

Video Article

Large Scale Non-targeted Metabolomic Profiling of Serum by Ultra Performance Liquid Chromatography-Mass Spectrometry (UPLC-MS)

Corey D. Broeckling¹, Adam L. Heuberger¹, Jessica E. Prenni¹

¹Proteomics and Metabolomics Facility, Colorado State University

Correspondence to: Jessica E. Prenni at Jessica.Prenni@colostate.edu

URL: <https://www.jove.com/video/50242>

DOI: [doi:10.3791/50242](https://doi.org/10.3791/50242)

Keywords: Chemistry, Issue 73, Biochemistry, Genetics, Molecular Biology, Physiology, Genomics, Proteins, Proteomics, Metabolomics, Metabolite Profiling, Non-targeted metabolite profiling, mass spectrometry, Ultra Performance Liquid Chromatography, UPLC-MS, serum, spectrometry

Date Published: 3/14/2013

Citation: Broeckling, C.D., Heuberger, A.L., Prenni, J.E. Large Scale Non-targeted Metabolomic Profiling of Serum by Ultra Performance Liquid Chromatography-Mass Spectrometry (UPLC-MS). *J. Vis. Exp.* (73), e50242, doi:10.3791/50242 (2013).

Abstract

Non-targeted metabolite profiling by ultra performance liquid chromatography coupled with mass spectrometry (UPLC-MS) is a powerful technique to investigate metabolism. The approach offers an unbiased and in-depth analysis that can enable the development of diagnostic tests, novel therapies, and further our understanding of disease processes. The inherent chemical diversity of the metabolome creates significant analytical challenges and there is no single experimental approach that can detect all metabolites. Additionally, the biological variation in individual metabolism and the dependence of metabolism on environmental factors necessitates large sample numbers to achieve the appropriate statistical power required for meaningful biological interpretation. To address these challenges, this tutorial outlines an analytical workflow for large scale non-targeted metabolite profiling of serum by UPLC-MS. The procedure includes guidelines for sample organization and preparation, data acquisition, quality control, and metabolite identification and will enable reliable acquisition of data for large experiments and provide a starting point for laboratories new to non-targeted metabolite profiling by UPLC-MS.

Video Link

The video component of this article can be found at <https://www.jove.com/video/50242/>

Introduction

The term "metabolomics" can encompass many things. For example, a metabolomics experiment can be performed using a variety of analytical platforms such as NMR and both gas and/or liquid chromatography coupled with mass spectrometry. Furthermore, metabolomics experiments can be performed in a targeted or non-targeted manner, or a combination of both. A targeted metabolomics experiment will involve directed analysis of a panel of molecules important to the biological question at hand (e.g. small molecules involved in the TCA cycle will allow for accurate quantitation of that pathway). In this situation, the biological hypothesis is dictating the choice of metabolites to be targeted in the analysis and the analytical steps are optimized for the detection of these molecules. Alternatively, a non-targeted metabolomics experiment is hypothesis generating. In this case, the experiment is performed in a broad and unbiased manner to enable detection of as many metabolites as possible. The results from a non-targeted experiment will drive the next step of the research (which in many cases may involve a targeted metabolomics workflow). It is also possible to combine the two approaches, in which case an experiment is performed in a non-targeted manner while concurrently a panel of known molecules are monitored within the data.

The tutorial presented here is focused specifically on non-targeted metabolite profiling of serum. As described above, the non-targeted approach provides an unbiased view of the detectable metabolites, can generate large amounts of information, and ultimately allow for novel discoveries. The use of this approach, specifically employing ultra performance liquid chromatography coupled with mass spectrometry (UPLC-MS), is becoming widespread^{1,2,3} and involves the following steps: (1) experimental design (2) sample collection (3) sample preparation (4) data acquisition by UPLC-MS (5) data pre-processing (peak detection, integration, alignment, and normalization) (6) statistical data analysis (both uni- and multivariate) (7) metabolite identification and (8) biological interpretation.

