analytical chemistry

Review

pubs.acs.org/ac

Bioinformatics: The Next Frontier of Metabolomics

Caroline H. Johnson, Julijana Ivanisevic, H. Paul Benton, and Gary Siuzdak*

Scripps Center for Metabolomics and Mass Spectrometry, The Scripps Research Institute, La Jolla, California 92037, United States

CONTENTS

Bioinformatic tools are required to carry out essential functions such as statistical analyses and database functionalities. Now, they are also needed for one of the most difficult tasks, helping researchers decide which metabolites are the most biologically meaningful. This can be achieved through aiding the identification process, reducing feature redundancy, putting forward better candidates for tandem mass spectrometry (MS/MS), speeding up or automating the workflow, deconvolving the feature list through meta-analysis or multigroup analysis, or using stable isotopes and pathway mapping. This review thus focuses on the most recent and innovative bioinformatic advancements for identifying metabolites.

A primary objective of metabolomics beyond biomarker discovery is to identify the most meaningful metabolites that correlate with disease pathogenesis or other perturbations of metabolism. Metabolites play important roles in biological pathways; their flux or differential regulation (dysregulation) can reveal novel insights into disease and environmental influences. Therefore, one of the most important goals of metabolomic analysis has been to assign metabolite identity so they can be used for further statistical and informed pathway analysis.[1,2] Over the past few years, technologies for analyzing metabolites by untargeted or targeted metabolomics have undergone extensive improvements. Strides to establish the most efficient protocols for experimental design, sample extraction techniques, and data acquisition have paid off providing robust complex data sets.[3−9] As more is being required of these data sets such as assigning identity and biological meaning to the features, bioinformatics is the area of metabolomics which is currently undergoing the most needed growth.

It is often the case that metabolomic analysis results in a list of metabolites with low specificity for the disease or stimulus being studied (Figure 1). Some of these metabolites seem to be

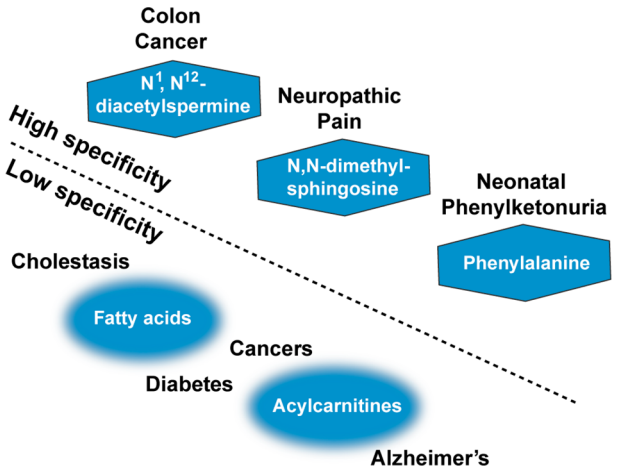

Figure 1. Biomarkers that have high vs low disease specificity.

dysregulated in a variety of diseases such as acylcarnitines[10−13] and fatty acids.[14−17] They may be more indicative of a perturbed systemic cause (appetite, physical activity, diurnal rhythm changes, etc..), sample contamination, or instrumental/bioinformatic noise, rather than a specific biomarker of disease. An example of this can be seen in the analysis of urinary biomarkers of ionizing radiation, where dicarboxylic acids were downregulated in the rat after radiation exposure. It was proven that this observation was actually caused by a decreased appetite after radiation exposure perturbing the β-oxidation pathway and not from radiation-induced cellular changes.[18,19] Furthermore, dicarboxylic acids can leach out from plastics during the extraction process, further adding to the ambiguity of their role in ionizing radiation.[20]

As well as identifying the correct source of the biomarkers, it is also important to identify their physiological role and how to utilize them as therapeutic targets. This first has to start with the identification of the metabolite and is determined by filtering thresholds set by the user which is intrinsically biased. These thresholds include those for fold change and p-value, which are highly dependent on the experiment; in vitro

Special Issue: Fundamental and Applied Reviews in Analytical Chemistry 2015

   dx.doi.org/10.1021/ac5040693 | *Anal. Chem.* XXXX, XXX, XXX−XXX

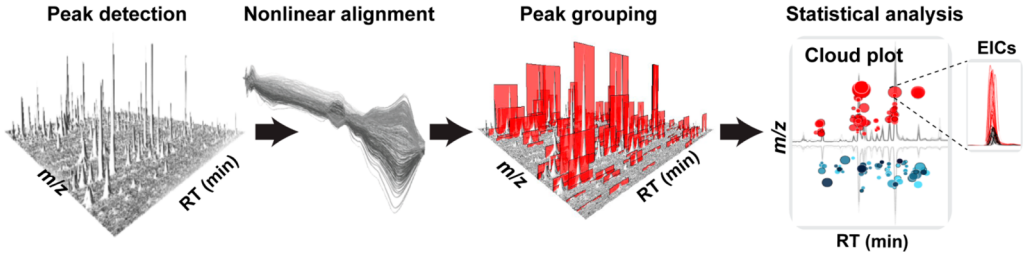

Figure 2. Nonlinear retention time alignment by XCMS allows for untargeted metabolomic analysis.

experiments would exhibit lower variation between biological replicates than in vivo. The ease of identifying the metabolite is also determined by its concentration in the sample and previous annotation in metabolite databases. Filtering thresholds for metabolite intensity that are set too high may omit important biologically meaningful metabolites rather than noise. Furthermore, a metabolite that is novel or not curated in a database may not be taken into consideration based on the chemical knowledge of the researcher and what they deem as meaningful.

In order to transform the complex list of identified metabolites into markers of disease, or assign what role they play, bioinformatic tools can aid in identifying the potential pathways that the metabolite may belong to. It is then that the researcher can use this knowledge surrounding the biology of the metabolite to probe the mechanism of the disease. Untargeted metabolomics has already been used in such a manner to find the source of neuropathic pain.[21] *N,N*-Dimethylsphingosine was dysregulated in a rat model of neuropathic pain, furthermore when dosed to control rats it induced mechanical hypersensitivity. This metabolite implicated the sphingomyelin-ceramide pathway as a potential therapeutic target. Antimetabolite inhibitors of enzymes in this pathway were tested and were able to ameliorate neuropathic pain (unpublished data). This study holds promise for other metabolomic studies to maximize the potential information contained within the data for finding therapeutics of disease rather than only providing lists of dysregulated metabolites.

STREAMLINING DATA ACQUISITION: THE FIRST STEP

Data Alignment. One of the most important developments for untargeted liquid chromatography/mass spectrometry (LC/MS)-based metabolomics was nonlinear retention time (RT) alignment with XCMS using endogenous metabolites.[22] This realignment for untargeted analysis is important to match peaks representing the same analytes from different samples for comparative analysis; these peaks naturally drift between sample runs due to sample build up on the columns and physical changes to the column from the mobile phase, both which change the nature of the sample-stationary phase interaction. An often used technique involved spiking internal standards into samples prior to acquisition, but this was based on the assumption that RT deviations were linear.[23] Thus, XCMS was particularly poignant as there were no options for carrying out nonlinear realignment and untargeted LC/MS-based untargeted metabolomics (Figure 2). Since XCMS there have been a number of notable alignment algorithms developed including MZmine2.[24,25] New developments in columns and LC systems have improved RT drift considerably producing more reliable data.

Automated Metabolomics. The feature identification process is the most time-consuming and complex part of the metabolomic workflow. After annotation of peaks and statistical analyses, MS/MS data need to be acquired for the feature of interest. This data is then compared to metabolite databases and commercial standards for definitive identification and validation, a process that can take weeks to carry out but can be potentially shortened through integration of metabolite profiling and identification into a single autonomous mass spectrometry method.

During the typical metabolomic workflow, MS^1 data is acquired for each of the samples and a feature list table is produced displaying the results of the statistical analysis. The investigator will search through the table and pick out the features of interest that need to be identified, then manually search metabolite databases for putative identification. To confirm the identification, further mass spectrometry setup and run time is needed for the subsequent MS/MS analysis. XCMS Online already speeds up this process through its integration to METLIN,[26] the world's largest metabolite database, which enables each feature, when possible, to have putative metabolite identification based on its mass-to-charge ratio (m/z) match. In addition METLIN has automated MS/MS matching to help confirm identity. RAMClust also has a semiautomated workflow where indiscriminant MS/MS (idMS/MS) data is used for automatic database searching.[27] Both the XCMS Online and RAMClust programs aid in the identification process but still require informed manual interpretation.

