# An Open Platform for Large Scale LC-MS-Based Metabolomics

Matthew R. Lewis*, Elena Chekmeneva, Stephane Camuzeaux, Caroline J. Sands, Ada H. Y. Yuen, Mark David, Ash Salam, Katie Chappell, Benjamin Cooper, Gordon A. Haggart, Lynn Maslen, María Gómez-Romero, Verena Horneffer-van der Sluis, Gonçalo Correia, Zoltan Takats

---

National Phenome Centre, Department of Metabolism, Digestion and Reproduction, Imperial College London, London W12 0NN, United Kingdom

and

Section of Bioanalytical Chemistry, Department of Metabolism, Digestion and Reproduction, Imperial College London, London SW7 2AZ, United Kingdom

* Correspondence: Matthew.Lewis@imperial.ac.uk

---

## Abstract

Metabolomics utilising liquid chromatography mass spectrometry (LC-MS) offers biomedical researchers a powerful means of assessing and comparing human phenotypes via measurement of the metabolome in biological samples. Platforms for LC-MS-based global profiling quantify hundreds or thousands of small molecule metabolites and/or lipids using combinations of distinct methods and analyses to develop broad coverage of the metabolome with high analytical sensitivity and specificity. However, the breadth of coverage provided by global profiling assays still outpaces efforts to characterise them by annotating profile signals with their respective metabolite identities. Fully realising the utility of metabolomics in biomedical research requires closing this gap by more accurately defining the finite metabolome coverage provided by common LC-MS-based global profiling methods. To date, method characterisation activities have progressed in the absence of broadly accepted standard LC methods as parallel efforts at building in-house libraries. While methodological diversity is a natural consequence of different design constraints and priorities observed across laboratories, it does tend to relegate in-house libraries to silos of information and

investment that fail to advance the broader metabolomics community. Here, the National Phenome Centre's (NPC) established platform for LC-MS-based global profiling of small molecule metabolites and lipids is made open in its entirety. Complete and detailed protocols for reversed-phase and hydrophilic interaction liquid chromatography LC-MS methods are offered alongside discussion of the rationale for their design specifics. In addition to the formal protocols used routinely within the Centre, the reader is provided with notes for replication and adaptation of the methodology, as well as guidance on the preparation of biofluid samples to ensure their suitability for the analytical platform. The Centre's accompanying open-source software for data extraction and pre-processing is also reviewed, and finally the method-specific identity of more than 700 small molecule and lipid species is disclosed. We hope that the substantial annotation information is useful to metabolomics practitioners of all experience levels and promotes the subsequent disclosure and constructive comparison (e.g. for validation and collective growth) of other in-house libraries and their associated methods. For interdisciplinary research teams looking to introduce LC-MS based metabolomics to their biomedical research programmes, we offer the open platform as a turnkey solution and welcome the growth in collective knowledge that may arise from its implementation in others' hands.

*Keywords: metabolomics, liquid chromatography, mass spectrometry, metabolite identification, lipidomics*

## Introduction

Metabolomics (including lipidomics, also known as metabolic phenotyping) is a powerful bioanalytical approach employed for the purposes of biomarker discovery in clinical and population-scale research.[1] Metabolomics studies are characterised by the inclusive measurement of the dynamic and versatile human metabolome(s)[2, 3] providing insight to physiological and pathological processes at a chemical level. This kind of measurement is enabled by a combination of specially adapted bioanalytical technologies, fit for purpose methods and protocols, and accompanying bioinformatic and chemometric tools for data extraction, pre-processing and interpretation. When used together, these resources constitute a platform for metabolomics.

Liquid chromatography mass spectrometry (LC-MS) technologies enable the isolation and measurement of small molecule metabolites and lipids in complex biological samples with a high degree of analytical sensitivity and specificity, making them a cornerstone in global profiling (un/non-targeted) metabolomics.[4] However, the resulting datasets are large and complex, typically containing tens of thousands of variables ranging in measurement quality and exhibiting a high degree of redundancy.[5, 6] Consequently, biological interpretation of global

profiling data is fruitless without considerable effort toward transforming analytical data into biochemical knowledge. This represents a substantial barrier to the success of LC-MS-based metabolomics in clinical and translational research. This knowledge can be built incrementally for each LC-MS method by establishing the chemical identity of each variable (or group of variables representing a metabolite's mass spectrum) and routinely assigning those identities to the variables observed in each LC-MS-based metabolomics study. This process can be arduous, however. The *de novo* annotation of metabolite-derived variables requires careful interpretation of experimental spectral data by experienced specialists, optionally aided by complex computational tools.[7] Furthermore, the procurement (or synthesis) and analysis of reference materials with accompanying wet lab experimentation is expected where possible, and required when reporting high confidence metabolite identification (MetID).[8, 9] Characterising the coverage of LC-MS profiling methods by this process is a mammoth task given the diversity and breadth of analytes present and detectable in biofluids.

To make matters worse, variety in the analytical instrumentation and consumable materials (LC stationary phases, predominantly) available to metabolomics practitioners, combined with their unique perceptions, considerations, and preferences in method design, has driven divergence among the LC-MS global profiling methods in use throughout the field. Chromatographic retention time values are empirically derived and specific to each assay, sensitive to even minute differences in column and stationary phase manufacture, mobile phase and gradient composition, and LC-MS system configuration. Unlike metabolite libraries containing mass spectral and collisional cross section data, the more context-specific nature of metabolite retention precludes the creation of global chromatographic libraries and ultimately hinders direct comparison of data between distinct LC-MS assays. Consequently, metabolite annotations cannot be easily transferred between laboratories using divergent methods, and a non-trivial amount of feature assignment work must be repeated for newly developed assays.

Standardisation of the chromatographic methods employed among laboratories for LC-MS-based global profiling is a practical solution to this problem. While methodological and broader platform diversity is helpful for the field in that it promotes healthy evolution and supports independent validation of findings (e.g., by independent cohorts analysed by differing platforms arriving at the same results), tension exists between the desire to develop cutting-edge methodology and its undermining effect on the characterization and consolidation of knowledge from stable (standardized) assays. Across laboratories, siloed development of method-specific in-house libraries is pursued at great expense in parallel efforts to break a field-wide bottleneck. Despite efforts toward standardisation in the metabolomics field[10-12], greater incentives appear to be required for success in standardising data-generating

methods. The open availability of such methods, disclosure of details and motivations driving key method-defining decisions, availability of quality documentation, demonstrated track record of performance and reproducibility, and accompanying open-source tools to facilitate interrogation of the resulting data are important factors for practitioners looking to adopt common methodology. However, gaining immediate knowledge of an extensive set of method-specific metabolite annotations and the potential for further development of that knowledge as a part of a larger collective of cooperative laboratories is arguably the most attractive incentive to adopt and use established methodology. Such an approach has been previously implemented for the substantial advancement of gas chromatography mass spectrometry (GC-MS) establishing a precedent for successful standardisation.[13] Sets of annotated metabolites are now being made publicly available for LC-MS methodology as well.[9, 14-16]

Here we summarise a conscientiously developed and extensively used set of LC-MS methodologies for human biofluid analysis (urine and blood products), complete with the sample handling and preparation protocols required to make those biofluids amenable to LC-MS analysis, a description of reference materials used for quality control (QC) assessment, and the open-source software used to extract and prepare the data generated for subsequent analysis. Furthermore, the *m/z* and retention time measurements for more than 700 biologically relevant metabolites detectable in biofluids are provided, bringing immediate value to researchers wishing to adopt these approaches to reversed-phase chromatography (RPC) and hydrophilic interaction chromatography (HILIC) separation for LC-MS analysis. Together, these materials (summarised in Figure 1 and Table 1) constitute the National Phenome Centre's open platform for LC-MS-based large scale metabolomics. Our aim in making this platform available in its entirety is to lower activation energy barriers involved in decentralising the application of metabolomics in biomedical research by distributing knowledge gained from centralised investment. By creating an open platform, we aim to help globalise metabolic profiling capabilities for greater impact across an increasingly interdisciplinary landscape. By adopting these methodologies for LC-MS global profiling, practitioners can benefit immediately from the knowledge contained herein, greatly reduce the resources spent on redundant assay development and MetID, and cooperate on building a more complete atlas of the metabolome observed in LC-MS-based global profiling.

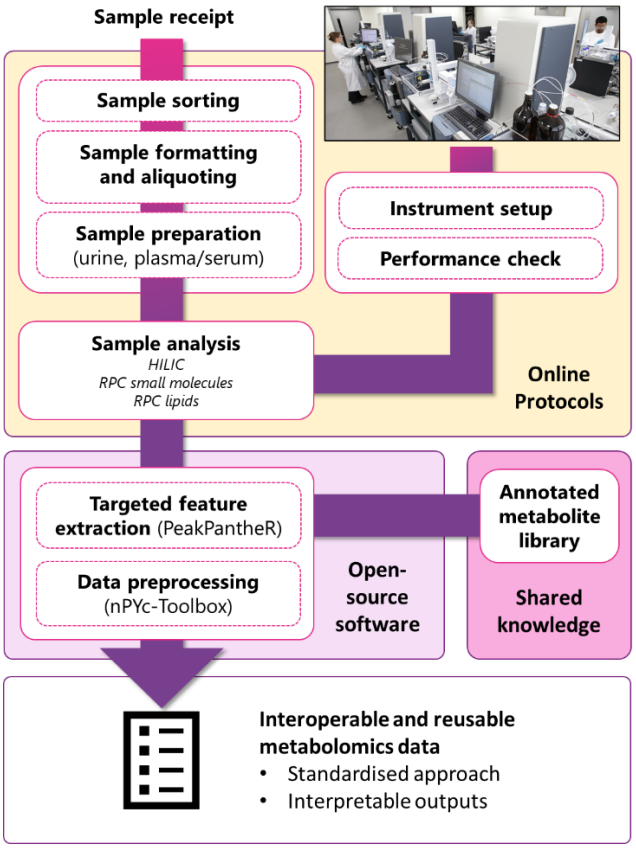


**Figure 1:** Summary of the workflow and materials available for the National Phenome Centre's open platform for LC-MS-based metabolomics.

