



Spectra Processing, Compound Annotation, Functional Insight and Causal Analysis using **MetaboAnalyst 6.0**

Jianguo (Jeff) Xia, Associate Professor

Canada Research Chair in Bioinformatics & Big Data Analytics

jeff.xia@mcgill.ca | www.xialab.ca

McGill University, Canada



XiaLab.ca

Empowering researchers through trainings, tools, and AI



Schedule

Part I: 2:15 pm – 4:15 pm

- **2:15 – 2:30:** General introduction
- **2:30 - 3:00:** Untargeted metabolomics
 - ✓ LC-MS & MS/MS spectral processing
 - ✓ From peaks to functions
- **3:05 – 3:25:** Live demo
- **3:25 – 4:15:** Hands on practice

Part II: 4:30 p.m. – 6:30 p.m.

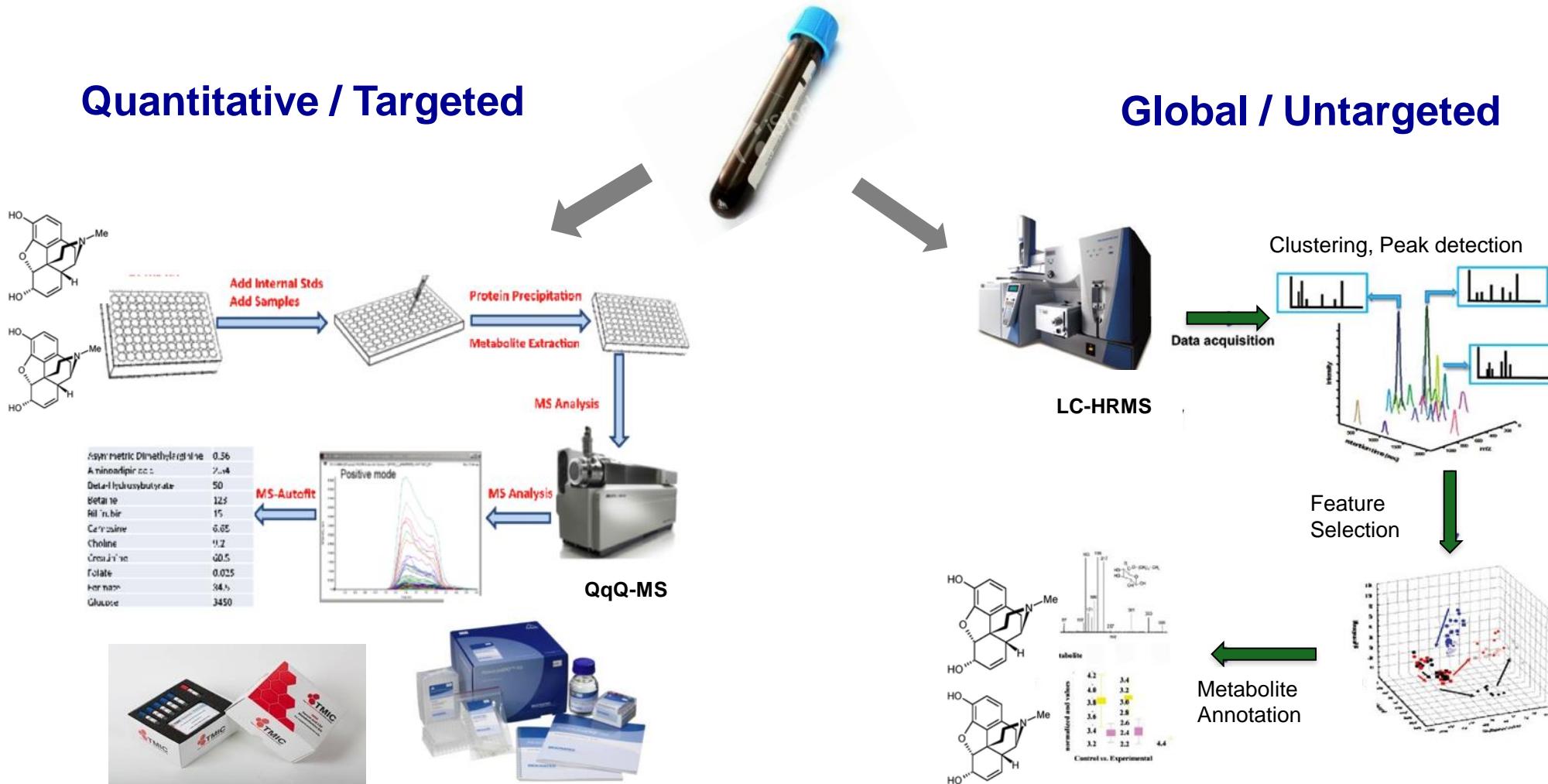
- **4:30 – 5:15:** Background
 - ✓ Statistical analysis
 - ✓ Causal analysis
- **5:15 – 5:35:** Live demo
- **5:40 – 6:20:** Hands on practice
- **6:20 – 6:30:** Summary & discussion

Github Repository

- https://github.com/xia-lab/Metabolomics_2024
- Slides (in PDF format);
- Example data;
- Reference literatures;
- Contact information.

LC-MS Spectra Processing

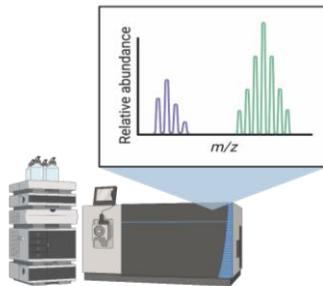
Two routes to metabolomics (LC-MS)



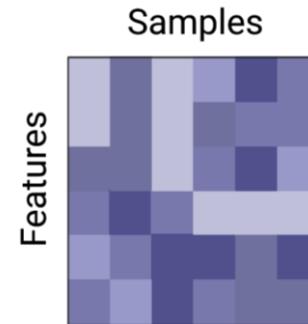
Untargeted Metabolomics Workflow

Quantitative analysis (LC-MS)

- Biological replicates
 - 10 control vs 10 disease
- Pooled QCs
- Blanks



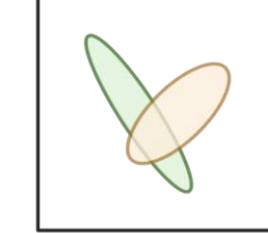
Raw data processing



Compound identification (MS/MS)

- 3 technical replicates from pooled QCs
 - Aliquots from all samples (better signals & coverage)
 - Spectra consensus to improve MS2 quality

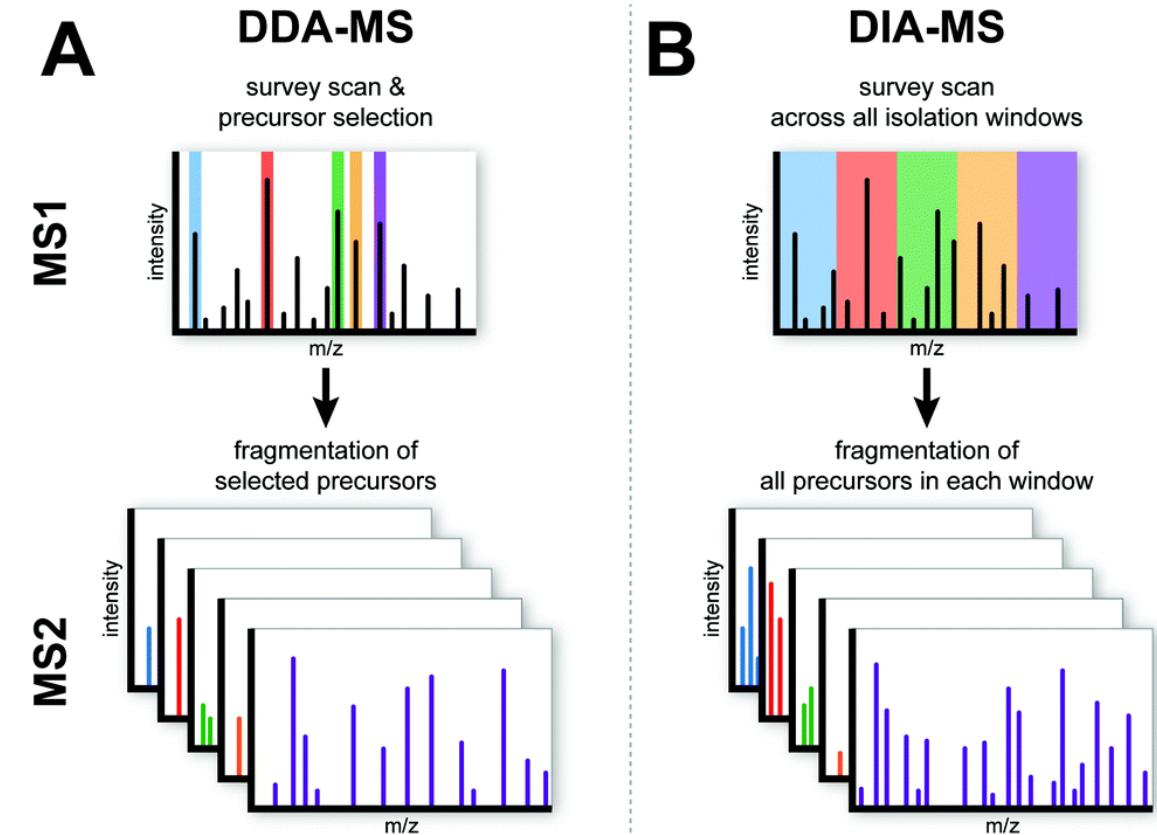
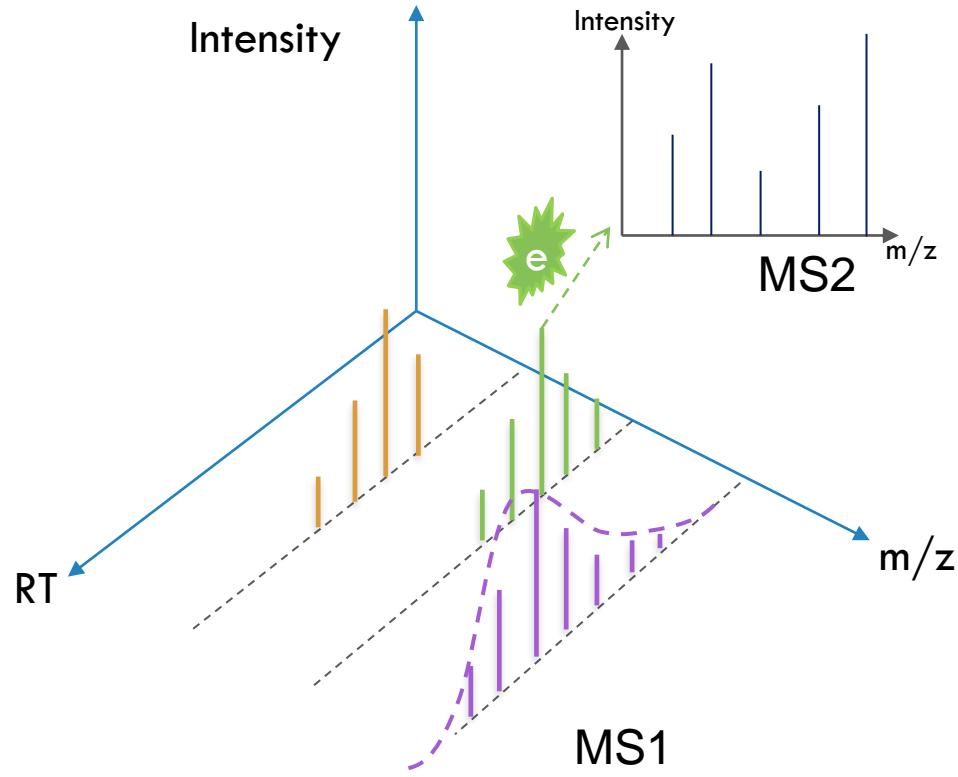
Statistical analysis



