Permits non-commercial access & re-use, provided that author attribution and integrity are maintained; but does not permit creation of adaptations or other derivative works (https://creativecommons.org/licenses/by-nc-nd/4.0/).

analytical chemistry

ACS AUTHORCHOICE

pubs.acs.org/ac

Perspective

Representing the Metabolome with High Fidelity: Range and Response as Quality Control Factors in LC-MS-Based Global Profiling

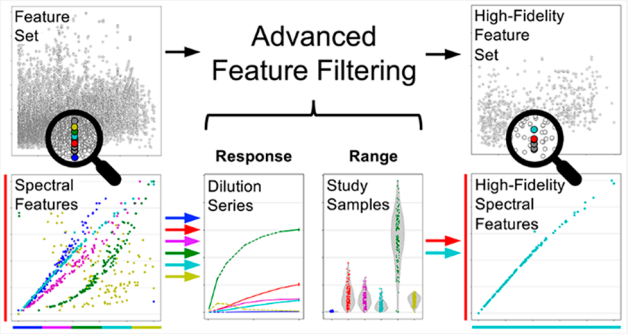
Caroline J. Sands,* María Gómez-Romero, Gonçalo Correia, Elena Chekmeneva, Stephane Camuzeaux, Chioma Izzi-Engbeaya, Waljit S. Dhillo, Zoltan Takats, and Matthew R. Lewis*

Cite This: https://dx.doi.org/10.1021/acs.analchem.0c03848

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Liquid chromatography–mass spectrometry (LC-MS) is a powerful and widely used technique for measuring the abundance of chemical species in living systems. Its sensitivity, analytical specificity, and direct applicability to biofluids and tissue extracts impart great promise for the discovery and mechanistic characterization of biomarker panels for disease detection, health monitoring, patient stratification, and treatment personalization. Global metabolic profiling applications yield complex data sets consisting of multiple feature measurements for each chemical species observed. While this multiplicity can be useful in deriving enhanced analytical specificity and chemical identities from LC-MS data, data set inflation and quantitative imprecision among related features is problematic for statistical analyses and interpretation. This Perspective provides a critical evaluation of global profiling data fidelity with respect to measurement linearity and the quantitative response variation observed among components of the spectra. These elements of data quality are widely overlooked in untargeted metabolomics yet essential for the generation of data that accurately reflect the metabolome. Advanced feature filtering informed by linear range estimation and analyte response factor assessment is advocated as an attainable means of controlling LC-MS data quality in global profiling studies and exemplified herein at both the feature and data set level.



Downloaded via IMPERIAL COLLEGE LONDON on January 15, 2021 at 16:37:51 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

Liquid chromatography–mass spectrometry (LC-MS) is a well-established technique for the quantitative measurement of chemical species in living systems (i.e., metabolomics and lipidomics).[1−3] Strategies for measuring the metabolome[4] can be targeted or untargeted with the former focusing on accurate quantification of predefined sets of metabolites and the latter aiming to more globally profile the chemical diversity present in biofluids and tissues.[5,6] While targeted methods are optimized to produce singular values for each chemical species (analyte) measured, untargeted approaches utilizing high resolution mass spectrometry record spectra composed of numerous distinct signals for each analyte. The result is a complex and inflated[7,8] data set, which can appear "overwhelming and impenetrable", especially to interdisciplinary researchers looking to metabolic profiling for novel insights.[9] To make matters worse, not all signals are equivalent in their representation of true analyte abundance, raising a key but broadly overlooked question: which signals most accurately characterize the variation in chemical species abundance across a set of samples? While addressed intrinsically in the analyte-specific design, validation, and rigorous quality control (QC) of targeted methods, the question is more difficult to answer post hoc when faced with a global profiling data set.

This Perspective details the challenges imposed by multiplicity and quantitative imprecision among features (grouped measurements of the same chemical species and ion type across all study samples) in LC-MS global profiling data and discusses QC strategies for ensuring the high-fidelity representation of the metabolome. These considerations are important for ensuring the accurate and efficient interpretation of results[10,11] and the translation of optimal biomarkers into the clinic where their reliability and performance are paramount.[12,13] To help orient the reader, an illustrative LC-MS metabolomics workflow is presented in Figure 1, which highlights key considerations for data QC and the reduction of feature multiplicity prior to analysis and interpretation. A comprehensive overview of key workflow stages including definitions, common terminology, and currently available open-source LC-MS metabolomics software is presented in Figure S1.

Received: September 10, 2020
Accepted: December 29, 2020

ACS Publications

© XXXX The Authors. Published by American Chemical Society

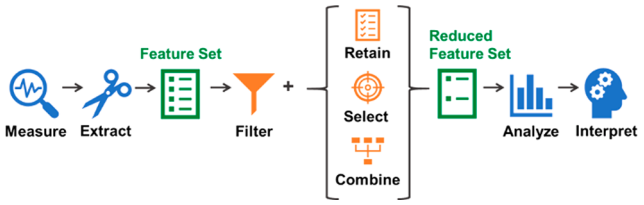


Figure 1. Simplified workflow for LC-MS metabolic profiling studies, highlighting key steps and logical options for reducing the feature set (highlighted in orange) prior to data analysis.

FEATURE MULTIPLICITY AND RESPONSE VARIATION

Chemical species are represented in LC-MS profiling data sets by multiple distinct features including those arising from (de)protonated molecules, isotopes, in-source (or in-instrument) fragments, adducts, oligomers, and multiply charged species.[8,14] This multiplicity confounds subsequent analysis by increasing the already large number of variables with knock-on effects to the number of simultaneous statistical tests performed, stringency of multiple testing correction procedures, and statistical power.[15−17] Collinearity (high correlation between variables) among these features imposes additional challenges in the form of biased estimators in regression and discriminant analyses.[18−20] However, LC-MS features arising from the same chemical species can also exhibit differing response functions (a response function in this context being a set of individual response factors, each of which is the ratio of analyte abundance to the magnitude of the corresponding recorded signal, that span a range of analyte abundances). Consequently, such features are not strictly redundant. This phenomenon is clearly visible in the analysis of minimally processed global profiling data sets, resulting in the dispersion of statistical significance among features derived from the same analyte (Figure 2). The observed variation in response is due to the idiosyncratic formation and transmission of ion species within the mass spectrometer and finite linear dynamic range of the LC-MS system, limited at the high end by saturation of the ion source and/or detector and at the low end by background interference and inconsistent detection.[21−25] Artifacts of on-instrument data processing (e.g., centroiding), the choice of feature extraction software and parameters,[26−28] and data preprocessing[7,29−31] compound the issue. Together, these introduce error into the data set that can further confound data modeling and the interpretation of results, even when using analysis strategies that deal well with modeling highly collinear data.[32,33]

To illustrate these effects at the level of individual chemical species, the quantitative responses of features within spectra derived from three common urinary metabolites were compared using data from an exemplar human metabolic profiling feature set (Figure 3). The variety of resulting pairwise relationships clearly illustrates the diversity of response functions possible among features derived from the same chemical species. Overcoming this phenomenon and enhancing the fidelity of global profiling feature sets require a consciously designed and executed strategy within the metabolomics workflow.

FEATURE REDUCTION

In order to improve data quality, statistical power, and interpretation clarity, it is attractive to consider reducing the multiplicity present within a data set prior to analysis. Yet, deciding whether or how to best select or combine features for these purposes can be challenging. The relation among distinct unidentified features within a global profiling data set is not obvious from the outset but can be inferred by the agreement in analyte retention time or elution profile, intensity correlation across a data set, or the detection of previously defined m/z differences among features (e.g., expected adduct types).[34,35] In practice, variations in response functions and in retention time, m/z, and intensity measurements naturally limit the accuracy and completeness of these approaches, especially when applied automatically to feature-rich global profiling data from chemically complex biological samples. Nonetheless, many tools have been developed to assist with the collation of spectral features, either for the purposes of feature reduction (Table S1) or more commonly to assist the identification of chemical species after analysis of the raw feature data (Figure S1).[34] Regardless of whether or how features are identified as belonging to the same spectrum, little consensus exists around what to do with their intensity measurements in metabolic profiling applications, despite the choice being formative for the final data set. The analyst is therefore faced with choosing one of the three logical approaches listed in the workflow (Figure 1).