Currently, there are no established standard methods for UPLC-MS based non-targeted metabolite profiling and subsequent data pre-processing steps. This lack of standardization is due in part to one of the primary analytical challenges of metabolite profiling; the chemical diversity of the metabolome. Because of this diversity, it is impossible for a single extraction method or mass spectrometry acquisition method to provide comprehensive coverage of all metabolites in a single analysis. In concept, metabolite coverage can be maximized by using multiple extractions (e.g. aqueous, methanol, chloroform:methanol, etc.) coupled with various chromatographic conditions (e.g. reverse phase, HILIC, etc.) and various ionization modes (e.g. positive ion, negative ion, chemical ionization, etc.). Often, however, researchers do not have a pre-determined bias for a specific chemical class and thus the expense of performing multiple extractions and instrument acquisitions is not warranted, especially for large-scale experiments. Thus, the video tutorial presented here was designed to provide a general procedure for large scale non-targeted metabolite profiling of serum by UPLC-MS. It will enable new and established laboratories to perform these types of experiments and the building blocks upon which they can expand the approach for various sample types, specific chemical classes, or targeted analysis. Specifically, this

protocol will include the steps of: serum sample preparation, sample organization for large scale studies, UPLC-MS data acquisition, quality control (QC) procedures, and metabolite identification. Strategies for data pre-processing and statistical analysis are also presented.

The protocol will not focus on the steps of experimental design, sample collection, or biological data interpretation as it is outside the scope of this tutorial. However, many resources exist in the literature for these topics and the authors encourage researchers new to metabolomics to explore these thoroughly^{4, 5, 6, 7, 8, 9}. In particular, experimental design is extremely important and is critical to the success of a non-targeted metabolomics experiment. Factors such as appropriate biological replication and consistency in sample collection procedure (e.g. time on bench, storage temperature, storage time, freeze-thaws, etc.) must be considered to ensure a viable study and to facilitate appropriate biological interpretation of the data.

Protocol

1. Sample Organization

1. To create a plate map for sample preparation on a spreadsheet, in the first column enter a sample list in order of loading. In a second column enter the 96 well plate locations using correct nomenclature for your autosampler software.
2. Save one well in each plate for QC samples.
3. If your LC autosampler can handle two plates at a time, separate the sample list into batches of 190 and save this information in a second worksheet.
4. Within each of these batches, use the randomize function [note: "=rand()" in Excel] to randomize the injection order. Now copy and paste your injection order into the LC-MS sample queue software system. Randomization should occur among batches, within batches, and be unique across injection replicates.

2. Serum and Quality Control (QC) Sample Preparation Procedures

1. Prepare a 10 ml stock of QC sample containing minimum of four compounds of known mass and retention time that span the chromatographic elution of the experiment. (note: caffeine, reserpine, sulfadimethoxine, and terfenadine at a concentration of 2 µg/ml each).
2. Thaw the serum samples on ice and gently vortex for ~5 sec.
3. Pre-wet the tips of a 12-channel pipettor with methanol to help prevent tip-dripping during sample dispensing. Then add 370 µl of cold methanol to each well of a 96 well plate.
4. Transfer 100 µl of each sample to the corresponding plate location from the plate maps. Cover the plates and incubating at -80 °C for 30 min to precipitate the proteins.
5. Spin the plates in a 4 °C centrifuge at 3,200 x g for 30 min. Transfer 250 µl of supernatant to a new 96-well and repeat the centrifugation.
6. Now transfer 60 µl aliquots of the supernatant to three 96-well plates suitable for the UPLC or LC autosampler.
7. Add 100 µl of QC sample mix to the reserved well position on the plate.
8. Seal each plate with adhesive film. Solvent may evaporate quickly after the well is pierced, and therefore injections should only be performed once from each plate. Plates in the sample queue can be stored in the short-term at -20 °C.

3. UPLC-MS Data Acquisition

1. Set the autosampler to 4 °C and place first two plates into autosampler.
2. Set up the acquisition method using one of the two alternative reverse phase gradients. Gradient A is biased towards non-polar molecules while gradient B will provide improved coverage of moderately polar molecules.

1. Gradient A:

Time (min)	%A	%B	curve
0.0	100	0	
0.1	100	0	6
1.0	60	40	6
3.0	30	70	6
11.0	0	100	6
17.0	0	100	6
17.1	100	0	6
23.0	100	0	6