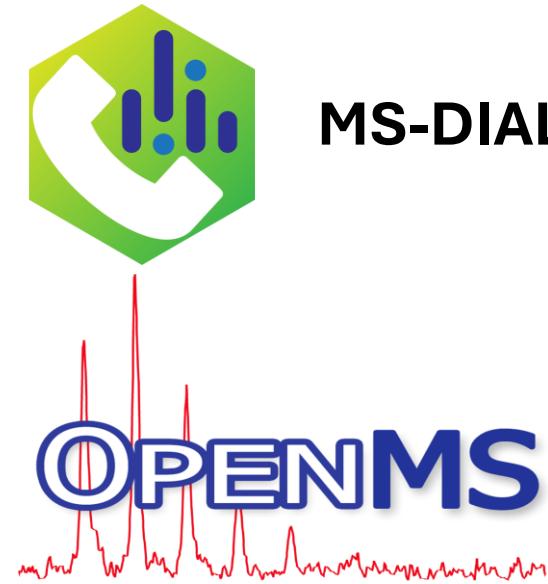
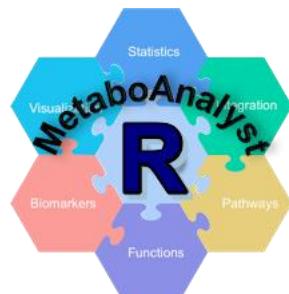
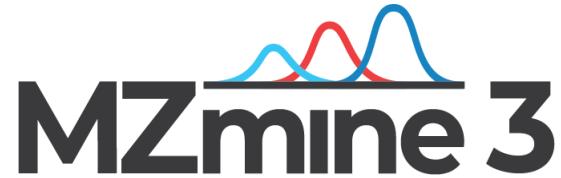
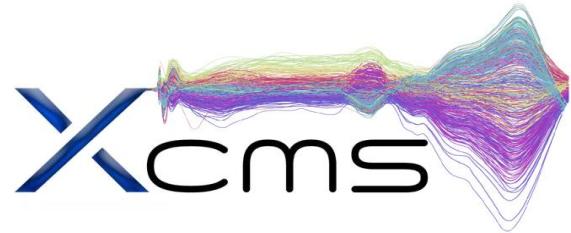
?

Functional interpretation

LC-MS & MS/MS



Open-source software for spectra processing



Asari

MetaboAnalystR & Asari

nature communications



Article

<https://doi.org/10.1038/s41467-023-39889-1>

Trackable and scalable LC-MS metabolomics data processing using asari

Received: 4 June 2022

Shuzhao Li^{1,2}✉, Amnah Siddiqi¹, Maheshwor Thapa¹, Yuanye Chi¹ & Shujian Zheng¹

Accepted: 30 June 2023

Published online: 11 July 2023

Check for updates

Significant challenges remain in the computational processing of data from liquid chromatography-mass spectrometry (LC-MS)-based metabolomic experiments into metabolite features. In this study, we examine the issues of provenance and reproducibility using the current software tools. Inconsistency among the tools examined is attributed to the deficiencies of mass alignment and controls of feature quality. To address these issues, we develop the open-source software tool asari for LC-MS metabolomics data processing. Asari is designed with a set of specific algorithmic framework and data structures, and all steps are explicitly trackable. Asari compares favorably to other tools in feature detection and quantification. It offers substantial improvement in computational performance over current tools, and it is highly scalable.

Metabolomics holds the promise to comprehensively measure and quantify small molecules in biological systems. Since the chemistry of these small molecules underlies most aspects of life science, metabolomics is recognized as critical to support missions in biomedical

disagreement between XCMS and OpenMS. It is common knowledge among users that the results also vary wildly based on parameter settings. Significant community efforts were spent on parameter optimization of XCMS^{13–18}. However, these post-hoc adjustments do not

nature communications



Article

<https://doi.org/10.1038/s41467-024-48009-6>

MetaboAnalystR 4.0: a unified LC-MS workflow for global metabolomics

Received: 15 September 2023

Zhiqiang Pang¹, Lei Xu¹, Charles Viau¹, Yao Lu², Reza Salavati¹, Niladri Basu¹ & Jianguo Xia^{1,2}✉

Accepted: 18 April 2024

Published online: 01 May 2024

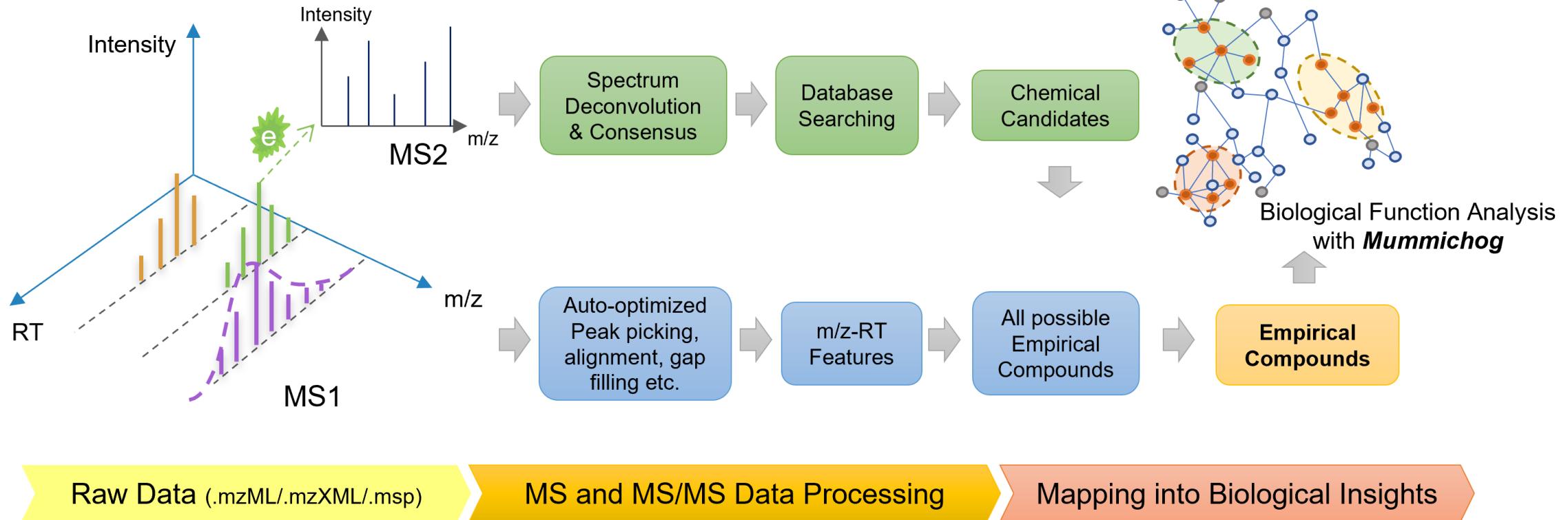
Check for updates

The wide applications of liquid chromatography - mass spectrometry (LC-MS) in untargeted metabolomics demand an easy-to-use, comprehensive computational workflow to support efficient and reproducible data analysis. However, current tools were primarily developed to perform specific tasks in LC-MS based metabolomics data analysis. Here we introduce MetaboAnalystR 4.0 as a streamlined pipeline covering raw spectra processing, compound identification, statistical analysis, and functional interpretation. The key features of MetaboAnalystR 4.0 includes an auto-optimized feature detection and quantification algorithm for LC-MS1 spectra processing, efficient MS2 spectra deconvolution and compound identification for data-dependent or data-independent acquisition, and more accurate functional interpretation through integrated spectral annotation. Comprehensive validation studies using LC-MS1 and MS2 spectra obtained from standards mixtures, dilution series and clinical metabolomics samples have shown its excellent performance across a wide range of common tasks such as peak picking, spectral deconvolution, and compound identification with good computing efficiency. Together with its existing statistical analysis utilities, MetaboAnalystR 4.0 represents a significant step toward a unified, end-to-end workflow for LC-MS based global metabolomics in the open-source R environment.

LC-MS spectra processing

- Identify, quantify, and align all possible features (peaks) across samples
- Output: A table of features (RT, m/z) with their quantitative information for subsequent statistical analysis

LC-MS & MS/MS processing in MetaboAnalyst



LC-MS1

- MetaboAnalystR
- Asari

MS2 - MetaboAnalystR

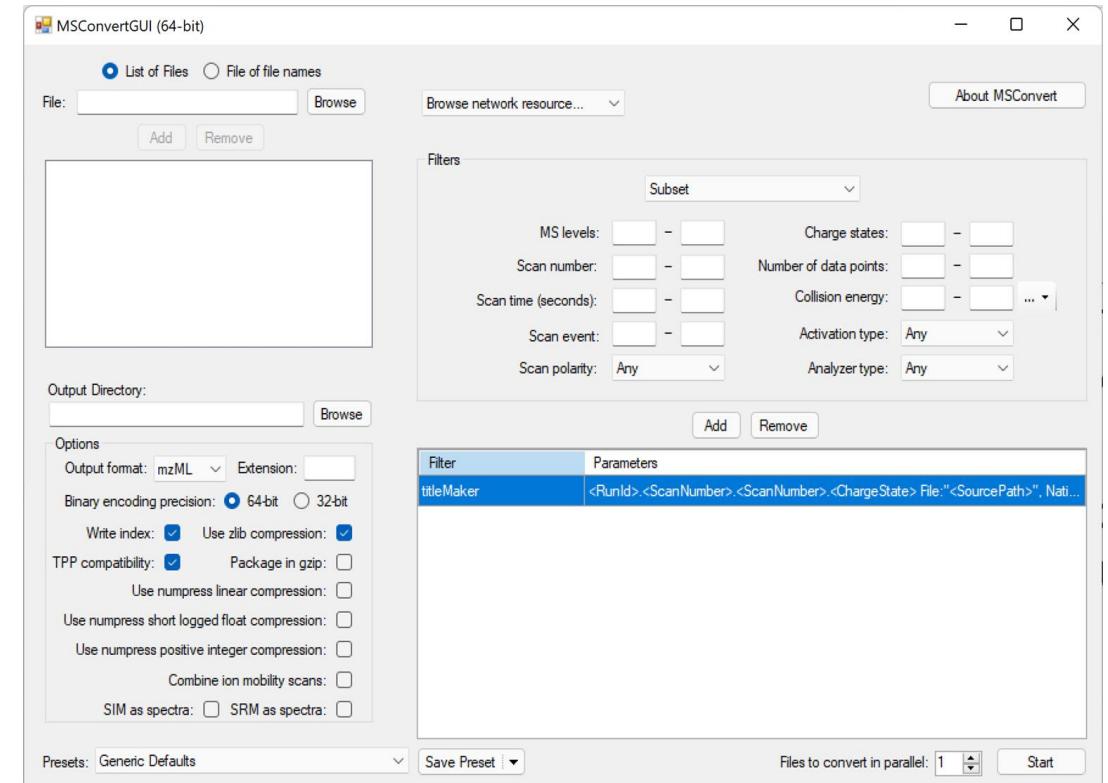
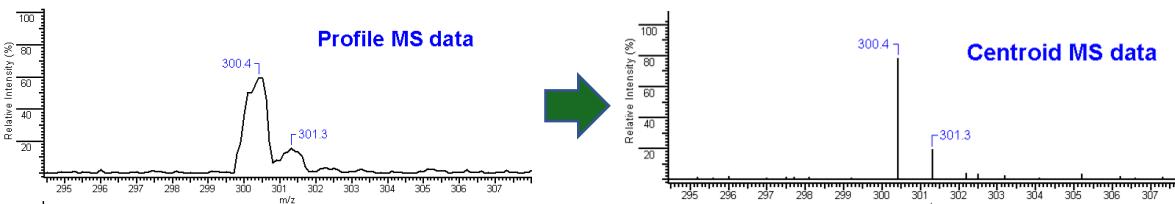
- DDA
- SWATH-DIA

