

Plant Metabolomics: A Cutting-Edge Approach

Dharitri Saikia¹, Pulakeswar Basumatary²

¹PhD Research Scholar Centre for the Environment
Indian Institute of Technology Guwahati

Abstract:

Metabolomics is a dynamic and multidisciplinary area of research focusing on the identification and quantification of metabolites, small molecules produced within cells, biofluids, tissues or organisms as a result of metabolic processes. Through the analyses of metabolites, such as amino acids, sugars, and lipids, researchers can understand the valuable insights into an organism's metabolic state and its response to environmental and genetic changes. It is a rapidly advancing area of scientific research that combines systems biology, biochemistry and analytical chemistry to find out the intricate inter-relationship between metabolites and their surroundings. The rise of metabolomics as a field has been driven by the advances in analytical technology. Various tools like gas chromatography-mass spectrometry (GC-MS), liquid chromatography-mass spectrometry (LC-MS), nuclear magnetic resonance (NMR) spectroscopy, high-resolution mass spectrometry (HRMS), and matrix-assisted laser desorption/ionization (MALDI) have unlocked new ways to study and understand the metabolites. These advanced techniques allow scientists to identify and quantify from extremely small molecules to large molecules in unprecedented detail, making it possible to explore the metabolic processes of living organisms more profoundly than ever before. In the case of plant metabolomics, the focus is specifically on the metabolites found in plant cells and tissues. By analyzing these compounds, researchers can get a deeper understanding of how plant metabolites function, influencing growth, development, and the ability to respond to stresses. It helps to identify the important metabolites tied to specific traits, key metabolic pathways, and explains how plants can adapt to environmental and genetic changes. In this review, we have briefly discussed the untargeted metabolomics, its workflow and applications with emphasis on plants.

Keywords: Plant metabolomics; untargeted metabolomics; mass spectrometry; data analysis; metabolites

INTRODUCTION

Metabolomics is a dynamic and multidisciplinary area of research focusing on the identification and quantification of metabolites, small molecules produced within cells, biofluids, tissues or organisms as a result of metabolic processes (Clish 2015a). Through the analyses of

metabolites, such as amino acids, sugars, and lipids, researchers can understand the valuable insights into an organism's metabolic state and its response to environmental and genetic changes (Rinschen et al. 2019). It is rapidly advancing area of scientific research that combines systems biology, biochemistry and analytical chemistry to

*Corresponding author's Email: dharitrisaikia@iitg.ac.in

find out the intricate inter-relationship between metabolites and their surrounding (Piazza et al. 2018; Manickam et al. 2023). The rise of metabolomics as a field has been driven by the advances in analytical technology. Various tools like gas chromatography-mass spectrometry (GC-MS), liquid chromatography-mass spectrometry (LC-MS), nuclear magnetic resonance (NMR) spectroscopy, high-resolution mass spectrometry (HRMS), and matrix-assisted laser desorption/ionization (MALDI) have unlocked new ways to study and understand the metabolites. These advanced techniques allow scientists to identify and quantify from extremely small molecules to large molecules in unprecedented detail, making it possible to explore the metabolic processes of living organisms more profoundly than ever before (Clish 2015b). This advanced approach allows scientists to detect and measure metabolites in complex samples such as tissue extracts, blood, serum, and urine etc. In the case of plant metabolomics, the focus is specifically on the metabolites found in plant cells and tissues. By analyzing these compounds, researchers can get a deeper understanding of how plant metabolites function, influencing growth, development, and the ability to respond to stresses. It helps to identify the important metabolites tied to specific traits, key metabolic pathways, and explains how plants can adapt to environmental and genetic changes (Manickam et al. 2023). Moreover, metabolomics in agricultural science investigates the metabolic pathways that facilitate crop growth and development. This also enables the identification of metabolites enhancing crop nutritional value and development of nutrient-enriched crop varieties (Ibarra-Estrada et al. 2016). Depending on the research approaches, metabolomics can be divided into two primary categories: targeted and untargeted (Lelli et al., 2021). Targeted metabolomics mainly involves in precise identification and quantification of a predefined set of well-characterized metabolites, typically those associated with a specific metabolic pathway or stress response mechanism (Roberts et

