appropriate use of RMs in untargeted metabolomics applications will provide confidence for such measurements and data standardization methods from different instrumental platforms, thereby ensuring suitable translation of biological discoveries through the elucidation of biomarkers or understanding of biological mechanisms.

Technological advances have allowed mass spectrometry (MS)-based untargeted metabolomics and lipidomics to be widely adopted in research laboratories. In addition to pushing the boundaries of biochemical research, including translational and precision medicine, untargeted analyses contribute to the advancement of nutritional assessment, fermentative optimizations, and agricultural productivity. Given its predominant usage in the metabolomics and lipidomics communities, MS hyphenated to chromatography separation techniques (e.g., liquid or gas chromatography with MS-based detection, LC–MS or GC–MS) represent a primary analytical method for untargeted metabolomics but also present unique challenges. The mQACC has recently defined the specific measurement challenges that different types of RMs may potentially pose and address best use practices for RMs by the metabolomics community (Evans et al., 2020). This effort directly builds upon prior considerations (Bowden et al., 2018; Broadhurst et al., 2018; Dudzik et al., 2018; Ribbenstedt et al., 2018; Schrimpe-Rutledge et al., 2016; Viant et al., 2019) of the analytical and QA/QC challenges faced in MS-based untargeted metabolomics and lipidomics, in contrast to the more common targeted metabolomics approaches. More recently, Alseekh et al. (2021) describe practical considerations for MS-based metabolomic workflows to improve the quality and comparability of resultant data and metadata. All of these efforts aim to demonstrate, disseminate and promote QA procedures and QC reference materials to be used across the community and enable metabolomics and lipidomics researchers to quickly adopt such practices to ultimately produce high-quality data and results.

QA/QC is critical to ensure that quality results are obtained from the diverse range of chromatographic separation approaches and MS-based detection methods that exist across laboratories. This diversity is due in part to available instrumentation, available processing software, the specific goals of the project, and the sample types used in the specific studies. Effective QA/QC in untargeted metabolomics requires the interplay between the two quality management processes (Broadhurst et al., 2018). QA is considered the processes that ensures quality results before actual measurements are conducted, such as the development and use of Standard Operating Procedures (SOPs) with corresponding training of metabolomics researchers and personnel. QC is the day-to-day operational techniques and processes of evaluating the quality of results and overall laboratory performance, which often includes the use of RMs. While current RMs cannot be expected to immediately solve all QC issues associated with MS-based untargeted metabolomics, they can be used in developing future approaches towards more confident compound identification (Levels 1 and 2 of the Metabolomics Standards Initiative) (Sumner et al., 2007), increased reproducibility of results and eventual quantification, while also acting as a necessary bridge for comparability of results across multiple analytical platforms (including nuclear magnetic resonance spectroscopy (NMR)-based untargeted metabolomics) and among other laboratories.

Accordingly, the mQACC is actively working with the broader metabolomics community to develop measurement designs, protocols, and methods together with supporting materials comprised of solution-based and matrix-relevant RMs that can be utilized across instrumentation platforms for routine QA/QC practices in untargeted metabolomics. The development of unified products that include associated reference data are a future goal and will be essential QA/QC tools for fully confident results to be obtained from untargeted metabolomics and lipidomics studies. The development of many of these RMs is being spearheaded by commercial organizations and government agencies in direct response to a community need, facilitating broader distribution among international research communities, which serves as a framework for increasingly coordinated generation of standardized and quality-controlled data.

Munafò et al. (2017) have highlighted that there is an urgent need to increase reproducibility in science, and that part of the problem is the lack of transparency in reporting studies. Therefore, having confidence in the data generated will reduce uncertainty within experimental pipelines and therefore improve laboratory performance and standardization across laboratories. For metabolomics, and indeed all the 'omics, transparency in data is central to this. The provision of open data accompanying any reported studies should be first encouraged and then expected. These freely available data should be readily interpretable and include the metabolomics data on the samples themselves, along with any associated metadata about these samples, as well as data on the QA and the QC samples that have been analyzed as part of the metabolomics pipeline.

Originating from its nonregulatory objectives, the mQACC consortium aims to present all potential QA/QC solutions that can improve overall laboratory performance and facilitate more comparable metabolomics and lipidomics measurements across the community, while maintaining neutrality to all potential products. For this review, the products are principally MS-based application centric and include commercial reference standard producers, RMs from government organizations, and any

associated reference data sources from commercial entities, academic institutions and/or other organizations. In each subsequent section, we will review how methodologies have emerged and evolved to the current state of the art, concluding with a look at what promising developments are on the horizon to further enable QC and drive data transparency and reproducibility, as well as integrity and confidence in untargeted MS-based metabolomics studies, both within and across laboratories.

2 Reference materials for untargeted metabolomics

The progression of metabolomics and lipidomics research to meet the ever-increasing demands of continuity and scale driven by large single and multi-site clinical and epidemiological studies, has created a need to demonstrate confidence in analytical performance. The same is true of other metabolomics and lipidomics research, where many thousands of samples may be analyzed; for example, for predicting gene-phenotype links in large-scale functional genomics studies or in plant breeding. One of the priority efforts of the mQACC is to envision the key characteristics for broadly applicable and sustainable RMs that the community can afford. An initial effort has focused on blood- and urine-based RMs, but other types of materials including (but not limited to) synthetic mixtures, endogenous materials, spiked and isotopically-labeled materials (including from bacterial, yeast or eukaryotic cell culture), disease- and species-specific materials, and various tissues (including from plants) with their associated extracts are an ultimate goal. Together with development of RMs, a consideration of what type of associated reference data (and metadata) for assessing metabolomics data quality is a critical parallel effort.

In a recent study (Evans et al., 2020), it was determined that only 33% of metabolomics laboratories use RMs regularly and that the use of RMs was not consistent across individual laboratories; some laboratories use RMs as a long-term reference QC samples, whereas others utilize them for cross-platform evaluations or interlaboratory studies. Similarly, a survey from over 125 laboratories in the lipidomics community (Bowden et al., 2018) indicated that a wide methodological diversity exists; less than half of laboratories formally establish and adhere to SOPs and QC practices. Further, most of the laboratories do not have standardized policies for the adoption of methods and protocols, including the use of measurement standards, software, and quantification procedures, and reporting of false positive results in lipid identification.

2.1 Definitions

Towards the development of solutions for improving QA/QC in untargeted metabolomics, the use of RMs aim to address overcome potential QC barriers in the various analytical and data acquisition steps. As previously introduced, the general concept of RMs can be broad, and the various forms of RMs are utilized for a range of applications (see Table 1). RMs include Certified Reference Materials (CRMs) (highly characterized RMs supplied with a certificate of analysis), synthetic reference standards, solutions and standard mixtures, and Reference Library (RL) products that are also comprised of higher purity standards. Often referred to as pooled QC samples, QCRMs can be study specific (Bijlsma et al., 2006; Sangster et al., 2006) or study independent (i.e. surrogate) (Dunn et al., 2011a, 2011b) including those intended for longer term use, for example across multiple studies and/or platforms within or across laboratories (termed "Long term Reference" or LTR samples) (Lewis et al., 2016). Each also include long-term reference QC samples (Broadhurst et al., 2018) and can be operationally defined to serve in a similar capacity to a RM for control and reporting of observed variation in untargeted profiling measurements. Other important QC samples include extraction or process blanks and system conditioning samples, but these are not considered RMs in a general sense. RMs have a range of applications as are described in this review and can be applied towards pre- or post-sample processing steps, including modifying instrument acquisition and MS-tuning parameters (Bouhifd et al., 2015) or the development of post-acquisition informatic approaches.

The terms RMs and CRMs are used throughout the various fields of analytics and bio-analytics and are explicitly defined by the International Organization for Standardization (ISO) (ISO, 2016). The purpose, intended use and scope of RMs and CRMs are also described in ISO Standard 17,034 (ISO, 2021) and its related Standards and Guides (Trapmann et al., 2017). Table 1 summarizes these definitions.

A measurement standard (and its derived term reference measurement standard) is the embodiment of the realization of a given quantity (with a stated value and associated measurement uncertainty) that is used as a reference (JCGM, 2012). This term is often used synonymously with calibrator and implies the determination of quantitative values through a calibration measurement procedure. Many reference measurement standards (i.e., reference standards) are utilized in quantitative determinations for targeted metabolomics; however, for untargeted metabolomics, these standards are used largely for chemical identification purposes.

Table 1 Definitions of RMs and CRMs, and descriptions of other measurement standards commonly employed for QC in untargeted metabolomics

Category	ISO definitions (ISO, 2016) and descriptions	Practical usage
Reference material (RM)	17,034 Sect. 3.3 Definition: A material, sufficiently homogeneous and stable with respect to one or more specified properties, which has been established to be fit for its intended use in a measurement process Notes: Reference material is a generic term. Properties can be quantitative or qualitative, e.g. identity of substances or species. Uses may include the calibration of a measurement system, assessment of a measurement procedure, assigning values to other materials, and quality control	RM is a generic term and is generally used to describe a wide range of materials and measurement standards used in QC
Certified reference material (CRM)	17,034 Sect. 3.2 Definition: A reference material characterized by a metrologically valid procedure for one or more specified properties, accompanied by a reference material certificate that provides the value of the specified property, its associated uncertainty, and a statement of metrological traceability Notes: The concept of value includes a nominal property or a qualitative attribute such as identity or sequence. Uncertainties for such attributes may be expressed as probabilities or levels of confidence	CRMs are considered within the broader definition of RMs. CRMs are highly specialized materials, which are generally only produced in a few highly specific areas, where critical measurement requirements and traceability considerations must be met In practice, pure chemical and solution CRMs are designed for calibration, chemical identification and SI traceability. Matrix-based CRMs are designed for method validation and accuracy control applications, but can also be used for intra-laboratory QC, interlaboratory assessments and method harmonization efforts
Quality control reference material (QCRM)	Description: Pooled materials comprised of subsets from all (or a representative subset) of the biological test samples in a specific study, that are well mixed into a homogenous pool and then aliquoted into subsamples. (Often termed pooled QC materials.)	QCRMs are used for general QC measurements for the study of origin but are also well suited to intra-laboratory assessment or routine analysis. They can also be utilized in other quality assurance purposes, such as quality management system training. Can also be repurposed for interlaboratory method harmonization
Standard mixtures	Description: Mixtures of reference standards of pure chemical compounds in a homogenous solution form that have been well characterized	Even though these standards are usually prepared for use as calibration standards for quantitative (targeted) analysis, they can be used in chemical identification for untargeted approaches. Standard mixtures are generally prepared with enough aliquots to be widely available and to be stable for a sufficient period of time
Reference library (RL)	Description: Collections of pure authenticated compounds prepared either in neat form or as individual solutions or as defined standard mixtures	The compounds and/or mixtures are used for chemical identification or system suitability applications. Often the individual chemical standards are provided initially as lyophilized (dried) to be reconstituted in an appropriate solvent

2.2 Design considerations

RMs are designed to be stable and homogenous, ideally with long-term stability with respect to the components of interest under defined storage conditions. Matrix-based RMs should also be designed with appropriate matrices for the intended QC applications. Further guidance on RMs in the context of alternatives to animal tests (Hoffmann et al., 2008) can be easily translated to the case for RMs for untargeted metabolomics and lipidomics. The characterization of these materials should include reliable and confident chemical identity, mass concentrations (when relevant) and metrological traceability of specific components. The characterization should also include descriptions of relevant impurities, physicochemical properties, use by dates, and safety considerations associated with their typical usage. The availability of high-quality reference results associated with typical laboratory usage would be highly desirable; this would include the adequate potency within response range of evaluation, such as detection limit and dynamic range of the instrumental platform, but also the historical performance and reproducibility of results. Universal access to these materials and their affordability (i.e., cost) should be a goal in production through suitable commercial vendors and private and governmental organizations. Furthermore, the resistance to the overall expense of these products and the limited guidance and "best practices" available on how to utilize them effectively also presents a challenge for widespread acceptance within the metabolomics and lipidomics communities. For standard mixtures of authentic standards as (certified) reference materials, composition, purity, and concentration are readily available, often supported by a certificate of analysis and easily reported. In the case of biological matrices (e.g., the 1950 plasma provided by NIST) many of these features are less readily available as such mixtures are complex and incompletely characterized in terms of all their constituents. Arguably, biological matrices not really qualify as CRM and especially QCRM as the composition and stability of components cannot be defined. However, pragmatically, often no more traceable mixtures can be created, but caution is advised that such ill-defined RM can serve as CRM or QCRM only for defined properties.

Note: while measurement reproducibility has a formal metrological definition (JCGM, 2012), herein it is used to mean that measurements on similar materials made by different laboratories and/or at different times obtain similar results.

As part of the development process, RMs are designed to be fit-for-purpose as documented by their intended usage statements (ISO, 2017) and thus appropriate RMs can be used to harmonize results obtained by individual instruments and procedures and directly compared to the reference results. As they are ideally commercially available, the more widely such RMs are used, the greater the potential exists for the comparability and harmonization of results across laboratories and over time. As a pooled matrix-based (e.g., serum, plasma, urine) RM becomes adopted and widely used, it is inevitable that the initial production batch will become depleted. Thus, the challenge exists for RM producers to replenish with contiguous replacement materials that maintain suitable characteristics for use in existing nontargeted metabolomics QC applications. Notably, as RMs become more specific in both how they are designed and characterized for intended usage (e.g., sample type and form), the likelihood that they become universally adopted is reduced. A limited selection of appropriate matrix-matched RMs for specific metabolomics and lipidomics applications remains an ever-present limitation.

2.3 Usage

An example of the range of reference materials employed in the field of metabolomics and lipidomics is provided in Fig. 1. RMs can be leveraged to determine and reduce measurement variability associated with the analytical sample preparation (and some select pre-analytical controls), instrumental analysis and system suitability, compound identification, and data processing and analysis steps for which the standards have been qualified. A further example (Evans et al., 2020) depicts the analysis of RMs as quality control (in addition to other QC tools) within a continuum of QC practices. The development and use of SOPs for a RM would fall under QA practices that comprise an overarching quality management system.

Routine usage of RMs with associated QA and QC protocols benefits not only the quality of their data for individual laboratories, but it also enhances the entire field with increased comparability of results and conclusions amongst laboratories and individual studies, which will lead to more open data and robustness and reproducibility for metabolomics and lipidomics. However, some analytical method variation across laboratories with respect to metabolite identification and/or profiles is still to be expected for these reference materials as a result of differences in procedures and instrumentation.

However, the wide adoption of RMs will increase our understanding of the sources of variation and help to minimize bias between method results. As an example, retention order can change using different column chemistry and/or mobile phase composition; Vaughan et al. discuss the application of contiguous samples run in 100 different laboratories using 100 different analytical methods and has generated a comparable dataset suitable for developing a widely useful retention time prediction model (Vaughan et al., 2012). It is also known that the use of different mass spectrometers can result in different relative intensities for

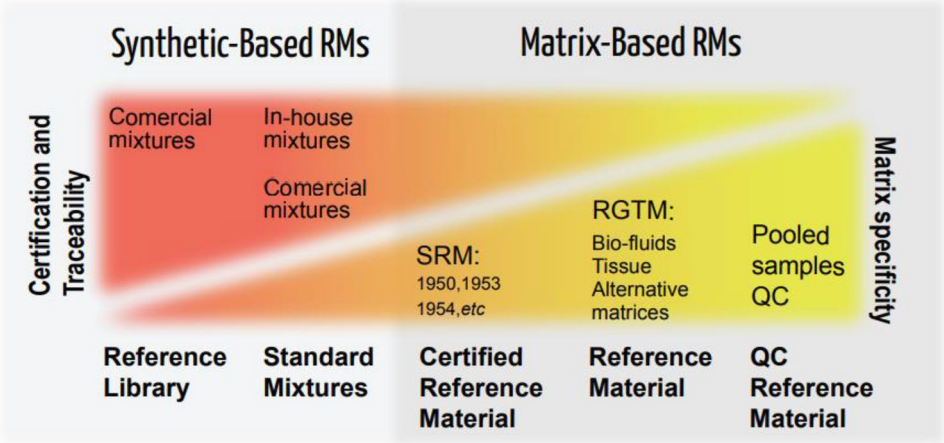


Fig. 1 Range of reference materials employed in the field of metabolomics and lipidomics. Gradient colors from yellow to red represent the inverse relationship between matrix specificity to the study samples and the metabolite traceability to certified standards. Each of these RMs can be applied to capture the inherent unwanted technical variance across the numerous steps that make up a metabolomics workflow. RM assessment is to be carried out before, during and after in accordance with the defined best practices and QA/QC system

the same metabolites, which can lead to differences in the results of multivariate analysis (Gika et al., 2010).

Use of a common RM in two datasets can serve as a key to aligning signals across an otherwise disparate sets of samples. The exact composition of the RM (e.g., a synthetic chemical standard mixture or a matrix-matched RM) will dictate its utility in alignment within the chromatographic retention time, *m/z*, and signal intensity (i.e., response factor) dimensions and the extent of its applicability to the global profile. Where successful, datasets aligned using a common RM can enable a higher degree of comparability and increased statistical power leading to more confidence in the biological knowledge gleaned from combined studies. Results obtained from a collective set are more extensible than that of the single study and may enable new questions to be asked of the data which were not anticipated in the original design. An important requirement for such an approach is that each sample set has clearly defined data and metadata organized consistent with the FAIR guidelines—Findable, Accessible, Interoperable, and Reproducible (Wilkinson et al., 2016). These guidelines do not act as a standard or specification as such, but rather as guiding principles for the reporting of data and metadata recently exemplified in system biology, drug discovery and other biomedical fields that can benefit from data reusability and hence further knowledge discovery.

