# Supplementary Information

# Integrative analysis of multimodal mass spectrometry data in MZmine 3

## Supplementary Note 1

### MZmine 3 development and features

MZmine's underlying data model was completely rewritten and modularized to achieve greater expandability. This is directly reflected in the graphical user interface, e.g., individual modules can add new data types or graphical columns into the processed feature tables. This modular architecture allows MZmine 3 to import and process various MS data types, including GC-MS, LC-MS, LC-IMS-MS, or MS-imaging datasets. The memory management was optimized, and the available RAM (Random Access Memory) is extended by memory mapping spectral data and ion feature data onto fast local drives, freeing more memory for data processing. Systems without RAM constraints can also process all data in memory if preferred. Previously, feature alignment and gap-filling used to be bottlenecks in the feature detection workflows. These key steps were optimized and parallelized in MZmine 3, increasing the overall sample throughput (**Supplementary Note 3**). MZmine 3 is a modern Java application that is packaged with a customized Java Virtual Machine, making it cross-platform compatible without the need for additional software package installation. The last nearly two decades have seen various additions to MZmine.[1–13] An up-to-date overview of all processing modules in MZmine 3 is available in the MZmine documentation:

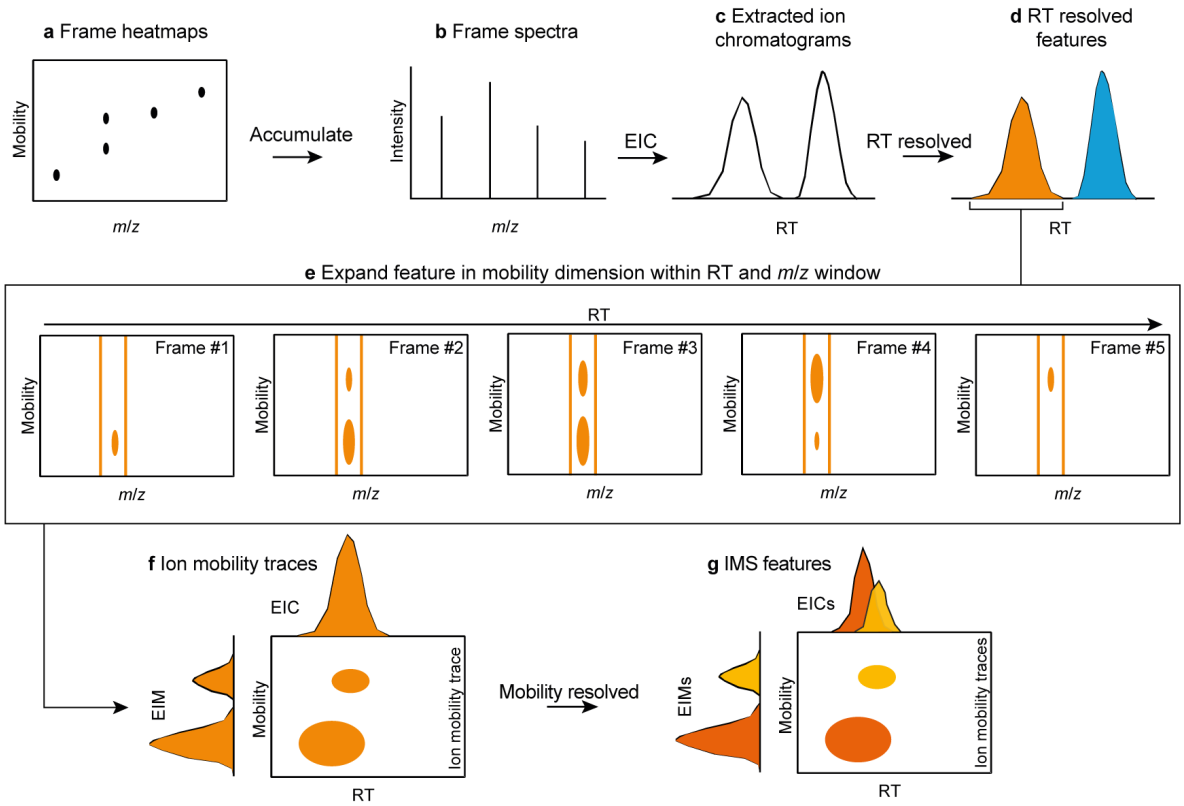
https://mzmine.github.io/mzmine_documentation/coding/module_list.html

Several other community tools, e.g., tidyMass,[14] MetDNA,[15] GNPS[16], NeatMS,[17] mzRAPP,[18] MetGem,[19], SLAW,[20] FERMO,[21] MEMO,[22] Inventa,[23] and NetID[15,24], can import MZmine results from its various data exchange formats:

https://mzmine.github.io/mzmine_documentation/tool_integration

**Overview of MZmine's 4D LC-IMS-MS feature detection workflow**

The MZmine LC-IMS-MS feature detection adds a few steps to the classical LC-MS workflow. IMS-MS data is supported through open MS data formats, e.g., .mzML, and vendor-specific formats, e.g., .tdf (Bruker Daltonics). At each retention time, multiple mobility-resolved mass spectra make up a frame; visualized as a heatmap in **Supplementary Fig. 1a**. All scans within a frame are accumulated into a merged frame spectrum (**b**), similar to the mass spectra in regular LC-MS datasets without ion mobility separation. Those frame spectra are used by the ADAP chromatogram builder[10] to create extracted ion chromatograms (EIC, **c**). MZmine then offers multiple algorithms to resolve chromatograms to RT-resolved features (**d**), defined by their boundaries in *m/z* and RT dimension. Features are then expanded into the mobility dimension by searching the original mobility-resolved scans for signals in the given *m/z* and RT boundaries (**e**), creating ion mobility traces (**f**). Every ion mobility trace is described by raw 4D data (signal distribution in *m/z*, RT, mobility, and intensity dimension) and not only by their 2D projections, namely EIC and extracted ion mobilogram (EIM). These projections are built from the 4D raw data as well, to allow either a simple (2D: EIC, EIM) or more complex (4D: ion mobility traces) visualization and investigation of every feature. To create mobility-resolved IMS features (**g**), the EIMs are resolved by the same modules as the RT features. After this step, the 2D projections are rebuilt from the respective 4D raw data of the ion mobility traces. Afterward, all features in the MZmine feature table are IMS features composed of the underlying 4D data, ensuring reproducibility and traceability to the raw data, and their 2D projections as EICs and EIMs. Depending on the underlying data, EICs and EIMs may be smoothed and EIMs may be binned before feature resolving steps to improve the shapes before peak detection.




**Supplementary Fig. 1 | Schematic LC-IMS-MS feature detection workflow.**

**a**, all mobility-resolved mass spectra measured at the same retention time (the same IMS ramp), typically represented as frame heatmaps, are accumulated into a merged frame spectrum, **b**. **c**, EICs connect signals within *m/z* tolerance across all merged frame spectra in the RT dimension. **d**, EICs are split into RT-resolved features by one of MZmine's resolving algorithms. **e**, the ion mobility dimension is added to features by searching signals in all mobility scans that fall within the *m/z* and RT boundaries of a feature, expanding them to 4D ion mobility traces with *m/z*, RT, mobility, and intensity dimensions. **f**, ion mobility features are resolved in the mobility dimension, using the same algorithms as in **d** and after optional smoothing and binning of multiple mobility values to improve the shape of EIMs.