MetaboAnalystR

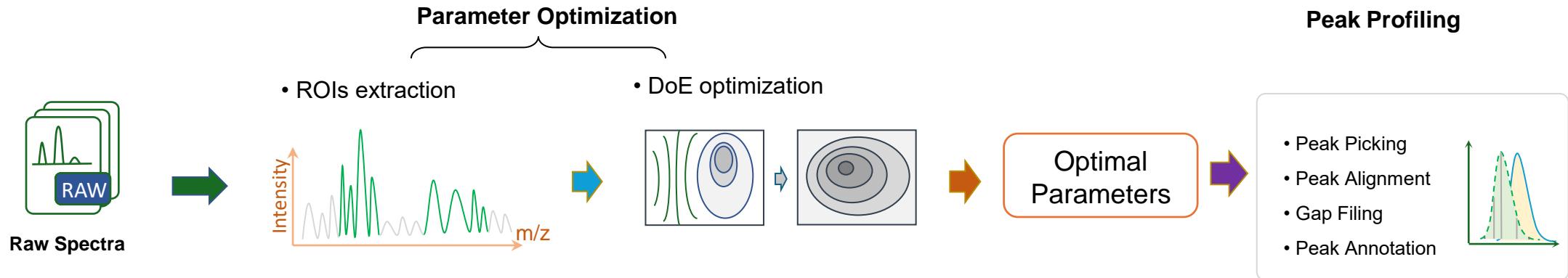
- Mummichog
- GSEA

Centroid mode & in open-source format

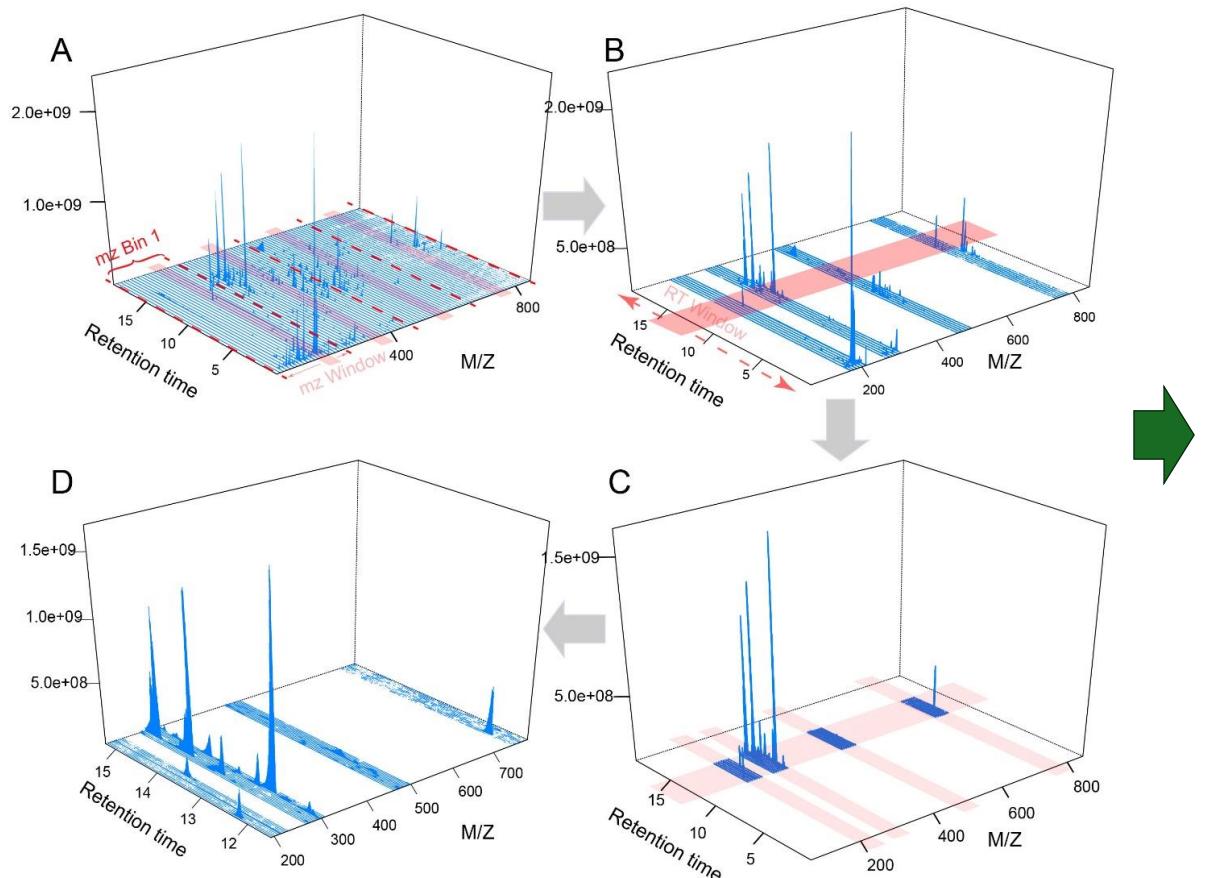
- The vendor raw spectra data is usually in profile format, which is redundant for regular LC-MS based metabolomics analysis;
- We need to convert the MS data into centroid mode to condense the Gaussian Profile peaks into centroids.
- Open formats (.mzML/ etc.)..



Auto-optimization workflow



ROI extraction and DOE-based optimization



DoE -- Central composite design

Order	Peakwith_min	Peakwith_max	mzdiff	snthresh	bw
1	-1	-1	-1	-1	-1
2	1	-1	-1	-1	-1
3	-1	1	-1	-1	-1
...
43	0	0	0	0	1
44	0	0	0	0	0

44 runs

3 level for every parameters (-1, 0, 1)

The most important parameters are evaluated with 44 DoE runs Instead of $3^8 = 6561$ one-variable-at-a-time runs.

Peak Quality Score (QS)

$$QS = \frac{RP^{3/2}}{'all\ peaks' - LIP} * GR^2 * QcoE$$

Relative reliable peaks ratio
(identified by their isotopes)

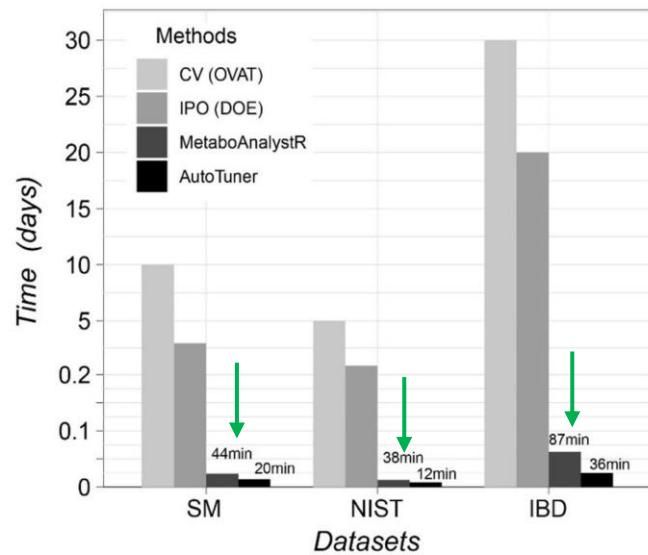
A co-efficient describing
the stability of a grouped
feature

Gaussian peaks
ratio

The diagram illustrates the components of the Peak Quality Score (QS) equation. The first term, $RP^{3/2}$, is highlighted with a red dashed box and has a green arrow pointing to the text 'Relative reliable peaks ratio (identified by their isotopes)'. The second term, $'all\ peaks' - LIP$, is highlighted with a red dashed box and has an orange arrow pointing to the text 'Gaussian peaks ratio'. The third term, GR^2 , is highlighted with a green dashed box. The fourth term, $QcoE$, is highlighted with a blue dashed box and has a blue arrow pointing to the text 'A co-efficient describing the stability of a grouped feature'.

Benchmarking: better identification

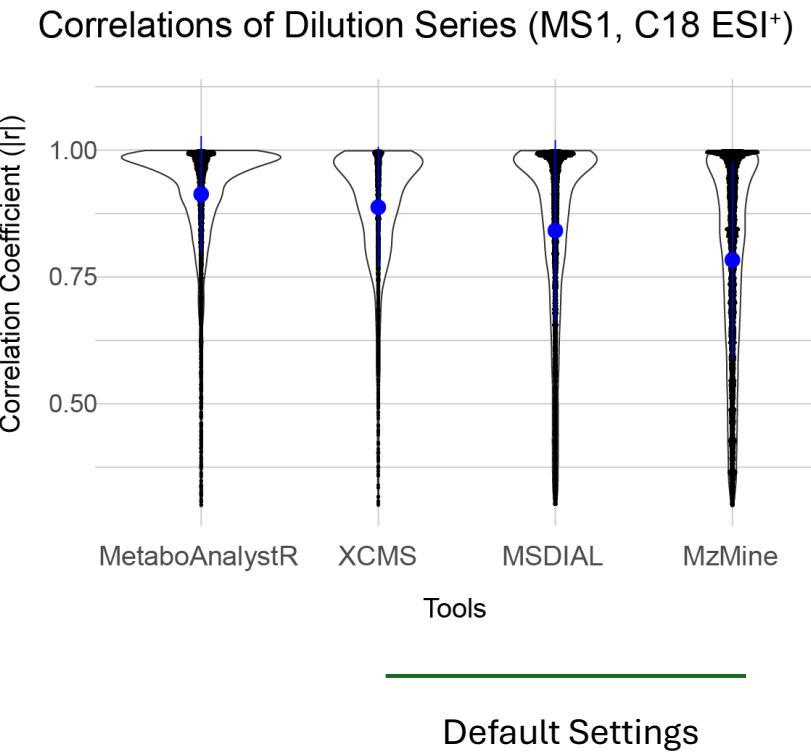
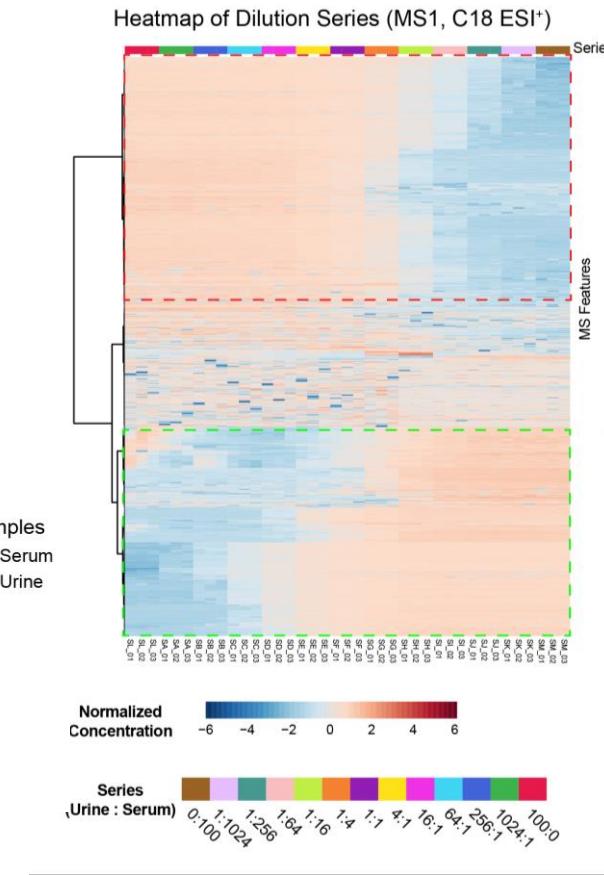
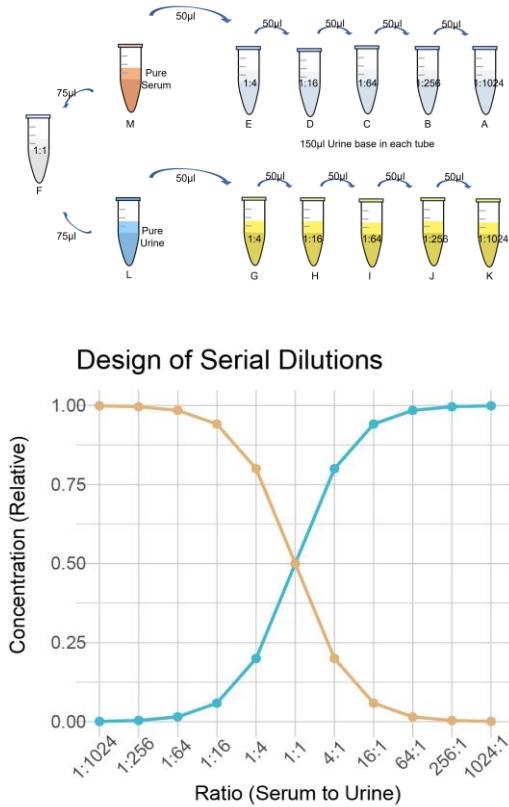
Benchmark with a standard mixture containing **1100** common compounds.