The first option is to combine features belonging to the same spectrum. Tools using this approach typically sum[36] or average[37] their measurements in each sample, yielding a single value for the respective metabolite and thereby reducing the feature set and potentially reducing the influence of random measurement error among the distinct features.[38] However, this approach is highly dependent on the accuracy of spectral feature collation,

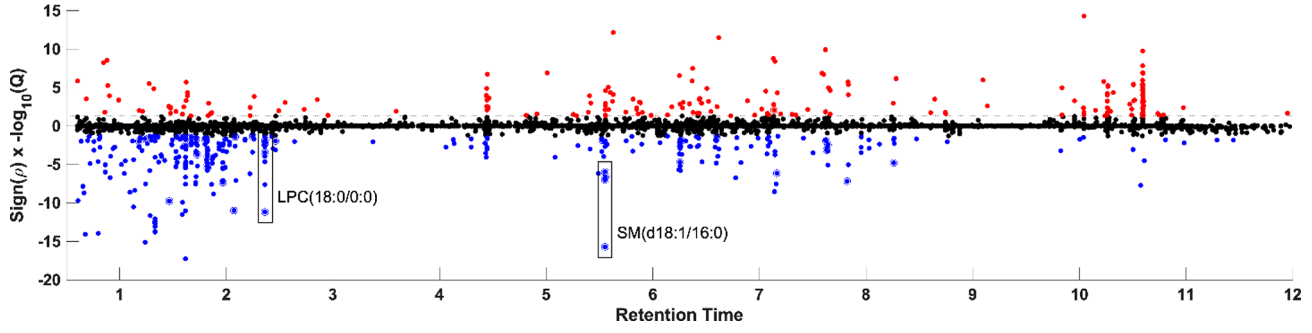


Figure 2. Manhattan-style plot illustrating the dispersion of test statistic values among metabolite-derived features in a comparative LC-MS lipid profiling study (see Methods S1 and S2 for more detail). Features passing false discovery rate correction are highlighted (blue and red). Features from two exemplifying lipid species are assigned for illustration: LPC(18:0/0:0) and SM(d18:1/16:0). Adapted with permission from Izzi-Engbeaya et al.[40] Copyright (2018) under the terms of the Creative Commons Attribution License (CC BY 4.0) John Wiley & Sons Ltd.

https://dx.doi.org/10.1021/acs.analchem.0c03848
Anal. Chem. XXXX, XXX, XXX−XXX

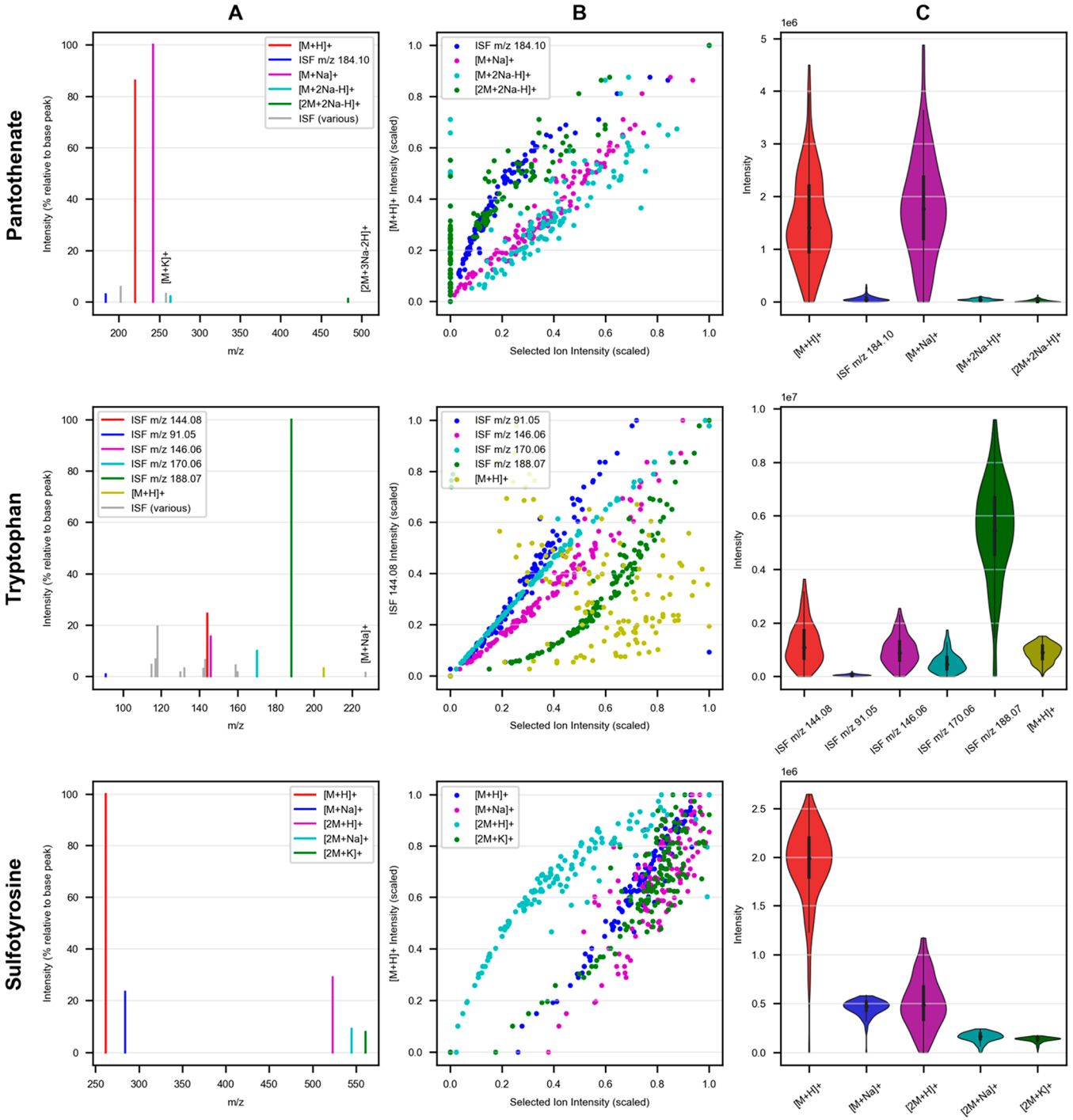


Figure 3. Variation in relative response among selected features including adducts and in-source fragments (ISF) originating from the same chemical species, illustrated across three common urinary metabolites (pantothenate, tryptophan, and sulfotyrosine; top, middle, and bottom rows, respectively) in a human metabolic profiling study (see Methods S3). Spectral annotations were confirmed against in-house reference spectra and those from the 2017 NIST Tandem Mass Spectral Library. Panel A contains mass spectra reconstructed from pooled QC sample m/z values and median intensities. For each analyte, the intensity values of selected features (various colors) across study samples were compared to those from a reference feature (shown in red) with the resulting pairwise relationships in intensity (scaled) illustrated in Panel B. Panel C shows the distribution in intensities observed across study samples for each selected feature. Features represented in each row of the panels are color-linked.

which must avoid both erroneous amalgamation of unrelated features and failure to collate all related features in a spectrum.[39] Furthermore, even where feature collation is performed accurately, the presence of poorer quality or lower fidelity measurements can directly compromise the combined metabolite measurement. This effect may be mitigated by more heavily weighting the contribution of selected (e.g., more intense) feature measurements to the total[38] or by utilizing factor decomposition methods,[37] provided that the highly weighted features are of high fidelity themselves (e.g., barring response saturation).