A simple but effective way to shorten this acquisition-to-identification process from weeks to hours is to simultaneously acquire MS^1 and MS/MS data over the duration of a LC/MS run, using an autonomous untargeted metabolomic workflow[26,28] (Figure 3). During this workflow MS^1 data are preprocessed by XCMS; features are extracted, realigned for RT correction, and undergo statistical analysis. The MS/MS data are acquired automatically using data dependent acquisition (DDA); the MS^1 data are scanned for precursor ions selected by predefined parameters. Automatic spectral matching of the MS/MS data to databases containing MS/MS spectra, such as METLIN, Human Metabolome Database (HMDB),[29] and MassBank[30] aid in putative identification. This autonomous approach has been optimized to achieve the correct balance of acquired MS^1 and MS/MS data. A high scan speed can allow for greater MS/MS spectral coverage, but a scan speed that is too high can affect the quality of the spectra due to the lack of data points for each extracted ion chromatogram (EIC). In both cases adequate time is required to obtain high-quality spectra, which is mitigated by using quadrupole time-of-flights (QTOFs) with fast scan rates and high sensitivity. The autonomous workflow has many benefits for untargeted metabolomics; it saves weeks of mass spectrometry time as well as inspection time for manually picking

dx.doi.org/10.1021/ac5040693 | *Anal. Chem.* XXXX, XXX, XXX−XXX

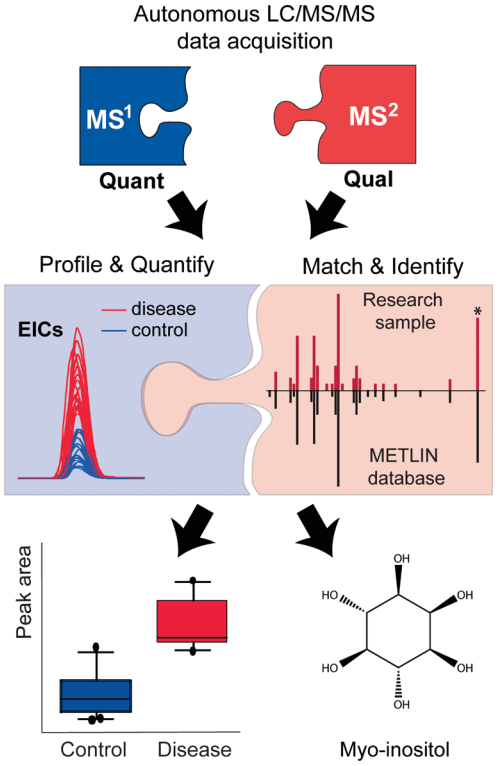

Figure 3. Overview of autonomous metabolomics. MS^1 and MS/MS spectra acquisition, relative comparative analysis, and metabolite identification are carried out simultaneously.

out significant peaks. It also saves sample as repeat injections are not needed. However, the downfall is its reliance on metabolite databases which are not yet fully comprehensive. METLIN has the largest number of MS/MS spectra (>12 000 metabolites with high-resolution MS/MS data) and is growing, which will improve the likelihood of successful matches with this workflow.

XCMS has been utilized in another semiautomated acquisition and processing workflow along with RT prediction to aid in metabolite identification.[31] This workflow performs feature detection, RT correction, gap filling, feature annotation, in silico fragmentation, and spectral matching to databases.[31] A *nearline* (nearly online) DDA and MS/MS processing step using MetShot (an R package) is also incorporated; MS/MS experiments are automatically generated from a ranked list of interesting precursor features within the same analysis, it uses defined filters which results in the acquisition of only relevant spectra.[32] The filters include sorting and prioritizing features by *p*-value or fold change, selecting features related to quasi-molecular ions, and the removal of ions that would be too low for MS/MS analysis. It also separates features from coeluting and cofragmenting compounds and places them into separate MS/MS files. To further aid in metabolite identification, the spectra can be compared to mass spectral databases.

In silico prediction tools for fragmentation such as MetFrag can further aid putative identification to help overcome the incompleteness of the databases.[33] More recently MetFusion has been developed which combines MassBank, METLIN, or the Golm Metabolome Database[34] libraries with MetFrag to improve the rank of the correct candidate.[35] For each candidate metabolite returned by MetFrag, RT prediction further limits the number of potential candidates based on physiochemical properties (lipophilicity) that may aid in metabolite identification.[36−39] Automated workflows that incorporate data upload, processing, identification, and pathway analyses will thus markedly improve the efficiency of the metabolomic workflow.

Data Streaming For Cloud-Based Metabolomics. To overcome the challenges involved in uploading metabolomic data files to servers, a streaming approach was developed.[40] Data upload can sometimes take more than 1 day due to limitations set by the speed of data transfer over the Internet. The software developed allows the acquired data to be uploaded from the instrument computer workstation directly to XCMS Online where they are converted and processed. This concept of data streaming reduces the mean wait time for complete data processing from 20 h to fewer than 3 h (Figure 4).

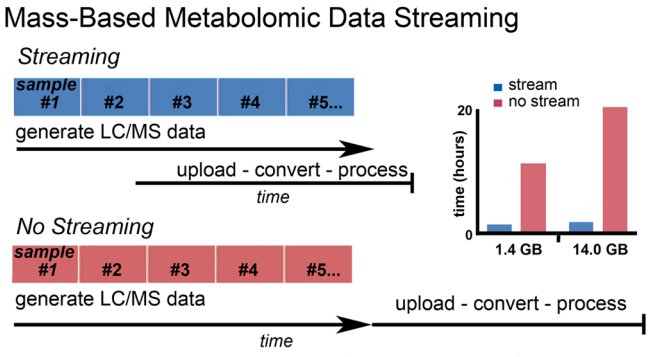

Figure 4. XCMS-based data streaming workflow (top left) allows data upload and processing after each LC/MS run is performed, dramatically reducing the processing time after the data are acquired for the final sample (top right). A thousand XCMS Online data sets were examined for their average processing time without streaming. For low-resolution data (~1.4 GB) and high-resolution data (~14.0 GB) over 10 and 20 h was required after the final LC/MS analysis was performed, respectively. Streaming allowed a 7-fold decrease in average processing time after data acquisition. Reprinted from ref 40. Copyright 2014 Nature America, Inc.

The upload speed is dependent on data size and the Internet connection, but there is a marked improvement in the efficiency of the untargeted metabolomic workflow, as the data upload occurs parallel to the data acquisition. In addition, simultaneous MS/MS data are only acquired for the features of interest based on statistical thresholds and matches to metabolite databases from the processed files already uploaded and analyzed by XCMS Online.

FEATURE ANALYSIS

Mass Spectral Annotations. In order to accurately identify metabolites from LC/MS data, it is first essential to elucidate which features are isotopologue ions, adduct ions ($[M + Na]^+$, $[M + Cl]^-$), multiply charged ions or fragment ions ($[M + H - H_2O]^+$). These all exist from the ionization process.[41,42] Through correct annotation, the complexity of the data sets can be reduced and features of biological interest identified. LC/MS data generated from untargeted metabolomics experiments are typically processed using freely available bioinformatics software such as XCMS,[22,43] OpenMS,[44] apLCMS,[45] xMSanalyzer,[46] mzMatch,[47] or MZmine.[24] These platforms create, in different manners, grouped feature lists across multiple samples and classes of samples in which only a percentage of the ions are unique; a feature is defined as a two-dimensional bounded signal, a chromatographic peak (RT), and a mass spectral peak (m/z).[48]

dx.doi.org/10.1021/ac5040693 | *Anal. Chem.* XXXX, XXX, XXX−XXX

Annotation software recently introduced, use clustering principles to deconvolve the data, where isotopic or adduct ions will coelute and improve confidence in assigning identity. The software uses within sample, sample-to-sample, or Bayesian "*priors*" correlation to find clusters. A Bioconductor package in R, called Collection of Algorithms for Metabolite pRofile Annotation (CAMERA) was recently designed to be used in conjunction with XCMS.[48] It facilitates the annotation of isotopic peaks, adducts, and fragment ions in peak lists using RT-based grouping, peak shape integration/analysis, and intensity correlation. CAMERA can reduce the number of redundant features in the processed LC/MS data by approximately 50%, decreasing the number of false identifications and unnecessary MS/MS experiments that would potentially be carried out.[49,50] One of the disadvantages of CAMERA is that it is biased toward the most abundant features,[27] but even so, very low abundant peaks would anyway have a low likelihood of being identified due to the concentration constraints of MS/MS analysis. Another spectral matching-based annotation software, RAMclust,[27] works on the basis that two features resulting from the same metabolite will have similar RTs and abundances across different samples within a sample set. A similarity function allows the generation of spectra through grouping features from a single metabolite into a single cluster. Another aspect of RAMclust is the feature finding parameter, which carries idMS/MS, enabling manual interpretation and automatic database searching of feature clusters.

