



# MetaboAnalyst 6.0

-- a unified platform for metabolomics data processing,  
analysis and interpretation

Spectra Processing [LC-MS w/wo MS2]

# Module Overview



## **An auto-optimized workflow for LC-MS raw spectra pre-processing**

- ✓ Support for LC-MS1 feature detection with centWave-based auto-optimized workflow;
- ✓ Support for LC-MS1 feature detection with a highly-efficient algorithm, Asari;
- ✓ Support for LC-MS2 (Data-Dependent Acquisition, DDA) data deconvolution and processing in an auto-optimized approach;
- ✓ Support for LC-MS2 (sequential windowed acquisition all theoretical MS data data-independent acquisition, SWATH-DIA) data deconvolution and processing in a highly-efficient way;
- ✓ Comprehensive MS2 reference spectra libraries for compound searching;
- ✓ MS2 spectra searching supports direct searching or searching based on neutral loss. MS2 spectrum similarity can be evaluated based on dot-product or spectral entropy method.



# Content

- ❑ 1. Introduction;
- ❑ 2. Spectra Preparation;
- ❑ 3. Choose the Module;
- ❑ 4. LC-MS1 Spectra Processing;
  - 4.1 Spectra Files Upload
  - 4.2 Integrity Check
  - 4.3 Parameter Setting
  - 4.4 Jos Status
  - 4.5 Result Exploration
- ❑ 5. LC-MS1 + DDA Spectra Processing;
  - 5.1 Spectra Files Upload
  - 5.2 Integrity Check
  - 5.3 Parameter Setting
  - 5.4 Jos Status
  - 5.5 Result Exploration
- ❑ 6. LC-MS1 + SWATH-DIA Spectra Processing;
  - 6.1 Spectra Files Upload
  - 6.2 Integrity Check
  - 6.3 SWATH window Check
  - 6.4 Parameter Setting
  - 6.5 Jos Status
  - 6.6 Result Exploration
- ❑ 7. Download Results.

# 1. Introduction



## Background

- Untargeted metabolomics data generated from LC-HRMS experiments are typically characterized by 1000s of peaks with unknown chemical identities.
- To assist with compound identification, tandem MS (called MS/MS or MS2) spectra are often collected.
- MS2 methods such as DDA and SWATH-DIA are commonly used

## Data Formats

Raw spectra files must be saved in common open-source formats and uploaded individually as separate zip files. LC-MS spectra data is mandatory, while MS2 is optional. There are four open-source formats supported:

- i. mzML (recommended);
- ii. mzXML;
- iii. cdf/CDF;
- iv. mzData (phasing-out).

## Expected Results

LC-MS raw spectra processing module provides user comprehensive results on LC-MS1 features and MS2-based compound identifications:

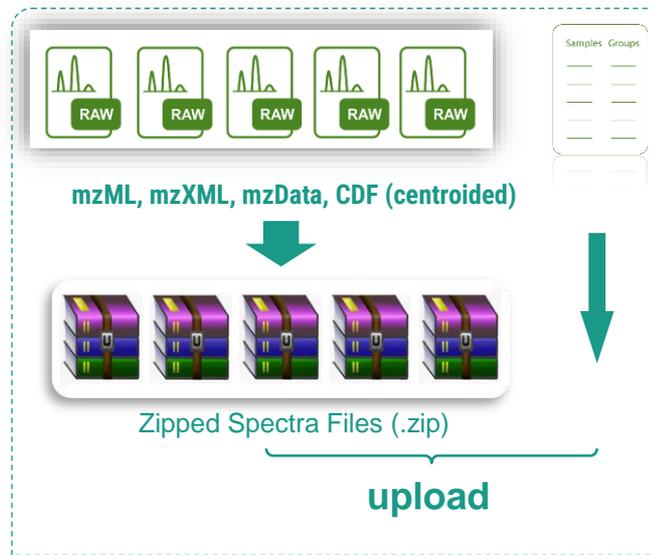
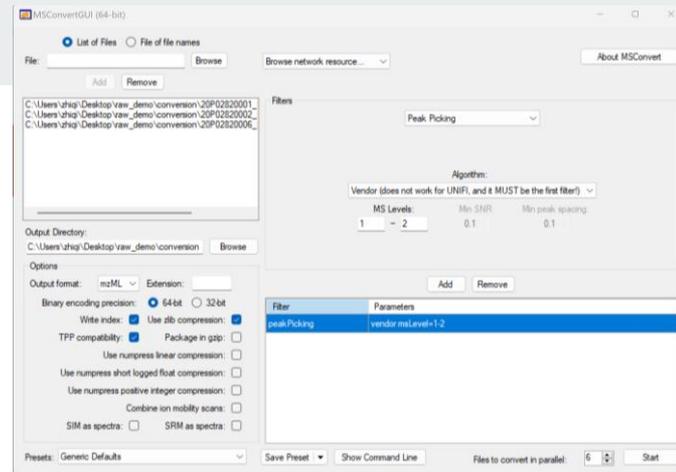
- i. Compound identification summary table;
- ii. Visualization on MS2 matching pattern and annotation of fragments;

## 2. Spectra data preparation

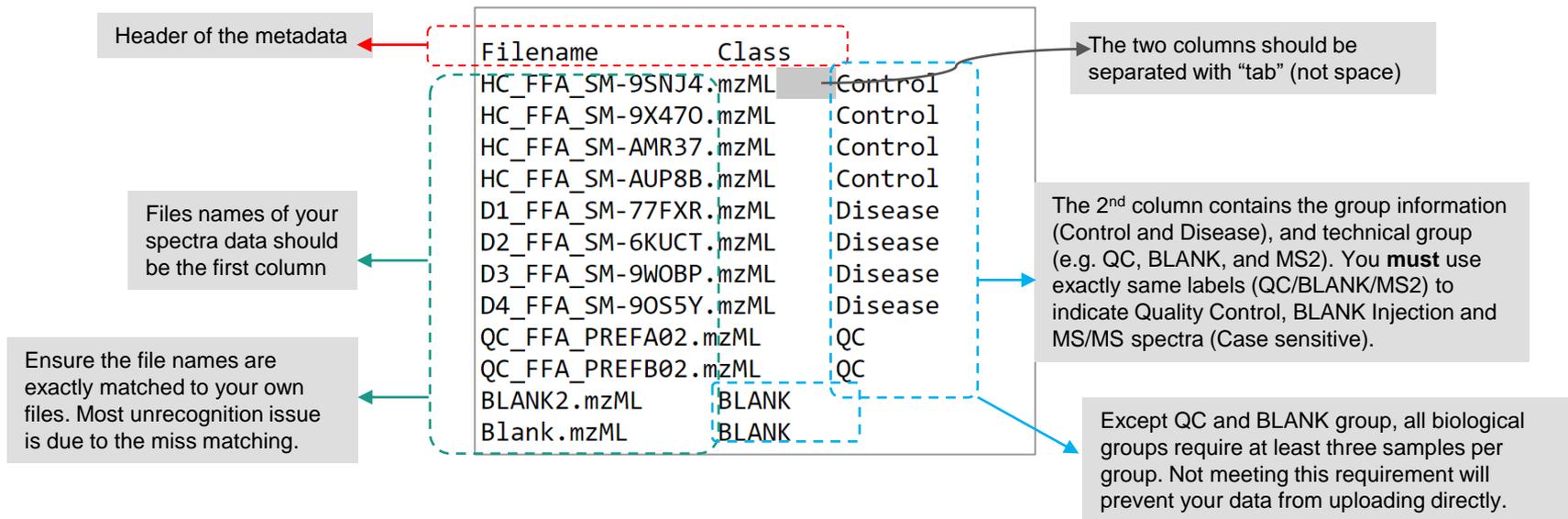
1. We highly recommend users to centroid MS files with **ProteoWizard**:

- **For GUI:** **Must** add 'Peak Picking' as the **1<sup>st</sup> filter** (see example at the left side);
- **For Command:** `docker run -it --rm -e WINEDEBUG=all -v /FILE_PATH/:/data chambm/pwiz-skyline-i-agree-to-the-vendor-licenses wine msconvert FILENAME -o OUTPUTDIR --mzML --filter "peakPicking true 1-" --filter "zeroSamples removeExtra" --filter "msLevel 1" --64 --zlib`

2. All centroid MS spectra files must be zipped individually prior to uploading. Please note, users should **NOT** zip all spectra files into one single zipped file (see the left workflow);
3. You should include a meta-data file to indicate the grouping information of your MS spectra files, as well as the technical sample information (like MS2, QC, and BLANK) by using a simple text file. The details about how to prepare the metadata file are shown in the next page;
4. You must upload your spectra files together with the metadata table. You cannot upload them separately;
5. Metadata table is not required, but highly recommended.



## 2. Metadata preparation



# 3. Choose the Module

Go to MetaboAnalyst (<https://www.metaboanalyst.ca>), and select the module



**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**)

Input Data Type	Available Modules (click on a module to proceed, or scroll down to explore a total of 18 modules including <b>utilities</b> )				
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]		





## **4. LC-MS1 Spectra Processing**

# 4.1 LC-MS1 spectra upload

Read Instructions before uploading

- Upload
- SWATH check
- Spectra check
- Spectra processing
- MS2 Spectra
- Job status
- Spectra result
- Download
- Exit

Log in

## LC-MS Spectra Upload

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 [example](#)
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.

Depending on our server load, spectra processing can take a long time to complete, to avoid waiting:

1. For guest users (default), after job submission, click **Create Bookmark URL** and save the URL so you can return later to check your job status.
2. For registered users, use the buttons on the left panel to manage your projects. Larger data processing and analysis support are available for [subscribed users](#).

Please **Select** all LC-MS1 and MS2 spectral data, then click **Upload** to start uploading. Once the uploading has completed, click **Proceed** to continue.

+ Select

Click "Select" to choose your files to upload

Select Mode:  MS1 Only  MS1+DDA  MS1+SWATH-DIA

Reset

Proceed

For LC-MS1 data processing, you should choose "MS1 only"

Click "Proceed" button when your uploading is completed

Try to use example data for testing and learning

### Try our example data

Data Type	Description	Download
<input checked="" type="radio"/> Quick Demo (MS1)	A small example dataset for demo purposes, containing 10 spectra (UPLC-Q/E-ESI, C18) organized into three groups (Healthy, Crohn's Disease and QC) from <a href="#">Lloyd-Price et al.</a>	<a href="#">Download GoFile</a>
<input type="radio"/> Malaria Data (MS1)	An experimental Malaria metabolomics dataset (UPLC-Q/E-ESI, HILIC) between two immune status (Native vs. Semi-immune) from <a href="#">Li et al.</a> 15 samples (12 Samples and 3 QC) are included.	<a href="#">Download GoFile</a>
<input type="radio"/> Blood Samples (MS1+DDA)	An whole blood metabolomics dataset (UPLC-Q/E-ESI, HILIC) among three blood types. DDA MS/MS file included. A total of 30 samples (serum, plasma and whole blood samples (n=6 for each), 6 QCs and 6 DDA MS/MS samples) are included.	<a href="#">Download GoFile</a>
<input type="radio"/> COVID-19 dataset (MS1+SWATH-DIA)	An clinical metabolomics dataset (UPLC-TripleTOF 5600+, AB Sciex, ESI, Acquity XSelect HSS T3 column) between control and COVID patients from <a href="#">Mahmoud S. et al.</a> 16 samples (12 MS1 Samples and 4 SWATH-DIA files) are included.	<a href="#">Download GoFile</a>

Submit

# 4.1 LC-MS1 spectra upload

Please **Select** all LC-MS1 and MS2 spectral data, then click **Upload** to start uploading. Once the uploading has completed, click **Proceed** to continue.

+ Select Upload

Click "Upload" to start uploading

Metadata.txt	340 Bytes	<input type="checkbox"/>	<input type="checkbox"/>
Semi_091.zip	9.2 MB	<input type="checkbox"/>	<input type="checkbox"/>
Semi_143.zip	9.4 MB	<input type="checkbox"/>	<input type="checkbox"/>
Semi_157.zip	9.7 MB	<input type="checkbox"/>	<input type="checkbox"/>
Naive_139.zip	8.7 MB	<input type="checkbox"/>	<input type="checkbox"/>
QC_001.zip	9.7 MB	<input type="checkbox"/>	<input type="checkbox"/>
QC_003.zip	9.6 MB	<input type="checkbox"/>	<input type="checkbox"/>
QC_005.zip	9.6 MB	<input type="checkbox"/>	<input type="checkbox"/>



Please note that you need to select and upload all files once at a time, you cannot upload in different batches.

Data uploading may take long time due to the connection speed and high volumes of users. Please be patient or avoid the peak time (EST 8:00AM-11:00AM, Mon-Fri).

Please **Select** all LC-MS1 and MS2 spectral data, then click **Upload** to start uploading. Once the uploading has completed, click **Proceed** to continue.

+ Select

Uploading progress and status is displayed here

Naive_109.zip	8.9 MB	<input type="checkbox"/>	<input type="checkbox"/>
Naive_127.zip	9.0 MB	<input type="checkbox"/>	<input type="checkbox"/>

Select Mode:  MS1 Only  MS1+DDA  MS1+SWATH-DIA

Reset Proceed

Successful  
Naive\_071.zip is uploaded.

Once all files get uploaded, Click "Proceed" to continue

If you want to discard all uploaded files, click "Reset" button

Try our example data

## 4.2 Integrity Check

Please note that online centroid may not succeed because some open-source files may not be able to be correctly identified. Users are highly recommended to use ProteoWizard to centroid their files before uploading (see section 3. Spectra data preparation).

If your data is not in centroid mode, click **Convert** wrench button to convert it online.

Results of the **Data Integrity Check** are shown here.

**Data Integrity Check:** Show R Cor

- Spectral Format - only mzML, mzXML, mzData and netCDF formats are currently supported.
- MS Mode - only spectra in **centroid mode** are supported in the online platform. Click **Convert** to centroid your profile data online. This conversion process will take some time, please be patient.
- If a meta data file is provided:
  - The first column (spectral names) must match the sample names in the meta-data file.
  - The second column (group labels) must contain at least two groups (not including QC), each containing ≥ 3 replicates.