A mixture solution of 838 standards

Methods	Total Peaks	True Peaks	Quantified	Gaussian Peak Ratio
Default centWave	16,896	382	350	47.8%
IPO	24,346	744	663	52.0%
AutoTuner	25,517	664	603	40.5%
MetaboAnalystR	18,044	799	754	64.4%

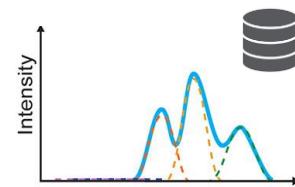
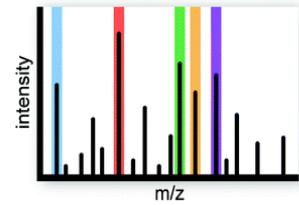
Benchmarking: better quantification



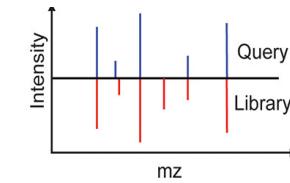
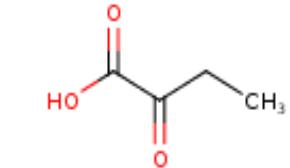
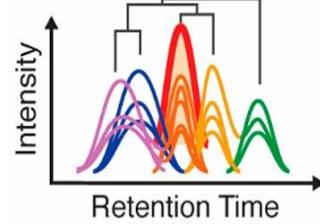
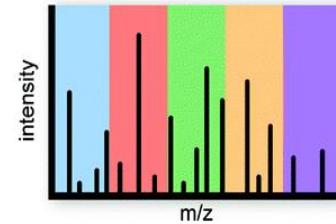
MS/MS spectra processing using MetaboAnalyst

MS2 spectra processing

DDA

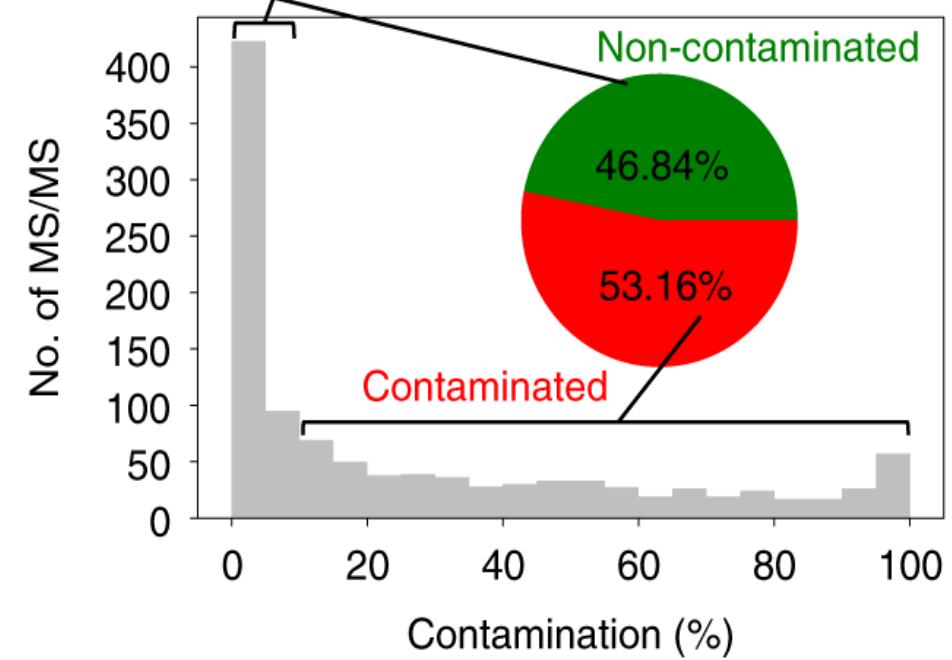
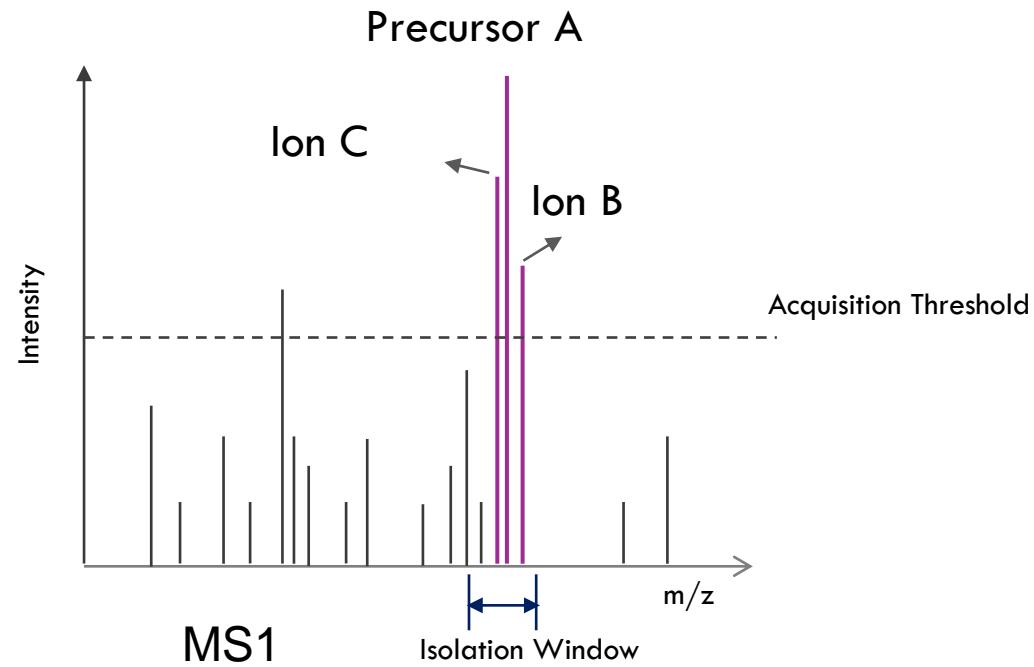


DIA

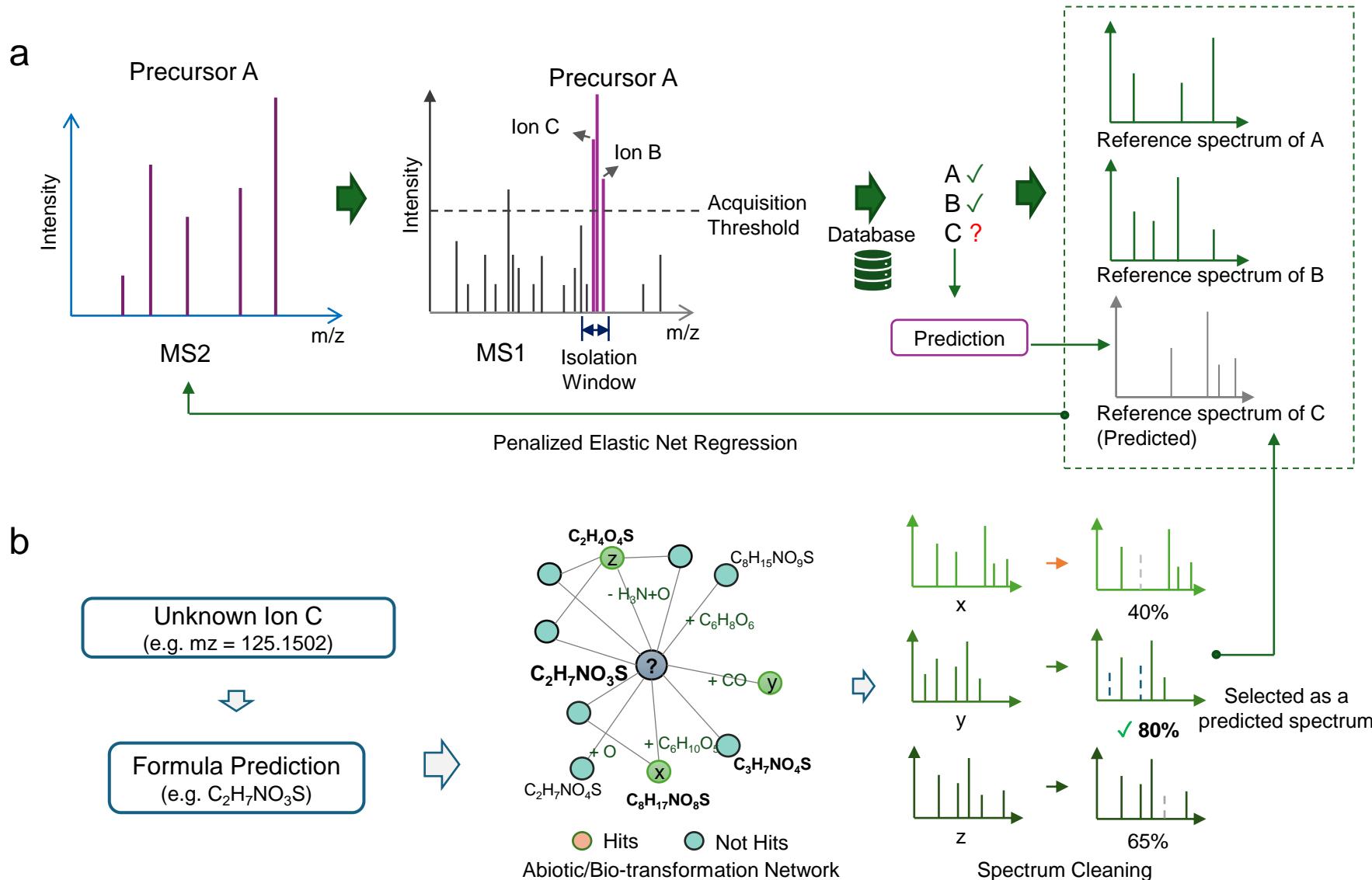


Spectra deconvolution + compound identification

Ion contaminations are prevalent in DDA



DDA spectra deconvolution



Penalized Linear Regression Model

- 1) Minimize the residue after linear regression.
- 2) Give more penalty to the “Predicted Candidate” from network-based approach;
- 3) 11 alpha (values) * 10 lambda (values) are permuted together as 110 parameters combination to minimize the residue -> automatically adaptive model.