**Table 1:** Available resources, including complete protocols, methods, software, and metabolite annotations.

| Resource | Description | Available At: |
|---|---|---|
| Experimental Protocols | Full experimental methods for NPC LC-MS profiling assays | https://github.com/phenomecentre/npc-open-lcms |
| LC-MS Metabolite Annotations | List of annotated compounds, retention time and m/z values for each assay | DOI: 10.5281/zenodo.5849038 |
| Targeted Extraction of Annotated Metabolites (PeakPantheR) | Comprehensive training materials and documentation for the targeted extraction of annotated metabolites from LC-MS global profiling data, including vignettes and exemplar data | https://github.com/phenomecentre/peakPantheR |
| Data Pre-processing and Quality Control (nPYc-Toolbox) | A Python implementation of the NPC toolchain for the import, pre-processing and quality control of metabolic profiling datasets | https://github.com/phenomecentre/nPYc-Toolbox |
| | Training materials and documentation for data pre-processing and QC, including Jupyter notebooks and exemplar data | https://github.com/phenomecentre/nPYc-toolbox-tutorials |

## Bioanalytical Methods

The analytical methods presented here are fit-for-purpose and performance-enhanced adaptions of preceding methods, each developed in 2012 and in continuous use since in a high-volume industrial-scale research environment. These methodologies are readily implemented for human urine and blood product analyses and adaptable to other clinical sample types. They are paired with bespoke open-source software for the targeted extraction of known metabolites and pre-processing of the data in order to generate high quality and high-fidelity datasets. The platform is scalable, enabling the analysis of 80 sample assays/day/instrument on a regular schedule which is operationally convenient for laboratories of any size.

■ *General parameters of bioanalytical methods:*

The methods were developed for now-conventional ultra-performance liquid chromatography systems with a maximum operating pressure of 15,000 psi. All methods are designed for ideal operation within 80% of this limit, providing some overhead for increases in system pressure often incurred during long sequences of sample analysis. The injection-to-injection cycle time was standardised to exactly 15 minutes across all methods, allowing the analysis of a complete 96-well plate of samples (80 study samples and 16 QC samples) in a 24-hour period. While convenient for day-to-day laboratory operations involving smaller projects, this feature becomes essential when applied to larger projects or concatenated smaller projects operated in a “continuous analysis” manner[17] during which plates are prepared and analysed sequentially without a break in analysis. This approach produces high precision data free from batch effects, with the added benefit of high instrument utilisation and efficiency. A regular 24-hour cycle under these conditions ensures predictable and easy to manage timing for the preparation and submission of samples for analysis. For example, one plate per instrument is prepared each morning and submitted for analysis in the early afternoon immediately following on from the previous plate, allowing some time for monitoring the plate’s analysis before leaving unattended overnight. This daily protocol is then repeated until all samples have been analysed.

■ *LC methods*

Three core separations are employed which together capture analytes present in human biofluids with wide ranging hydrophobicity (Figure 2). The first is an RPC method for the analysis of a broad range of analytes including smaller more hydrophilic species through to more moderately hydrophobic species. Very polar and ionic species (especially basic analytes) that are not well retained under these conditions are the targets of the HILIC

separation. Finally, more hydrophobic species, principally neutral and complex lipid species ranging from lysophospholipids to triglycerides that would not be cleanly eluted from the RPC method, are the targets of an adapted RPC method for lipidomics. The first two are applied routinely to the analysis of urine and blood products, while the latter is generally applied to blood products only (as well as tissue extracts and other clinical sample types outside the scope of this paper). Notes for use, standard operating procedures and proformas are available at https://github.com/phenomecentre/npc-open-lcms.

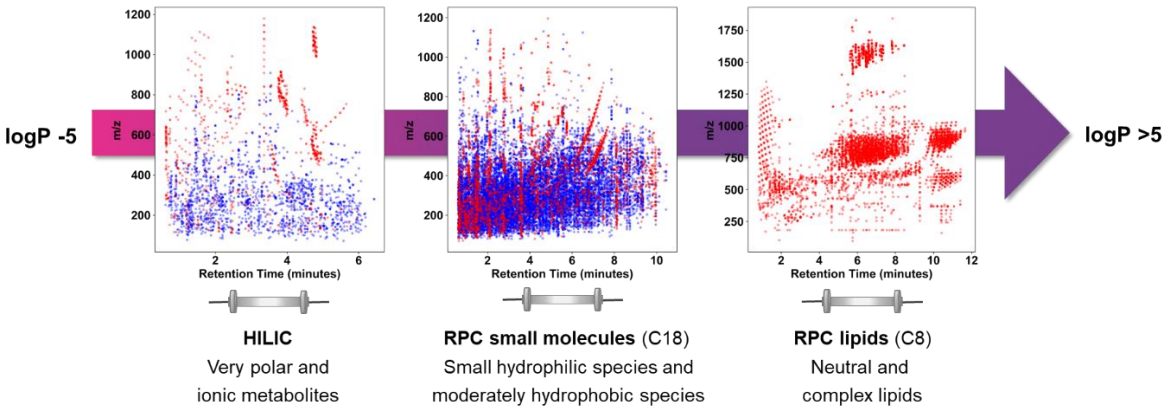


**Figure 2:** Metabolite profiles produced by the methods described herein, illustrating the field of metabolome capture across chemical space from smaller, polar, and ionic molecules (HILIC) to larger and more moderately hydrophobic molecules (RPC small molecules) and lipid species (RPC lipids). Two-dimensional feature maps (x = chromatographic retention time; y = *m/z*) show noise-filtered data (using the previously defined criteria of feature precision and correlation to dilution[17]) from selected studies where urine (blue features) and blood products (red features) were analysed. LogP values were calculated using the RDKit, an open-source cheminformatics toolkit (http://www.rdkit.org).

The RPC method adopts the low ligand density HSS (high strength silica) T3 ($C_{18}$) stationary phase (Waters Corp., Milford MA, USA) and mobile phase conditions previously employed by others including Wong et al[18] and later Want et al[19] for the metabolic profiling of urine, proving highly suitable for the retention of smaller and more hydrophilic analytes in highly aqueous environments as well as moderately hydrophobic analytes including drugs and drug metabolites. The adapted method substitutes a 150 mm column in place of the then-conventional and ubiquitously used 100 mm columns, improving the separation of early eluting metabolites.[17] This expanded column length was paired with a higher 0.6 mL/min flowrate, producing sharp peaks (approximately 1.5 seconds at baseline) across the gradient. This method, eventually published in 2016 after extensive testing and use, was important in establishing a new benchmark and “convention”.[20] It is used with both positive and negative mode ion detection.

The HILIC method was adapted from Want et al.[19], employing a simple and reliable unbonded ethylene bridged hybrid (BEH) particle (Waters Corp., Milford MA, USA) as the stationary phase. The mobile phase composition was adapted to use a pair of unblended solvents, eliminating the need for accurate volumetric preparation of a disproportionate (95:5 organic solvent-to-water) mix sensitive to even minor compositional variation.[21] The ultra-high performance liquid chromatography (UHPLC) instrument is instead used to mix the solvent pair during use in an accurate and reproducible manner, achieving the desired initial conditions for maximal analyte retention. The aqueous solvent is prepared with 20mM ammonium formate and both solvents are prepared with 0.1% formic acid, resulting in an ionic strength gradient across the separation. Importantly, this precludes the need to force volatile buffer salts into bulk solution with a mostly organic solvent which in our experience is an unreliable practice. The result is easy and efficient solvent preparation that yields exceptional separation reproducibility among solvent batches and experiments. The gradient used follows the same principle employed in the RPC method of higher flowrate and longer column, the latter being especially useful for HILIC because of the near-isocratic conditions used. The method is used with positive mode ion detection only.

The lipidomics method utilises an RPC approach that separates individual lipids and other hydrophobic metabolite species (as opposed to class-based lipid separations and direct infusion methods[22]). The method used herein was designed as an adaptation of a now-widely used approach (seemingly originating in the work of Castro-Perez et al[23]) which employs a C18 stationary phase and 90/10 mixture of 2-propanol and acetonitrile as the strong eluent. In an effort to enable higher solvent flow rate and improved peak capacity (benefitting isomeric structure separation), the strong eluent 2-propanol/acetonitrile ratio was modified from the original method (90/10) to 50/50, resulting in a less viscous eluent better suited to high flowrate delivery for higher peak capacity. The lower elution strength of the resulting mix was compensated by use of a shorter acyl chain bonded phase (C8 instead of C18), allowing the clean elution of triglycerides and longer chain lipids. Phosphoric acid was added to the mobile phase for the chromatographic peak shape enhancement of acidic phospholipids.[24] The addition of phosphoric acid requires special care of and attention to the composition of the LC pumps and pump heads (in our case, replacing more common grade 316 stainless steel parts with MP35N alloy parts available from the original equipment manufacturer) to reduce pump surface corrosion. The method is used with both positive and negative mode ion detection.