Daly et al.[51] use a Bayesian "*priors*" approach called MetAssign. They point out that the level of confidence in annotations varies across metabolites and data sets. To improve confidence in peak annotation, probabilistic estimates of the presence/absence of metabolites are provided based on integration of information from multiple peaks. MetAssign is also based on the principle that several peaks of an isotopic series at the same RT should provide a higher confidence in a putative identification than a single noisy peak. When validated on an experimental data set, this software performed better than CAMERA and mzMatch[47] (another annotation software) for peak annotation. It also provided a measure of confidence for its putative annotations. Fernandez-Albert et al.[52] showed that peak aggregation (clustering) improved the statistical power of LC/MS data when using analysis of variance (ANOVA) to select features and multivariate methods such as partial least-squares discriminant analysis (PLS-DA)[53] and support vector machines (SVM).[54] They used four peak aggregation methods to take advantage of and solve high variable collinearity. Two of the clustering methods, "Non-Negative Matrix Factorization (NMF) Reduction" and "Principal Component Analysis (PCA) Decomposition" significantly improved the detection of significant features. The recent surge of papers for improving peak annotation shows the importance of reducing redundant features and integrating a confidence level for the annotations may aid in more efficient identification.

Credentialing. Even though these tools can aid in annotating multiple features for a single metabolite and thus reduce the redundancy of many of the metabolite features, a new tool introduced by Mahieu et al.[55] aids in identifying features, which are of biological origin only. The idea is based on the principle that untargeted metabolomic analysis results in thousands of biological and artifactual features. Omitting these artifacts that can arise from contaminants (from sample extraction or carryover from previous experiments), chemical/background noise detected by the MS, and bioinformatic noise (misannotation during processing) will allow those more important biological features to be assessed. This process of distinguishing features of biological origin from artifactual ones is called "credentialing". To assess the efficacy of the credentialing algorithm, *Escherichia coli* (*E. coli*) was grown in media containing natural-abundance ^{12}C glucose or media containing U−^{13}C glucose. The cultures were mixed together in defined ratios and analyzed by untargeted LC/MS metabolomics. The credentialing algorithm identified and credentialed features based on isotope-intensity ratios; intensities from coeluting isotopologue pairs compared to the values expected in the culture volume ratios. Signals of biological origin could thus be distinguished from artifactual ones. This tool allowed the authors to optimize the bioinformatic analysis, reducing overall noise features by 15% and increasing biological feature detection by 20%. Another advantage to this tool, like the others aforementioned that annotate unwanted feature peaks, is that the list of features for MS/MS is dramatically reduced; when credentialing was applied to the *E. coli* data set it reduced the number of candidates from 23 567 to 2 912. Even if all these metabolites cannot be correctly identified, knowing that the ones targeted for analysis are of biological origin effectively improves the metabolomic workflow, and moves toward finding those that are meaningful. Similarly, others have used stable isotopes for peak annotation but do not provide enough specificity to remove all spurious peaks.[56−59] Unlike these methods, the ^{13}C and ^{12}C samples are run together to reduce RT variation, and the absolute mass differences of U−^{13}C and U−^{12}C metabolites are filtered rather than using predicted molecular formulas. Therefore, the credentialing approach limits the amount of noise and enhances the annotation of biologically relevant peaks, meanwhile the other workflows are better for improving formula annotation which would be useful for identification and have a lower false discovery rate.

Calculating Mass Measurement Errors. Metabolite identification can also be problematic in high throughput or large-scale LC/MS runs. During these long run times the mass accuracy suffers and the number of incorrectly assigned or redundant peaks dramatically increases. The mass accuracy is crucial for matching experimental accurate masses to those found in databases, an increase of 10 ppm (ppm) in the mass accuracy window results in a 10-fold increase in database hits.[60] The major factor in maintaining a high accuracy window of less than 5 ppm is the intensity of the ion signal.[61−64] This can be demonstrated when measuring the mass error of the lock mass signal; its two isotopic peaks which are at lower concentrations often have a larger ppm than the parent ion. Conversely when a sample is too concentrated and peaks are saturated, the mass accuracy can also suffer. When the mass accuracy shifts the mass error window needs to be widened to perhaps 10−22 ppm, which greatly increases the false positive rate. Methods to correct the mass accuracy while data is being acquired include using a reference mix of known ions which the mass spectrometer uses to calibrate during the run. Ions that occur naturally in most of the experimental runs can also be used. There are also prediction models that have been designed to estimate the mass measurement errors. These models do not change the acquired data but aid in reducing the false positive rate. One such model by Shahaf et al.[60] uses XCMS and CAMERA to process the data and annotate peaks. Mass measurement errors are estimated using an annotated library of reference metabolites which are obtained from multiple runs on

dx.doi.org/10.1021/ac5040693 | *Anal. Chem.* XXXX, XXX, XXX−XXX

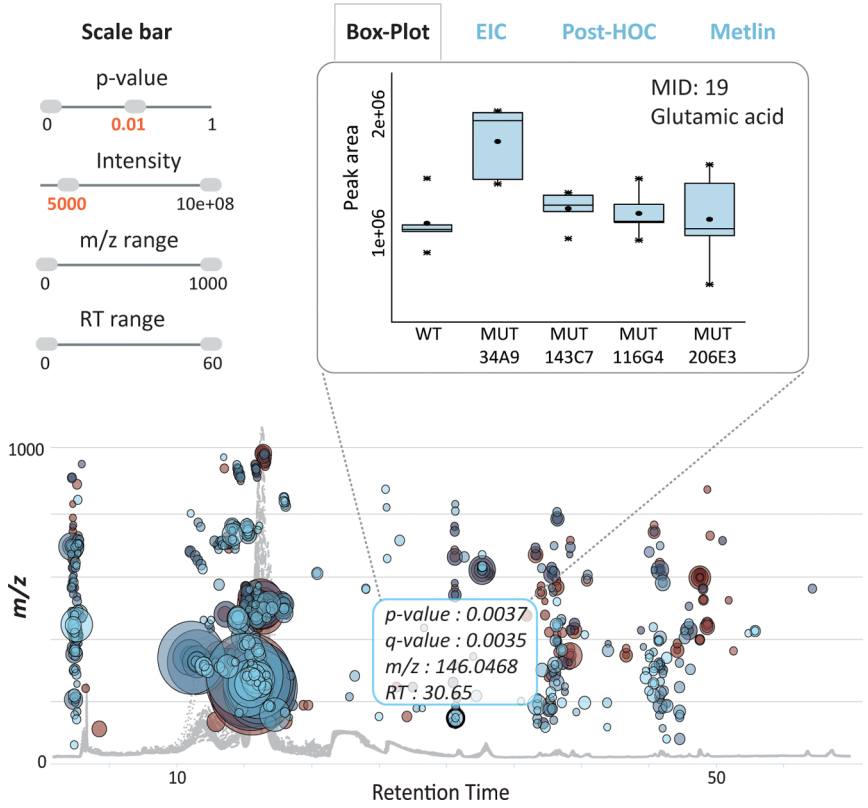

Figure 5. Interactive multigroup cloud plot. Metabolite features whose level varies significantly ($p < 0.01$) across wild-type and mutant bacteria are projected on the cloud plot depending on their RT (x-axis) and m/z (y-axis). Each metabolite feature is represented by a bubble. Statistical significance (p-value) is represented by the bubble's color intensity. The size of the bubble denotes feature intensity. When the user scrolls the mouse over a bubble, feature assignments are displayed in a pop-up window (m/z, RT, p-value, fold change). When a bubble is selected by a "mouse click", the EIC, Box–Whisker plot, Posthoc, and METLIN hits appear on the main panel. Each bubble is linked to the METLIN database to provide putative identifications based on accurate m/z. The variation pattern of glutamic acid (m/z 146.0468, MS/MS METLIN match) across different mutants is shown by a box–whisker plot. Reprinted from ref 65. Copyright 2014 American Chemical Society.

the same MS instrument, containing information on peaks of related features such as isotopes and adducts. The data are grouped into bins that cover the available mass and intensity range and a prediction model applied to the data which takes into account the one-sided confidence levels of all the mass measurement errors. The model predicts the error for an ion peak's mass-intensity pair within a range and is specific for the instrument on which it was carried out, making it a reusable model for the next experiment. A reduction in the false positive rate of 21% on a small data set was seen and could be very effective for larger high-throughput studies given that the prediction models are optimized specifically for an instrument.

