Integrative analysis of multimodal mass spectrometry data in MZmine 3

Schmid, Robin ; Heuckeroth, Steffen ; Korf, Ansgar ; Smirnov, Aleksandr ; Myers, Owen ; Dyrlund, Thomas S ;
Bushuiev, Roman ; Murray, Kevin J ; Hoffmann, Nils ; Lu, Miaoshan ; Sarvepalli, Abinesh ; Zhang, Zheng ;
Fleischauer, Markus ; Dührkop, Kai ; Wesner, Mark ; Hoogstra, Shawn J ; Rudt, Edward ; Mokshyna, Olena ;
Brungs, Corinna ; Ponomarov, Kirill ; Mutabdžija, Lana ; Damiani, Tito ; Pudney, Chris J ; Earll, Mark ; Helmer,
Patrick O ; Fallon, Timothy R ; Schulze, Tobias ; Rivas-Ubach, Albert ; Bilbao, Aivett ; Richter, Henning ; et al

Integrative analysis of multimodal mass spectrometry data in MZmine 3

Robin Schmid[1,2,3,&], **Steffen Heuckeroth**[2,&], **Ansgar Korf**[2,&], **Aleksandr Smirnov**[4], **Owen Myers**[4], **Thomas S. Dyrlund**[5], **Roman Bushuiev**[3], **Kevin J. Murray**[6], **Nils Hoffmann**[7], **Miaoshan Lu**[8], **Abinesh Sarvepalli**[9], **Zheng Zhang**[1], **Markus Fleischauer**[10], **Kai Dührkop**[10], **Mark Wesner**[2], **Shawn J. Hoogstra**[11], **Edward Rudt**[2], **Olena Mokshyna**[3], **Corinna Brungs**[3],

[*]Corresponding author. tomas.pluskal@uochb.cas.cz.

[&]These authors contributed equally

Kirill Ponomarov[3], **Lana Mutabdžija**[3], **Tito Damiani**[3], **Chris J. Pudney**[12], **Mark Earll**[13], **Patrick O. Helmer**[2], **Timothy R. Fallon**[14], **Tobias Schulze**[15], **Albert Rivas-Ubach**[16], **Aivett Bilbao**[17], **Henning Richter**[18], **Louis-Félix Nothias**[19], **Mingxun Wang**[20], **Matej Orešič**[21,22], **Jing-Ke Weng**[23,24], **Sebastian Böcker**[10], **Astrid Jeibmann**[25], **Heiko Hayen**[2], **Uwe Karst**[2], **Pieter C. Dorrestein**[1], **Daniel Petras**[26], **Xiuxia Du**[4], **Tomáš Pluskal**[3,*]

[1]Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA, 92093, USA

[2]Institute of Inorganic and Analytical Chemistry, University of Münster, Münster, 48149, Germany

[3]Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague, 160 00, Czech Republic

[4]Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC, 28223, USA

[5]Steno Diabetes Center Copenhagen, Gentofte, 2820, Denmark

[6]Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota - Twin Cities, Minneapolis, MN, 55409, USA

[7]Institute for Bio- and Geosciences (IBG-5), Forschungszentrum Jülich GmbH, Jülich, 52425, Germany

[8]School of Engineering, Westlake University, Hangzhou, Zhejiang, 310000, China

[9]BlockLab, Center for Large Datasystems Research, San Diego Supercomputer Center, La Jolla, CA, 92093, USA

[10]Chair for Bioinformatics, Friedrich Schiller University Jena, Jena, 07743, Germany

[11]Agriculture and Agri-Food Canada, London Research and Development Centre, London, ON, N5V 4T3, Canada

[12]Datacraft Technologies, Mosman Park, WA, 6012, Australia

[13]Analytical Solutions Group, Product Technology and Engineering, Jealott's Hill International Research Centre, Bracknell, Berkshire, RG42 6EY, United Kingdom

[14]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, 92037, USA

[15]Department of Effect-Directed Analysis, Helmholtz Centre for Environmental Research - UFZ, Leipzig, 04318, Germany

[16]Ecology and Forest Genetics, Institute of Forest Sciences (ICIFOR-INIA-CSIC), Madrid, 28040, Spain

[17]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA, 99352, USA

[18]Clinic for Diagnostic Imaging, Diagnostic Imaging Research Unit (DIRU), University of Zurich, Zürich, CH-8057, Switzerland

[19]School of Pharmaceutical Sciences, University of Geneva, Geneva, CH-1211, Switzerland