3 Reference materials of biological origin

Biological reference materials are RMs with biological (rather than synthetic) origin that are characterized for specific biological or chemical properties and frequently serve as QC materials to support numerous QC practices including: assessment of analytical system suitability, evaluation of measurement reproducibility, and fusion of batched data. They are often pooled materials and designed to be representative of the natural biochemical complexity observed in a given sample type (e.g., biofluid or tissue). These RMs address a different need than spike-in QC materials added for the evaluation samples at a specific stage during processing or to pooled QC materials that are applied in each analytical evaluation (Broadhurst et al., 2018). In recent years, the use of biological RMs with defined SOPs has become a cornerstone of metabolomics and lipidomics profiling interlaboratory studies, allowing the highly complex data generated to be compared across numerous sites to either generate (untargeted studies) or answer (targeted studies) hypotheses. Use of biological RMs and deposition of the resulting data to repositories also enable multi-laboratory data compilations and benchmarking of community practices and measurement methodologies for advancing intra- and interlaboratory harmonization.

3.1 NIST SRM 1950

One of the first CRMs designed specifically for targeted metabolomics and made widely available to the research community was the National Institute of Standards and Technology (NIST) Standard Reference Material® (SRM) 1950 Metabolites in Frozen Human Plasma. This CRM was designed as a "universal matrix" to include plasma from 100 individuals from an equal number of men and women in a narrow adult age range (40–50 years) together with a racial distribution of the donors reflecting the distribution in the US population at the time of implementation. The CRM is

specifically intended for the validation of analytical methods used in the determination of various nutritional and health status markers for clinically relevant metabolites in human plasma and similar materials. It has been value-assigned for nearly 100 electrolytes, amino acids, vitamins, hormones, and fatty acids (Phinney et al., 2013; Simón-Manso et al., 2013). SRM 1950 has since been used to validate novel measurement methodologies, protocols and workflows for metabolomics and lipidomics in the profiling and quantification of targeted compounds (Azab et al., 2019; Colas et al., 2014; Hermann et al., 2018; Lange & Fedorova, 2020; Misra & Olivier, 2020; Rampler et al., 2018; Rathod et al., 2020; Ribbenstedt et al., 2018; Roy et al., 2016; Schoeny et al., 2020; Schwaiger et al., 2018; Triebl et al., 2020; Ulmer et al., 2018; Wang et al., 2018). It has also recently been used for comparing methods, platforms, and data analysis for untargeted metabolomics and lipidomics applications (Cajka et al., 2017; Di Giovanni et al., 2020; Drotleff et al., 2019; Koelmel et al., 2020; Nichols et al., 2018; Ribbenstedt et al., 2018; Telu et al., 2016).

Many researchers performing untargeted metabolomics and lipidomics have adopted SRM 1950 as a long-term reference QC sample. It has been employed in evaluating instrumental performance, correction for batch variance, and ensuring comparability within and across laboratory studies (Liu et al., 2020). The extension of SRM 1950 as a control material for lipidomics evaluations is more recent (Aristizabal-Henao et al., 2020) and has been fundamental for the development of the human plasma lipidomic field. In 2010, Quehenberger et al. (2010) published a semi-quantitative description of SRM 1950, reporting the levels of over 500 distinct molecular species distributed among the major lipid classes. In this work, supported by the LIPID MAPS initiative, SRM 1950 promoted the use of adequate analytical methodologies to quantify the large spectrum of plasma lipids. The multiple targeted approaches adopted to generate the results were based on multiple reaction monitoring (MRM) detection and could almost cover the same number of species normally reported in untargeted experiments after validation of features.

The interest in the use of biological matrix-based RMs in the lipidomic community was furthered with the inclusion of SRM 1950 as a common QC material in a recent interlaboratory study that generated both lipid identifications and quantitative and/or semi-quantitative estimates of lipid concentrations (Bowden et al., 2017). While primarily a targeted MS-based lipidomics effort, this study highlighted significant disparities in the lipid concentrations and profiles reported by the participants, possibly due to the use of different internal standards, extraction methods and MS techniques. This has furthered efforts to develop guidelines and harmonize lipidomic workflows (Bowden et al., 2018) for the entire community. In the meanwhile, lipids in SRM 1950 have been characterized and quantified by other metabolomics global initiatives and in more controlled conditions, as part of a comprehensive list of metabolites that were reported in these efforts (Koelmel et al., 2020).

3.2 Plasma and urine RM suites

SRM 1950 was originally designed as a standalone "normal" human plasma pooled material to validate analytical methods for targeted metabolite determinations rather than for untargeted profiling purposes. As this SRM contains nearly 50 measurands that are certified (and another 50 that are non-certified) by NIST, the cost remains relatively high, and is thus not practical for use as a routine and forever-sustained metabolomics QC material. NIST has thus recently established the Metabolomics Quality Assurance and Quality Control Materials (MetQual) program in an effort to improve the comparability of untargeted metabolomics measurements across all sectors, including industry, government, and academic laboratories, and provide access to new NIST RMs for metabolomics with a cost-friendly option. This includes the four-part candidate RM 8231 Frozen Human Plasma Suite for Metabolomics that can be used as QC materials for untargeted, differential metabolomics composed of pooled plasma: Part A. Diabetic Plasma, Part B. High Triglyceride Plasma, Part C. Young, African–American Plasma, and Part D. Normal Human Plasma from the same source as SRM 1950. This RM suite has been designed specifically for untargeted metabolomic analysis. Furthermore, the lipidomic profiles of the RM plasma suite have been characterized (Aristizabal-Henao et al., 2020), and its utility to benchmark the performance of data processing tools has been established (Riquelme et al., 2020).

A complementary suite of pooled urine RMs composed of both female and male smokers and non-smokers is also under development. These suites are intended to provide the metabolomics community with additional options to single-point QCRMs, such as SRM 1950. Moreover, in contrast to the extensive (and costly) value assignment of SRM 1950, these plasma and urine RM suites will be characterized by NIST and metabolomics stakeholders in through community-based interlaboratory studies. Characterization will include metabolite identification, annotation, fold changes and percent differences for purposes of underpinning differential metabolomics (Bearden et al., 2019) and lipidomics (Aristizabal-Henao et al., 2020) studies (see Sect. 6.1).

3.3 Tissue-based reference materials

In contrast to metabolomics studies that employ common biofluids, direct analysis of a tissue provides a specific understanding of the biochemical function of a specialized organ or local environment. Such studies may seek to

correlate the more invasive tissue diagnostic markers with profile signatures identified in biofluids. Reproducibility and QC practices are equally important for tissue research, and to the best of our knowledge, these practices rely on in-house developed pooled materials, a limitation to broader harmonization. To this end, NIST is developing tissue based RMs specifically for untargeted and differential based lipidomic, metabolomic and proteomic measurements in model and non-model species.

Research involving metabolomics of tissue commonly use liver, heart, kidney, skeletal muscle, or lungs, to name a few; however, with the complex process involving the design and production of RMs (Sect. 2.2), it is improbable to produce a RM for every common matrix. Challenges in RM development include acquiring large quantities of material that can be available to the community for upwards of a decade. Obtaining limited supply, human-sourced tissues amplifies this challenge. Furthermore, the time and expense of adding metrological traceability to the chemical characterization of these highly-complex matrix materials adds to the costs of RM production which are ultimately transferred to the customer.

In researching options for the first tissue-based omics material, NIST prioritized minimizing cost and production time and concentrated on a common study tissue. A fortunate opportunity presented itself with a collection of human livers cryogenically preserved and archived at the NIST Biorepository. The livers were preserved for a collaborative project between NIST and the Environmental Protection Agency (EPA) in 1979 to establish an environmental specimen banking system aimed at evaluating risks to human health and the environment due to the influx of man- made substances into the ecosystem. Though these tissues were originally collected for biomonitoring purposes, the goals of the NIST Biorepository have expanded to include the use of archived tissues for additional applications such as genetics, metabolomics, and proteomics. The livers are a source of several materials mentioned within this section.

3.3.1 Liver suite for untargeted metabolomics analysis

The primary rationale for developing RM suites with distinct metabolic profiles is to promote measurement harmonization through the detection of differences when used within and across studies. Candidate RM 8462 Frozen Human Liver Suite for Proteomics and Metabolomics is currently in development. Pathological data and body mass index (BMI) calculations were used to categorize into three cohort liver materials: Normal Liver, Congested Liver and Fatty Liver (Fig. 2). The RM will be characterized for differential expression of lipids, metabolites, and proteins. It should be highly valuable in determination of definitive levels of analytes for quality control in differential based studies. Use of such suites may help to ensure that actual differences can be detected between sample groups regardless of the instrumentation, statistical approaches, and software tools.

Candidate RM 8461 Human Liver for Proteomics is another cryogenically homogenized and freeze-dried liver tissue originally developed as a qualitative material for complex bottom-up LC–MS proteomic analysis (Davis et al., 2019c). However, RM 8461 has been assessed as a stand-alone untargeted metabolomics material (Davis et al., 2019a, 2019b) and should be a good candidate for use as the only currently available tissue-based control material in the field. Preliminary LC–MS data demonstrated over 12,000 features with CV $\leq$ 20% from an initial assessment of 4 vials with over 2000 putative annotations resulting from spectral library matches of both mzCloud and NIST20 databases (NIST, 2020).

3.3.2 Liver extracts for system suitability

The importance of evaluating the performance of analytical platforms in advance of conducting metabolomic measurements has been illuminated in the recent literature (Broadhurst et al., 2018; Dunn et al., 2011a, 2011b; Rattray et al., 2019; Viant et al., 2019). These system suitability QC samples are typically a mixture of a small number of metabolites in a solution absent of a sample matrix and at known concentrations. These metabolites can be monitored over time and acceptance criteria (peak shape, retention time, peak area/height) checked before the start of each analytical batch (see Sect. 4).

However, the metabolomics community is suffering from the lack of a common, everyday system suitability standard by which to benchmark instrument performance for untargeted MS-based approaches. NIST has developed a research grade test material (RGTM) 10,122 Metabolomics System Suitability Sample as a large quantity, biological extract from human livers which incorporates the entirety of a metabolome, resulting in an encompassing system suitability sample. The design of a tissue extract as a suitability standard eliminates the sample preparation variation observed within biological samples, while offering simplicity of use (rehydrate and inject) and analyte complexity for analysis of all metabolomics platforms. This sample will require initial benchmarking in conjunction with the currently used and defined standard mixtures described in Sect. 4. Once suitability is established, either or both samples can be evaluated in system suitability testing. In addition to the acceptance criteria mentioned above, the metrics used to establish suitability may include the number of detected features, total ion chromatogram, MS resolution (ppm error), and LC resolution of critical pairs (e.g., leucine/isoleucine).

In addition to use in evaluating platform readiness, RGTM 10,122 is presently available, and can be a QC tool

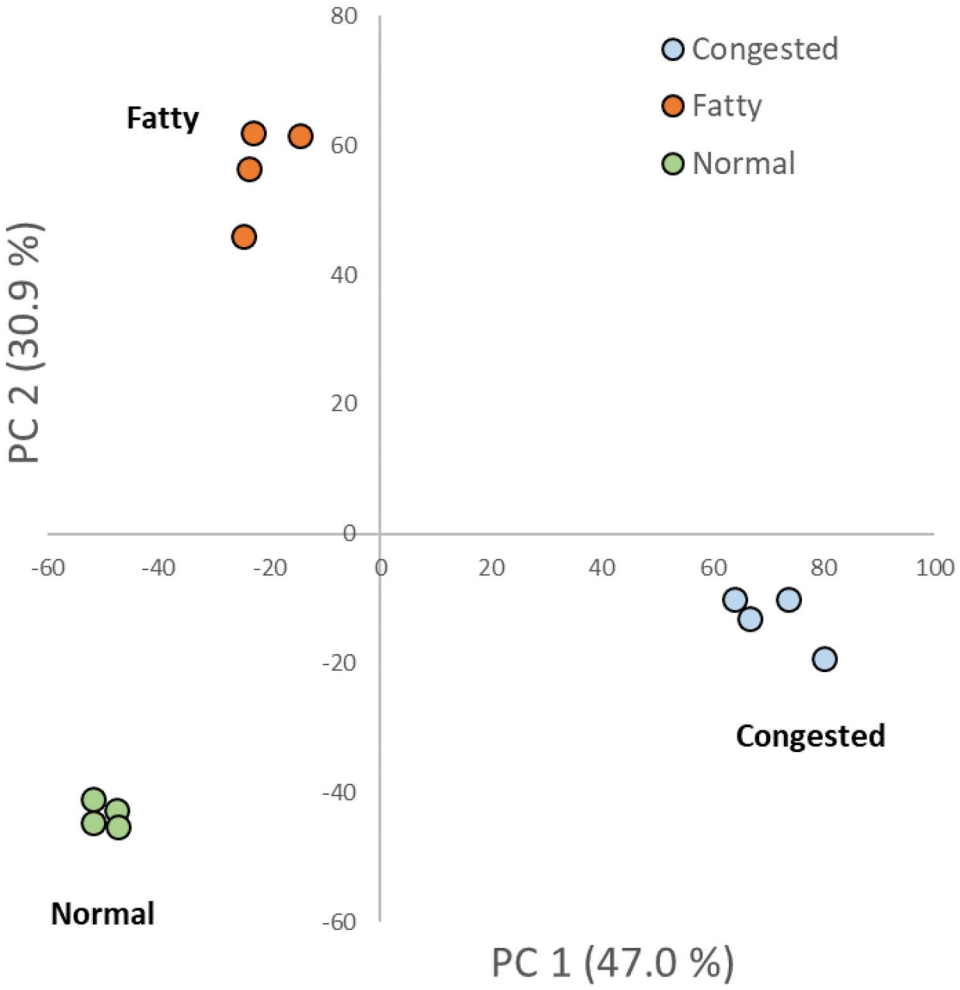


Fig. 2 Principal component analysis (PCA) scores plot for the liver suite for differential analysis. High resolution accurate mass (HRAM) of each health state ($n=4$) includes normal (green filled circle), fatty (orange filled circle) and congested (dark blue filled circle) liver. The values in paratheses in the axes refer to the percentage total explained variance

in the comparability of instrument performance across batches, studies, and laboratories, such as large, multi-center collaborations.

3.4 In-house matrix-based reference materials

Lower-cost QCRMs can be produced in-house to initially address QA/QC needs, however production and sampling variability limit their applicability and long-term use. As noted, as a significant challenge in design of RMs (Sect. 2.2), there is a recognized need for contiguous supply of stable, matrix-specific materials. As an alternative, an iterative batch averaging method (IBAT) (Gouveia et al., 2021) may be used to produce stable in-house RMs over the course of time with relatively low variance. The IBAT process reduces the production and sampling contributions to variance by creating a common source of material from which homogeneous aliquots are produced. The advantage of this method is that instead of producing a single large batch, which will have its own challenges in achieving homogeneity and longevity, the material is continuously generated over time. Aliquots from different batches are combined into a single tube of which only minor amounts are newly produced material. Each combined material captures small changes over time while having minimal variance between different IBAT iterations. This method is flexible, easily adjusted to the production throughput and applicable to any type of matrix, thus, suitable for QC applications but also to establish RMs for various metabolomes as an important component for metabolite annotation across analytical platforms, methods, and laboratories.

3.5 Alternative matrices as future reference materials

The traditional biofluid-based materials such as plasma, serum and urine commonly used in clinical diagnostics are the most dominant type of samples used for metabolomics and lipidomics studies. However, many substitute biological matrices such as saliva, cerebrospinal fluid, mucous, bronchial lavage, breast milk and feces are being considered and subsequently investigated as alternative RMs to

support (relatively) non-invasive metabolomics and lipidomics screening techniques. As this is a nascent area for metabolomics and lipidomics research, prototype RMs are either in the concept phase or in the early stages of development and thus are not readily available.

3.5.1 Saliva and mucosa

Salivary metabolomics has been mostly focused on biomarker discovery (Gardner et al., 2020), but the limited supply and production of saliva represent a significant bottleneck for using it and other non-blood biofluids as RMs in clinical research. The term "saliva" mainly has been used for the fluids produced in the oral cavity by glands, which includes whole-mouth saliva, gingival–crevicular fluid, parotid saliva, and submandibular/sublingual saliva. The collection of whole-mouth saliva is non-invasive and does not require special training. The complexity of saliva composition and its natural variability per individual (and secretion over time) is a significant driver for the development of a RM for use as a QC material for untargeted metabolomics analysis.

The evaluation of microbial-derived metabolites (Brown et al., 2019; Scott et al., 2020; Song et al., 2020) within mucosa and stroma such as hydrogen sulfide, ceramides, tryptophan and bile acid derivatives and their association with non-communicable chronic diseases (i.e., diabetes, non-alcoholic fatty liver disease, obesity, Crohn's and inflammatory bowel disease and cancer) is another emerging area in metabolomics. This evidence suggests that some of these microbial-derived metabolites can affect mucosa permeability and induce localized proinflammatory response (i.e., along the mucosa lining in the gut) or are able to enter the colonic epithelial cells (i.e., chronic systemic inflammation in obesity). Both targeted and untargeted metabolomic approaches, in conjunction with other multi-omics platforms, have been used to determine the effects of diet or dietary components in mucosa associated microbiota linked to disease and evaluate the efficacy of therapeutic strategies (Aden et al., 2019; Mars et al., 2020). The development of RMs for the harmonization of the measurements of microbial-derived metabolites associated with mucosa integrity represents a necessary first step in the identification of potential biomarkers of disease risk for both preclinical and clinical studies.