2. Gradient B

Time (min)	%A	%B	curve
0.0	100	0	
1	100	0	6
13	5	95	6
16	5	95	6

16.05	100	0	6
20	100	0	6

- Set mass spectrometer to collect data in positive mode scanning from 50-1,200 m/z. Additional specific instrument conditions will vary according to the system used. The following represents the conditions use on our Waters Xevo G2 Q-TOF mass spectrometer: Collision Energy = 6V; Scan Time =0.2 sec per scan; Capillary Voltage =2,200 V; Source Temperature =150 °C Desolvation Temperature =350 °C Desolvation Gas Flow (N2) = 800 L/hr).
- Inject the QC samples five times at the beginning of each 2-plate batch of samples. After conditioning the UPLC-MS system with the first two injections, monitor for peak area (<25% RSD), retention time (+/- 0.05 min), and mass accuracy (+/- 3 ppm) of the third through fifth injections. These parameters may need to be adjusted to fit the specifications of the instrument being used for data acquisition.
- If QC samples pass, acquire data on serum samples for those two plates. If QC samples fail, sample data is not collected until the issue is resolved and the QC samples pass. In our laboratory we have found that these parameters are consistently stable over three days of injections (i.e. the injection of 2 96 well plates). If your system is found to be unstable over this time period, more frequent QC injections may be appropriate. The QC sample is used not to correct poor quality data, but to prevent poor quality data from ever being collected.

4. Peak Detection, Integration, Alignment, and Normalization

There are a variety of options to perform these steps including freeware and vendor specific tools. This step has been skipped in the video tutorial but our approach is described below as an example.

For each sample, the analytical data generated from this type of experiment is a profile of ions as described by a retention time, m/z value, and a spectral intensity. A non-targeted metabolite profiling experiment will be comprised of many samples and thus peak detection, retention time alignment, and normalization must be performed to enable subsequent statistical analysis of the dataset.

- Export raw data into .cdf format using the Waters DataBridge software. Perform peak detection and retention time alignment using the XCMS software package¹⁰ (freeware available at <http://metlin.scripps.edu/xcms/>) and is executed in the R statistical environment¹¹). The website contains documentation on how to download and run the software.
- Use the "matchedfilter" method within XCMS with the following parameters: full width at half max = 8 sec, S/N ration = 3, maximum of 100 peaks for each extracted ion chromatogram. Perform feature grouping, retention time correction, regrouping, and fillPeaks steps. The output of the XCMS software is a data matrix with all detectable features listed as rows, each experimental sample as a column, and cells are populated with the peak area of each feature in each sample. Transposition of this dataset may be necessary for incorporation into downstream statistical workflows.
- Normalize the data matrix in R by dividing the peak area of each feature in a given sample by the sum of the extracted ion peak area (referred to as total ion current, or TIC for that sample (multiplied by a 100,000 to make the numbers more amenable for visual data interpretation).
- Remove outliers. Outliers in the data can result from either a bad injection of a good sample or a good injection of a bad sample.
 - Perform outlier detection in R using the outliers package with the following parameters: (1) TIC values and (2) PC scores with 99.9% confidence.
 - If outliers are present, the corresponding data files are removed and peak detection, alignment, and normalization procedures are performed again.
- Average the normalized peak areas by sample identifier (to average multiple injections). This must be done prior to performing statistical analysis (note injection replicates account for analytical variation, but cannot be used as true biological replicates).

5. Statistical Analysis

For metabolomics experiments, both multivariate and univariate statistical techniques are necessary for data interpretation. Two techniques that are commonly used include Principal Component Analysis and Analysis of Variance (ANOVA). There are multiple ways to approach statistical analysis of the data and both open source and commercial software tools are available. This step has been skipped in the video tutorial but our approach is described below as an example.

- Perform ANOVA by applying an aov function to the dataset in R.
- Account for false positives using a Benjamini-Hochberg adjustment (p.adjust function in R).
- Conduct PCA with the Pareto-scaled (unit variance can also be applied) and mean-centered dataset. Both PC scores and loadings are calculated using the pcaMethods package in R.
- Use ANOVA p-values, loading scores, and fold-change parameters to determine molecular features that best described variation associated with the experimental treatment.
- Manually interrogate each molecular feature of interest to determine if it corresponds to a molecular ion, adduct (e.g. sodium, potassium, or dimer), or neutral loss.

6. Metabolite Identification

The workflow presented here is very general and can be applied to results from any instrument platform. An alternative strategy is presented in the Discussion section.

