Analyst

PAPER

View Article Online
View Journal | View Issue

Cite this: *Analyst*, 2024, **149**, 5423

Evaluating protocols for reproducible targeted metabolomics by NMR†

Darcy Cochran, [a,b] Panteleimon G. Takis, *[c,d,e] James L. Alexander,[f,g,h] Benjamin H. Mullish, [f,g] Nick Powell,[f,g] Julian R. Marchesi[h] and Robert Powers *[a,b]

Metabolomics aims to study the downstream effects of variables like diet, environment, or disease on a given biological system. However, inconsistencies in sample preparation, data acquisition/processing protocols lead to reproducibility and accuracy concerns. A systematic study was conducted to assess how sample preparation methods and data analysis platforms affect metabolite susceptibility. A targeted panel of 25 metabolites was evaluated in 69 clinical metabolomics samples prepared following three different protocols: intact, ultrafiltration, and protein precipitation. The resulting metabolic profiles were characterized by 1D ^{1}H nuclear magnetic resonance (NMR) spectroscopy and analyzed with Chenomx v8.3 and SMolESY software packages. Greater than 90% of the metabolites were extracted more efficiently using protein precipitation than filtration, which aligns with previously reported results. Additionally, analysis of data processing software suggests that metabolite concentrations were overestimated by Chenomx batch-fitting, which only appears reliable for determining relative fold changes rather than absolute quantification. However, an assisted-fit method provided sufficient guidance to achieve accurate results while avoiding a time-consuming fully manual-fitting approach. By combining our results with previous studies, we can now provide a list of 5 common metabolites [2-hydroxybutyrate (2-HB), choline, dimethylamine (DMA), glutamate, lactate] with a high degree of variability in reported fold changes and standard deviations that need careful consideration before being annotated as potential biomarkers. Our results show that sample preparation and data processing package critically impact clinical metabolomics study success. There is a clear need for an increased degree of standardization and harmonization of methods across the metabolomics community to ensure reliable outcomes.

Received 24th July 2024,
Accepted 1st October 2024

DOI: 10.1039/d4an01015a

rsc.li/analyst

1 Introduction

Metabolomics is the process of identifying and quantifying the small molecular weight (<1000 Da) metabolites present in a given organism, system, or sample.[1–3] Metabolomics has been applied to clinical, environmental, and agricultural projects to investigate metabolic networks of interest within humans, plants and animals.[4–7] The increased sensitivity of the metabolome to stressors can be leveraged to learn how diet, the environment and diseases affect underlying biochemical pathways.[8–12] This heightened sensitivity of the metabolome relative to the genome or proteome allows researchers to gain deeper insights and to obtain reliable molecular mechanisms of diseases, drugs, and further our understanding of a variety of biological systems. However, a downside of the increased sensitivity of metabolomics is that biologically irrelevant changes in response to storage and preparation conditions during analysis are easily detected.[13–16]

The typical workflow for metabolomics studies involves isolating and extracting fractions of metabolites from the biologi-

[a]*Department of Chemistry, University of Nebraska-Lincoln, Lincoln, Nebraska, 68588-0304, USA*
[b]*Nebraska Center for Integrated Biomolecular Communication, University of Nebraska-Lincoln, Lincoln, Nebraska, 68588-0304, USA. E-mail: rpowers3@unl.edu*
[c]*Department of Chemistry, University of Ioannina, Ioannina GR 451 10, Greece. E-mail: ptakis@uoi.gr*
[d]*Section of Bioanalytical Chemistry, Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London SW7 2AZ, UK*
[e]*National Phenome Centre, Department of Metabolism, Digestion and Reproduction, Imperial College London, London W12 0NN, UK. E-mail: p.takis@imperial.ac.uk*
[f]*Departments of Gastroenterology and Hepatology, St Mary's Hospital, Imperial College Healthcare NHS Trust, South Wharf Road, Paddington London, W2 1NY, UK*
[g]*Division of Digestive Diseases, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, W2 1NY, UK*
[h]*Department of Gastroenterology, St Mark's Hospital and Academic Institute, Middlesex, UK*

† Electronic supplementary information (ESI) availableSupplementary Fig. S1–S6 showing methods, PCA scores plots and standard deviation scatterplots. Supplementary Tables S1–S7 providing scaled data, *p*-value calculations and standard deviation analysis as referenced in text. See DOI: **https://doi.org/10.1039/d4an01015a**

cal samples to be evaluated by an analytical instrumentation before finally assessing the generated data or spectra with univariate or multivariate statistical methods. While the process may seem simple, there are many opportunities for inconsistency when making choices about factors like storage temperature, extraction method, sample preparation and instrumentation type.[17] In this regard, the metabolomics process comprises numerous decision points that may impact the composition of the detectable metabolome and lead to reproducibility and accuracy concerns.[18]

Several studies have shown that metabolites are susceptible to variation due to both pre-analytic and post-analytic protocols. For example, Breier *et al.* (2014) investigated the effects of collection, transport, and sample processing methods on a panel of metabolites and found that amino acids were particularly fragile during simulated transportation.[19] Residual enzymatic activity was also observed to be present even after the sample was frozen, which suggested the potential need for a single freeze–thaw cycle to achieve a stable metabolomics sample.[19] Nevertheless, long term storage even at ultra-low −80 °C temperatures resulted in statistically significant (±15%) changes in 50% of the analyzed metabolites.[14] Want *et al.* (2009) demonstrated that varying the extraction method and the solvent composition produced unique metabolite panels and yielded distinct extraction efficiencies.[20] A total of 14 different extraction methods were examined and compared in which a 100% methanol extraction of serum was deemed to provide the best comprehensive metabolite profile with a high degree of protein removal.[20] Despite similar on-going efforts by COSMOS: Coordination of Standards in Metabolomics, MSI: Metabolomics Standards Initiative, mQACC: Metabolomics Quality Assurance & Quality Control Consortium, and MANA: Metabolomics Association of North America to establish community standards and best practices, there continues to be a proliferation of metabolomics methodologies populating the scientific literature.[21–27] Simply, the metabolomics community has yet to broadly adopt well-vetted general guidelines for metabolomics studies.[20–22,28–30]

