

Using MetaboAnalyst 5.0 Part II: Obtaining functional insights from peak list data

Zhiqiang Pang

McGill University

Guangyan Zhou

McGill University

Jessica Ewald

McGill University

Le Chang

McGill University

Orcun Hacariz

McGill University

Niladri Basu

McGill University

Jianguo Xia (✉ jeff.xia@mail.mcgill.ca)

McGill University

Method Article

Keywords: Global metabolomics, exposomics, HRMS, spectra processing, parameter optimization

Posted Date: June 17th, 2022

DOI: <https://doi.org/10.21203/rs.3.pex-1919/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Liquid-chromatography coupled with high-resolution mass spectrometry (LC-HRMS) has become a workhorse in global metabolomics studies with growing applications across biomedical and environmental sciences. However, outstanding bioinformatics challenges in terms of data processing, statistical analysis and functional interpretation remain critical barriers to the wider adoption of this technology. To help the user community overcome these barriers, we have made major updates to the well-established MetaboAnalyst platform (www.metaboanalyst.ca). This protocol is part two of four and provides stepwise instructions on how to use MetaboAnalyst to obtain functional insights from peak list data.

Introduction

Please see the “Using MetaboAnalyst 5.0 Part I: Optimizing parameters for LC-HRMS spectra processing” protocol for more general background information.

Peak annotation

A typical LC-HRMS spectrum of common biofluids (such as blood or urine samples) can often produce over 10,000 peaks. However, a critical point is that this number is not equivalent to the number of compounds detected. The correspondence between these peaks and the actual metabolites remains elusive. Many peaks are derived from the same compounds in the forms of adducts, isotopes, fragmentations, *etc.* (real or biological peaks), and other peaks may reflect background noise (artifacts or noise peaks)³⁸. Therefore, the first step in peak annotation aims to identify real peaks, and to clarify the relationships among them. Many empirical and statistical rules have been developed to address this problem, including CAMERA and CliqueMS^{39,40} which are two popular R packages. The next step is to assign putative compound IDs to those peaks. This is a challenging task as even with high resolution instruments, a single peak can potentially match multiple compounds⁴¹. It is important to keep in mind that, unlike targeted metabolomics, the primary goal of global metabolomics is to understand the overall patterns and to identify promising features to inform the design of follow-up studies for more targeted analysis. As discussed in the sections below, accurate peak annotation, although important, is not an absolute prerequisite for functional interpretation of global metabolomics data.

Functional interpretation of global metabolomics data

Identifying a list of compounds and performing enrichment analysis with pre-defined libraries of pathways or metabolite sets⁴² is well-established for targeted metabolomics data^{43,44}. Please refer to many excellent tutorials as well as our previous published protocols^{28,32}. For global metabolomics, since

the mapping of peaks to compounds is often inaccurate, researchers often wonder if we can still identify meaningful functional changes by performing enrichment analysis based on those fuzzy annotations. This question has been answered by the *mummichog* algorithm⁴⁵, which has clearly demonstrated that the collective behavior (i.e. changes across multiple compounds involved in a pathway) are robust to random errors introduced during individual compound assignments. In other word, enrichment analysis of putatively annotated peaks is a valid approach to identify pathway or network-level changes, as long as the annotation errors are random. In general, HRMS peaks are preferred as they provide better coverage of pathways with reduced error rates in putative annotation, as compared to using peaks from low-resolution MS instruments⁴¹.

Reagents

Equipment

Computer requirements

- Browser requirements: MetaboAnalyst 5.0 runs on all modern web browsers. For the best results, we recommend Google Chrome 92+, Firefox 92+, Safari 12+, and Microsoft Edge v93+. JavaScript must be enabled in your browser.
- Internet connection requirements: A fast connection is highly recommended.
- Hardware requirements: >4 GB of RAM and a screen resolution of at least 1200 x 800 is preferred. At least 8 GB hard drive space is needed to store the raw spectra files.

Data files

- *Input Files.* MetaboAnalyst accepts multiple data formats for different modules. For raw spectral processing, the supported data formats include mzML, mzXML, mzData and NetCDF, while other modules usually require comma-separated values (.csv) or tab delimited (.txt) file format.
- *Example Datasets.* Multiple built-in example datasets are offered in each MetaboAnalyst module. Users can simply choose an example dataset and click the 'Submit' button to start exploring the tool. Six datasets are used in this protocol:
 - 1) A raw spectra data (*malaria_raw.zip*) consisting of 12 plasma samples from healthy and malaria semi-immune subjects along with 3 Quality Controls (QC)⁵³.