- For all statistically significant features (as determined from analysis in Step 6) review the raw data and confirm that parent ions, sodium adducts, isotopes, and fragments follow similar trends across the dataset. For example, the ANOVA-derived p-value for a parent ion should be similar to its isotope. This indicates that both molecular features correspond to the same metabolite.
- Search accurate mass measurements and/or neutral losses of statistically significant molecular features (choose C12 isotope if multiple isotopes are significant) against in-house (if available) or publically available metabolite databases such as:
 - Human Metabolome Database (<http://www.hmdb.ca/>)
 - Metlin (http://metlin.scripps.edu/metabo_search_alt2.php)
 - Mass Bank (<http://www.massbank.jp/?lang=en>)
 - Lipid Maps (<http://www.lipidmaps.org/data/standards/index.html>)
 - National Institute of Standards and Technology MS Search (<http://chemdata.nist.gov/mass-spc/ms-search/>)
- Use software to predict a molecular formula from the accurate mass and isotopic distribution of your significant molecular features. In Waters software, this is performed using the Elemental Comp tool in the MassLynx package.
- Filter candidate metabolite identifications by: mass error (as suitable for your instrument platform), predicted molecular formula, biological relevance in the trends observed with respect to experimental design, and retention time (e.g. lipids and other non-polar compounds will have longer retention times when employing reverse phase chromatography).
- Perform subsequent experiments to acquire the MS/MS fragmentation for statistically significant molecular features. Whenever possible, metabolite identification should be based on matching experimental fragmentation peaks and retention time to that of an authentic standard acquired on the same instrument. Alternatively, experimental fragmentation can be matched to a spectral database.
- Identification confidence should be assigned to all reported metabolite identifications based on metabolomics standards initiative recommendations¹².
 - Level 1: confident molecular identification based on orthogonal analytical parameters (accurate mass, retention time, and MS/MS fragmentation) relative to an authentic compound.
 - Level 2 refers to a putative identification base on physicochemical properties and/or spectral similarity with literature or spectral libraries.
 - Level 3 refers to the putative identification of a compound class based on physicochemical properties or spectral similarity.
 - Level 4 refers to an unknown compound.

Representative Results

The basic analytical steps of a non-targeted metabolite profiling experiment by UPLC-MS are outlined in **Figure 1**. The raw data for each sample can be visualized as a base peak chromatogram. **Figure 2** shows an example base peak chromatogram of a serum sample analyzed by gradient option (a) in the tutorial. Following statistical analysis as described above, metabolite identification is attempted for all statistically significant molecular features. Confident identification (level 1) requires matching of chromatographic retention time, accurate mass, and fragmentation pattern, acquired on the same analytical instrumentation, between the experimental molecular feature and an authentic standard of the putative metabolite. **Figure 3** shows an example of a level 1 metabolite identification of caffeine in human serum.

The protocol presented here is focused on sample organization and preparation, UPLC-MS data acquisition and quality control, and analytical metabolite identification. The steps of peak detection, alignment, and normalization and statistical analysis are critically important to the output of a non-targeted metabolite profiling experiment, however, we chose to skip them in the video tutorial as there are multiple options and tools available for performing these tasks. Many of the tools are vendor specific, although freeware options are available. An example of our approach to these steps is presented in the protocol above.

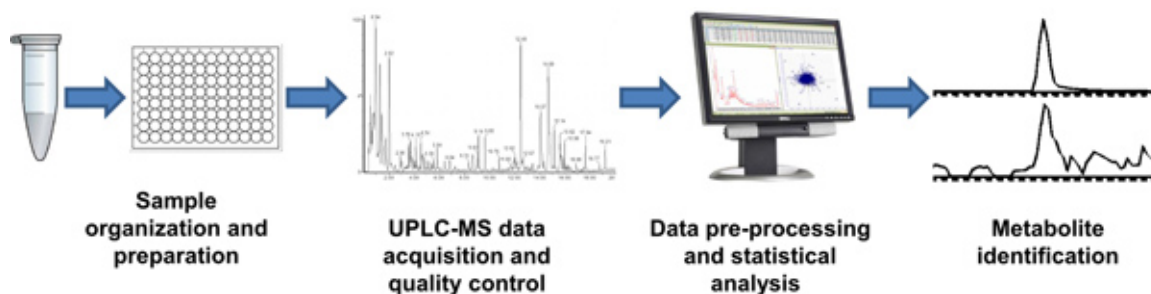


Figure 1. Analytical workflow for a non-targeted metabolite profiling experiment. Steps include: sample organization and preparation, UPLC-MS data acquisition, quality control, data pre-processing and statistical analysis, and metabolite identification.