Choosing the correct sample preparation technique depends on several factors that include both the type and the amount of the biological samples, and the metabolites being investigated. The choice of analytical platform may also be guided by these and other considerations. Nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) are commonly used in metabolomics studies and provide complementary results.[31,32] NMR detects the most abundant metabolites (>1 μM) while providing accurate and highly reproducible absolute quantification. Conversely, MS, which is typically combined with liquid- or gas-chromatography, provides a broader coverage of the metabolome with a significantly lower limit of detection and a larger dynamic range. However, one of the important benefits of NMR over MS as an analytical platform for metabolomics is the limited need for sample manipulation, which lowers the variability and inconsistency across and between data sets.[33,34] Specifically, for liquid biological samples like blood, urine, or other biofluids, an intact sample preparation approach can often be effective in providing reliable and reproducible metabolome concentration changes. An intact sample preparation method requires the simple addition to the sample of a deuterated buffer containing an internal standard. Additional processing, such as the removal of other endogenous biomolecules, is not needed. However, an intact sample approach risks metabolite signal suppression due to metabolite–protein (or other biomolecules) binding, which yields lower resolution and lower intensity peaks with a broadened baseline. The bound metabolites may remain completely undetected in the resulting NMR spectra, which can lead to erroneous conclusions about the perceived changes in the system. While intact sample preparation has advantages, alternative protocols such as solvent-induced protein precipitation or ultrafiltration extract the metabolites of interest while simultaneously removing unwanted biomaterials before analysis.[28,35–40]

Another potential source of variation or error in a metabolomics study comes from the myriad of data processing tools used to analyze a given spectral data set.[41] For NMR, a few options for metabolomics software toolkits include Chenomx (Chenomx, Inc., Alberta Canada), MetaboAnalyst 5.0 (**https://www.metaboanalyst.ca**), MVAPACK (**https://mvapack.unl.edu**), and SMolESY (**https://github.com/pantakis/SMolESY_platform**).[42–44] Each of these software platforms uniquely handles the NMR data by relying on a variety of spectral binning, metabolite identification algorithms, or statistical analysis options. Especially among academia, there is an understandable preference for freely available software like MetaboAnalyst 6.0, MVAPACK, or SMolESY, although commercial software does provide valuable benefits such as customer support and routine updates. The broad range of available algorithms in metabolomics software allows researchers the flexibility to choose and optimize data processing pipelines to meet their individual needs, but difficulties often arise when making inter-laboratory comparisons or assessing data accuracy and reproducibility. Consequently, the harmonization of metabolomics methods, data, and results is an ongoing effort that has only made modest advancements with plenty of opportunities for improvement.[21,24–26]

Despite the potential influence of other factors on metabolic profiles, such as sample storage, our primary objective was to explore only the effects of sample treatment and the choice of software on the accurate identification and quantification of a common set of plasma metabolites. With our approach, we tried to simulate the typical pipeline followed in conducting an NMR-based metabolomics study, specifically without prior knowledge of the samples content. The reproducibility of metabolome measurements was assessed with a 1D ^{1}H NMR data set of human plasma samples since blood samples are commonly used in clinical metabolomics studies. The data set consists of plasma samples obtained from patients diagnosed with inflammatory bowel disease (IBD) that were age and sex matched to healthy controls. An important feature of our study is that our data was combined with previously published results to achieve a comprehensive view of

the degree of variation inherent to individual plasma metabolites. In this regard, the outcomes of our research will contribute to the ongoing metabolomics harmonization efforts by determining how sample preparation methods and NMR acquisition and data processing parameters affect the absolute concentration, statistical significance of between-group and within group metabolite changes, and the reliability and reproducibility of metabolite quantification and annotation.

2 Material and methods

Details for the study design and cohort demographics, preparation of pooled QC samples, preparation of NMR samples using ultrafiltration, preparation of NMR samples using methanol precipitation, preparation of NMR samples using intact plasma, NMR data acquisition and processing, representative 1D ^{1}H NMR spectra with annotations, and statistical analyses are included in the ESI.†

3 Results and discussion

3.1 Metabolic profile is determined by protein removal protocol and analysis method

Principal components analysis (PCA) of the 1D ^{1}H NMR data sets yielded a scores plot containing two distinct clusters comprising either the intact plasma samples or the ultrafiltered and methanol-precipitated plasma samples (see ESI Fig. S1 and S2†). Conversely, there was no separation in the PCA scores plot according to clinical group (see ESI Tables S1 and S2, Fig. S3 and S4†). The PCA models created from the 1D ^{1}H NMR data sets for each preparation method indicate that the pooled samples are in the center of the 2D PCA scores plot. Additionally, the lack of separation based on pathology increases the level of confidence that the observed group differences were truly derived from only the sample preparation method. Finally, it is encouraging to note that these trends persist across all data irrespective of whether the PCA was modelled with binned raw spectral data (see ESI Fig. S3a and S4a†), Chenomx data (see ESI Fig. S3b and S4b†), or SMolESY data (see ESI Fig. S3c and S4c†).

3.2 Methanol protein precipitation leads to higher overall metabolite concentrations

Fig. 1 and Table S3† summarize the scaled measured concentrations for the 25 targeted metabolites grouped by sample preparation protocol and data processing method. Table S3† also lists the coefficient of variation (CV) and percent differences in the measured metabolite concentrations, which are calculated relative to the manual analysis of protein precipitation plasma samples. Fig. 2 summarizes the average metabolite concentration and the number of metabolites that were significantly altered ($p < 0.05$) in a pairwise comparison of sample preparation protocols and according to processing method. A detailed overview of the comparisons can be seen in Fig. S5 and ESI Table S4.† In nearly all cases, almost all the metabolites in the panel were detected. The Chenomx batch-fit approach missed the greatest number of metabolites, omitting choline, histidine, and lysine when analyzing the plasma NMR samples following methanol-induced protein precipitation. In general, the assisted-fit Chenomx analysis of the 1D ^{1}H NMR spectra acquired for the methanol-induced protein precipitation plasma samples had the most accurate (lowest % difference from Chenomx manual-fit data) metabolite concentrations where the average percent difference was 7% (for the 19 metabolite concentrations lower in manual-fit) or −16% (5 higher in assisted-fit) (see ESI Table S3†). The same Chenomx analysis of the plasma NMR samples prepared by ultrafiltration yielded a nearly uniform decrease in the measured metabolite concentrations corresponding to an average decrease of 57% for manual-fit and a decrease of 67% for assisted-fit. The observed decrease in ultrafiltered plasma samples aligns well with previous results.[35] Citrate, an endogenous metabolite, was the only notable exception. Overall, 96% (24) of the 25 targeted metabolites were determined to have at a higher concentration in the methanol-precipitated plasma samples with 92% (23) of these concentration differences being statistically significant (Fig. 2, Fig S5, see ESI Table S4†).

