



MetaboAnalyst 5.0

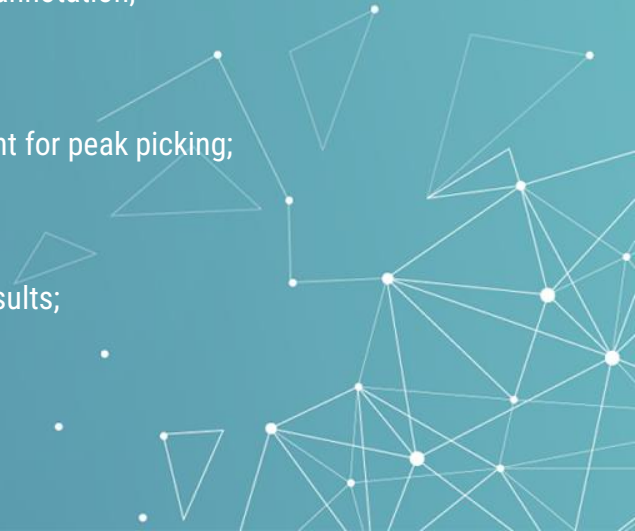
A Web-based Tool for streamlined
Metabolomics Data Analysis

2022.07.12

1. MS Spectral Processing

The **MS Spectral Processing** module of MetaboAnalyst fills the important gap of raw spectral processing of high-resolution LC-MS data that was previously only available for users in our MetaboAnalystR package.

Highlights:

- Support raw spectra data processing for peak picking, alignment, gap filling and annotation;
 - Support for fast and automated parameters optimization;
 - Support for customized parameters and centwave, matchedFilter and Massifquant for peak picking;
 - Multiple common formats are supported (mzML, mzXML, mzData and NetCDF);
 - Resumable pipeline was embedded for users to manually and quickly tune the results;
- 
- A decorative network diagram on the right side of the slide, consisting of white dots connected by thin white lines, forming a complex web-like structure against the teal background.

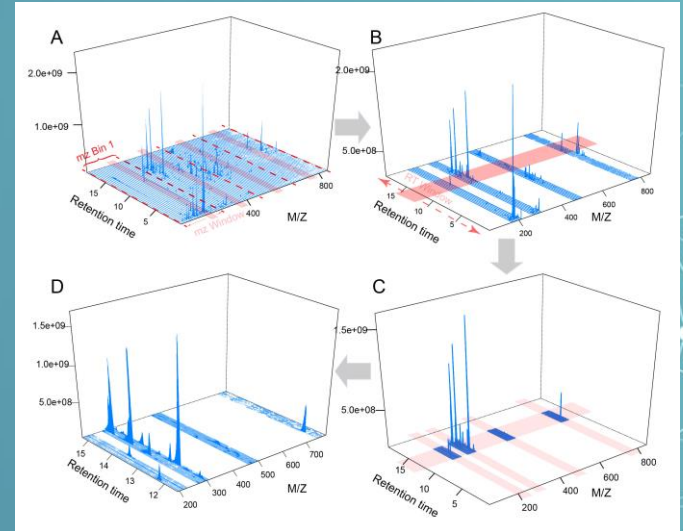
1.0 Knowledge & Background

- This module is designed to process the raw spectra data file with an R package, [OptiLCMS](#), as the core processing engine.
- Three algorithms are supported to do the pre-processing (peak picking), including *centWave* (for high-resolution Mass Spectrometer), *mathedFilter* (for low-resolution MS) and *Massifquant* (which is more sensitive to low-intensity peaks).
- The automated optimization option could optimize the parameters for *centWave* automatically to give the optimal results based on users' data. The optimization pipeline was initially published in [MetaboAnalystR 3.0](#), which is briefly described as below,

The 'automated optimization' pipeline would extract the most abundant MS areas (Regions of Interest, ROIs) across the whole spectra as the training spectra (as shown in the left Figure).