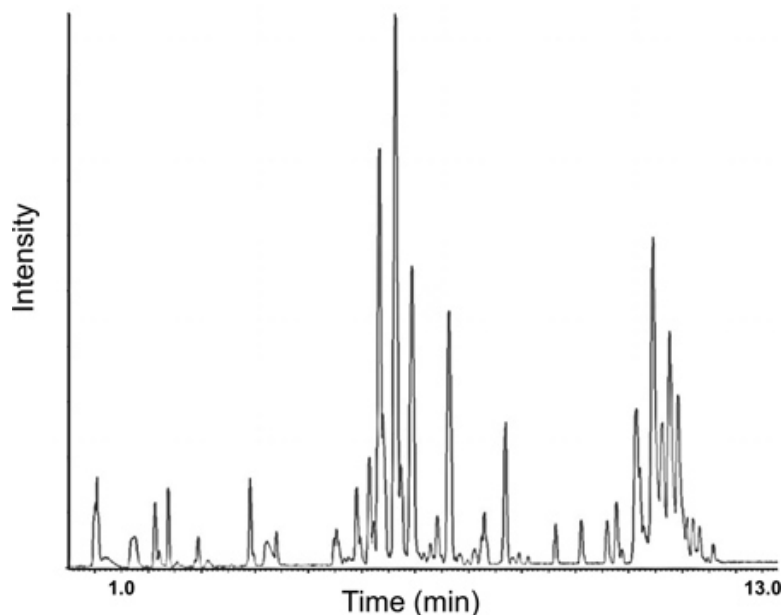


Figure 2. Example UPLC-MS base peak chromatogram of human serum using gradient option a as described in the protocol. Peaks represent the base peak (largest) ion signal at each point in time.

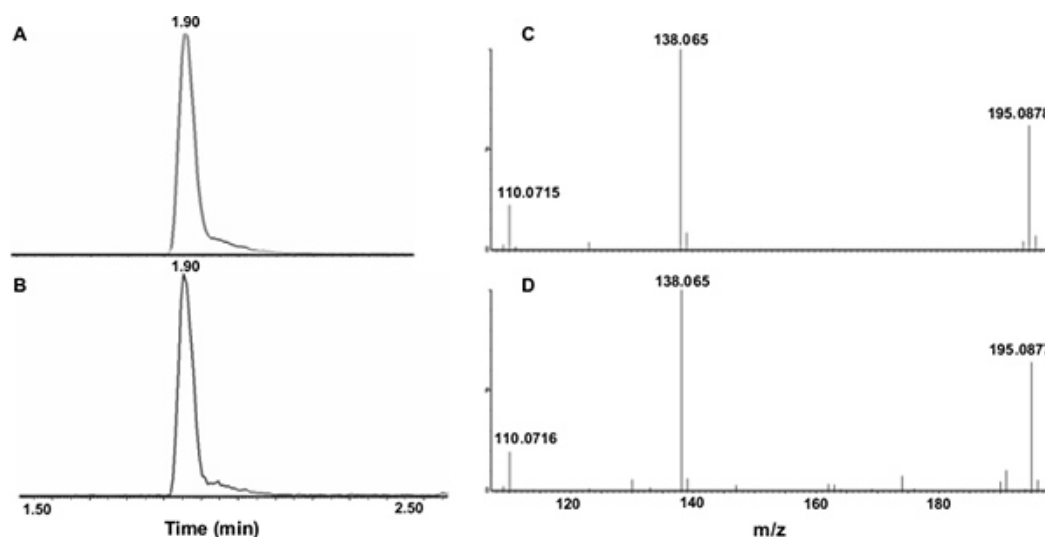


Figure 3. Level 1 metabolite identification of caffeine in human serum based on chromatographic retention time, accurate mass, and fragmentation pattern. Extracted ion chromatogram of caffeine (parent ion 195.0878) from an authentic standard (A) and an experimental serum sample (B). Fragmentation spectra of caffeine from an authentic standard (C) and an experimental serum sample (D). [Click here to view larger figure.](#)

Discussion

This tutorial is meant to serve as a starting point for conducting large scale non-targeted metabolite profiling by UPLC-MS. The workflow is focused on metabolites that can be extracted with an aqueous methanol solvent, retained on a C8 or C18 UPLC column, and detected as positive ions. In the situation where there is not a pre-determined bias towards a specific metabolite class and a hypothesis generating global profile is desired, this protocol is valuable as it will result in the detection of a large percentage of serum metabolites from a variety of chemical classes (*i.e.* moderately polar to non-polar compounds). If budget and time allow, metabolite coverage can be easily expanded by the incorporation of various extraction methods, chromatographic separations, and ionization modes^{13,14,15}.