3.3 SMolESY and Chenomx batch-fit tend to overestimate metabolite concentrations

The batch-fit Chenomx analysis and SMoIESY processing protocols of both the methanol-precipitated and ultrafiltered NMR data sets yielded similar, but mixed results relative to the manual analysis of the methanol-precipitated data set. In general, metabolite concentrations tended to increase with batch-fit and SMoIESY. For example, batch-fit analysis produced 15 metabolites with an average increase in concentration of 98% (see ESI Table S3†). This trend is clearly apparent for 3-hydroxybutyrate (3-HB), alanine, creatinine, dimethyl sulfone (DMSO2), DMA, formate, glutamate, glutamine, glycine, isoleucine, lactate, phenylalanine, tyrosine, and valine (Fig. 1).

Alternatively, the batch-fit Chenomx analysis measured 7 metabolites with a 29% decrease in concentration for the methanol-precipitated NMR data set. SMoIESY analysis yielded nearly identical results to Chenomx corresponding to 19 metabolites with an average increase in concentration of 98% and 6 metabolites with a 48% decrease in concentration. Similarly, the batch-fit Chenomx analysis of the ultrafiltered NMR data set resulted in 12 metabolites with a 99% average increase in concentration compared to 13 metabolites with a 52% decrease in concentration. Again, SMoIESY processing yielded comparable results with 15 metabolites exhibiting a 121% average increase in concentration compared to 10 metabolites with a 38% decrease in concentration. The metabolite concentrations measured by the batch-fit Chenomx analysis varied significantly ($p < 0.05$) from the values determined by the assisted-fit or manual-fit methods in 52% of the ultrafiltered plasma samples and in 41% of the methanol-

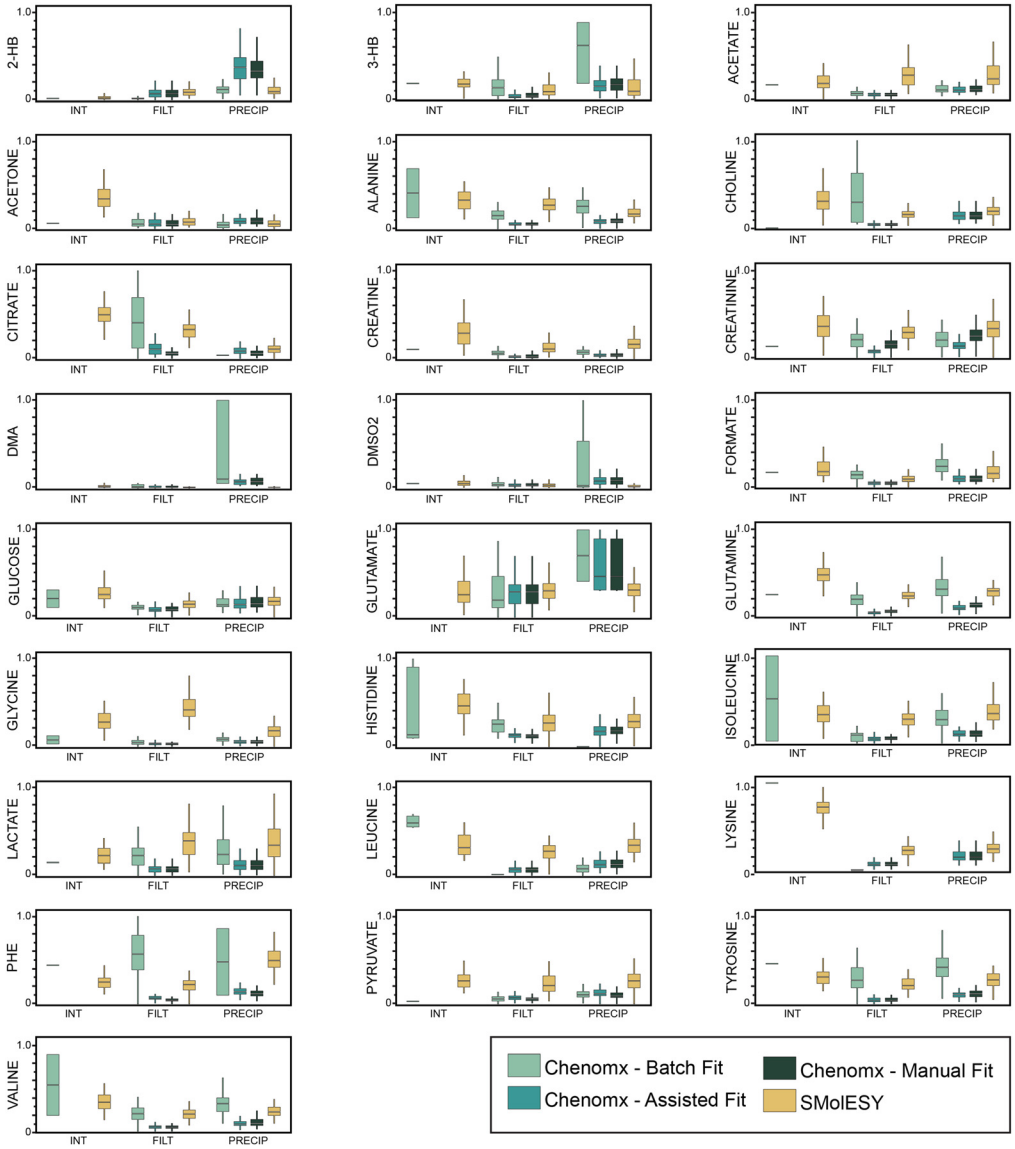

Fig. 1 Scaled metabolite concentrations organized by fit and sample preparation method. Box and whisker plots of range scaled (0–1) data for each of the 25 metabolites. Plots are placed in alphabetical order. Each individual box and whisker plot is organized left-to-right by preparation method from intact (INT), to ultrafiltered (FILT), to methanol-induced protein precipitated (PRECIP). The color of each box corresponds to the processing method used to generate the data in the following order: Chenomx – batch-fit (light blue), Chenomx-assisted-fit (teal), Chenomx – manual-fit (dark green), SMolESY (gold). Abbreviations: 2-HB: 2-hydroxybutyrate, 3-HB: 3-hydroxybuytrate, DMSO2: dimethyl sulfone, DMA: dimethylamine.

precipitated plasma samples (Fig. 2, see ESI Table S4†). Overall, Chenomx determined 80% of the metabolites to be significantly different between the methanol-precipitated plasma samples and the ultrafiltered plasma samples, but SMolESY only found a statistical difference in 36% of the metabolites.