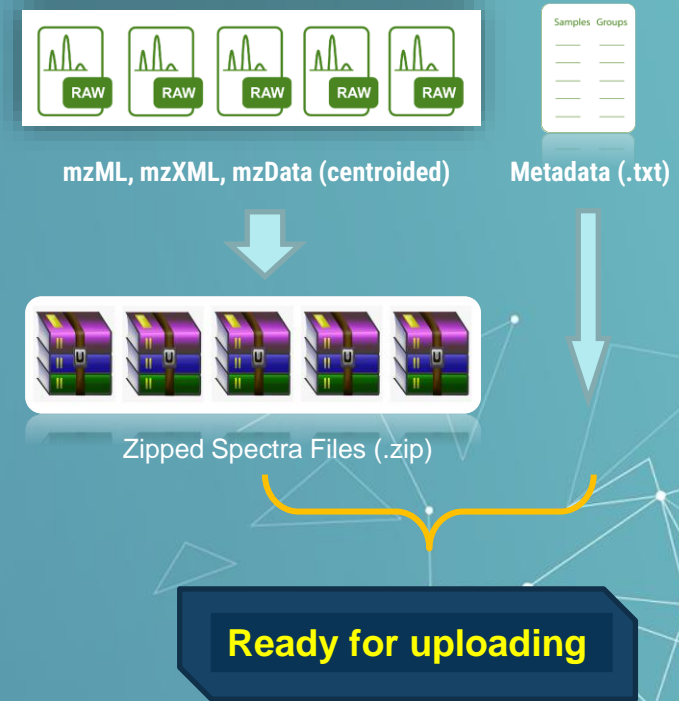
Then, a design-of-experiment (DoE) optimization will be executed to find out the combination of parameters with the most well-behaved shape and stable peak groups for the following whole spectrum detection.

Besides, users could avoid the potential overwhelming on the optimization steps from long-retention time signals (usually contaminants or noise) by removing them (see [1.5](#)).



1.1 Preparation for MS Spectral Processing

1. Users must upload their spectra as individual zip files - one zip (.zip) per spectrum [max: 200 spectra].
2. Optionally but strongly recommended, users can upload a metadata file uploaded as a plain text (.txt) file containing two columns - spectral names and group labels.
3. After their data is successfully uploaded, a data integrity check is performed to verify the correct data format (mzML, mzXML, mzData + centroided) and metadata information. Please check [1.4](#) about how to centroid your data.



1.2 Register & Login (Optional)

NOTE: Register or Login is optional. You can upload your files directly, but the jobs for registered users will be kept for 180 days.

MetaboAnalyst 5.0 - user-friendly, streamlined metabolomics data analysis

Module Overview

| Input Data Type | Available Modules (click on a module to proceed, or scroll down for more details) | | | | | |
|--|---|--|--------------------------|---------------------------|------------------|-----------------|
| Raw Spectra (mzML, mzXML or mzData) | | | LC-MS Spectra Processing | | | |
| MS Peaks (peak list or intensity table) | | | Functional Analysis | Functional Meta-analysis | | |
| Annotated Features (compound list or table) | | Enrichment Analysis | Pathway Analysis | Joint-Pathway Analysis | Network Analysis | |
| Generic Format (csv or .txt table files) | Statistical Analysis (one factor) | Statistical Analysis (metadata table) | Biomarker Analysis | Statistical Meta-analysis | Power Analysis | Other Utilities |

Click here to start

LC-MS Spectra Upload

MetaboAnalyst currently supports mzML, mzXML, CSV or mzData formats in centroid mode. Quality control (QC) spectra are not required but recommended. QC should start with "QC_" or marked as "QC" in meta data. BLANK should be marked as "BLANK" in meta data for subtraction. The following two data types are allowed:

- spectrum (max: 200 spectra),
- columns - spectral names and group labels (example)

Please **Select** all files, then click **Upload** to start. Once the upload has completed, click **Proceed** to continue.

Click login to register/login

Log in

MetaboAnalyst 5.0 - user-friendly, streamlined metabolomics data analysis

Project Initialized successfully!
Please click "Start" to start

Use the panel below to add a new project. Start by clicking **Add New** to add a new project. In order to purge a project, see the **Delete** button in the **Projects Available** table. Once you start a project, you will be prompted to name the project.