**Multidimensional feature alignment across *m/z*, ion mobility, and RT**

Feature alignment (Join aligner) scores features from multiple samples against a master feature list in parallel and the best matches are aligned: If multiple features from the same sample match to a base row, the one with the lowest weighted deviation in all available dimensions $p$ (*m/z*, RT, ion mobility) is added. For each feature-pair within tolerances, all available descriptors are divided by the maximum allowed alignment difference ($tolerance_p$), weighted by a user-defined factor $weight_p$, and summed to a score.

$$score = \sum_{p} \left( \left( 1 - \frac{\Delta p}{tolerance_p} \right) \cdot weight_p \right)$$

Additional pre-filters are available to score only features that share a similar isotope pattern, $MS^2$ spectral similarity, or the same identifier from different annotation modules in MZmine, including the spectral library search and the local database compound matching (from a csv compound table). The $MS^2$ spectral similarity can provide a more stringent filtering approach to feature alignment. However, this approach is limited to datasets with a high fragmentation spectra-coverage across samples and features. Currently, the absence of DDA acquisition of fragmentation spectra in (IMS)-MS imaging usually prevents the usage of this strategy.

The current workflow aligns all samples from LC-(IMS)-MS into one aligned feature list. Afterward, all (IMS)-MS imaging samples are aligned onto this feature list. Starting with MZmine 3.3.0, multiple metrics are available in the feature table, reflecting on the alignment of each aligned feature (row). This includes the alignment rate (aligned/total samples) and the mean distances between the features *m/z*, RT, and ion mobility compared for each sample to their average value. These mean distances can provide valuable insight into the mass accuracy and potential RT and mobility shifts. Further, they can guide the optimization of the alignment parameters. A weighted distance score is calculated for the final aligned feature based on the equation above. Each feature falls within different *m/z*, RT, and mobility ranges which might contain more or fewer isomers or other interferences. By summing the total number of features within those ranges across all samples, minus the aligned features, a sum of additional features is listed for each row. This number reflects on closely eluting features in RT and ion mobility with a similar *m/z* and can help identify possible alignment issues or raise confidence in the alignment if there are no additional features found.

### Overview of MZmine's feature annotation modules

MZmine comes with dedicated modules for the annotation of ion types (adducts), molecular formulas, and compounds listed in databases. Furthermore, export formats allow exporting MZmine results into other annotation tools, such as SIRIUS. The IMS-MS-derived CCS values are currently considered by the spectral library search, local compound database search, and the FBMN/IIMN (GNPS) export. Other modules can be applied to mobility-resolved features. The annotations from multiple modules are reflected in the feature table by multiple columns (see **Supplementary Fig. 2**)

| Aligned feature list 13C gaps corr | | | | | | | | | | | NI... |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ID | ▸ RT | ▸ m/z | CCS / Å² | Height | Spectral match | | Compound DB | | Lipid Annotation | | Height |
| | | | | | Spectral match ▼ | Adduct | Compound DB | Adduct | Lipid Annotation | Adduct | |
| 471 | 6.62 | 828.5743 | 297 | 2.7E5 | PC 18:1_18:2 (0.735) | [M+HCOO]- | PC 36:3 | [M+FA]- | PC 16:0_20:3 | [M+HCOO]- | 2.7E5 |
| 407 | 6.51 | 878.5910 | 305 | 5.4E4 | PC 18:0_22:6 (0.845) | [M+HCOO]- | PC 40:6 | [M+FA]- | PC 18:1_22:5 | [M+HCOO]- | 5.3E4 |
| 425 | 6.54 | 882.6194 | 307 | 1.4E4 | PC 18:0_22:4 (0.750) | [M+HCOO]- | PC 40:4 | [M+FA]- | PC 18:0_22:4 | [M+HCOO]- | 1.4E4 |
| 424 | 6.54 | 854.5910 | 302 | 2.8E5 | PC 18:0_20:4 (0.771) | [M+HCOO]- | PC 38:4 | [M+FA]- | PC 18:0_20:4 | [M+HCOO]- | 2.8E5 |
| 443 | 6.57 | 856.6032 | 303 | 8.8E4 | PC 18:0_20:3 (0.727) | [M+HCOO]- | PC 38:3 | [M+FA]- | PC 18:0_20:3 | [M+HCOO]- | 8.8E4 |
| 480 | 6.63 | 830.5910 | 298 | 4.2E5 | PC 18:0_18:2 (0.789) | [M+HCOO]- | PC 36:2 | [M+FA]- | PC 18:0_18:2 | [M+HCOO]- | 4.2E5 |

**Supplementary Fig. 2 | Example of a feature list annotated by multiple methods.**

Annotations originate from the spectral library search, local compound database search, and lipid annotation modules in the columns *Spectral match*, *Compound DB*, and *Lipid Annotation*, respectively.

The following list describes the most used annotation modules:

**Ion identity networking**[7] builds ion type libraries based on user-defined adducts (e.g., $H^+$ and $Na^+$) and in-source modifications (e.g., $-H_2O$ and +ACN). All combinations of an adduct and a modification are combined to represent monomers (e.g., $[M+H]^+$) and multimers (e.g., $[2M+H]^+$). Before annotating ions, features are grouped based on various criteria, including a retention time tolerance and correlation of the feature shapes and intensities across all samples. Pairs of grouped features are then matched against an ion library to annotate combinations that point to the same neutral mass. Options are available to define and add new adducts and modifications and to refine the results to retain only larger ion identity networks of ions that originate from the same molecule. More information is available in the online MZmine documentation:

https://mzmine.github.io/mzmine_documentation/visualization_modules/interactive_ion_id_netw/interactive_ion_id_netw.html

**Formula prediction** is available to run on a whole feature list or on individual rows. There is an option to use the ion type from ion identity networking or to overwrite all ions by a fixed ionization type. Filters include an *m/z* tolerance, a list of elements with minimum and maximum occurrence, element count heuristics and ring-and-double-bond equivalent restrictions[25], isotope pattern scoring, and fragmentation pattern scoring.[9] The final results are sorted based on a weighted combination of the mass deviation, isotope pattern score, and fragmentation pattern score. More information is available in the documentation:

https://mzmine.github.io/mzmine_documentation/module_docs/id_spectra_chem_formula/chem-formula-pred.html

**Spectral library search** matches experimental fragmentation spectra against local spectral libraries using spectral alignment and spectral similarity (e.g., cosine similarity for $MS^2$ spectra or composite cosine similarity for GC-EI-MS spectra). Filters include *m/z* tolerances for precursor matching and

signal alignment, a minimum number of matched signals, and a similarity threshold. A retention time tolerance is available for chromatography-MS data and a CCS tolerance can prefilter ion mobility data. A $^{13}C$ deisotoper can be applied to experimental and reference spectra before matching, to remove residual $^{13}C$ isotope signals, that were co-isolated. Supported formats for reference spectra include .json (MassBank of North America (MoNA), GNPS, MZmine), .mgf (GNPS, MoNA, and others), .msp (NIST), and .jdx (JCAMP-DX). Recommended downloads are provided by GNPS (https://gnps-external.ucsd.edu/gnpslibrary) and MoNA (https://mona.fiehnlab.ucdavis.edu/downloads). More information is available in the documentation:

https://mzmine.github.io/mzmine_documentation/module_docs/id_spectral_library_search/spectral_library_search.html