Statistical Analysis: Visualization and Deconvolution. Multivariate and univariate statistical analyses can further aid in finding meaningful metabolites from the hundreds of filtered features postannotation. However, choosing the correct test is challenging for those without a background in bioinformatics. Recently, journals have become more stringent in making sure that articles submitted for publication use the correct statistical tests. Indeed *Nature* requires a statistical checklist to ensure articles have statistical adequacy. Gowda et al.[65] highlight the appropriate use of different statistical tests, such as when to use parametric vs nonparametric tests and paired vs unpaired tests for metabolomic analysis. Paired tests can be especially useful in human studies where there is high interindividual variability, comparing differences between two measurements (e.g., metabolic response before and after drug treatment) for each subject across the sample set.

Developments in visualizing results and filtering features have facilitated characterization and structural identification in untargeted metabolomics. MetaboAnalyst[66] and XCMS Online both provide comprehensive statistical analysis tools which include univariate, multivariate, high-dimensional feature selection, clustering, and supervised classification analysis. Most recently the interactive cloud plot, interactive PCA, and interactive heat map were introduced to XCMS Online.[65,67] These interactive statistical read-outs let the user customize the display and choose the most valuable features. The cloud plot allows for instant visualization of each feature processed from the untargeted data.[67] The interactive version of this plot can be manipulated to remove isotopes, change thresholds and ion-intensity ranges, and zooms in on areas of feature overlap.[65] By clicking on a feature it is possible to instantly obtain information about the feature; p-value, q-value, m/z, RT, intensity, and peak group. One of its strengths is that coeluting features and the hydrophilicity/hydrophobicity can be easily observed. An example of an interactive cloud plot can be seen in Figure 5. PCA and heatmaps have been used widely in metabolomic research; however, the interactive versions of these tools allow for instant modification by user definitions of criteria and visualization of underlying information (Figure 6).

One of the most useful tools for finding meaningful metabolites is meta-analyses and multigroup comparisons. With these analyses it is possible to observe shared metabolic patterns across multiple experiments and metabolite variation patterns across multiple data groups. Meta-analysis can prioritize

dx.doi.org/10.1021/ac5040693 | *Anal. Chem.* XXXX, XXX, XXX–XXX

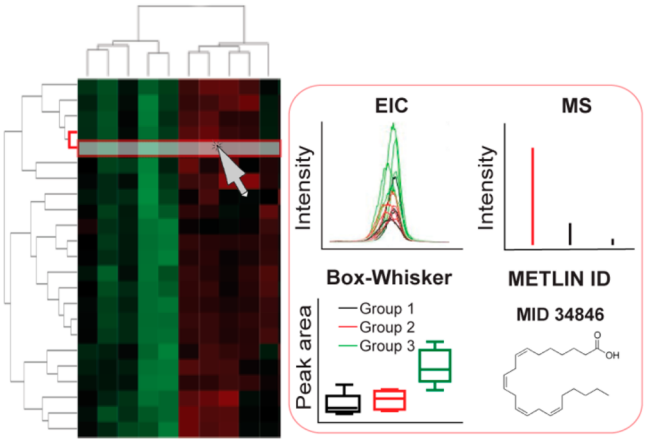

Figure 6. Interactive heatmap with metabolomic data visualization. Each row represents a metabolite feature, and each column represents a sample. The *Z*-scale of each feature is plotted on the red-green color scale. When a feature annotation tile (*m*/*z*, RT, or *p*-value) is selected, its Box–Whisker plot, EIC (extracted ion chromatogram), MS spectrum, and putative METLIN matches appear.

interesting features by integrating data from multiple studies and help identify shared homologous patterns of metabolic variation across the results of multiple different experiments. Multigroup analysis on the other hand is an extension of the two-group/pairwise analysis that allows for the comparison of multiple classes and identifies features whose variation pattern is statistically significant across them. This type of analysis is particular useful for a time-course experiment. An example of multigroup analysis can be seen in Figure 5 where the metabolic response to stress was analyzed across different types of bacteria.

Targeted Validation. While not an aspect of the bioinformatic solution per se, it is worth noting that all untargeted metabolomic analyses require further validation to remove false positives and provide an additional level of confidence in the follow up biological experiments. To accomplish this, typically quantitative targeted analysis (triple quadrupole mass spectrometry (QqQ-MS)) are performed using multiple reaction monitoring (MRM). False positives can occur during untargeted analysis therefore carrying out targeted QqQ-MS with standards for each metabolite and can provide assurance that an accurate fold change and *p*-value is being reported.

PATHWAY ANALYSIS: PUTTING METABOLOMIC DATA INTO A BIOLOGICAL CONTEXT

On the quest to find meaningful metabolites in metabolomic data, the ability to relate the identified metabolites to the biological question at hand is imperative. Pathway/network tools can aid in elucidating the roles the metabolites play in multiple pathways. However, it is often the case that metabolomic analysis can result in a number of metabolites that are not related to each other, i.e., they are not in the same pathways or a thin coverage is providing for one particular pathway; this can be due to some metabolites being in flux, while others may be at a steady concentration and would not be differentially regulated. Furthermore, some metabolites cannot be mapped to any pathways. Therefore, it is far from trivial to visualize metabolites with respect to their presence and interactions in biochemical networks.

There have been a number of recently developed programs which are designed to help with biological pathway and network analyses. As part of the streaming approach aforementioned, simultaneous MS/MS data are acquired for features of interest based on statistical thresholds and matches to metabolite databases from files already uploaded and analyzed by XCMS Online. When two or more putatively assigned metabolites are observed in the same pathway, the MS^1 data are mined for other metabolites in that pathway and targeted for MS/MS analysis, essentially aiding pathway and network analysis. Thus, streaming allows for biology-dependent data acquisition (BDDA) to take place. BDDA differs from DDA as MS/MS is not triggered on the basis of ion intensity.[68] The BDDA concept was validated in the analysis of tumor samples, where four metabolites were found belonging to the same pathway (Figure 7).

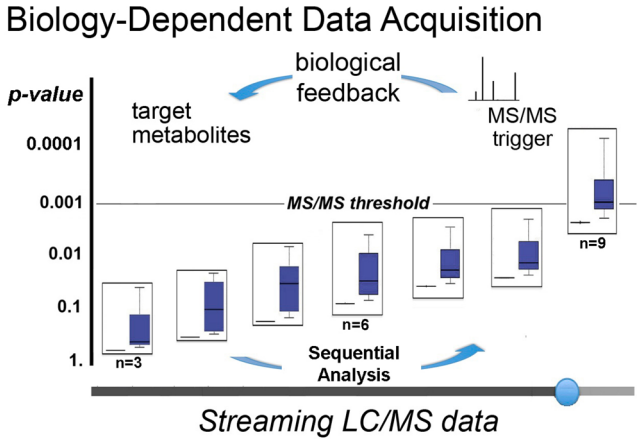

Figure 7. Biology-dependent data acquisition relies on statistics generated after each sample run for mass spectrometry data acquisition decision making. The representative example shows a decreasing *P* value for a feature of interest over the time-course of data streaming. When the *P* value for the features reaches 0.001, MS/MS is performed. A two-tailed Wilcoxon signed-rank test was used to calculate the statistical significance for $n = 28$. Box and whisker plots display the full range of variation (whiskers, median with minimum–maximum; boxes, interquartile range). Reprinted from ref 40. Copyright 2014 Nature America, Inc.

Another notable program is *mummichog*, which can aid in finding biological pathway activity.[69] Instead of identifying the metabolites before carrying out pathway/network analysis, *mummichog* predicts biological activity using the *m*/*z* values of both the statistically dysregulated and unchanged features. Genome-scale human metabolic networks are then mined for enriched pathways. These metabolic networks include KEGG,[70] Recon1,[71] and Edinburgh human metabolic network.[72] However, *mummichog* can incorrectly assign metabolites since one *m*/*z* could account for several different metabolites, but the program's design is to find networks rather than individual metabolites. Indeed one of the major benefits of this program is that one can bypass the initial laborious metabolite identification process and move directly onto investigating whole pathways that are disrupted and can be targeted in future studies. It also eliminates user bias, where features are traditionally picked for identification based on interest and biochemical relevance.[50] The successes of this technology will likely progress as metabolic models improve and as metabolomic data becomes integrated into genome-scale metabolic models.