- 2) A feature table (*malaria_feature_table.csv*) generated from processing the above raw spectra for functional analysis.
- 3) A gene list (*integ_genes_1.txt*) and a compound list (*integ_cmpds.txt*) to demonstrate the integration of targeted metabolomics data with transcriptomics data.
- 4) A gene list (*integ_genes_2.txt*) and a peak list (*integ_peaks.txt*) to demonstrate the integration of global metabolomics data with transcriptomics data.
- 5) Three peak intensity tables (*A1_pos.csv*, *B1_pos.csv*, *C1_pos.csv*) from global metabolomics studies of serum samples from COVID-19 patients to demonstrate the integration of multiple global metabolomics datasets⁵⁴.
- 6) A feature table (*TCE_feature_table.csv*) and its associated metadata table (*TCE_metadata.csv*) from an exposomics study on metabolic changes associated with occupational exposure to trichloroethylene (TCE) ⁵⁵.

Equipment Setup

(Optional) Download the example datasets. Go to the MetaboAnalyst home page and click 'Data Formats' link from the left panel. Users can download all the example datasets used in the protocol. These datasets are also available as built-in examples in each corresponding module. Users can directly select those datasets and follow the protocols without downloading and uploading them.

Procedure

Stage 2: Functional analysis of LC-HRMS peaks (Timing ~ 20min)

19. *Starting up.* Go to the MetaboAnalyst homepage and select the 'Click here to start' button to enter the modules overview page. Locate the 'Functional Analysis' button (on the 2nd tier). Click to enter the module.

20. *Data uploading.* Users can upload two types of data: a peak list or a complete peak intensity table. In this protocol, we will use the 'A peak intensity table' tab with the following parameters: Ion Mode: 'Positive Mode'; Mass Tolerance (ppm): 4.3 (based on the optimized value from the Raw Spectra Processing module); Retention time: 'Yes - Seconds'. Next, locate the data #2 (*malaria_feature_table.csv*) using the

'Choose File' option and then click 'Submit' button. Alternatively, users can simply select the third example, 'Malaria', which is the same data and click 'Submit' at the bottom of the page.

Critical Step: *Users should provide a peak list or a peak intensity table containing the complete features detected from the raw spectra processing step, instead of just significant features. Functional enrichment analysis of global metabolomics data requires permutation tests to calculate empirical p values to measure the significance of the pathways. Please refer to the **Figure 1** for more details on functional enrichment analysis.*

21. *Data integrity check.* The data integrity check summarizes the key information of this data. Note that features containing all zero values or a single non-zero values are excluded automatically. This is not uncommon for global metabolomics data with a small number of samples. Click 'Proceed' button.

22. *Data filtering.* The data filtering step is used to remove the non-informative variables. Select 'None' and click 'Submit' button or click the 'Proceed' button to keep all features for this data.

23. *Data normalization.* MetaboAnalyst offers comprehensive options for data normalization. Different purposes and implementation principals are introduced in the previous protocol of MetaboAnalyst³². In this case, we select 'Log transformation' and click 'Normalize' to perform normalization. Then, click 'View Result' to check and compare the normalization results. Click 'Proceed' button to continue.

24. *Group removal.* Functional analysis is mainly designed to detect functional changes between two experimental groups. In our case, there are three groups (naive, semi-immune, and QC) in the data table, the QC group should be excluded. To do this, click the 'Data Editor' from the navigation tree. Click 'Edit Groups' tab. Select 'QC' from the 'Available' box and move it into the 'Exclude' box by clicking the left arrow. Click 'Submit' to finish editing.

Critical Step: *Users need to do the normalization again after editing data.*

25. *Parameter setting.* The page is organized into two sections - algorithm parameters and library selection. Two algorithms are available for functional analysis. The default option is *mummichog*. Users could manually specify the p-value cut-off and the version of the algorithm. The default p-value cut-off is to keep the top 10% peaks and using *mummichog* 2.0. The results can be viewed in a scatter plot or heatmap (for peak intensity tables only). The library selection is based on the organism under study. The

parameter at the bottom of the page is to exclude small pathways, because very few compounds usually do not provide sufficient information to allow reliable identification of the underlying pathways. Here we keep the default options and click 'Submit'. See **Figure 1** for more details on the concepts of functional enrichment analysis.