Spectra	Centroid	Size (MB)	Group	Convert	Include
Naive_007.mzML	True	15.4	Naive		<input checked="" type="checkbox"/>
Naive_139.mzML	True	14.7	Naive		<input checked="" type="checkbox"/>
Naive_027.mzML	True	15.0	Naive		<input checked="" type="checkbox"/>
Naive_071.mzML	True	15.0	Naive		<input checked="" type="checkbox"/>
Naive_109.mzML	True	15.0	Naive		<input checked="" type="checkbox"/>
Naive_127.mzML	True	15.0	Naive		<input checked="" type="checkbox"/>
Semi_025.mzML	True	15.7	Semi_Immune		<input checked="" type="checkbox"/>
Semi_061.mzML	True	15.6	Semi_Immune		<input checked="" type="checkbox"/>
Semi_143.mzML	True	15.7	Semi_Immune		<input checked="" type="checkbox"/>
Semi_157.mzML	True	16.0	Semi_Immune		<input checked="" type="checkbox"/>
Semi_091.mzML	True	15.3	Semi_Immune		<input checked="" type="checkbox"/>
Semi_045.mzML	True	15.6	Semi_Immune		<input checked="" type="checkbox"/>
QC_005.mzML	True	15.8	QC		<input checked="" type="checkbox"/>
QC_001.mzML	True	16.1	QC		<input checked="" type="checkbox"/>
QC_003.mzML	True	15.9	QC		<input checked="" type="checkbox"/>

« < 1 > » 20

**Next**

Click **Next** to move on to the Parameters Selection page (At least 3 samples included for next)

# 4.3 Parameter Setting

## LC-MS/MS Spectra Processing

MetaboAnalyst 6.0 supports multiple algorithms for raw spectral peak picking (including [centWave](#), [Asari](#), [MatchedFilter](#) and [Massifquant](#)):

An auto-optimized workflow has been offered for [centWave](#). The auto-optimized procedure can significantly improve both the quality of peak detection and quantification (see [benchmarking results](#)). There are two options for users to choose for [centWave](#) algorithm.

- Default/manual option will use the parameters in the current display. You can manually overwrite these settings;
- Auto-optimized will automatically select the best parameter combination (for the [centWave](#) only).

The source code for raw spectral processing is now 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: 0.8

Polarity:  Positive  Negative

Adducts: View

More options: View

4. Contaminant Removal  View

5. Blank Subtraction

User could customize the parameters for spectra processing. By default, the auto-optimized workflow will be applied. You can use other algorithm, like **Asari** for a higher computational efficiency.



LC-MS Platform: Generic

1. Peak Picking Algorithms: Asari

ppm: centWave-auto

More options: Asari

minFraction: centWave-manual

Polarity:  Positive  Negative

Adducts: View

More options: View

4. Contaminant Removal  View

5. Blank Subtraction

**Contaminant Removal** is designed to remove contaminating signal for parameter optimization step ONLY. All MS signals will be retained and detected at the later peak detection step.

Click "**Submit Job**" button to perform the spectra processing.

Submit Job

# 4.4 Jos Status

**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: 9681  
Bookmark Link: [Create Job URL](#)  
Current Status: Running  
Priority: Level 1  
Parameters: Save  
Job Progress:  22%

**Text Output:**

```
QC_001.mzML import done!  
QC_003.mzML import done!  
QC_005.mzML import done!  
Semi_025.mzML import done!  
Semi_045.mzML import done!  
Semi_061.mzML import done!  
Semi_091.mzML import done!  
Semi_143.mzML import done!  
Semi_157.mzML import done!  
Raw file initialized Successfully!  
To reduce memory usage BPIS and TICS plots will be created using only 10 samples per group.  
Plotting BPIS and TICS.
```

**Output File:** Status Text 2024-03-09 18:12:38

[Refresh Status](#) [Cancel Job](#) [Proceed](#)

The job may take some time to complete, so click **“Create Bookmark URL”** to save the job link to check the job status at a later time. Otherwise, once the web session expired after 40min, your job will be lost forever.

The status of the job will update here in real-time.

**Job Status Link**

Please save this link to resume the process in the future. The link will expire in 14 days.

Note: it is advised to have only a single window running MetaboAnalyst open at a time.

[https://new.metaboanalyst.ca/MetaboAnalyst/faces/Share?ID=blcicz22p\\_9681](https://new.metaboanalyst.ca/MetaboAnalyst/faces/Share?ID=blcicz22p_9681)

[Copy](#) [Cancel](#)

**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: 9681  
Bookmark Link: [Create Job URL](#)  
Current Status: Finished  
Priority: Level 1  
Parameters: Save  
Job Progress:  100%

**Text Output:**

```
decompose, spectrum  
The following object is masked from 'packagebase':  
union  
Raw spectral data processing has been finished completely!
```

**Output File:** Status Text 2024-03-09 18:11:40

[Refresh Status](#) [Cancel Job](#) [Proceed](#)

Once the job is complete (**Job Progress 100%**), click **Proceed** to view the results.

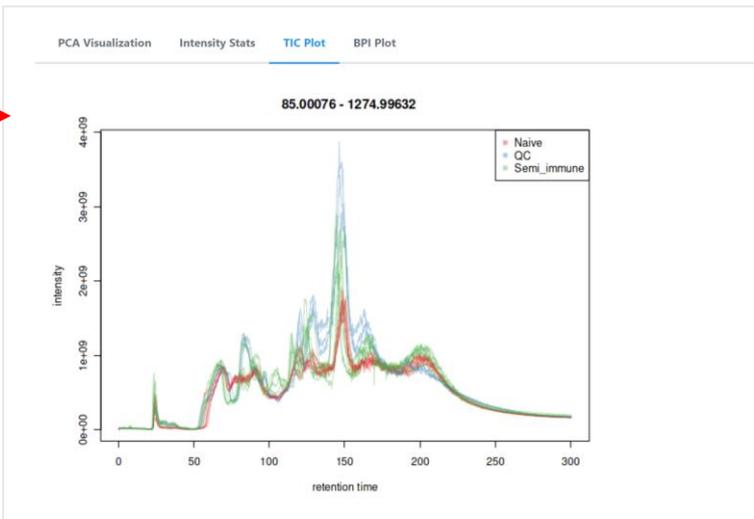
# 4.5 Result Exploration

The screenshot displays the 'Processing Results' section of a software interface. On the left is a navigation sidebar with options like 'Upload', 'SMATH check', 'Spectra check', 'Spectra processing', 'MS2 Spectra', 'Job status', 'Spectra result', 'Download', and 'Exit'. The main area shows 'Processing Results' with tabs for 'PCA Visualization', 'Intensity Stats', 'TIC Plot', and 'BPI Plot'. The 'PCA Visualization' tab is active, showing a 3D PCA score plot with axes PC1 (26.8%), PC2 (6.2%), and PC3 (6.2%). Data points are colored by group: Naive (red), QC (blue), and Semi\_immune (green). A smaller 2D PCA loading plot is visible in the bottom left of the plot area. Below the plot is a 'Raw Spectra Processing Result Summary' panel with the following text:

**Raw Spectra Processing Result Summary:**  
Metabolabz has finished raw spectra processing with OptiLCMS (1.1.0):  
There are 15 samples of 3 groups (Naive, QC, Semi\_Immune) included for processing!  
Total of 11433 features have been detected and aligned across the whole sample list.  
The mass deviation of this study was estimated/set as 5 ppm.  
5228 features (45.72%) have been annotated as isotopes.  
5085 features (44.47%) have been annotated as adducts.  
346 unique formulas have been matched to HMDB database.

Explore other graphical summaries of the spectral processing results here.

Try to use these functionalities to switch the PCA view, compute ellipse, and download the result figures.