Another approach aids visualization of metabolites in related pathways, through the creation of network "MetaMapp" graphs in Cytoscape.[73,74] These graphs integrate biochemical

dx.doi.org/10.1021/ac5040693 | *Anal. Chem.* XXXX, XXX, XXX–XXX

pathway and chemical relationships using KEGG reactant pair database,[70] Tanimoto chemical and NIST mass spectral similarity scores.[75] Differential expression of metabolite nodes are superimposed onto network graphs to help with the interpretation of their involvement in metabolic networks and facilitate biological interpretation. To overcome the issue of incomplete mapping and also metabolites that lack any putative identification or network mapping in current reaction databases, metabolites are associated together based on their chemical similarity as these compounds should be in theory metabolized to each other. However, this can result in a loss of overall biochemical clarity but still allows for visualization of their involvement in pathways, which would otherwise not be possible. The other advantages to this program are that it can be used on any type of acquired metabolomic data (MS or nuclear magnetic resonance spectroscopy) allowing for integration between multiple analytical platforms. Multiple species can be mapped at the same time as it is not constricted by genomic information, and the maps can be automatically updated as more additional information regarding the identification of the metabolites is gathered.

Another notable method for network analysis is Metabolite Set Enrichment Analysis (MSEA), part of the MetaboAnalyst program.[66,76,77] MSEA is based on gene set enrichment analysis (GSEA) and is used to investigate the enrichment of predefined groups of functionally related metabolites rather than individual metabolites. This program was developed through the creation of one encompassing metabolite library using HMDB, PubChem,[78] ChEBI,[79] KEGG, BiGG,[80] METLIN, BioCyc,[81] Reactome,[82] and Wikipedia, and importantly includes reference concentrations for many metabolites. The main limitation of MSEA is that it is biased to mammalian studies therefore metabolomic experiments using other species will require separate metabolite sets. However, MSEA can automatically direct the investigator to biologically important pathways, and remove manual searching of pathway databases. When a list of compound names is entered into MetaboAnalyst, Over Representation Analysis (ORA) is performed and evaluates whether a metabolite set is represented more than expected by chance within the given compound list. Figure 8 shows an example of the output from a MSEA analysis of metabolites that were dysregulated after the treatment of rats with ionizing radiation exposure. A number of metabolic pathways were dysregulated, specifically, two metabolites, taurine and isethionic acid were mapped onto the taurine metabolism pathway.[19] MeltDB, like MetaboAnalyst is a software platform that processes raw data, carries out peak picking, statistical analysis, and visualization. It was first introduced in 2008 but has recently undergone a number of improvements including pathway mapping via ProMeTra.[83,84] ProMeTra allows visualization and mapping of metabolite, transcript, and protein ratios to metabolic pathways maps defined by the user.

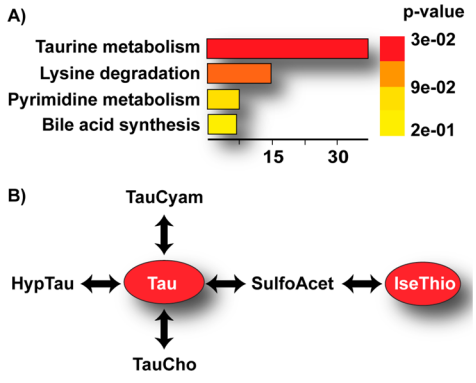

Figure 8. An example of metabolite set enrichment analysis (MSEA) using MetaboAnalyst.[66] (A) Enrichment of metabolic pathways after hypergeometric test to evaluate whether urinary metabolites upregulated in rats after radiation exposure are represented more than expected by chance within a compound list. (B) Taurine (Tau) and isethionic acid (IseThio) were found to be involved in the same pathway; taurine metabolism. Other metabolites in the taurine pathway are not changed; taurocyamine (TauCyam), sulfoacetaldehyde (SulfoAcet), hypotaurine (HypTau), and taurocholic acid (TauCho).

The new bioinformatic tools developed for pathway analysis use a number of different methodologies to achieve the same end-goal but have different advantages over each other. For example, some do not require the features to be identified first such as *mummichog*, which finds metabolic pathways based on *m/z*. Some require feature identification such as MetaMapp but can fill in gaps for metabolites not found in any pathways.[74] Others use the network analysis function as part of their workflow as in the autonomous/BDDA approach and also as part of the MetaboAnalyst software package.[66]

THE PAST AND THE FUTURE: STABLE ISOTOPES IN METABOLIC ANALYSIS

Stable isotopes have been used in targeted MS-based metabolomics as internal standards to validate the identity of a metabolite and also for elucidating and tracing the involvement of metabolites in pathways for the past 60 years.[85−94] The stable isotopes have the same physicochemical properties as natural abundance metabolites but contain at least one atom that is one mass unit greater, thus the isotope can be distinguishable from its natural abundance counterpart. A known metabolite such as glutamine for example can be introduced into an in vitro or in vivo experiment with at least one of its atoms labeled with a stable isotope, such as [1-^{13}C] or [5-^{13}C]. The incorporation of the ^{13}C-labeled atom(s) into other metabolites can show the potential impact of the labeled metabolite in pathways and on the biology of the system being studied. This is exemplified in a study where labeled ^{13}C was transferred from [5-^{13}C]glutamine to acetyl coenzyme A for lipid biosynthesis during hypoxia. The conversion of glutamine was traced by reductive flux, catalyzed by isocitrate dehydrogenase (IDH).[95] This study revealed a novel biological role for glutamine in lipogenesis during hypoxia. A follow-up study using a quantitative flux model with [U−^{13}C] glutamine and glucose showed that fatty acid labeling from glutamine does occur, but simultaneously with oxidative flux, and not by net reductive IDH flux.[96]

A logical progression from the stable isotope targeted metabolomic technology is to follow the full conversion of the labeled isotopes in an untargeted unbiased manner, allowing observation of the metabolic fate of the metabolites. For metabolites such as glucose, glutamine, or acetyl coenzyme A which are involved in multiple pathways, they can be used as surrogate markers to infuse into the experiment and assess widespread metabolic flux within a perturbed system (influenced by a disease, environmental factor, genetic change, microbial influence, xenobiotic use). However, infusing these precursors will also result in precursor-related metabolic perturbations. Furthermore, it is somewhat problematic to introduce labeled precursors at biologically relevant concentrations. In rodent models most of these metabolites will be

dx.doi.org/10.1021/ac5040693 | *Anal. Chem.* XXXX, XXX, XXX−XXX

completely metabolized within a few minutes and undetectable in biofluids by LC/MS. Designing an experiment that shows the direct involvement of a metabolite with the outcome of the disease pathogenesis may be more useful. For example a metabolite shown through conventional untargeted metabolomics to be downregulated in a disease, could be administered as a stable isotope, and reverse the disease phenotype. This would allow the metabolic fate to be revealed which may not have been seen by targeted analysis (due to only a small number of metabolites being targeted) or conventional untargeted analysis (due to the complexity of the data set). The simplest experiment would be to have a fully labeled metabolite, where all the ^{13}C atoms would be transferred to its products and cofactors. More complicated experiments arise when the precursor metabolite is partially labeled, since the metabolite may split and those unlabeled atoms would not provide a labeled isotopomer. Glucose, for example, splits into glyceraldehyde 3-phosphate and dihydroxyacetone phosphate during glycolysis; these metabolites are converted into pyruvate or used for lipid biosynthesis, respectively. Partial labeling of glucose would allow the observation of how these atoms are distributed during glycolysis. Another issue is synchronizing metabolism. According to Bar-Joseph et al.,[97] cells need to be arrested so that they start at the same phase, and even then they may lose their synchronization after awhile. Therefore, determining what phase the cells are in will help decide what time point is best for sampling; this however is somewhat impossible for in vivo experiments but useful for in vitro validation studies.

Similar to conventional untargeted metabolomics, which ultimately provides a list of dysregulated features, stable isotope untargeted metabolomics would provide the same list of metabolites but containing a metabolically derived stable isotope perturbed to the same extent (fold change and statistical significance) as the natural abundance metabolite. This would allow direct comparison of the paired isotopologues. This concept is one that many metabolomic researchers would seize upon to use but has been technologically difficult to set up due to many constraints; algorithms capable of detecting metabolite isotopologues are needed, and databases with MS/MS fragmentation data are required for identifying the isotopologues and isotopomers. Huang et al.[98] have made great strides into providing the bioinformatics tools to make this possible. Using the XCMS platform they have extended its capabilities to identify isotopologue groups that correspond to isotopically labeled compounds. The aptly named X^{13}CMS program can thus be used to track the metabolism of isotopically labeled substrates in an untargeted manner revealing valuable insights into metabolic mechanisms. Another major advancement to stable isotope untargeted metabolomics has been the recent development of a database containing thousands of metabolite isotopes. This isotope-based database, isoMETLIN will allow users to find all possible isotopologues derived from METLIN and obtain MS/MS fragmentation data on isotopomers[99] (Figure 9).