**Local compound database (CSV) search** loads compound information from a tabular comma-separated file and annotates features that match multiple filters. Based on a provided neutral mass, multiple ion types (adducts and in-source modifications) are searched. All matches are added to a feature within *m/z* tolerance, RT tolerance, and ion mobility or CCS tolerance for IMS-MS. More information is available in the documentation:

https://mzmine.github.io/mzmine_documentation/module_docs/id_prec_local_cmpd_db/local-cmpd-db-search.html

**Lipid annotation** follows a rule-based approach, which is common in computational lipidomics.[26] For each lipid class, a set of ion notation-specific fragmentation rules are provided, reflecting the differing fragmentation behavior of ions, e.g., [M+H]+ or [M+Na]+. After matching the accurate *m/z*, $MS^2$ fragment ions and neutral losses are annotated by those rules, covering head group and chain level information. If lipid class-specific fragments are matched, a species level or molecular species level annotation will be added to the feature; otherwise, the putative $MS^1$ annotation is discarded (**Supplementary Fig. 3**). More information is available in the documentation:

https://mzmine.github.io/mzmine_documentation/module_docs/id_lipid_annotation/lipid-annotation.html

**SIRIUS export (SIRIUS,[27] ZODIAC,[28] CSI:FingerID,[29] CANOPUS)**[30] creates an .mgf file with representative spectra per feature. Both $MS^1$ and $MS^2$ data are exported to allow the full use of the SIRIUS software that comprises multiple tools for the prediction of molecular formulas, ion types, compound structures, and compound classes. For each feature, a representative $MS^1$ spectrum and multiple $MS^2$ spectra are exported. The $MS^2$ level includes either all available spectra or different options of merged spectra. The SIRIUS results can be merged back into MZmine results using the MZmine feature ID. More information is available in the documentation:

https://mzmine.github.io/mzmine_documentation/module_docs/io/data-exchange-with-other-software.html#sirius-csi-fingerid

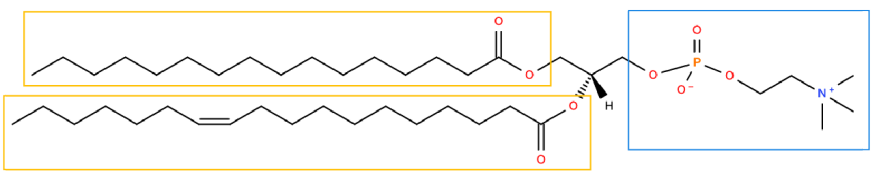
**FBMN[6]/IIMN[7] export (and the GNPS ecosystem[16])** creates all files needed for running Feature-based Molecular Networking and Ion Identity Molecular Networking on GNPS. This includes the quantification table (.csv), an .mgf file with one representative $MS^2$ spectrum per feature, and the additional ion identity networking edges (optionally). The $MS^2$ spectrum is either the most abundant with the highest TIC or a merged spectrum. Those molecular networking (MN) workflows represent features as nodes in a network of (cosine) similarities between fragmentation spectra. GNPS annotates compounds by matching against their open public spectral libraries. As an open community project,

many tools use the MN output for annotation propagation or for further analysis. Using the MZmine feature ID, results from GNPS, SIRIUS, and other tools can be merged. CCS values are exported for IMS-MS data and can provide coloring-schemes for the visual interpretation of networks. More information is available in the documentation:

https://mzmine.github.io/mzmine_documentation/module_docs/io/data-exchange-with-other-software.html#gnps-fbmniimn-export

**a** **Example lipid:**

1-hexadecanoyl-2-(11Z-octadecenoyl)-sn-glycero-3-phosphocholine, **PC 16:0/18:1(11Z)**



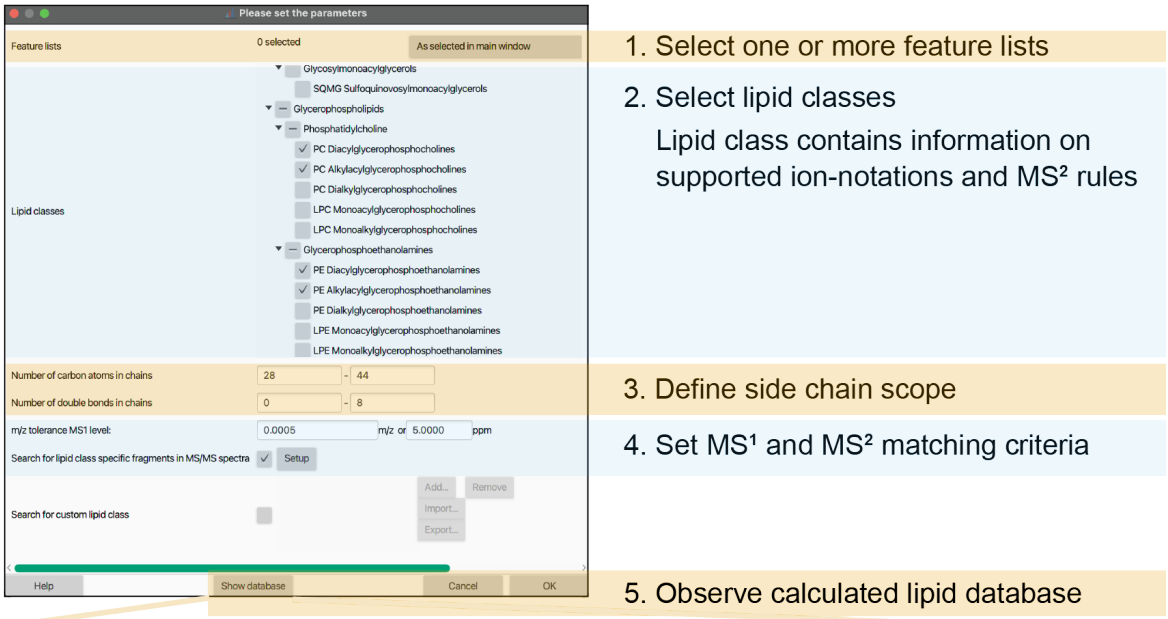
Use rule sets to avoid overannotation

**Supported annotation levels based on MS² information:**

Species level, headgroup information — **PC 34:1**

Molecular species level, side chain information — **PC 16:0_18:1**

**b**




1. Select one or more feature lists
2. Select lipid classes
   Lipid class contains information on supported ion-notations and MS² rules
3. Define side chain scope
4. Set MS¹ and MS² matching criteria
5. Observe calculated lipid database