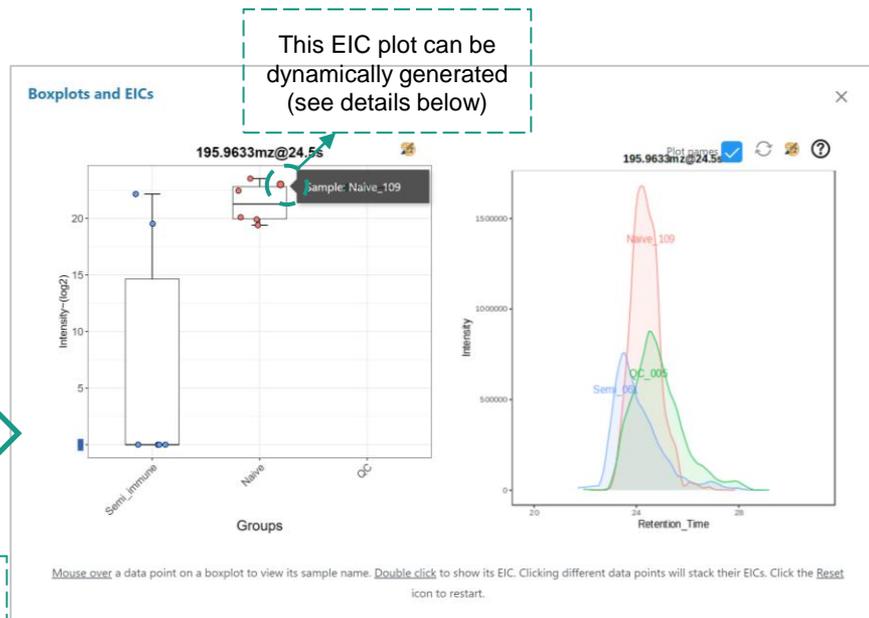
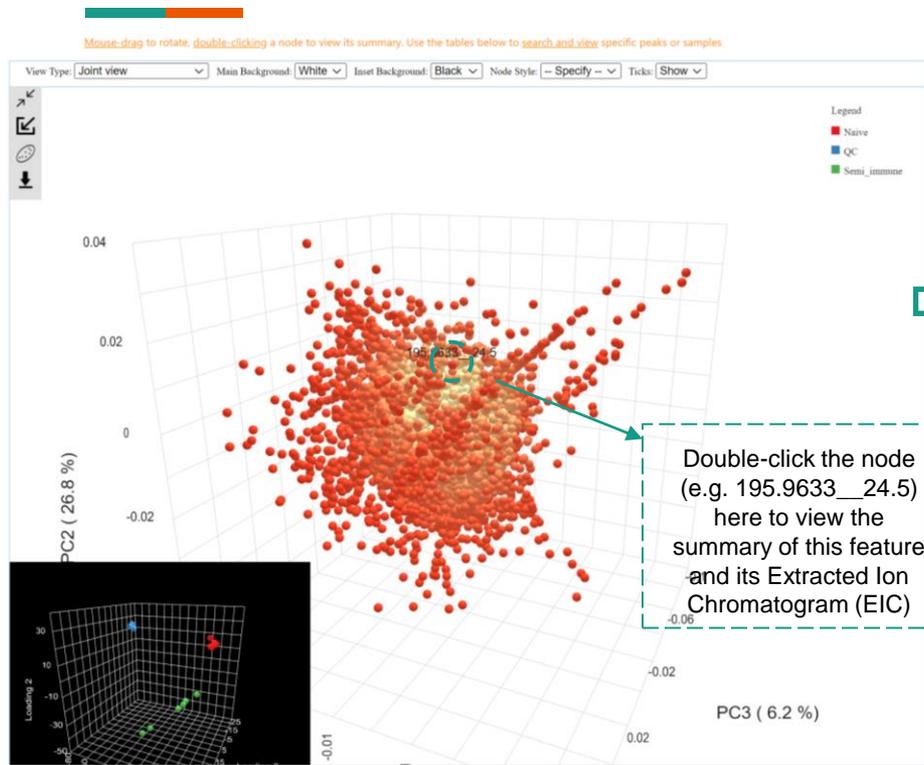


Spectra processing results are visually summarized in 3D PCA view, with all samples displayed in score plot, and all features in loading plot.

Detailed information including samples and features are summarized at the bottom panel.



# 4.5 Result Exploration



## How to use the dynamic EIC generation,

1. **Select MS feature.** From the PCA loading plot, double click the node of interest. Then a dialog will be pop-up automatically;
2. **View the boxplot and click the sample node:** The intensity of all samples are statistically summarized as a boxplot. Double any node to add the EIC of the sample to the right panel. By clicking multiple samples, you can dynamically add more layers into the EIC panel.
3. **Reset the EIC:** If you want to reset the EIC plot, click the reset circle to empty the EIC panel.

# 4.5 Result Exploration

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

**Raw Spectra Processing Result Summary:**  
MetaboAnalyst has finished raw spectra processing with OptiLCMS (1.1.0):  
There are 15 samples of 3 groups (Naive, QC, Semi\_immune) included for processing!  
Total of 11433 features have been detected and aligned across the whole sample list.  
The mass deviation of this study was estimated/set as 5 ppm.  
5228 features (45.72%) have been annotated as isotopes.  
5085 features (44.47%) have been annotated as adducts.  
346 unique formulas have been matched to HMDB database.  
2813 potential compounds have been matched to HMDB database.



- 1. Result Summary:** This summary box is used to summarize the basic information of this job;
- 2. Spectra/Sample Table:** This table offers detailed information on all samples (see the example at the right side).
- 3. Feature/Peak Table:** This table provides detailed information on all detected features and its putative annotation (see the example from the next page).

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

Click the "View" icon to view the corresponding Total Ion Chromatogram of the sample

Spectra ↑↓	Group ↑↓	Peaks No. ↑↓	Missing (%) ↑↓	RT Range	m/z Range	View
<input type="text"/>						
Semi_157	Semi_immune	5850	48.83	1.91~296.45	85.048~1264.1306	
Naive_007	Naive	4465	60.95	1.91~297.5	85.0844~1274.5201	
Naive_027	Naive	4497	60.67	1.91~297.5	85.0762~1274.5201	
Naive_071	Naive	4076	64.35	1.91~297.5	85.0844~1270.6334	
Naive_109	Naive	4518	60.48	1.91~297.5	85.0844~1270.6334	
Naive_127	Naive	4500	60.64	1.91~297.5	85.048~1274.5201	
Naive_139	Naive	4364	61.83	1.91~297.5	85.0844~1274.5201	
QC_001	QC	5777	49.47	1.91~296.45	85.048~1264.1306	
QC_003	QC	5836	48.95	19.25~296.45	85.048~1264.1306	
QC_005	QC	5636	50.7	19.25~296.45	85.048~1264.1306	
Semi_025	Semi_immune	5634	50.72	1.91~297.5	85.048~1264.1306	

# 4.5 Result Exploration

Result Summary Spectra / Sample Table **Feature / Peak Table**

- For isotopes/adducts annotation, the matching is based on the m/z value of its corresponding parent ion. Otherwise, it is considered as in the format of the primary ion.
- All compounds/formulas are matched to [HMDB](#) (v5) based on the mass error (ppm value) for raw spectra processing.
- Intensity is average of all samples. Coefficient of variation (CV) is also the summarized based on all samples.
- When group information is provided, p values will be calculated with t-test/ANOVA based on log transformed data.