■ *MS coupling*

To achieve the intended system pressures when coupling to MS, the volume of post-column tubing is kept to an absolute minimum (33 cm or 38 cm of 0.10 mm ID tubing when used with

150 mm or 100 mm long columns, respectively), avoiding any post-column fluidics (i.e., those incorporated in the MS system to facilitate automated switching of inputs) to ensure highest chromatographic performance and reliability of the system. These methods were initially used with Q-ToF instruments (Waters Xevo G2-S) utilising Z-spray configured electrospray sources but have been compatibility tested with mass spectrometers from other instrument manufacturers. Any high-resolution MS instrument capable of adequately desolvating a 0.6 mL/min flow (challenging at 99% aqueous effluent) and a rapid scanning rate capable of producing more than 10 points per peak for accurate quantitation [25] (excluding any duty cycle consumed for alternative purposes including structural characterisation by MS/MS or dynamic range extension by interleaving scans with an attenuated beam) should be suitable for utilising these methods. When paired with Waters Xevo G2-S instruments, we observed a propensity towards in source fragmentation (ISF) owing to the nature of the ion optics. Consequently, acquisition of routine profiling data with simple MS1 is sufficient to generate rich spectra inclusive of diagnostic fragmentation patterns. This acquisition strategy is complemented with data-dependent analysis (DDA) MS/MS analyses on pooled QC samples (QC materials are discussed below in greater detail) and may be further augmented with targeted MS/MS analyses on specific samples of interest as needed. Finally, it should be noted that the time-of-flight MS instrument is regularly tuned to produce the highest resolution achievable at or near maximal sensitivity. By prioritising sensitivity over other instrument attributes (i.e. resolution in orthogonal acceleration time-of-flight mass spectrometry), the amount of sample required for each analysis can be minimised. In turn, larger scale analysis (batches of one to two thousand study samples) can be more practically achieved without stopping and disassembling for cleaning and maintenance. This is a central tenet of our “continuous analysis” approach to delivering high quality LC-MS raw data.

■ *Notes for replication and adaptation.*

The ability to completely reproduce and confidently adapt methodology is a fundamental tenet of an open platform (analogous to open-source software). Notes to promote the successful replication and adaption of these methods are given below.

When replicating these methods, special attention should be paid to the system pressure profiles produced with a new column. In our experience, despite the emphasis placed by column manufacturers on consistency of performance, different column and packing batches can result in columns that perform considerably better (albeit it with higher system pressure) or worse for profiling applications. In some cases, the pressure profiles for a column can be prohibitive for large project use. We routinely evaluate column pressure profiles and select columns with average (for their type) performance, ensuring that similar performing columns

can be regularly procured. Representative pressure traces are provided with the experimental protocols. Furthermore, if replicating these methods on different instrumentation, attention should be paid to the extra-column volume of the LC-MS system and consequential holdup time (see quality control section Table 2), as this will affect the accuracy of retention times, with a knock-on effect to method performance and the accuracy of method-specific analyte databases.

The methods presented here can be scaled for longer or shorter analysis times. To facilitate this, the methods adhere to simple linear gradients as much as possible, with only slight deviations in the HILIC and lipid methods. This differs substantially from more highly tailored methods which, in our experience for metabolic profiling of complex biofluids, are less reproducible and reliable. However, when shortening methods, it must be noted that a reduction in the number of features detected and a reduction in the precision of their measurement has been well described.[18] For this reason, shorter methods utilising the same UHPLC technologies may suffer a reduced ability to capture more subtle metabolic effects, although more rapid methods have been shown to be of use where the expected effect sizes are large and the analytes studied are dominant in the profiles.[26] For best results, the throughput provided by a platform should be carefully matched to (not needlessly exceeding) the actual needs of the laboratory. Researchers concerned with data quality should not overlook the addition of LC-MS instruments as a viable, albeit more costly, alternative to reliance on higher throughput methodology alone when looking to increase overall analysis capacity. Finally, the methods presented here can also be adapted for larger scale separations (e.g., using HPLC columns) to allow isolation of metabolites from biological samples in support of MetID efforts. We have achieved this using repeated HPLC-scale separations of urine samples without difficulty using a rescaled version of our RPC assay for small molecules.[27]

Beyond global profiling applications, the methods presented here can also be adapted for use as targeted assays. For example, when developing an assay for the direct analysis of bile acids without laborious sample preparation[28], the lipid RPC method was used as a foundation owing to its ability to cleanly elute very hydrophobic lipid species in blood, bile and other sample types of interest. Doing so precluded the need to separate lipids in the sample preparation step, streamlining the whole analytical procedure. Similarly, when developing a small molecule assay for tryptophan catabolism pathway analysis, the established lower hydrophobicity RPC profiling method (column chemistry and dimensions, column temperature, solvent system and initial conditions) was adapted, substantially accelerating the development of the targeted method.[29]

■ *Sample preparation*

Given the extensive characterisation of the methods, the concept followed is that samples should be adapted to a consistent and well-characterised analytical platform by conscientious preparation. This approach deviates from the practice of adapting specific methods to specific sample types, which multiplies the number of analytical methods which require characterisation in the form of MetID. Instead, by minimising the number of methods maintained within a laboratory, it becomes easier to deploy them more consistently (e.g., in the form of dedicating an instrument to a specific method type) and allows investment made in any one method in the form of characterising the elution of analytes (MetID) to be amortised across more sample analyses and types. Sample preparation SOPs for blood product and urine analysis by each method described above (RPC, HILIC, and Lipid RPC) are presented in the appendices and recommended as points of departure for other biofluids.

The RPC method is well suited to the analysis of aqueous samples lacking appreciable lipid and protein content. Human urine fits this description well, requiring only an optional dilution prior to direct analysis by RPC LC-MS.[17] The analytical method can also be used for more complex sample types where the lipids and proteins present are first removed as to not compromise the analysis by accumulating or precipitating, respectively, on the column leading to system pressure increases, unpredictable hydrophobic metabolite elution and ionisation suppression effects. However, the complexity of the sample preparation and potential for introducing selectivity bias rapidly increases. For example, precipitation of protein (required in samples with appreciable protein to prevent on-column precipitation, increased LC system pressure and ultimately column blockage) by addition of organic solvent either requires drying and solvent exchange (i.e. reconstitution in water) or injection of the organic-solvent-containing sample in sub-optimal conditions for the analysis, given that retention of small molecules in biofluid samples generally requires loading and initial chromatographic conditions to be as close to 100% aqueous as possible given the composition and tolerance of the stationary phase. Similarly, depletion of lipids can be conducted by laborious methods such as liquid/liquid biphasic separation or consumables-driven approaches such as solid-phase extraction (SPE). However, perturbation of the small molecule profile is a risk that must be minimised or avoided. Consequently, not all depletion methods are accepted as suitable for analysis of the small molecule content of a biological sample.[30-32] Finally, note that this method would not be suitable for analysis of lipophilic species (e.g. by extending the gradient to higher strength elution) owing to their insolubility in the requisite aqueous sample diluent and chromatographic initial conditions.

The HILIC method is more intrinsically adaptable to a wider variety of biological sample types including those rich with lipid and protein content. HILIC methods generally do not retain neutral lipids which therefore form a part of the injection peak.[33] The HILIC method presented here retains and elutes complex lipids within a compressed band to minimise overlap with the small molecule profile. Proteins are precipitated in the normal course of preparing the sample for HILIC which typically requires the addition of an organic solvent to the aqueous biofluid in order to better match the initial mobile phase conditions of the method which are typically between 90% and 95% acetonitrile. The method presented here utilises a 1:3 precipitation with ACN as a compromise approximation of the 95:5 starting conditions. Caution must be taken to avoid partitioning at very high percentage of acetonitrile specifically, and especially with high salinity samples, as this can lead to phase separation.[34, 35] Working at 4°C instead of -20°C (as might be preferred for protein precipitation alone) can help.

The lipid method, being a RPC method that initiates in a mixture of aqueous and organic solvents, is well served by organic extracts of biological samples which again generally precludes issues with protein precipitation on-column. Furthermore, our experience shows that the presence of small molecules in the sample does not pose a risk to the analysis of lipids by RPC LC-MS, as the former are handily eluted from the column prior to lipid elution and do not contribute a matrix effect (note that this would not be true for other methods e.g. DIMS that are outside our platform and beyond the scope of this manuscript). It is therefore unnecessary to perform complex and laborious biphasic separations, although lipid extracts prepared as such can certainly be analysed by the method. Rather, protein precipitation with 2-propanol as a strong organic solvent capable of solubilising complex and neutral lipids, e.g. from their native structures in blood products, has proven a simple and effective method for preparing blood products and is generally extendable to other lipid-containing biofluids.[36]

Finally, depending on the sensitivity achieved by the ionisation and MS system used, samples may require greater or lesser dilution to ensure that the majority of analytes are well situated within the linear dynamic range of the instrument for a given method.[37] Range and response assessment involving pilot sample sets and dilution series data from pooled representative samples can be a useful tool to guide the most appropriate sample dilution for a given LC-MS system (see more details of the quality control system below).[5]

## Quality Control

Measurement precision is monitored and assessed using three types of quality control reference materials: method reference mixtures, internal standards, and pooled QC samples. The composition and the role of each type of these materials is summarised in Table 2, and their integration with the sample handling and preparation workflow is illustrated in Figure 3.