3.4 Biological variance is a primary factor affecting precision in measured metabolite concentrations

Unsurprisingly for biological samples, the overall variability in the measured metabolite concentrations was high as assessed by CVs (see ESI Tables S3 and S5†). The Chenomx analysis of the 69 methanol-precipitated plasma samples yielded average CVs ranging from 97 to 102%. SMoIESY processing yielded a slightly improved but not statistically significant ($p > 0.4$) average CV of 86%. Given the overall lower metabolite concentrations measured for the ultrafiltration samples, the resulting CVs were equally reduced and ranged from an average of 57 to 82% for Chenomx and a comparable average CV of 75% for SMoIESY. Notably, only the manual-fit and assisted-fit Chenomx analysis led to average CV values that were significantly different ($p < 0.0001$) between the methanol-precipitated plasma samples and ultrafiltered plasma samples. The overall average 25% decrease in CV values is partly attributed to the average 55% decrease in metabolite concentrations.

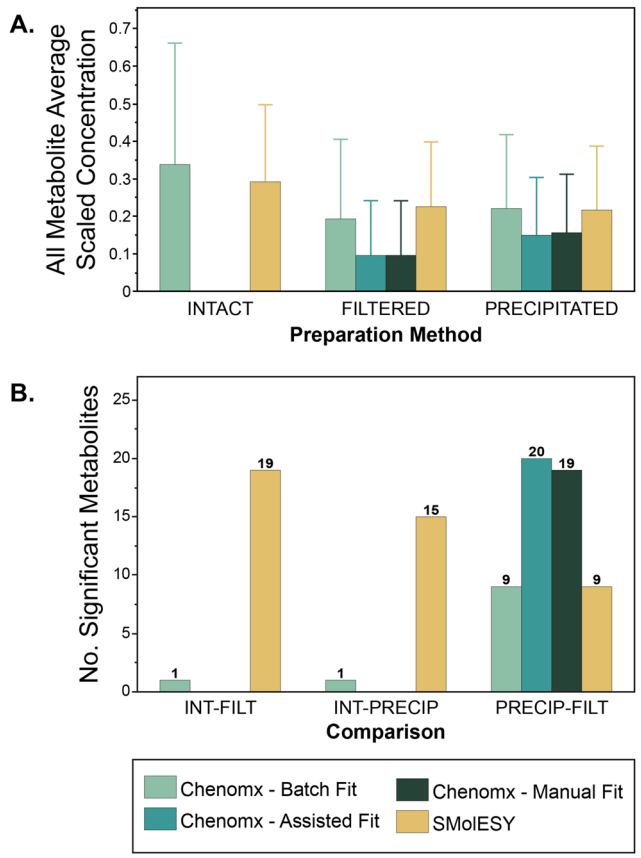

Fig. 2 Summary bar charts showing (A) general trends in metabolite concentration and (B) the total number of metabolites that had statistically significant concentration variation between sample preparation type.

Overall, there is no clear evidence to suggest that sample preparation protocol or data processing method produced an inherently better precision. Instead, the natural biological variation appears to be the dominating factor.

3.5 Analysis of intact plasma samples required specialized protocols and algorithms

The analysis of the intact plasma sample was completed using both the Chenomx and SMolESY software platforms. SMolESY was used to analyze the NOESY and CPMG NMR datasets and Chenomx was only used to analyze the CPMG NMR spectra. Chenomx is not able to account for the large protein background signal present in the NOESY NMR spectra. Conversely, SMolESY was specifically designed to evaluate intact plasma samples acquired with a NOESY pulse sequence. SMolESY mimics the effects of the CPMG NMR experiment and computationally removes the macromolecule baseline from the NMR spectra before evaluating peak areas. In this regard, SMolESY was able to successfully detect 100% of the metabolite panel in 100% of the 69 NOESY and CPMG spectra acquired for the intact plasma samples. Encouragingly, the measured metabolites using both the NOESY and CPMG spectra yielded similar results from the SMolESY platform. The average concentration difference ranged from 17% to 19% with only 6 statistically significant differences ($p < 0.05$). In total, eleven metabolites were measured at a higher concentration using the CPMG NMR spectra with only the differences in 2-HB and formate being statistically significant. A total of 14 metabolites resulted in a lower concentration with the CPMG spectra but only the measured concentration differences for 3-HB, acetone, glutamate, and lysine were statistically significant. Again, the overall variability in these measured metabolite concentrations were high as assessed by CVs and ranged from 45 to 322%. Overall, the variability in the measured metabolite concentrations appear to be determined by a combination of inherent biological variations and the absolute metabolite concentration.

The SMolESY processing of the intact plasma samples resulted in statistically significant ($p < 0.05$) differences in the measured metabolite concentrations relative to manual-fit Chenomx analysis of the methanol-precipitated plasma samples (see ESI Tables S3 and S5†). Specifically, the SMolESY processing of the NOESY data set resulted in an average increase in measured metabolite concentrations by 174% for 20 of the 25 metabolites. Five metabolites were measured with an average 57% decrease in concentration. 16 of these metabolite concentration differences were statistically significant ($p < 0.05$).

Although the CPMG pulse sequence is not recommended for quantitative analyses due to its interference with signal intensities, we also analyzed the CPMG profiles since they are more similar to the profiles of the extracted samples (*i.e.*, without the broad macromolecular signals). The intact plasma samples acquired with the CPMG pulse sequence exhibited the poorest Chenomx metabolic profiling across several metrics. Metabolites were only detected in 3% of the intact samples and required extensive manual analysis to retrieve any semblance of usable data. The very sparse results that were produced by this analysis were also clearly incorrect. Chenomx batch-fit results of the CPMG experiments suggested that in the few samples metabolites were detected in, that they were found to be at a much higher or much lower concentration than the NOESY experiments. For example, when measured with the CPMG pulse sequence, glutamine was found to be at 1003.8 μM on average but only at 164 μM in the NOESY experiments and DMA was detected at 0.6 μM in CPMG experiments but 4.85 μM in NOESY experiments.

3.6 Reproducibility of metabolomics sample preparation protocols

Gowda *et al.* (2014) previously reported an increase in metabolite concentrations in serum and plasma samples prepared by protein precipitation methods compared to either an ultrafiltration protocol or to an intact sample.[35] Additionally, they determined that intact sample preparation was not appropriate for metabolite quantitation. Thus, the protein precipitation approach was identified as a robust and best-choice method.[35] The higher percentage of metabolites recovered through precipitation is probably a contributing factor to its relative popularity as a metabolite extraction method compared to other

options.[45] Precipitating proteins from a biofluid sample removes 98% of the proteins and the addition of organic solvent to a biofluid sample will disrupt any metabolite–protein binding.[45] However, the results from Tiziani *et al.* (2008) directly contradicts these observations.[35,46] Specifically, 3-HB, alanine, creatine, citrate, isobutyrate, succinate, threonine, glucose, glutamate, and ethanol were all found by Tiziani *et al.* (2008) to be significantly increased in filtered samples compared to precipitated samples.[46] A recently published study by Madrid-Gambin *et al.* (2023) found an increase in concentration for 63% of the metabolites measured from protein-filtered samples compared to an increase in concentrations for 37% of the metabolites measured from the protein-precipitated samples.[47] The methanol-induced protein precipitation protocol was observed to produce a non-ideal baseline in the 1D ^{1}H NMR spectra. As a result, glycerophospholipid solid-phase extraction (g-SPE) was recommended as the preferred sample preparation method. These divergent results and a corresponding lack in reproducibility show a clear gap in our fundamental knowledge of metabolomics and the effects of sample preparation method.