[20]Department of Computer Science, University of California Riverside, Riverside, CA, 92521, USA

[21]School of Medical Sciences, Örebro University, Örebro, 701 82, Sweden

[22]Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, 20520, Finland

[23]Whitehead Institute for Biomedical Research, Cambridge, MA, 02142, USA

[24]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA

[25]Institute of Neuropathology, University Hospital Münster, Münster, 48149, Germany

[26]CMFI Cluster of Excellence, University of Tuebingen, Tuebingen, 72076, Germany

To the editor:

Innovation in mass spectrometry (MS) and the rapidly increasing throughput and sensitivity of MS instrumentation require adaptations and innovations in data processing tools. Here, we introduce MZmine 3, a scalable MS data analysis platform that supports hybrid datasets from various instrumental setups, including gas and liquid chromatography (GC and LC)-MS, ion mobility spectrometry (IMS)-MS, and MS imaging. In particular, the integration of IMS-MS imaging and LC-IMS-MS datasets provides opportunities for spatial metabolomics analyses with increased annotation confidence.

Over the past decade, the MZmine project has evolved into a community-driven, collaborative effort. As an open-source ecosystem for MS data processing, MZmine is a cross-platform software (Supplementary Note 1) that can be tuned for robust, scalable, and reproducible data analysis on personal computers as well as high-performance super computers. The project has seen continuous development since its inception in 2004.[1,2] Community additions (Fig. 1a) introduced various functions, such as performant feature detection workflows,[3,4] modules for lipid annotation,[5] and strong ties to other community projects (Fig. 1b). Here, data exchange formats and direct interfaces (listed in *Tool integration* in the documentation) enable downstream analysis in external tools, such as compound annotation in SIRIUS,[6] statistical analysis in MetaboAnalyst,[7] and directly bind MZmine results into the molecular networking ecosystem of the Global Natural Products Social Molecular Networking (GNPS) web-platform (Supplementary Note 2).[8–10]

Recent advances in MS instrumentation push sensitivity, resolving power, and data acquisition speed, resulting in increased data volume and complexity. Notably, IMS gains traction in the field by including an additional separation dimension to LC-MS or imaging-based techniques like matrix-assisted laser desorption/ionization (MALDI)-MS. These advances introduce new acquisition modes (e.g., parallel accumulation-serial fragmentation - PASEF)[11], or enable hyphenation of IMS and imaging, which was shown to improve annotation quality in MS imaging.[12] Furthermore, the number of large-scale cohort and multifactorial studies in clinical, environmental, and other fields is growing, as registered in the three major metabolomics data repositories, MassIVE/GNPS,[8] MetaboLights, and Metabolomics Workbench.[13] The need for scalable, reproducible, and flexible data analysis workflows that can combine mass spectrometry data from various sources, remains

unaddressed by existing tools. For example, to combine LC- and imaging-(IMS)-MS results from the same sample, users are forced to master multiple software tools[12] that divide the workflow and are specialized in either chromatography-MS (e.g., MS-DIAL, XCMS, OpenMS)[14–16] or MS imaging (e.g., METASPACE, rMSI, Cardinal MSI, SpectralAnalysis).[17]

The integrative spatial metabolomics workflow in MZmine 3 (Fig. 1c) imports LC-IMS-MS and IMS-MS imaging datasets stored in either open or vendor-specific formats and processes them by non-targeted feature detection. This entails resolving peak shapes for ion features in both the retention time (RT) and ion mobility dimension in LC-IMS-MS and extracting mobility-resolved ion image features with spatial distributions in IMS-MS imaging (Supplementary Fig. 1). Individual features from both methodologies are subsequently represented and aligned by their RT (LC only), *m/z*, and ion mobility values. The resulting aligned feature list combines the strengths of the individual analytical methods by integrating the compound annotation capabilities of modern chromatography-based MS with spatial metabolite distributions that can be mapped to histological data, addressing the issue of missing MS^2 data in most imaging studies. For data evaluation, MZmine organizes annotations in a feature table with interactive charts, exemplified in Fig. 1d for one ion feature detected in LC-IMS-MS samples and aligned to an ion image from one MALDI-IMS-MS imaging dataset. An exemplary spatial metabolomics workflow leading to LC-IMS-MS resolved molecular networks, enriched with spatial ion feature information is described in Supplementary Note 2 (Supplementary Fig. 4). Additional visualization modules (Supplementary Fig. 5) connect all available data dimensions; a fast memory-mapped data backend enables interactive exploration.