3.5.2 Breath and volatile analyses

There has been increasing interest in the use of breath analysis (also called breathomics), as the metabolome of the volatile organic compounds (VOCs) in breath can be captured in a facile non-invasive manner. Breath, and indeed VOCs from other sources (e.g., skin, wounds, bacteria, foodstuffs), are generally captured using sorbent materials due to the high vapour pressure of these chemical species, and as a consequence of their low concentrations (ppb is typical), sorbents also pre-concentrates them. Once captured on materials like (e.g.) Tenax, polydimethylsiloxane, or the Carbopack and Carboxen series, VOCs are released using thermal desorption (TD) and subsequently analysed by GC–MS (Lawal et al., 2017; Rattray et al., 2014). In addition to these off-line methods, some MS analyses can be performed directly on VOCs in the headspace using selected ion flow tube-mass spectrometry (SIFT-MS) or proton transfer reaction-mass spectrometry (PTR-MS) (Bruderer et al., 2019; Smith & Spanel, 2005). The breathomics community are investigating standardization of sampling and analysis of breath samples (Herbig & Beauchamp, 2014) and a recent study by Wilkinson and colleagues generated benchmark values for TD-GC–MS analysis of human breath samples containing peppermint-derived VOCs using data collected from several different research groups (Wilkinson et al., 2020). Due to the nature of the analysis and capture of VOCs, the RM cannot be a QCRM sample as it is technically very challenging to mix breath from different people (Broadhurst et al., 2018). Therefore, RMs will need to be comprised of standard mixtures of reference VOCs, with known vapor pressures and known chemistries so that they are absorbed by the sorbent materials used for VOC capture.

3.5.3 Breast milk and other fluids

There has been significant interest in the characterization of complex lipids in human breast milk beyond classical fatty acid compositional studies (George et al., 2018). Although human breast milk RMs do exist with certified values for organic contaminants (NIST SRM 1953 Organic Contaminants in Non-Fortified Human Milk and SRM 1954 Organic Contaminants in Fortified Human Milk), it does not appear that recent lipidomic studies have used these SRMs. There are known differences in the lipid content of human breast milk throughout the course of lactation, due to diet, health status, and depending on the time of sampling (i.e., foremilk vs. hindmilk) (Jensen, 1996). It is thus imperative that future studies consider adopting such materials for QC purposes in order to maximize the translatability of the results within this burgeoning area of infant nutrition. Likewise, other biofluids that have been used in recent lipidomic studies but do not presently have commercially-available RMs include cerebrospinal fluid (Reichl et al., 2020), synovial fluid (Leimer et al., 2017), and amniotic fluid (Cao et al., 2020) among others (Acera et al., 2019; Agatonovic-Kustrin et al., 2019; Gregory et al., 2013; Höfner et al., 2020; Nilsson et al., 2019; Yang et al., 2019).

3.5.4 Feces

Another emerging area in metabolomics is the characterization of microbial metabolites in human and other mammalian feces (aka stool) and its association with gut microbiota and metabolism biomarkers of health and disease. Both targeted and untargeted metabolomic profiling of human stool have been successfully used to identify a dysbiotic metabolic signature associated with disease and for the assessment of dietary intake in diverse communities (Jain et al., 2019; Kim et al., 2020; Lloyd-Price et al., 2019). However, relatively few studies have reported the optimization and QC of microbial metabolites from stool, likely due to the complexity and heterogeneity of stool specimens. Deda et al. (2017) illustrated that the pH and volume ratio for the fecal sample weight to extraction solvent in the sample preparation are critical to obtain more comprehensive metabolic profiles (untargeted and targeted), which are also platform dependent. In another study (Moosmang et al., 2019), different extraction methods and the effect of solvents were used to compare the variability in metabolome analysis from stool specimens; water as the extraction solvent yielded the best results in terms of coverage, number of detected features and reproducibility. A recent systematic study (Cui et al., 2020) found that nearly 70% of the systematic variation is related to the extraction solvent; freeze drying caused a relative loss of short chain fatty acids and lower reproducibility than wet methods utilizing raw fecal slurry stool material. Volatile metabolites in collected stool specimens present a serious challenge for preservation and require development of standardized protocols for extraction (Bosch et al., 2018). Since freeze-drying may have a strong influence on metabolite degradation, QCRMs are needed to measure the efficacy of extraction and integrity of any novel approaches used for compound preservation.

Currently, the Gut Microbiome Committee of the Institute for the Advancement of Food and Nutrition Sciences (IAFNS; formerly the International Life Sciences Institute) aims to identify and eventually quantify gut microbial metabolites that have been linked to diet and health. Following a October 2019 workshop (Mandal et al., 2020), NIST and IAFNS are now collaborating to develop a suite of human whole stool RMs to validate both metagenomics measurements associated with Fecal Microbiome Transplants (FMTs) and other live biotherapeutic products as well as metabolomic measurements to identify new biomarkers associated with the health of the human gut microbiome. Two pooled whole stool prototype RMs collected from vegans and omnivores has been developed for use in method harmonization and QC for next generation sequence (NGS) metagenomics and MS- and NMR-based untargeted metabolomics. These RMs have been demonstrated to be homogeneous with respect to both the microbial taxa (DNA) and key metabolites. They are being evaluated for longer-term (> 6 months) stability in addition to any differences in the aqueous and lyophilized storage conditions. They are being used as test materials in an ongoing gut microbiome metabolomics NIST interlaboratory study (NIST, 2021).

4 Synthetic chemical standard mixtures

The need for high purity RMs comprised of individual chemical components as standard mixtures has increased over the past several years. These mixtures are used for performing routine metabolomic assay fitness evaluations, broad-based metabolite quantification and for construction of retention time and multi-stage MS (MS/MS) spectral libraries. Four major components for ensuring data quality in bioanalytical MS-based measurements are instrument qualification, analytical method validation, system suitability and quality control checks (Briscoe et al., 2007). Related to untargeted metabolomics, system suitability tests are commonly performed before, during and after an experiment to assess the system performance using a set of standard mixtures of compounds, which are either natural or stable isotope labeled compounds. In theory, the standard mixture(s) should be adaptable to a variety of metabolomic methods and applicable to different analytical workflows, whilst satisfying the aims necessary for qualification and ideally quantification. A control chart of results can reveal gas chromatography (GC), LC or MS system performance deficits (e.g., signal drift or offset, peak tailing or splitting, column degradation, mass calibration) and can alert the user to corrective maintenance that would be necessary prior to the measurement of precious experimental samples. QC check standards, comprised of an experimental sample spiked with a known concentration of stable isotope-labeled internal standards, are then incorporated into the analysis run and are designed to monitor the data quality of each sample. By establishing acceptance criteria and requirements for use as an internal standard, both qualification and quantification are inherently possible.

4.1 In-house standard mixtures

Synthetic mixtures of chemical reference standards prepared within individual laboratories ("in-house" standard mixtures) can provide a fit-for-purpose solution to the immediate and often unique needs of untargeted metabolomics and lipidomics studies, or where RMs for specific sample types are not available. These mixtures do have limitations but are affordable and can be applied for different QA/QC processes once they have been shown to be fit for purpose. The construction of in-house mixtures generally leverages the wide variety of neat chemical standards available to

tailor mixtures that meet the specific needs of the locally employed technologies and methods. Table 2 describes a variety of standard mixtures used in metabolic profiling and clearly illustrates the diversity of applications, methods, and purposes. A review of the relevant literature suggests that synthetic mixtures themselves broadly fit one of two types: (1) those intended to be used on their own (generally constituting unlabeled materials) and (2) those intended to be added to biological sample material (generally constituting stable-isotope labelled materials). These mixtures appear to be used broadly in one or more of three main contexts: (1) assurance of system suitability before, during and after a metabolomics experiment, (2) QC during an analysis, and/or (3) correction of data after an analysis. Specific applications include the monitoring of chromatographic retention time accuracy and/or precision, monitoring signal intensity accuracy and/or precision, controlling preparative or technical processes (e.g., extraction efficiency, injection volume) and/or correcting after acquisition of data.

The maintenance of in-house mixtures for routine QC applications requires procurement and characterization of neat chemicals, consideration of the stability of these chemicals individually and in mixtures, rationalized mixture design, construction of SOPs for the repeatable construction of mixtures for use as routine reagents, and validation of mixture stability over the course of intended use. When prepared utilizing well-characterized pure chemicals as starting materials and adhering to a rigorous SOP, the quality of the prepared standard mixture may even meet the ISO qualifications (ISO, 2016) and be considered equivalent to a RM. While the costs of producing in-house standard mixtures is more controllable and may be minimized for a specific application, material creation and maintenance procedures can require significant laboratory resource investment (e.g., in maintaining the capability of personnel and condition of requisite measurement instruments).

Technologies supporting metabolic profiling are numerous and varied, including LC, GC, capillary electrophoresis, and NMR. Methodology can be equally varied, for example with common modes of LC separation for metabolomics applications. These modes include reversed-phase LC (RPLC), hydrophilic interaction liquid chromatography (HILIC) and ion exchange chromatography (IEC), each of which presents additional diversity in stationary and liquid phase combinations. For this reason, QC in metabolic profiling approaches is largely supported by in-house standard mixtures.

Table 2 describes standard mixtures used in metabolic profiling and illustrates the diversity of applications, methods, and purposes. A review of the relevant literature suggests that synthetic mixtures themselves broadly fit one of two types: (1) those intended to be used on their own (generally constituting unlabeled materials) and (2) those intended to be added to biological sample material (generally constituting stable-isotope labeled materials). Similarly, these mixtures appear to be used broadly in one or more of three main contexts: (1) assurance of system suitability before, during and after a metabolomics experiment, (2) QC during an analysis, and/or (3) correction of data after an analysis. Specific applications include the monitoring of chromatographic retention time accuracy and/or precision, monitoring signal intensity accuracy and/or precision, controlling preparative or technical processes (e.g. extraction efficiency, injection volume) and/or correcting after acquisition of data.

The maintenance of in-house mixtures for routine QC applications requires procurement and characterization of neat chemicals, consideration of the stability of these chemicals individually and in mixtures, rationalized mixture design, construction of SOPs for the repeatable construction of mixtures for use as routine reagents, and validation of mixture stability over the course of intended use. Although these steps do require significant laboratory resource investment (e.g., maintaining the capability of personnel, dedicated laboratory instrument usage), they have several inherent advantages over use of commercially available mixtures. The wide variety of neat chemical standards to tailor the preparation of mixtures can meet the needs of specific situations or methods and can provide better fitness for purpose for a specific technology or methodology, such as a specific mixture to match biochemical panels or to extend across broad profiling applications.

4.2 Commercially-available standard mixtures

When available and suitable for the application, commercially-available standard mixtures can offer laboratories ready and continuous access to qualified standard mixtures that have been consistently produced with high lot-to-lot reproducibility and have well-established stability criteria. Both convenience and third-party accreditation may be important in some applications including those in more regulated environments. In addition to the technical requirements, standard mixtures are often designed to meet any number of important factors including cost minimization, control of reagent purity, ease of preparation, and chemical stability. These benefits result at the expense of convenience and third-party accreditation or guarantee of quality, which may be important in some more regulated applications. When available and suitable for the application, commercially-available standard mixtures can offer laboratories ready (and continuous) access to qualified standard mixes that have been consistently produced with high lot-to-lot reproducibility and have well-established stability criteria.

Standards such as those discussed in this section assure that there is enough reproducibility in the data that they are worth all of the time, money and effort of further data

Table 2 Examples of synthetic chemical standard mixtures

Mixture composition (no. of components)	Mixture type		Purpose	Application	Application context			Performance checks				References
	Un-labeled	Labeled			Before	During	After	RT	Signal Intensity	*m/z*	Yield	
Amino acids, bile acids, sugars, organic acids (7)	Y		System suitability	Run separately	Y		Y	Y	Y	Y		(Zelena et al., 2009)
Small polar metabolites (8)	Y		Quantification	Spike in sample		Y		Y	Y	Y	Y	(Zelena et al., 2009)
Small molecule metabolites, nonpolar species (13)		Y	RI QC (for POS)	Spike in sample		Y		Y	Y	Y	Y	(Evans et al., 2009)
Amino acid metabolites, organic acids (11)		Y	RI QC (for NEG)	Spike in sample		Y		Y	Y	Y	Y	(Evans et al., 2009)
Small molecule metabolites, bile acids, organic acids (14)	Y		System suitability	Run separately	Y	Y	Y	Y	Y	Y		(Gika et al., 2016)
Small molecule metabolites, organic acids (4)	Y		System suitability (for POS)	Run separately	Y	Y	Y	Y	Y	Y		(Gika et al., 2016)
Small molecule metabolite, sugar, bile acids (4)	Y		System suitability (for NEG)	Run separately	Y	Y	Y	Y	Y	Y		(Gika et al., 2007)
Small molecule metabolites, organic acids (8)		Y	RP QC	Spike in sample	Y	Y	Y	Y	Y	Y		(Lewis et al., 2016)
Amino acid, small polar metabolites (6)		Y	HILIC QC	Spike in sample	Y	Y	Y	Y	Y	Y		(Lewis et al., 2016)
Amino acid, organic acid (2)		Y	Quantification for GC–MS	Spike in sample		Y		Y	Y	Y		(Lewis et al., 2016)
Sugars (2)	Y	Y	Quantification	Spike in sample	Y	Y	Y	Y	Y	Y	Y	(Papadimitropoulos et al., 2018)
Small molecule metabolites, organic acids (6)	Y		System suitability	Run separately	Y	Y	Y	Y	Y	Y		(Pandher et al., 2009)
Amino acids, bile acids, small molecule metabolites, xenobiotics (11)	Y		System suitability	Run separately	Y	Y	Y	Y	Y	Y		(Pereira et al., 2010)
Amino acids, small molecule metabolites, lipids (22)	Y		Validation data across platforms	Run separately	Y	Y	Y	Y	Y	Y		(Naz et al., 2013)
Amino acids, lipids, xenobiotics (44)	Y		System suitability	Run separately	Y	Y	Y	Y	Y	Y		(Barri et al., 2012)
amino acids, bile acids, small molecule metabolites, lipids (7)		Y	Quantification	Spike in sample		Y		Y	Y	Y		(Barri et al., 2012)
Amino acids, lipids (7)	Y		System suitability	Run separately	Y	Y	Y	Y	Y	Y		(Broadhurst et al., 2018)
Small polar metabolites (4)	Y		System suitability	Run separately	Y	Y	Y	Y	Y	Y		(Gika et al., 2012)
Amino acids, fatty acids, sugars, organic acids (11)		Y	Quantification	Spike in sample		Y		Y	Y	Y		(Dunn et al., 2011a, 2011b)
Amino acids, lipids, small polar metabolites (14)		Y	Quantification	Spike in sample		Y		Y	Y	Y	Y	(Soltow et al., 2013)

workup. This point of view is not just relevant to the isotopic standards discussed but is equally true for all of the materials discussed in this paper. One could avoid the cost of the quality control elements, but to do so puts the quality of the data at risk. Products such as QReSS, IROA, and the other standards discussed in this review provide methods to assure and enhance data quality. The understanding and correction of all instrumentation, and source-created quantitative error, the ability to sample-to-sample normalize data, and to correctly name compounds are all critical to developing the reproducible, high-quality datasets that metabolomics needs to move forward and attain the utility it should. The elements in this section, being more chemically defined than biological mixtures and yet being standard mixtures available to all researchers provide a consistent foundation by which instrumental performance, and ultimately the quality of the data, can be understood.