A direct comparison of these published outcomes with our results can be challenging due to missing or variations in experimental parameters and because of fundamental differences in the scale of reported metabolite concentrations. Fortunately, Gowda *et al.* (2014) and Madrid-Gambin *et al.* (2023) provided sufficient information to enable proper scaling of study data for a direct comparison with our results (see ESI Table S6†). Fig. 3 (top) plots the fold change (FC) in metabolite concentrations derived by comparing protein precipitation samples (PRECIP) to intact plasma samples (INT) for the commonly reported metabolites. Similarly, FCs were also calculated between PRECIP and ultrafiltration (FILT) samples (Fig. 3, middle). An average of these previously published FCs was then compared to our results (Fig. 3, bottom). Since Gowda *et al.* (2014) determined an intact plasma sample was not appropriate for metabolite quantitation, this data was excluded from Fig. 3.[35] The Chenomx analysis of the intact plasma sample was similarly determined to be unreliable and was not included in the fold change comparison presented in Fig. 3. Encouragingly, a qualitative examination of all the FCs suggests similar overall trends that are independent of study. The overall consistency in the FC calculated from the three independent studies and the sample preparation methods is readily apparent in the overlay of the average FC values (Fig. 3, bottom). Please note the overlap of the average metabolite concentrations with the error bars and the close clustering around or slightly above an FC of 1 in most cases. This suggests the relative FCs likely negates sources of error and/or variability readily apparent in the individual measurements of absolute metabolite concentrations.

For the PRECIP/INTACT FC comparison from (Fig. 3, top), it is particularly encouraging to note that our SMolESY derived FC values closely aligns with the results from Madrid-Gambin *et al.* (2023) with only a few notable deviations. Specifically, SMolESY measured a significantly higher FC for 2-HB. The FC

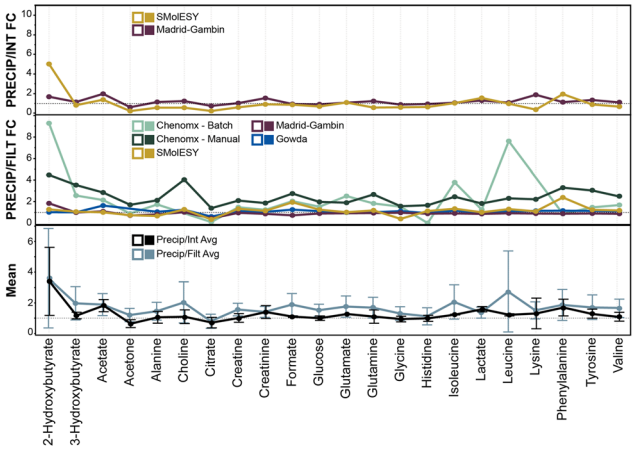

Fig. 3 Comparative analysis of results. Line graph plotting the fold change of each metabolite as reported by previously published studies and present results. Fold changes (FC) are organized by group with top panel containing precipitated sample/intact sample fold changes from SMolESY data (gold) and Madrid-Gambin *et al.* (2023) data (maroon). The middle panel contains methanol-induced protein precipitated (PRECIP) and ultrafiltered (FILT) sample fold changes from Chenomx – batch-fit results, Chenomx – manual-fit results, SMolESY results, Madrid-Gambin *et al.* (2023) results, and Gowda *et al.* (2014) results. The bottom panel contains line plots of calculated averages from the top two panels (Chenomx results were excluded from the Precip/Filt Avg line since they were identified as unreliable). The color of the line is correlated to the method or previous publication.[26,50] A dashed line is present at FC = 1 for reference. All fold changes were calculated from the data generated by each processing method or publication. For example, the gold SMolESY lines are generated from SMolESY data. No cross-comparative fold changes were calculated or plotted.

values for 3-HB, alanine, choline, creatine, creatinine, glutamine, leucine, lysine, tyrosine and valine were opposite in the relative directions (*i.e.*, increased or decreased). Madrid-Gambin *et al.* (2023) did not report a concentration for glutamate. Overall, the Madrid-Gambin *et al.* (2023) concentrations averaged 39% higher than the SMolESY derived FCs.

In the PRECIP/FILT comparison (Fig. 3, middle), the manually-fit Chenomx FC values were systematically increased relative to SMolESY and the other studies. Chenomx consistently underestimated metabolite concentrations from ultrafiltered plasma samples and/or overestimated concentrations for the methanol-precipitated plasma samples. The manually-fit Chenomx FC values averaged 2.47 ± 0.81 compared to 1.09 ± 0.18 for Gowda *et al.* (2014), 0.92 ± 0.25 for Madrid-Gambin *et al.* (2023), and 1.15 ± 0.44 for SMolESY. A few notable deviations were the relatively large (>200%) FC increases for 2-HB, 3-HB, choline, and citrate. Despite these systematic offsets, the manual-fit Chenomx metabolite concentrations appear to follow a similar trend as the other studies and SMolESY. For example, the spike in the FC of choline, the relative decrease in the citrate FC, and the relative increase in isoleucine FC were mirrored by SMolESY.

The batch-fit Chenomx analysis provided metabolite concentrations that were also systematically higher than the values determined by other studies and SMolESY with an average FC

value of 2.40 ± 2.28. Again, Chenomx consistently underestimated metabolite concentrations from ultrafiltered plasma samples and/or overestimated concentrations for the methanol-precipitated plasma samples. A relatively large FC increase (236% to 589%) was observed for 2-HB, isoleucine, and leucine. Also, the batch-fit Chenomx analysis did not report a concentration for choline, histidine or lysine. It was also notable that the batch-fit Chenomx FC values had the highest variability relative to the other methods and studies. The lower correlation in the Chenomx measured metabolite concentrations from the ultrafiltered and methanol-precipitated plasma samples and the resulting higher FC values were surprising since Chenomx is more compatible with extracted samples. Conversely, the FC values calculated from the Gowda *et al.* (2014), Madrid-Gambin *et al.* (2023), and SMolESY analysis yielded average values near one suggesting a good correlation in the measured metabolite concentrations from the ultrafiltered and methanol-precipitated plasma samples. Nevertheless, there are still a few deviations, with a higher FC for 2-HB (Madrid-Gambin *et al.* (2023)), acetate (Gowda *et al.* (2014),) formate and phenylalanine (SMolESY). A lower FC was observed for acetone (Madrid-Gambin *et al.* (2023) and SMolESY), citrate (all three), and glycine (SMolESY).