The second option is to select a single feature for each chemical species as a representative measurement. An intuitive strategy may be used whereby the "main" spectral feature is

https://dx.doi.org/10.1021/acs.analchem.0c03848
Anal. Chem. XXXX, XXX, XXX−XXX

selected, colloquially referring to either the (de)protonated molecule or the base (most intense) peak, which can be the same. The approach has the benefit of avoiding low abundance features such as in-source fragments or adducts that may be scattered about the limit of detection, and thus potentially confounding statistical analysis due to a high proportion of missing, imputed, or baseline integration values. However, there is no assurance that the main feature for any given spectrum provides the highest fidelity measurement. On a broader scale, when the feature set is biased toward more intense peaks, nonlinearity and saturation may become more prevalent, contributing to range compression[41] and peak picking artifacts.[42] Indeed, without established metrics to determine which feature best represents the abundance of a chemical species in solution, making an objective choice is impractical. Alternative approaches to feature selection have been proposed that additionally take clustering information into consideration, selecting the most representative feature from each "pseudo-spectrum" on the basis of both abundance and connectivity.[43,44]

The third option is to retain all features that pass QC and analyze the set without further selection or combination. Although this approach does not attempt to directly address the issues posed by feature set inflation, it avoids intervention and eliminates any risk of erroneous feature selection or combination. Statistical and machine learning analysis methods[20] such as PLS,[45] random forests,[46,47] LASSO,[48] and other penalized regression models[20,49] can produce reliable predictive models from multivariate and multicollinear metabolic profiling data. Furthermore, it can also be reassuring to the analyst when multiple features thought to belong to the same chemical species independently pass a statistical test. On the other hand, because of varying response factors and finite instrument linear dynamic range, the approach also runs the risk of leaving the analyst with a split within features derived from the same chemical species, with some failing to meet a threshold for statistical significance while others appear significant. In these scenarios, it is often tempting to ignore the former in favor of the latter, interpreting any statistically significant feature as indicative of a chemical's significance and opening a potential pitfall in the resulting biological interpretation.

Applications of various feature reduction approaches prior to statistical modeling have shown a potential in increasing predictive power for classification,[37,50] but regardless of the approach taken, the fidelity of the underlying measurements remains key. Individual spectral features with more accurate and therefore comparable response factors are more concordant in statistical testing, more interchangeable when selected, and both easier to identify as representing the same chemical species and less error prone to combine. Here, we focus on feature filtering as an opportunity to simultaneously reduce multiplicity and establish prerequisite control of feature quality, acting as a gatekeeper that ultimately yields more reliable and interpretable data.

FEATURE FILTERING STRATEGIES AND ADAPTING BEST PRACTICES FROM BIOANALYSIS

Measurement quality standards are well-established and rigorously defined for bioanalytical methods,[51,52] providing an aspirational framework for assessing and controlling data quality in global profiling studies.[53,54] Whereas data quality can be dictated in targeted analysis by prospective validation of assay performance for selected targets and the use of QC materials to monitor adherence to those expectations, QC practices in untargeted profiling have evolved from the same guidance for post hoc application.[55] In practice, data quality in global profiling is controlled by feature filtering, which removes feature set entries failing to exhibit one or more characteristics of the desired data quality. One of the earliest[56,57] and certainly one of the most common strategies used in metabolic profiling studies[54] is the assessment of technical precision (e.g., relative standard deviation, RSD) across repeated injections of a pooled QC sample and exclusion of features failing to meet a predefined threshold. This procedure is effective for ensuring a minimum of technical reproducibility among features in the resulting feature set; however, as a repeated measurement of a single sample, it does not take into account the variation in both precision and analyte response factor across the dynamic range of the instrument. Furthermore, in our experience with time-of-flight mass spectrometry, the approach suffers from the severe pitfall that noise regions and saturated peaks alike tend to produce reproducible signal intensities.

Other common filtering practices include the removal of features appearing in blank samples (e.g., background subtraction) and those appearing infrequently across the sample set (i.e., prevalence-based filtering).[54] A summary of filtering approaches employed by open-source software can be found in Table S2. However, none of these approaches directly select for features that most accurately represent the variation present in analyte abundance across a set of samples. It seems that, despite being a fundamental tenet of targeted quantitative LC-MS analysis and other bioanalytical techniques, the assessment of measurement accuracy and linearity of analyte response is broadly overlooked in metabolic profiling applications[11,41] and software tools (see Table S2). This is largely owing to both perceived theoretical and practical difficulties in adapting the practice to profiling data, where features of interest cover a wide range of concentrations and chemical classes and are determined post hoc, generally precluding the otherwise standard practice of using chemical standards to construct calibration curves.[54,57]

Nonetheless, a practice that integrates a more traditional bioanalytical assessment of the analyte response factor with the intent of establishing additional criteria for feature filtering in metabolic profiling studies is gaining traction. The approach, as proposed by Croixmarie et al., employs the systematic dilution of a dried pooled QC sample (e.g., in 3:1, 1:1, and 1:3 concentrations with respect to the original pool).[58] These "dilution series" samples are analyzed as an integral part of the full sample set, and correlation coefficients between analyte response and relative sample concentration are subsequently calculated for each feature. The resulting coefficients are then used to identify and remove features lacking a "linear or at least monotonous relationship" with sample analyte concentration, depleting uninformative, uninterpretable, and potentially misleading features from the data set.[58−60] While not yet widely used, the approach has been shown to improve robustness and quality of the data, discerning true biological signals from noise, background, and sample processing artifacts.[58−65] Yet, despite the simplicity and utility of the dilution series approach for feature quality assessment and filtering, at the time of writing, support for this is absent from virtually all major commercial and open source software packages with few notable exceptions (see Table S2).[63,66,67] This may be due to a lack of underlying consensus or guidance for how a diluted QC (dQC) sample-based strategy is best implemented.

It is also noteworthy that the calculation of correlation is inherently sensitive to the design of the QC sample dilution