$$\min_{\beta_0, \beta} \left(\frac{1}{2N} \sum_{i=1}^N (y_i - \beta_0 - \mathbf{x}_i^T \beta)^2 + \lambda P_\alpha(\beta) \right),$$

where

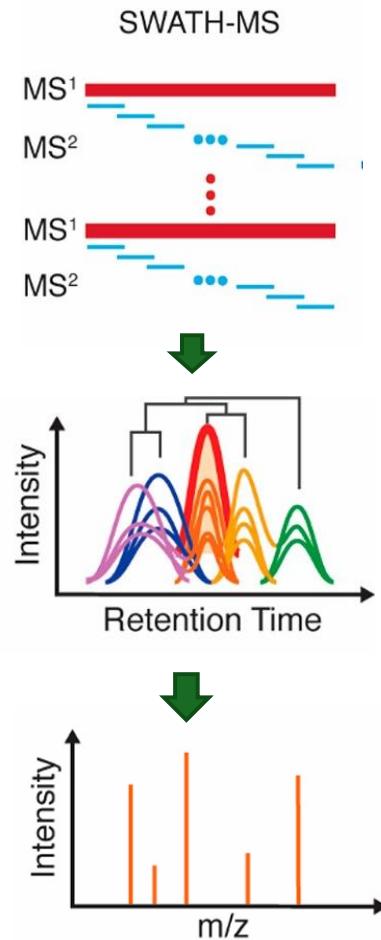
$$P_\alpha(\beta) = \frac{(1-\alpha)}{2} \|\beta\|_2^2 + \alpha \|\beta\|_1 = \sum_{j=1}^p \left(\frac{(1-\alpha)}{2} \beta_j^2 + \alpha |\beta_j| \right).$$

Elastic net is the same as lasso when $\alpha = 1$. As α shrinks toward 0, elastic net approaches [ridge](#) regression. For other values of α , the penalty term $P_\alpha(\beta)$ interpolates between the L^1 norm of β and the squared L^2 norm of β .

$$(4) \quad \nabla_{(\mu, \beta)} \mathcal{L}_{PS}(\mu, \beta) = 2 \left[-\frac{1}{n} \sum_{i=1}^n r_i(\mu, \beta) w_i(\mu, \beta) \begin{pmatrix} 1 \\ \mathbf{x}_i \end{pmatrix} + \frac{\lambda_S}{2} \begin{pmatrix} 0 \\ \nabla_\beta P_\alpha(\beta) \end{pmatrix} \right],$$

where $P_\alpha(\beta) = \frac{1}{2}(1-\alpha)\|\beta\|_2^2 + \alpha\|\beta\|_1$ is the elastic net penalty, $r_i(\mu, \beta) = y_i - \mu - \mathbf{x}_i^T \beta$ are the residuals and the weights $w_i(\mu, \beta)$ are proportional to $\rho'(\tilde{r}_i(\mu, \beta)) / \tilde{r}_i(\mu, \beta)$ where $\tilde{r}_i(\mu, \beta) = r_i(\mu, \beta) / \sigma(\mu, \beta)$.

SWATH-DIA data processing in MetaboAnalyst



Adapted from DecoMetDIA, but
>10X faster!

Targeted MS/MS Spectrum

MS2 EIC Peaks

Clustering based on
peaks similarity & RT
information

Select model peaks of
different clusters

Model peaks based
on deconvolution of all
MS2 EICs

MS/MS Peaks
Reconstruction

Pseudo MS/MS Spectrum
(Deconvoluted peaks)

MS2 Databases

- HMDB (Experimental + Predicted [CFM-ID])
- GNPS
- MassBank
- MINE
- MoNA
- RIKEN
- ReSpect
- Vaniya Natural Product
- MS-DIAL
- LipidBlast
- BMDMS (Positive mode only)



MoNA - MassBank of North America



MS2 Databases in MetaboAnalyst 6

Name	Records	Unique compounds	Size (MS2 neutral loss)
Complete library	10,420,215	1,551,012	7.2 6.4GB
Pathway library	172,370	3456	138.2 94.1MB
Biology library	864,386	49,055	744.0 491.0MB
Exposome library	1,883,828	106,387	1.5 1.1GB
Lipid library	3,221,409	878,220	1.9 1.1GB

Database SQLite files: <https://metaboanalyst.ca/docs/Databases.xhtml>

Open-source schema

ID	CompoundName	DBID	PrecursorMZ	PrecursorType	Formula	Smiles	InchiKey	InstrumentType	CollisionEnergy	RetentionTime	Ontology	NumberOfPeak	MS2Peaks
Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1 Pyruvic acid	BMDM...		89.02 [M+H]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		7 68.952	2...
2 Pyruvic acid	BMDM...		111.01 [M+Na]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		4 110.975	69...
3 Pyruvic acid	BMDM...		89.02 [M+H]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		18 68.452	21...
4 Pyruvic acid	BMDM...		111.01 [M+Na]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		4 110.975	64...
5 Pyruvic acid	BMDM...		89.02 [M+H]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		21 67.951	7...
6 Pyruvic acid	BMDM...		111.01 [M+Na]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		5 110.975	76...
7 Pyruvic acid	BMDM...		89.02 [M+H]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		20 68.452	14...
8 Pyruvic acid	BMDM...		111.01 [M+Na]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		7 83.049	10...
9 Pyruvic acid	BMDM...		89.02 [M+H]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		19 68.452	77...
10 Pyruvic acid	BMDM...		111.01 [M+Na]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		8 83.049	15...
11 Pyruvic acid	BMDM...		89.02 [M+H]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		19 68.452	28...
12 Pyruvic acid	BMDM...		111.01 [M+Na]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		6 110.975	94...
13 Pyruvic acid	BMDM...		89.02 [M+H]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		21 67.951	12...
14 Pyruvic acid	BMDM...		111.01 [M+Na]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		4 110.975	74...
15 Pyruvic acid	BMDM...		89.02 [M+H]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		20 67.951	10...
16 Pyruvic acid	BMDM...		111.01 [M+Na]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		5 110.975	31...
17 Pyruvic acid	BMDM...		89.02 [M+H]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		18 67.951	6...
18 Pyruvic acid	BMDM...		111.01 [M+Na]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		7 110.975	29...
19 Pyruvic acid	BMDM...		89.02 [M+H]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		20 68.452	25...
20 Pyruvic acid	BMDM...		111.01 [M+Na]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		6 110.975	65...
21 Pyruvic acid	BMDM...		89.02 [M+H]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		22 68.452	48...
22 Pyruvic acid	BMDM...		111.01 [M+Na]+	C3H4O3	O=C(O)C...	LCTONWC...	Orbitrap	10.0		2.2 Alpha-keto...		3 110.975	88...

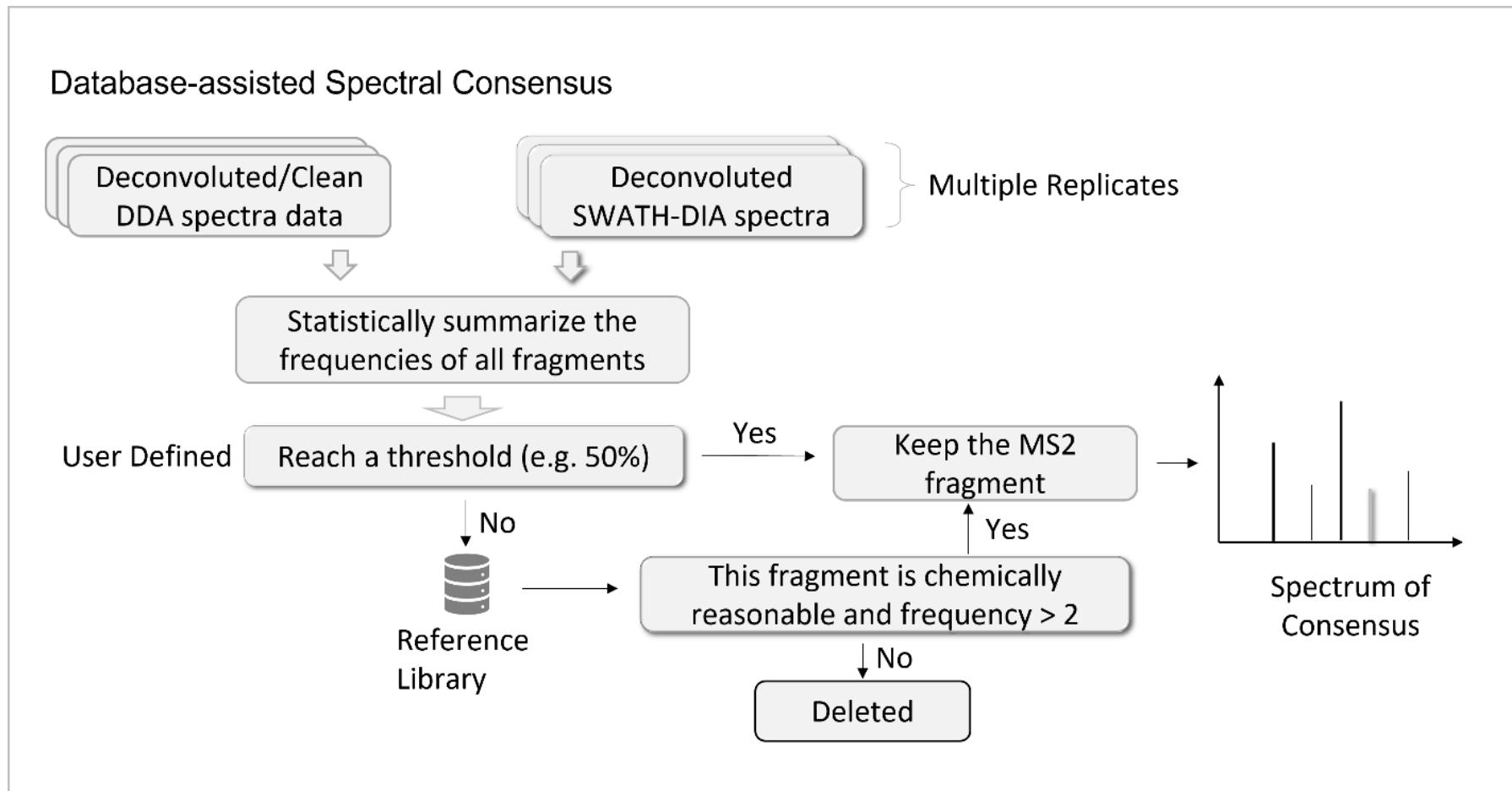
Spectral Similarity Evaluation

$$\frac{MS2\ Similarity + MS1\ Similarity + RT\ Similarity + 0.5 \times Isotope\ similarity}{3.5} \times 100$$

- MS2 similarity measures
 - Dot product
 - Spectral entropy
- MS1 and RT similarity
 - Based on exponential distribution
 - RT similarity is not considered by default
- Isotope similarity
 - Adapted from MS-DIAL
 - Isotope elements considered here include carbon (C^{13}), hydrogen (H^2), nitrogen (N^{15}), oxygen (O^{17} , O^{18}), and sulfur (S^{33} , S^{34})

- The matching scores range between 0 and 100, where 0 indicates no matching and 100 indicates a perfect match.
- The top N candidates can be exported as the database search results.
- If the matching score is below 10, can optionally perform a neutral loss scan