The metabolites most sensitive to sample preparation conditions may be identified by a further comparison of the three independent studies. The average FC and standard deviation values were calculated from the Gowda *et al.* (2014), Madrid-Gambin *et al.* (2023), and SMolESY reported metabolite concentrations and plotted in Fig. 3 (bottom) and Table S6.† The Chenomx derived FC values were excluded from the analysis because of the systematically elevated values relative to the other published studies and to the corresponding SMolESY analysis. As evident from the visual evaluation of Fig. 3 (bottom) and confirmed by a pairwise Student *t*-test ($p > 0.1$, see ESI Table S6†) the results were statistically equivalent except for lactate ($p < 0.02$). As desired, most FC values were approximately one. There were some notable differences in the variability or precision in the measured metabolite concentrations and associated FC values as evident by %CV. The average %CV was 20% ± 14% for the FC values calculated from the comparison of the methanol-induced protein precipitation and ultrafiltration plasma samples. Formate, glycine, and phenylalanine exhibited significantly higher %CV values ranging from 46% to 54%. The average %CV increased to 38% ± 25% for the FC values calculated from the comparison of the methanol-induced protein precipitation and intact plasma samples. 2-HB, acetone, alanine, choline, citrate, glutamine, and lysine exhibited significantly higher %CV values ranging from 48% to 96%. Again, Madrid-Gambin *et al.* (2023) did not report glutamate.

Taken together, these results clearly demonstrate that fold-change measurements provide an overall higher reproducibility in measured metabolite changes compared to absolute metabolite concentrations. FC values were statistically equivalent regardless of the plasma sample preparation protocol, methanol-induced protein precipitation, protein removal by ultrafiltration, or an intact plasma sample. Presumably, calculating a ratio negates constant sources of errors resulting from the processing, handling, and analysis of the plasma samples. In our study, measurements taken from the methanol-precipitated plasma samples served as the gold standard. However, we posit that employing reference compounds (such as the addition of non-endogenous metabolites to the samples) and/or electronic signals (such as ERETIC[48]) could also effectively mitigate errors stemming from the processing, handling, and analysis of the plasma samples as well as multicentered cohorts. Conversely, Chenomx processing was sensitive to sample preparation protocols and data processing methods, resulting in over a 100% difference in measured FC values. Recall, the manual-fit Chenomx analysis of the plasma samples following methanol-induced protein precipitation produced the highest overall metabolite concentrations, which is likely the primary source of the poor correlation between sample preparation protocols. The results also highlight the metabolites that should be carefully scrutinized to determine if the observed concentration change is truly caused by the condition of interest or is an artifact of sample preparation. For example, metabolites with %CVs and FCs significantly higher than average may be suspect.

3.7 2-HB and glutamate are sensitive to sample preparation method

To evaluate metabolite stability, we compared average metabolite standard deviations and %CV across each processing method (see ESI Fig. S6 and Table S7†). Metabolites with a standard deviation above the group average as determined by Tukey's IQR outlier test were determined to be less stable. Conversely, standard deviations below the group average were determined to be more stable. While %CV is a valuable metric, it is not the most precise measurement of reproducibility since its magnitude varies with absolute concentration. When evaluating each metabolite's %CV, DMA and DMSO2 were the most consistent outliers with %CV's ranging from 158%–519%. However, these two metabolites were also present at the lowest concentration in the plasma samples and their standard deviations were similar to other metabolites. Thus, the lower absolute concentration systematically inflated the %CV to being a statistical outlier. Therefore, our criteria for determining a metabolite to be significantly more or less stable than average was based exclusively on quantile range outlier testing of the standard deviations.

Lysine and phenylalanine have the highest standard deviation of 0.22 when all the data was analyzed together as one group (see ESI Fig. S6a†). However, lysine and phenylalanine were not statistically different after grouping the data by processing software. The Chenomx data shows an increase in the standard deviation for 2-HB and glutamate (see ESI Fig. S6b†) while the SMolESY analysis resulted in no standard deviation values for the combined dataset being identified as an outlier (see ESI Fig. S6c†). The data from the Chenomx batch-fit analysis was excluded from the pooled Chenomx group since it was previously determined to be an unreliable analysis

method, which was further confirmed by the widespread in standard deviation values as seen in Fig. S6d and e.† However, when examining the Chenomx assisted-fit and manual-fit standard deviations, a clear trend can be seen in both the ultrafiltered and methanol-precipitated plasma samples. 2-HB, DMA and glutamate consistently exhibited standard deviations that were outliers from the rest of the metabolites. Glutamate, in particular, was a clear outlier in all four data groups (see ESI Fig. S6f–i†). Overall, these results suggest that 2-HB, DMA and glutamate are especially sensitive to sample preparation conditions. Additionally, the standard deviations from the methanol-precipitated plasma samples have a narrower range (excluding outliers, see ESI Fig. S6g and i†) compared to the ultrafiltered plasma samples (see ESI Fig. S6f and h†). This trend is highlighted by the tight distribution around the dashed average line in the scatter plots shown in Fig. S6d–i.† These results confirm previous studies which concluded that methanol-induced protein precipitation to be the most robust protein removal method for plasma or serum metabolomics.[35,47]

There were no outliers when examining the standard deviation values from the combined SMolESY data set (see ESI Fig. S6c†), but when grouping the data by sample preparation method, 2-HB, choline, and lactate became outliers. Interestingly, 2-HB had a higher standard deviation in the Chenomx analysis of the methanol-precipitated plasma samples, but it had a non-significant standard deviation in the SMolESY analysis of these same samples. A significantly lower standard deviation in the SMolESY analysis was obtained with the intact samples. This wide range of low, normal, and high standard deviations suggests 2-HB may be more sensitive to sample preparation conditions than the other metabolites. For similar reasons, choline and lactate are also suspect.