4.2.1 QReSS kits

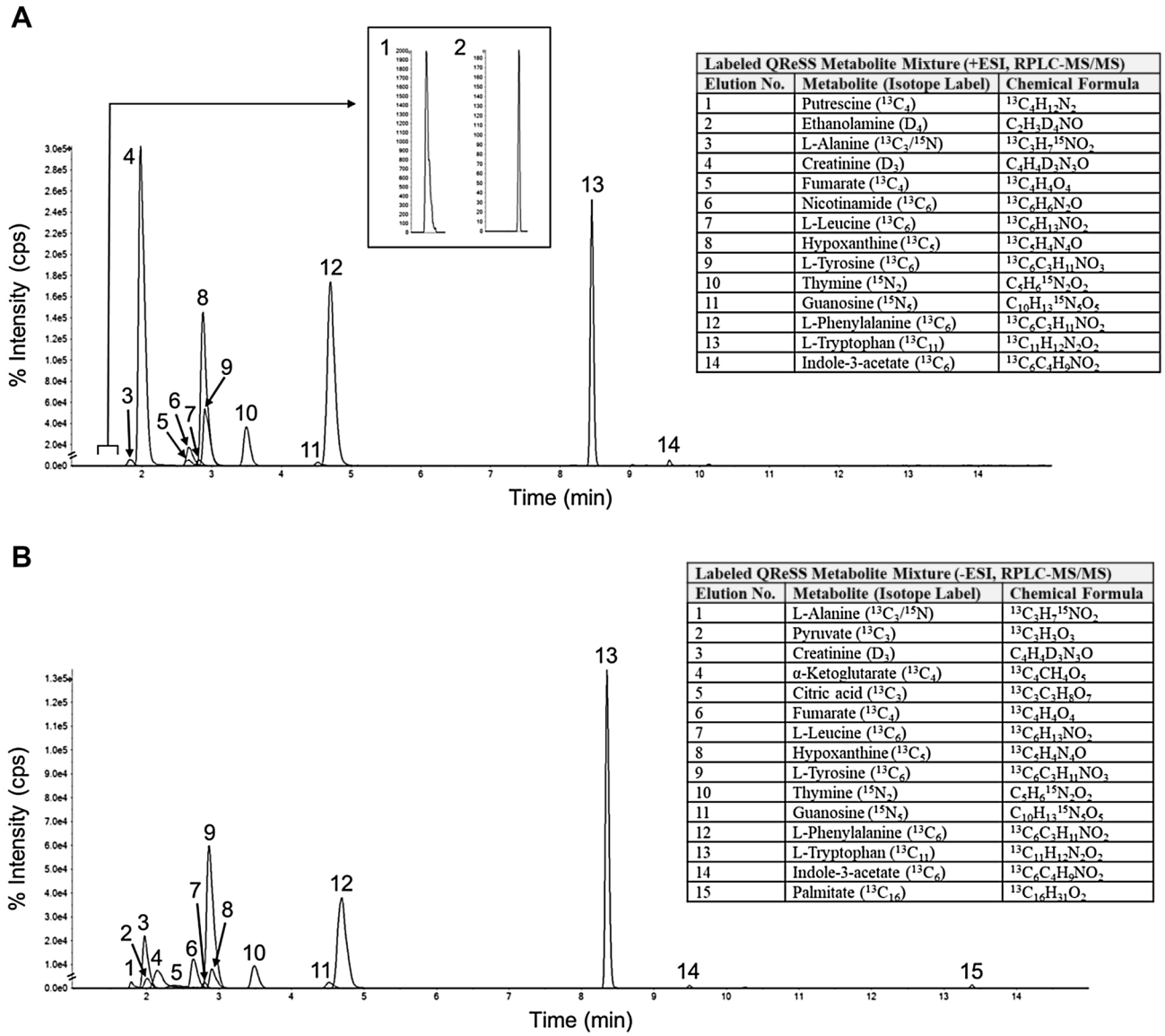
Cambridge Isotope Laboratories (CIL) has developed multi-component standard mixes that can be applied to MS metabolomic experiments following simple solvent dissolution and mix workup. In one recent example, CIL (in collaboration with SCIEX) developed a metabolomics kit named QReSS (Quantification, Retention, and System Suitability). This kit was designed to help aid performance assessments (i.e., method QC and system suitability) and, in tandem, enable metabolite quantification in MS metabolomic workflows (untargeted, semi-targeted and targeted). This kit comprises two dried-down mixtures of 18 metabolites (in their stable isotope-labeled or unlabeled form (CIL, 2021) that span molecular weights and metabolic classes as well as chromatographic retention range. These standard mixtures are well suited for such applications due to the carefully selected compounds, inherent characteristics, and their experimental tendencies [e.g., diverse elution behavior (Fig. 3), devoid of solubility issues or stability concerns]. In one application example of system suitability with QReSS, an aliquot of the combined QReSS mixes is analyzed directly by LC–MS or LC–MS/MS using an untargeted or targeted metabolomics workflow, with performance metrics tracked over time. The nature of the MS metric tracking is predicated on the technique (e.g., targeted *vs*. untargeted), with the results ideally being displayed pictorially using Pareto plots or Shewhart control charts (González-Riano et al., 2020). Through the longitudinal monitoring of performance metrics, deviations in data quality relating to the LC and/or MS system can be illuminated and immediately addressed.

4.2.2 IROA TruQuant measurement system

The IROA TruQuant measurement system (see Fig. 4A–D) relies on an isotopically labeled Long Term Reference Standard (IROA-LTRS, Fig. 4D), which is paired with a chemically identical but isotopically different Internal Standard (IROA-IS, Fig. 4B). The IROA-LTRS provides a daily measure of platform QC (Evans et al., 2020), including measures of MS source function, MS instrument performance, chromatographic separation, and quantification. The IROA-LTRS is a Standard Reference Material that is run qualitatively (using data dependent fragmentation techniques) rather than quantitatively; alternate data independent fragmentation, or ion mobility scans yield validated identification of all IROA signed peaks and provide daily retention time (RT) and amplitude for all the same compounds in the IROA-IS. The IROA-IS is an internal standard that is chemically identical to the LTRS but has only the C13 isotopic (i.e., all C13 dominant isotopomers and isotopologues -Fig. 4B) which when spiked into the experimental samples (Fig. 4A) provides a mechanism for correction of ion suppression and other source-induced variances, calibration standard-based quantification, and sample-to-sample normalization for the analytical samples (Fig. 4C). The randomization of injections of the IROA-LTRS and analytical samples (containing the IROA-IS) at an injection frequency of 1–10, respectively, corrects for any within-experiment drift and provides a measure of daily instrument reproducibility if used as an injection standard, or combined sample preparation and instrument reproducibility if used as a recovery standard. Because of their biological origin these standards can be quite cost effective as injection standards and yet are quite justifiable as recovery standards because they allow for the correction of all variances imposed onto the original samples by either instrumental or preparative variability.

The IROA-IS is inserted into all analytical samples (Fig. 4C) and uses the RT and identification from the IROA-LTRS (Fig. 4D) to locate, identify, and to quantify all the natural abundance peaks associated with the IROA-IS. Quantification of metabolites is enhanced because the IROA-IS is always present at the constant concentration, and therefore (1) ion-suppression, and other source-related errors, may be corrected, and (2) the natural abundance analytical sample may be normalized to the IROA-IS (Fig. 4E; always constant). The normalization factors may be applied to all compounds in each sample, even for compounds without a specific standard.

The cost of most isotopic labels is not in the cost of the isotopic label, but rather in the cost of the isolation, and purification of individual compounds; therefore, IROA's approach is to use standardized biologically produced compound mixtures to make the internal standards as broadly



Labeled QReSS Metabolite Mixture (+ESI, RPLC-MS/MS)		
Elution No.	Metabolite (Isotope Label)	Chemical Formula
1	Putrescine ($^{13}C_4$)	$^{13}C_4H_{12}N_2$
2	Ethanolamine (D_4)	$C_2H_3D_4NO$
3	L-Alanine ($^{13}C_3/^{15}N$)	$^{13}C_3H_7{}^{15}NO_2$
4	Creatinine (D_3)	$C_4H_4D_3N_3O$
5	Fumarate ($^{13}C_4$)	$^{13}C_4H_4O_4$
6	Nicotinamide ($^{13}C_6$)	$^{13}C_6H_6N_2O$
7	L-Leucine ($^{13}C_6$)	$^{13}C_6H_{13}NO_2$
8	Hypoxanthine ($^{13}C_5$)	$^{13}C_5H_4N_4O$
9	L-Tyrosine ($^{13}C_6$)	$^{13}C_6C_3H_{11}NO_3$
10	Thymine ($^{15}N_2$)	$C_5H_6{}^{15}N_2O_2$
11	Guanosine ($^{15}N_5$)	$C_{10}H_{13}{}^{15}N_5O_5$
12	L-Phenylalanine ($^{13}C_6$)	$^{13}C_6C_3H_{11}NO_2$
13	L-Tryptophan ($^{13}C_{11}$)	$^{13}C_{11}H_{12}N_2O_2$
14	Indole-3-acetate ($^{13}C_6$)	$^{13}C_6C_4H_9NO_2$

Labeled QReSS Metabolite Mixture (-ESI, RPLC-MS/MS)		
Elution No.	Metabolite (Isotope Label)	Chemical Formula
1	L-Alanine ($^{13}C_3/^{15}N$)	$^{13}C_3H_7{}^{15}NO_2$
2	Pyruvate ($^{13}C_3$)	$^{13}C_3H_3O_3$
3	Creatinine (D_3)	$C_4H_4D_3N_3O$
4	α-Ketoglutarate ($^{13}C_4$)	$^{13}C_4CH_4O_5$
5	Citric acid ($^{13}C_3$)	$^{13}C_3C_3H_8O_7$
6	Fumarate ($^{13}C_4$)	$^{13}C_4H_4O_4$
7	L-Leucine ($^{13}C_6$)	$^{13}C_6H_{13}NO_2$
8	Hypoxanthine ($^{13}C_5$)	$^{13}C_5H_4N_4O$
9	L-Tyrosine ($^{13}C_6$)	$^{13}C_6C_3H_{11}NO_3$
10	Thymine ($^{15}N_2$)	$C_5H_6{}^{15}N_2O_2$
11	Guanosine ($^{15}N_5$)	$C_{10}H_{13}{}^{15}N_5O_5$
12	L-Phenylalanine ($^{13}C_6$)	$^{13}C_6C_3H_{11}NO_2$
13	L-Tryptophan ($^{13}C_{11}$)	$^{13}C_{11}H_{12}N_2O_2$
14	Indole-3-acetate ($^{13}C_6$)	$^{13}C_6C_4H_9NO_2$
15	Palmitate ($^{13}C_{16}$)	$^{13}C_{16}H_{31}O_2$

Fig. 3 Total ion chromatogram (TIC) of a matrix-free, combined QReSS mix measured by RPLC-MS (Phenomenex Kinetex F5 column, SCIEX TripleTOF® 6600 LC–MS/MS System). Acquisition from + ESI and -ESI are shown in **A** and **B** respectively, with the annotations corresponding to the metabolite elution order in its corresponding table inset

applicable as possible. In the case of the IROA IS and LTRS, the standards may be used as either recovery standards or as injection standards depending on when they are introduced, i.e., either before or after the sample preparation, with the recovery standard requiring approximately three times as much isotopic standard as the amount used as an injection standard. To put this in perspective, at current pricing, these standards cost $2 (as an Injection standard) or $6 (as a recovery standard) per sample. The "recovery cost" of most of the NIH Metabolomics Centers, and most of the academic metabolomic centers, appears to be averaging approximately $200 per sample. The major contributors to these costs are the annual instrumentation costs (including depreciation), and the personnel costs, thus the costs of the isotopic standards are running between 1 and 3 percent of the costs of the analysis. In addition, the cost of running the MS analysis is a small portion of the additional cost of the personnel time to analyze, extract, and interpret the data; all of which would be wasted if the data quality was not sufficient to justify this additional investment.

4.2.3 Standard mixtures for lipidomics

Broad-based absolute quantification is difficult in lipidomics as there are estimated to be between 10,000 and 100,000 different lipid molecular species (van Meer, 2005, van Meer,

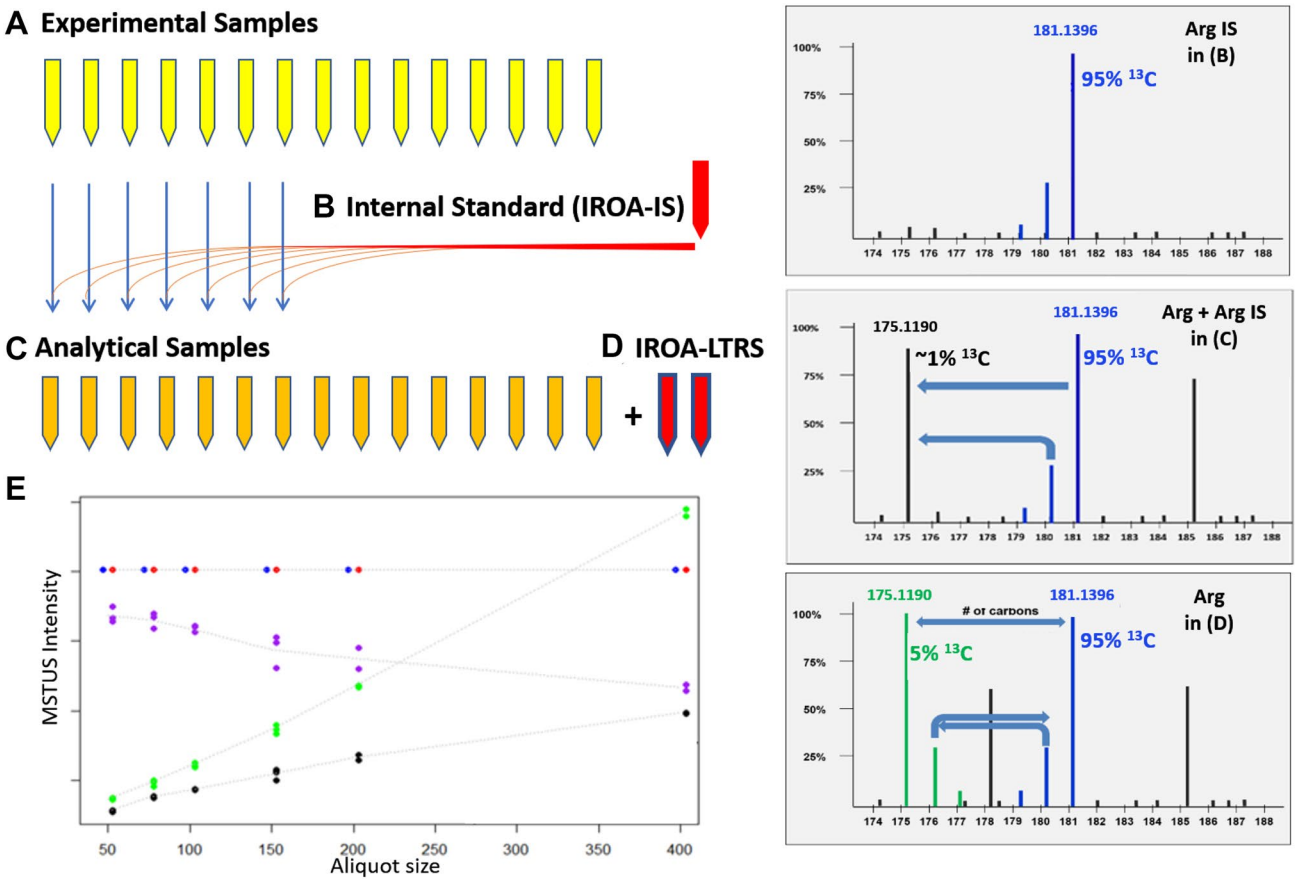


Fig. 4 The IROA TruQuant measurement system with the experimental samples **A** is spiked with the **B** internal standard (IROA-IS) to generate **C** analytical samples that are also paired with an **D** isotopically labeled Long Term Reference Standard (IROA-LTRS). Example mass spectra of an analytical sample **C** for arginine (Arg) with the corresponding IROA-IS **B** and IROA-LTRS **D** are illustrated on the right panels. The triple-redundancy of the LTRS assures more accurate identification. An example quantification result as a normalized intensity for arginine to the IROA-IS is also provided **E**

et al., 2008, Yetukuri et al., 2008; Wenk, 2010). As only a small number of commercially available isotope-labelled standards are available, it is necessary to assume that the behavior of these standards is representative for each lipid class, with respect to extraction recovery, matrix effect and mass spectrometric detection. New internal standards mixtures that could qualitatively and quantitatively represent the endogenous lipid classes distribution in plasma have been added to the catalogue of the major manufacturers, such as Avanti Lipids. The series of SPLASH® LIPIDOMIX®, tailored on either human or mouse plasma, is produced with varying amounts of labelled standards (one standard per class) and premixed in organic solvent, which can be added to a defined amount of sample prior to lipid extraction. The mixtures aid relative quantification using HILIC or direct infusion-based methods, since all the members of each class co-elute hence the matrix suppression effect is comparable. Because of this relative quantification approach, samples employing this standard can only be compared for relative differences. The new, more diverse lipid standard mixture, UltimateSPLASH, includes 69 deuterated lipids covering multiple classes with the inclusion of several molecular species (3–9) for each class. These deuterated standards can better mimic the structural characteristics of the endogenous analytes that affect their signal intensity and thus result in more accurate quantification of lipids.

Lipid subclass-specific components of the UltimateSPLASH mix have been recently made commercially available, including, for example, a phosphatidylcholine mix of five mass-labeled standards across varying fatty acyl-chain lengths and degrees of unsaturation at different concentrations. Application-specific standard mixes within the SPLASH series have also begun to be developed, such as the OxysterolSPLASH mixture, which contains 13 deuterated oxysterol standards. This standard mix is particularly interesting as it enables a bridge to be drawn between a relatively niche area within lipid research (targeted or semi-targeted sterol analyses) and the vast majority of nontargeted

lipidomics research. Many of these sterol analyses are based on GC–MS and require analyte derivatization [e.g., oxysterols (Dias et al., 2018)], in contrast to the majority of the lipidomics analyses which primarily utilize LC–MS or direct infusion (shotgun) MS approaches with nonderivatized samples. The implementation of different analyte handling and preparation techniques as well as the intrinsic differences between complementary technologies and platforms has important implications for ensuring QC. The potential for further community-wide efforts, such as interlaboratory studies for GC and LC–MS-based lipid analyses, should continue to be explored, similar to what has previously been achieved in GC–MS-based metabolomics (Lin et al., 2020) and fatty acid methyl ester analysis (Metherel et al., 2019; Schantz et al., 2016). The creation of novel labeled standards that are designed for specific lipid classes, matrices and/or applications will facilitate future harmonization studies through more accurate quantification.

4.2.4 Lipidyzer™ kits

The Lipidyzer™ platform uses an expanded set of internal standards, containing over 50 deuterium-labeled lipid molecular species across 13 lipid classes to mimic the biochemistry found in human plasma. The standards were developed by SCIEX, a mass spectrometry vendor in collaboration with Avanti Polar Lipids and Metabolon, a service provider in the metabolomics industry. This approach normalizes the quantitative bias that occurs across lipids with different chain lengths and degrees of unsaturation to allow for more accurate measurements. In addition to the labelled internal standards, additional kits are available which allow a user to assess the reproducibility and sensitivity of their system before running samples. The System Suitability kit enables the user to assess the sensitivity of the assay and the reproducibility (robustness) of the platform. The SelexION® Technology Tuning kit allows the automated optimization of the differential mobility spectrometry (DMS) cell which aids in definitive lipid identification. Finally, the QC Spike Standards kit, which contains unlabeled molecular lipid species, can be added to the QC control plasma at a known concentration and monitored throughout the analysis as a QC sample. Even though these standards were developed using the above technology, the Lipidyzer Platform has since been discontinued. The internal standards are still commercially available and can be applied to any mass spectrometry platform and ion mobililty technology such as FAIMS (Thermofisher Scientific), SLIM (MOBILion Systems Inc.), etc.