Data stability over time is an important analytical consideration for large scale experiments in which data may be collected over many weeks or months. Given the unbiased and holistic nature of the approach, the incorporation of an internal standard (or even a panel of internal standards) for the purpose of data normalization is not appropriate and may in fact skew the data, as the assumption that an internal standard is representative of the entire metabolome is invalid. Additionally, given the natural variation inherent to serum, the use of an endogenous internal control should be approached with caution. Proper sample preparation and randomization combined with the normalization strategy described

below can ensure normality across the dataset. In addition, the use of easily quantifiable and traceable QC metrics during data acquisition ensures analytical stability (e.g. retention time, sensitivity, and mass accuracy).

While the unbiased nature of non-targeted metabolite profiling allows for the generation of large amounts of data and the potential for novel discoveries, it is marked by a substantial limitation in the challenge of metabolite identification. Small molecule spectral libraries are well developed for gas chromatography (GC)-MS applications utilizing electron impact ionization, but have been slow to develop for LC-MS small molecule applications. Unlike peptides, which fragment predictably, metabolites are structurally diverse, making their behavior in the mass spectrometer less predictable, and this has greatly inhibited the generation of theoretical spectral databases (as used in proteomic applications). Recently, some progress has been made in the generation of small molecule LC-MS spectral libraries through the analysis of authentic standards^{10, 16}, and these libraries continue to grow. However, given the complexity of the metabolome, variation of the metabolome between biological sources, and the variation in molecule fragmentation with MS conditions, the generation of comprehensive spectral libraries through the analysis of authentic standards is a slow and tedious process. Thus, the process of molecular identification remains a significant bottleneck in non-targeted LC-MS based metabolite profiling studies^{1, 17}.

An alternative approach to the metabolite identification workflow described in the protocol is to incorporate *indiscriminant* MS/MS acquisition (idMS/MS) as part of the initial metabolite profiling workflow. In this approach, MS and MS/MS acquisitions are performed simultaneously in parallel alternating scans, and the MS/MS acquisition is an indiscriminant process, *i.e.* no precursor ion isolation is employed. Thus, using an idMS/MS workflow, both the MS and MS/MS signals are acquired concurrently for every sample, eliminating the need for subsequent targeted MS/MS acquisition. The challenge of this approach is the assignment of precursor-product ion relationships. A workflow for this analysis was recently described by our group and can be implemented in R¹⁸. The output of this workflow is reconstructed MS/MS spectra for each detectable metabolite that can be used to compare against an authentic standard and/or searched against public or in-house spectral databases. The advantage of this approach is that it does not require subsequent mass spectrometry experimentation to facilitate the metabolite identification process. Mass spectrometers from Waters are capable of performing this type of data acquisition in what is called MS^E mode. Alternative approaches are possible for tandem MS/MS instruments from other vendors. For example, two injections of each sample can be performed, one using low collision energy and a second at high collision energy. If using a non-tandem instrument (e.g. ESI-TOF), the data can be collected using two injections at two different cone voltage settings, where higher cone voltage induces increased in-source fragmentation. Several instrument platforms do allow scan to scan parameter switching, which often includes both collision energy as well as cone voltage settings¹⁸.

The workflow for metabolite identification presented in the protocol, as well as the alternative strategy described above, are both focused on the assignment of molecular features to known metabolites. Thus, there still remains the significant challenge of unknown metabolite identification. At this time, the key strategy for structural elucidation of unknown metabolites remains focused on isolation and concentration of the metabolites with subsequent analysis by NMR¹⁷. This is not practically feasible for a large number of metabolites and thus the molecular identification of unknowns represents a major analytical hurdle that will continue to be a priority for the field. An example of current efforts in this area is the development of in-silico spectral libraries of small molecule metabolites. Kangas *et al.* have recently published a tool for the generation of in-silico spectra of lipid compounds¹⁹. Work such as this has the potential to significantly expand coverage in MS/MS spectral libraries to all known biological compounds, with the caveat that some compounds classes will be more difficult to predict than the lipids.

Disclosures

The authors declare that they have no competing financial interests.

Acknowledgements

The presented tutorial was performed and developed within the Proteomics and Metabolomics Facility at Colorado State University which is partially funded by the CSU Research Administration Resources for Scholarly Projects.