Troubleshooting.

26. *Pathway activity prediction.* The results from the *mummichog* algorithm are displayed in the 'Mummichog Pathway Activity Profile' page. The top half page is an interactive scatter plot with 'Enrichment Factor' as x-axis and '-log₁₀(p)' as y-axis (**Figure 2A**). Hovering a mouse on a node shows the pathway name and the corresponding statistics. Clicking on a node displays the matched compounds (**Figure 2B**). Mummichog has identified several fatty acid metabolism related pathways which are consistent with the literature on the roles of fatty acid metabolism in the pathogenesis of malaria⁵⁷⁻⁵⁹. User can download the 'Pathway Hits' and 'Compound Hits' file directly. The pathway results are displayed at the bottom of this page. Users can click the 'View' link from the last column of the table to view all potentially matched compounds (**Figure 2B**).

27. *(Optional) GSEA analysis.* Use the navigation tree to go back to the 'Set Parameters' page. Uncheck the mummichog option and select GSEA algorithm, click 'Submit' for further exploration. Different from *mummichog* results, GSEA returns a volcano plot with normalized enrichment score (NES) as x-axis and '-log(p)' as y-axis for scatter plot visualization (**Figure 2C**).

28. *(Optional) Joint analysis with mummichog and GSEA.* Users can combine results from both algorithms by make both options selected from the 'Set parameters' page. The results would be a joint scatter plot with GSEA plotted on the x-axis and *mummichog* on the y-axis (**Figure 2D**). The p values from both algorithms are merged with Fisher's method.

29. *Heatmap-based metabolic pattern analysis.* Click 'Set parameters' from the navigation tree to go back to the parameter configuration page and select 'Heatmaps' as the visual analytics option and click 'Submit'. As shown in **Figure 3**, the page is divided into three vertical sections – Overview, Focus View and Enrichment Analysis. Users can use mouse to drag-and-select peaks of interest from the Overview to the Focus View and perform enrichment analysis based on the selected peaks. By default, the Overview displays all peaks ranked by their p-values and the Focus View shows the significant peaks. Users should first perform cluster analysis using different cluster algorithms (from 'Cluster peaks' menu) and visually select any regions of interest for enrichment analysis (**Figure 3**). Watch the second part video demo of

No. 2 (https://www.metaboanalyst.ca/resources/data/7_Video_Tutorial.pdf) to learn how to manually select multiple regions of interest to compile a composite pattern for enrichment analysis. All patterns and results can be downloaded easily from the 'Download' menu.

Critical Step: *The interactive heatmaps run on users' browser rather than on the MetaboAnalyst server. Please directly download the images using the "Download" menu at the current page, as the final Download page will not be able to capture these heatmap images.*

30. *(Optional) Network visualization.* From the 'Pathway Activity Profile' page, click the 'Network Explorer' button. Users can intuitively explore the enriched pathways within the KEGG global metabolic network.

31. *Result downloading.* Click 'Download' node from the navigation tree at the left panel. At the download page, generate the analysis report, download all results and exit.

Troubleshooting

Time Taken

Anticipated Results

The top significantly perturbed pathways are related with fatty acid metabolism based on the *mummichog* algorithm, while the GSEA algorithm also reports some amino acid metabolism related pathways. The combined approach of GSEA and *mummichog* confirm two significant pathways (Fatty acid oxidation and Carnitine shuttle). All intermediate normalized data, matching compound list and predicted pathways are available as different tables from the result download.

References

- 1 Alseekh, S. *et al.* Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nat Methods* **18**, 747-756, doi:10.1038/s41592-021-01197-1 (2021).
- 2 Alseekh, S. & Fernie, A. R. Metabolomics 20 years on: what have we learned and what hurdles remain? *Plant J* **94**, 933-942, doi:10.1111/tpj.13950 (2018).
- 3 Doerr, A. Global metabolomics. *Nature Methods* **14**, 32-32, doi:10.1038/nmeth.4112 (2017).