**c**

| ID | Lipid class | Molecular formula | Abbreviation | Exact mass | Info | Status | MS/MS fragments positive ionization |
|---|---|---|---|---|---|---|---|
| 1 | Diacylglycerophosphocholines | C36H72NO8P | PC 28:0 | [M+Na]+ 700.4888<br>[M+HCOO]- 722.4978<br>[M+CH3COO]- 736.51...<br>[M+H]+ 678.5068 | [M+CH3COO]- interference with PC 29:0 [M+HCOO]-<br>[M+Na]+ possible interference with PC 30:3 [M+H]+ Δ 0.0024<br>[M+Na]+ interference with PE 31:0 [M+Na]+<br>[M+H]+ interference with PE 31:0 [M+H]+<br>[M+Na]+ possible interference with PE 33:3 [M+H]+ Δ 0.0024 | | [M+H]+, HEADGROUP_FRAGMENT C5H15NO...<br>[M+H]+, ACYLCHAIN_FRAGMENT_NL<br>[M+H]+, ACYLCHAIN_MINUS_FORMULA_FRA...<br>[M+Na]+, HEADGROUP_FRAGMENT_NL C3H...<br>[M+Na]+, HEADGROUP_FRAGMENT_NL C5H1...<br>[M+Na]+, HEADGROUP_FRAGMENT C2H5O4...<br>[M+HCOO]-, ACYLCHAIN_FRAGMENT<br>[M+CH3COO]-, ACYLCHAIN_FRAGMENT |
| 2 | Diacylglycerophosphocholines | C36H70NO8P | PC 28:1 | [M+Na]+ 698.4731<br>[M+HCOO]- 720.4821<br>[M+CH3COO]- 734.4...<br>[M+H]+ 676.4912 | [M+CH3COO]- interference with PC 29:1 [M+HCOO]-<br>[M+Na]+ possible interference with PC 30:4 [M+H]+ Δ 0.0024<br>[M+Na]+ interference with PE 31:1 [M+Na]+<br>[M+H]+ interference with PE 31:1 [M+H]+<br>[M+Na]+ possible interference with PE 33:4 [M+H]+ Δ 0.0024 | | [M+H]+, HEADGROUP_FRAGMENT C5H15NO...<br>[M+H]+, ACYLCHAIN_FRAGMENT_NL<br>[M+H]+, ACYLCHAIN_MINUS_FORMULA_FRA...<br>[M+Na]+, HEADGROUP_FRAGMENT_NL C3H...<br>[M+Na]+, HEADGROUP_FRAGMENT_NL C5H1...<br>[M+Na]+, HEADGROUP_FRAGMENT C2H5O4...<br>[M+HCOO]-, ACYLCHAIN_FRAGMENT |

**Supplementary Fig. 3 | Example lipid PC 16:0/18:1(11Z), highlighting supported annotation levels and screenshots of the lipid annotation module in MZmine 3.**

**a**, MZmine 3 supports species and molecular species annotation levels following guidelines recommended by the Lipidomics Standard Initiative.[31] **b**, The module can be applied to any feature list (1). The user defines the scope of the database by selecting lipid classes (2) and side-chain constraints as the number of carbons and double bonds (3). Matching criteria can be set for $MS^1$ and $MS^2$ levels (4). Before running the module, the on-the-fly created lipid species database can be viewed in a table (5). Listing the calculated exact *m/z* for supported ion notations (lipid class-specific) together with the implemented $MS^2$ rules. Information on potential $MS^1$ overlaps with other lipid species is provided in the *Info* and *Status* columns (orange: potential interference; cyan: clear) of **c,** the generated lipid database.

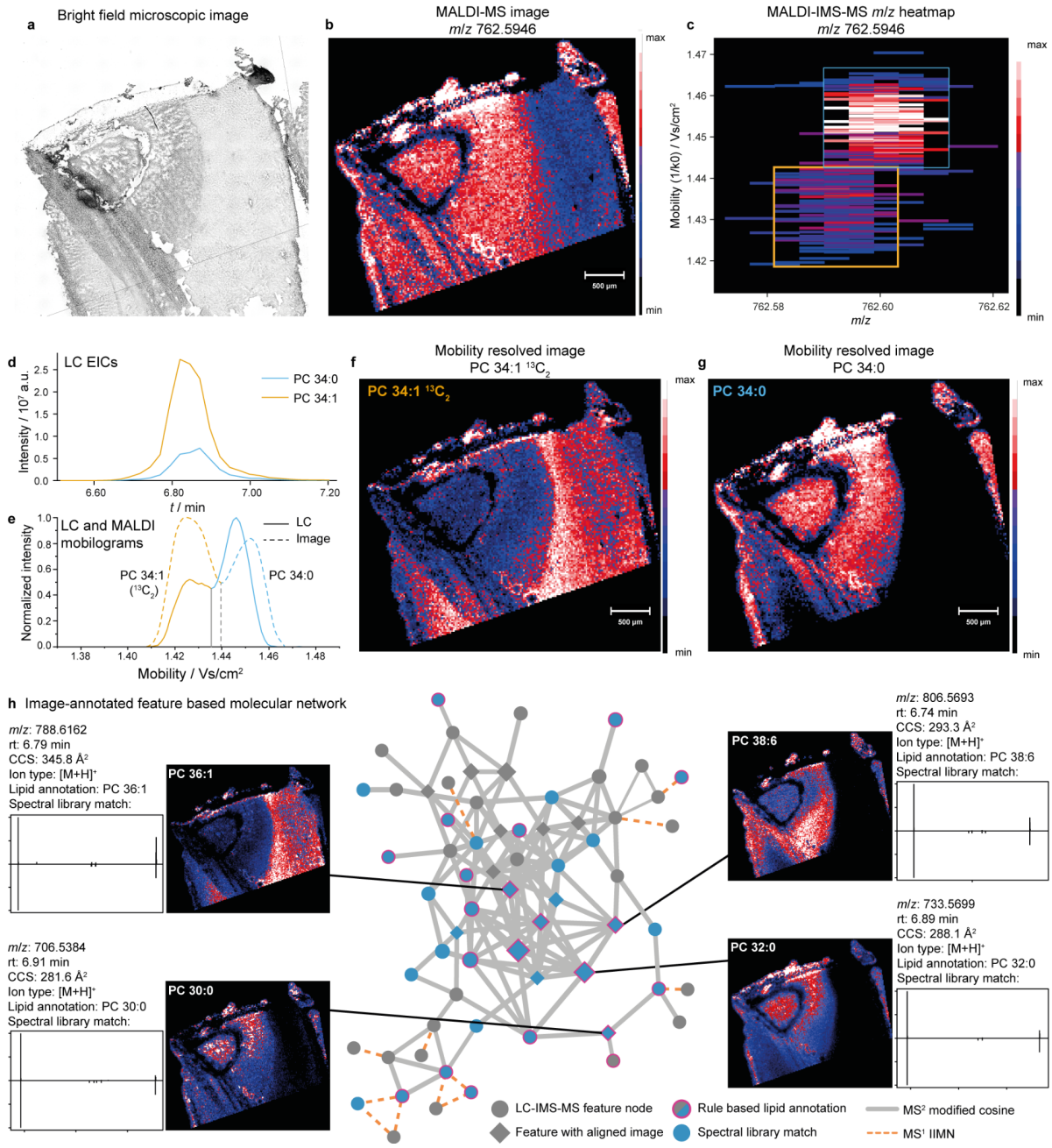
## Supplementary Note 2

### Example for integrative LC-IMS-MS and MALDI-IMS-MS imaging analysis in MZmine 3

Over the past years, MZmine has evolved from a raw data processing tool into a mature analysis pipeline. Implementing annotation capabilities based on various characteristics of LC-$MS^2$ data, such as $MS^1$ mass accuracy, $MS^2$ spectra matching, or the isotope pattern scoring, contributes towards a comprehensive matured data analysis solution. With the introduction of IMS data processing in MZmine 3, annotation tools such as the lipid annotation module,[3] can now leverage mobility-resolved features and $MS^2$ spectra obtained by PASEF data acquisition. For example, IMS can separate chimeric spectra produced from isobaric precursor ions.