https://dx.doi.org/10.1021/acs.analchem.0c03848
Anal. Chem. XXXX, XXX, XXX−XXX

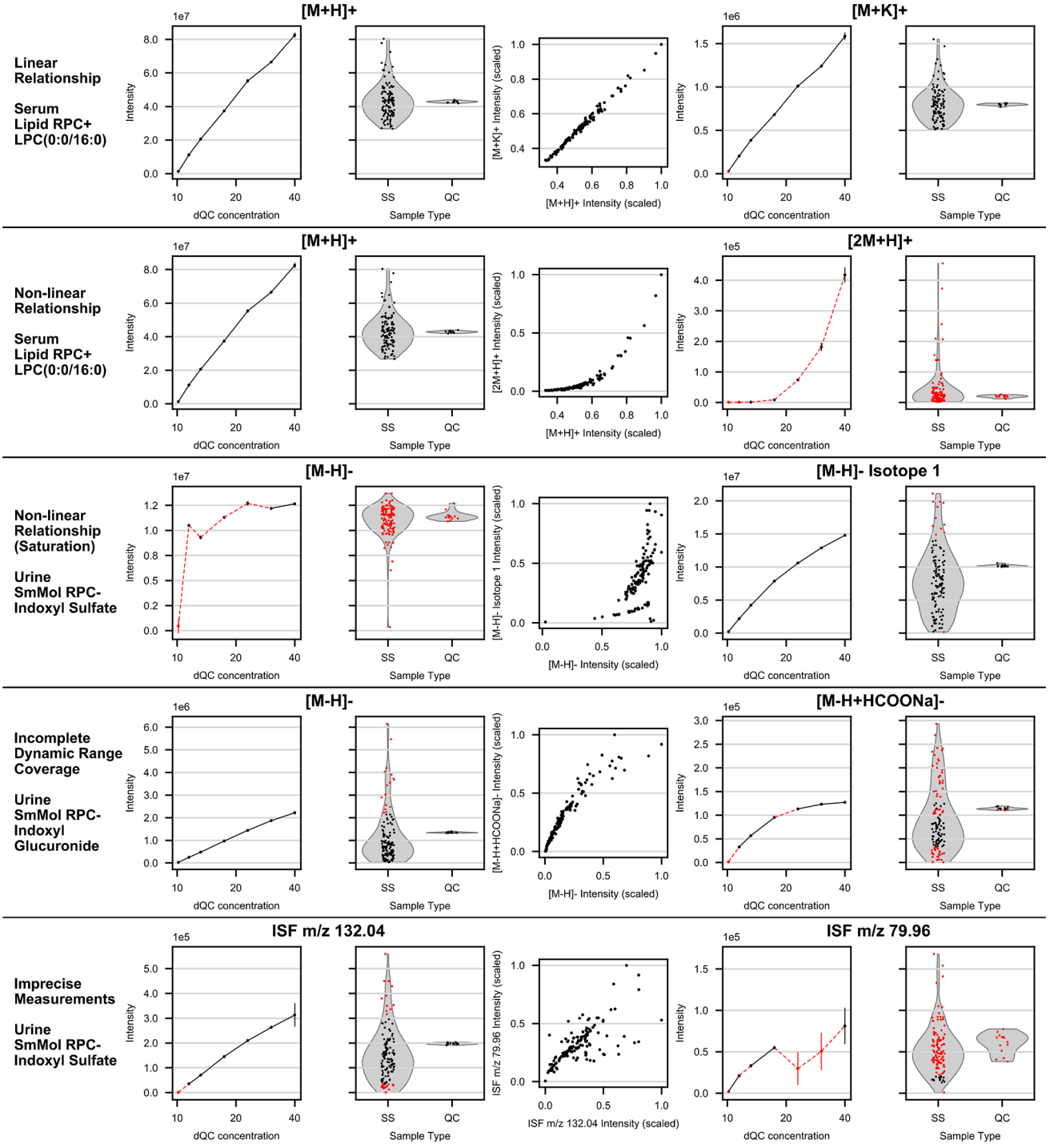


Figure 4. Exemplar feature relationships representing classes of observed behavior, illustrated for feature pairs for five chemical species from two LC-MS data sets: serum lipid RPC+ and urine small molecule (SmMol) RPC− (see Methods S1 and S3). Precision (RSD) was calculated for each set of dQC samples, and fold-change error (Method S4) was calculated between consecutively increasing dQC pairs. For each feature, plots show dQC median intensities (left) and violin plots of the observed intensities (right) in study samples and QCs. In the dQC plots, black points and connecting lines represent RSD $\leq$ 30 and fold-change error $\leq$ 20%, while measurements not meeting these criteria are highlighted in red. In the associated violin plots, any samples that fall in the regions with high RSD/fold-change error are similarly highlighted. The central panel for each feature pair illustrates the observed relationship between the selected features.

scheme. For example, the inclusion of dQC samples at both very low and very high concentrations can result in high correlation coefficients even when the behavior of intervening dQC samples is nonlinear with respect to response. Thus, high correlation coefficients can still be attained despite an obvious nonlinear response at the high and low ends of the response function provided the correlation is measured across enough points to dilute those effects. Consequently, it has been our experience that features exhibiting substantially different response functions (such as those illustrated in Figure 3) can still pass dQC series correlation filtering when reasonable thresholds (e.g., passing

https://dx.doi.org/10.1021/acs.analchem.0c03848
Anal. Chem. XXXX, XXX, XXX−XXX

features with ≥0.7 Pearson correlation coefficient) are used, warranting the further development of the design concept.

ADVANCED FEATURE FILTERING IN METABOLIC PROFILING

The improvement of dQC series correlation filtering requires an alternative assessment of the dilution series data based on the measurement of the proportionality of response at each dilution series point. Drawing from outside the field of metabolomics, in 2009, Popa-Burke et al. advocated the calculation of an average fold-change error between consecutive serial dilutions as a measure of relative accuracy when setting quality assurance criteria for serial dilution operations in liquid-handling equipment.[68] The method is particularly attractive for implementation in metabolic profiling studies because of its ability to generate measurements of relative accuracy across the dQC range, even for a priori undefined compounds. In metabolomics, however, applications utilizing fold-change between dQC/QC samples for error estimations are relatively sparse. Fold-change comparisons between QC and dQC samples were first used by Veselkov et al. to benchmark the performance of normalization and data transformations for LC-MS data analysis.[69] The concept was later adapted for the purposes of feature filtering in metabolomics data by calculating the dQC/QC ratio for each feature after regular injections of QC and dQC samples and removing features with poor stability (by ratio RSD) and response to dilution (by deviation from the expected ratio).[70] A conceptually similar approach has recently been explored to correct quantitative bias resulting from nonlinear ESI responses in metabolic profiling data.[41] In this work, Yu et al.[41] established calibration curves for every feature by modifying the injection volumes of a QC sample, mimicking a serially diluted QC sample. Subsequently, measured MS signals were fitted to these calibration curves, correcting signal intensity and generating projections of sample loading amounts for improved accuracy.

Although assessing measurement accuracy is impossible in the absence of absolute quantification, the inclusion of repeated dQC series samples provides a framework of known relative abundances that enables the adaptation of QC practices from bioanalytical guidelines for evaluating both accuracy and precision.[51,52,71] Doing so (1) enables the assessment of precision across an increased measurement range (beyond the average pooled QC), (2) allows the estimation of relative quantification accuracy and linear range limits by the calculation of fold-change error (Methods S4), and (3) defines the boundaries for QC evaluation relevant for interpreting study sample measurements.

The utility of this approach is readily demonstrated at the metabolite level when applied to selected features of known ion type and chemical assignment from real-world global profiling feature sets in which the dQC series design has been employed (Figure 4 and Methods S1 and S3 for experimental details). The acquisition of such samples allows the assessment of precision (RSD) at each dQC concentration and linearity by the calculation of pairwise fold-change error between consecutive dQC sample sets at different concentrations, with fold-change error defined as

$$\text{fold-change error}_{ij} = \frac{(\text{FC}_o - \text{FC}_e)}{\text{FC}_e} \times 100$$

where the expected and observed fold changes between two consecutive dQC sample sets at dilutions i and j are defined, respectively, as

$$\text{FC}_e = \frac{\text{dilution factor dQC}_j}{\text{dilution factor dQC}_i} \text{ and } \text{FC}_o = \frac{\text{median intensity dQC}_j}{\text{median intensity dQC}_i}$$

In addition, the comparison between the dQC series and acquired study sample measurements can be used to calculate the percentage of study samples that fall within the defined linear range. As illustrated in Figure 4, despite each feature passing basic filtering thresholds for precision (RSD in pooled QC samples ≤30) and the dQC series correlation to the dilution factor ($r \geq 0.7$), not all paired features provide equivalent representations of their respective analyte abundances. In the first example, the serum lipid LPC(0:0/16:0)'s protonated molecule and potassium adduct show excellent agreement in their response functions with virtually all study samples captured within the range covered by the dilution series and a good precision (RSD ≤ 30) and linear response (fold-change error ≤ 20) among all QC dilutions spanning the range occupied by the study samples. In the second example, response functions between the same protonated molecule and the $[2M + H]^+$ ion species reveal a nonlinear relationship. An inspection of the latter ion's dQC response function shows fold-change errors greater than 20% across the entire dilution series. This appears to be caused at the lower end by a change in the response factor as the signals in more diluted QC's fall outside of the linear dynamic range and ultimately below the detection limits and at the upper end by the disproportionately favored formation of oligomers.[72] As virtually all study sample measurements fall within this compromised range, this feature would be a good candidate for removal from the feature set. Consequently, the exemplar lipid LPC(0:0/16:0) is left to be more faithfully represented in the feature set by its protonated ion and potassium adduct features, which can be more confidently combined, selected, or retained for analysis.