m/z ↑↓	RT/s ↑↓	Intensity ↑↓	CV (%) ↑↓	P values ↑↓	FDR	Annotations	Putative IDs	View
544.370236354959	38.6911764	1.1206938E7	45.27	6.2481402E-23	0.0			
684.830841124181	91.503288	4.3811566E7	49.65	1.1195928E-22	0.0			
237.170126537048	28.9694562	8.1026999E7	51.62	8.2754042E-22	0.0	M0 M-H-		
293.196509329594	23.9772402	1.1601194E7	38.33	8.1127794E-21	0.0	13C/12C ACN		
336.945587666283	59.710875	3.4089651E7	48.97	1.199339E-20	0.0	M0 M-H-		
238.173669143381	28.4439942	8563085.0	44.38	1.7304435E-20	0.0	13C/12C M-H-		
830.36552376317	72.322668	4.2872149E7	40.62	1.9462206E-20	0.0			
125.060050295032	275.9517	5525526.0	42.88	2.8978219E-20	0.0			

Putative IDs

Formulas	Compounds
	<a href="#">1-Hydroxyepiacetone</a> <a href="#">Aconitidin</a> <a href="#">24-&gt;5-Abec-</a>
	<a href="#">4-11-d-12-dieoxy-3-epidemanol</a> <a href="#">4R-5S-7R-11R-</a>
	<a href="#">11-12-Dihydroxy-110-soirovetiven-2-one</a>
	<a href="#">Anotrichodinol</a> <a href="#">Saloba-Carissanol</a> <a href="#">aloba-Carissanol</a>
	<a href="#">Enoxyubimin</a> <a href="#">Dihydroxyocorone</a> <a href="#">Piperalol</a>
	<a href="#">Zedoaronidin</a> <a href="#">Hydroxytelonolide</a> <a href="#">Toxin F52</a>
	<a href="#">Urodolone</a> <a href="#">Bisacurone B</a> <a href="#">2-3-Dihydroabscisic</a>
	<a href="#">alcohol</a> <a href="#">2-Methyl-5-oxethyl-2-furanacetanoic acid</a>
	<a href="#">2-Methyl-5-oxethyl-2-furanheptanoic acid</a>

Click the **"Putative ID"** icon to view the putative ID results. Please note, this chemical ID is generated based on the MS1 information.

Features without the putative IDs from HMDB database would not show the icon.

Click the **"View"** icon to start the corresponding dynamic EIC plot panel.



## 5. LC-MS1 + DDA Spectra Processing

In this section, we only emphasize the functionalities specifically related to the DDA data processing. All other MS1 associated features have been included with details in the previous section 4, therefore they won't be repeated. Please read the section 3 and 4 at first if you are not familiar with MetaboAnalyst.

# 5.1 Spectra Files Upload

## Meta-data Table

Samples Groups	
PLASMA01.mzML	Plasma
PLASMA02.mzML	Plasma
PLASMA03.mzML	Plasma
PLASMA05.mzML	Plasma
PLASMA06.mzML	Plasma
PLASMA07.mzML	Plasma
QC1.mzML	QC
QC2.mzML	QC
QC3.mzML	QC
QC4.mzML	QC
QC5.mzML	QC
QC6.mzML	QC
QCDDA_SCAN1_01.mzML	MS2
QCDDA_SCAN1_02.mzML	MS2
QCDDA_SCAN2_01.mzML	MS2
QCDDA_SCAN2_02.mzML	MS2
QCDDA_SCAN3_01.mzML	MS2
QCDDA_SCAN3_02.mzML	MS2
SERUM01.mzML	Serum
SERUM02.mzML	Serum
SERUM03.mzML	Serum
SERUM04.mzML	Serum
SERUM05.mzML	Serum
SERUM06.mzML	Serum
WB01.mzML	whole_blood
WB02.mzML	whole_blood
WB03.mzML	whole_blood
WB04.mzML	whole_blood
WB05.mzML	whole_blood
WB06.mzML	whole_blood

### LC-MS Spectra Upload

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.

- [Required] MS1 Spectra uploaded as individual zip files - one zip (.zip) per spectrum [max: 200 spectra].
- [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.
- [Optional] Meta data uploaded as a plain text (.txt) file containing two columns - spectral names and group labels [\[example\]](#)
- [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.

Depending on our server load, spectra processing can take a long time to complete, to avoid waiting:

- For guest users (default), after job submission, click **Create Bookmark URL** and save the URL so you can return later to check your job status.
- For registered users, use the buttons on the left panel to manage your projects. Larger data processing and analysis support are available for [subscribed users](#).

Please **Select** all LC-MS1 and MS2 spectral data, then click **Upload** to start uploading. Once the uploading has completed, click **Proceed** to continue.

Select Mode:

- MS1 Only  MS1+DDA  MS1+SWATH-DIA

Reset

Proceed

For LC-MS1 + DDA data processing, you should choose "MS1 + DDA"

All MS2 spectra (DDA/SWATH-DIA) files must be grouped as "MS2". Otherwise, these files won't be detected as MS2 files.

# 5.2 Integrity Check

## Data Integrity Check:

1. Spectral Format - only mzML, mzXML, mzData and netCDF formats are currently supported;
2. MS Mode - only spectra in **centroid mode** are supported in the online platform. Click **Convert** to centroid your profile data online. **This conversion process will take some time, please be patient ...**
3. If a meta data file is provided;
  - o The first column (spectral names) must match the sample names in the meta-data file;
  - o The second column (group labels) must contain at least two groups (not including QC), each containing  $\geq 3$  replicates.

For MS2 data processing,  
MS level is indicated at the  
Integrity Check page

Spectra	Centroid	Size (MB)	MS Level	Group	Convert	Include
QCDDA_SCAN1_01.mzML	True	6.9	MS2	MS2		<input checked="" type="checkbox"/>
QCDDA_SCAN2_02.mzML	True	3.9	MS2	MS2		<input checked="" type="checkbox"/>
QCDDA_SCAN1_02.mzML	True	6.9	MS2	MS2		<input checked="" type="checkbox"/>
QCDDA_SCAN3_02.mzML	True	2.2	MS2	MS2		<input checked="" type="checkbox"/>
QC5.mzML	True	6.6	MS1	QC		<input checked="" type="checkbox"/>
QC3.mzML	True	6.6	MS1	QC		<input checked="" type="checkbox"/>
QC4.mzML	True	6.6	MS1	QC		<input checked="" type="checkbox"/>
QC6.mzML	True	6.6	MS1	QC		<input checked="" type="checkbox"/>
QC2.mzML	True	6.6	MS1	QC		<input checked="" type="checkbox"/>

## MS Levels,

1. **MS1:** These files will be used for MS1 feature detection only. Detected features could be used as the target features list for MS2 data processing;
2. **MS2:** MS1 feature detection will not be performed for these files. They will be only used for MS2 feature detection.