Click "start" action to begin the desired project or "load" a previously saved project.

| Project ID | Title | Description | Type | Date created | Action |
|------------|-------------|----------------------|------|--------------|--------------|
| 1 | bug_removal | test | raw | 2022/02/03 | Load Delete |
| 2 | test demo | this is test project | raw | 2022/06/12 | Start Delete |

Click buttons to operate your projects

Click to create new projects

MetaboAnalyst 5.0 - user-friendly, streamlined metabolomics data analysis

Log in to start a new analysis or resume your previous analysis

Email *

Password *

Log in

Register here

Create account

Forgot password?

1.3 Spectra Upload

The screenshot shows the 'LC-MS Spectra Upload' page. On the left is a navigation menu with options: Upload, Spectra check, Spectra processing, Job status, Spectra result, Download, and Exit. The main content area is titled 'LC-MS Spectra Upload' and contains instructions on supported file formats (mzML, mzXML, CDF, or mzData in centroid mode) and quality control (QC) requirements. It lists two steps: 1. [Required] Spectra uploaded as individual zip files (max 200 spectra) and 2. [Optional] Meta data uploaded as a plain text file with two columns. Below this, it states that processing can take a long time and provides instructions for guest users (click 'Create Bookmark URL') and registered users (use left panel buttons). A section titled 'Please Select all files, then click Upload to start. Once the upload has completed, click Proceed to continue.' features a '+ Select' button, a 'Reset' button, and a 'Proceed' button. A 'Try our example data' section offers two datasets with radio buttons and 'Download' links for 'Dropbox' and 'GoFile'. A 'Submit' button is at the bottom.

LC-MS Spectra Upload

MetaboAnalyst currently supports [mzML](#), [mzXML](#), [CDF](#) or [mzData formats in centroid mode](#). Quality control (QC) spectra are not required but recommended. QC should start with "QC_" or marked as "QC" in meta data. BLANK should be marked as "BLANK" in meta data for subtraction. The following two data types are allowed:

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

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.

Please **Select** all files, then click **Upload** to start. Once the upload has completed, click **Proceed** to continue.

+ Select

Reset **Proceed**

Try our example data

| Description | Download |
|---|---|
| <input checked="" type="radio"/> 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 Lloyd-Price et al. | Dropbox GoFile |
| <input type="radio"/> An experimental Malaria metabolomics dataset (UPLC-Q/E-ESI ⁺ , HILIC) between two immune status (Native vs. Semi-immune) from Li et al. 15 samples (12 Samples and 3 QCs) are included. | Dropbox GoFile |

Submit

Click **"Select"** to start uploading your zipped spectra files.

Click **"Proceed"** to move onto the Data Integrity Check.

TIPS: You can choose to upload multiple spectra files at once, but please upload your spectra all at a time to avoid any potential exceptions caused by internet connection issues.

Click **"Submit"** to Submit an example data.

1.4 Data Integrity Check

R Command History
appears in real-time and is
ordered sequentially

TIPS: We encourage users to
centroid data before uploading.
Here are several approaches
recommended to centroid your data.

1. ProteoWizard :

For GUI: Add 'Peak Picking' as the 1st filter;
For Command: `docker run -it --rm -e
WINEDEBUG=all -v /FILE_PATH/:/data
chamb/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. OptiLCMS:

Install this R package from [here](#) and do the
centroiding with function "**CentroidMSData**".

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

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

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

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 ≥ 3 replicates.

| Spectra | Centroid | Size (MB) | Group | Convert | Include |
|----------------|----------|-----------|------------|---------|-------------------------------------|
| Semi_025.mzML | True | 15.7 | Semi_immue | | <input checked="" type="checkbox"/> |
| Semi_091.mzML | True | 15.3 | Semi_immue | | <input checked="" type="checkbox"/> |
| Semi_157.mzML | True | 16.0 | Semi_immue | | <input checked="" type="checkbox"/> |
| Semi_061.mzML | True | 15.6 | Semi_immue | | <input checked="" type="checkbox"/> |
| Semi_143.mzML | True | 15.7 | Semi_immue | | <input checked="" type="checkbox"/> |
| Semi_045.mzML | True | 15.6 | Semi_immue | | <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"/> |
| Naive_109.mzML | True | 15.0 | Naive | | <input checked="" type="checkbox"/> |
| Naive_127.mzML | | | Naive | | <input checked="" type="checkbox"/> |
| Naive_139.mzML | | | Naive | | <input checked="" type="checkbox"/> |
| Naive_007.mzML | | | Naive | | <input checked="" type="checkbox"/> |
| Naive_027.mzML | | | Naive | | <input checked="" type="checkbox"/> |
| Naive_071.mzML | | | Naive | | <input checked="" type="checkbox"/> |