An example of the internal strategy used for the phosphatidylcholines (PC) class is provided in Fig. 5A. The *sn*-1 (top carbon of glycerol backbone) stereospecific numbering position is a labeled palmitate and then the *sn*-2 (middle carbon) position is changed for every fatty acid from a short chain palmitoleic acid to a long chain docosahexaenoic acid. Therefore, there are multiple internal standards to reflect the diversity of the lipid molecular species. The remainder of the 12 lipids classes have a similar strategy.

An example of the quantification data produced by the Lipidyzer™ platform is provided in Fig. 5B. Twenty-five human serum samples with known total cholesterol esters (CE) and CE fatty acid compositions were profiled using the Lipidyzer™ Platform (Ubhi, 2018). Figure 5B (left panel) highlights the Lipidyzer™ Platform quantified total CE with less than 10% bias, compared to 100% bias in the estimate made using a single internal standard. The results indicated that using a single internal standard greatly overestimated the concentration of CE, likely by overestimating the contribution of the major unsaturated fatty acids. Figure 5B (right panel) illustrates the individual fatty acid profiles of CE (expressed as a mole % of total CE) when quantified using the Lipidyzer™ Platform and the single internal standard. A recent study (Contrepois et al., 2018) utilized these standards for a comparative untargeted vs targeted lipidomics approach.

4.2.5 Biocrates absoluteIDQ

Biocrates incorporates a standard mixture of internal standards as part of a method for targeted quantification of both hydrophilic and lipid metabolites. The AbsoluteIDQ p180 kit is the most commonly used and widely adopted by laboratories conducting targeted quantification studies. It comprises analysis of 180 metabolites, mainly amino acids, biogenic amines, and lipids. The Biocrates AbsoluteIDQ p400HR kit offers a ready-to-use, standardized approach for broad lipid and metabolic profiling based on high resolution mass spectrometers, either by LC–MS data acquisition (amino acids and biogenic amines) or flow injection analysis methods (FIA-MS) (lipids). It provides quantitative or semiquantitative information for over 400 metabolites from 11 analyte groups: amino acids, biogenic amines, acylcarnitines, monosaccharides (hexose), diglycerides, triglycerides, lysophosphatidylcholines, phosphatidylcholines, sphingomyelins, ceramides, and cholesteryl esters. All kits contain calibration standards, internal standards, QC samples, and system suitability test mixes and an SOP with detailed instructions for sample preparation, instrument setup, system suitability testing, and data analysis. While the Biocrates kits are primarily designed to enable quantification, they may also be leveraged for use as reference standards for QC and chemical identification purposes in untargeted methods.

An international ring trial, with data collected by 14 laboratories, was established to evaluate the intra- and inter-laboratory precision and accuracy of the AbsoluteIDQ p400HR kit (Thompson et al., 2019) using plasma test materials from

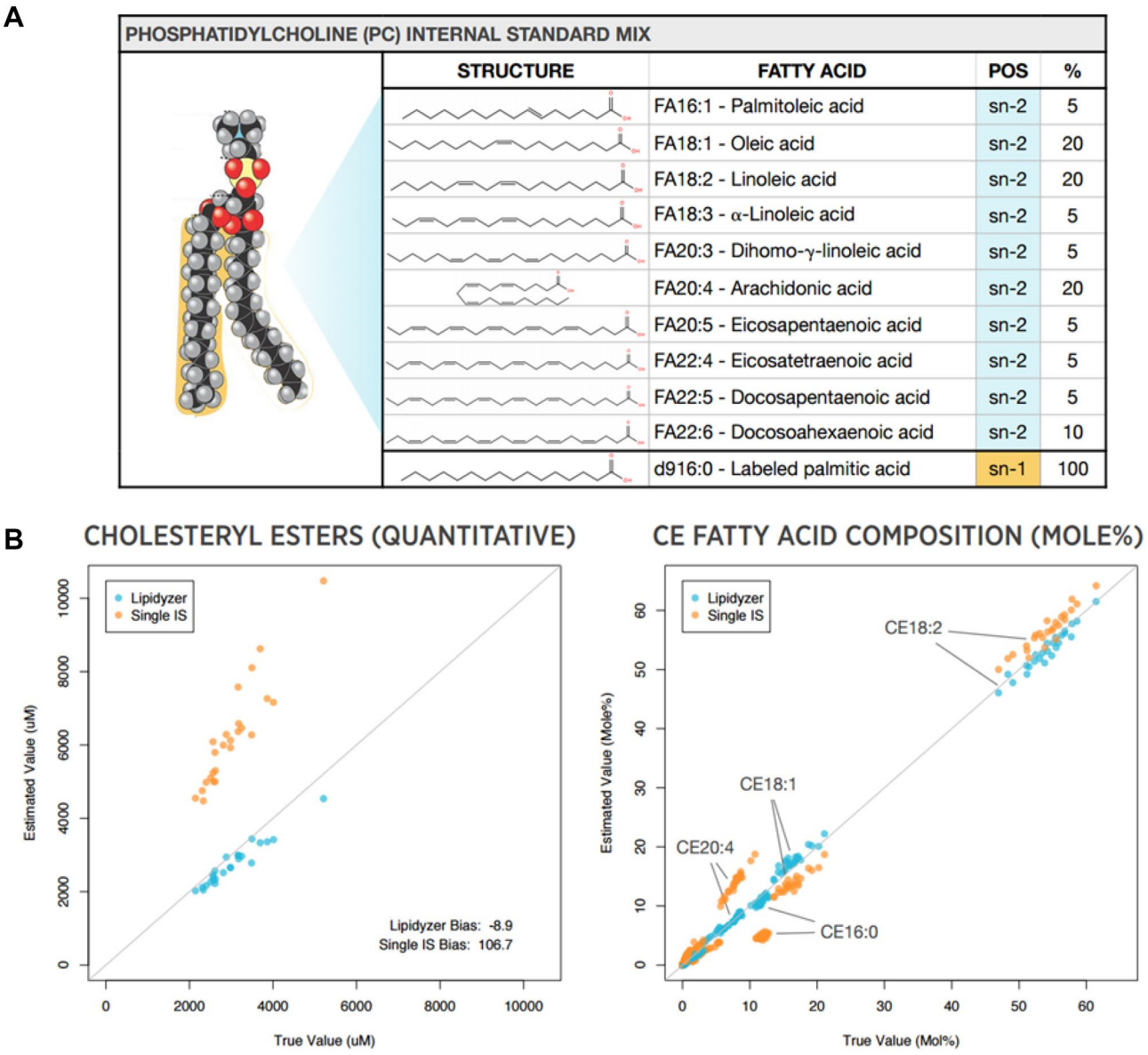


Fig. 5 **A** An example of the Lipidyzer phosphatidylcholine (PC) internal lipid class labeling strategy. The *sn*-1 (top carbon of glycerol backbone) stereospecific numbering position is a labeled palmitate and then the *sn*-2 (middle carbon) position is changed for every fatty acid from a short chain palmitoleic acid to a long chain docosahexaenoic acid. Therefore, there are multiple internal standards to reflect the diversity of the lipid molecular species. The remainder of the 12 lipids classes have a similar strategy. **B** The Lipidyzer™ internal standards (yellow filled circle) were compared to the use of a single internal standards (blue filled circle) for their ability to accurately calibrate the concentration (μM) of total cholesteryl esters (CE) in human serum (left panel). The estimated value (using the current lipidyzer platform) versus true value (known, historical concentrations using an orthogonal LC–MS/MS platform) of the fatty acid composition of cholesteryl esters expressed as mole% fatty acid composition in human serum is also illustrated (right panel)

humans and rodents and NIST SRM 1950 as QC material. System suitability testing was performed prior to sample analysis, and 41 analytes in the LC–MS test mix and 17 analytes in the FIA-MS test mix were used to evaluate instrument performance, including signal abundance, mass accuracy, retention time, and peak shapes. As anticipated, both the intra- and inter-laboratory variance measured in this ring trial were far less that the biological variance observed across the study samples. Intra-laboratory variance was low for all analytes ranging from 5 to 15%, whereas inter-laboratory variance across laboratories was analyte class-dependent (amino acids, cholesteryl esters, sphingolipids, and total hexoses were below 20% median CV; biogenic amines, glycerolipids, and glycerophospholipids were below 25% median CV and acylcarnitines had a median CV of 38%). Ultimately, the AbsoluteIDQ p400HR ring trial

demonstrated that through system suitability testing, SOPs, and RMs, reproducible quantitative metabolomics data could be obtained across different instruments and laboratories. The performance of specific lipidomics platforms was also explored by Siskos et al. (Siskos et al., 2017) in which six different laboratories measured shared materials including SRM 1950 as a harmonization RM with AbsoluteIDQ p180 Biocrates kit. Likewise, most metabolites were observed with interlaboratory variance of below 20% CV (median).

5 Reference libraries and other data harmonization approaches

As described in Sect. 3, biological RMs applied in metabolomic and lipidomic applications contain hundreds or thousands of metabolites present in a complex matrix that are sourced from a range of biological subjects. The mass concentration and chemical identity of some or all metabolites are not known, and for some molecules the stability is also undetermined. As an alternative, reference libraries (RLs) comprised of more limited number of specific metabolites or lipids as reference standards or standard mixtures can be applied that are present and/or detectable in various biological RMs. As synthetically prepared solutions or neat standards, they can also be spiked into biological matrices prior to analysis. These RLs can be applied in different approaches including to aid metabolite quantification, to apply as system suitability test samples or to construct retention time and MS/MS mass spectral libraries for metabolite identification. Some RLs offer library-specific software to find peaks and simplify library construction. A summary of RLs and related standard mixtures currently available is provided in Table 3.

RLs comprised of standard samples or mixtures are constructed with authenticated chemical standards representing naturally occurring metabolites or lipids. RLs can focus on a specific metabolic pathways or processes (e.g., glycolysis/gluconeogenesis kits) or metabolite classes (e.g., amino acids, organic acids, lipids, bile acids), or compounds from more general primary metabolism. These products distributed as individual standards or kits are also provided in Table 3. While some RLs are designed for targeted quantitative applications, all are leveraged by the untargeted metabolomics and lipidomics communities to generate suitable databases for metabolite identification and cross site comparison of chromatographic retention times and reference mass spectra. These authentic compound libraries may be found both as natural abundance and isotopic libraries, depending on need.

Apart from RMs developed specifically for metabolomics, other chemical property standards of wider application may assist metabolomics researchers to improve reporting and harmonization of method and results. As the retention time of analytes is not molecule specific, but instead it is highly dependent on system conditions, reporting retention indices (RIs) have been used as a gold standard in GC analysis for some decades to correct for instrument or column variability. In metabolomics applications, the GC–MS Metabolomics RTL Library (Kind et al., 2009) incorporates RIs based on the use of fatty acid methyl esters (FAMES). Recently the National Research Council Canada developed a reference material (RM-RILC) (NRC, 2021) designed for the measurement of LC RIs based on 20 homologous *N*-alkylpyridium-3 sulfonates in methanol that bear increasing hydrophobicity. Its use in interlaboratory study of five different LC–MS systems was shown to minimize relative deviation and improve cross laboratory comparisons (Quilliam et al., 2015).

Similar developments have been independently reported in untargeted LC–MS analysis from the NORMAN network of reference laboratories (NORMAN, 2021) for monitoring of emerging environmental substances. Overall, analysis of such standards and calculation of RIs for known analytes and for recurrent unknowns shows promise to help method validation, improve comparison of data obtained in different laboratories and thus harmonize reporting, but further collaborative interlaboratory work is needed. Similarly, the EPA's non-targeted analysis (NTA) collaborative trial (ENTACT) (Ulrich et al., 2019) was conducted with the aim to move the untargeted analysis community towards an improved comparison of methods and results, and ultimately to devise performance benchmark standards. The recent formation of the Benchmarking and Publications for Non-Targeted Analysis (BP4NTA) Working Group (BP4NTA, 2021) aims to motivate the NTA community toward competency for proficiency testing. Furthermore, significant challenges exist to reconcile the inconsistencies in compound identification that result from incompleteness of databases and varied results from different data pipelines. The successes (and failures) of the standardization efforts within these corresponding NTA communities may be leveraged and implemented into practice within metabolomics and lipidomics communities.

6 Implementation and outreach

In addition to publication efforts, the mQACC is committed to promote, highlight and disseminate the needs and utility of RMs to the broader metabolomics community through leadership and active participation in interlaboratory studies, cooperation with other consortia, and engagement in symposia and workshops at relevant society meetings and international conferences. The mQACC will also seek to generate online tutorials and videos. With respect to social media, mQACC has a twitter account (https://twitter.com/mQACC) which we shall also use to highlight advances

Table 3 Examples of commercially-available reference libraries and related standards

Reference library/standard mixes	Manufacturer/supplier (product code)	Composition	QC purpose
Glycolysis/gluconeogenesis metabolite library	MilliporeSigma (ML0013)	10 mg Each of 18 glycolysis and gluconeogenesis pathway metabolite neat standards	Chemical identification
Amino acid standard	MilliporeSigma (AAS18)	(1.25–2.5) µM solutions (1 mL) of 17 amino acids	Chemical Identification/system Suitability
Cell free amino acid mixture—13C,15 N	MilliporeSigma (767,964)	(5–100) mM solutions (1 mL) of 20 stable isotope-labeled amino acids	Chemical identification
Mass spectrometry metabolite library	MilliporeSigma/IROA technologies (MSMLS)	600 Metabolites (inc. carboxylic acids, amino acids, biogenic amines, polyamines, nucleotides, coenzymes, vitamins, mono/disaccharides, fatty acids, lipids, steroids, hormones) in 96-well format	Chemical identification
QRESS standard kit	Cambridge isotope laboratories, Inc./SCIEX (MSK-QReSS)	Lyophilized for reconstitution to (2 to 100) µg/mL of 18 stable isotope-labeled metabolites. Companion unlabeled mix separately available	Chemical identification/system suitability and quantification
Amino acid standard mixes	Cambridge isotope laboratories, Inc. (e.g., MSK-CAA)	Lyophilized for reconstitution to (1.25 to 2.5) mM of 20 canonical, stable isotope-labeled amino acids. Companion unlabeled mix separately available	Chemical identification/system suitability and quantification
Organic acid standard mixes	Cambridge isotope laboratories, Inc. (MSK-OA)	Lyophilized for reconstitution to 250 mM of 33 stable isotope-labeled organic acids. Companion unlabeled mix separately available	Chemical identification/system suitability and quantification
Bile acid standard mixes	Cambridge isotope laboratories, Inc. (MSK-BA)	Lyophilized for reconstitution to 100 mM of 16 (6 unconjugated and 10 conjugated) stable isotope-labeled bile acids. Companion unlabeled mix separately available	Chemical identification/system suitability and quantification
SPLASH LIPIDOMIX® Mass spec standard	Avanti polar lipids (330,707)	(2 to 350) µg/mL solutions of 14 lipids	Chemical identification/system suitability/quantification
Human Metabolites V12.0	MetaSci (https://www.metasci.ca/products)	5–10 mg of 1200 metabolites (inc. organic acids, fatty acids and esters, vitamins and co-factors, purine metabolism, amino acids, food additives and components, bacterial (E. coli), plant and fecal sources)	Chemical identification
AbsoluteIDQ p180 and p400 and Quant 500 Kits	Biocrates (MxP, AbsoluteIDQ)	Patented 96-well plate with internal standards, calibrators, QCs and system suitability tests. SOP included as well as software for data processing and interpretation	Chemical Identification and quantification
Lipidyzer Kits	SCIEX (XXXISTLPV-100)	Internal Standards Kit (over 50 labeled standards), SelexION® Tuning Kit (for ion mobility), System Suitability Kit, and QC Spike Standards with Control Plasma Kit (reference material). SOP included	Chemical identification/system suitability and quantification

in standards for quality assurance and quality control, disseminate important QA/QC scientific publications and make recommendations on best practices from mQACC to the metabolomics community.

6.1 Interlaboratory studies

Often referred to as ring trials and round-robin comparisons, interlaboratory studies are useful in allowing metabolomics and lipidomics researchers to assess differences and validate their measurement processes and methodology. Similar to proficiency testing schemes, interlaboratory studies can be used to demonstrate measurement competency, which can be advantageous for QA applications of qualifying recently trained personnel and ensure that a new method performs as anticipated. More importantly, interlaboratory studies that utilize RMs and other QC materials are an effective tool to determine sources of variation or challenges that impact metabolomics measurements and untargeted profiling efforts.

In recent years, several interlaboratory studies have been conducted with SRM 1950 as a QCRM but has been applied for largely targeted metabolomics and lipidomics (Bowden et al., 2017; Cheema et al., 2015; Siskos et al., 2017; Thompson et al., 2019). While designed to define metabolite specific (targeted) reproducibility and accuracy, the results can also inform good laboratory practices that benefit untargeted analysis techniques. Interlaboratory studies have been launched recently (2019) and in cooperation with the mQACC as part of NIST's MetQual program. The primary goals for the interlaboratory study were: (1) to support measurement comparability of untargeted metabolomic profiling in human plasma, (2) assess measurement variability within the untargeted metabolomics community, (3) to evaluate and ascertain a (qualitative) consensus characterization of metabolites present in the candidate RM 8231 plasma suite, and (4) to gather feedback from the untargeted metabolomics community on the potential functions and implementation practicality (fit-for-purpose) nature of the RM suite. Participants agreed to measure metabolomic profiles using the routine metabolomics sample preparation protocols and analytical methods and data acquisition employed by their labs, in addition to using their customary data processing, data curation, and multivariate analysis methodology. The collection of results from the participants and the resultant analysis of the data is ongoing.