**Table 2:** Quality Control reference materials composition and purpose in LC-MS assays

| Type of QC | Composition | Assays | Purpose |
|---|---|---|---|
| **Method Reference (MR)** | Mixture of heavy labelled small molecules added to pooled QC | Small molecules RPC<br>HILIC | LC-MS system monitoring |
| | Mixture of odd-chain lipids added to pooled QC and study samples | Lipid RPC | LC-MS system monitoring<br>Sample preparation<br>Injection precision<br>Matrix effect in study samples |
| **Internal Standards (IS)** | Mixture of heavy labelled small molecules added to pooled QC and study samples | Small molecules RPC<br>HILIC | Sample preparation<br>Injection precision<br>Matrix effect in study samples |
| **Study Reference (SR)** | Pooled QC sample composed of equal parts of all study samples | Small molecules RPC<br>HILIC<br>Lipid RPC | Analytical reproducibility<br>Precision<br>Data pre-processing<br>QC dilution series |
| **Long Term Reference (LTR)** | Externally procured pooled QC sample of studied biofluid | Small molecules RPC<br>HILIC<br>Lipid RPC | Analytical reproducibility<br>Precision<br>LC-MS system monitoring<br>Reference across multiple studies |

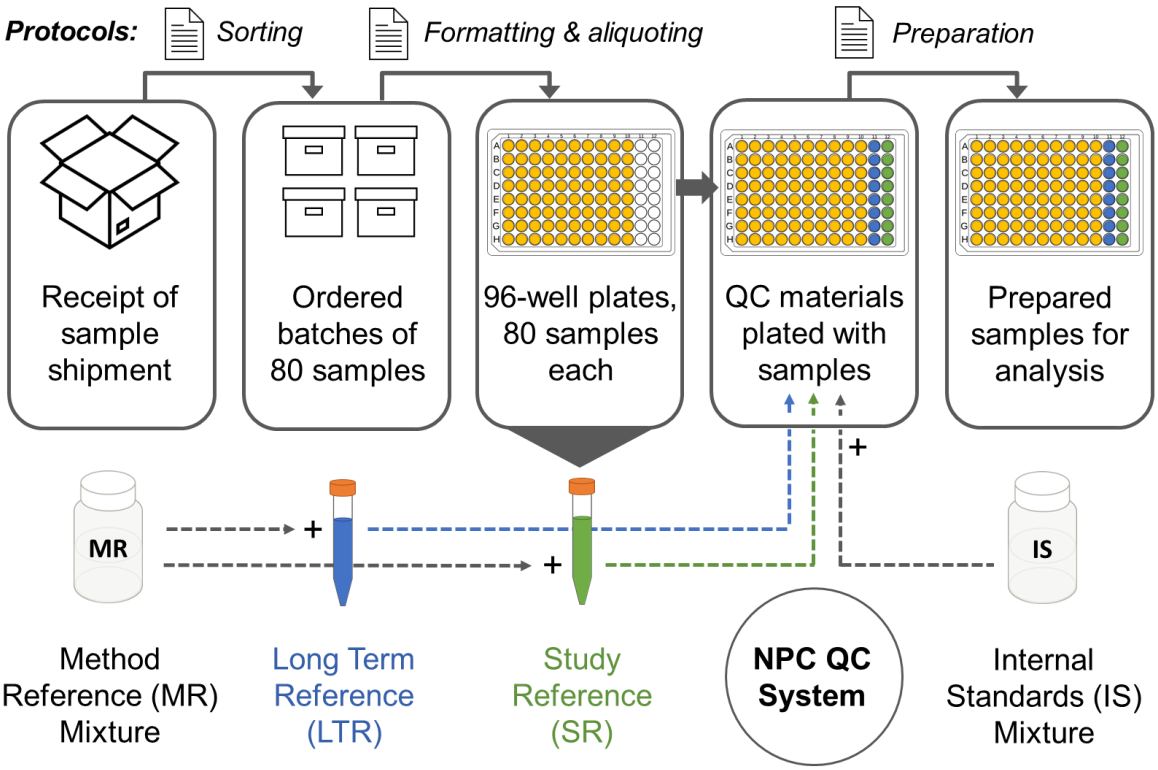


**Figure 3:** Schematic representation of the NPC quality control (QC) system for small molecule metabolite analysis, illustrating how reference materials are incorporated in the sample handling and preparation workflow. For the lipidomics workflow, method reference (MR) lipids are incorporated directly within the internal standards (IS) mixture.

Method Reference (MR) mixtures are method-specific sets of reference chemicals not expected to be present in the biological samples being studied (i.e., heavy labelled small molecules and non-endogenous odd-chain lipids). Each MR is specific to a chromatographic method (small molecule RPC, HILIC and lipid RPC) and formulated for use with both ionization modes, where applicable. The MR mixture is added to all pooled QC samples providing metabolite targets that represent the wider observable metabolome while facilitating a more real-time assessment of data quality. To allow assessment of sample preparation and injection precision as well as some limited assessment of matrix effect across individual study samples, internal standards (IS) are added to all pooled QC and study samples. When preparing samples for small molecule RPC or HILIC profiling, MS and IS mixtures are prepared in aqueous solutions and added to the sample prior to further preparative steps. For lipidomic profiling, a combined MR/IS mixture (referred to as LipidMix in the lipid method protocol) is added to all samples during protein precipitation with 2-propanol owing to the inability to solubilise and add them in an aqueous solution prior. The chemical composition of each MR and IS, the ions routinely monitored, and their method-specific retention times are reported in Table 3.

**Table 3:** Retention times and ion types of internal standards (IS) and method reference (MR) standards for each LC-MS method

| RPC (small molecules) | RT [min] | Positive mode ion | *m/z* | Negative mode ion | *m/z* |
|---|---|---|---|---|---|
| **Holdup time** | 0.44 +/- 0.02 | --- | --- | --- | --- |
| **L-Glutamine-13C5** | 0.56 +/- 0.01 | $[M+H]^+$ | 152.0937 | $[M-H]^-$ | 150.0781 |
| **L-Glutamic Acid-13C5** | 0.58 +/- 0.01 | $[M+H]^+$ | 153.0777 | $[M-H]^-$ | 151.0621 |
| **Creatinine-Methyl-d3** | 0.59 +/- 0.01 | $[M+H]^+$ | 117.0856 | --- | --- |
| **Cytidine-5,6-d2** | 0.90 +/- 0.04 | $[M+H]^+$ | 246.1059 | $[M-H]^-$ | 244.0902 |
| **L-Isoleucine-13C6,15N** | 1.49 +/- 0.02 | $[M+H]^+$ | 139.1196 | $[M-H]^-$ | 137.1039 |
| **L-Leucine-13C6** | 1.56 +/- 0.02 | $[M+H]^+$ | 138.1226 | $[M-H]^-$ | 136.1069 |
| **L-Phenylalanine-13C9,15N †** | 2.11 +/- 0.02 | $[M+H]^+$ | 176.1140 | $[M-H]^-$ | 174.0983 |
| **L-Tryptophan-13C11, 15N2** | 2.78 +/- 0.02 | $[M+H]^+$ | 218.1286 | $[M-H]^-$ | 216.1130 |
| **Hippuric Acid-d5 †** | 3.59 +/- 0.02 | $[M+H]^+$ | 185.0975 | $[M-H]^-$ | 183.0818 |
| **Benzoic Acid-Ring-13C6** | 5.35 +/- 0.04 | --- | --- | $[M-H]^-$ | 127.0491 |
| **Octanoic Acid-13C8** | 9.24 +/- 0.04 | --- | --- | $[M-H]^-$ | 151.1340 |

*† Compounds acting as IS*

*Continued on following page*

**Table 3:** *Continued from previous page*

| HILIC | RT [min] | Positive mode ion | *m/z* |
|---|---|---|---|
| **Holdup time** | 0.53 +/- 0.08 | --- | --- |
| **Uracil-2-13C,15N2** | 0.87 +/- 0.02 | $[M+H]^+$ | 116.0325 |
| **Hippuric Acid-13C6 †** | 1.18 +/- 0.03 | $[M+H]^+$ | 186.0862 |
| **Hippuric Acid-d5 ‡** | 1.18 +/- 0.03 | $[M+H]^+$ | 185.0975 |
| **Adenosine-2-d1 ‡** | 1.75 +/- 0.03 | $[M+H]^+$ | 269.1109 |
| **Adenine-2-d1** | 2.17 +/- 0.05 | $[M+H]^+$ | 137.0686 |
| **Taurine-15N** | 2.59 +/- 0.05 | $[M+H]^+$ | 127.0195 |
| **L-Tryptophan-d5-(indole)** | 3.74 +/- 0.21 | $[M+H]^+$ | 210.1291 |
| **L-Phenylalanine-13C9,15N †‡** | 3.84 +/- 0.24 | $[M+H]^+$ | 176.1140 |
| **Creatine-(methyl-d3)** | 5.05 +/- 0.08 | $[M+H]^+$ | 135.0961 |
| **L-Arginine-13C6** | 5.95 +/- 0.07 | $[M+H]^+$ | 181.1396 |

