

mpactR: an R adaptation of the metabolomics peak analysis computational tool (MPACT) for use in reproducible data analysis pipelines

Allison R. Mason,¹ Gregory Johnson, Jr,¹ Joseph Krampen,¹ Jennifer N. T. Nguyen,² Marcy J. Balunas,^{1,2} Patrick D. Schloss¹

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT mpactR automates pre-processing of liquid chromatography-tandem mass spectrometry (LC-MS/MS) data from microbiological samples to correct mispicked peaks, resolve inter-sample variation in abundance across technical replicates, account for in-source ion fragmentation, and remove background noise to yield high-quality mass spectrometry features. The package is available through CRAN and GitHub.

KEYWORDS metabolomics, software, R package, data curation

There has been increased interest in performing metabolomic analyses of microbial communities with the goals of understanding the relationship between the structure and function of these communities and linking changes in the metabolome with changes in health. Combined with DNA and RNA-based sequencing of microbiomes, the broader availability of liquid chromatography-tandem mass spectrometry (LC-MS/MS) has made untargeted metabolomics a powerful addition to multi-omics analyses (e.g., 1–3). Traditionally, LC-MS/MS has been used to identify novel bioactive natural products. Incorporating metabolomics into microbiome studies presents several novel challenges. These include accessing software that can process large numbers of complex biological samples. Although there are numerous options for analyzing LC-MS/MS data, the software tends to be expensive, proprietary, has poor interoperability and is accessible only through a graphical user interface (GUI) as a website or stand-alone application. Additionally, analysis of large metabolomics data sets is complicated by inherent data variability. These factors limit the ability to foster transparency and reproducibility.

Initial efforts to overcome these challenges have included packages including the tidymass ensemble of R packages (4) and the metabolomics peak analysis computational tool (MPACT) (5). The tidymass ensemble is an open source R package seeking to centralize the full data workflow within the R ecosystem. However, existing peak-picking algorithms are error-prone, reducing our ability to identify high-quality ions. MPACT is an open source Python package allowing researchers to reconcile mispicked peaks, resolve inter-sample variation in abundance across technical replicates, account for in-source ion fragmentation, and remove background noise using data collected from solvent and/or media blanks. MPACT can only be used as a GUI and has limited options for visualizing and analyzing the processed data; therefore, we developed mpactR as an open source R package that provides users the useful functionality of MPACT while integrating the processed data in a manner that satisfies their research needs using the R ecosystem.

mpactR's website provides detailed installation instructions to install the package (<https://www.mums2.org/mpactr/>) from the Comprehensive R Archive Network (CRAN) or GitHub. The website provides tutorials on how to use the mpactR package as well as tutorials for using R with the output of mpactR to generate the types of figures that were originally produced in MPACT. For consistency, these tutorials use an example data

Editor Catherine Putonti, Loyola University Chicago, Chicago, Illinois, USA

Address correspondence to Patrick D. Schloss, pschloss@umich.edu.

Allison R. Mason and Gregory Johnson, Jr contributed equally to this article. Author order was determined based on seniority in the research group.

The authors declare no conflict of interest.

See the funding table on p. 2.

Received 10 September 2024

Accepted 4 December 2024

Published 15 January 2025

Copyright © 2025 Mason et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

set characterizing the metabolome of the tunicate-associated bacterium *Streptomyces* sp. PTY08712 used in the original MPACT publication (5). mpactR is designed using the R6 object-oriented programming system with reference semantics, allowing for minimal memory footprint while executing at least as fast as MPACT.

mpactR was designed to be straightforward to use and easy to integrate in automated reproducible analysis pipelines. mpactR requires two comma-separated values files, including a peak table and a metadata file containing relevant sample information that can be read in using the `import_data` function. Peak tables from multiple upstream analysis tools are accepted, including but not limited to Progenesis and Metabocape. Instructions on how to generate these files are provided at the mpactR website. Filtering is performed using any of four functions, and the package provides four summary functions to allow users to evaluate the effects of filtering. The accessor returns the processed data in the `data.table` package format facilitating rapid and memory-efficient analyses with other `data.table` functions, used with functions from R packages in the tidyverse, including `ggplot2`, or written to a file for uploading to the GNPS website (6).

ACKNOWLEDGMENTS

This project was supported, in part, by a grant from the US National Institutes of Health (U01CA264071) to P.D.S. and M.J.B.

AUTHOR AFFILIATIONS

¹Department of Microbiology & Immunology, University of Michigan, Ann Arbor, Michigan, USA
²Department of Medicinal Chemistry, University of Michigan, Ann Arbor, Michigan, USA

AUTHOR ORCIDs

Marcy J. Balunas  <http://orcid.org/0000-0003-2374-4048>
Patrick D. Schloss  <http://orcid.org/0000-0002-6935-4275>

FUNDING

Funder	Grant(s)	Author(s)
HHS National Institutes of Health (NIH)	U01CA264071	Marcy J. Balunas Patrick D. Schloss

AUTHOR CONTRIBUTIONS

Allison R. Mason, Conceptualization, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Writing – review and editing | Gregory Johnson, Jr, Conceptualization, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Writing – review and editing | Joseph Krampen, Methodology, Validation, Writing – review and editing | Jennifer N. T. Nguyen, Methodology, Validation, Writing – review and editing | Marcy J. Balunas, Conceptualization, Funding acquisition, Methodology, Writing – review and editing | Patrick D. Schloss, Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

Stable releases of mpactR are available through CRAN and developmental versions are available through the project’s GitHub website (<https://github.com/mums2/mpactr>). The package is available under the MIT Open Source License.

REFERENCES

1. Jenior ML, Leslie JL, Young VB, Schloss PD. 2018. *Clostridium difficile* alters the structure and metabolism of distinct cecal microbiomes during initial infection to promote sustained colonization. *mSphere* 3:e00261-18. <https://doi.org/10.1128/mSphere.00261-18>
2. Kyle KE, Puckett SP, Caraballo-Rodríguez AM, Rivera-Chávez J, Samples RM, Earp CE, Raja HA, Pearce CJ, Ernst M, van der Hooft JJJ, Adams ME, Oberlies NH, Dorrestein PC, Klassen JL, Balunas MJ. 2023. *Trachymyrmex septentrionalis* ants promote fungus garden hygiene using *Trichoderma*-derived metabolite cues. *Proc Natl Acad Sci U S A* 120:e2219373120. <https://doi.org/10.1073/pnas.2219373120>
3. Griffiths JA, Yoo BB, Thuy-Boun P, Cantu VJ, Weldon KC, Challis C, Sweredoski MJ, Chan KY, Thron TM, Sharon G, Moradian A, Humphrey G, Zhu Q, Shaffer JP, Wolan DW, Dorrestein PC, Knight R, Gradinaru V, Mazmanian SK. 2024. Peripheral neuronal activation shapes the microbiome and alters gut physiology. *Cell Rep* 43:113953. <https://doi.org/10.1016/j.celrep.2024.113953>
4. Shen X, Yan H, Wang C, Gao P, Johnson CH, Snyder MP. 2022. TidyMass an object-oriented reproducible analysis framework for LC-MS data. *Nat Commun* 13. <https://doi.org/10.1038/s41467-022-32155-w>
5. Samples RM, Puckett SP, Balunas MJ. 2023. Metabolomics peak analysis computational tool (MPACT): an advanced informatics tool for metabolomics and data visualization of molecules from complex biological samples. *Anal Chem* 95:8770–8779. <https://doi.org/10.1021/acs.analchem.2c04632>
6. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, et al. 2016. Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nat Biotechnol* 34:828–837. <https://doi.org/10.1038/nbt.3597>