Next

1.5 Parameter Selection

LC-MS Spectra Processing

The parameter optimization is based on regions of interest (ROIs) to avoid the time-consuming step of recursive peak detection using complete spectra. The procedure can significantly improve both the quality of peak detection and quantification compared to results obtained using the default XCMS parameters ([benchmarking results](#)). The algorithm is now available as the [OptiLCMS R](#) package for local installation or further extension.

- 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).

LC-MS Platform Generic

Parameter Setting Default/manual Auto-optimized

Method: centWave

min_peakwidth: 5.0

max_peakwidth: 30.0

ppm: 5.0

mzdiff: 0.01

More options View

Method: loess

Bandwidth: 10.0

minFraction: 0.8

More options View

Polarity: positive

Adducts: View

More options View

1. Peak Picking

2. Peak Alignment

3. Peak Annotation

4. Contaminant Removal

5. Blank Subtraction

Submit Job

Show R Commands

1. Adjust the following parameters according to the LC-MS instrument/extraction methods used.

2. Click **Submit Job** to perform the spectra processing.

TIP1: Default Parameters setting option is 'customized'. If you are not a parameter expert, please try to use the automated optimization pipeline.

TIP2: The automated pipeline only optimizes the 'centWave' algorithm. Other algorithms, like Massifquant, is more sensitive to spectral signals and is only available from 'Default/Manual' mode.

TIP3: Contaminant removal is only functional for the 'Automated' pipeline. It will automatically remove potential contaminants before performing parameters' optimization. Please view your data before you decide to submit your job.

1.6 Job Status View

TIPs: If you are not logged in as a registered user, you can create a URL link to check your result. Otherwise, you will lose your result immediately once you exit or close this page. The bookmark URL will expire 14 days later.

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 FAQs or other static web pages), as **multiple tabs/windows will interfere with each other**, leading to unpredictable results.

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.

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

The screenshot displays the 'Job Status View' interface. At the top left is a navigation menu with options: Upload, Spectra check, Spectra processing, Job status, Spectra result, Download, and Exit. The main content area shows job details for Job ID 10702. A 'Job Status' button is positioned above the details. The details include: 'Bookmark Link: Create Job URL', 'Current Status: Running', 'Priority: Level 1', 'Parameters: Save', and 'Job Progress: 10%' with a progress bar. Below this is a 'Text Output' section showing a list of import completion messages for various files. At the bottom, there are three buttons: 'Refresh Status', 'Cancel Job', and 'Proceed'.

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

1.7 Exploring the Results -1

Processing Results:

PCA Visualization | Intensity Stats | RT Correction | TIC Plot | BPI Plot | Aligned BPI

PCA Scores: Based on all peaks | PCA Loadings: 4075 | Update | Download: --Please Select--

Mouse-drag to rotate, double-clicking a node to view its summary. Use the tables below to search and view specific peaks or samples

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

Raw Spectra Processing Result Summary:
MetaboAnalyst has finished raw spectra processing with OptiLIMS (1.0.5):
There are 15 samples of 3 groups (Naive, QC, Semi_immue) included for processing!
All of 4075 features have been detected and aligned across the whole sample list.
The mass deviation of this study was estimated/set as 5 ppm.
740 features (18.16%) have been annotated as isotopes.
899 features (22.06%) have been annotated as adducts.
611 unique formulas have been matched to HMDB database.
1561 potential compounds have been matched to HMDB database.

Explore other graphical summaries of the spectral processing results here.

TIPs: If you are not satisfied with the result, return to the Parameters' page from the 'Navigation Tree' at the Left side of this page to resubmit your job with different algorithms or parameters.

You could interactively view the data results with 3D PCA and check the main loadings for the distribution (Double click the nodes to view TIC/EIC)

There are one summary box and two results tables, one for the Spectra and the other for the Features.

Introduce more about these Result tables from the next pages.