- 4 Cajka, T. & Fiehn, O. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Anal Chem* **88**, 524-545, doi:10.1021/acs.analchem.5b04491 (2016).
- 5 Vermeulen, R., Schymanski, E. L., Barabasi, A. L. & Miller, G. W. The exposome and health: Where chemistry meets biology. *Science* **367**, 392-396, doi:10.1126/science.aay3164 (2020).
- 6 Beger, R. D. *et al.* Metabolomics enables precision medicine: "A White Paper, Community Perspective". *Metabolomics* **12**, 149, doi:10.1007/s11306-016-1094-6 (2016).
- 7 Want, E. J. *et al.* Global metabolic profiling of animal and human tissues via UPLC-MS. *Nat Protoc* **8**, 17-32, doi:10.1038/nprot.2012.135 (2013).
- 8 Zhou, G. *et al.* NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res* **47**, W234-W241, doi:10.1093/nar/gkz240 (2019).
- 9 Chong, J., Liu, P., Zhou, G. & Xia, J. Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. *Nat Protoc* **15**, 799-821, doi:10.1038/s41596-019-0264-1 (2020).
- 10 Pang, Z. *et al.* MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Research* **49**, W388-W396, doi:10.1093/nar/gkab382 (2021).
- 11 Xia, J., Psychogios, N., Young, N. & Wishart, D. S. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res* **37**, W652-660, doi:10.1093/nar/gkp356 (2009).
- 12 Xia, J., Mandal, R., Snelnikov, I. V., Broadhurst, D. & Wishart, D. S. MetaboAnalyst 2.0—a comprehensive server for metabolomic data analysis. *Nucleic Acids Res* **40**, W127-133, doi:10.1093/nar/gks374 (2012).
- 13 Xia, J., Snelnikov, I. V., Han, B. & Wishart, D. S. MetaboAnalyst 3.0—making metabolomics more meaningful. *Nucleic Acids Res* **43**, W251-257, doi:10.1093/nar/gkv380 (2015).
- 14 Chong, J. *et al.* MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res* **46**, W486-W494, doi:10.1093/nar/gky310 (2018).
- 15 Stanstrup, J. *et al.* The metaRbolomics Toolbox in Bioconductor and beyond. *Metabolites* **9**, doi:10.3390/metabo9100200 (2019).
- 16 Gardinassi, L. G., Xia, J., Safo, S. E. & Li, S. Bioinformatics Tools for the Interpretation of Metabolomics Data. *Current Pharmacology Reports* **3**, 374-383, doi:10.1007/s40495-017-0107-0 (2017).
- 17 Tautenhahn, R., Patti, G. J., Rinehart, D. & Siuzdak, G. XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal Chem* **84**, 5035-5039, doi:10.1021/ac300698c (2012).

- 18 Giacomoni, F. *et al.* Workflow4Metabolomics: a collaborative research infrastructure for computational metabolomics. *Bioinformatics* **31**, 1493-1495, doi:10.1093/bioinformatics/btu813 (2015).
- 19 Yang, Q. *et al.* NOREVA: enhanced normalization and evaluation of time-course and multi-class metabolomic data. *Nucleic Acids Res* **48**, W436-W448, doi:10.1093/nar/gkaa258 (2020).
- 20 Chambers, M. C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **30**, 918-920, doi:10.1038/nbt.2377 (2012).
- 21 Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395, doi:10.1186/1471-2105-11-395 (2010).
- 22 Du, X. X., Smirnov, A., Pluskal, T., Jia, W. & Sumner, S. Metabolomics Data Preprocessing Using ADAP and MZmine 2. *Computational Methods and Data Analysis for Metabolomics* **2104**, 25-48, doi:Book_Doi 10.1007/978-1-0716-0239-3 (2020).
- 23 Tsugawa, H. *et al.* A lipidome atlas in MS-DIAL 4. *Nat Biotechnol* **38**, 1159-1163, doi:10.1038/s41587-020-0531-2 (2020).
- 24 Tsugawa, H. *et al.* Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. *Anal Chem* **88**, 7946-7958, doi:10.1021/acs.analchem.6b00770 (2016).
- 25 Rost, H. L. *et al.* OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat Methods* **13**, 741-748, doi:10.1038/nmeth.3959 (2016).
- 26 Kenar, E. *et al.* Automated label-free quantification of metabolites from liquid chromatography-mass spectrometry data. *Mol Cell Proteomics* **13**, 348-359, doi:10.1074/mcp.M113.031278 (2014).
- 27 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).
- 28 Xia, J. & Wishart, D. S. Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nat Protoc* **6**, 743-760, doi:10.1038/nprot.2011.319 (2011).
- 29 Xia, J. & Wishart, D. S. Metabolomic data processing, analysis, and interpretation using MetaboAnalyst. *Curr Protoc Bioinformatics* **Chapter 14**, Unit 14 10, doi:10.1002/0471250953.bi1410s34 (2011).
- 30 Xia, J. & Wishart, D. S. Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. *Curr Protoc Bioinformatics* **55**, 14 10 11-14 10 91, doi:10.1002/cpbi.11 (2016).