Spectral consensus



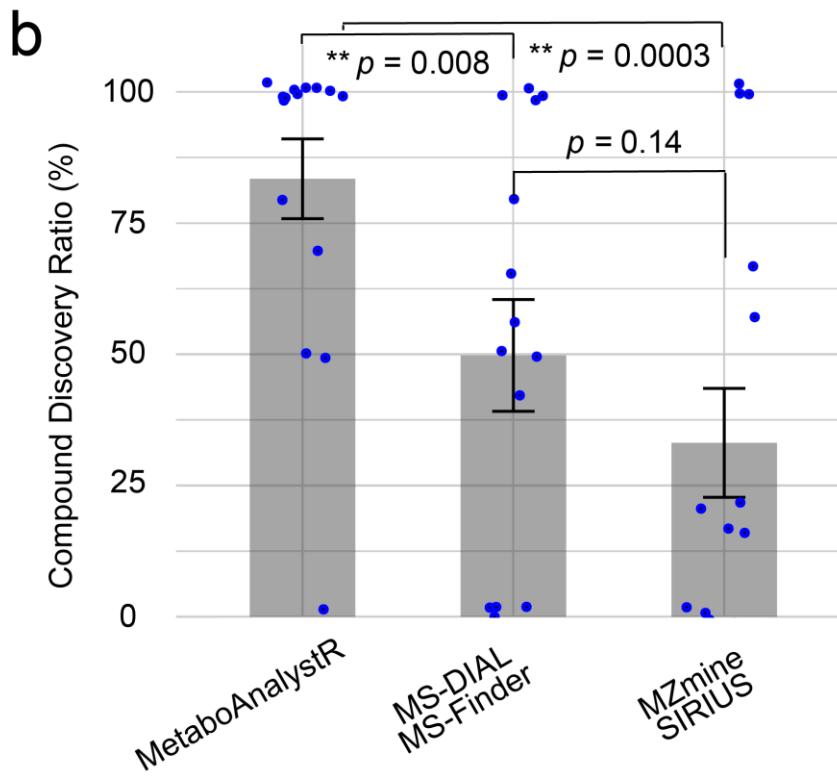
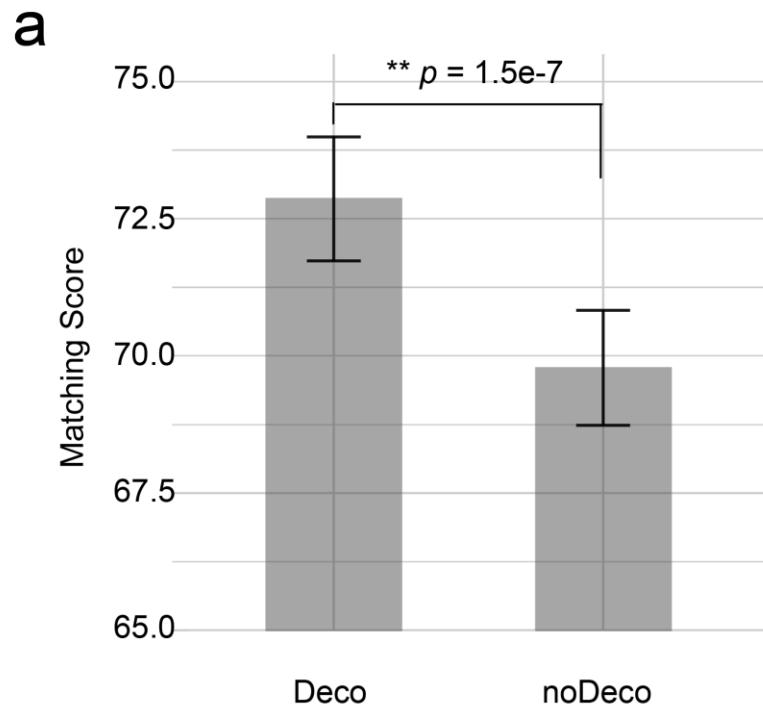
DDA benchmark – IROA mixture (ESI+)

406 Compounds – ESI+, Isolation window: 1Da

Tools	Number of detected standards (MS1)	Compounds correctly annotated (MS2)	Percentage	Time elapsed (1 CPU core)
MS-DIAL + MS-FINDER	165	82	20.2%	32 min
MZmine + SIRIUS	221	77	19.0%	4 hours
MetaboAnalyst*	239	159	39.1%	22 min
MetaboAnalyst (nonDeco)	239	146	36.0%	12 min

* Based on **Complete Database** with deconvolution enabled

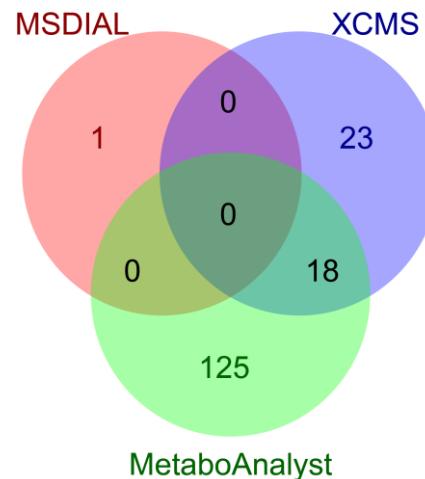
DDA benchmark



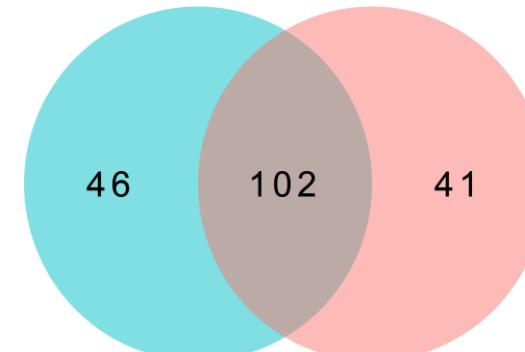
SWATH-DIA (ESI⁺) benchmark

406 Compounds.

Tools	Number of detected standards (MS1)	Compounds correctly annotated (MS2)	Percentage	Time elapsed (1 CPU core)
MS-DIAL + MS-FINDER	5	1	0.25%	3 min
XCMS + SIRIUS	108	42	10.3%	~ 12 h
MetaboAnalyst	324	143	35.22%	14 min
MetaboAnalyst (PathwayDB)	324	148	36.45%	5 min



Complete DB vs. Pathway DB



Levels of Metabolite Identification

1. Positively identified compounds
 - Confirmed by match to known standard
2. Putatively identified compounds
 - Match to MS + RT or MS/MS + RT
3. Compounds putatively identified in a compound class
4. Unknown compounds

In-house reference
standards

LC-MS & MS/MS

MS/MS raw spectra processing

- MS2 data acquisition could be carried out independently from MS1 in DDA or DIA method.
- MS2 spectra-based compound identification is a commonly used approach for global (untargeted) metabolomics;
- There are multiple public reference libraries curated for metabolomics/exposomics community;

Accepted Formats and expected results

To accommodate application scenario and offer compatibility with MS2 spectra results from other popular tools. There are three formats supported:

- i. Simple text file (m/z and intensity separated by tab);
- ii. MGF file format (standard);
- iii. MSP file format (MS-DIAL);

The MS2 spectra/spectrum searching provides results including comprehensive compound identification summary and visualization of the matching pattern:

- i. Compound identification summary table;
- ii. Visualization on MS2 matching pattern and annotation of fragments;

Peak Annotation



MetaboAnalyst 6.0 - from raw spectra to biomarkers, patterns, functions and systems biology

Module Overview

Input Data Type	Available Modules (click on a module to proceed, or scroll down to explore a total of 18 modules including utilities)				
LC-MS Spectra (mzML, mzXML or mzData)			Spectra Processing [LC-MS1 w/wo MS2]		
MS Peaks (peak list or intensity table)		Peak Annotation [MS2-DDA/DIA]	Functional Analysis [LC-MS1]	Functional Meta-analysis [LC-MS1]	
Generic Format (.csv or .txt table files)	Statistical Analysis [one factor]	Statistical Analysis [metadata table]	Biomarker Analysis	Statistical Meta-analysis	Dose Response Analysis
Annotated Features (metabolite list or table)		Enrichment Analysis	Pathway Analysis	Network Analysis	
Link to Genomics & Phenotypes (metabolite list)			Causal Analysis [Mendelian randomization]		

A red arrow points to the "Peak Annotation [MS2-DDA/DIA]" button.

[Home](#)
[Data Formats](#)
[Tutorials](#)
[User Forum](#)
[MetaboAnalystR](#)
[Publications](#)
[Update History](#)
[Databases](#)
[APIs](#)
[User Stats](#)
[About](#)
[Contact](#)

 NSERC
CRSNG

Xia Lab @ McGill (last updated 2024-03-07)

36

MS/MS spectral batch processing

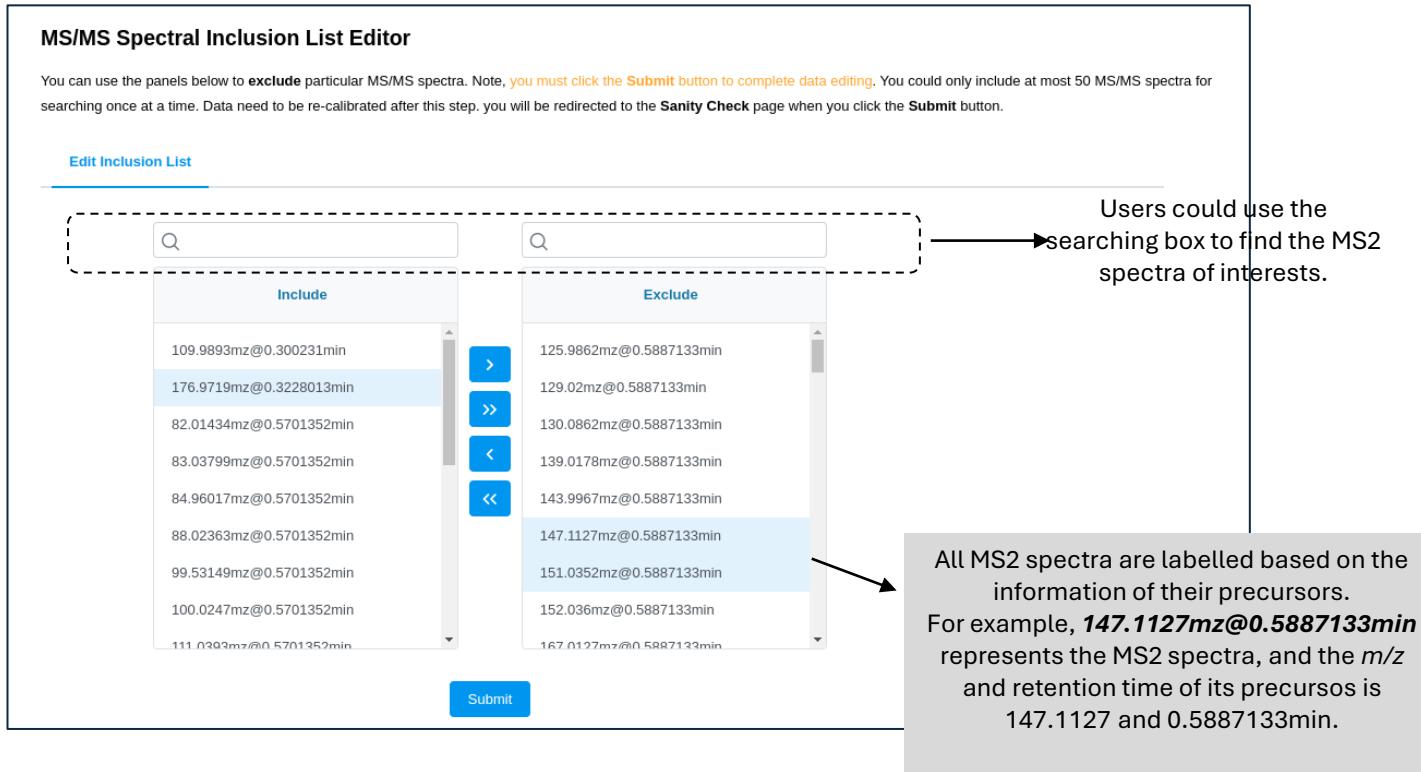
Spectra inclusion editor,

1. Since MetaboAnalyst only support at most 20 spectra searching once at a time, user could manually customize the inclusion list for MS2 database search;
2. By default, the first 20 spectra will be listed into “Include” list to be included for searching;
3. User could move MS2 spectra features between two lists by using the blue moving arrows; 
4. Once the editing is done, Click “Submit” button to confirm.