# 5.3 Parameter Setting

## 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 quantification (see [benchmarking results](#)). The source code for raw spectral processing is now 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

### 3. Peak Annotation

Polarity:  Positive  Negative

Adducts: [View](#)

More options: [View](#)

ppm for MS2: 10.00

Filtering value: 200.00

Window Size: 1.50

Threshold: 100,000.00

Deconvolution:

Similarity Method:  Dot Product  Spectral Entropy

Target Peaks:  Significant Ones  All Features

MS2 Database: HMDB Experimental

### 4. MS2 Processing

### 5. Contaminant Removal

[View](#)

### 6. Blank Subtraction

Submit Job

Parameters for MS2 data processing needs to be offered based on MS instruments

### 4. MS2 Processing

ppm for MS2: 10.00

Filtering value:

Window Size:  All Database  HMDB Experimental  HMDB Predicted

Threshold:

Deconvolution:  GNPS  Spectral Entropy

Similarity Method:  MINEs  Entropy

Target Peaks:

MS2 Database: HMDB Experimental

## MS2 Processing Parameters,

- Window Size:** For DDA data acquisition, there must a fixed window size setup to ion acquisition. It is usually a narrow window, e.g. 1~3Da;
- Threshold:** MS/MS acquisition usually requires a threshold to exclude the low-intensity ions to ensure the quality of acquired spectrum;
- Deconvolution:** Deconvolution on DDA spectra can be optionally disabled;
- Target Peaks:** MS2 spectra data processing is based on the MS1 feature detection results. User could choose use significantly different features ( $p < 0.05$ ) or all features as targets.
- MS2 Databases:** Multiple MS2 spectra databases have been curated and included for database searching. User could use one or more to search. Please note that use "All Database" option will significantly increase the data processing time.

## 5.4 Jos Status

**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: 9687

Bookmark Link: [Create Job URL](#)

Current Status: Running

Priority: Level 1

Parameters: [Save](#)

Job Progress:

Text Output:

```
+ mes = paste0('Step 7/12: MS/MS data imported completely!')
+
+ ');
+ ecol = "
+ progress = 110
+ };
Step 7/12: MS/MS data imported completely!
mSet <- PerformDDAdeconvolution(mSet,
+ ppm1 = 5.0, ppm2 = 10.0, sn = 12, filtering = 200.0, window_size = 1.5, intensity_thresh = 100000.0,
database_path = ~/home/glassfish/sqlite/MS2ID_Complete_v09102023.sqlite',
+ ncores = 4L, decoOn = TRUE, useEntropy = FALSE);
Loading required package: parallel
```

Output File: [Status Text](#) 2024-03-10 00:54:05

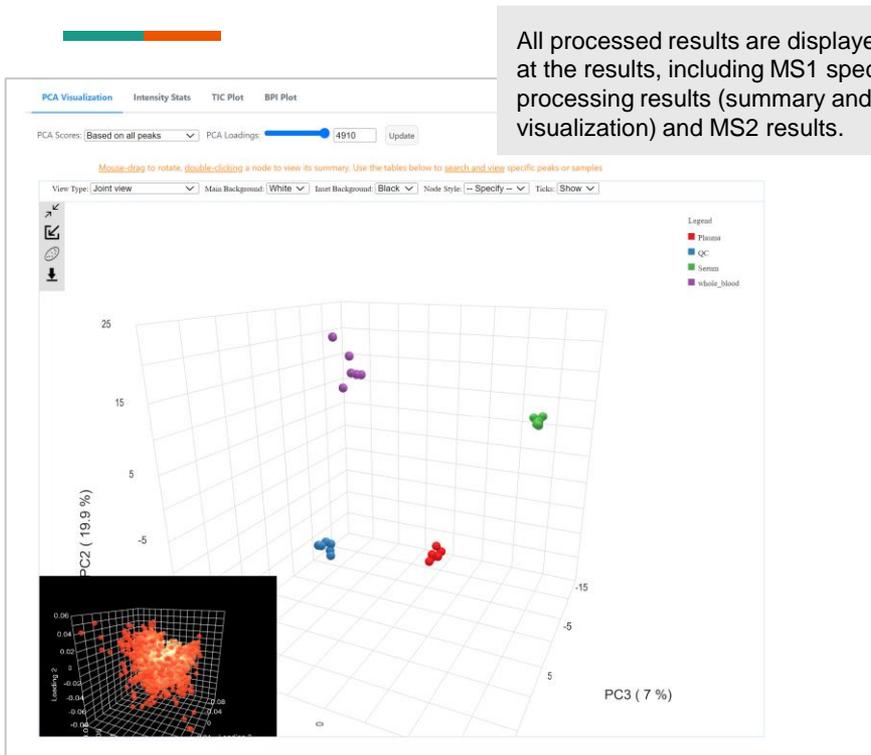
[Refresh Status](#) [Cancel Job](#) [Proceed](#)

The status and processing logs are displayed here. A total of 12 steps will appear for MS1 + DDA:

**Step 1 to 6:** LC-MS1 spectra processing, including data import, peak picking, alignment, gap filling, peak annotation and LC-MS1 results export;

**Step 7 to 12:** LC-MS2 spectra processing, including data import, spectra deconvolution, spectra consensus, database searching and results export.

# 5.5 Result Exploration



All processed results are displayed at the results, including MS1 spectra processing results (summary and visualization) and MS2 results.

Result Summary   Spectra / Sample Table   Feature / Peak Table   **MS/MS Results**

**Raw Spectra Processing Result Summary:**

MetaboAnalyst has finished raw spectra processing with OptiLCMS (1.1.0):

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.

2421 features (49.3%) have been annotated as isotopes.

2319 features (47.22%) have been annotated as adducts.

191 unique formulas have been matched to HMDB database.

1725 potential compounds have been matched to HMDB database.

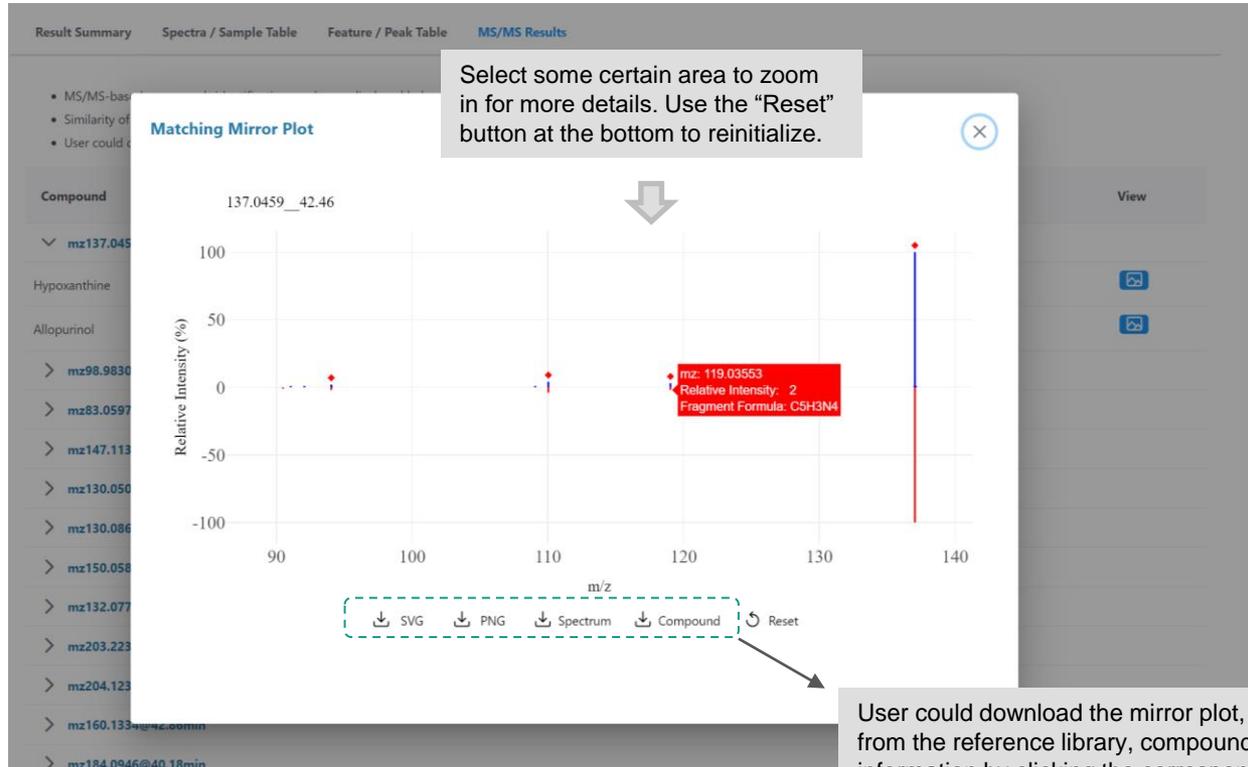
Result Summary   Spectra / Sample Table   Feature / Peak Table   **MS/MS Results**