References

1. Theodoridis, G., Gika, H.G., & Wilson, I.D. Mass Spectrometry-Based Holistic Analytical Approaches for Metabolite Profiling in Systems Biology Studies. *Mass Spectrom. Rev.* **30**, 884-906, doi:10.1002/Mas.20306 (2011).
2. Zelena, E., *et al.* Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. *Analytical Chemistry*. **81**, 1357-1364, doi:10.1021/ac8019366 (2009).
3. Michopoulos, F., Lai, L., Gika, H., Theodoridis, G., & Wilson, I. UPLC-MS-based analysis of human plasma for metabonomics using solvent precipitation or solid phase extraction. *Journal of Proteome Research*. **8**, 2114-2121 (2009).
4. Kamburov, A., Cavill, R., Ebbels, T.M., Herwig, R., & Keun, H.C. Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA. *Bioinformatics*. **27**, 2917-2918, doi:10.1093/bioinformatics/btr499 (2011).
5. Xia, J. & Wishart, D.S. MetPA: a web-based metabolomics tool for pathway analysis and visualization. *Bioinformatics*. **26**, 2342-2344, doi:10.1093/bioinformatics/btq418 (2010).
6. Johnson, C.H. & Gonzalez, F.J. Challenges and opportunities of metabolomics. *J. Cell. Physiol.* **227**, 2975-2981, doi:10.1002/jcp.24002 (2012).
7. Issaq, H.J., Waybright, T.J., & Veenstra, T.D. Cancer biomarker discovery: Opportunities and pitfalls in analytical methods. *Electrophoresis*. **32**, 967-975, doi:10.1002/elps.201000588 (2011).
8. Zhou, B., Xiao, J.F., Tuli, L., & Ransom, H.W. LC-MS-based metabolomics. *Mol. Biosyst.* **8**, 470-481, doi:10.1039/c1mb05350g (2012).
9. Weckwerth, W. & Morgenthal, K. Metabolomics: from pattern recognition to biological interpretation. *Drug Discov. Today*. **10**, 1551-1558, doi:10.1016/S1359-6446(05)03609-3 (2005).
10. Smith, C.A., *et al.* METLIN - A metabolite mass spectral database. *Ther. Drug Monit.* **27**, 747-751 (2005).

11. Culpepper, S.A. & Aguinis, H. R is for Revolution: A Cutting-Edge, Free, Open Source Statistical Package. *Organ Res. Methods*. **14**, 735-740, doi:10.1177/1094428109355485 (2011).
12. Sumner, L.W., *et al.* Proposed minimum reporting standards for chemical analysis. *Metabolomics*. **3**, 211-221, doi:10.1007/s11306-007-0082-2 (2007).
13. Vuckovic, D. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. *Anal. Bioanal. Chem.* **403**, 1523-1548, doi:10.1007/s00216-012-6039-y (2012).
14. Garcia-Canaveras, J.C., Donato, M.T., Castell, J.V., & Lahoz, A. A comprehensive untargeted metabonomic analysis of human steatotic liver tissue by RP and HILIC chromatography coupled to mass spectrometry reveals important metabolic alterations. *Journal of Proteome Research*. **10**, 4825-4834, doi:10.1021/pr200629p (2011).
15. Nordstrom, A., Want, E., Northen, T., Lehtio, J., & Siuzdak, G. Multiple ionization mass spectrometry strategy used to reveal the complexity of metabolomics. *Analytical Chemistry*. **80**, 421-429, doi:10.1021/ac701982e (2008).
16. Horai, H., *et al.* MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **45**, 703-714, doi:10.1002/Jms.1777 (2010).
17. Wishart, D.S. Advances in metabolite identification. *Bioanalysis*. **3**, 1769-1782, doi:10.4155/Bio.11.155 (2011).
18. Broeckling, C., Heuberger, A., Prince, J., Ingelsson, E., & Prenni, J. Assigning precursor-product ion relationships in indiscriminant MS/MS data from non-targeted metabolite profiling studies. *Metabolomics*. 1-11, doi:citeulike-article-id:10630185 (2012).
19. Kangas, L.J., *et al.* In silico identification software (ISIS): a machine learning approach to tandem mass spectral identification of lipids. *Bioinformatics*. **28**, 1705-1713, doi:10.1093/bioinformatics/bts194 (2012).