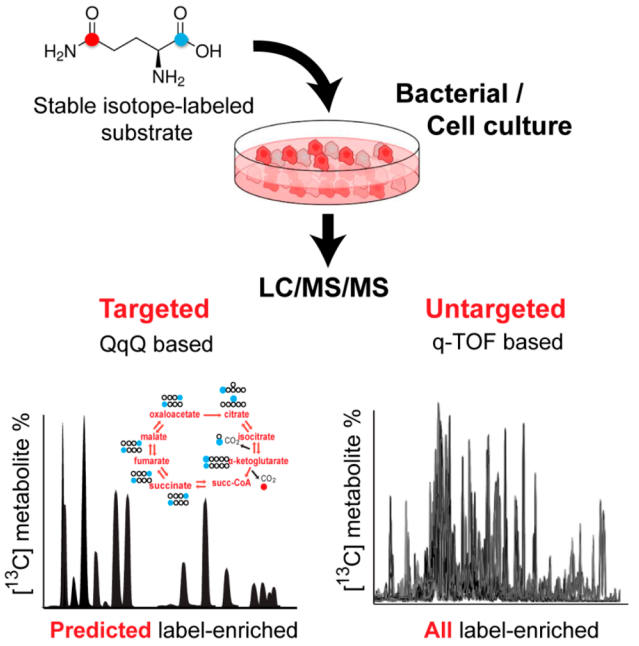

Figure 9. Stable isotope untargeted and targeted metabolomics.

CONCLUSIONS

Recent developments in bioinformatic tools have enhanced the untargeted metabolomic workflow, enabling researchers to identify metabolite features from LC/MS data and assign their biological roles by identifying their involvement in chemical pathways. Experiments carried out on high-resolution mass spectrometers result in thousands of dysregulated features. One of the biggest obstacles has been deconvolution and identification of these features, the latter requiring highly biased manual interpretation by the researcher. Automated multistep workflows have alleviated this process by incorporating functions that remove redundant features and enhance the efficiency and efficacy of metabolite identification. The novel use of stable isotopes for untargeted metabolomics and feature annotation has further enhanced the ability of the investigator to recover biologically relevant metabolites. Another major advancement has been in the development of metabolite databases and in silico fragmentation tools to help identify these metabolites for functionality in biological pathways. Indeed the area of growth for bioinformatics in metabolomic research will be in finding the role of these metabolites, rather than creating lists of biomarkers without mechanistic implications. This is somewhat dependent on the further curation of metabolite MS/MS fragmentation data in metabolite databases as well as the development of network mapping tools.

AUTHOR INFORMATION

Corresponding Author

*Phone: 858-784-9415. E-mail: siuzdak@scripps.edu.

Author Contributions

The manuscript was 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.

Biographies

Caroline H. Johnson received her BSc. in biomedical sciences from Keele University, U.K., her MSc. in Forensic Archaeological Science from University College London, U.K., and her Ph.D. in analytical chemistry from Imperial College London, U.K. She has carried out metabolomic research in the laboratories of Gary Siuzdak at The Scripps Research Institute, CA, and Frank J. Gonzalez at the National Cancer Institute, MD. Her research interests are in untargeted and targeted mass spectrometry-based metabolomics for the study of cancer metabolism and xenobiotics.

dx.doi.org/10.1021/ac5040693 | *Anal. Chem.* XXXX, XXX, XXX−XXX

Julijana Ivanisevic is currently working as a postdoctoral associate in Gary Siuzdak's Center for Metabolomics at The Scripps Research Institute, CA. She focuses her research interests and expertise in the field of mass spectrometry-based metabolomics, including technology development, biofluid screening, and bacterial and brain tissue metabolism. She has received her BSc. in Biology from University of Zagreb, Croatia, her MSc. in Life and Environmental Sciences from University of Pierre and Marie Curie (Paris VI), France, and her Ph.D. in Environmental Sciences from Aix-Marseille University, France.

H. Paul Benton received his BSc. in Biochemistry from the University of Birmingham, U.K., in 2004. After which he worked as a Research Technician in the Siuzdak laboratory at The Scripps Research Institute. He received his Ph.D in 2013 under Dr. Ebbels at Imperial College London in the Department of Surgery and Cancer. He is currently a Visiting Scientist in the Siuzdak laboratory.

Gary Siuzdak is Professor and Director at the Center for Metabolomics and Mass Spectrometry at the Scripps Research Institute in La Jolla, CA. He received his B.S. in Chemistry and B.A. in Mathematics from Rhode Island College and his Ph.D. in Physical Chemistry at Dartmouth College in New Hampshire. His research interests have been in whole virus analysis, nanostructure imaging mass spectrometry, and mass spectrometry-based metabolomics.

■ ACKNOWLEDGMENTS

This work was supported by grants from California Institute of Regenerative Medicine Grant No. TR1-01219 and the U.S. National Institutes of Health Grants R01 CA170737, R24 EY017540, P30 MH062261, RC1 HL101034, and P01 DA026146. The authors would also like to thank Nathaniel Mahieu for helpful discussions on stable isotopes.