- 31 Xia, J., Broadhurst, D. I., Wilson, M. & Wishart, D. S. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics* **9**, 280-299, doi:10.1007/s11306-012-0482-9 (2013).
- 32 Chong, J., Wishart, D. S. & Xia, J. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. *Curr Protoc Bioinformatics* **68**, e86, doi:10.1002/cpbi.86 (2019).
- 33 Chong, J. & Xia, J. Using MetaboAnalyst 4.0 for Metabolomics Data Analysis, Interpretation, and Integration with Other Omics Data. *Methods Mol Biol* **2104**, 337-360, doi:10.1007/978-1-0716-0239-3_17 (2020).
- 34 Tautenhahn, R., Bottcher, C. & Neumann, S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics* **9**, 504, doi:10.1186/1471-2105-9-504 (2008).
- 35 Yu, T., Park, Y., Johnson, J. M. & Jones, D. P. apLCMS-adaptive processing of high-resolution LC/MS data. *Bioinformatics* **25**, 1930-1936, doi:10.1093/bioinformatics/btp291 (2009).
- 36 Alboniga, O. E., Gonzalez, O., Alonso, R. M., Xu, Y. & Goodacre, R. Optimization of XCMS parameters for LC-MS metabolomics: an assessment of automated versus manual tuning and its effect on the final results. *Metabolomics* **16**, 14, doi:10.1007/s11306-020-1636-9 (2020).
- 37 Pang, Z., Chong, J., Li, S. & Xia, J. MetaboAnalystR 3.0: Toward an Optimized Workflow for Global Metabolomics. *Metabolites* **10**, doi:10.3390/metabo10050186 (2020).
- 38 Sindelar, M. & Patti, G. J. Chemical Discovery in the Era of Metabolomics. *J Am Chem Soc* **142**, 9097-9105, doi:10.1021/jacs.9b13198 (2020).
- 39 Senan, O. *et al.* CliqueMS: a computational tool for annotating in-source metabolite ions from LC-MS untargeted metabolomics data based on a coelution similarity network. *Bioinformatics* **35**, 4089-4097, doi:10.1093/bioinformatics/btz207 (2019).
- 40 Kuhl, C., Tautenhahn, R., Bottcher, C., Larson, T. R. & Neumann, S. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal Chem* **84**, 283-289, doi:10.1021/ac202450g (2012).
- 41 Kind, T. & Fiehn, O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *Bmc Bioinformatics* **8**, doi:10.1186/1471-2105-8-105 (2007).
- 42 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
- 43 Xia, J. & Wishart, D. S. MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Res* **38**, W71-77, doi:10.1093/nar/gkq329 (2010).

- 44 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).
- 45 Li, S. *et al.* Predicting network activity from high throughput metabolomics. *PLoS Comput Biol* **9**, e1003123, doi:10.1371/journal.pcbi.1003123 (2013).
- 46 Xia, J. *et al.* INMEX—a web-based tool for integrative meta-analysis of expression data. *Nucleic Acids Res* **41**, W63-70, doi:10.1093/nar/gkt338 (2013).
- 47 Zhou, G. & Xia, J. OmicsNet: a web-based tool for creation and visual analysis of biological networks in 3D space. *Nucleic Acids Res* **46**, W514-W522, doi:10.1093/nar/gky510 (2018).
- 48 Schriml, L. M. *et al.* COVID-19 pandemic reveals the peril of ignoring metadata standards. *Sci Data* **7**, 188, doi:10.1038/s41597-020-0524-5 (2020).
- 49 Nygaard, V., Rodland, E. A. & Hovig, E. Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. *Biostatistics* **17**, 29-39, doi:10.1093/biostatistics/kxv027 (2016).
- 50 Kahan, B. C., Jairath, V., Dore, C. J. & Morris, T. P. The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. *Trials* **15**, 139, doi:10.1186/1745-6215-15-139 (2014).
- 51 Chong, J. & Xia, J. MetaboAnalystR: an R package for flexible and reproducible analysis of metabolomics data. *Bioinformatics* **34**, 4313-4314, doi:10.1093/bioinformatics/bty528 (2018).
- 52 Chong, J., Yamamoto, M. & Xia, J. MetaboAnalystR 2.0: From Raw Spectra to Biological Insights. *Metabolites* **9**, doi:10.3390/metabo9030057 (2019).
- 53 Gardinassi, L. G. *et al.* Integrative metabolomics and transcriptomics signatures of clinical tolerance to Plasmodium vivax reveal activation of innate cell immunity and T cell signaling. *Redox Biol* **17**, 158-170, doi:10.1016/j.redox.2018.04.011 (2018).
- 54 Pang, Z., Zhou, G., Chong, J. & Xia, J. Comprehensive Meta-Analysis of COVID-19 Global Metabolomics Datasets. *Metabolites* **11**, doi:10.3390/metabo11010044 (2021).
- 55 Walker, D. I. *et al.* High-resolution metabolomics of occupational exposure to trichloroethylene. *Int J Epidemiol* **45**, 1517-1527, doi:10.1093/ije/dyw218 (2016).
- 56 Conley, C. J. *et al.* Massifquant: open-source Kalman filter-based XC-MS isotope trace feature detection. *Bioinformatics* **30**, 2636-2643, doi:10.1093/bioinformatics/btu359 (2014).
- 57 Vaughan, A. M. *et al.* Type II fatty acid synthesis is essential only for malaria parasite late liver stage development. *Cell Microbiol* **11**, 506-520, doi:10.1111/j.1462-5822.2008.01270.x (2009).