*† Compounds acting as IS in urine; ‡ Compounds acting as IS in blood products*

| RPC (lipids) † | RT [min] | Positive mode ion | *m/z* | Negative mode ion | *m/z* |
|---|---|---|---|---|---|
| **Holdup time** | 0.35 +/- 0.05 | --- | --- | --- | --- |
| **LPC (9:0/0:0)** | 0.57 +/- 0.01 | $[M+H]^+$ | 398.2308 | $[M+CH_3COO]^-$ | 456.2362 |
| **PC (11:0/11:0)** | 3.03 +/- 0.07 | $[M+H]^+$ | 594.4135 | $[M+CH_3COO]^-$ | 652.4190 |
| **C17:0** | 3.18 +/- 0.07 | --- | | $[M-H]^-$ | 269.2481 |
| **PG(15:0/15:0)** | 4.81 +/- 0.13 | $[M+Na]^+$ | 717.4682 | $[M-H]^-$ | 693.4707 |
| **PE(15:0/15:0)** | 5.91 +/- 0.11 | $[M+H]^+$ | 664.4917 | $[M-H]^-$ | 662.4761 |
| **PS(17:0/17:0)** | 6.01 +/- 0.17 | $[M+H]^+$ | 764.5442 | $[M-H]^-$ | 762.5285 |
| **PA(17:0/17:0)** | 6.3 +/- 0.16 | $[M+Na]^+$ | 699.4941 | $[M-H]^-$ | 675.4965 |
| **Cer(d18:1/17:0)** | 6.92 +/- 0.11 | $[M+H-H_2O]^+$ | 534.5249 | $[M-H]^-$ | 550.5199 |
| **DG(19:0/0:0/19:0)** | 9.26 +/- 0.1 | $[M+H-H_2O]^+$ | 675.5903 | --- | --- |
| **PC(23:0/23:0)** | 10.1 +/- 0.09 | $[M+H]^+$ | 930.7891 | $[M+CH_3COO]^-$ | 914.7578 |
| **TG(15:0/15:0/15:0)** | 10.38 +/- 0.1 | $[M+NH_4]^+$ | 787.6792 | --- | --- |
| **TG(17:0/17:0/17:0)** | 11.25 +/- 0.08 | $[M+NH_4]^+$ | 871.7731 | --- | --- |

*† LipidMix: all compounds act as both IS and MR standards*

To facilitate quality control and assessment on the scale of the true chemical complexity of biological samples, two types of biological pooled QC samples are used for global profiling studies. The first, a study reference (SR), is a pooled QC sample composed of equal parts of all samples in the study except any purposefully excluded. Here the SR is used in the traditional sense for assessing the reproducibility of feature intensity measurement.[38] Furthermore, by diluting the SR across a range of concentrations to form a set of "dilution series" QC samples, features can also be assessed for how well they are captured with linear dynamic range of the instrument, and for the proportionality of their response to dilution.[5] The SR also forms the basis of pre-processing steps such as batch and run-order correction (see data extraction and pre-processing methods section below).

A second pooled QC sample set was developed and is routinely integrated in profiling studies as a Long Term Reference (LTR)[17], allowing independent monitoring of precision within a study,[39] providing a point of reference for relating measurements across multiple profiling studies, and allowing the general monitoring of system performance across studies and years of operation. Over time, LTR materials (human urine, plasma and serum) have also become important as a system suitability test mixtures (SSTMs), allowing rapid assessment of an LC-MS system's accuracy (retention time, m/z and signal intensity measurement of spiked in MR and IS as well as some most abundant endogenous metabolites commonly detected in biofluid samples) after method-specific setup but before acquisition of study data. Each pooled QC sample is injected regularly throughout the sample analysis sequence (see Figure 4, following page).[17]

## Analysis Sequence

To provide the reader with a clear description of how independent studies are composed and analysed on the LC-MS system (regardless of the method used), an example is provided below (Figure 4). Each study begins with injections of blank, SR and dilution series (SRD) samples. Then, study sample plates as constructed in Figure 3 are prepared daily and appended to the analysis without a break in continuity. As many as 12 plates are consecutively analysed on a routine basis in our laboratory, yielding sample batches of nearly 1000 samples. In some cases, larger batch sizes have been implemented, but care must be taken to ensure good chromatographic peak shape is maintained and that signal is not lost to source fouling. Finally, the study concludes with the analysis of additional blank, SR and dilution series (SRD) samples. The dilution series is repeated here in a manner similar to best practice for targeted bioanalytical methods whereby calibration curves are analysed at the start and end of an analysis to assess changes in instrument response over the course of the study. The final region of the sequence is reserved for MS/MS data generation from repeat analysis of the pooled undiluted study reference sample. Data dependent analysis (DDA) is routinely used in our laboratory; however, data independent analysis (DIA) methods are equally applicable.

For small studies, the amount of system conditioning and the extent of replication within the dilution series sets may be disproportionately large to the number of study samples analysed. In this case, or in cases where less conditioning is thought to be necessary, these objectives can be met with less repetitive sample injection. The development of automatic gain control (AGC) for the purposes of compensating detector wear and stabilising the detection system's output [17] has indeed alleviated our dependence on heavy initial conditioning to ensure measurement precision, however we have maintained the practice.


**Figure 4:** A visual representation of the sequence of sample analysis within a study, illustrated as the sum of all signal intensity (y axis, with 100% representing that obtained from the undiluted study reference) for each sample analysed in the sequence (x axis). The delineated sequence regions are described as follows: (1) blank injections for assessing the stability and baseline of the LC-MS system, (2) undiluted study reference sample injections for conditioning the LC-MS system, (3) blank injections to assess and eliminate carryover, (4) dilution series sample injections for establishing feature filtering criteria based on range and response at the outset of the study, (5) study sample and interleaved study reference and long term reference sample analysis (two plates of a typical maximum of 12 plates shown), (6) final study reference sample injections, (7) reverse order dilution series sample injections for establishing feature filtering criteria based on range and response at the conclusion of the study, (8) blank injections for assessing the baseline of the LC-MS system at the conclusion of the study, and (9) undiluted study reference sample injections to support MS/MS data acquisition.

## Bioinformatics Methodology

■ *Data extraction and pre-processing methods:*

Data-processing and feature extraction of profiling LC-MS assays is usually performed using untargeted peak detection workflows that aim to comprehensively detect all the LC-MS peaks in a sample set. These workflows generate large data matrices comprised of thousands of features requiring *a posteriori* metabolite annotation. This is commonly achieved using a combination of computational methods to “group” features into “pseudo-spectra”, followed by spectral matching to databases or other experimental data sources. There are already many commercial and open-source software available for untargeted peak detection [40-43] and post-processing.[44, 45]

Investment in assay-specific chromatographic and spectral library building unlocks the possibility of deploying a more targeted feature extraction workflow. In this context, feature extraction is performed solely with the objective of detecting and integrating a set of well characterized and previously annotated signals. This approach provides users with a consistent output that includes a fixed set of variables and conclusive assessment of detection or non-detection of those metabolites in individual samples. We have developed an open-source R package, peakPantheR[46], which leverages the knowledge gained from assay

characterisation and enables targeted detection, peak fitting and integration of pre-annotated LC-MS features. The software also enables convenient visual review of individual peak performance across a study (Figure 5). The approach produces high-quality annotated datasets that are accessible to a broader range of biomedical researchers and more immediately interpretable than untargeted feature tables. Further details on PeakPantheR installation and usage can be found at https://github.com/phenomecentre/peakPantheR.

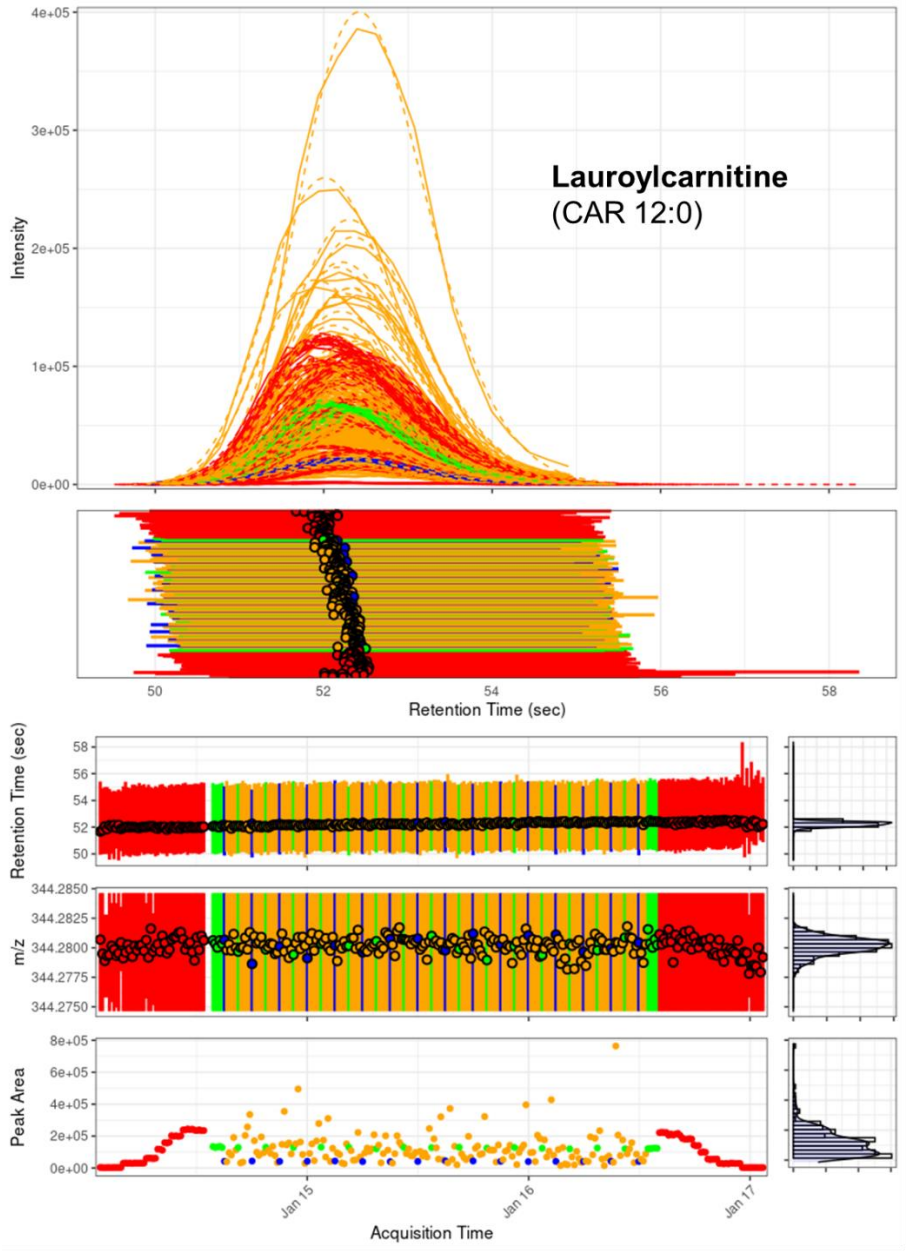


**Figure 5:** A real-world example of the visual summary provided for an individual metabolite extracted from global profiling data using PeakPantheR. From top to bottom, the output clearly illustrates the chromatographic peak shape, chromatographic peak width and apex tracking, and the trends and distribution for retention time, *m/z* and integrated peak area for all samples across the study.