■ REFERENCES

(1) Dunn, W. B.; Broadhurst, D. I.; Atherton, H. J.; Goodacre, R.; Griffin, J. L. *Chem. Soc. Rev.* **2011**, *40*, 387−426.
(2) Wishart, D. S. *Bioanalysis* **2011**, *3*, 1769−1782.
(3) Ivanova, P. T.; Milne, S. B.; Byrne, M. O.; Xiang, Y.; Brown, H. A. *Methods Enzymol.* **2007**, *432*, 21−57.
(4) Yang, J.; Eiserich, J. P.; Cross, C. E.; Morrissey, B. M.; Hammock, B. D. *Free Radical Biol. Med.* **2012**, *53*, 160−171.
(5) Kingsley, P. J.; Marnett, L. J. *Methods Enzymol.* **2007**, *433*, 91−112.
(6) Signorelli, P.; Hannun, Y. A. *Methods Enzymol.* **2002**, *345*, 275−294.
(7) Murphy, R. C.; Leiker, T. J.; Barkley, R. M. *Biochim. Biophys. Acta* **2011**, *1811*, 776−783.
(8) Koek, M. M.; Muilwijk, B.; van der Werf, M. J.; Hankemeier, T. *Anal. Chem.* **2006**, *78*, 1272−1281.
(9) Ivanisevic, J.; Zhu, Z.; Plate, L.; Tautenhahn, R.; Chen, S.; O'Brien, P. J.; Johnson, C. H.; Marletta, M. A.; Patti, G. J.; Siuzdak, G. *Anal. Chem.* **2013**, *85*, 6876−6884.
(10) Bi, H. C.; Pan, Y. Z.; Qiu, J. X.; Krausz, K. W.; Li, F.; Johnson, C. H.; Jiang, C. T.; Gonzalez, F. J.; Yu, A. M. *Carcinogenesis* **2014**, *35*, 2264−2272.
(11) Gonzalez-Dominguez, R.; Garcia, A.; Garcia-Barrera, T.; Barbas, C.; Gomez-Ariza, J. L. *Electrophoresis* **2014**, DOI: 10.1002/elps.201400196.
(12) Kalim, S.; Clish, C. B.; Wenger, J.; Elmariah, S.; Yeh, R. W.; Deferio, J. J.; Pierce, K.; Deik, A.; Gerszten, R. E.; Thadhani, R.; Rhee, E. P. *J. Am. Heart Assoc.* **2013**, *2*, e000542.
(13) Sampey, B. P.; Freemerman, A. J.; Zhang, J.; Kuan, P. F.; Galanko, J. A.; O'Connell, T. M.; Ilkayeva, O. R.; Muehlbauer, M. J.; Stevens, R. D.; Newgard, C. B.; Brauer, H. A.; Troester, M. A.; Makowski, L. *PLoS One* **2012**, *7*, No. e38812.
(14) Ishihara, K.; Katsutani, N.; Asai, N.; Inomata, A.; Uemura, Y.; Suganuma, A.; Sawada, K.; Yokoi, T.; Aoki, T. *Basic Clin. Pharmacol. Toxicol.* **2009**, *105*, 156−166.
(15) Dudzik, D.; Zorawski, M.; Skotnicki, M.; Zarzycki, W.; Kozlowska, G.; Bibik-Malinowska, K.; Vallejo, M.; Garcia, A.; Barbas, C.; Ramos, M. P. *J. Proteomics* **2014**, *103*, 57−71.
(16) Abaffy, T.; Moller, M. G.; Riemer, D. D.; Milikowski, C.; DeFazio, R. A. *Metabolomics: Off. J. Metabolomic Soc.* **2013**, *9*, 998−1008.
(17) Bahado-Singh, R. O.; Akolekar, R.; Mandal, R.; Dong, E.; Xia, J.; Kruger, M.; Wishart, D. S.; Nicolaides, K. *J. Matern.-Fetal Neonat. Med.* **2012**, *25*, 1840−1847.
(18) Lanz, C.; Patterson, A.; Slavík, J.; Krausz, K.; Ledermann, M.; Gonzalez, F.; Idle, J. *Radiat. Res.* **2009**, *172*, 198−212.
(19) Johnson, C. H.; Patterson, A. D.; Krausz, K. W.; Lanz, C.; Kang, D. W.; Luecke, H.; Gonzalez, F. J.; Idle, J. R. *Radiat. Res.* **2011**, *175*, 473−484.
(20) Bennett, M. J.; Ragni, M. C.; Hood, I.; Hale, D. E. *J. Inherited Metab. Dis.* **1992**, *15*, 220−223.
(21) Patti, G. J.; Yanes, O.; Shriver, L. P.; Courade, J. P.; Tautenhahn, R.; Manchester, M.; Siuzdak, G. *Nat. Chem. Biol.* **2012**, *8*, 232−234.
(22) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779−787.
(23) Frenzel, T.; Miller, A.; Engel, K. H. *Eur. Food Res. Technol.* **2003**, *216*, 335−342.
(24) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. *BMC Bioinf.* **2010**, *11*, 395.
(25) Katajamaa, M.; Oresic, M. *BMC Bioinf.* **2005**, *6*, 179.
(26) Tautenhahn, R.; Cho, K.; Uritboonthai, W.; Zhu, Z.; Patti, G. J.; Siuzdak, G. *Nat. Biotechnol.* **2012**, *30*, 826−828.
(27) Broeckling, C. D.; Afsar, F. A.; Neumann, S.; Ben-Hur, A.; Prenni, J. E. *Anal. Chem.* **2014**, *86*, 6812−6817.
(28) Benton, H. P.; Wong, D. M.; Trauger, S. A.; Siuzdak, G. *Anal. Chem.* **2008**, *80*, 6382−6389.
(29) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A. *Nucleic Acids Res.* **2013**, *41*, D801−D807.
(30) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T. *J. Mass Spectrom.* **2010**, *45*, 703−714.
(31) Stanstrup, J.; Gerlich, M.; Dragsted, L. O.; Neumann, S. *Anal. Bioanal. Chem.* **2013**, *405*, 5037−5048.
(32) Neumann, S.; Thum, A.; Bottcher, C. *Metabolomics: Off. J. Metabolomic Soc.* **2013**, *9*, S84−S91.
(33) Wolf, S.; Schmidt, S.; Muller-Hannemann, M.; Neumann, S. *BMC Bioinf.* **2010**, 11.
(34) Kopka, J.; Schauer, N.; Krueger, S.; Birkemeyer, C.; Usadel, B.; Bergmuller, E.; Dormann, P.; Weckwerth, W.; Gibon, Y.; Stitt, M.; Willmitzer, L.; Fernie, A. R.; Steinhauser, D. *Bioinformatics* **2005**, *21*, 1635−1638.
(35) Gerlich, M.; Neumann, S. *J. Mass Spectrom.* **2013**, *48*, 291−298.
(36) Creek, D. J.; Jankevics, A.; Breitling, R.; Watson, D. G.; Barrett, M. P.; Burgess, K. E. V. *Anal. Chem.* **2011**, *83*, 8703−8710.
(37) Menikarachchi, L. C.; Cawley, S.; Hill, D. W.; Hall, L. M.; Hall, L.; Lai, S.; Wilder, J.; Grant, D. F. *Anal. Chem.* **2012**, *84*, 9388−9394.
(38) Hall, L. M.; Hall, L. H.; Kertesz, T. M.; Hill, D. W.; Sharp, T. R.; Oblak, E. Z.; Dong, Y. W.; Wishart, D. S.; Chen, M. H.; Grant, D. F. *J. Chem. Inf. Model.* **2012**, *52*, 1222−1237.
(39) Boswell, P. G.; Schellenberg, J. R.; Carr, P. W.; Cohen, J. D.; Hegeman, A. D. *J. Chromatogr., A* **2011**, *1218*, 6732−6741.
(40) Rinehart, D.; Johnson, C. H.; Nguyen, T.; Ivanisevic, J.; Benton, H. P.; Lloyd, J.; Arkin, A. P.; Deutschbauer, A. M.; Patti, G. J.; Siuzdak, G. *Nat. Biotechnol.* **2014**, *32*, 524−527.