al. 2012a). This utilizes the already established internal standards for semi-quantitative or quantitative analysis. In contrast, untargeted metabolomics deals with a discovery-driven approach, looking for comprehensively identification and quantification of all metabolites present in a biological sample, regardless of their known or unknown identity, function, or biological significance (Schrimpe-Rutledge et al., 2016; Snart et al., 2015). This produces vast datasets, requiring advanced statistical tools like multivariate analysis to extract meaningful insights and simplify the data (Chen *et al.* 2022). In this article, we have briefly discussed the workflow of untargeted plant metabolomic analysis, from sample preparation to data analysis. The protocols for targeted metabolomics have been detailed well by Roberts and team (Roberts et al. 2012b).

1. Sample preparation

Sample preparation is critical in metabolomic studies, significantly impacting result reliability. Small variations in sample collection, metabolite extraction, or storage can compromise metabolite stability, leading to substantial changes in the observed results. To mitigate this, metabolomics samples must be collected uniformly and rapidly, considering the rapid enzymatic turnover rate of the metabolome (Salem et al. 2020). Plants parts like leaf, root, stems etc. can be used for metabolite extraction. Rapid freezing (flash-freezing) methods using liquid nitrogen (or dry ice) is strongly recommended for preserving fresh samples during harvesting to stop enzyme-induced metabolic changes. This immediate freezing step effectively halts metabolic activity, ensuring the integrity of the sample for subsequent metabolomics analysis (Sumner et al. 2007). The preparation of plant samples for metabolomics analysis typically encompasses four distinct phases: harvesting, drying (optional, fresh samples can also be used for extraction), extraction (if required), and sample preparation for analysis. Minor changes in sample

preparation is allowed depending on the analytical method and experimental design (Kim and Verpoorte 2010). Change in environmental condition induces variation in metabolome composition of the samples to be studied. Therefore, samples are generally grown under controlled environmental condition or collected from the same environment at the same time and condition to minimize the variation. Post-harvest, immediate freezing is crucial to halt enzymatic reactions triggered by handling and wounding. The most effective method is to flash-freeze the samples in liquid nitrogen to preserve metabolite integrity and quenching (Sumner et al. 2007). In some cases, such as NMR analysis, drying is essential before extraction of the plant metabolome. Flash-frozen samples can be stored in -80°C for a short period of time. Before metabolite extraction, samples can be lysed, lyophilized or homogenized according to the experimental requirement (Salem et al. 2020).

2. Metabolite extraction

Effective metabolite extraction is crucial for successful metabolomics experiments. For untargeted metabolomics, optimal extraction methods are vital to release a diverse spectrum of metabolites, ensuring comprehensive analysis and accurate representation of the sample's metabolome (Roberts et al. 2012b). An appropriate extraction solvent is chosen based on polarity, which governs metabolite solubility ('like-dissolves-like' principle); selectivity, targeting specific metabolite classes and toxicity and inertness, ensuring safe and unbiased extraction. For novel, exploratory studies, a series of solvents may require to be tested for ensuring maximum metabolite extraction. For LC/MS based practices, methanol or acetonitrile acidified with 0.1% formic acid is widely used for untargeted metabolite extraction. For GC/MS, however, volatile organic solvents such as ethyl acetate, dichloromethane, toluene and hexane etc. are used (Gaylor et al. 2018). Derivatization, a chemical modification that enhances the separation and

sensitivity of detection of analytes is mandatory in GC/MS based practices. A set of biological replicates and a pooled QC stock sample prepared by pooling a specific volume from each sample of the experiment is highly recommended. This increases the reproducibility of the experiment and helps in signal normalization and identification and annotation of unknown metabolite (Souza and Patti 2021). Plant tissues such as leaves, stems or roots are homogenized well using liquid nitrogen, vortexed, ultrasonicated and centrifuged in the extraction solvent and the supernatant is collected for analysis.