Regardless of the feature extraction method used, quality control and feature filtering procedures should be applied to assess and ensure the quality of the dataset and individual LC-MS features. For this purpose, we have implemented routines for metabolomic data pre-processing and quality control in an open source Python software, the nPYc-Toolbox.[47] The nPYc-Toolbox provides functionality for batch and run order correction, feature filtering procedures based on repeated injections and dilution series of pooled QC samples, exploratory multivariate quality control, and summary reporting (Figure 6). Further details on nPYc-Toolbox installation and usage can be found at https://github.com/phenomecentre/nPYc-Toolbox and detailed tutorials at https://github.com/phenomecentre/nPYc-toolbox-tutorials.

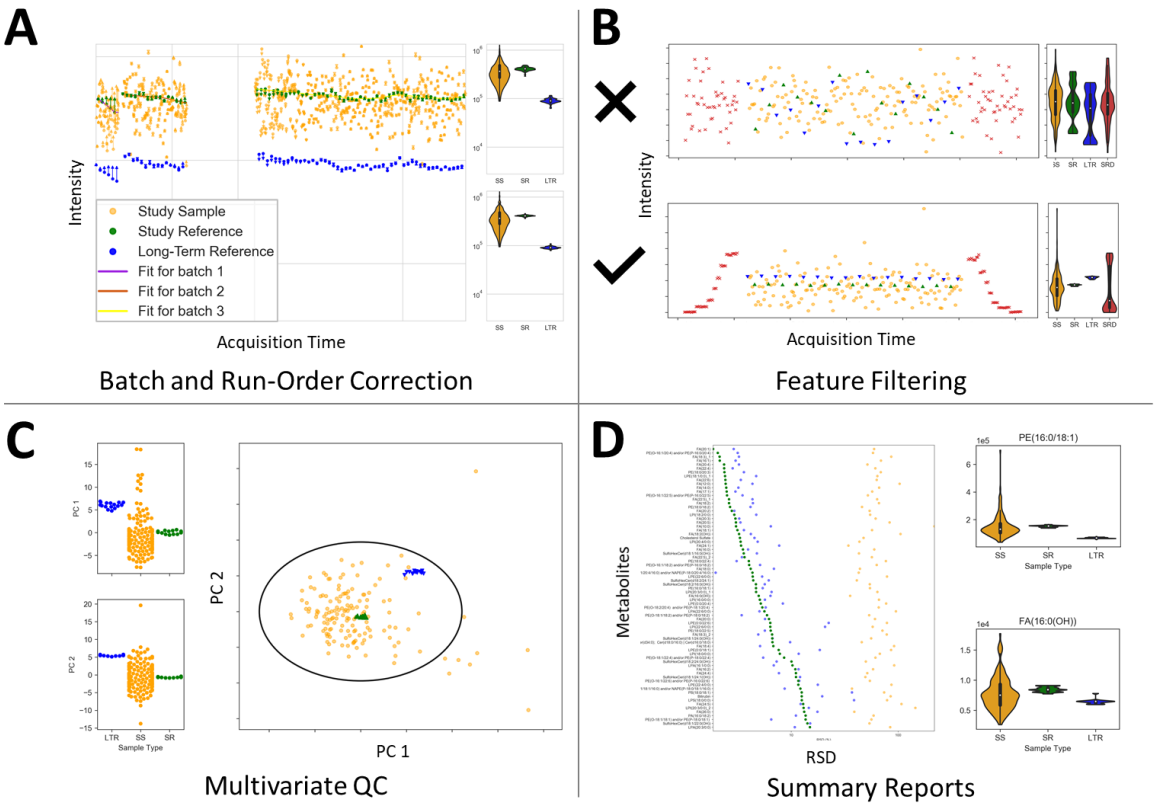


**Figure 6:** Real-world example outputs from the nPYc-Toolbox that aid the user in clearly visualising batch and run-order feature intensity correction on a per-feature basis (A), feature filtering on a per-feature basis (B), comprehensive study multivariate quality assessment (C), and a summary report of all named metabolites and their measurement quality characteristics (D). Study samples (SS) are shown in yellow, study reference samples (SR) in green, long-term reference samples (LTR) in blue and dilution series samples (SRD) in red.

## Method Characterisation

As discussed above, the classical metabolomics approach follows a workflow whereby complex analytical profiles are captured in sample analysis, optionally pre-processed (e.g., for quality control purposes) and analysed using univariate or multivariate statistical methods. This approach is powerful for biomarker discovery but can be frustrating for both metabolomics practitioners and their collaborators if important biomarkers are slow or difficult to identify *post-hoc*. For this reason, metabolic profiling laboratories often build in-house method-specific databases, empirical in nature and matching their chosen chromatographic separation conditions, for the characterisation (broad annotation) of their profiling methods. This action requires the procurement and analysis of chemical reference standards, building an investment-based activation energy barrier to adopting or even developing newer or different methodology. Beyond the personal preference of analytical chemists, this phenomenon might be one of the principal reasons that a wide variety of otherwise similar RPC and HILIC methods persist in the field. It therefore follows that standardisation requires the complete disclosure of the value associated with a given method or set of methods. For this reason, we provide here more than 700 annotated biologically relevant metabolites detectable in human biofluid samples in the context of the methods described (available at https://github.com/phenomecentre/npc-open-lcms). All annotations have been manually curated and whenever possible (for the majority of small molecules and some lipids) compared to the reference chemical material achieving the highest annotation confidence level of 1 according to MSI.[48] An overview of the small molecules and lipid species described is provided in Figure 7 (following page).

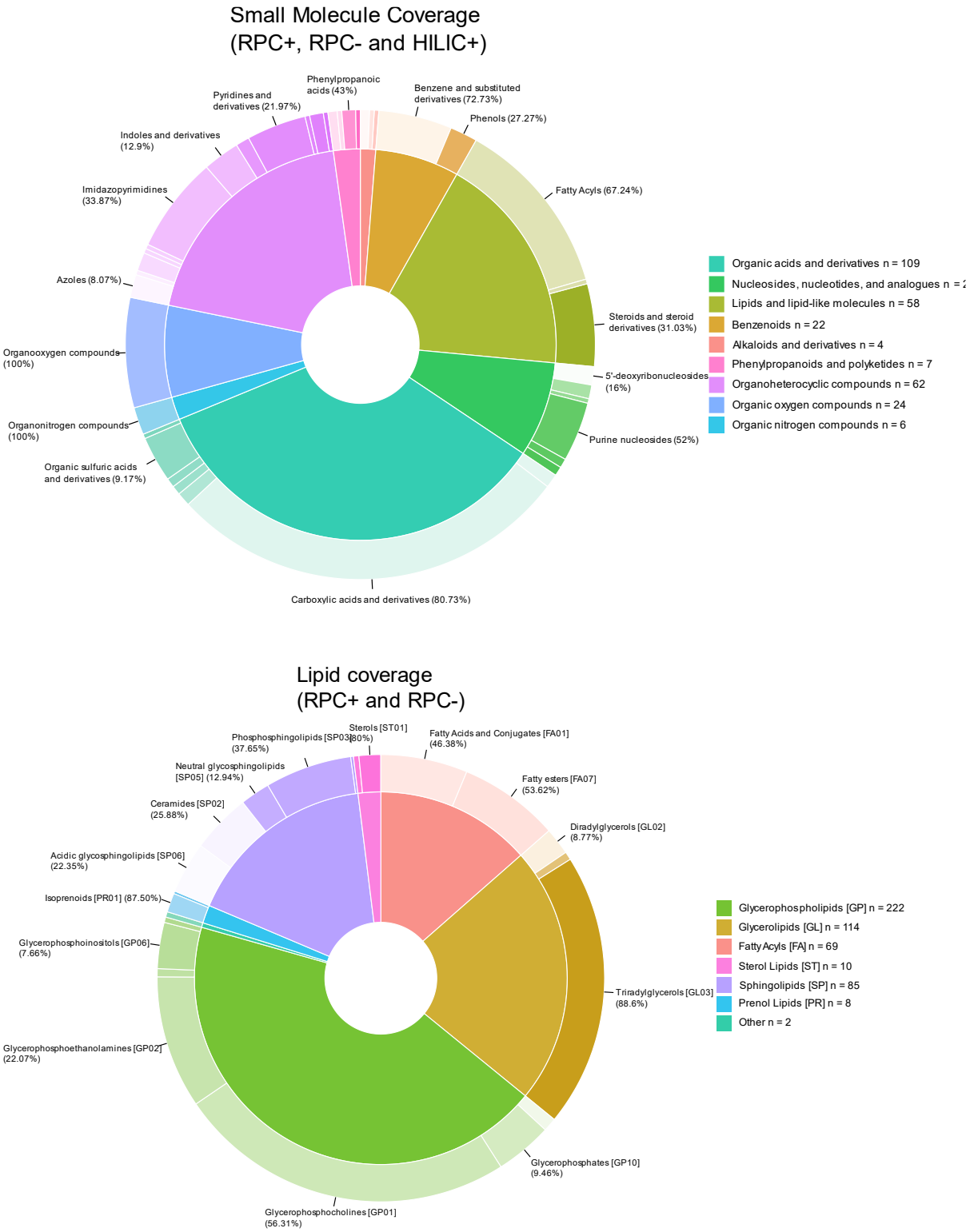


**Figure 7:** Summary of the annotated chemical coverage for the small molecule and lipid profiling assays. Small molecule annotations are grouped by their ClassyFire[49] superclass and class, and lipids by LIPID MAPS[50] main class and sub class (inner and outer pie chart, respectively).