MS/MS Spectral Inclusion List Editor

You can use the panels below to **exclude** particular MS/MS spectra. Note, you must click the **Submit** button to complete data editing. You could only include at most 50 MS/MS spectra for searching once at a time. Data need to be re-calibrated after this step. you will be redirected to the **Sanity Check** page when you click the **Submit** button.

Edit Inclusion List



Include
109.9893mz@0.300231min
176.9719mz@0.3228013min
82.01434mz@0.5701352min
83.03799mz@0.5701352min
84.96017mz@0.5701352min
88.02363mz@0.5701352min
99.53149mz@0.5701352min
100.0247mz@0.5701352min
111.0303mz@0.5701352min

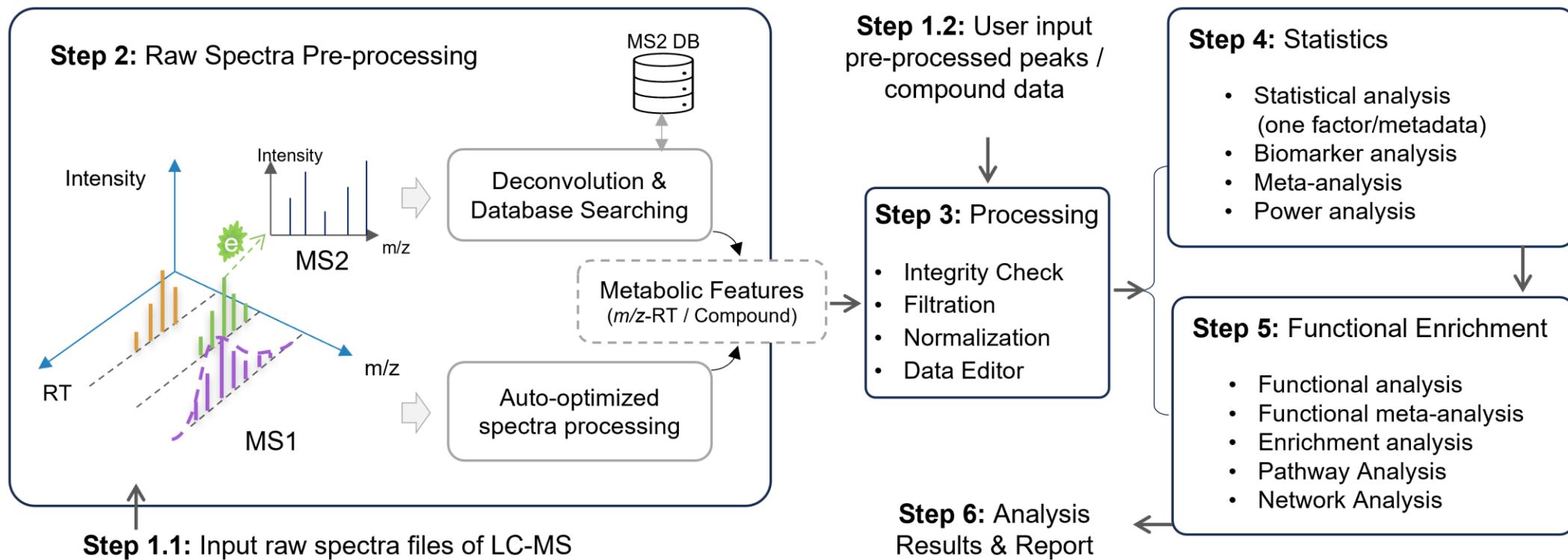
Exclude
125.9862mz@0.5887133min
129.02mz@0.5887133min
130.0862mz@0.5887133min
139.0178mz@0.5887133min
143.9967mz@0.5887133min
147.1127mz@0.5887133min
151.0352mz@0.5887133min
152.036mz@0.5887133min
167.0127mz@0.5887133min

Submit

Users could use the searching box to find the MS2 spectra of interests.

All MS2 spectra are labelled based on the information of their precursors.
For example, **147.1127mz@0.5887133min** represents the MS2 spectra, and the **m/z** and retention time of its precursors is **147.1127 and 0.5887133min**.

A unified workflow



Live Demo

LC-MS Spectra Preparation

MetaboAnalyst currently supports mzML, mzXML, CDF or mzData formats in centroid mode. For MS2 data, spectra should be acquired in either **DDA** or **SWATH-DIA** mode for each job. Mixed mode is not supported.

1. [Required] MS1 Spectra uploaded as individual zip files - one zip (.zip) per spectrum [max: 200 spectra].
2. [Optional] Either **DDA**- or **SWATH-DIA**-based LC-MS/MS Spectra should be uploaded as individual zip files (same as MS1) [max: 50 spectra]. MS2 data must start with "**MS2_**" or marked as "MS2" in meta data file.
3. [Optional] Meta data uploaded as a plain text (.txt) file containing two columns - spectral names and group labels
4. [Optional] Quality control (QC) spectra should start with "**QC_**" or marked as "QC" in meta data. BLANK should be marked as "BLANK" in meta data for subtraction.

Parameter setting & job submission

LC-MS/MS Spectra Processing

MetaboAnalyst currently supports four algorithms for raw spectral peak picking - [centWave](#), [Asari](#), [MatchedFilter](#) and [Massifquant](#).

An auto-optimized workflow has been implemented for [centWave](#). The auto-optimized procedure can significantly improve both the quality of peak detection and the speed of processing. It is available as the [OptiCMS R package](#) for local installation or further extension.

LC-MS Platform: Generic

1. Peak Picking: Algorithms: centWave-auto

2. Peak Alignment: minFraction: 0.80, Polarity: Positive

3. Peak Annotation: Adducts: View, More options: View, ppm for MS2: 10.00, Filtering value: 200.00, Window Size: 1.50

4. MS2 Processing: Threshold: 100,000.00, Deconvolution: checked, Similarity Method: Dot Product, Target Peaks: Significant Ones, MS2 Database: HMDB Experimental

5. Contaminant Removal: checked

6. Blank Subtraction: unchecked

Submit Job



Job Status View

Depending on the current server load and the size of your data, it can take a few hours up to several days to complete your job.

- If you have not logged in, please click [Create Job URL](#) and save the job link. You can then close the current page and come back later using this link.
- At any time during data analysis, keep only one active web page open (except static web pages), as multiple tabs/windows will interfere with each other, leading to unpredictable results.

Job Status

Job ID: 13177
Bookmark Link: [Create Job URL](#)
Current Status: Running
Priority: Level 1
Parameters: Save
Job Progress: 5%

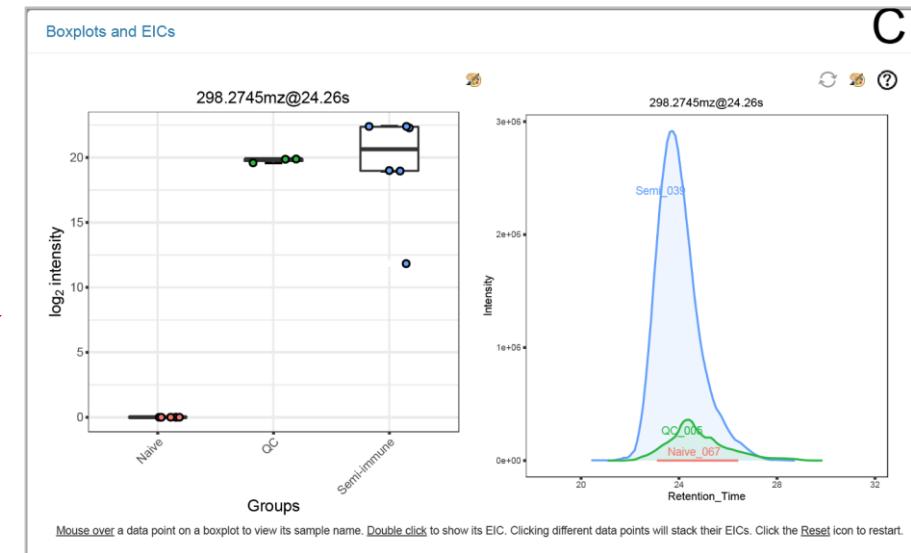
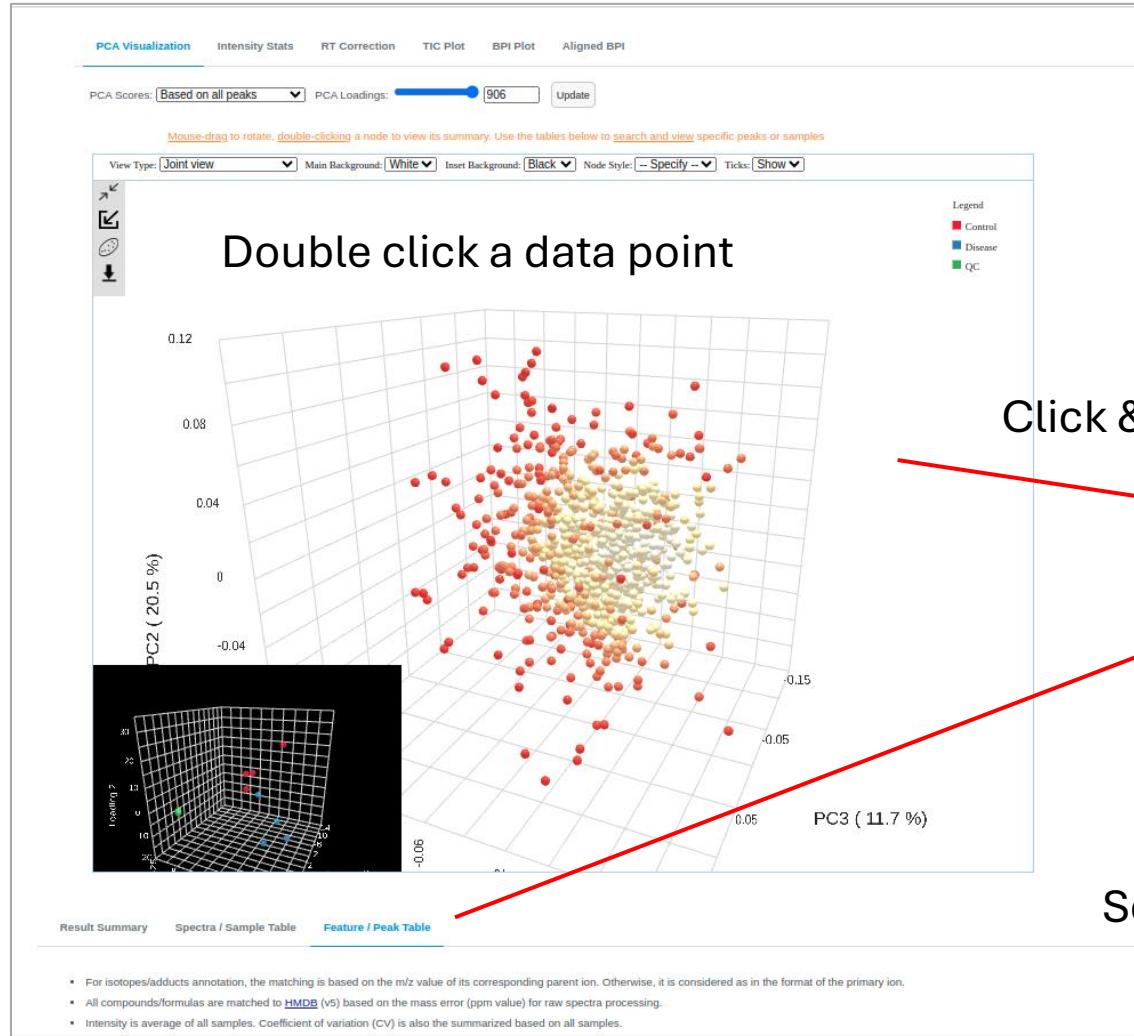
Text Output:

```
QC1.mzML import done!
QC2.mzML import done!
QC3.mzML import done!
QC4.mzML import done!
QC5.mzML import done!
QC6.mzML import done!
SERUM01.mzML import done!
SERUM02.mzML import done!
SERUM03.mzML import done!
SERUM04.mzML import done!
SERUM05.mzML import done!
SERUM06.mzML import done!
```

Output File: Status Text 2024-06-15 07:33:53

Buttons: Refresh Status, Cancel Job, Proceed

Result Exploration



Single spectrum upload

At the first page, user can upload single spectrum or multiple spectra. We used the “**Single Tandem Spectrum**” at this stage.