In MZmine 3, special attention was directed towards scalability due to the ever-increasing study sizes that lead to large raw data volumes, particularly in the case of LC-IMS-MS datasets. Efficient memory management and parallelization removed bottlenecks, resulting in an 89% reduction in processing time for 250 dissolved organic matter (DOM) samples when compared to MZmine 2. A stress test demonstrated in high sample throughput, where the mean processing times elapsed to 0.1% to 0.3% of the total data acquisition time for six different LC-MS datasets (Supplementary Note 3; Supplementary Fig. 6). Further, MZmine 3 was benchmarked using 8273 fecal $LC-MS^2$ samples, requiring just 47 min of processing time (see hardware specifications in Supplementary Note 3).

The improved performance of MZmine 3 over previous MZmine versions now allows processing of large datasets, including large-volume LC-IMS-MS data. For new users, the MZmine website contains detailed manuals and video tutorials, and the new processing wizard in MZmine provides starting points for various standard workflows and mass spectrometer types. In addition, a development tutorial is available for potential new contributors, and the modular design of MZmine enables testing and implementing new ideas within the MZmine framework.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Data availability

Datasets are available on MassIVE[8] with their accession IDs:

- MSV000088054, human cohort study, LC-MS, neg

- MSV000087728, diverse plant extracts, LC-MS^2, top-3 DDA, pos

- MSV000090079, DOM, LC-MS^2, top-5 DDA, pos

- MSV000090328, sheep brain, LC-tims-MS, PASEF, pos

- MSV000090327, piper plant extracts, LC-tims-MS, PASEF, pos

IMS resolved ion identity molecular networking results are available through GNPS: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=7a06fa3dfadd4158bcb4ee300b574747