3. Sample analysis

Although untargeted metabolomics aims to identify the global metabolome in the studied sample, it faces limitations due to the diverse chemical properties of metabolites. Experimental conditions, such as solvents and instrumentation, affect which metabolites are detectable. In general, methods that are suitable for only a specific class or few are avoided as this decreases the global metabolome coverage of the sample. Liquid Chromatography/Mass Spectrometry (LC/MS) is the gold standard for untargeted metabolomic study due to its capability to detect numerous metabolites and highly sensitive nature (Souza and Patti 2021). A general untargeted plant metabolomic analysis starts with extracting the frozen plant tissue (e.g. 100 mg) in methanol/water (80:20, v/v) using 10-minute ultrasonication and 13,000 rpm centrifugation for 10 minutes. The resulting supernatant is evaporated and reconstituted in LC-MS grade solvent (50% acetonitrile, 0.1% formic acid). For LC-MS-based metabolomics analysis, pre-filtering of the samples using 0.22 µm nylon syringe filters is crucial for removing contaminants, preventing column clogging and to enhance the efficacy of the experiment. LC-MS analysis generally utilizes a reverse-phase C18 column with a 0-20 minute gradient elution (5-100% acetonitrile) at 0.3 mL/min. Electrospray ionization mass spectrometry (ESI-MS) within a scanning range of 50-1000 m/z

in both positive and negative polarities is suggested (Want et al. 2013).

4. Processing of raw files

In untargeted plant metabolomics, raw LC-MS files are processed through a series of steps to extract meaningful information. Initially, raw files are imported into specialized software (e.g., Progenesis QI, MetaboAnalyst, XCMS, MZmine) for peak detection, alignment, and feature extraction. Noise, isotopes, and adducts are removed through data filtering, followed by normalization (e.g., total sum normalization) to correct for variability. A typical MS data processing workflow consists of importing the raw data file, mass detection, smoothing/filtering (optional), peak picking, deisotoping, alignment, gap filling, and annotation. Metabolite identification is performed using databases (e.g., KEGG, PlantCyc). Finally, pathway analysis is conducted to understand the overall metabolic network. Statistical analysis (e.g., PCA, ANOVA) is then applied to identify significant metabolites. MetaboAnalyst (Pang et al. 2024), MZmine (Schmid et al. 2023), XCMS (Tautenhahn et al. 2012) are widely used user-friendly, open-source platforms that enable investigators to pre-process, analyse and visualize large metabolomic datasets easily. A graphic representation of the data-processing workflow is given in **Figure 1**. All the tools follow almost the same operating protocol, from raw data processing to statistical analysis. Here, we have discussed briefly the steps followed in MZmine3.3.0. This tool supports .mzML, .mzXML, .imzML, .netCDF, and .aird file formats. Raw files can be converted to any of these formats utilizing msConvert, a command-line utility for converting mass-spectrometry file format. msConvert enhances the quality of mass spectrometry data through peak picking, which extracts prominent peaks and minimizes noise and irrelevant information. This refined data facilitates accurate identification and quantification of target compounds by downstream analysis tools. Following conversion to .mzXML files, the data is

imported into MZmine 3.3.0 software via the "Raw Data Import" tool under the "Raw Data Methods" tab. Once all files are selected, mass detection is performed by navigating to "Raw Data Methods" and accessing "Feature Detection" and then "Mass Detection", enabling comprehensive analysis.

In MZmine 3.3.0, the centroid mass detector processes raw mass spectrometry data to identify and extract individual peaks. This centroiding process converts complex mass spectra into discrete peaks representing single compounds or ions. For feature detection and chromatogram construction, the ADAP chromatogram builder module identifies target compounds' m/z values based on user-defined lists. It then extracts corresponding intensity values and constructs chromatograms. Adjustments can be made as needed to suit experimental requirements. Chromatogram resolution is achieved through the Local Minimum Resolver, accessible via Feature Detection > Chromatogram Resolving > Local Minimum Resolver. This critical feature-resolving step separates overlapping and co-eluting peaks, ensuring accurate data pre-processing. In this process, parameters like chromatographic threshold, minimum absolute height, peak duration range etc. are to be adjusted based on the experiment. To eliminate the repetitive features, for example, the ones generated due to the presence of isotopes, isotope filters is applied which removes the possible isotope signals. In mass spectrometry, ionization of a compound can generate multiple peaks representing different isotopes. For instance, a compound with one carbon and two nitrogen atoms (C1N2) may produce peaks at m/z 42, 43, and 44, corresponding to carbon-12, carbon-13, nitrogen-14, and nitrogen-15 isotopes. The Isotope Pattern Finder tool, accessible via "Feature List Methods" > "Isotopes" > "Isotope Pattern Finder", identifies these isotopic peaks to confirm compound identity. Following isotopic analysis, the Join Aligner module aligns corresponding features

across samples based on mass, retention time, and user-defined tolerance ranges. This module is accessed through "Feature List Methods" > "Alignment" > "Join Aligner". To optimize the results, various parameters within the "Join Aligner" module can be adjusted according to the research objectives.