1.8 Exploring the Results -2

Result Summary Spectra / Sample Table Feature / Peak Table

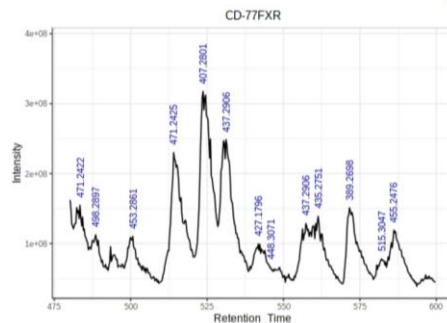
| Spectra ↑↓ | Group ↑↓ | Peaks No. ↑↓ | Missing (%) ↑↓ | RT Range | m/z Range | View |
|------------|------------|--------------|----------------|-------------|-----------------|------|
| | | | | | | |
| Naive_007 | Naive | 3510 | 13.87 | 9.15~292.71 | 85.065~1273.52 | |
| Naive_027 | Naive | 3517 | 13.69 | 9.15~292.71 | 85.065~1273.52 | |
| Naive_071 | Naive | 3433 | 15.75 | 9.15~292.71 | 85.065~1273.52 | |
| Naive_109 | Naive | 3167 | 22.28 | 9.15~292.71 | 85.065~1273.52 | |
| Naive_127 | Naive | 3438 | 15.63 | 9.15~292.71 | 85.065~1273.52 | |
| Naive_139 | Naive | 3450 | 15.34 | 9.15~292.71 | 85.065~1273.52 | |
| QC_001 | QC | 3349 | 17.82 | 9.15~292.71 | 85.084~1264.131 | |
| QC_003 | QC | 3385 | 16.93 | 9.15~292.71 | 85.065~1264.131 | |
| QC_005 | QC | 3397 | 16.64 | 9.15~292.71 | 85.065~1266.511 | |
| Semi_025 | Semi_immue | 3588 | 11.95 | 9.15~292.71 | 85.065~1264.131 | |
| Semi_045 | Semi_immue | 3667 | 10.01 | 9.15~292.71 | 85.065~1273.52 | |
| Semi_061 | Semi_immue | 3631 | 10.9 | 9.15~292.71 | 85.065~1264.131 | |
| Semi_091 | Semi_immue | 3573 | 12.32 | 9.15~292.71 | 85.065~1264.131 | |
| Semi_143 | Semi_immue | 3596 | 11.75 | 9.15~292.71 | 85.065~1264.131 | |
| Semi_157 | Semi_immue | 3620 | 11.17 | 9.15~292.71 | 85.065~1264.131 | |

<< < 1 > >> 20 ▾

> Download Page

Click the **View** button to see TIC of the corresponding spectra.

Total Ion Chromatogram



The labels marked in the TIC is the corresponding m/z value of the base ion in the peak.

1.9 Exploring the Results -3

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 |
|-----------|---------|--------------|-----------|---------------|-----|--|--------------|------|
| 1190.7142 | 113.32 | 192755.1 | 54.28 | 2.3205802E-15 | 0.0 | | | |
| 768.4118 | 71.34 | 298137.8 | 43.58 | 2.5673008E-15 | 0.0 | [M+Na+NaCOOH]+ 677.43 [M+H-CH2]+ 781.419 | | |
| 759.723 | 60.93 | 124919.2 | 52.69 | 3.624273E-15 | 0.0 | | | |
| 438.6319 | 72.01 | 4262468.0 | 48.17 | 5.093338E-15 | 0.0 | [2M+Na]+ 207.821 | | |
| 1008.5172 | 116.55 | 378720.3 | 48.79 | 6.2724633E-15 | 0.0 | | | |
| 913.7906 | 108.08 | 128345.3 | 52.76 | 2.3088964E-14 | 0.0 | | | |
| 1200.3439 | 71.5 | 92447.1 | 51.53 | 2.6098027E-14 | 0.0 | | | |

This table is showing all MS features. Click the button of Putative IDs show the potential Chemical IDs of the features towards HMDB.