For single spectrum uploading,

1. It should be a text containing two columns. The first column is *m/z* values, while the second column is intensity values.
2. The two columns must be separated by tab (not space).
3. It is unnecessary to normalize the intensity values, we will automatically do it.

Please enter your data below

[Single Tandem Spectrum](#) [Multiple Tandem Spectra](#)

This module is designed to provide an easy tandem MS spectrum annotation functionalities for single MS2 spectrum.

- The input data should be a two-column list, containing *m/z* and intensity of MS/MS spectrum;
- Two columns should be separated with tab. Each row represents a fragment (e.g. 157.9023 3415);
- *m/z* of the precursor ion is required;
- Specify the ion mode for the MS/MS spectrum is optional but highly-recommended to improve the accuracy;

135.0802 9.23
147.0807 27.55
149.0965 8.74
153.091 22.39
159.0806 9.47
161.0966 8.84
171.0805 15.77
215.1071 13.62
235.1112 12.59
237.1279 23.62
267.138 11.0
277.1586 27.9
279.1744 77.14
309.1851 30.04
325.1792 20.22
337.1802 100.0
393.21 44.44

Precursor Ion Mass (Da):

Precursor Mass Tolerance: PPM ▾

Fragment Mass Tolerance: PPM ▾

MS/MS Database: [? HMDB Experimental](#) ▾

Use Neutral Loss [?](#)

Ion Mode: ▾

Similarity Method: ▾

Try Our Example:

[Submit](#) Click submit to proceed

Parameters for searching,

1. **Precursor Ion Mass** is required, please input the value as precisely as possible;
2. **Tolerance**: both tolerance values are recommended to be optimized based on MS instrument’
3. **MS/MS Databases**: user could customize their database option (see 3.2 for more information);
4. **Use Neutral Loss**: user could optionally use Neutral Loss for database search by use the option. Please note, this is only encouraged for unknown new compound discovery;
5. **Similarity Method**: User could choose traditional way (dot-product) or a new strategy ([spectral entropy](#)).

MS2 spectra databases

- Users can choose one or multiple MS2 spectra data for searching.
- Different databases contain significantly different number of MS2 spectra records.
- Choose “All Database” will search the entire MS2 spectra database, but it may take significantly longer to complete.

The image shows a user interface for searching MS2 spectra. On the left, there's a search form with fields for Precursor Ion Mass (Da) (393.2072), Precursor Mass Tolerance (10.0 PPM), Fragment Mass Tolerance (30.0 PPM), and an MS/MS Database dropdown menu. The dropdown menu is open, showing options: All Database, HMDB Experimental (which is checked), HMDB Predicted, GNPS, and MINEs. A large black arrow points from the dropdown menu to a detailed statistics modal on the right. The modal has a dark header and contains a summary of database options and a table of database statistics.

It is allowed to customize the database options based on your needs. You could select one or multiple database for MS/MS searching. It is noted that if you selected "All Database", the searching may take a quite long time based on your MS/MS spectrum and server loading. Here is a summary on all database option:

Database	Total Records	Unique Compounds	Version/Date
HMDB Experimental	64711	4049	v5 (Sep 4th 2022)
HMDB Predicted	1786690	204969	v5 (Sep 4th 2022)
GNPS	80802	11852	Aug 30th 2022
MoNA	1545053	614391	Aug 15th 2022
MassBank	90190	16879	Sep 4th 2022
MINEs	3009477	537799	Feb 2nd 2023
LipidBlast	1832594	778501	Feb 7th 2023
RIKEN	12771	1231	Sep 7th 2022
ReSpect	4310	691	Sep 7th 2022
VaniyaNP	39937	2681	Sep 4th 2022
MSDIAL	1726174	606878	v4.8 (Sep 11th 2022)
BMDMS	227307	2739	Aug 20th 2022

Mouse hover the help tip to view the detailed information on different database options.

MS2 spectra searching results

MS2 results explanation,

1. Database search results would be summarized as a table;
2. User could expand a row to visually explore the matching results of a MS2 spectrum;
3. Information of fragments will be automatically displayed when mouse hover the fragment;
4. The top (blue) part are users' input, while the bottom (red) parts are from the reference library;
5. All matched fragments will be marked with red diamond at the top.



Demo Datasets

- **DDA dataset**
 - Whole blood metabolomics dataset
 - A total of 30 samples (serum, plasma and whole blood samples (n=6 for each), 6 QCs and 6 DDA MS/MS samples)
- **SWATH-DIA dataset (homework)**
 - Clinical metabolomics dataset ;
 - Control and COVID patients: 16 samples (12 MS1 Samples and 4 SWATH-DIA files) are included.

Output: peak abundance table

- Uniquely identify each peak: retention time and m/z value
- Calculate the relative intensity in each sample

Retention time											
Sample	X1014	X1049	X1068	X1070	X1071	X1073	X1074	X1075	X1076	X1078	
85.02798773_398.845656	91281.129	295971.19	244257.92	82883.828	357387.91	314793.29	296933.07	259134.23	316398.3	298981.38	
85.03918591_540.7198895	20368705	23645645	27541993	20197810	20698441	27700133	18903295	21151136	22135283	23551889	
85.03934182_206.8491361	100801.73	147630.84	128838.32	48201.572	14503.911	94388.175	147840.04	94226.848	47368.725	86117.51	
85.05850153_553.5489174	28578.672	NA	42871.286	45854.92	31862.665	42511.683	16638.517	21645.293	42802.335	47630.422	
85.06447722_552.8676506	64506.008	36993.153	64365.242	21970.254	22431.698	42717.702	49608.002	61113.878	45457.694	31242.437	
85.07557123_503.1977875	5185552	6545664.8	4849575.1	7455068.2	4687812.5	8568037.4	5092330.2	3961282.2	6480194.6	7331818.4	
85.07616337_141.9029172	82899.952	207861.36	50610.657	79208.885	225161.43	NA	347408.98	236485.2	776251.79	164112	
85.0838011_198.0411769	85303.336	123532.16	91254.97	66497.463	172721.72	236255.05	47396.288	78663.557	189683.64	245493.04	
85.50950642_172.3411474	339908.68	321187.16	322001.53	255557.48	330914.06	254245.84	NA	NA	290287.6	298955.37	
85.51517772_50.65023803	118159.94	112972.04	114059.62	113950.95	167858.69	103292.57	86749.39	82707.461	119298.44	107657.2	
85.5363475_41.45434989	53482.821	17514.179	35163.947	36411.914	59951.47	51123.602	41371.083	30019.615	22520.943	47343.966	
85.96264165_42.73935005	81788.089	78215.738	50882.903	65819.686	73752.586	57479.55	71399.888	42905.115	49373.813	68847.43	
86.00545485_545.3171583	46468.886	40671.699	23324.775	36142.339	31310.553	56563.276	26034.229	NA	NA	29480.762	
86.01779309_54.25356378	57728.236	36204.919	31645.834	63374.773	42848.297	70339.755	NA	46788.918	78406.509	49801.696	
86.03613685_568.0578201	120163.19	121293.45	137159.94	118697.36	114696.1	147598.85	95348.512	97339.544	120371.54	117616.77	
86.04255464_546.3279646	773051.95	675716.91	764306.84	716529.31	614985.95	775433.46	527588.69	666915.52	719938.04	659466.81	
86.05953662_575.4776799	1305749	986112.65	1107787.1	896955.61	623282.13	622941.45	627053.74	507228.47	1017792.5	491771.67	
86.05955314_395.2147633	1151506.4	827450.26	484189.22	252791.25	1586988.1	522492.9	1083396.6	410343.24	291013.34	591663.48	
86.0596265_321.9552286	2306641.6	2636648.8	2057971	2244866.3	2813936.3	2650464.4	2521397.2	2291594.2	2794708.7	2986888.4	
86.07101485_545.5074342	18024.92	48694.834	39266.12	NA	21814.652	14367.843	NA	16065.358	11001.248	26206.676	
86.07888004_507.2294891	186762.52	274866.34	292333	168433.73	130364.59	257889.76	129553.62	137593.85	315715.95	134660.58	
86.08334857_524.9644006	NA	NA	51854.327	NA	55064.237	84586.362	38654.123	45651.322	54524.784	40857.812	

Key features

- Raw Data Uploading (.mzML/.mzXML/.mzData/.cdf);
- Parameters Optimization (automatically);
- Peak Profiling (Peak Picking/Alignment/Gap filling);
- Peak Annotation (Adducts + Isotopes);
- Putative Compound Mapping;
- Result Visualization...

Learning Questions ..

- Which software is often used for raw data centroiding?
- What is the underlying algorithm for MS data preprocessing in MetaboAnalyst?
- Generate an Extracted Ion Chromatogram (EIC) plot (overlay of at least one sample from the three groups) for the most significant peak?

Cautions & Tips

1. For raw spectra processing, use 1st example rather than the 2nd/3rd/4th example to avoid waiting in queue for learning purpose;
2. Avoid downloading and uploading any example raw spectra data due to the limited bandwidth.
3. Default MetaboAnalyst includes www.metaboanalyst.ca, new.metaboanalyst.ca and dev.metaboanalyst.ca. please use either of them.

Results summary

[Result Summary](#)[Spectra / Sample Table](#)[Feature / Peak Table](#)[MS/MS Results](#)

Raw Spectra Processing Result Summary:

MetaboAnalyst has finished raw spectra processing with OptiLCMS (1.2.0):

There are 24 samples of 4 groups (Plasma, QC, Serum, whole_blood) included for processing!

Total of 4910 features have been detected and aligned across the whole sample list.

The mass deviation of this study was estimated/set as 5 ppm.

2413 features (49.13%) have been annotated as isotopes.

2312 features (47.08%) have been annotated as adducts.

189 unique formulas have been matched to HMDB database.

958 potential compounds have been matched to HMDB database.

> Download Page

MS/MS Results

Result Summary	Spectra / Sample Table	Feature / Peak Table	MS/MS Results		
Compound	Formula	Matching Score ↑	InchiKey	Database	View
• MS/MS-based compounds identification results are displayed below.					
• Similarity of MS/MS are evaluated based on dot-product or spectral entropy methods. Top 5 compounds are listed from high to low (100, perfect match; 0, not matched).					
• User could click View button below to view the MS/MS pattern matching results.					
mz137.0459@42.46min	C5H4N4O	84.79	FDGQSTZJBFJUBT-UHFFFAOYSA-N	HMDB_experimental	
Allopurinol	C5H4N4O	83.53	OFCNXPNDARWKPPY-UHFFFAOYSA-N	HMDB_experimental	
mz98.9930@51.97min					
mz83.0597@45.80min					
mz147.1130@35.24min					
mz150.0585@43.66min					