- 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:

Compound	Formula	Matching Score	InchiKey	Database	View
▼ <b>mz137.0459@42.46min</b>					
Hyposanthine	CSH4NO	84.79	FDGQSTZBFUJBT-UHFFFAOYSA-N	HMDB_experimental	<a href="#">View</a>
Allopurinol	CSH4NO	83.53	OFCNXPDRWKKPPY-UHFFFAOYSA-N	HMDB_experimental	<a href="#">View</a>
> <b>mz98.9830@51.97min</b>					
> <b>mz83.0597@45.80min</b>					
> <b>mz147.1130@35.24min</b>					
> <b>mz130.0501@43.26min</b>					
> <b>mz130.0864@36.31min</b>					
> <b>mz150.0585@43.66min</b>					
> <b>mz132.0770@40.85min</b>					
> <b>mz203.2232@33.63min</b>					

MS2 results are summarized in the specific tab, MS/MS results. All detected features are shown in different rows. User could click the expand icon to see the chemical candidate and click "View" button to see the matching pattern (see next page).

# 5.5 Result Exploration



## MS2 results visualization,

1. Information of fragments will be automatically displayed when mouse hover the fragment;
2. The top (blue) part are users' input, while the bottom (red) parts are from the reference library;
3. All matched fragments will be marked with red diamond at the top.



## 6. LC-MS1 + SWATH-DIA Spectra Processing

In this section, we only emphasize the functionalities specifically related to the SWATH-DIA data processing. All other MS1 associated features have been included with details in the previous section 4, therefore they won't be repeated here. Please read the section 3 and 4 at first if you are not familiar with MetaboAnalyst.

# 6.1 Spectra Files Upload

## LC-MS Spectra Upload

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 [\[example\]](#)
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.

Depending on our server load, spectra processing can take a long time to complete, to avoid waiting:

1. For guest users (default), after job submission, click **Create Bookmark URL** and save the URL so you can return later to check your job status.
2. For registered users, use the buttons on the left panel to manage your projects. Larger data processing and analysis support are available for [subscribed users](#).

Please **Select** all LC-MS1 and MS2 spectral data, then click **Upload** to start uploading. Once the uploading has completed, click **Proceed** to continue.

**Select Mode:**  MS1 Only  MS1+DDA  MS1+SWATH-DIA

For LC-MS1 + SWATH-DIA data processing, you should choose "MS1 + SWATH-DIA"

Reset Proceed

For metadata file, all MS2 spectra (DDA/SWATH-DIA) files must be grouped as "MS2". Otherwise, these files won't be detected as MS2 files.

# 6.2 Integrity Check

## Data Integrity Check:

1. Spectral Format - only mzML, mzXML, mzData and netCDF formats are currently supported;
2. MS Mode - only spectra in **centroid mode** are supported in the online platform. Click **Convert** to centroid your profile data online. **This conversion process will take some time, please be patient ...**
3. If a meta data file is provided:
  - o The first column (spectral names) must match the sample names in the meta-data file;
  - o The second column (group labels) must contain at least two groups (not including QC), each containing  $\geq 3$  replicates.

For MS2 data processing, MS level is indicated at the Integrity Check page

Spectra	Centroid	Size (MB)	MS Level	Group	Convert	Include
Covid_Cov_19_MS.mzML	True	21.8	MS1	COVID		
Covid_Cov_18_MS.mzML	True	21.8	MS1	COVID		
Covid_Cov_23_MS.mzML	True	21.6	MS1	COVID		
Covid_Cov_24_MS.mzML	True	22.0	MS1	COVID		
Covid_Cov_21_MS.mzML	True	22.0	MS1	COVID		
Covid_Cov_22_MS.mzML	True	21.6	MS1	COVID		
Covid_Ct_1_MS2.mzML	True	78.5	MS2	MS2		
Covid_Ct_2_MS2.mzML	True	76.5	MS2	MS2		
Covid_Cov_17_MS2.mzML	True	83.0	MS2	MS2		
Covid_Cov_16_MS2.mzML	True	84.2	MS2	MS2		

## 6.3 SWATH window check

### SWATH-DIA Design Integrity Check

MetaboAnalyst will detect the design of SWATH-DIA spectral acquisition automatically. However, some spectral data below is all 0.0, user must manually specify design.

Index	Lower M/Z	Upper M/Z
1	50.00	100.00
2	99.00	150.00
3	149.00	200.00
4	199.00	250.00
5	249.00	300.00
6	299.00	350.00
7	349.00	400.00
8	399.00	450.00
9	449.00	500.00
10	499.00	550.00

Lower and upper m/z value of the specific SWATH window

Index of SWATH windows

### SWATH-DIA Design window:

This is the most critical parameter for SWATH-DIA spectra processing. MetaboAnalyst could automatically detect the SWATH window information from uploaded files. However, some raw spectra files may not contain the information. In this case, MetaboAnalyst could only identify the total number of windows in an experiment cycle, and users must manually specify the SWATH Window information here.

11	549.00	600.00
12	599.00	650.00
13	649.00	700.00
14	699.00	750.00
15	749.00	800.00
16	799.00	850.00
17	849.00	900.00
18	899.00	950.00
19	949.00	1,000.00

[Confirm](#) Click "Confirm" to continue

# 6.4 Parameter Setting

## 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 quantification (see [benchmarking results](#)). The source code for raw spectral processing is now available as the [OptiLCMS](#) 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  Negative

**3. Peak Annotation** Adducts: [View](#)  
More options: [View](#)

**4. MS2 Processing**

ppm for MS2: 10.00

Filtering value: 200.00

Similarity Method:  Dot Product  Spectral Entropy

Target Peaks:  Significant Ones  All Features

MS2 Database: HMDB Experimental

**5. Contaminant Removal**  [View](#)

**6. Blank Subtraction**

Parameters for MS2 data processing. Please refer to section 5.3 for more explanations on these parameters.