4 Conclusions

In the present study, we systematically reviewed the effects of sample preparation methods and data analysis platforms on measured metabolite concentrations. The quickly expanding library of metabolomics software provides many alternative options for spectral analysis, but also adds new complications for on-going harmonization efforts by the metabolomics community. We clearly observed that both the choice of sample preparation method and data analysis platform significantly perturbed the resulting metabolic profile, requiring careful consideration of an appropriate metabolomics processing pipeline. Consistent with prior results, we also found that methanol-induced protein precipitation yielded a higher overall concentration of metabolites recovered from plasma samples. Clearly, methanol-induced protein precipitation should be the preferred choice for sample preparation by the scientific community. Importantly, a high level of reproducibility was achieved across our study and the two prior studies by using fold-change measurements relative to absolute concentration changes.[35,46,47] Calculating metabolite concentration ratios cancels out sample processing errors, making an additional "normalization step" (*e.g.*, ratio calculation) essential when merging datasets from multicenter cohorts, various studies or laboratories.

Our parallel investigation of Chenomx and SMolESY as analysis platforms offered several valuable insights. Chenomx presents several degrees of automation, batch-fit, assisted-fit, and manual-fit, which can be useful for high-throughput data analysis as the number of samples, or the cohort size increases to comprise hundreds to thousands of NMR spectra. Encouragingly, assisted-fit, and manual-fit Chenomx analysis were found to yield essentially identical results. Assisted-fit Chenomx processing is the preferred choice, especially for a large number of samples, since it drastically reduces the analysis time proportional to the increase in cohort size.

The manual-fit Chenomx processing of the plasma samples prepared by protein-precipitation yielded the highest overall metabolite concentrations relative to SMolESY and the other Chenomx protocols. Accordingly, this increase in measured metabolite concentrations is likely a closer representation of the true metabolite concentrations in the plasma samples. Thus, it is important to recognize that manually adjusting Chenomx profiles will likely lead to a positive increase in metabolite concentrations. However, the comparative analysis of Chenomx to SMolESY and other studies from the scientific literature revealed fold changes that were systematically elevated. This may suggest that Chenomx either overestimates metabolite concentrations from the methanol-precipitated plasma samples or significantly underestimates concentrations from the ultrafiltered plasma samples. Conversely, the overall consistency in the results obtained by Gowda *et al.* (2014), Madrid-Gambin *et al.* (2023), and SMolESY is quite encouraging and indicates NMR-based metabolomics data is reproducible. Further, the consistent outcome also suggests that intact plasma samples may be a viable choice.

Finally, while our results align with previously reported data, 2-HB, choline, DMA, lactate, and glutamate were found to deviate from the measurements obtained for other metabolites. 2-HB and glutamate exhibit significant deviations regardless of sample preparation and processing methods. Thus, these metabolites may be specifically sensitive to sample preparation method and may need to be carefully considered before being classified as a biomarker or a group-differentiating metabolite. This work is not the first to investigate the effect of sample preparation on metabolomics studies and while a handful of other studies have similarly explored sample handling, processing and preparation concerns, the underlying issues to these protocols are complex. Accordingly, a robust and consensus best practice has not yet been reached, but our results reported herein provide an important next step to achieving accurate and reproducible metabolomics data.[29,35,45–47,49–51] As a final note, there is a natural biological variance in the metabolome due to a variety of factors that may be exacerbated by the choice of experimental protocols. Recognizing this challenge and minimizing non-biological var-

iance by the proper design and implementation of a metabolomics study is critical to its success.

Author contributions

R.P. and P.G.T. conceptualized the study and acquired funding. J.L.A., B.H.M., N.P, J.M., and P.G.T. acquired samples. D.C. completed NMR experiments. D.C. and P.G.T. completed data analysis. D.C., P.G.T., and R.P. wrote the manuscript. All authors read, revised, and approved the manuscript.

Data availability

The data supporting this article have been included as part of the ESI.†

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported in part by funding from the Nebraska Center for Integrated Biomolecular Communication (P20 GM113126, NIGMS) and by the Medical Research Council (MRC) and National Institute for Health Research (NIHR) (grant number MC_PC_12025) and the MRC UK Consortium for MetAbolic Phenotyping (MAP/UK) (grant number MR/S010483/1). J. L. A. was the recipient of a National Institute for Heath and Care Research (NIHR) Academic Clinical Lectureship, funded by Imperial College London and The Joyce and Norman Freed Charitable Trust. B. H. M. was the recipient of a NIHR Academic Clinical Lectureship (CL-2019-21-002) and is now the recipient of an MRC Clinician Scientist Fellowship. The Department of Metabolism, Digestion and Reproduction at Imperial College London receives financial and infrastructure support from the NIHR Imperial Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London.