In MS imaging analysis, a crucial drawback is the lack of large-scale $MS^2$-based annotations, because images are usually acquired on the $MS^1$ level only and data-dependent acquisition (DDA) methods are still lacking. Furthermore, even with $MS^2$ acquisition, the windows for precursor ion isolation typically span unit resolution and cannot resolve all ions. This is particularly challenging when analyzing lipids, where the overlap of isomers and isobars must be considered.[32] The annotation confidence can be increased by acquiring LC-IMS-$MS^2$ data of the same sample to create a tissue-specific, $MS^2$ curated, and CCS value enriched target list.[33] MZmine 3 combines the annotations obtained from LC-IMS-MS analysis with the imaging data of a MALDI-IMS-MS experiment. The feature alignment module groups features from different data types using *m/z* and CCS, as demonstrated in **Supplementary Fig. 4**. A sheep brain tissue was analyzed by hydrophilic interaction liquid chromatography (HILIC)-IMS-MS and MALDI-IMS-MS. A microscopic bright field image is depicted in **Supplementary Fig. 4a**. A common challenge in annotating lipid MS imaging data is the differentiation of naturally occurring $^{13}C_2$ isotopes from species differing by double bonds (e.g., $[PC\ 34{:}1+H+^{13}C_2]^+ \approx [PC\ 34{:}0+H]^+$, $\Delta m/z$ 0.009). The distribution of 762.5946 ±0.01 *m/z* is shown in **Supplementary Fig. 4b**, which covers $[PC\ 34{:}0+H]^+$ and $[PC\ 34{:}1+H+^{13}C_2]^+$ and therefore cannot clearly be localized. However, the mobility-*m/z* heatmap in **Supplementary Fig. 4c** reveals the presence of two different compounds. While this distribution can be resolved by MALDI-IMS-MS alone, a confident $MS^2$-backed annotation requires the feature alignment to the lipid extracts analyzed by HILIC-IMS-MS (**Supplementary Fig. 4d**). $MS^2$ spectra were annotated by MZmine's rule-based lipid annotation module[3], revealing that both lipids are present in the sample and are resolved in the IMS dimension (**Supplementary Fig. 4e**). By aligning the complementary datasets using *m/z* and mobility, as shown in **Supplementary Fig. 4e**, the annotation obtained by HILIC-IMS-$MS^2$ analysis can be transferred to the mobility-resolved images shown in **Supplementary Fig. 4f and Supplementary Fig. 4g**. In contrast to the non-mobility-resolved distribution in **Supplementary Fig. 4b**, two distinct, anti-correlated distributions are found for $[PC\ 34{:}0+H]^+$ and $[PC\ 34{:}1+H+^{13}C_2]$. Potentially, the PC 34:1 distribution could be deduced from the monoisotopic mass alone, and the two ions at *m/z* 762.59 could be distinguished by ultrahigh resolution MS (R > 85,000). However, the PC 34:0 annotation was only possible due to the mobility separation and integrative analysis of complementary HILIC-IMS-$MS^2$ and MALDI-IMS-MS data, highlighting the potential of such a combined analysis.[33] Using IIMN in the GNPS environment, the molecular network in **Supplementary Fig. 4h** illustrates the aligned LC and MALDI results by representing ion features (*m/z*, CCS, RT, spatial distribution, and $MS^2$) as nodes, connected by pairwise matching of fragmentation spectra by modification-aware cosine similarity (grey edges). Ions from the same molecule are connected by feature shape correlation (red dashed edges). If the features are identified by rule-based lipid annotation or spectral library matching, pink borders or blue fillings are used, respectively. Features with images from MALDI-IMS-MS analysis show a diamond shape.

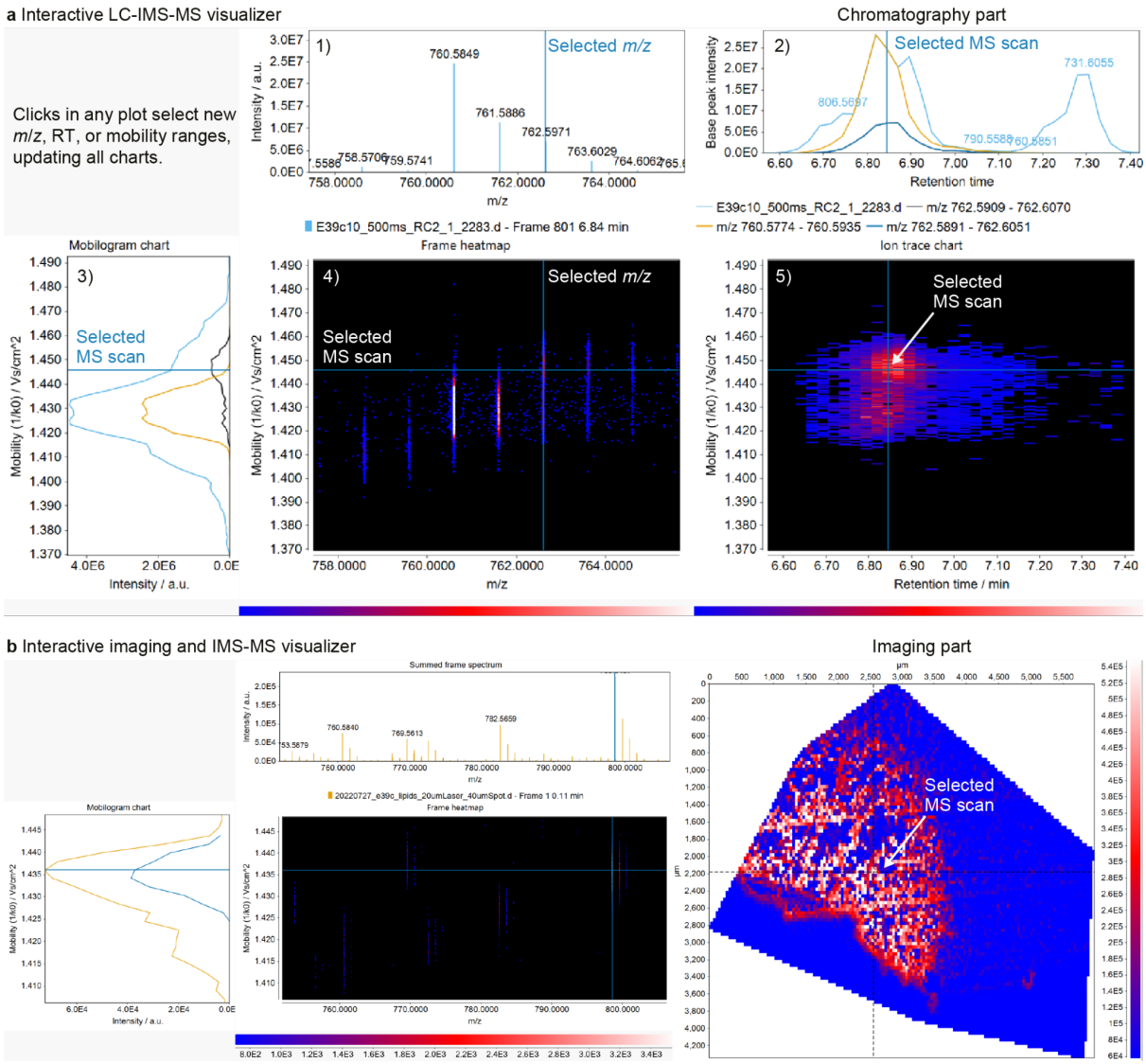
For manual investigation of LC-IMS-MS and IMS-MS data, two new interactive visualizers were developed for MZmine 3 (**Supplementary Fig. 5a** and **5b).** Both visualizers allow easily accessible point-and-click investigation of the multidimensional data.