Subsequent examples illustrate alternative response function relationships observed between pairs of features and provide insight into the distinct underlying phenomena affecting the fidelity of those feature measurements. These include response factor changes due to exceeding the linear dynamic range and saturation (e.g., the deprotonated molecule of indoxyl sulfate in urine), imprecision due to suspected saturation with incomplete dynamic range coverage (indoxyl glucuronide), and more general measurement imprecision, perhaps caused by background interference or analyte coelution, leading to both high RSDs and high fold-change error (e.g., indoxyl sulfate in-source fragment m/z 79.96). Together, these data clearly demonstrate that, although traditional QC RSD filtering and the calculation of dQC series correlation coefficients are not sufficient to remove all poorly measured features, the QC dilution series design does support a more thorough interrogation. With this in mind, we advocate two additional feature filtering criteria that further leverage dQC series data: (1) response filtering for the identification and removal of features where a high proportion of study samples fall within an intensity range that does not respond as expected to varying concentration (indicating poor measurement quality) and (2) range filtering for the identification and removal of features where a high proportion of study samples fall outside of the assessed dilution series range (and therefore measurement quality is unknown). The application of these advanced filtering steps in addition to those more traditionally used should ensure greater congruency

https://dx.doi.org/10.1021/acs.analchem.0c03848
Anal. Chem. XXXX, XXX, XXX−XXX

in the final feature set used for data analysis. While we propose the terms response filtering and range filtering for these purposes, we hesitate to propose strict mathematical definitions in favor of leaving them open to future implementation development by the metabolomics community as a whole.

To assess the broader impact of response and range filtering, a rudimentary implementation was applied to complete profiling feature sets (Figures 5 and S2) alongside dQC series correlation to dilution factor filtering and pooled QC RSD filtering. Across all feature sets, regardless of biofluid, assay, or polarity, the data reduction impact of advanced filtering is striking. Response filtering, when set to remove features with less than 80% of study sample measurements falling within an intensity range where dQC series fold-change error is <20%, is the dominant factor in data reduction. Range filtering reduces the feature set further, removing features with less than 80% of study sample measurements falling within the range covered by the dQC series samples. In the six data sets tested, the combined range and response filtering resulted in feature set reduction between 65% and 98%, similar in scale to previously reported results achieved by different means.[7] These two approaches capture virtually all features filtered by RSD and the correlation to dilution factor filtering, indicating that advanced filtering approaches may functionally supersede even the most established means of filtering-based QC in global profiling studies.

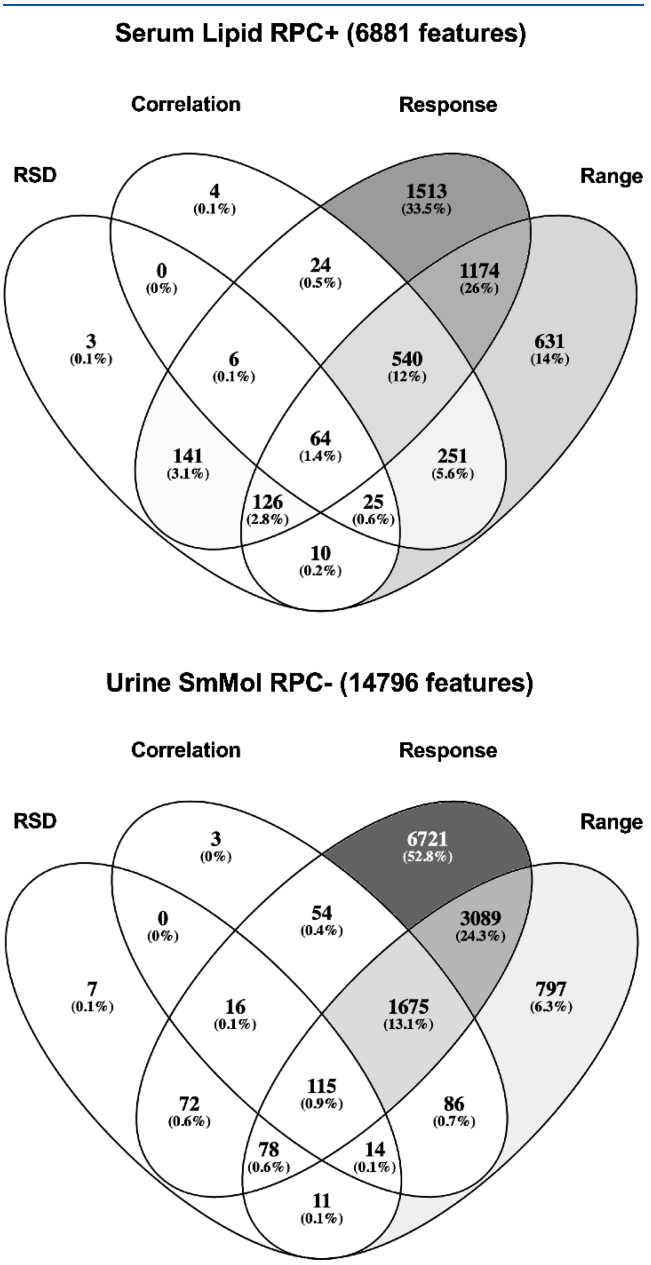


Figure 5. Assessment of response and range-based filtering methods applied to serum lipid RPC+ and urine small molecule (SmMol) RPC− LC-MS data sets (see Methods S1 and S3). Additional data sets are provided in Figure S2. In each Venn diagram,[73] the number of features not meeting each filtering strategy is given according to the following inclusion criteria: RSD in pooled QC samples ≤ 30 (RSD); Pearson correlation coefficient between dQC series and dilution factor ≥ 0.7 (correlation); greater than 80% of study samples within an intensity range where fold-change error ≤ 20% (response); greater than 80% of study samples within a range covered by the dQC series samples (range).

CHALLENGES AND DESIGN LIMITATIONS FOR THE IMPLEMENTATION OF ADVANCED FILTERING METHODS

It is our opinion that the application of the QC dilution series and an advanced filtering approach should be a standard prerequisite for generating high-fidelity untargeted LC-MS-based metabolic profiling data sets. However, numerous challenges, weaknesses, and limitations require consideration when developing and implementing such workflows. Chief among these are fundamental weaknesses inherited from the pooled QC sample approach from which the dQC series is derived. Despite being a widely used material in metabolic profiling, the use of the pooled QC sample itself[56,57] (an amalgamation of practices adapted from clinical chemistry[74,75] and bioanalysis in pharmacology[51,52]) is not a flawless cornerstone of QC. As an equal part composite of all study samples, it suffers the same drawbacks as an arithmetic mean in that it is susceptible to the influence of outliers. In practice, this can produce a disproportionate representation of analytes with an excessive abundance in a few samples (e.g., xenobiotics), creating a sample that, together with an average profile of all endogenous metabolites and lipids observed in all study samples, is unrepresentative in terms of its complexity. Conversely, lower abundance biomarkers present in fewer samples (e.g., potential biomarkers of subpopulations within nested case-control study designs, particularly in large scale molecular epidemiology applications) may not be well represented in the pooled QC sample. For these reasons, their repeated assessment in the composite average sample is a limited representation of the true quality of measurement in study samples. For use in response filtering, there is a further requirement that analytes are present in the pooled QC sample with sufficient abundance to withstand dilution over an appropriate (with respect to the abundance of analytes within the study samples) measurement range.