References

1. Katajamaa M, Miettinen J & Oresic M Bioinformatics 22, 634–636 (2006). [PubMed: 16403790]

2. Pluskal T, Castillo S, Villar-Briones A & Oresic M BMC Bioinformatics 11, 395 (2010). [PubMed: 20650010]

3. Smirnov A. et al. Anal. Chem 91, 9069–9077 (2019). [PubMed: 31274283]

4. Du X, Smirnov A, Pluskal T, Jia W & Sumner S Methods Mol. Biol 2104, 25–48 (2020). [PubMed: 31953811]

5. Korf A, Jeck V, Schmid R, Helmer PO & Hayen H Anal. Chem 91, 5098–5105 (2019). [PubMed: 30892876]

6. Dührkop K. et al. Nat. Methods 16, 299–302 (2019). [PubMed: 30886413]

7. Pang Z. et al. Nucleic Acids Res. 49, W388–W396 (2021). [PubMed: 34019663]

8. Wang M et al. Nat. Biotechnol 34, 828–837 (2016). [PubMed: 27504778]

9. Nothias L-F et al. Nat. Methods 17, 905–908 (2020). [PubMed: 32839597]

10. Schmid R et al. Nat. Commun 12, 3832 (2021). [PubMed: 34158495]

11. Meier F et al. J. Proteome Res 14, 5378–5387 (2015). [PubMed: 26538118]

12. Helmer PO et al. Anal. Chem 93, 2135–2143 (2021). [PubMed: 33416303]

13. Aksenov AA, da Silva R, Knight R, Lopes NP & Dorrestein PC Nature Reviews Chemistry 1, 1–20 (2017).

14. Smith CA, Want EJ, O'Maille G, Abagyan R & Siuzdak G Anal. Chem 78, 779–787 (2006). [PubMed: 16448051]

15. Tsugawa H et al. Nat. Biotechnol 38, 1159–1163 (2020). [PubMed: 32541957]

16. Röst HL et al. Nat. Methods 13, 741–748 (2016). [PubMed: 27575624]

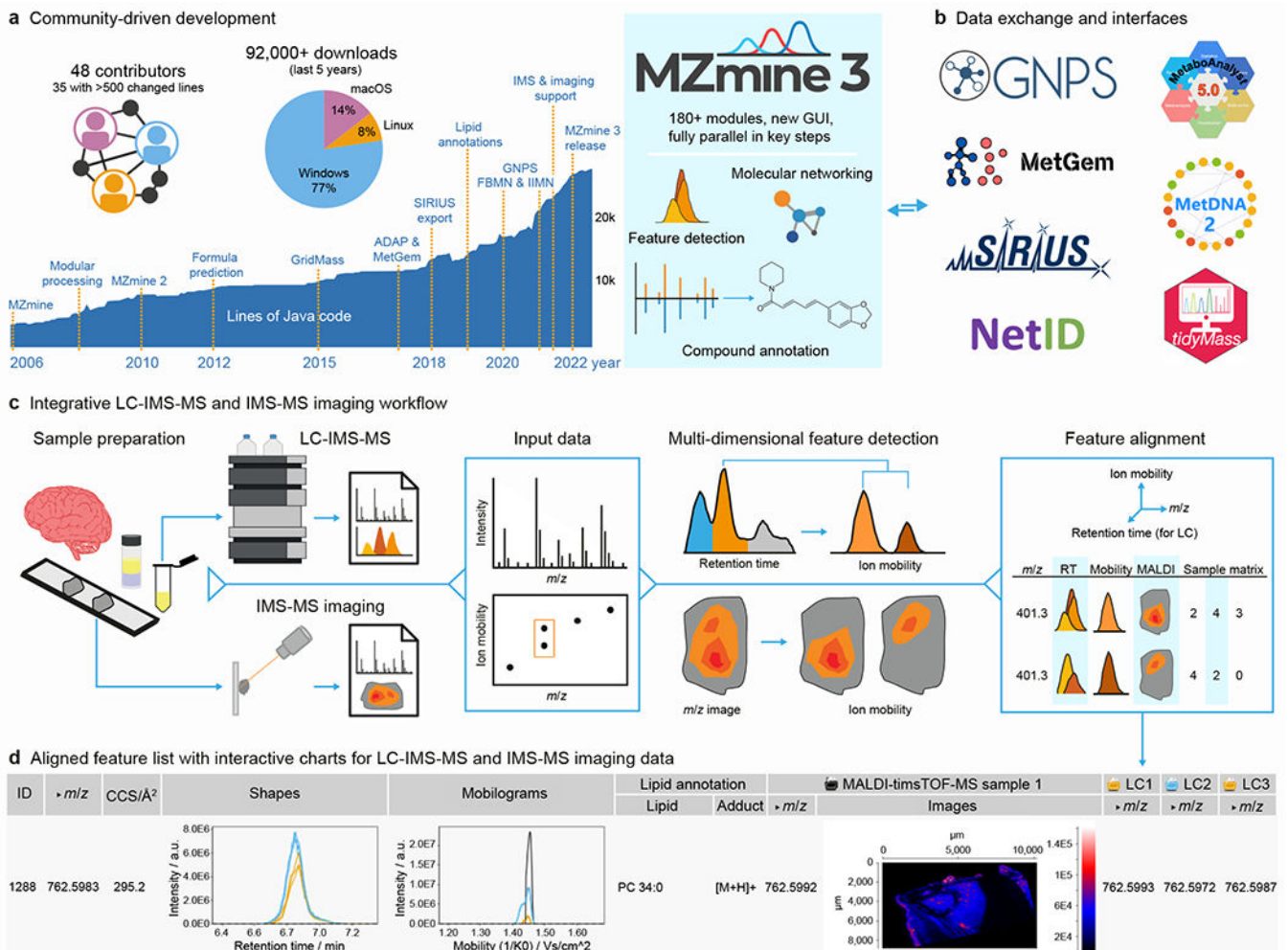17. Weiskirchen R, Weiskirchen S, Kim P & Winkler RJ Cheminform. 11, 16 (2019).

18. https://github.com/mzmine/mzmine3/

Fig. 1 |. MZmine, an open-source community project for integrative LC-IMS-MS and IMS-MS data processing.

a, Overview of active developments and key additions to MZmine since the first publication, which led to over 180 modules that now drive interactive, reproducible, and efficient data processing and visualization in MZmine 3. **b**, Data exchange formats and direct interfaces enable downstream analysis with strong ties to projects like GNPS, SIRIUS, and MetaboAnalyst. **c,** The integrative LC- and IMS-MS imaging workflow applies feature detection in RT, ion mobility, and *m/z* dimension to MS data stored in open or vendor formats. Comprehensive processing and annotation results are merged into **d**, an aligned feature list with one ion feature detected in LC-IMS-MS samples and aligned to one MALDI-IMS-MS ion feature image. Annotation results (*Lipid annotation column*) and interactive charts include the table columns *Shapes* (extracted ion chromatograms), *Mobilograms* (extracted ion mobilograms), and *Images* (extracted ion images).