**Supplementary Fig. 4 | Alignment of ion mobility-resolved imaging and chromatography MS data for improved annotations in spatial omics workflows.**

**a,** A microscopic bright field image of a sheep cerebrum and **b,** an exemplary signal distribution of *m/z* 762.5946 ±0.01 are shown. **c,** The ion mobility *m/z* heatmap indicates the presence of an isobaric interference. **d,** HILIC-IMS-MS extracted ion chromatograms (EIC) of PC 34:0 and PC 34:1 obtained from a Matyash extraction,[34] indicating that both lipids are present in the sample. **e,** The mobilograms of *m/z* 762.5946 from the imaging (dotted) and HILIC (solid) data are shown. $MS^2$ spectra acquired during the HILIC separation allow compound annotation as PC 34:0 and the $^{13}C_2$ isotope of PC 34:1. This annotation can be confidently transferred to the MALDI MS imaging data by aligning the HILIC features with mobility resolved images (panel **f** and **g**). **h,** An IIMN subnetwork of the HILIC-IMS-MS data, complemented by the images obtained from the MALDI-IMS-MS analysis. HILIC-IMS-MS features

aligned with an image are shown with a diamond shape. Interactive networks are available through the GNPS web platform.




**Supplementary Fig. 5 | IMS-MS data inspection interfaces in MZmine 3.**

**a,** A screenshot of the interactive LC-IMS-MS visualizer. All charts are connected by their data and by selected *m/z*, RT, and mobility ranges. 1) Summed frame spectrum, 2) EIC and BPC, 3) extracted ion- and base peak mobilograms, 4) frame heatmap, and 5) RT/mobility ion trace. Clicks in any chart change the selected *m/z*, retention time, and mobility ranges and trigger updates of the other visualizers. **b,** Combined screenshots of the IMS-MS and imaging visualizers.

## Methods

### Chemicals and materials

Methanol and acetonitrile were obtained from VWR International GmbH and were of LC-MS grade. HPLC grade methyl *tert*-butyl ether (MTBE) was obtained from Merck KGaA. 98% 2,5-dihydroxybenzoic acid (DHB), ammonium formate (>99%), and trifluoroacetic acid were ordered from Sigma Aldrich. Water was purified by a Milli-Q Academic system (18.2 MΩ cm; 0.2 µm filter; Millipore). Formic acid (LC-MS grade) was obtained from Fisher Scientific.

### Sheep brain tissue preparation

Sheep brain tissue for this study was harvested from Swiss-Alpine sheep (animal license number ZH235/17, Swiss Animal Welfare Act). The details of the study have been described previously.[35] A lipid extract of the brain tissue was prepared following the Matyash protocol.[34] Briefly, 43.7 mg of brain tissue were cut on dry ice and rinsed with sterile filtered ammonium acetate buffer (0.1%, w/v). 750 µL of buffer were added, and the tissue was homogenized with an Ultra Turrax T8 (IKA Labortechnik) on ice. Subsequently, 1.4 mL of methanol, 100 µL of BHT in methanol (65 mM), and 5 mL of MTBE were added, and the mixture was incubated at 100 rpm on ice for 1 h. Afterwards, 1.25 mL of water were added, and the sample was centrifuged at 1000 g. The organic phase was collected and the aqueous phase was re-extracted with 2 mL of the organic phase MTBE/methanol/water (10:3:2.5, *v*/*v*/*v*). After centrifugation, the combined organic phase was dried under a gentle nitrogen stream and redissolved in 600 µL of isopropanol.

### HILIC-IMS-MS

The sample was analyzed using an UltiMate 3000 system with a dual gradient pump (Thermo Scientific) and a timsTOF fleX (Bruker Daltonics). An iHILIC Fusion(+) column (20 × 2.1 mm, 5 µm, 100 Å) (HILICON AB) was employed for lipid class separation. The LC-IMS-$MS^2$ method was adapted from Helmer et al.[33] Briefly, 35 mM ammonium formate in water/acetonitrile (95:5, *v*/*v*, pH 3.5) (A) and acetonitrile (B) were used as eluents. Eluent B was held at 97% until 0.2 min, ramped to 93% at 0.5 min, held until 2.75 min, ramped to 60 % at 8.0 min, held until 11.5 min, ramped to 97 % at 12 min, and re-equilibrated until 18.0 min. The column oven was set to 40 °C. The timsTOF fleX was operated in positive ESI mode, with 2.0 bar nebulizer pressure, 9.0 L/min dry gas, 200°C dry heater, 4000 V capillary voltage, -500 V end plate offset, 360 Vpp funnel 1 RF, 250 Vpp funnel 2 RF, 80 V deflection delta, 5 eV ion energy, 10 eV collision energy, 1000 Vpp collision RF, 5 µs pre pulse storage, 65 µs transfer time. TIMS delta values were set to -20 V (delta 1), -120 V (delta 2), 80 V (delta 3), 100 V (delta 4), 0 V (delta 5), and 100 V (delta 6). The $1/K_0$ (inverse reduced ion mobility) range was set from 0.9 Vs/$cm^2$ to 1.6 Vs/$cm^2$, the mass range was *m/z* 100-1350. $MS^2$ spectra were acquired using the PASEF DDA mode with a collision energy of 40 eV. Ion charge control (ICC) was enabled and set to 7.5 mio counts. The samples were analyzed with TIMS ramp times of 100 ms and 500 ms.

### MALDI-IMS-MS

10 µm thin sections were mounted on indium tin oxide coated glass slides. Bright field microscopic images were acquired BIOREVO BZ-9000 digital microscope (Keyence) in 2 and 10-fold magnification.

150 mg of DHB were dissolved in acetonitrile/water (9:1, v/v) with 0.1 % trifluoroacetic acid. The matrix was applied by a TM-Sprayer (HTX Imaging) with 40 mm nozzle height, 10 psi nitrogen pressure, 60 °C, 0.125 mL/min solvent flow, 1200 mm/min z-arm velocity, 14 passes, a CC moving pattern, 3 mm track spacing, and 0 s drying time. The timsTOF fleX was operated in positive ion mode, with 350 laser shots, 10,000 Hz laser frequency, 50 µm spot size. The TIMS ramp time was set to 350 ms. Other parameters were retained from the HILIC measurements.

**Ion mobility raw data pre-processing**

LC-IMS-MS raw data was recalibrated in DataAnalysis 5.2 (Bruker Daltonics) using in-line infusion of sodium formate and ESI-L low concentration tuning mix (Agilent Technologies Manufacturing) mixed in a 1:1 ratio. Bruker raw data was directly imported from the native *.tdf* file format. MZmine then stores spectral data in temporary files that are memory mapped for efficient data retrieval. Mobility-resolved mass spectra at one retention time are grouped in *frames*. An accumulated mass spectrum representing the whole frame was calculated.