To improve the quality of the raw feature data, a "Feature List Filtering" step is used to remove low-quality or noisy features based on set criteria, ensuring that the dataset remains reliable. Further, the "Gap Filling" step is applied to address missing data, which could result from instrument errors, sample handling problems, or other limitations. This is an important step which helps in making the dataset more complete, enabling better statistical analysis and easier identification of unknown metabolites. Once this process is completed, the cleaned-up feature list is matched against the various databases like, KEGG database for molecular annotations. Lastly, the annotated results can be exported for further analysis and visualization, providing a more complete picture of the metabolomic findings.

5. Statistical Analysis

Analyzing metabolomic data involves different statistical techniques based on the type of experiment being conducted. Researchers often use both the univariate and multivariate analyses. Univariate analysis provides a straightforward look at the data, making it easier to identify important conditions. For comparing two groups, methods such as t-tests, fold change analysis, and volcano plots are commonly used. When looking at multiple groups, one-way ANOVA is usually the go-to method, and researchers often follow up with post hoc tests and correlation analysis to dig deeper into the relationships between the variables (Cambiaghi et al. 2017). With many metabolites present in each sample, the risk of false positives may increase during multiple testing. To avoid this, researchers mostly use the correction methods like Bonferroni adjustment and Benjamini-Hochberg's False Discovery Rate (FDR) to refine p-values and minimize false positives.

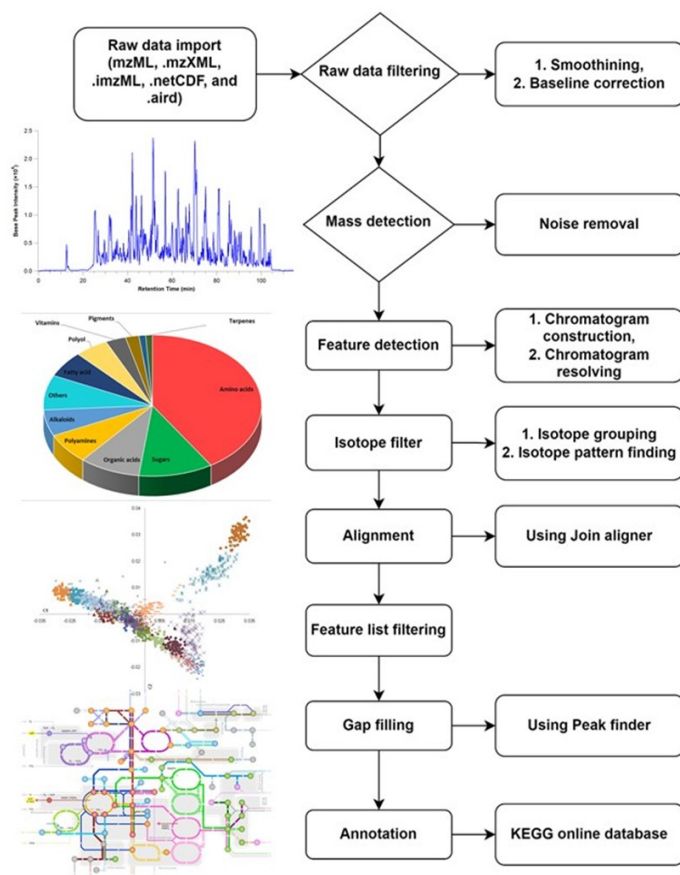
Multivariate analyses play a crucial role in making sense of complex data by looking at several statistical variables at the same time. This approach gives researchers a clearer picture of the underlying patterns. These methods are particularly helpful when we are dealing with exploratory datasets, as they simplify large amounts of information into few and to more significant variables. The toolbox for multivariate analysis includes techniques like multiple regression, principal component analysis (PCA), factor analysis, ANOVA, and partial least squares discriminant analysis (PLS). Moreover, machine learning methods such as support vector machines (SVM) and random forests are also part of this toolkit. Cluster analysis adds another dimension by revealing hidden patterns and groupings within the data. By utilizing these advanced statistical techniques, researchers can uncover valuable insights from intricate datasets, paving the way for new discoveries and innovations (Chen et al. 2022). In untargeted plant metabolomics, PCA is a key tool for handling complex datasets. This powerful statistical method helps ensure data quality and consistency by identifying patterns, trends, outliers, and anomalies, which in turn validates both instrument performance and sample reliability. It also reduces its dimensionality, making it easier to visualize relationships between samples and metabolites. This allows researchers to uncover correlated metabolic networks and understanding the deeper scientific explorations. (Chen et al. 2022).