## Conclusion

LC-MS-based global profiling technologies enable measurement of the small molecule and lipid content of biofluids and tissue extracts with excellent coverage and precision. However, the bioanalytical data produced (sets of retention time, *m/z* and signal intensity coordinates) is at least one step removed from biologically interpretable information, requiring linkage of those measurements to the small molecule metabolites and lipids they represent before the data are usable in the context of biomedical research and clinical practice. The process of MetID remains a bottleneck in the metabolomics workflow, limiting the perceived utility of LC-MS-based global profiling platforms in key application areas including clinical diagnostics development and molecular epidemiology. Given the resource and expertise requirements essential for MetID, cooperative efforts and information sharing across laboratories are clearly needed to break this bottleneck. However, the diversity in chromatographic methodology in use across the field poses some challenges to straightforward sharing of metabolite annotations, creating a case for methodological standardisation that we believe is reasonable only when substantial value is offered in the form of a highly characterised and open platform.

Here the National Phenome Centre's established platform for metabolic profiling is made open in its entirety, including complete protocols, methods, software, and metabolite annotations. The open platform leverages the experience of dedicated metabolomics practitioners, providing both a foundation for cooperation among like-minded experts and a turnkey solution for novice users. The latter is specifically helpful in promoting the application of metabolomics within multidisciplinary research teams as it helps to quickly establish reliable methodology and the means to interrogate the data it produces to derive biochemical insights. We advocate the suitability of the methods presented here for a wide variety of clinical sample types and encourage researchers to adapt new sample types to these methods rather than create new uncharacterised methods. However, their use does not preclude more hypothesis and need-driven expansion via the development of augmenting methods and technologies including those oriented toward more sensitive detection (e.g., for low abundance metabolites) or increased analytical specificity (e.g., chiral separations).

The sharing of method-specific annotations is core to the purpose of our promoting the National Phenome Centre's open platform for metabolomics. These annotations, each linked to the well-established methods detailed here, represent a substantial investment in method characterisation that we hope will not require repeating by our colleagues in the field (beyond that prudent and necessary for any further validation of their accuracy). We feel strongly that biochemical annotations in metabolomics should not be considered as protectable intellectual property, but rather as public knowledge which can be used, updated, and expanded by a

broader community of researchers. In this manner, the disclosure of global profiling methods and associated method-specific annotations is essential for facilitating this process, accelerating the unlocking of the real value held within global profiling, and democratising its development and utility across fields of application.

## Acknowledgements

This work was supported by the Medical Research Council UK Consortium for MetAbolic Phenotyping (MAP/UK) [grant number MR/S010483/1]. Infrastructure support was provided by the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (BRC). The authors wish to acknowledge the key role of Professors Jeremy Nicholson, Paul Elliott, and Frank Kelly in establishing the National Phenome Centre with funding provided by the Medical Research Council and National Institute for Health Research [grant number MC_PC_12025], allowing the methods described here to be piloted, published, and in the years since, developed further into the highly characterised standard platform described here.

## References

1. Trivedi, D. K.; Hollywood, K. A.; Goodacre, R., Metabolomics for the masses: The future of metabolomics in a personalized world. *New Horiz Transl Med* **2017,** *3* (6), 294-305.

2. Bouatra, S.; Aziat, F.; Mandal, R.; Guo, A. C.; Wilson, M. R.; Knox, C.; Bjorndahl, T. C.; Krishnamurthy, R.; Saleem, F.; Liu, P.; Dame, Z. T.; Poelzer, J.; Huynh, J.; Yallou, F. S.; Psychogios, N.; Dong, E.; Bogumil, R.; Roehring, C.; Wishart, D. S., The human urine metabolome. *PLoS One* **2013,** *8* (9), e73076.

3. Psychogios, N.; Hau, D. D.; Peng, J.; Guo, A. C.; Mandal, R.; Bouatra, S.; Sinelnikov, I.; Krishnamurthy, R.; Eisner, R.; Gautam, B.; Young, N.; Xia, J.; Knox, C.; Dong, E.; Huang, P.; Hollander, Z.; Pedersen, T. L.; Smith, S. R.; Bamforth, F.; Greiner, R.; McManus, B.; Newman, J. W.; Goodfriend, T.; Wishart, D. S., The human serum metabolome. *PLoS One* **2011,** *6* (2), e16957.

4. Gika, H. G.; Wilson, I. D.; Theodoridis, G. A., LC-MS-based holistic metabolic profiling. Problems, limitations, advantages, and future perspectives. *J Chromatogr B Analyt Technol Biomed Life Sci* **2014,** *966*, 1-6.

5. Sands, C. J.; Gomez-Romero, M.; Correia, G.; Chekmeneva, E.; Camuzeaux, S.; Izzi-Engbeaya, C.; Dhillo, W. S.; Takats, Z.; Lewis, M. R., Representing the Metabolome

with High Fidelity: Range and Response as Quality Control Factors in LC-MS-Based Global Profiling. *Anal Chem* **2021,** *93* (4), 1924-1933.

6. Mahieu, N. G.; Patti, G. J., Systems-Level Annotation of a Metabolomics Data Set Reduces 25000 Features to Fewer than 1000 Unique Metabolites. *Anal Chem* **2017,** *89* (19), 10397-10406.

7. Chen, L.; Lu, W.; Wang, L.; Xing, X.; Chen, Z.; Teng, X.; Zeng, X.; Muscarella, A. D.; Shen, Y.; Cowan, A.; McReynolds, M. R.; Kennedy, B. J.; Lato, A. M.; Campagna, S. R.; Singh, M.; Rabinowitz, J. D., Metabolite discovery through global annotation of untargeted metabolomics data. *Nat Methods* **2021,** *18* (11), 1377-1385.

8. Viant, M. R.; Kurland, I. J.; Jones, M. R.; Dunn, W. B., How close are we to complete annotation of metabolomes? *Curr Opin Chem Biol* **2017,** *36*, 64-69.

9. Blazenovic, I.; Kind, T.; Sa, M. R.; Ji, J.; Vaniya, A.; Wancewicz, B.; Roberts, B. S.; Torbasinovic, H.; Lee, T.; Mehta, S. S.; Showalter, M. R.; Song, H.; Kwok, J.; Jahn, D.; Kim, J.; Fiehn, O., Structure Annotation of All Mass Spectra in Untargeted Metabolomics. *Anal Chem* **2019,** *91* (3), 2155-2162.

10. Viant, M. R.; Ebbels, T. M. D.; Beger, R. D.; Ekman, D. R.; Epps, D. J. T.; Kamp, H.; Leonards, P. E. G.; Loizou, G. D.; MacRae, J. I.; van Ravenzwaay, B.; Rocca-Serra, P.; Salek, R. M.; Walk, T.; Weber, R. J. M., Use cases, best practice and reporting standards for metabolomics in regulatory toxicology. *Nat Commun* **2019,** *10* (1), 3041.

11. Rampler, E.; Abiead, Y. E.; Schoeny, H.; Rusz, M.; Hildebrand, F.; Fitz, V.; Koellensperger, G., Recurrent Topics in Mass Spectrometry-Based Metabolomics and Lipidomics-Standardization, Coverage, and Throughput. *Anal Chem* **2021,** *93* (1), 519-545.

12. Spicer, R. A.; Salek, R.; Steinbeck, C., A decade after the metabolomics standards initiative it's time for a revision. *Sci Data* **2017,** *4*, 170138.

13. Kind, T.; Wohlgemuth, G.; Lee, D. Y.; Lu, Y.; Palazoglu, M.; Shahbaz, S.; Fiehn, O., FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal Chem* **2009,** *81* (24), 10038-48.

14. Phapale, P.; Palmer, A.; Gathungu, R. M.; Kale, D.; Brugger, B.; Alexandrov, T., Public LC-Orbitrap Tandem Mass Spectral Library for Metabolite Identification. *J Proteome Res* **2021,** *20* (4), 2089-2097.

15. Meister, I.; Zhang, P.; Sinha, A.; Skold, C. M.; Wheelock, A. M.; Izumi, T.; Chaleckis, R.; Wheelock, C. E., High-Precision Automated Workflow for Urinary Untargeted Metabolomic Epidemiology. *Anal Chem* **2021,** *93* (12), 5248-5258.

16. Vorkas, P. A., Expanding lipidome coverage using MS/MS-aided untargeted data-independent RP-UPLC-TOF-MS(E) acquisition. *Bioanalysis* **2018,** *10* (5), 307-319.

17. 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., Development and Application of Ultra-Performance Liquid Chromatography-TOF MS for Precision Large Scale Urinary Metabolic Phenotyping. *Anal Chem* **2016,** *88* (18), 9004-13.