**LC-IMS-MS feature detection workflow**

Construction of extracted ion chromatograms (EICs) and resolving individual features, termed *feature detection,* was performed on the accumulated frame spectra. Ion chromatograms were created with the *ADAP chromatogram builder*.[10] The EICs were then resolved in the RT dimension by the *ADAP continuous wavelet transform* to create individual features defined by an *m/z* and retention time range.

The mobility-resolved data was searched for data points within the detected *m/z* and retention time limits. This created three-dimensional *ion mobility traces.* The ion mobility traces represent the intensity distribution of a feature in mobility and retention time dimension within an *m/z* range. Accumulated EICs (in mobility dimension) and accumulated mobilograms (in retention time dimension) were then created from the data in the ion mobility traces. The mobilograms were finally resolved in mobility dimension using feature detection algorithms such as *local minimum search* (100 ms TIMS ramp) or *ADAP continuous wavelet transform* (500 ms TIMS ramp).[10] Feature list deisotoping, alignment, and gap filling of ion mobility data were performed within given *m/z*, RT, and mobility tolerances. PASEF $MS^2$ spectra were assigned to the features based on their RT and ion mobility ranges. Results were exported to FBMN on the GNPS web service.

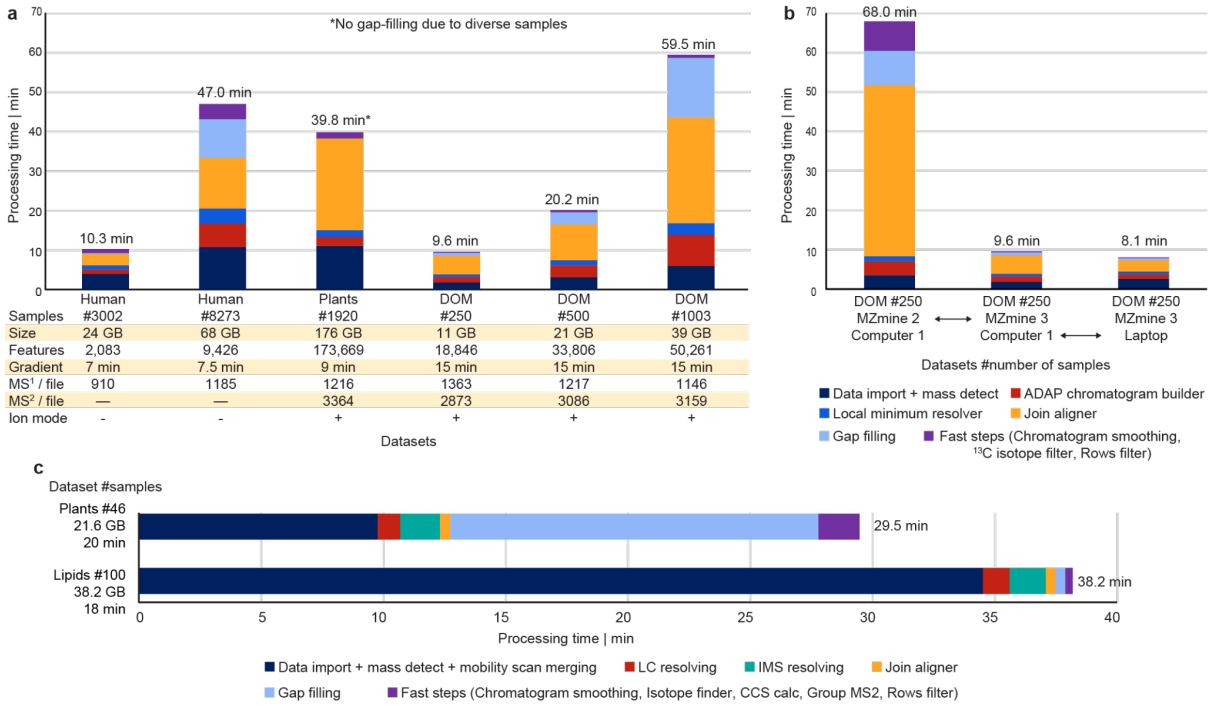
**MALDI-IMS-MS imaging feature detection workflow**

Images were created with an adapted ADAP chromatogram builder algorithm.[10] For every detected *m/z*, an image was created based on accumulated frame spectra. Afterwards, the mobility-resolved sub-scans were queried to build mobility-resolved images. The images were resolved in IMS dimension by the feature detection algorithm *local minimum search*.[10] The mobility resolved images were aligned to an LC-IMS-MS feature list by *m/z* and mobility values, retaining LC-based annotations.

## Supplementary Note 3

### Performance stress test

Processing large-scale MS datasets requires significant computational resources. MZmine 3 improves the practical scalability of analysis from hundreds to tens of thousands of samples. To demonstrate the scalability of data analysis in MZmine 3, we processed various public MS datasets on common hardware. The stress test resulted in high sample throughput, where the mean processing times only took 0.1% to 0.3% of the total data acquisition time for 6 different LC-MS datasets. A breakdown of the analysis (**Supplementary Fig. 6**) shows that parallelization of all computational steps removed significant bottlenecks. The alignment step, which creates a matrix of samples × features, remains the most time-consuming step for larger datasets. For example, the alignment of 250 complex dissolved organic matter (DOM) samples from ocean waters was reduced by 89% (from 43.4 min to 4.7 min), processed on data analysis computer 1 by MZmine 2.53 and MZmine 3.2.0, respectively (refer to **Methods** for hardware specifications). Due to the increased complexity of ion mobility-resolved data, LC-IMS-MS datasets demand even more efficient processing. MZmine 3 performed the complete four-dimensional feature detection workflow in preparation of molecular networking in a fraction of the data acquisition time, e.g., in 29.5 min for 40 *Piper* plant extracts. For large datasets, MZmine 3 was able to process data from 8273 fecal LC-$MS^2$ samples in 47.0 min and data from 1003 DOM LC-$MS^2$ samples, which are significantly more complex, in 59.5 min.




**Supplementary Fig. 6 | Performance evaluation of MZmine 3.**

**a,** MZmine 3.2.0 processed six different LC-MS datasets of varying study targets and sizes. On average, processing took 0.1% to 0.3% of the data acquisition times. **b,** Processing of the same 250 sample dataset in MZmine 2 reveals a 7x speed up in MZmine 3. Analysis on a laptop and on the data analysis computer 1 gave similar performance (see specifications in Methods). Running MZmine 2.53 on other datasets, with similar parameters, revealed severe limitations; MZmine 2.53 crashed on all larger studies during chromatogram building or gap-filling. **c,** Benchmarks of the LC-IMS-MS workflow on 2 datasets with 40 samples and 100 samples (the dataset size was increased by copying data files, see **Methods**) resulted in processing times that are a fraction of the total data acquisition time. See hardware specifications in the Methods section.

## Methods

MZmine 3 was written with performance and throughput in mind. We performed comparative tests (MZmine 2.53 → 3.2.0) on six datasets with various study backgrounds. A large public environmental study (MSV000090079) of DOM, measured by LC-MS$^2$ with top-5 DDA, was split into three datasets with 250, 500, and 1,003 samples, using the largest mzML files for the smaller subsets. Two human cohort studies (MSV000088054) with 3,008 and 8,273 samples were acquired by LC-MS$^2$. A set of 1,600 diverse plant extracts, blanks, and quality controls created the last public dataset (MSV000087728) with 1,920 LC-MS$^2$ samples measured with top-3 DDA.