Early community-wide harmonization efforts through interlaboratory studies in lipidomics were focused more on broadly defining the lipidome and less on a centralized or accepted workflow (Bowden et al., 2017; Quehenberger et al., 2010). More recently, the Reference Material and Biological Reference Ranges interest group of the International Lipidomics Society (ILS), in collaboration with NIST and Avanti Lipids, launched in January 2020 a series of international ring trials, each one focused on measuring the absolute concentration of different selected lipid species of clinical relevance in RMs. Initially, SRM 1950 and the suite of NIST plasma RMs described in Sect. 3.2 were distributed to approximately 40 laboratories around the world (located in Europe, the Americas, Asia and Australia) that were asked to quantify four ceramides (Cer d18:1/16:0, d18:1/18:0, d18:1/24:0 and d18:1/24:1), based on their clinical utility (PMID: 27,125,947) in the cardiovascular disease field. The first phase of the trials is centered on RMs to establish a global network of laboratories that will then collect samples from different human cohorts in healthy and diseased conditions and measure the same lipids that were quantified previously in RMs. The initial RM evaluation is anticipated to enumerate interlaboratory differences, which provides a critical first step to provide confidence in lipid concentration reference ranges from cohort data and thus drive translation towards more clinical applications. Even though these are designed and evaluated as targeted interlaboratory studies, the aim is to pinpoint specific aspects of workflows that cause non-agreement in community results, which will ultimately improve untargeted lipidomics applications.

In addition to the use of common RMs across metabolomics and lipidomics laboratories, the knowledge gained by these community-based activities can foster the development of best practice guidelines, and ultimately reduce measurement variance within metabolomics and lipidomics. Given the foremost goal of clinical translation of scientific findings, these efforts are anticipated to significantly reduce measurement variance while improving community measurement agreement, thus facilitating the fields of metabolomics and lipidomics to reach their full potential of producing results of clinical significance.

6.2 Stakeholder outreach

A significant driver is also to inform stakeholders and provide educational outreach on current and prospective RMs and associated QA/QC tools available for safeguarding untargeted metabolomics and lipidomics measurements. This effort is aimed at both newcomers as well as experienced analysts and researchers. For example, mQACC members that are also a part of the ILS Reference Materials and Biological Reference Ranges working group engage and support the researchers in this field to implement the use of RMs as common practice. The initial effort is focused on the correct use of SRM 1950 and the new plasma RM suite described previously; further activities will promote the characterization of other materials, such as urine or specific lipid classes (e.g., bile acids in stools, triacylglycerol in liver). Corresponding guidelines for standardization, method validation and reporting of lipid molecules are currently

being developed within ILS as community-wide effort (Liebisch et al., 2019). The development and implementation of novel RMs with matrix-specific lipidomic profiles and the identification of novel lipids will also have important implications in agriculture and nutrition applications.

Other stakeholder clinical organizations and working groups such as the International Federation of Clinical Chemistry (IFCC) Scientific Division also support the development of RMs and reference measurement procedures that are traceable to these materials to promote standardization in laboratory medicine. There is a keen interest in RMs that can be easily integrated into clinical application workflows to allow for reliable clinical biomarker measurements. However, the development of appropriate RMs for clinical applications requires the collaboration of clinical researchers in the field of untargeted metabolomics to ensure their suitability in clinical practice. Because differential metabolite and lipid signatures can be caused by a disease state, there may be some interest in the development of RMs that represent the pooled serum and/or plasma from chronic disease patients (i.e., patients suffering from autoimmune diseases, diabetes, cancer, kidney disease, cardiovascular disease, etc.). A defined or standardized biological matrix in the form of a commutable RM will also aid in finding biomarker signatures to help elucidate clinical questions and provide reliability in clinical measurements established methods (Dias and Koal, 2016). The issue of commutability (or equivalence through an established method) between the applied matrix-matched RM to the representative metabolomics test samples always remains.

Lastly and equally as important, mass spectrometry vendor and academic research organizations who are at the forefront of developing and innovating new technology and applications have a major stake in the implementation of reference materials, reference standards and system suitability test mixtures. These organizations utilize such materials and products for standardization of the technology and methodologies, as well as validation and verification at product development stages for a range of applications. Collaborative engagement between commercial reagents and consumables manufacturers and metabolomic researchers and leaders will enable technological solutions to address the many challenges described here.

The descriptions and associated references provided in this review offer a foundational knowledge on available QA/QC tools and best practices to aid the untargeted metabolomics community. Some might prefer didactic learning methods to the experiential (self-taught) approach, thus engagements with experts or hands on training may better instill the justification of the suggested practices. Such opportunities include a specific QA/QC in Metabolomics course by the University of Birmingham Training Center and topic sections within general metabolomic training courses offered by institutions such as European Bioinformatics Institute (EMBL-EBI) or the NIH Metabolomic Program Regional Comprehensive Metabolomics Resource Cores (RCMRCs). Workshops, webinars [e.g., Chemical & Engineering News webcast of a QA/QC Practices in Untargeted Metabolomics (C&EN, 2020)] and any other recorded materials also offer exposure to a broad audience and are additional opportunities for training and engagement on the appropriate usage of RMs and the current state of QA/QC practices. The mQACC Best Practices working group also began hosting a series of interactive discussion workshops (Metabolomics Association of North America (MANA) 2019, with the Human Health Exposure Analysis Resource (HHEAR), focusing on current practices of QA/QC regarding topics such as pooled QC use and system suitability testing.

7 Summary

We have highlighted numerous and diverse RMs from well characterized mixtures of authentic standards of complex and less well characterized biosamples that are currently available for use, in a variety of ways, as QC tools to help ensure that metabolomic and lipidomic research is as robust and reproducible as possible. Clearly numerous challenges and barriers still exist however with the continued engagement of the metabolomics community, the development of new RMs, commercially available chemical standards and reference data products will help to address them. This will require the adoption of agreed methods for their use, supplemented through well-designed SOPs which will further promote sound QA/QC practices.

8 Conclusions

RM's, whether libraries of standards, mixtures or biologically based material represent a useful source of materials that can be used to calibrate, standardize, and compare the results obtained by different laboratories. Whilst they do not represent total solutions to the many problems posed by the attempt to use untargeted metabolic phenotyping for the comprehensive characterization of the metabolomes of all of the systems being studied they nevertheless represent an important tool in the armament of the analyst undertake such work. Here we have reviewed the currently available materials and indicated some of their applications.

Acknowledgements We are grateful to Krista Zanetti of the National Institutes of Health (NIH) National Cancer Institute (NCI) for her contributions to the mQACC Reference and Test Material Working Group since its inception. We are also grateful to other members of the

mQACC that have provided additional reviews. The opinions expressed in this manuscript do not necessarily represent those of the National Institute of Standards and Technology, the Centers for Disease Control and Prevention, the National Institutes of Health or the Food and Drug Administration. Certain commercial equipment, instruments, software, or materials are identified in this paper in order to specify the experimental procedure or describe the results adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, the Centers for Disease Control and Prevention, the National Institutes of Health or the Food and Drug Administration, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

Author contributions All authors contributed specific text and/or figure content to the manuscript. All authors read, edited, and approved the manuscript. KAL and BKU produced and copyedited the manuscript content.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Acera, A., Pereiro, X., Abad-Garcia, B., Rueda, Y., Ruzafa, N., Santiago, C., Barbolla, I., Duran, J. A., Ochoa, B., & Vecino, E. (2019). A simple and reproducible method for quantification of human tear lipids with ultrahigh-performance liquid chromatography-mass spectrometry. *Molecular Vision, 25*, 934–948.

Aden, K., Rehman, A., Waschina, S., Pan, W. H., Walker, A., Lucio, M., Nunez, A. M., Bharti, R., Zimmerman, J., Bethge, J., Schulte, B., Schulte, D., Franke, A., Nikolaus, S., Schroeder, J. O., Vandeputte, D., Raes, J., Szymczak, S., Waetzig, G. H., & Rosenstiel, P. (2019). Metabolic functions of gut microbes associate with efficacy of tumor necrosis factor antagonists in patients with inflammatory bowel diseases. *Gastroenterology, 157*, 1279-1292. e11. https://doi.org/10.1053/j.gastro.2019.07.025

Agatonovic-Kustrin, S., Morton, D. W., Smirnov, V., Petukhov, A., Gegechkori, V., Kuzina, V. N., Gorpinchenko, N., & Ramenskaya, G. (2019). Analytical strategies in lipidomics for discovery of functional biomarkers from human saliva. *Disease Markers, 2019*, 6741518. https://doi.org/10.1155/2019/6741518

Alseekh, S., Aharoni, A., Brotman, Y., Contrepois, K., D'Auria, J., Ewald, J., Ewald, C. J., Fraser, P. D., Giavalisco, P., Hall, R. D., Heinemann, M., Link, H., Luo, J., Neumann, S., Nielsen, J., Perez Souza, L., Saito, K., Sauer, U., Schroeder, F. C.,… Fernie, A. R. (2021). Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nature Methods, 18*(7), 747–756. https://doi.org/10.1038/s41592-021-01197-1

Aristizabal-Henao, J. J., Jones, C. M., Lippa, K. A., & Bowden, J. A. (2020). Nontargeted lipidomics of novel human plasma reference materials: Hypertriglyceridemic, diabetic, and African-American. *Analytical and Bioanalytical Chemistry, 412*, 7373–7380. https://doi.org/10.1007/s00216-020-02910-3

Azab, S., Ly, R., & Britz-McKibbin, P. (2019). Robust method for high-throughput screening of fatty acids by multisegment injection-nonaqueous capillary electrophoresis-mass spectrometry with stringent quality control. *Analytical Chemistry, 91*, 2329–2336. https://doi.org/10.1021/acs.analchem.8b05054

Barri, T., Holmer-Jensen, J., Hermansen, K., & Dragsted, L. O. (2012). Metabolic fingerprinting of high-fat plasma samples processed by centrifugation- and filtration-based protein precipitation delineates significant differences in metabolite information coverage. *Analytica Chimica Acta, 718*, 47–57. https://doi.org/10.1016/j.aca.2011.12.065

Bearden, D. W., Sheen, D. A., Simón-Manso, Y., Benner, B. A., Jr, Rocha, W., Blonder, N., Lippa, K. A., Beger, R. D., Schnackenberg, L. K., Sun, J., Mehta, K. Y., Cheema, A. K., Gu, H., Marupaka, R., Nagana Gowda, G. A., & Raftery, D. (2019). Metabolomics test materials for quality control: A study of a urine materials suite. *Metabolites*. https://doi.org/10.3390/metabo9110270

Beger, R. D., Dunn, W. B., Bandukwala, A., Bethan, B., Broadhurst, D., Clish, C. B., Dasari, S., Derr, L., Evans, A., Fischer, S., Flynn, T., Hartung, T., Herrington, D., Higashi, R., Hsu, P. C., Jones, C., Kachman, M., Karuso, H., Kruppa, G., Lippa, K., … Zanetti, K. A. (2019). Towards quality assurance and quality control in untargeted metabolomics studies. *Metabolomics, 15*, 4. https://doi.org/10.1007/s11306-018-1460-7

Bijlsma, S., Bobeldijk, I., Verheij, E. R., Ramaker, R., Kochhar, S., Macdonald, I. A., van Ommen, B., & Smilde, A. K. (2006). Large-scale human metabolomics studies: A strategy for data (Pre-) processing and validation. *Analytical Chemistry*. https://doi.org/10.1021/ac051495j

Bosch, S., El Hassani, S. M., Covington, J. A., Wicaksono, A. N., Bomers, M. K., Benninga, M. A., Mulder, C. J. J., Boer, N. K. H., & Meij, T. G. J. (2018). Optimized sampling conditions for fecal volatile organic compound analysis by means of field asymmetric ion mobility spectrometry. *Analytical Chemistry, 90*, 7972–7981. https://doi.org/10.1021/acs.analchem.8b00688

Bouhifd, M., Beger, R., Flynn, T., Guo, L., Harris, G., Hogberg, H., Kaddurah-Daouk, R., Kamp, H., Kleensang, A., Maertens, A., Odwin-DaCosta, S., Pamies, D., Robertson, D., Smirnova, L., Sun, J., Zhao, L., & Hartung, T. (2015). Quality assurance of metabolomics. *Altex, 32*, 319–326. https://doi.org/10.14573/altex.1509161

Bowden, J. A., Heckert, A., Ulmer, C. Z., Jones, C. M., Koelmel, J. P., Abdullah, L., Ahonen, L., Alnouti, Y., Armando, A. M., Asara, J. M., Bamba, T., Barr, J. R., Bergquist, J., Borchers, C. H., Brandsma, J., Breitkopf, S. B., Cajka, T., Cazenave-Gassiot, A., Checa, A., Cinel, M. A., … Zhou, S. (2017). Harmonizing lipidomics: NIST interlaboratory comparison exercise for lipidomics using SRM 1950-metabolites in frozen human plasma. *Journal of Lipid Research, 58*, 2275–2288. https://doi.org/10.1194/jlr.M079012

Bowden, J. A., Ulmer, C. Z., Jones, C. M., Koelmel, J. P., & Yost, R. A. (2018). NIST lipidomics workflow questionnaire: An assessment of community-wide methodologies and perspectives. *Metabolomics, 14*, 53. https://doi.org/10.1007/s11306-018-1340-1

BP4NTA (2021). Benchmarking and publications fro non-targeted analysis. https://nontargetedanalysis.org. Accessed August 2021.

Briscoe, C. J., Stiles, M. R., & Hage, D. S. (2007). System suitability in bioanalytical LC/MS/MS. *Journal of Pharmaceutical and Biomedical Analysis, 44*, 484–491. https://doi.org/10.1016/j.jpba.2007.03.003

Broadhurst, D., Goodacre, R., Reinke, S. N., Kuligowski, J., Wilson, I. D., Lewis, M. R., & Dunn, W. B. (2018). Guidelines and

considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics, 14*, 72. https://doi.org/10.1007/s11306-018-1367-3

Brown, E. M., Ke, X., Hitchcock, D., Jeanfavre, S., Avila-Pacheco, J., Nakata, T., Arthur, T. D., Fornelos, N., Heim, C., Franzosa, E. A., Watson, N., Huttenhower, C., Haiser, H. J., Dillow, G., Graham, D. B., Finlay, B. B., Kostic, A. D., Porter, J. A., Vlamakis, H., Clish, C. B., … Xavier, R. J. (2019). Bacteroides-derived sphingolipids are critical for maintaining intestinal homeostasis and symbiosis. *Cell Host & Microbe, 25*, 668-680.e7. https://doi.org/10.1016/j.chom.2019.04.002

Bruderer, T., Gaisl, T., Gaugg, M. T., Nowak, N., Streckenbach, B., Müller, S., Moeller, A., Kohler, M., & Zenobi R. (2019). Online analysis of exhaled breath focus review. *Chemical Reviews, 119*, 10803–10828. https://doi.org/10.1021/acs.chemrev.9b00005

C&EN (2020). Chemical & engineering news, quality assurance and quality control practices in untargeted metabolomics. Thermo fisher scientific. https://cen.acs.org/media/webinar/thermo_040220.html?utm_source=Webinar&utm_medium=Webinar&utm_campaign=CEN. Accessed August 2021.

Cajka, T., Smilowitz, J. T., & Fiehn, O. (2017). Validating quantitative untargeted lipidomics across nine liquid chromatography-high-resolution mass spectrometry platforms. *Analytical Chemistry, 89*, 12360–12368. https://doi.org/10.1021/acs.analchem.7b03404

Cao, Z., Liu, J., Xie, X., Zhan, S., Song, W., Wu, S., Sun, Z., Dong, Y., Tang, G., Liu, Y., Li, L., Shen, M., Zhai, Y., Zou, J., & Liu, X. (2020). Lipidomic profiling of amniotic fluid and its application in fetal lung maturity prediction. *Journal of Clinical Laboratory Analysis, 34*, e23109. https://doi.org/10.1002/jcla.23109

Cheema, A. K., Asara, J. M., Wang, Y., Neubert, T. A., Tolstikov, V., & Turck, C. W. (2015). The ABRF metabolomics research group 2013 study: Investigation of spiked compound differences in a human plasma matrix. *Journal of biomolecular techniques*, *26*(3), 83–89. https://doi.org/10.7171/jbt.15-2603-001

CIL (2021). Cambridge isotope laboratories, Inc. QRESS Kits In. https://shop.isotope.com/advancedsearchresults.aspx?keyword2=QRESS&searchType=ALL+Keywords&x=0&y=0. Accessed June 2021

Colas, R. A., Shinohara, M., Dalli, J., Chiang, N., & Serhan, C. N. (2014). Identification and signature profiles for pro-resolving and inflammatory lipid mediators in human tissue. *American Journal of Physiology Cell Physiology, 307*, C39–C54. https://doi.org/10.1152/ajpcell.00024.2014

Contrepois, K., Mahmoudi, S., Ubhi, B. K., Papsdorf, K., Hornburg, D., Brunet, A., & Snyder, M. (2018). Cross-platform comparison of untargeted and targeted lipidomics approaches on aging mouse plasma. *Science and Reports, 8*, 17747. https://doi.org/10.1038/s41598-018-35807-4

Cui, M., Trimigno, A., Aru, V., Khakimov, B., & Engelsen, S. B. (2020). Human faecal (1)H NMR metabolomics: Evaluation of solvent and sample processing on coverage and reproducibility of signature metabolites. *Analytical Chemistry, 92*, 9546–9555. https://doi.org/10.1021/acs.analchem.0c00606

Davis, W., Neely, B., Schock, T., LE, K., & Ellisor D., (2019) Identification commutability in proteomics and metabolomics utilizing human tissue reference materials. In: ASMS 67th conference on mass spectrometry and allied topics, Atlanta, GA.