Second, when creating the dilution series from the pooled QC sample, it must be considered that all analytes are diluted simultaneously, affecting any given analyte together with its matrix. This approach satisfies the aim of creating a set of known relative abundances for each analyte, but it fails on two important fronts: (1) to model the response function of an analyte in a broadly unchanging matrix (e.g., as recommended for targeted analyses[51,52]) and (2) to account for matrix effects[22] introduced by sample specific variation in overall composition.[53]

https://dx.doi.org/10.1021/acs.analchem.0c03848
Anal. Chem. XXXX, XXX, XXX−XXX

Addressing the former is not possible using current profiling workflows as the exact chemical species coverage and analytes of interest are determined post hoc and the analyte-free matrix cannot be sourced.[76] However, the latter concern may be at least partially addressable, for example, by creating and analyzing (either sparsely or in totality) dilution series sets from subpools (e.g., of case and control samples) or individual study samples to assess the robustness of the analyte response to matrix effects and validate the QC dilution series-derived data.[77]

Third, the dilution of a pooled QC sample does not encapsulate the higher-than-average analyte concentrations present in study samples, and therefore, more concentrated dQC samples are needed. The approach proposed by Croixmarie et al. required the drying and reconstitution of the pooled QC sample, facilitating the creation of a 3-fold concentrated sample.[58] In cases where drying is not a routine component of the sample preparation workflow, the analyst may wish to avoid the step. Fortunately, the sensitivity of modern LC-MS instrumentation can be leveraged to circumvent this need for sample manipulation. A successful expansion of the effective concentration range coverage can be achieved by diluting all study samples and the pooled QC sample (e.g., by a factor of 2), retaining undiluted pooled QC sample material for use in creating the dilution series (illustrated in Figure S3). It has been our experience that this straightforward adaptation greatly improves the capture of analyte abundance present in most study samples within the range covered by the dilution series. A greater relative range can be created in the dilution series by further dilution of the study samples, albeit at the possible expense of low abundance metabolite coverage.

A final challenge relates to the actual strategic implementation of dQC sample analysis. In our laboratory where each study is bookended by dQC series samples as shown in Figure S3, we have observed that the extrapolation of the response precision is more complicated across larger study sizes where feature intensity fluctuations[59,78] are more likely and prominent. On-instrument stabilization of signal detection[59] is an ideal solution where available, but bioinformatics approaches (e.g., by applying a correction factor to the dilution series samples) and adaptations to the study design (e.g., by acquiring both QC and dQC series samples more regularly throughout the analysis run) could be valuable alternatives.

CONCLUDING REMARKS AND FUTURE DIRECTIONS

LC-MS global metabolic profiling assays generate large feature sets where individual chemical species are represented by multiple signals. While this multiplicity can enhance analytical specificity and be exploited to assist chemical assignment, it is a hindrance to downstream analysis and interpretation of a metabolic profiling data set. Ideally, steps in the metabolic profiling workflow should ensure that a final feature set is produced where each chemical species is represented by a single feature comprising a set of measurements that best represent true analyte concentration and exploit the instrument's available linear dynamic range. While a common consideration in bioanalytical techniques, most state-of-the-art experimental QC procedures, feature filtering approaches, and feature reduction methods for global profiling data overlook the otherwise key consideration of analyte linearity. Support for this is notably underrepresented in the myriad of options for feature filtering and reduction proposed and encoded in LC-MS preprocessing software. Consequently, we believe that there is an opportunity on this front to facilitate the improvement of global profiling data quality. Specifically, we believe the simplicity of implementation and value of the data generated from dQC series warrant their routine incorporation in metabolic profiling studies. The information revealed by this approach is critical for rigorous assessment and possible ranking of multiple ions from a single chemical species in terms of their quality and fidelity to the metabolome (within or even across multiple global profiling assays used in concert).

While the discussion and examples shown largely focus on advanced feature filtering, we also envision that the data provided by range and response calculations will be useful in the development of improved feature combination algorithms that go beyond the summation or averaging of features. Using detection limit information provided by the dilution series, it may become possible to extend the linear dynamic range in global profiling applications via a combination of isotopic measurements or fitting of intensity−calibration models between different ions.[79−81] An improved understanding of the underlying patterns and physical chemistry mechanisms responsible for diversity in feature response functions may avoid the loss of information by allowing a more sophisticated amalgamation of data.

For now, the assessment and assurance of feature quality remain paramount for the generation of high-fidelity metabolomics data sets. Here, we advocate that leveraging a dQC series design element to estimate the linear range and analyte response enables the implementation of advanced feature filtering methods inspired by best practices in bioanalysis. Support for these developing trends and their wider application relies on algorithms and options implemented within commercial and open-source software to more fully exploit the value of dilution series data. We believe efforts in these directions will help bridge the gap in quality between targeted bioanalysis and global metabolic profiling, advancing the field and practical applications in clinical biomarker discovery and beyond.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.0c03848.

Experimental details of serum and urine feature sets; further description of the data shown in Figure 2; overview of key stages, definitions, and software in the metabolomics workflow; details of current strategies and software for feature reduction and feature filtering (PDF)

AUTHOR INFORMATION

Corresponding Authors

Caroline J. Sands − *National Phenome Centre, Department of Metabolism, Digestion and Reproduction, Imperial College London, London W12 0NN, United Kingdom; Section of Bioanalytical Chemistry, Department of Metabolism, Digestion and Reproduction, Imperial College London, London SW7 2AZ, United Kingdom*; Email: Caroline.Sands01@imperial.ac.uk

Matthew R. Lewis − *National Phenome Centre, Department of Metabolism, Digestion and Reproduction, Imperial College London, London W12 0NN, United Kingdom; Section of Bioanalytical Chemistry, Department of Metabolism, Digestion and Reproduction, Imperial College London, London SW7*

https://dx.doi.org/10.1021/acs.analchem.0c03848
Anal. Chem. XXXX, XXX, XXX−XXX

2AZ, United Kingdom; orcid.org/0000-0001-5760-5359; Email: Matthew.Lewis@imperial.ac.uk

Authors

María Gómez-Romero − *National Phenome Centre, Department of Metabolism, Digestion and Reproduction, Imperial College London, London W12 0NN, United Kingdom; Section of Bioanalytical Chemistry, Department of Metabolism, Digestion and Reproduction, Imperial College London, London SW7 2AZ, United Kingdom*

Gonçalo Correia − *National Phenome Centre, Department of Metabolism, Digestion and Reproduction, Imperial College London, London W12 0NN, United Kingdom; Section of Bioanalytical Chemistry, Department of Metabolism, Digestion and Reproduction, Imperial College London, London SW7 2AZ, United Kingdom*

Elena Chekmeneva − *National Phenome Centre, Department of Metabolism, Digestion and Reproduction, Imperial College London, London W12 0NN, United Kingdom; Section of Bioanalytical Chemistry, Department of Metabolism, Digestion and Reproduction, Imperial College London, London SW7 2AZ, United Kingdom;* orcid.org/0000-0003-1807-2398

Stephane Camuzeaux − *National Phenome Centre, Department of Metabolism, Digestion and Reproduction, Imperial College London, London W12 0NN, United Kingdom; Section of Bioanalytical Chemistry, Department of Metabolism, Digestion and Reproduction, Imperial College London, London SW7 2AZ, United Kingdom*

Chioma Izzi-Engbeaya − *Section of Endocrinology and Investigative Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London W12 0HS, United Kingdom*

Waljit S. Dhillo − *Section of Endocrinology and Investigative Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London W12 0HS, United Kingdom*

Zoltan Takats − *National Phenome Centre, Department of Metabolism, Digestion and Reproduction, Imperial College London, London W12 0NN, United Kingdom; Section of Bioanalytical Chemistry, Department of Metabolism, Digestion and Reproduction, Imperial College London, London SW7 2AZ, United Kingdom;* orcid.org/0000-0002-0795-3467

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.0c03848

Author Contributions

The manuscript was conceived and written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Medical Research Council and National Institute for Health Research [grant number MC_PC_12025] and 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 Section of Endocrinology and Investigative Medicine is funded by grants from the MRC, BBSRC, and NIHR and is supported by the NIHR Biomedical Research Centre Funding Scheme. The views expressed are those of the authors and not necessarily those of the MRC, NIHR, or the Department of Health. The following authors are also funded as noted: C.I.-E., Imperial College-BRC Postdoctoral Fellowship [grant number P79696]; W.S.D., NIHR Research Professorship [grant number RP-2014-05-001]. The authors wish to thank Martin Green, Scientific Fellow at Waters Corporation, for his insights and critical reading of this manuscript.