6. Applications of untargeted plant metabolomics

As plant science continues to evolve, metabolomics has emerged as a powerful tool that offers profound insights into the complex cellular processes that underpin plant biology. The wide variety of metabolites produced by plants is essential for numerous functions, including regulating growth and development, producing vibrant pigments, creating aromatic compounds, and enhancing resilience against environmental stressors. Gaining a deeper understanding of these metabolites is crucial for unraveling how factors

like the environment and genetic traits influence plant physiology and biochemistry. By identifying and quantifying these metabolites, researchers can interpret the intricate language of plant metabolism, revealing how plants respond to stress, adapt to changing conditions, and improve their growth and productivity. This knowledge is vital for developing crops that are more resilient, sustainable, and high-yielding, ultimately supporting global food security and promoting environmental sustainability (Bozza *et al.* 2024). Scientists are exploring to understand how plants respond to different environmental stressors through metabolically, and aiming to understand what makes crops resilient. Exposing to challenges

like extreme weather or pests, plants activate intricate defense strategies, adjusting their internal chemistry and producing special protective compounds (metabolites). Understanding these natural adaptations became the key to developing crops which can better endure climate change and resist to diseases (Litvinov *et al.* 2021). Untargeted metabolomics may also be utilized for rapid identification and quantification of phytonutrients (known as phytochemicals or antioxidants) (Girija *et al.* 2021). In recent years, it is extensively used to investigate biochemical changes occurring before and after the harvesting of various crops, including vegetables, fruits, grains and ornamental plants.



7. Conclusion

Plant metabolomics has emerged as a powerful tool of research in recent years because it provides a comprehensive view of the metabolic status of plant physiology and metabolism. However, the metabolomics analysis presents several challenges such as the complexity and diversity of the metabolome, a dynamic range of metabolite concentrations and variability in sample preparation. The cost of procuring the analytical instruments and high purity chemicals pose another challenge for the budding researchers or scientists. Knowledge on computational analysis and statistics is also an indispensable prerequisite. However, this has been substantially reduced through the development of several easy-to-use, open-source, web-based or installable tools in the recent years. To be completely comprehended, untargeted metabolomic data must be combined and processed with other omic data. Nowadays, the most popular method is to perform integration analysis to extract structural and dynamic changes in the underlying biological network of interest while concurrently monitoring the levels of transcripts, proteins, and metabolites. In future, advances in technology and analytics will propel plant metabolomics forward, deepening our understanding of plant systems and unlocking innovative, sustainable solutions.

References:

- Bozza D, Barboni D, Spadafora ND, et al (2024) Untargeted metabolomics approaches for the characterization of cereals and their derived products by means of liquid chromatography coupled to high resolution mass spectrometry. *Journal of Chromatography Open* 6:100168. <https://doi.org/10.1016/J.JCOA.2024.100168>
- Cambiaghi A, Ferrario M, Masseroli M (2017) Analysis of metabolomic data: Tools, current strategies and future challenges for omics data integration. *Brief Bioinform* 18:498–510. <https://doi.org/10.1093/bib/bbw031>
- Chen Y, Li EM, Xu LY (2022) Guide to Metabolomics Analysis: A Bioinformatics Workflow. *Metabolites* 12
- Clish CB (2015a) Metabolomics: an emerging but powerful tool for precision medicine. *Cold Spring Harb Mol Case Stud* 1:a000588. <https://doi.org/10.1101/MCS.A000588>
- Clish CB (2015b) Metabolomics: an emerging but powerful tool for precision medicine. <https://doi.org/10.1101/mcs.a000588>
- Gaylor MO, Juntunen HL, Hazelwood D, Videau P (2018) Assessment of Multiple Solvents for Extraction and Direct GC–MS Determination of the Phytochemical Inventory of *Sansevieria Extrafoliar Nectar Droplets*. *J Chromatogr Sci* 56:293–299. <https://doi.org/10.1093/CHROMSCI/BMY008>
- Girija A, Yadav R, Corke F, et al (2021) Untargeted Metabolomic Profiling Reveals Variation in Metabolites Associated with Nutritional Values in Tef Accessions. *Plant Foods for Human Nutrition* 76:536–539. <https://doi.org/10.1007/s11130-021-00931-6>
- Ibarra-Estrada E, Palma-Tenango RMS-H and M, Ibarra-Estrada E, Palma-Tenango RMS-H and M (2016) Metabolomics as a Tool in Agriculture. *Metabolomics - Fundamentals and Applications*. <https://doi.org/10.5772/66485>
- Kim HK, Verpoorte R (2010) Sample preparation for plant metabolomics. *Phytochemical Analysis* 21:4–13. <https://doi.org/10.1002/PCA.1188>
- Lelli V, Belardo A, Timperio AM, et al (2021) From Targeted Quantification to Untargeted Metabolomics. *Metabolomics - Methodology and Applications in Medical Sciences and Life Sciences*. <https://doi.org/10.5772/INTECHOPEN.96852>
- Litvinov DY, Karlov GI, Divashuk MG (2021) Metabolomics for Crop Breeding: General Considerations. *Genes (Basel)* 12: <https://doi.org/10.3390/GENES12101602>
- Manickam S, Rajagopalan VR, Kambale R, et al (2023) Plant Metabolomics: Current Initiatives and Future Prospects. *Curr Issues Mol Biol* 45:8894–8906
- Pang Z, Lu Y, Zhou G, et al (2024) MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation. *Nucleic Acids Res* 52:W398–W406. <https://doi.org/10.1093/NAR/GKAE253>

- Piazza I, Kochanowski K, Cappelletti V, et al (2018) A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. *Cell* 172:358-372.e23. <https://doi.org/10.1016/J.CELL.2017.12.006>
- Rinschen MM, Ivanisevic J, Giera M, Siuzdak G (2019) Identification of bioactive metabolites using activity metabolomics. *Nature Reviews Molecular Cell Biology* 20:6 20:353–367. <https://doi.org/10.1038/s41580-019-0108-4>
- Roberts LD, Souza AL, Gerszten RE, Clish CB (2012a) Targeted Metabolomics. *Curr Protoc Mol Biol* CHAPTER:Unit30.2. <https://doi.org/10.1002/0471142727.MB3002S98>
- Roberts LD, Souza AL, Gerszten RE, Clish CB (2012b) Targeted Metabolomics. *Curr Protoc Mol Biol* CHAPTER:Unit30.2. <https://doi.org/10.1002/0471142727.MB3002S98>
- Salem MA, De Souza LP, Serag A, et al (2020) Metabolomics in the context of plant natural products research: From sample preparation to metabolite analysis. *Metabolites* 10
- Schmid R, Heuckeroth S, Korf A, et al (2023) Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nature Biotechnology* 2023 41:4 41:447–449. <https://doi.org/10.1038/s41587-023-01690-2>
- Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA (2016) Untargeted metabolomics strategies – Challenges and Emerging Directions. *J Am Soc Mass Spectrom* 27:1897. <https://doi.org/10.1007/S13361-016-1469-Y>
- Souza AL, Patti GJ (2021) A Protocol for Untargeted Metabolomic Analysis: From Sample Preparation to Data Processing. *Methods Mol Biol* 2276:357. https://doi.org/10.1007/978-1-0716-1266-8_27
- Sumner LW, Amberg A, Barrett D, et al (2007) Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* 3:211–221. <https://doi.org/10.1007/S11306-007-0082-2>
- Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G (2012) XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal Chem* 84:5035. <https://doi.org/10.1021/AC300698C>
- Want EJ, Masson P, Michopoulos F, et al (2013) Global metabolic profiling of animal and human tissues via UPLC-MS. *Nat Protoc* 8:17–32. <https://doi.org/10.1038/NPROT.2012.135>