- 58 Shears, M. J., Botte, C. Y. & McFadden, G. I. Fatty acid metabolism in the Plasmodium apicoplast: Drugs, doubts and knockouts. *Mol Biochem Parasitol* **199**, 34-50, doi:10.1016/j.molbiopara.2015.03.004 (2015).
- 59 Cumnock, K. *et al.* Host Energy Source Is Important for Disease Tolerance to Malaria. *Curr Biol* **28**, 1635-1642 e1633, doi:10.1016/j.cub.2018.04.009 (2018).
- 60 Libiseller, G. *et al.* IPO: a tool for automated optimization of XCMS parameters. *BMC Bioinformatics* **16**, 118, doi:10.1186/s12859-015-0562-8 (2015).
- 61 McLean, C. & Kujawinski, E. B. AutoTuner: High Fidelity and Robust Parameter Selection for Metabolomics Data Processing. *Anal Chem* **92**, 5724-5732, doi:10.1021/acs.analchem.9b04804 (2020).
- 62 Zhou, G., Ewald, J. & Xia, J. OmicsAnalyst: a comprehensive web-based platform for visual analytics of multi-omics data. *Nucleic Acids Res* **49**, W476-W482, doi:10.1093/nar/gkab394 (2021).
- 63 Rappoport, N. & Shamir, R. Multi-omic and multi-view clustering algorithms: review and cancer benchmark. *Nucleic Acids Res* **47**, 1044, doi:10.1093/nar/gky1226 (2019).
- 64 Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **78**, 779-787, doi:10.1021/ac051437y (2006).
- 65 Prince, J. T. & Marcotte, E. M. Chromatographic alignment of ESI-LC-MS proteomics data sets by ordered bijective interpolated warping. *Anal Chem* **78**, 6140-6152, doi:10.1021/ac0605344 (2006).

Acknowledgements

Acknowledgements

The authors would like to thank Genome Canada, Génome Québec, Natural Sciences and Engineering Research Council of Canada (NSERC), and Canada Research Chairs (CRC) Program for funding support.

Author Contributions

Z.P., J.E., N.B. and J.X. prepared the manuscript. Z.P., G.Z., J.E., L.C., O.H. and J.X. contributed to the development and testing of the MetaboAnalyst. All authors read and approved the final manuscript.

Competing Financial Interests

The authors declare no competing financial interests.

Data Availability Statement

All example datasets used in the protocol are integrated as example datasets in their respective modules and are also available for download from the “Format” page of MetaboAnalyst (<https://www.metaboanalyst.ca/MetaboAnalyst/docs/Format.xhtml>). There are no restrictions on their use.

Code Availability Statement

MetaboAnalyst is freely accessible as a web-based application. The underlying R code is freely available at GitHub as the MetaboAnalystR (<https://github.com/xia-lab/MetaboAnalystR>) and OptiLCMS (<https://github.com/xia-lab/OptiLCMS>) packages under the GNU General Public License version 2 or later.

Figures

Figure 1

Figure 2

Figure 3