■ REFERENCES

(1) Gika, H.; Virgiliou, C.; Theodoridis, G.; Plumb, R. S.; Wilson, I. D. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2019**, *1117*, 136−147.
(2) Cajka, T.; Fiehn, O. *TrAC, Trends Anal. Chem.* **2014**, *61*, 192−206.
(3) Roux, A.; Lison, D.; Junot, C.; Heilier, J. F. *Clin. Biochem.* **2011**, *44* (1), 119−35.
(4) Kell, D. B.; Oliver, S. G. *Metabolomics* **2016**, *12* (9), 148.
(5) Lu, W.; Bennett, B. D.; Rabinowitz, J. D. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2008**, *871* (2), 236−42.
(6) Cajka, T.; Fiehn, O. *Anal. Chem.* **2016**, *88* (1), 524−45.
(7) Mahieu, N. G.; Patti, G. J. *Anal. Chem.* **2017**, *89* (19), 10397−10406.
(8) Sindelar, M.; Patti, G. J. *J. Am. Chem. Soc.* **2020**, *142* (20), 9097−9105.
(9) Clarke, C. J.; Haselden, J. N. *Toxicol. Pathol.* **2008**, *36* (1), 140−7.
(10) Burton, L.; Ivosev, G.; Tate, S.; Impey, G.; Wingate, J.; Bonner, R. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2008**, *871* (2), 227−35.
(11) Engskog, M. K. R.; Haglöf, J.; Arvidsson, T.; Pettersson, C. *Metabolomics* **2016**, *12* (7), 114.
(12) Wishart, D. S. *Nat. Rev. Drug Discovery* **2016**, *15* (7), 473−84.
(13) Xia, J.; Broadhurst, D. I.; Wilson, M.; Wishart, D. S. *Metabolomics* **2013**, *9* (2), 280−299.
(14) Want, E. *Bioanalysis* **2009**, *1* (4), 805−19.
(15) Chadeau-Hyam, M.; Ebbels, T. M.; Brown, I. J.; Chan, Q.; Stamler, J.; Huang, C. C.; Daviglus, M. L.; Ueshima, H.; Zhao, L.; Holmes, E.; Nicholson, J. K.; Elliott, P.; De Iorio, M. *J. Proteome Res.* **2010**, *9* (9), 4620−7.
(16) Chong, E. Y.; Huang, Y.; Wu, H.; Ghasemzadeh, N.; Uppal, K.; Quyyumi, A. A.; Jones, D. P.; Yu, T. *Sci. Rep.* **2015**, *5*, 17221.
(17) Broadhurst, D. I.; Kell, D. B. *Metabolomics* **2007**, *2* (4), 171−196.
(18) Næs, T.; Mevik, B.-H. *J. Chemom.* **2001**, *15* (4), 413−426.
(19) Tolosi, L.; Lengauer, T. *Bioinformatics* **2011**, *27* (14), 1986−94.
(20) Hastie, T.; Tibshirani, R.; Friedman, J. H. *The elements of statistical learning: data mining, inference, and prediction*, 2nd ed.; Springer: New York, NY, 2009; p xxii, 745 p.
(21) Cech, N. B.; Enke, C. G. *Mass Spectrom. Rev.* **2001**, *20* (6), 362−87.
(22) Chamberlain, C. A.; Rubio, V. Y.; Garrett, T. J. *Metabolomics* **2019**, *15* (10), 135.
(23) Raji, M. A.; Schug, K. A. *Int. J. Mass Spectrom.* **2009**, *279* (2), 100−106.
(24) Cajka, T.; Smilowitz, J. T.; Fiehn, O. *Anal. Chem.* **2017**, *89* (22), 12360−12368.
(25) Wu, Z. E.; Kruger, M. C.; Cooper, G. J. S.; Poppitt, S. D.; Fraser, K. *Metabolites* **2019**, *9* (7), 124.
(26) Brodsky, L.; Moussaieff, A.; Shahaf, N.; Aharoni, A.; Rogachev, I. *Anal. Chem.* **2010**, *82* (22), 9177−87.
(27) Myers, O. D.; Sumner, S. J.; Li, S.; Barnes, S.; Du, X. *Anal. Chem.* **2017**, *89* (17), 8689−8695.
(28) Coble, J. B.; Fraga, C. G. *J. Chromatogr A* **2014**, *1358*, 155−64.
(29) Gurdeniz, G.; Kristensen, M.; Skov, T.; Dragsted, L. O. *Metabolites* **2012**, *2* (1), 77−99.
(30) Castillo, S.; Gopalacharyulu, P.; Yetukuri, L.; Orešič, M. *Chemom. Intell. Lab. Syst.* **2011**, *108* (1), 23−32.
(31) Eliasson, M.; Rannar, S.; Madsen, R.; Donten, M. A.; Marsden-Edwards, E.; Moritz, T.; Shockcor, J. P.; Johansson, E.; Trygg, J. *Anal. Chem.* **2012**, *84* (15), 6869−76.