References

1 J. R. Idle and F. J. Gonzalez, *Cell Metab.*, 2007, **6**, 248–351.
2 M. J. Jeppesen and R. Powers, *Magn. Reson. Chem.*, 2023, **61**, 628–653.
3 I. Subramanian, S. Verma, S. Kumar, A. Jere and K. Anamika, *Bioinf. Biol. Insights*, 2020, **14**, 1177932219899051.
4 R. P. Teklab Gebregiworgis, *Comb. Chem. High Throughput Screening*, 2012, **15**, 595–610.
5 J. Hong, L. Yang, D. Zhang and J. Shi, *Int. J. Mol. Sci.*, 2016, **17**, 767.
6 H. Tran, M. McConville and P. Loukopoulos, *J. Vet. Diagn. Invest.*, 2020, **32**, 635–647.
7 S. Muroya, S. Ueda, T. Komatsu, T. Miyakawa and P. Ertbjerg, *Metabolites*, 2020, **10**, 188.
8 C. H. Johnson, J. Ivanisevic and G. Siuzdak, *Nat. Rev. Mol. Cell Biol.*, 2016, **17**, 451–459.
9 M. Guasch-Ferre, S. N. Bhupathiraju and F. B. Hu, *Clin. Chem.*, 2018, **64**, 82–98.
10 Q. Jin, A. Black, S. N. Kales, D. Vattem, M. Ruiz-Canela and M. Sotos-Prieto, *Nutrients*, 2019, **11**, 207.
11 M. C. Kido Soule, K. Longnecker, W. M. Johnson and E. B. Kujawinski, *Mar. Chem.*, 2015, **177**, 374–387.
12 B. P. Lankadurai, E. G. Nagato and M. J. Simpson, *Environ. Rev.*, 2013, **21**, 180–205.
13 E. Valo, M. Colombo, N. Sandholm, S. J. McGurnaghan, L. A. K. Blackbourn, D. B. Dunger, P. M. McKeigue, C. Forsblom, P. H. Groop, H. M. Colhoun, *et al.*, *Sci. Rep.*, 2022, **12**, 4571.
14 M. Haid, C. Muschet, S. Wahl, W. Romisch-Margl, C. Prehn, G. Moller and J. Adamski, *J. Proteome Res.*, 2018, **17**, 203–211.
15 J. A. Cancelas and N. Rugg, *EBioMedicine*, 2016, **12**, 32–33.
16 R. E. Pincock, *Acc. Chem. Res.*, 1969, **2**, 97–103.
17 J. L. Alexander, N. J. Wyatt, S. Camuzeaux, E. Chekmeneva, B. Jimenez, C. J. Sands, H. Fuller, P. Takis, T. Ahmad, J. A. Doyle, *et al.*, *Gut*, 2024, **73**, 379.
18 H. E. Roth and R. Powers, *Cancers*, 2022, **14**, 3992.
19 M. Breier, S. Wahl, C. Prehn, M. Fugmann, U. Ferrari, M. Weise, F. Banning, J. Seissler, H. Grallert, J. Adamski, *et al.*, *PLoS One*, 2014, **9**, e89728.
20 E. J. Want, G. O'Maille, C. A. Smith, T. R. Brandon, U. Wilasinee, Q. Chuan, S. A. Trauger and G. Siuzdak, *Anal. Chem.*, 2006, **78**, 743–752.
21 D. W. Bearden, R. D. Beger, D. Broadhurst, W. Dunn, A. Edison, C. Guillou, R. Trengove, M. Viant and I. Wilson, *Metabolomics*, 2014, **10**, 539–540.
22 R. D. Beger, *Metabolomics*, 2018, **15**, 1.
23 R. D. Beger, W. Dunn, M. A. Schmidt, S. S. Gross, J. A. Kirwan, M. Cascante, L. Brennan, D. S. Wishart, M. Oresic, T. Hankemeier, *et al.*, *Metabolomics*, 2016, **12**, 149.
24 R. D. Beger, W. B. Dunn, A. Bandukwala, B. Bethan, D. Broadhurst, C. B. Clish, S. Dasari, L. Derr, A. Evans, S. Fischer, *et al.*, *Metabolomics*, 2019, **15**, 4.
25 J. A. Kirwan, H. Gika, R. D. Beger, D. Bearden, W. B. Dunn, R. Goodacre, G. Theodoridis, M. Witting, L. R. Yu, I. D. Wilson, *et al.*, *Metabolomics*, 2022, **18**, 70.
26 O. Fiehn, D. Robertson, J. Griffin, M. van der Werf, B. Nikolau, N. Morrison, L. W. Sumner, R. Goodacre, N. W. Hardy, C. Taylor, *et al.*, *Metabolomics*, 2007, **3**, 175–178.
27 R. Powers, E. R. Andersson, A. L. Bayless, R. B. Brua, M. C. Chang, L. L. Cheng, C. S. Clendinen, D. Cochran, V. Copié, J. R. Cort, *et al.*, *TrAC, Trends Anal. Chem.*, 2024, **171**, 117478.

28 D. G. Sitnikov, C. S. Monnin and D. Vuckovic, *Sci. Rep.*, 2016, **6**, 38885.
29 D. Vuckovic, *Anal. Bioanal. Chem.*, 2012, **403**, 1523–1548.
30 A. Hirayama, M. Sugimoto, A. Suzuki, Y. Hatakeyama, A. Enomoto, S. Harada, T. Soga, M. Tomita and T. Takebayashi, *Electrophoresis*, 2015, **36**, 2148–2155.
31 D. D. Marshall and R. Powers, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2017, **100**, 1–16.
32 S. Z. Abd Ghafar, A. Mediani, M. Maulidiani, R. Rudiyanto, H. Mohd Ghazali, N. S. Ramli and F. Abas, *Food Res. Int.*, 2020, **136**, 109312.
33 P. G. Takis, V. Ghini, L. Tenori, P. Turano and C. Luchinat, *TrAC, Trends Anal. Chem.*, 2019, **120**, 115300.
34 A. H. Emwas, R. Roy, R. T. McKay, L. Tenori, E. Saccenti, G. A. N. Gowda, D. Raftery, F. Alahmari, L. Jaremko, M. Jaremko, *et al.*, *Metabolites*, 2019, **9**, 123.
35 G. A. Nagana Gowda and D. Raftery, *Anal. Chem.*, 2014, **86**, 5433–5440.
36 A. L. Souza and G. J. Patti, *Methods Mol. Biol.*, 2021, **2276**, 357–382.
37 E. G. Bligh and W. J. Dyer, *Can. J. Biochem. Physiol.*, 1959, **37**, 911–917.
38 J. Folch, M. Lees and G. H. S. Stanley, *J. Biol. Chem.*, 1957, **226**, 497–509.
39 S. Dietmair, N. E. Timmins, P. P. Gray, L. K. Nielsen and J. O. Kromer, *Anal. Biochem.*, 2010, **404**, 155–164.
40 N. Matheus, S. Hansen, E. Rozet, P. Peixoto, E. Maquoi, V. Lambert, A. Noel, M. Frederich, D. Mottet and P. de Tullio, *Phytochem. Anal.*, 2014, **25**, 342–349.
41 B. B. Misra, *Metabolomics*, 2021, **17**, 49.
42 Z. Pang, J. Chong, G. Zhou, D. A. de Lima Morais, L. Chang, M. Barrette, C. Gauthier, P. E. Jacques, S. Li and J. Xia, *Nucleic Acids Res.*, 2021, **49**, W388–W396.
43 B. Worley and R. Powers, *ACS Chem. Biol.*, 2014, **9**, 1138–1144.
44 P. G. Takis, B. Jimenez, C. J. Sands, E. Chekmeneva and M. R. Lewis, *Chem. Sci.*, 2020, **11**, 6000–6011.
45 D. Vuckovic, in *Proteomic and Metabolomic Approaches to Biomarker Discovery*, 2020, pp. 53–83.
46 S. Tiziani, A. H. Emwas, A. Lodi, C. Ludwig, C. M. Bunce, M. R. Viant and U. L. Gunther, *Anal. Biochem.*, 2008, **377**, 16–23.
47 F. Madrid-Gambin, S. Oller, S. Marco, O. J. Pozo, C. Andres-Lacueva and R. Llorach, *Front. Mol. Biosci.*, 2023, **10**, 1125582.
48 S. Akoka, L. Barantin and M. Trierweiler, *Anal. Chem.*, 1999, **71**, 2554–2557.
49 S. Tulipani, R. Llorach, M. Urpi-Sarda and C. Andres-Lacueva, *Anal. Chem.*, 2013, **85**, 341–348.
50 D. Vuckovic and J. Pawliszyn, *Anal. Chem.*, 2011, **83**, 1944–1954.
51 S. Bruce, I. Tavazzi, V. Parisod, S. Rezzi, S. Kochhar and P. Guy, *Anal. Chem.*, 2009, **91**, 3285–3296.