dx.doi.org/10.1021/ac5040693 | *Anal. Chem.* XXXX, XXX, XXX−XXX

(41) Scheltema, R. A.; Decuypere, S.; Dujardin, J. C.; Watson, D. G.; Jansen, R. C.; Breitling, R. *Bioanalysis* **2009**, *1*, 1551−1557.
(42) Brown, M.; Dunn, W.; Dobson, P.; Patel, Y.; Winder, C.; Francis-McIntyre, S.; Begley, P.; Carroll, K.; Broadhurst, D.; Tseng, A.; Swainston, N.; Spasic, I.; Goodacre, R.; Kell, D. *Analyst* **2009**, *134*, 1322−1332.
(43) Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G. *Anal. Chem.* **2012**, *84*, 5035−5039.
(44) Sturm, M.; Bertsch, A.; Gropl, C.; Hildebrandt, A.; Hussong, R.; Lange, E.; Pfeifer, N.; Schulz-Trieglaff, O.; Zerck, A.; Reinert, K.; Kohlbacher, O. *BMC Bioinf.* **2008**, *9*, 163.
(45) Yu, T. W.; Park, Y.; Johnson, J. M.; Jones, D. P. *Bioinformatics* **2009**, *25*, 1930−1936.
(46) Uppal, K.; Soltow, Q. A.; Strobel, F. H.; Pittard, W. S.; Gernert, K. M.; Yu, T.; Jones, D. P. *BMC Bioinf.* **2013**, *14*, 15.
(47) Scheltema, R. A.; Jankevics, A.; Jansen, R. C.; Swertz, M. A.; Breitling, R. *Anal. Chem.* **2011**, *83*, 2786−2793.
(48) Kuhl, C.; Tautenhahn, R.; Bottcher, C.; Larson, T. R.; Neumann, S. *Anal. Chem.* **2012**, *84*, 283−289.
(49) Bottcher, C.; von Roepenack-Lahaye, E.; Schmidt, J.; Schmotz, C.; Neumann, S.; Scheel, D.; Clemens, S. *Plant Physiol.* **2008**, *147*, 2107−2120.
(50) Cho, K.; Mahieu, N. G.; Johnson, S. L.; Patti, G. J. *Curr. Opin. Biotechnol.* **2014**, *28*, 143−148.
(51) Daly, R.; Rogers, S.; Wandy, J.; Jankevics, A.; Burgess, K. E.; Breitling, R. *Bioinformatics* **2014**, *30*, 2764−2771.
(52) Fernandez-Albert, F.; Llorach, R.; Andres-Lacueva, C.; Perera-Lluna, A. *Anal. Chem.* **2014**, *86*, 2320−2325.
(53) Trygg, J.; Holmes, E.; Lundstedt, T. *J. Proteome Res.* **2007**, *6*, 469−479.
(54) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*, 2nd ed.; Springer: New York, 2009.
(55) Mahieu, N. G.; Huang, X.; Chen, Y. J.; Patti, G. J. *Anal. Chem.* **2014**, *86*, 9583−9589.
(56) Giavalisco, P.; Kohl, K.; Hummel, J.; Seiwert, B.; Willmitzer, L. *Anal. Chem.* **2009**, *81*, 6546−6551.
(57) Giavalisco, P.; Hummel, J.; Lisec, J.; Inostroza, A. C.; Catchpole, G.; Willmitzer, L. *Anal. Chem.* **2008**, *80*, 9417−9425.
(58) Giavalisco, P.; Li, Y.; Matthes, A.; Eckhardt, A.; Hubberten, H. M.; Hesse, H.; Segu, S.; Hummel, J.; Kohl, K.; Willmitzer, L. *Plant J.: Cell Mol. Biol.* **2011**, *68*, 364−376.
(59) Baran, R.; Bowen, B. P.; Bouskill, N. J.; Brodie, E. L.; Yannone, S. M.; Northen, T. R. *Anal. Chem.* **2010**, *82*, 9034−9042.
(60) Shahaf, N.; Franceschi, P.; Arapitsas, P.; Rogachev, I.; Vrhovsek, U.; Wehrens, R. *Rapid Commun. Mass Spectrom.* **2013**, *27*, 2425−2431.
(61) Kind, T.; Fiehn, O. *BMC Bioinf.* **2006**, *7*, 234.
(62) Ojanpera, S.; Pelander, A.; Pelzing, M.; Krebs, I.; Vuori, E.; Ojanpera, I. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 1161−1167.
(63) Wolff, J. C.; Eckers, C.; Sage, A. B.; Giles, K.; Bateman, R. *Anal. Chem.* **2001**, *73*, 2605−2612.
(64) Wolff, J. C.; Fuentes, T. R.; Taylor, J. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1216−1219.
(65) Gowda, H.; Ivanisevic, J.; Johnson, C. H.; Kurczy, M. E.; Benton, H. P.; Rinehart, D.; Nguyen, T.; Ray, J.; Kuehl, J.; Arevalo, B.; Westenskow, P. D.; Wang, J.; Arkin, A. P.; Deutschbauer, A. M.; Patti, G. J.; Siuzdak, G. *Anal. Chem.* **2014**, *86*, 6931−6939.
(66) Xia, J. G.; Psychogios, N.; Young, N.; Wishart, D. S. *Nucleic Acids Res.* **2009**, *37*, W652−W660.
(67) Patti, G. J.; Tautenhahn, R.; Rinehart, D.; Cho, K.; Shriver, L. P.; Manchester, M.; Nikolskiy, I.; Johnson, C. H.; Mahieu, N. G.; Siuzdak, G. *Anal. Chem.* **2013**, *85*, 798−804.
(68) Liu, H.; Sadygov, R. G.; Yates, J. R., 3rd *Anal. Chem.* **2004**, *76*, 4193−4201.
(69) Li, S.; Park, Y.; Duraisingham, S.; Strobel, F. H.; Khan, N.; Soltow, Q. A.; Jones, D. P.; Pulendran, B. *PLoS Comput. Biol.* **2013**, *9*, e1003123.
(70) Ogata, H.; Goto, S.; Sato, K.; Fujibuchi, W.; Bono, H.; Kanehisa, M. *Nucleic Acids Res.* **1999**, *27*, 29−34.
(71) Duarte, N. C.; Becker, S. A.; Jamshidi, N.; Thiele, I.; Mo, M. L.; Vo, T. D.; Srivas, R.; Palsson, B. O. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1777−1782.
(72) Ma, H.; Sorokin, A.; Mazein, A.; Selkov, A.; Selkov, E.; Demin, O.; Goryanin, I. *Mol. Syst. Biol.* **2007**, *3*, 135.
(73) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. *Genome Res.* **2003**, *13*, 2498−2504.
(74) Barupal, D. K.; Haldiya, P. K.; Wohlgemuth, G.; Kind, T.; Kothari, S. L.; Pinkerton, K. E.; Fiehn, O. *BMC Bioinf.* **2012**, *13*, 99.
(75) Rogers, D. J.; Tanimoto, T. T. *Science* **1960**, *132*, 1115−1118.
(76) Xia, J.; Wishart, D. S. *Bioinformatics* **2010**, *26*, 2342−2344.
(77) Xia, J.; Wishart, D. S. *Nucleic Acids Res.* **2010**, *38*, W71−77.
(78) Austin, C. P.; Brady, L. S.; Insel, T. R.; Collins, F. S. *Science* **2004**, *306*, 1138−1139.
(79) Degtyarenko, K.; de Matos, P.; Ennis, M.; Hastings, J.; Zbinden, M.; McNaught, A.; Alcantara, R.; Darsow, M.; Guedj, M.; Ashburner, M. *Nucleic Acids Res.* **2008**, *36*, D344−350.
(80) Feist, A. M.; Herrgard, M. J.; Thiele, I.; Reed, J. L.; Palsson, B. O. *Nat. Rev. Microbiol.* **2009**, *7*, 129−143.
(81) Karp, P. D.; Ouzounis, C. A.; Moore-Kochlacs, C.; Goldovsky, L.; Kaipa, P.; Ahren, D.; Tsoka, S.; Darzentas, N.; Kunin, V.; Lopez-Bigas, N. *Nucleic Acids Res.* **2005**, *33*, 6083−6089.
(82) Matthews, L.; Gopinath, G.; Gillespie, M.; Caudy, M.; Croft, D.; de Bono, B.; Garapati, P.; Hemish, J.; Hermjakob, H.; Jassal, B.; Kanapin, A.; Lewis, S.; Mahajan, S.; May, B.; Schmidt, E.; Vastrik, I.; Wu, G.; Birney, E.; Stein, L.; D'Eustachio, P. *Nucleic Acids Res.* **2009**, *37*, D619−622.
(83) Neuweger, H.; Albaum, S. P.; Dondrup, M.; Persicke, M.; Watt, T.; Niehaus, K.; Stoye, J.; Goesmann, A. *Bioinformatics* **2008**, *24*, 2726−2732.
(84) Neuweger, H.; Persicke, M.; Albaum, S. P.; Bekel, T.; Dondrup, M.; Huser, A. T.; Winnebald, J.; Schneider, J.; Kalinowski, J.; Goesmann, A. *BMC Syst. Biol.* **2009**, *3*, 82.
(85) Rodgers, R. P.; Blumer, E. N.; Hendrickson, C. L.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 835−840.
(86) Fischer, E.; Sauer, U. *Eur. J. Biochem.* **2003**, *270*, 880−891.
(87) Lamonte, G.; Tang, X.; Chen, J. L.; Wu, J.; Ding, C. K.; Keenan, M. M.; Sangokoya, C.; Kung, H. N.; Ilkayeva, O.; Boros, L. G.; Newgard, C. B.; Chi, J. T. *Cancer Metab.* **2013**, *1*, 23.
(88) Harris, D. M.; Li, L.; Chen, M.; Lagunero, F. T.; Go, V. L.; Boros, L. G. *Metabolomics: Off. J. Metabolomic Soc.* **2012**, *8*, 201−210.
(89) Zimmermann, M.; Sauer, U.; Zamboni, N. *Anal. Chem.* **2014**, *86*, 3232−3237.
(90) Yao, D.; Shi, X.; Wang, L.; Gosnell, B. A.; Chen, C. *Drug Metab. Dispos.* **2013**, *41*, 79−88.
(91) Harada, K.; Fukusaki, E.; Bamba, T.; Sato, F.; Kobayashi, A. *Biotechnol. Prog.* **2006**, *22*, 1003−1011.
(92) Bloch, K.; Rittenberg, D. *J. Biol. Chem.* **1942**, *145*, 625.
(93) Bloch, K.; Rittenberg, D. *J. Biol. Chem.* **1945**, *159*, 45.
(94) Schoenheimer, R.; Rittenberg, D. *Science* **1938**, *87*, 221−226.
(95) Metallo, C. M.; Gameiro, P. A.; Bell, E. L.; Mattaini, K. R.; Yang, J.; Hiller, K.; Jewell, C. M.; Johnson, Z. R.; Irvine, D. J.; Guarente, L.; Kelleher, J. K.; Vander Heiden, M. G.; Iliopoulos, O.; Stephanopoulos, G. *Nature* **2012**, *481*, 380−384.
(96) Fan, J.; Kamphorst, J. J.; Rabinowitz, J. D.; Shlomi, T. *J. Biol. Chem.* **2013**, *288*, 31363−31369.
(97) Bar-Joseph, Z. *Bioinformatics* **2004**, *20*, 2493−2503.
(98) Huang, X.; Chen, Y. J.; Cho, K.; Nikolskiy, I.; Crawford, P. A.; Patti, G. J. *Anal. Chem.* **2014**, *86*, 1632−1639.
(99) Cho, K.; Mahieu, N.; Ivanisevic, J.; Uritboonthai, W.; Chen, Y. J.; Siuzdak, G.; Patti, G. J. *Anal. Chem.* **2014**, *86*, 9358−9361.

dx.doi.org/10.1021/ac5040693 | *Anal. Chem.* XXXX, XXX, XXX−XXX