https://dx.doi.org/10.1021/acs.analchem.0c03848
Anal. Chem. XXXX, XXX, XXX−XXX

(32) Gromski, P. S.; Muhamadali, H.; Ellis, D. I.; Xu, Y.; Correa, E.; Turner, M. L.; Goodacre, R. *Anal. Chim. Acta* **2015**, *879*, 10−23.
(33) Trygg, J.; Holmes, E.; Lundstedt, T. *J. Proteome Res.* **2007**, *6* (2), 469−79.
(34) Domingo-Almenara, X.; Montenegro-Burke, J. R.; Benton, H. P.; Siuzdak, G. *Anal. Chem.* **2018**, *90* (1), 480−489.
(35) Brown, M.; Wedge, D. C.; Goodacre, R.; Kell, D. B.; Baker, P. N.; Kenny, L. C.; Mamas, M. A.; Neyses, L.; Dunn, W. B. *Bioinformatics* **2011**, *27* (8), 1108−12.
(36) DeFelice, B. C.; Mehta, S. S.; Samra, S.; Cajka, T.; Wancewicz, B.; Fahrmann, J. F.; Fiehn, O. *Anal. Chem.* **2017**, *89* (6), 3250−3255.
(37) Fernandez-Albert, F.; Llorach, R.; Andres-Lacueva, C.; Perera-Lluna, A. *Anal. Chem.* **2014**, *86* (5), 2320−5.
(38) Broeckling, C. D.; Afsar, F. A.; Neumann, S.; Ben-Hur, A.; Prenni, J. E. *Anal. Chem.* **2014**, *86* (14), 6812−7.
(39) Mahieu, N. G.; Spalding, J. L.; Gelman, S. J.; Patti, G. J. *Anal. Chem.* **2016**, *88* (18), 9037−46.
(40) Izzi-Engbeaya, C.; Comninos, A. N.; Clarke, S. A.; Jomard, A.; Yang, L.; Jones, S.; Abbara, A.; Narayanaswamy, S.; Eng, P. C.; Papadopoulou, D.; Prague, J. K.; Bech, P.; Godsland, I. F.; Bassett, P.; Sands, C.; Camuzeaux, S.; Gomez-Romero, M.; Pearce, J. T. M.; Lewis, M. R.; Holmes, E.; Nicholson, J. K.; Tan, T.; Ratnasabapathy, R.; Hu, M.; Carrat, G.; Piemonti, L.; Bugliani, M.; Marchetti, P.; Johnson, P. R.; Hughes, S. J.; James Shapiro, A. M.; Rutter, G. A.; Dhillo, W. S. *Diabetes, Obes. Metab.* **2018**, *20* (12), 2800−2810.
(41) Yu, H.; Xing, S.; Nierves, L.; Lange, P. F.; Huan, T. *Anal. Chem.* **2020**, *92* (10), 7011−7019.
(42) Lommen, A. *Anal. Chem.* **2009**, *81* (8), 3079−86.
(43) Tikunov, Y. M.; Laptenok, S.; Hall, R. D.; Bovy, A.; de Vos, R. C. *Metabolomics* **2012**, *8* (4), 714−718.
(44) Fraisier-Vannier, O.; Chervin, J.; Cabanac, G.; Puech, V.; Fournier, S.; Durand, V.; Amiel, A.; Andre, O.; Benamar, O. A.; Dumas, B.; Tsugawa, H.; Marti, G. *Anal. Chem.* **2020**, *92* (14), 9971−9981.
(45) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J., III *SIAM Journal on Scientific and Statistical Computing* **1984**, *5* (3), 735−743.
(46) Breiman, L. *Machine Learning* **2001**, *45* (1), 5−32.
(47) Genuer, R.; Poggi, J.-M.; Tuleau-Malot, C. *Pattern Recognition Letters* **2010**, *31* (14), 2225−2236.
(48) Tibshirani, R. *Journal of the Royal Statistical Society. Series B (Methodological)* **1996**, *58* (1), 267−288.
(49) van Erp, S.; Oberski, D. L.; Mulder, J. *Journal of Mathematical Psychology* **2019**, *89*, 31−50.
(50) Wong, G.; Chan, J.; Kingwell, B. A.; Leckie, C.; Meikle, P. J. *Bioinformatics* **2014**, *30* (19), 2832−3.
(51) U.S. FDA *Bioanalytical Method Validation Guidance for Industry*; U.S. FDA: Rockville, MD, 2018.
(52) European Medicines Agency *Guideline on Bioanalytical Method Validation*; 2011; https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-bioanalytical-method-validation_en.pdf.
(53) Naz, S.; Vallejo, M.; Garcia, A.; Barbas, C. *J. Chromatogr A* **2014**, *1353*, 99−105.
(54) Broadhurst, D.; Goodacre, R.; Reinke, S. N.; Kuligowski, J.; Wilson, I. D.; Lewis, M. R.; Dunn, W. B. *Metabolomics* **2018**, *14* (6), 72.
(55) Gika, H. G.; Theodoridis, G. A.; Wingate, J. E.; Wilson, I. D. *J. Proteome Res.* **2007**, *6* (8), 3291−303.
(56) Bijlsma, S.; Bobeldijk, I.; Verheij, E. R.; Ramaker, R.; Kochhar, S.; Macdonald, I. A.; van Ommen, B.; Smilde, A. K. *Anal. Chem.* **2006**, *78* (2), 567−74.
(57) Sangster, T.; Major, H.; Plumb, R.; Wilson, A. J.; Wilson, I. D. *Analyst* **2006**, *131* (10), 1075−8.
(58) Croixmarie, V.; Umbdenstock, T.; Cloarec, O.; Moreau, A.; Pascussi, J. M.; Boursier-Neyret, C.; Walther, B. *Anal. Chem.* **2009**, *81* (15), 6061−9.
(59) 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. *Anal. Chem.* **2016**, *88* (18), 9004−13.
(60) Kuehnbaum, N. L.; Kormendi, A.; Britz-McKibbin, P. *Anal. Chem.* **2013**, *85* (22), 10664−9.
(61) Vorkas, P. A.; Isaac, G.; Anwar, M. A.; Davies, A. H.; Want, E. J.; Nicholson, J. K.; Holmes, E. *Anal. Chem.* **2015**, *87* (8), 4184−93.
(62) Duan, L.; Molnar, I.; Snyder, J. H.; Shen, G. A.; Qi, X. *Mol. Plant* **2016**, *9* (8), 1217−1220.
(63) Jankevics, A.; Merlo, M. E.; de Vries, M.; Vonk, R. J.; Takano, E.; Breitling, R. *Metabolomics* **2012**, *8* (S1), 29−36.
(64) Saigusa, D.; Okamura, Y.; Motoike, I. N.; Katoh, Y.; Kurosawa, Y.; Saijyo, R.; Koshiba, S.; Yasuda, J.; Motohashi, H.; Sugawara, J.; Tanabe, O.; Kinoshita, K.; Yamamoto, M. *PLoS One* **2016**, *11* (8), e0160555.
(65) Kouassi Nzoughet, J.; Bocca, C.; Simard, G.; Prunier-Mirebeau, D.; Chao de la Barca, J. M.; Bonneau, D.; Procaccio, V.; Prunier, F.; Lenaers, G.; Reynier, P. *Anal. Chem.* **2017**, *89* (3), 2138−2146.
(66) Sands, C. J.; Wolfer, A. M.; Correia, G. D. S.; Sadawi, N.; Ahmed, A.; Jimenez, B.; Lewis, M. R.; Glen, R. C.; Nicholson, J. K.; Pearce, J. T. M. *Bioinformatics* **2019**, *35* (24), 5359−5360.
(67) Guitton, Y.; Tremblay-Franco, M.; Le Corguille, G.; Martin, J. F.; Petera, M.; Roger-Mele, P.; Delabriere, A.; Goulitquer, S.; Monsoor, M.; Duperier, C.; Canlet, C.; Servien, R.; Tardivel, P.; Caron, C.; Giacomoni, F.; Thevenot, E. A. *Int. J. Biochem. Cell Biol.* **2017**, *93*, 89−101.
(68) Popa-Burke, I.; Lupotsky, B.; Boyer, J.; Gannon, W.; Hughes, R.; Kadwill, P.; Lyerly, D.; Nichols, J.; Nixon, E.; Rimmer, D.; Saiz-Nicolas, I.; Sanfiz-Pinto, B.; Holland, S. *J. Biomol. Screening* **2009**, *14* (8), 1017−30.
(69) Veselkov, K. A.; Vingara, L. K.; Masson, P.; Robinette, S. L.; Want, E.; Li, J. V.; Barton, R. H.; Boursier-Neyret, C.; Walther, B.; Ebbels, T. M.; Pelczer, I.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Anal. Chem.* **2011**, *83* (15), 5864−72.
(70) Gagnebin, Y.; Tonoli, D.; Lescuyer, P.; Ponte, B.; de Seigneux, S.; Martin, P. Y.; Schappler, J.; Boccard, J.; Rudaz, S. *Anal. Chim. Acta* **2017**, *955*, 27−35.
(71) Lee, J. W.; Devanarayan, V.; Barrett, Y. C.; Weiner, R.; Allinson, J.; Fountain, S.; Keller, S.; Weinryb, I.; Green, M.; Duan, L.; Rogers, J. A.; Millham, R.; O'Brien, P. J.; Sailstad, J.; Khan, M.; Ray, C.; Wagner, J. A. *Pharm. Res.* **2006**, *23* (2), 312−28.
(72) Stefansson, M.; Sjöberg, P. J. R.; Markides, K. E. *Anal. Chem.* **1996**, *68* (10), 1792−1797.
(73) Oliveros, J. C. *Venny: An interactive tool for comparing lists with Venn's diagrams (2007−2015)*; https://bioinfogp.cnb.csic.es/tools/venny/index.html (accessed Dec 7, 2020).
(74) Benson, E. S.; Hendricks, E. J.; Freier, E. F. *Postgrad. Med.* **1966**, *40* (6), A-70−A-82.
(75) Whitehead, T. P.; Morris, L. O. *Ann. Clin. Biochem.* **1969**, *6* (4), 94−103.
(76) Redestig, H.; Kobayashi, M.; Saito, K.; Kusano, M. *Anal. Chem.* **2011**, *83* (14), 5645−51.
(77) Panuwet, P.; Hunter, R. E., Jr.; D'Souza, P. E.; Chen, X.; Radford, S. A.; Cohen, J. R.; Marder, M. E.; Kartavenka, K.; Ryan, P. B.; Barr, D. B. *Crit. Rev. Anal. Chem.* **2016**, *46* (2), 93−105.
(78) Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R. *Nat. Protoc.* **2011**, *6* (7), 1060−1083.
(79) Liu, H.; Lam, L.; Dasgupta, P. K. *Talanta* **2011**, *87*, 307−10.
(80) Miller, A. M.; Goggin, M. M.; Nguyen, A.; Gozum, S. D.; Janis, G. C. *J. Anal. Toxicol.* **2017**, *41* (5), 355−359.
(81) Trobbiani, S.; Stockham, P.; Scott, T. *Drug Test. Anal.* **2017**, *9* (10), 1630−1636.

https://dx.doi.org/10.1021/acs.analchem.0c03848
Anal. Chem. XXXX, XXX, XXX−XXX