18. Wong, M. C.; Lee, W. T.; Wong, J. S.; Frost, G.; Lodge, J., An approach towards method development for untargeted urinary metabolite profiling in metabonomic research using UPLC/QToF MS. *J Chromatogr B Analyt Technol Biomed Life Sci* **2008,** *871* (2), 341-8.

19. Want, E. J.; Wilson, I. D.; Gika, H.; Theodoridis, G.; Plumb, R. S.; Shockcor, J.; Holmes, E.; Nicholson, J. K., Global metabolic profiling procedures for urine using UPLC-MS. *Nat Protoc* **2010,** *5* (6), 1005-18.

20. Letertre, M.; Munjoma, N. C.; Slade, S. E.; Plumb, R. S.; Swann, J.; Coen, M.; Nicholson, J. K.; Wilson, I. D., Metabolic Phenotyping Using UPLC–MS and Rapid Microbore UPLC–IM–MS: Determination of the Effect of Different Dietary Regimes on the Urinary Metabolome of the Rat. *Chromatographia* **2020,** *83* (7), 853-861.

21. Gray, N.; Heaton, J.; Musenga, A.; Cowan, D. A.; Plumb, R. S.; Smith, N. W., Comparison of reversed-phase and hydrophilic interaction liquid chromatography for the quantification of ephedrines using medium-resolution accurate mass spectrometry. *J Chromatogr A* **2013,** *1289*, 37-46.

22. Lisa, M.; Cifkova, E.; Khalikova, M.; Ovcacikova, M.; Holcapek, M., Lipidomic analysis of biological samples: Comparison of liquid chromatography, supercritical fluid chromatography and direct infusion mass spectrometry methods. *J Chromatogr A* **2017,** *1525*, 96-108.

23. Castro-Perez, J. M.; Kamphorst, J.; DeGroot, J.; Lafeber, F.; Goshawk, J.; Yu, K.; Shockcor, J. P.; Vreeken, R. J.; Hankemeier, T., Comprehensive LC-MS E lipidomic

analysis using a shotgun approach and its application to biomarker detection and identification in osteoarthritis patients. *J Proteome Res* **2010,** *9* (5), 2377-89.

24. Ogiso, H.; Suzuki, T.; Taguchi, R., Development of a reverse-phase liquid chromatography electrospray ionization mass spectrometry method for lipidomics, improving detection of phosphatidic acid and phosphatidylserine. *Anal Biochem* **2008,** *375* (1), 124-31.

25. Roemmelt, A. T.; Steuer, A. E.; Poetzsch, M.; Kraemer, T., Liquid chromatography, in combination with a quadrupole time-of-flight instrument (LC QTOF), with sequential window acquisition of all theoretical fragment-ion spectra (SWATH) acquisition: systematic studies on its use for screenings in clinical and forensic toxicology and comparison with information-dependent acquisition (IDA). *Anal Chem* **2014,** *86* (23), 11742-9.

26. Gray, N.; Adesina-Georgiadis, K.; Chekmeneva, E.; Plumb, R. S.; Wilson, I. D.; Nicholson, J. K., Development of a Rapid Microbore Metabolic Profiling Ultraperformance Liquid Chromatography-Mass Spectrometry Approach for High-Throughput Phenotyping Studies. *Anal Chem* **2016,** *88* (11), 5742-51.

27. Whiley, L.; Chekmeneva, E.; Berry, D. J.; Jimenez, B.; Yuen, A. H. Y.; Salam, A.; Hussain, H.; Witt, M.; Takats, Z.; Nicholson, J.; Lewis, M. R., Systematic Isolation and Structure Elucidation of Urinary Metabolites Optimized for the Analytical-Scale Molecular Profiling Laboratory. *Anal Chem* **2019,** *91* (14), 8873-8882.

28. Sarafian, M. H.; Lewis, M. R.; Pechlivanis, A.; Ralphs, S.; McPhail, M. J.; Patel, V. C.; Dumas, M. E.; Holmes, E.; Nicholson, J. K., Bile acid profiling and quantification in biofluids using ultra-performance liquid chromatography tandem mass spectrometry. *Anal Chem* **2015,** *87* (19), 9662-70.

29. Whiley, L.; Nye, L. C.; Grant, I.; Andreas, N.; Chappell, K. E.; Sarafian, M. H.; Misra, R.; Plumb, R. S.; Lewis, M. R.; Nicholson, J. K.; Holmes, E.; Swann, J. R.; Wilson, I. D., Ultrahigh-Performance Liquid Chromatography Tandem Mass Spectrometry with Electrospray Ionization Quantification of Tryptophan Metabolites and Markers of Gut Health in Serum and Plasma-Application to Clinical and Epidemiology Cohorts. *Anal Chem* **2019,** *91* (8), 5207-5216.

30. Chaker, J.; Kristensen, D. M.; Halldorsson, T. I.; Olsen, S. F.; Monfort, C.; Chevrier, C.; Jegou, B.; David, A., Comprehensive Evaluation of Blood Plasma and Serum Sample Preparations for HRMS-Based Chemical Exposomics: Overlaps and Specificities. *Anal Chem* **2022,** *94* (2), 866-874.

31. Ulmer, C. Z.; Jones, C. M.; Yost, R. A.; Garrett, T. J.; Bowden, J. A., Optimization of Folch, Bligh-Dyer, and Matyash sample-to-extraction solvent ratios for human plasma-based lipidomics studies. *Anal Chim Acta* **2018,** *1037*, 351-357.

32. Simon-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., 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. *Anal Chem* **2013,** *85* (24), 11725-31.

33. Wolrab, D.; Chocholouskova, M.; Jirasko, R.; Peterka, O.; Holcapek, M., Validation of lipidomic analysis of human plasma and serum by supercritical fluid chromatography-mass spectrometry and hydrophilic interaction liquid chromatography-mass spectrometry. *Anal Bioanal Chem* **2020,** *412* (10), 2375-2388.

34. Li, M.; Zhuang, B.; Lu, Y.; An, L.; Wang, Z. G., Salt-Induced Liquid-Liquid Phase Separation: Combined Experimental and Theoretical Investigation of Water-Acetonitrile-Salt Mixtures. *J Am Chem Soc* **2021,** *143* (2), 773-784.

35. Renard, J. A.; Oberg, A. G., Ternary Systems: Water-Acetonitrile-Salts. *Journal of Chemical & Engineering Data* **1965,** *10* (2), 152-155.

36. Sarafian, M. H.; Gaudin, M.; Lewis, M. R.; Martin, F. P.; Holmes, E.; Nicholson, J. K.; Dumas, M. E., Objective set of criteria for optimization of sample preparation procedures for ultra-high throughput untargeted blood plasma lipid profiling by ultra performance liquid chromatography-mass spectrometry. *Anal Chem* **2014,** *86* (12), 5766-74.

37. Cajka, T.; Smilowitz, J. T.; Fiehn, O., Validating Quantitative Untargeted Lipidomics Across Nine Liquid Chromatography-High-Resolution Mass Spectrometry Platforms. *Anal Chem* **2017,** *89* (22), 12360-12368.

38. Broadhurst, D.; Goodacre, R.; Reinke, S. N.; Kuligowski, J.; Wilson, I. D.; Lewis, M. R.; Dunn, W. B., Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* **2018,** *14* (6), 72.

39. van der Kloet, F. M.; Bobeldijk, I.; Verheij, E. R.; Jellema, R. H., Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping. *J Proteome Res* **2009,** *8* (11), 5132-41.

40. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M., MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **2010,** *11*, 395.

41. Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G., XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **2006,** *78* (3), 779-87.

42. Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M., MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods* **2015,** *12* (6), 523-6.

43. Yu, T.; Park, Y.; Johnson, J. M.; Jones, D. P., apLCMS--adaptive processing of high-resolution LC/MS data. *Bioinformatics* **2009,** *25* (15), 1930-6.

44. Kuhl, C.; Tautenhahn, R.; Bottcher, C.; Larson, T. R.; Neumann, S., CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal Chem* **2012,** *84* (1), 283-9.

45. Broeckling, C. D.; Afsar, F. A.; Neumann, S.; Ben-Hur, A.; Prenni, J. E., RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Anal Chem* **2014,** *86* (14), 6812-7.

46. Wolfer, A. M.; Correia, G. D. S.; Sands, C. J.; Camuzeaux, S.; Yuen, A. H. Y.; Chekmeneva, E.; Takats, Z.; Pearce, J. T. M.; Lewis, M. R., peakPantheR, an R package for large-scale targeted extraction and integration of annotated metabolic features in LC-MS profiling datasets. *Bioinformatics* **2021**.

47. Sands, C. J.; Wolfer, A. M.; Correia, G. D. S.; Sadawi, N.; Ahmed, A.; Jiménez, B.; Lewis, M. R.; Glen, R. C.; Nicholson, J. K.; Pearce, J. T. M., The nPYc-Toolbox, a Python module for the pre-processing, quality-control and analysis of metabolic profiling datasets. *Bioinformatics* **2019,** *35* (24), 5359-5360.

48. 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.; Thaden, J. J.; Viant, M. R., Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **2007,** *3* (3), 211-221.

49. Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; Greiner, R.; Wishart, D. S.,

ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J Cheminform* **2016,** *8*, 61.

50. Fahy, E.; Subramaniam, S.; Murphy, R. C.; Nishijima, M.; Raetz, C. R.; Shimizu, T.; Spener, F.; van Meer, G.; Wakelam, M. J.; Dennis, E. A., Update of the LIPID MAPS comprehensive classification system for lipids. *J Lipid Res* **2009,** *50 Suppl*, S9-14.