Data processing workflow consisted of the main steps that were optimized for each dataset separately, considering the MS scan acquisition rate, the LC peak width, and the number of spectra-per-feature. The DOM datasets, which show high spectral complexity and have an emphasis on annotation and molecular networking based on the top-5 DDA method, were processed with low noise levels and few feature constraints. The human cohort studies were processed with higher feature constraints, considering the higher MS acquisition speed and higher number of data points per

feature. The workflow for the plant dataset used chromatogram smoothing to improve the feature shapes before resolving, and skipped the gap-filling step given the diversity of samples. This step is usually performed as a secondary informed feature detection, preparing the feature list for statistical analysis. With the diversity of tested specimens and the top-3 DDA method, this is a perfect example of a dataset suited for molecular networking for novel compound discovery. All parameters and steps are provided as batch files in **Supplementary File S1.**

First, performance tests were performed comparing the processing speed in MZmine 3.2.0 to the data acquisition speed, as summarized in **Supplementary Fig. 6a**. On data analysis computer 1 with 16 threads and 128 GB memory and on a laptop with 16 threads and 32 GB memory (system specifications below), processing took 0.1%–0.3% of the time needed to acquire the LC-MS data of the whole study, not accounting for delays between sample injections. With all steps running in parallel, the throughput of samples scales with the increasing complexity of more features being aligned into one feature list. Alignment and gap-filling used to be bottlenecks of the feature detection workflow, however, with the current optimization, they produce reproducible results in a high throughput manner. This represents a 7-fold increase in processing speed in comparison with the previous generation, MZmine 2 (**Supplementary Fig. 6b**).

Performance of the LC-IMS-MS workflow was tested on two public datasets of 46 *Piper* plant extracts (21.6 GB) and sheep brain. To simulate a larger dataset, the 3 sheep brain samples and 2 blanks were copied until reaching 10 blanks and 90 samples, totaling 100 LC-IMS-MS files (38.2 GB). MZmine 3 imported, memory-mapped spectral data, and extracted LC-IMS-MS resolved ion features in a fraction of the total data acquisition time on data analysis computer 2 (**Supplementary Fig. 6c)**.

MZmine 3.2.0 was built and packaged with Java Development Kit (JDK) 17.0.4 (Temurin), eliminating the need to install a Java Virtual Machine on the target machine. MZmine 3.3.0 was packaged with the updated JDK 19.0.1 (Temurin).

**Hardware specifications**

The systems used to process LC-MS studies were:

- Data analysis computer 1, Intel Core i7-9700K, 8 cores, 16 threads, 128GB RAM, SSD storage
- Laptop, Dell XPS15 9510, Intel Core i7-11800H, 8 cores, 16 threads, 32GB RAM, SSD storage

The system used to process LC-IMS-MS and MALDI-IMS-MS studies was:

- Data analysis computer 2, Dell Optiplex 7080, Intel Core i9 10900K, 10 cores, 20 threads, 64 GB RAM, NVMe SSD storage

Performance tuning and options are described in the MZmine documentation:

https://mzmine.github.io/mzmine_documentation/performance.html

## Supplementary References


1. Treviño, V. et al. *J. Mass Spectrom.* **50**, 165–174 (2015).
2. Korf, A. et al. *Rapid Commun. Mass Spectrom.* **32**, 981–991 (2018).
3. Korf, A., Jeck, V., Schmid, R., Helmer, P.O. & Hayen, H. *Anal. Chem.* **91**, 5098–5105 (2019).
4. Korf, A. et al. *Sci. Rep.* **10**, 767 (2020).
5. Korf, A., Fouquet, T., Schmid, R., Hayen, H. & Hagenhoff, S. *Anal. Chem.* **92**, 628–633 (2020).
6. Nothias, L.-F. et al. *Nat. Methods* **17**, 905–908 (2020).
7. Schmid, R. et al. *Nat. Commun.* **12**, 3832 (2021).
8. Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. *BMC Bioinformatics* **11**, 395 (2010).
9. Pluskal, T., Uehara, T. & Yanagida, M. *Anal. Chem.* **84**, 4396–4403 (2012).
10. Myers, O.D., Sumner, S.J., Li, S., Barnes, S. & Du, X. *Anal. Chem.* **89**, 8696–8703 (2017).
11. Smirnov, A., Jia, W., Walker, D.I., Jones, D.P. & Du, X. *J. Proteome Res.* **17**, 470–478 (2018).
12. Smirnov, A. et al. *Anal. Chem.* **91**, 9069–9077 (2019).
13. Du, X., Smirnov, A., Pluskal, T., Jia, W. & Sumner, S. *Computational Methods and Data Analysis for Metabolomics* 25–48 (2020).
14. Shen, X. et al. *Nat. Commun.* **13**, 1–12 (2022).
15. Shen, X. et al. *Nat. Commun.* **10**, 1–14 (2019).
16. Wang, M. et al. *Nat. Biotechnol.* **34**, 828–837 (2016).
17. Gloaguen, Y., Kirwan, J.A. & Beule, D. *Anal. Chem.* **94**, 4930–4937 (2022).
18. El Abiead, Y., Milford, M., Salek, R.M. & Koellensperger, G. *Bioinformatics* (2021).doi:10.1093/bioinformatics/btab231
19. Olivon, F. et al. *Anal. Chem.* **90**, 13900–13908 (2018).
20. Delabriere, A., Warmer, P., Brennsteiner, V. & Zamboni, N. *Anal. Chem.* **93**, 15024–15032 (2021).
21. Zdouc, M.M. et al. *bioRxiv* 2022.12.21.521422 (2022), doi:10.1101/2022.12.21.521422
22. Gaudry, A. et al. *Front Bioinform* **2**, 842964 (2022).
23. Quiros-Guerrero, L.-M. et al. *bioRxiv* 2022.08.25.505324 (2022), doi:10.1101/2022.08.25.505324
24. Chen, L. et al. *Nat. Methods* **18**, 1377–1385 (2021).
25. Kind, T. & Fiehn, O. *BMC Bioinformatics* **8**, 105 (2007).
26. Damiani, T. et al. *Anal. Chem.* **95**, 287–303 (2023).
27. Dührkop, K. et al. *Nat. Methods* **16**, 299–302 (2019).
28. Ludwig, M. et al. *Nature Machine Intelligence* **2**, 629–641 (2020).
29. Dührkop, K., Shen, H., Meusel, M., Rousu, J. & Böcker, S. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 12580–12585 (2015).
30. Dührkop, K. et al. *Nat. Biotechnol.* **39**, 462–471 (2021).
31. Liebisch, G. et al. *J. Lipid Res.* **61**, 1539–1555 (2020).
32. Köfeler, H.C. et al. *J. Lipid Res.* **62**, 100138 (2021).
33. Helmer, P.O. et al. *Anal. Chem.* **93**, 2135–2143 (2021).
34. Matyash, V., Liebisch, G., Kurzchalia, T.V., Shevchenko, A. & Schwudke, D. *J. Lipid Res.* **49**, 1137–1146 (2008).
35. Richter, H. et al. *Radiology* **301**, 637–642 (2021).