Submit Job

# 6.5 Jos Status

**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: 9689  
Bookmark Link: [Create Job URL](#)  
Current Status: Running  
Priority: Level 1  
Parameters: [Save](#)  
Job Progress:

**Step 10/12: MS/MS spectra database searching is starting ...**  
this step may take some time.  
mSet <- PerformDBSearchingBatch (mSet,  
+ ppm1 = 5.0,  
+ ppm2 = 10.0,  
+ rt\_tol = 5,  
+ database\_path = '/home/glassfish/sqlite/MS2ID\_Complete\_v09102023.sqlite',  
+ use\_rt = FALSE, enableNL = FALSE, ncores = 4L, useEntropy = FALSE,  
+ databaseOptions =c('hmdb\_exp'));  
Loading required package: parallel  
==== Database searching against MS2ID\_Complete\_v09102023.sqlite started ====

Text Output:

Output File: [Status Text](#) 2024-03-10 01:20:18

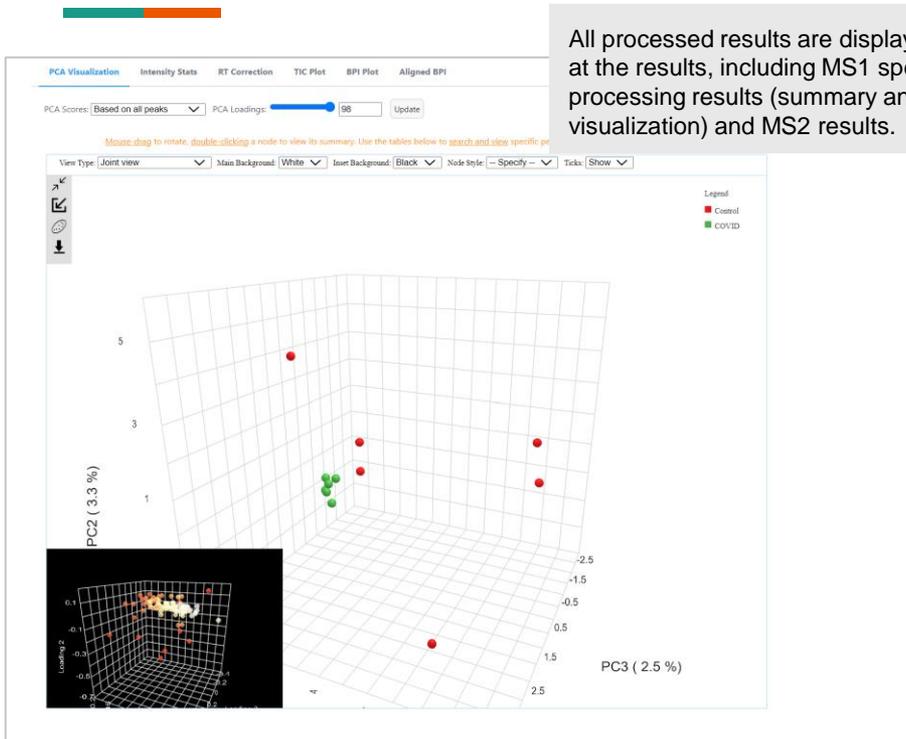
[Refresh Status](#) [Cancel Job](#) [Proceed](#)

The status and processing logs are displayed here. A total of 12 steps will appear for MS1 + SWATH-DIA:

**Step 1 to 6:** LC-MS1 spectra processing, including data import, peak picking, alignment, gap filling, peak annotation and LC-MS1 results export;

**Step 7 to 12:** LC-MS2 spectra processing, including data import, spectra deconvolution, spectra consensus, database searching and results export. Different from DDA, spectra deconvolution is required for DIA spectra dataset.

# 6.6 Result Exploration



All processed results are displayed at the results, including MS1 spectra processing results (summary and visualization) and MS2 results.

Result Summary | Spectra / Sample Table | **Feature / Peak Table** | MS/MS Results

- For isotopes/adducts annotation, the matching is based on the m/z value of its corresponding parent ion. Otherwise, it is considered as in the format of the primary ion.
- All compounds/formulas are matched to [HMDB](#) (v-5) based on the mass error (ppm value) for raw spectra processing.
- Intensity is average of all samples. Coefficient of variation (CV) is also the summarized based on all samples.
- When group information is provided, p values will be calculated with t-test/ANOVA based on log transformed data.

m/z	RT/s	Intensity	CV (%)	P values	FDR	Annotations	Putative IDs	View
160.8422	79.23	11036.9	46.45	1.5068802E-4	0.01476743			<a href="#">View</a>
217.0301	84.55	215276.0	10.02	0.0027741963	0.0659002			<a href="#">View</a>
117.0555	138.64	20089.0	25.41	0.0029012625	0.0659002		<a href="#">View</a>	<a href="#">View</a>
96.9598	72.13	38909.9	14.77	0.0036270983	0.0659002			<a href="#">View</a>
678.5083	81.66	9218.9	69.89	0.0065294412	0.0659002		<a href="#">View</a>	<a href="#">View</a>
215.0331	84.66	577626.3	8.77	0.0082779721	0.0659002		<a href="#">View</a>	<a href="#">View</a>

MS2 results are summarized in the tab, MS/MS results. (see more details from section 5.4 and 5.5).

Result Summary | Spectra / Sample Table | Feature / Peak Table | **MS/MS Results**

- 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.

Compound	Formula	Matching Score	InchiKey	Database	View
▼ <b>mz146.0457@74.06sec</b>					
L-Glutamic acid	CSH9ND4	57.16	WHUUTDBJRXKMK-VIHMHEASA-N	HMDB_experimental	<a href="#">View</a>
O-Acetyserine	CSH9ND4	57.16	VZXPDPZARILFGX-BYPYZUCNSA-N	HMDB_experimental	<a href="#">View</a>
▶ <b>mz160.9202@83.59sec</b>					

« < 1 > » 50

[Download Page](#)

Click "Download Page" to download results

# 7 Result Downloading

**Download Results & Start New Journey**

Please download the results (tables and images) from the **Results Download** tab below. The **Download.zip** contains all the files in your home directory. You can also generate a report.

[Results Download](#) [Start New Journey](#)

[Generate Report](#)

Download.zip	compound_msn_results.csv
Rhistory.R	mirror_plotting_1_0_72.json
spectra_3d_loading.json	mirror_plotting_1_0_72.png
metaboanalyst_input.csv	BPIS_72.png
TICS_72.png	Peak_Intensity.png
peak_feature_summary.csv	PCA.png

[Logout](#)

All results can be downloaded here.



## In summary

If you have any questions, please read/post into OmicsForum ([www.omicsforum.ca](http://www.omicsforum.ca))

Or contact us:

[zhiqiang.pang\[at\]xialab.ca](mailto:zhiqiang.pang@xialab.ca)

[jeff.xia\[at\]xialab.ca](mailto:jeff.xia@xialab.ca)

- Raw spectra files must be saved in common open-source formats and uploaded individually as separate zip files.
- LC-MS spectra data is mandatory, while MS2 is optional. Upon data uploading, MetaboAnalyst 6.0 automatically validates the status of MS files.
- For SWATH-DIA data, the SWATH window design is automatically extracted from the spectra. If the related information is missing, users will be prompted to enter the window design manually.
- On the parameters setting page, users are given the option to choose the default auto-optimized centWave algorithm or use the *asari* algorithm for LC-MS data processing.
- If MS2 data is included, spectra deconvolution, consensus, and database searching are performed automatically, using the MS features as target list. Once the spectra processing is complete, users can explore both MS and MS2 data processing results.