Davis, W., Neely, B. A., Kilpatrick, L. E., & Schock, T. B., (2019) Development of human liver quality control materials for metabolomics and proteomics, metabolomics association of North America. In: Metabolomics association of North America, Atlanta, GA.

Davis, W. C., Kilpatrick, L. E., Ellisor, D. L., & Neely, B. A. (2019c). Characterization of a human liver reference material fit for proteomics applications. *Scientific Data, 6*, 324. https://doi.org/10.1038/s41597-019-0336-7

Deda, O., Chatziioannou, A. C., Fasoula, S., Palachanis, D., Raikos, N., Theodoridis, G. A., & Gika, H. G. (2017). Sample preparation optimization in fecal metabolic profiling. *Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences, 1047*, 115–123. https://doi.org/10.1016/j.jchromb.2016.06.047

Di Giovanni, N., Meuwis, M. A., Louis, E., & Focant, J. F. (2020). Untargeted serum metabolic profiling by comprehensive two-dimensional gas chromatography-high-resolution time-of-flight mass spectrometry. *Journal of Proteome Research, 19*, 1013–1028. https://doi.org/10.1021/acs.jproteome.9b00535

Dias, D. A., & Koal, T. (2016). Progress in metabolomics standardisation and its significance in future clinical laboratory medicine. *Ejifcc, 27*, 331–343.

Dias, I. H. K., Wilson, S. R., & Roberg-Larsen, H. (2018). Chromatography of oxysterols. *Biochimie, 153*, 3–12. https://doi.org/10.1016/j.biochi.2018.05.004

Drotleff, B., Illison, J., Schlotterbeck, J., Lukowski, R., & Lämmerhofer, M. (2019). Comprehensive lipidomics of mouse plasma using class-specific surrogate calibrants and SWATH acquisition for large-scale lipid quantification in untargeted analysis. *Analytica Chimica Acta, 1086*, 90–102. https://doi.org/10.1016/j.aca.2019.08.030

Dudzik, D., Barbas-Bernardos, C., Garcia, A., & Barbas, C. (2018). Quality assurance procedures for mass spectrometry untargeted metabolomics. A review. *Journal of Pharmaceutical and Biomedical Analysis, 147*, 149–173. https://doi.org/10.1016/j.jpba.2017.07.044

Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., Brown, M., Knowles, J. D., Halsall, A., Haselden, J. N., Nicholls, A. W., Wilson, I. D., Kell, D. B., & Goodacre, R. (2011a). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols, 6*, 1060–1083. https://doi.org/10.1038/nprot.2011.335

Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., Brown, M., Knowles, J. D., Halsall, A., Haselden, J. N., Nicholls, A. W., Wilson, I. D., Kell, D. B., & Goodacre, R. (2011). 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. *Nature Protocols*. https://doi.org/10.1038/nprot.2011.335

Evans, A. M., DeHaven, C. D., Barrett, T., Mitchell, M., & Milgram, E. (2009). Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Analytical Chemistry, 81*, 6656–6667. https://doi.org/10.1021/ac901536h

Evans, A. M., O'Donovan, C., Playdon, M., Beecher, C., Beger, R. D., Bowden, J. A., Broadhurst, D., Clish, C. B., Dasari, S., Dunn, W. B., Griffin, J. L., Hartung, T., Hsu, P. C., Huan, T., Jans, J., Jones, C. M., Kachman, M., Kleensang, A., Lewis, M. R., Monge, M. E., Mosley, J.D., Taylor, E., Tayyari, F., Theodoridis, G., Torta, F., Ubhi, B. K., & Vuckovic, D. (2020). Dissemination and analysis of the quality assurance (QA) and quality control (QC) practices of LC-MS based untargeted metabolomics practitioners. *Metabolomics, 16*, 113. https://doi.org/10.1007/s11306-020-01728-5

Gardner, A., Carpenter, G., & So, P. W. (2020). Salivary metabolomics: From diagnostic biomarker discovery to investigating biological function. *Metabolites*. https://doi.org/10.3390/metabo10020047

George, A. D., Gay, M. C. L., Trengove, R. D., & Geddes, D. T. (2018). Human milk lipidomics: Current techniques and methodologies. *Nutrients*. https://doi.org/10.3390/nu10091169

Gika, H. G., Theodoridis, G. A., Earll, M., Snyder, R. W., Sumner, S. J., & Wilson, I. (2010). Does the mass spectrometer define the marker? A comparison of global metabolite profiling data generated simultaneously via UPLC-MS on two different mass spectrometers. *Analytical Chemistry, 82*, 8226–8234. https://doi.org/10.1021/ac1016612

Gika, H. G., Theodoridis, G. A., Earll, M., & Wilson, I. D. (2012). 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*, 2239–2247. https://doi.org/10.4155/bio.12.212

Gika, H. G., Theodoridis, G. A., Wingate, J. E., & Wilson, I. D. (2007). Within-day reproducibility of an HPLC−MS-based method for metabonomic analysis: Application to human urine. *Journal of Proteome Research, 6*, 3291–3303. https://doi.org/10.1021/pr070183p

Gika, H. G., Zisi, C., Theodoridis, G., & Wilson, I. D. (2016). Protocol for quality control in metabolic profiling of biological fluids by U(H)PLC-MS. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences, 1008*, 15–25. https://doi.org/10.1016/j.jchromb.2015.10.045

González-Riano, C., Dudzik, D., Garcia, A., Gil-de-la-Fuente, A., Gradillas, A., Godzien, J., López-Gonzálvez, A., Rey-Stolle, F., Rojo, D., Ruperez, F. J., Saiz, J., & Barbas, C. (2020). Recent developments along the analytical process for metabolomics workflows. *Analytical Chemistry, 92*, 203–226. https://doi.org/10.1021/acs.analchem.9b04553

Gouveia, G. J., Shaver, A. O., Garcia, B. M., Morse, A. M., Andersen, E. C., Edison, A. S., & McIntyre, L. M. (2021). Long-term metabolomics reference material. bioRxiv. 436834. https://doi.org/10.1101/2021.03.24.436834

Gregory, K. E., Bird, S. S., Gross, V. S., Marur, V. R., Lazarev, A. V., Walker, W. A., & Kristal, B. S. (2013). Method development for fecal lipidomics profiling. *Analytical Chemistry, 85*, 1114–1123. https://doi.org/10.1021/ac303011k

Herbig, J., & Beauchamp, J. (2014). Towards standardization in the analysis of breath gas volatiles. *Journal of Breath Research, 8*, 037101. https://doi.org/10.1088/1752-7155/8/3/037101

Hermann, G., Schwaiger, M., Volejnik, P., & Koellensperger, G. (2018). (13)C-labelled yeast as internal standard for LC-MS/MS and LC high resolution MS based amino acid quantification in human plasma. *Journal of Pharmaceutical and Biomedical Analysis, 155*, 329–334. https://doi.org/10.1016/j.jpba.2018.03.050

Hoffmann, S., Edler, L., Gardner, I., Gribaldo, L., Hartung, T., Klein, C., Liebsch, M., Sauerland, S., Schechtman, L., Stammati, A., & Nikolaidis, E. (2008). Points of reference in the validation process: The report and recommendations of ECVAM workshop 66a. *Alternatives to Laboratory Animals, 36*, 343–352. https://doi.org/10.1177/026119290803600311

Höfner, L., Luther, A. M., Palladini, A., Fröhlich, T., & Waberski, D. (2020). Tolerance of stored boar spermatozoa to autologous seminal plasma: A proteomic and lipidomic approach. *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms21186474

ISO. (2016). *International organization for standardization. ISO 17034: General requirements for the competence of reference material producers*. Geneva: Routledge.

ISO (2017). ISO GUIDE 35: Reference materials—guidance for characterization and assessment of homogeneity and stability. NIST 2017-08 (105).

ISO (2021). International organization for standardization. In http://www.iso.org/iso/iso_catalogue/. Accessed June 2021.

Jain, A., Li, X. H., & Chen, W. N. (2019). An untargeted fecal and urine metabolomics analysis of the interplay between the gut microbiome, diet and human metabolism in Indian and Chinese adults. *Science and Reports, 9*, 9191. https://doi.org/10.1038/s41598-019-45640-y

JCGM (2012). Joint committee for guides in metrology (JCGM), International vocabulary of metrology—basic and general concepts and associated terms (VIM). Joint Committee for Guides in Metrology committee a part of the Bureau International des Poids et Mesures.

Jensen, R. G. (1996). The lipids in human milk. *Progress in Lipid Research, 35*, 53–92. https://doi.org/10.1016/0163-7827(95)00010-0

Kim, M., Vogtmann, E., Ahlquist, D. A., Devens, M. E., Kisiel, J. B., Taylor, W. R., White, B. A., Hale, V. L., Sung, J., Chia, N., Sinha, R., & Chen, J. (2020). Fecal metabolomic signatures in colorectal adenoma patients are associated with gut microbiota and early events of colorectal cancer pathogenesis. *mBio*. https://doi.org/10.1128/mBio.03186-19

Kind, T., Wohlgemuth, G., Lee, D. Y., Lu, Y., Palazoglu, M., Shahbaz, S., & Fiehn, O. (2009). FiehnLib: Mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Analytical Chemistry, 81*, 10038–10048. https://doi.org/10.1021/ac9019522

Koelmel, J. P., Li, X., Stow, S. M., Sartain, M. J., Murali, A., Kemperman, R., Tsugawa, H., Takahashi, M., Vasiliou, V., Bowden, J. A., Yost, R. A., Garrett, T. J., & Kitagawa, N. (2020). Lipid annotator: Towards accurate annotation in non-targeted liquid chromatography high-resolution tandem mass spectrometry (LC-HRMS/MS) lipidomics using a rapid and user-friendly software. *Metabolites*. https://doi.org/10.3390/metabo10030101

Lange, M., & Fedorova, M. (2020). Evaluation of lipid quantification accuracy using HILIC and RPLC MS on the example of NIST® SRM® 1950 metabolites in human plasma. *Analytical and Bioanalytical Chemistry, 412*, 3573–3584. https://doi.org/10.1007/s00216-020-02576-x

Lawal, O., Ahmed, W. M., Nijsen, T. M. E., Goodacre, R., & Fowler, S. J. (2017). Exhaled breath analysis: A review of "breath-taking" methods for off-line analysis. *Metabolomics, 13*, 110. https://doi.org/10.1007/s11306-017-1241-8

Leimer, E. M., Pappan, K. L., Nettles, D. L., Bell, R. D., Easley, M. E., Olson, S. A., Setton, L. A., & Adams, S. B. (2017). Lipid profile of human synovial fluid following intra-articular ankle fracture. *Journal of Orthopaedic Research, 35*, 657–666. https://doi.org/10.1002/jor.23217

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

Liebisch, G., Ahrends, R., Arita, M., Arita, M., Bowden, J. A., Ejsing, C. S., Griffiths, W. J., Holčapek, M., Köfeler, H., Harald, M., Mitchell, T. W., Wenk, M. R., Ekroos, K., Lipidomics Standards Initiative Consortium. (2019). Lipidomics needs more standardization. *Nature Metabolism, 1*, 745–747. https://doi.org/10.1038/s42255-019-0094-z

Lin, Y., Caldwell, G. W., Li, Y., Lang, W., & Masucci, J. (2020). Inter-laboratory reproducibility of an untargeted metabolomics GC-MS assay for analysis of human plasma. *Science and Reports, 10*, 10918. https://doi.org/10.1038/s41598-020-67939-x

Liu, K. H., Nellis, M., Uppal, K., Ma, C., Tran, V., Liang, Y., Walker, D. I., & Jones, D. P. (2020). Reference standardization for quantification and harmonization of large-scale metabolomics.

Analytical Chemistry, 92, 8836–8844. https://doi.org/10.1021/acs.analchem.0c00338

Lloyd-Price, J., Arze, C., Ananthakrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W., Andrews, E., Ajami, N. J., Bonham, K. S., Brislawn, C. J., Casero, D., Courtney, H., Gonzalez, A., Graeber, T. G., Hall, A. B., Lake, K., Landers, C. J., Mallick, H., Plichta, D. R., Prasad, M., … Huttenhower, C. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature, 569*, 655–662. https://doi.org/10.1038/s41586-019-1237-9

Mandal, R., Cano, R., Davis, C. D., Hayashi, D., Jackson, S. A., Jones, C. M., Lampe, J. W., Latulippe, M. E., Lin, N. J., Lippa, K. A., Piotrowski, P., Da Silva, S. M., Swanson, K. S., & Wishart, D. S. (2020). Workshop report: Toward the development of a human whole stool reference material for metabolomic and metagenomic gut microbiome measurements. *Metabolomics, 16*, 119. https://doi.org/10.1007/s11306-020-01744-5

Mars, R., Yang, Y., Ward, T., Houtti, M., Priya, S., Lekatz, H. R., Tang, X., Sun, Z., Kalari, K. R., Korem, T., Bhattarai, Y., Zheng, T., Bar, N., Frost, G., Johnson, A. J., van Treuren, W., Han, S., Ordog, T., Grover, M., Sonnenburg, J., … Kashyap, P. C. (2020). Longitudinal multi-omics reveals subset-specific mechanisms underlying irritable bowel syndrome. *Cell, 182*, 1460-1473.e17. https://doi.org/10.1016/j.cell.2020.08.007

Metherel, A. H., Harris, W. S., Ge, L., Gibson, R. A., Chouinard-Watkins, R., Bazinet, R. P., Liu, L., Brenna, J. T., Aristizabal-Henao, J. J., Stark, K. D., & Block, R. C. (2019). Interlaboratory assessment of dried blood spot fatty acid compositions. *Lipids, 54*, 755–761. https://doi.org/10.1002/lipd.12203

Misra, B. B., & Olivier, M. (2020). High resolution GC-Orbitrap-MS metabolomics using both electron ionization and chemical ionization for analysis of human plasma. *Journal of Proteome Research, 19*, 2717–2731. https://doi.org/10.1021/acs.jproteome.9b00774

Moosmang, S., Pitscheider, M., Sturm, S., Seger, C., Tilg, H., Halabalaki, M., & Stuppner, H. (2019). Metabolomic analysis-addressing NMR and LC-MS related problems in human feces sample preparation. *Clinica Chimica Acta, 489*, 169–176. https://doi.org/10.1016/j.cca.2017.10.029

Munafò, M. R., Nosek, B. A., Bishop, D., Button, K. S., Chambers, C. D., du Sert, N. P., Simonsohn, U., Wagenmakers, E. J., Ware, J. J., & Ioannidis, J. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*, 0021. https://doi.org/10.1038/s41562-016-0021

Naz, S., Garcia, A., & Barbas, C. (2013). Multiplatform analytical methodology for metabolic fingerprinting of lung tissue. *Analytical Chemistry, 85*, 10941–10948. https://doi.org/10.1021/ac402411n

Nichols, C. M., Dodds, J. N., Rose, B. S., Picache, J. A., Morris, C. B., Codreanu, S. G., May, J. C., Sherrod, S. D., & McLean, J. A. (2018). Untargeted molecular discovery in primary metabolism: Collision cross section as a molecular descriptor in ion mobility-mass spectrometry. *Analytical Chemistry, 90*, 14484–14492. https://doi.org/10.1021/acs.analchem.8b04322

Nilsson, A. K., Sjöbom, U., Christenson, K., & Hellström, A. (2019). Lipid profiling of suction blister fluid: Comparison of lipids in interstitial fluid and plasma. *Lipids in Health and Disease, 18*, 164. https://doi.org/10.1186/s12944-019-1107-3

NIST (2020). NIST20: Updates to the NIST tandem and electron ionization spectral libraries.

NIST (2021). National institute of standards and technology, gut microbiome metabolomics. In. https://www.nist.gov/programs-projects/gut-microbiome-metabolomics-interlaboratory-program. Accessed June 2021.

NORMAN (2021). Network of reference laboratories, research centres and related organisations for monitoring of emerging environmental substances. In. https://www.norman-network.net/. Accessed June 2021.

NRC (2021). National research council Canada, RM-RILC. In. https://nrc.canada.ca/en/certifications-evaluations-standards/certified-reference-materials/list/115/html. Accessed June 2021.

Pandher, R., Ducruix, C., Eccles, S. A., & Raynaud, F. I. (2009). Cross-platform Q-TOF validation of global exo-metabolomic analysis: Application to human glioblastoma cells treated with the standard PI 3-kinase inhibitor LY294002. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences, 877*, 1352–1358. https://doi.org/10.1016/j.jchromb.2008.12.001

Papadimitropoulos, M.-E.P., Vasilopoulou, C. G., Maga-Nteve, C., & Klapa, M. I. (2018). Untargeted GC-MS metabolomics. In G. A. Theodoridis, H. G. Gika, & I. D. Wilson (Eds.), *Metabolic profiling: Methods and protocols* (pp. 133–147). New York: Springer.

Pereira, H., Martin, J.-F., Joly, C., Sébédio, J.-L., & Pujos-Guillot, E. (2010). Development and validation of a UPLC/MS method for a nutritional metabolomic study of human plasma. *Metabolomics, 6*, 207–218. https://doi.org/10.1007/s11306-009-0188-9

Phinney, K. W., Ballihaut, G., Bedner, M., Benford, B. S., Camara, J. E., Christopher, S. J., Davis, W. C., Dodder, N. G., Eppe, G., Lang, B. E., Long, S. E., Lowenthal, M. S., McGaw, E. A., Murphy, K. E., Nelson, B. C., Prendergast, J. L., Reiner, J. L., Rimmer, C. A., Sander, L. C., Schantz, M. M., … Castle, A. L. (2013). Development of a standard reference material for metabolomics research. *Analytical Chemistry, 85*, 11732–11738. https://doi.org/10.1021/ac402689t

Quehenberger, O., Armando, A. M., Brown, A. H., Milne, S. B., Myers, D. S., Merrill, A. H., Bandyopadhyay, S., Jones, K. N., Kelly, S., Shaner, R. L., Sullards, C. M., Wang, E., Murphy, R. C., Barkley, R. M., Leiker, T. J., Raetz, C. R., Guan, Z., Laird, G. M., Six, D. A., Russell, D. W., … Dennis, E. A. (2010). Lipidomics reveals a remarkable diversity of lipids in human plasma. *Journal of Lipid Research, 51*, 3299–3305. https://doi.org/10.1194/jlr.M009449

Quilliam, M., Bekri, K., McNamara, C., Giddings, S., & Hui, J., (2015). A new retention index system for liquid chromatography-mass spectrometry. In Conference: 42nd international symposium on high performance liquid phase separations and related techniques (HPLC2015), Geneva, Switzerland.

Rampler, E., Criscuolo, A., Zeller, M., El Abiead, Y., Schoeny, H., Hermann, G., Sokol, E., Cook, K., Peake, D. A., Delanghe, B., & Koellensperger, G. (2018). A novel lipidomics workflow for improved human plasma identification and quantification using RPLC-MSn methods and isotope dilution strategies. *Analytical Chemistry, 90*, 6494–6501. https://doi.org/10.1021/acs.analchem.7b05382

Rathod, R., Gajera, B., Nazir, K., Wallenius, J., & Velagapudi, V. (2020). Simultaneous measurement of tricarboxylic acid cycle intermediates in different biological matrices using liquid chromatography-tandem mass spectrometry; quantitation and comparison of TCA cycle intermediates in human serum, plasma, kasumi-1 cell and murine liver tissue. *Metabolites*. https://doi.org/10.3390/metabo10030103

Rattray, N. J., Hamrang, Z., Trivedi, D. K., Goodacre, R., & Fowler, S. J. (2014). Taking your breath away: Metabolomics breathes life in to personalized medicine. *Trends in Biotechnology, 32*, 538–548. https://doi.org/10.1016/j.tibtech.2014.08.003

Rattray, N., Trivedi, D. K., Xu, Y., Chandola, T., Johnson, C. H., Marshall, A. D., Mekli, K., Rattray, Z., Tampubolon, G., Vanhoutte, B., White, I. R., Wu, F., Pendleton, N., Nazroo, J., & Goodacre, R. (2019). Metabolic dysregulation in vitamin E and carnitine shuttle energy mechanisms associate with human frailty.

Nature Communications, 10, 5027. https://doi.org/10.1038/s41467-019-12716-2

Reichl, B., Eichelberg, N., Freytag, M., Gojo, J., Peyrl, A., & Buchberger, W. (2020). Evaluation and optimization of common lipid extraction methods in cerebrospinal fluid samples. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences, 1153*, 122271. https://doi.org/10.1016/j.jchromb.2020.122271

Ribbenstedt, A., Ziarrusta, H., & Benskin, J. P. (2018). Development, characterization and comparisons of targeted and non-targeted metabolomics methods. *PLoS ONE, 13*, e0207082. https://doi.org/10.1371/journal.pone.0207082

Riquelme, G., Zabalegui, N., Marchi, P., Jones, C. M., & Monge, M. E. (2020). A python-based pipeline for preprocessing LC-MS data for untargeted metabolomics workflows. *Metabolites*. https://doi.org/10.3390/metabo10100416

Roy, C., Tremblay, P. Y., Bienvenu, J. F., & Ayotte, P. (2016). Quantitative analysis of amino acids and acylcarnitines combined with untargeted metabolomics using ultra-high performance liquid chromatography and quadrupole time-of-flight mass spectrometry. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences, 1027*, 40–49. https://doi.org/10.1016/j.jchromb.2016.05.006

Sangster, T., Major, H., Plumb, R., Wilson, A. J., & Wilson, I. D. (2006). A pragmatic and readily implemented quality control strategy for HPLC-MS and GC-MS-based metabonomic analysis. *The Analyst*. https://doi.org/10.1039/b604498k

Schantz, M. M., Powers, C. D., Schleicher, R. L., Betz, J. M., & Wise, S. A. (2016). Interlaboratory analytical comparison of fatty acid concentrations in serum or plasma. *Clinica Chimica Acta, 462*, 148–152. https://doi.org/10.1016/j.cca.2016.09.013

Schoeny, H., Rampler, E., Hermann, G., Grienke, U., Rollinger, J. M., & Koellensperger, G. (2020). Preparative supercritical fluid chromatography for lipid class fractionation-a novel strategy in high-resolution mass spectrometry based lipidomics. *Analytical and Bioanalytical Chemistry, 412*, 2365–2374. https://doi.org/10.1007/s00216-020-02463-5

Schrimpe-Rutledge, A. C., Codreanu, S. G., Sherrod, S. D., & McLean, J. A. (2016). Untargeted metabolomics strategies-challenges and emerging directions. *Journal of the American Society for Mass Spectrometry, 27*, 1897–1905. https://doi.org/10.1007/s13361-016-1469-y

Schwaiger, M., Schoeny, H., El Abiead, Y., Hermann, G., Rampler, E., & Koellensperger, G. (2018). Merging metabolomics and lipidomics into one analytical run. *The Analyst, 144*, 220–229. https://doi.org/10.1039/c8an01219a

Scott, S. A., Fu, J., & Chang, P. V. (2020). Microbial tryptophan metabolites regulate gut barrier function via the aryl hydrocarbon receptor. *Proc Natl Acad Sci U S A, 117*, 19376–19387. https://doi.org/10.1073/pnas.2000047117

Simón-Manso, Y., Lowenthal, M. S., Kilpatrick, L. E., Sampson, M. L., Telu, K. H., Rudnick, P. A., Mallard, W. G., Bearden, D. W., Schock, T. B., Tchekhovskoi, D. V., Blonder, N., Yan, X., Liang, Y., Zheng, Y., Wallace, W. E., Neta, P., Phinney, K. W., Remaley, A. T., & Stein, S. E. (2013). Metabolite profiling of a NIST standard reference material for human plasma (SRM 1950): GC-MS, LC-MS, NMR, and clinical laboratory analyses, libraries, and web-based resources. *Analytical Chemistry, 85*, 11725–11731. https://doi.org/10.1021/ac402503m

Siskos, A. P., Jain, P., Römisch-Margl, W., Bennett, M., Achaintre, D., Asad, Y., Marney, L., Richardson, L., Koulman, A., Griffin, J. L., Raynaud, F., Scalbert, A., Adamski, J., Prehn, C., & Keun, H. C. (2017). Interlaboratory reproducibility of a targeted metabolomics platform for analysis of human serum and plasma. *Analytical Chemistry, 89*, 656–665. https://doi.org/10.1021/acs.analchem.6b02930

Smith, D., & Spanel, P. (2005). Selected ion flow tube mass spectrometry (SIFT-MS) for on-line trace gas analysis. *Mass Spectrometry Reviews, 24*, 661–700. https://doi.org/10.1002/mas.20033

Soltow, Q. A., Strobel, F. H., Mansfield, K. G., Wachtman, L., Park, Y., & Jones, D. P. (2013). High-performance metabolic profiling with dual chromatography-Fourier-transform mass spectrometry (DC-FTMS) for study of the exposome. *Metabolomics, 9*, S132–S143. https://doi.org/10.1007/s11306-011-0332-1

Song, X., Sun, X., Oh, S. F., Wu, M., Zhang, Y., Zheng, W., Geva-Zatorsky, N., Jupp, R., Mathis, D., Benoist, C., & Kasper, D. L. (2020). Microbial bile acid metabolites modulate gut RORγ(+) regulatory T cell homeostasis. *Nature, 577*, 410–415. https://doi.org/10.1038/s41586-019-1865-0

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., Fan, T. W., Fiehn, O., Goodacre, R., Griffin, J. L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A. N., Lindon, J. C., Marriott, P., Nicholls, A. W., Reily, M. D., … Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics, 3*, 211–221. https://doi.org/10.1007/s11306-007-0082-2

Telu, K. H., Yan, X., Wallace, W. E., Stein, S. E., & Simón-Manso, Y. (2016). Analysis of human plasma metabolites across different liquid chromatography/mass spectrometry platforms: Cross-platform transferable chemical signatures. *Rapid Communications in Mass Spectrometry, 30*, 581–593. https://doi.org/10.1002/rcm.7475

Thompson, J. W., Adams, K. J., Adamski, J., Asad, Y., Borts, D., Bowden, J. A., Byram, G., Dang, V., Dunn, W. B., Fernandez, F., Fiehn, O., Gaul, D. A., Hühmer, A. F., Kalli, A., Koal, T., Koeniger, S., Mandal, R., Meier, F., Naser, F. J., O'Neil, D., … Moseley, M. A. (2019). International ring trial of a high resolution targeted metabolomics and lipidomics platform for serum and plasma analysis. *Analytical Chemistry, 91*, 14407–14416. https://doi.org/10.1021/acs.analchem.9b02908

Trapmann, S., Botha, A., Linsinger, T. P. J., Mac Curtain, S., & Emons, H. (2017). The new International Standard ISO 17034: General requirements for the competence of reference material producers. *Accreditation and Quality Assurance, 22*, 381–387. https://doi.org/10.1007/s00769-017-1285-5

Triebl, A., Burla, B., Selvalatchmanan, J., Oh, J., Tan, S. H., Chan, M. Y., Mellet, N. A., Meikle, P. J., Torta, F., & Wenk, M. R. (2020). Shared reference materials harmonize lipidomics across MS-based detection platforms and laboratories. *Journal of Lipid Research, 61*, 105–115. https://doi.org/10.1194/jlr.D119000393

Ubhi, B.K., 2018. Clinical metabolomics methods and protocols. Springer ISBN 978-1-4939-7591-4

Ulmer, C. Z., Jones, C. M., Yost, R. A., Garrett, T. J., & Bowden, J. A. (2018). Optimization of Folch, Bligh-Dyer, and Matyash sample-to-extraction solvent ratios for human plasma-based lipidomics studies. *Analytica Chimica Acta, 1037*, 351–357. https://doi.org/10.1016/j.aca.2018.08.004

Ulrich, E. M., Sobus, J. R., Grulke, C. M., Richard, A. M., Newton, S. R., Strynar, M. J., Mansouri, K., & Williams, A. J. (2019). EPA's non-targeted analysis collaborative trial (ENTACT): Genesis, design, and initial findings. *Analytical and Bioanalytical Chemistry, 411*, 853–866. https://doi.org/10.1007/s00216-018-1435-6

van Meer, G. (2005). Cellular lipidomics. *EMBO Journal*. https://doi.org/10.1038/sj.emboj.7600798

van Meer, G., Voelker, D. R., & Feigenson, G. W. (2008). Membrane lipids: Where they are and how they behave. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm2330

Vaughan, A. A., Dunn, W. B., Allwood, J. W., Wedge, D. C., Blackhall, F. H., Whetton, A. D., Dive, C., & Goodacre, R. (2012). Liquid chromatography-mass spectrometry calibration transfer and metabolomics data fusion. *Analytical Chemistry, 84*, 9848–9857. https://doi.org/10.1021/ac302227c

Viant, M. R., Ebbels, T., Beger, R. D., Ekman, D. R., Epps, D., Kamp, H., Leonards, P., Loizou, G. D., MacRae, J. I., van Ravenzwaay, B., Rocca-Serra, P., Salek, R. M., Walk, T., & Weber, R. (2019). Use cases, best practice and reporting standards for metabolomics in regulatory toxicology. *Nature Communications, 10*, 3041. https://doi.org/10.1038/s41467-019-10900-y

Wang, L., Su, B., Zeng, Z., Li, C., Zhao, X., Lv, W., Xuan, Q., Ouyang, Y., Zhou, L., Yin, P., Peng, X., Lu, X., Lin, X., & Xu, G. (2018). Ion-Pair Selection method for pseudotargeted metabolomics based on SWATH MS acquisition and its application in differential metabolite discovery of type 2 diabetes. *Analytical Chemistry, 90*, 11401–11408. https://doi.org/10.1021/acs.analchem.8b02377

Wenk, M. (2010). Lipidomics: New tools and applications. *Cell*. https://doi.org/10.1016/j.cell.2010.11.033

Wilkinson, M., White, I., Hamshere, K., Holz, O., Schuchardt, S., Bellagambi, F. G., Lomonaco, T., Biagini, D., Di, F. F., & Fowler, S. J. (2020). The peppermint breath test: A benchmarking protocol for breath sampling and analysis using GC-MS. *Journal of Breath Research*. https://doi.org/10.1088/1752-7163/abd28c

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci Data, 3*, 160018. https://doi.org/10.1038/sdata.2016.18

Yang, R., Zhang, Y., Qian, W., Peng, L., Lin, L., Xu, J., Xie, T., Ji, J., Zhan, X., & Shan, J. (2019). Surfactant lipidomics of alveolar lavage fluid in mice based 0on ultra-high-performance liquid chromatography coupled to hybrid quadrupole-exactive orbitrap mass spectrometry. *Metabolites*. https://doi.org/10.3390/metabo9040080

Yetukuri, L., Ekroos, K., Vidal-Puig, A., & Oresic, M. (2008). Informatics and computational strategies for the study of lipids. *Molecular BioSystems*. https://doi.org/10.1039/b715468b

Zelena, E., Dunn, W. B., Broadhurst, D., Francis-McIntyre, S., Carroll, K. M., Begley, P., O'Hagan, S., Knowles, J. D., Halsall, A., Consortium, H., Wilson, I. D., & Kell, D. B. (2009). (2009). Development of a robust and repeatable UPLC−MS method for the long-term metabolomic study of human serum. *Analytical Chemistry, 81*, 1357–1364. https://doi.org/10.1021/ac8019366

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Katrice A. Lippa[1] · Juan J. Aristizabal-Henao[2,3] · Richard D. Beger[4] · John A. Bowden[2] · Corey Broeckling[5] · Chris Beecher[6] · W. Clay Davis[7] · Warwick B. Dunn[8] · Roberto Flores[9] · Royston Goodacre[10] · Gonçalo J. Gouveia[11] · Amy C. Harms[12] · Thomas Hartung[13] · Christina M. Jones[1] · Matthew R. Lewis[14] · Ioanna Ntai[15] · Andrew J. Percy[16] · Dan Raftery[17] · Tracey B. Schock[7] · Jinchun Sun[4] · Georgios Theodoridis[18] · Fariba Tayyari[19] · Federico Torta[20] · Candice Z. Ulmer[21] · Ian Wilson[22] · Baljit K. Ubhi[23]

1 Chemical Sciences Division, National Institute of Standards and Technology (NIST), Gaithersburg, MD 20899, USA

2 Department of Physiological Sciences, Center for Environmental and Human Toxicology, College of Veterinary Medicine, University of Florida, Gainesville, FL 32610, USA

3 BERG LLC, 500 Old Connecticut Path, Building B, 3rd Floor, Framingham, MA 01710, USA

4 Division of Systems Biology, National Center for Toxicological Research, U.S. Food and Drug Administration (FDA), Jefferson, AR 72079, USA

5 Analytical Resources Core: Bioanalysis and Omics Center, Colorado State University, Fort Collins, CO 80523, USA

6 IROA Technologies, Chapel Hill, NC 27517, USA

7 Chemical Sciences Division, National Institute of Standards and Technology (NIST), Charleston, SC 29412, USA

8 School of Biosciences, Institute of Metabolism and Systems Research and Phenome Centre Birmingham, University of Birmingham, Birmingham B15, 2TT, UK

9 Division of Program Coordination, Planning and Strategic Initiatives, Office of Nutrition Research, Office of the Director, National Institutes of Health (NIH), Bethesda, MD 20892, USA

10 Department of Biochemistry and Systems Biology, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, BioSciences Building, Crown St., Liverpool L69 7ZB, UK

11 Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA

12 Biomedical Metabolomics Facility Leiden, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands

13 Bloomberg School of Public Health, Environmental Health and Engineering, Johns Hopkins University, Baltimore, MD 21205, USA

14 National Phenome Centre, Imperial College London, London SW7 2AZ, UK

15 Thermo Fisher Scientific, San Jose, CA 95134, USA

16 Cambridge Isotope Laboratories, Inc., Tewksbury, MA 01876, USA

17 Northwest Metabolomics Research Center, University of Washington, Seattle, WA 98109, USA

18 Department of Chemistry, Aristotle University, 54124 Thessaloniki, Greece

19 Department of Internal Medicine, University of Iowa, Iowa City, IA 52242, USA

20 Centre for Life Sciences, National University of Singapore, 28 Medical Drive, Singapore 117456, Singapore

[21] Centers for Disease Control and Prevention (CDC), Atlanta, GA 30341, USA

[22] Computational & Systems Medicine, Imperial College, Exhibition Rd, London SW7 2AZ, UK

[23] MOBILion Systems Inc., 4 Hillman Drive Suite 130, Chadds Ford, PA 19317, USA