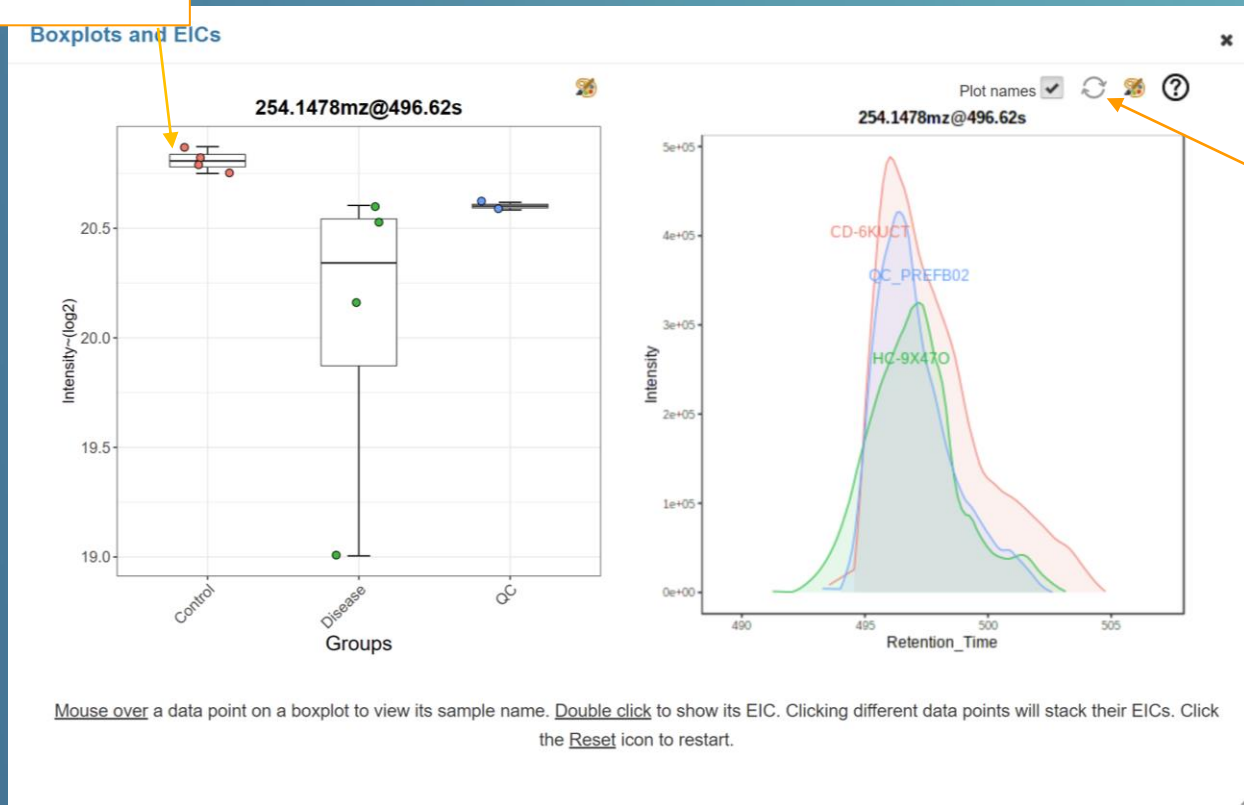
Putative IDs ×

| Formulas | Compounds |
|-----------------------------|---|
| C24H54NO10P | PS(14:0/14:1/92%); PS(14:1/92/14:0) |

2. Click the button under **View** to see a dynamic Extracted Ion Chromatogram for the selected feature (see next page).

1.10 Exploring the Results -3

1. Click the node in the boxplot to generate the EIC cumutatively and dynamically.



TIP1: If the plotting failed, please clean the cache of your browser or use another browser.

2. Click this 'reset' icon to restart the generation of EIC plot.

3. Scroll down to the bottom of page and click "**Proceed**" to view the Downloads page.

1.11 Result Downloading & New Journey

Download Results & Start New Journey

Please download the results (tables and images) from the **Results Download** tab below. The **Download.zip** also generate a **PDF analysis report** using the button. Finally, you can continue to explore other compatible m

Results Download | **Start New Journey**

Generate Report

| | |
|-----------------------|--------------------------|
| Download.zip | spectra_3d_loading.json |
| Rhistory.R | PCA.png |
| BPIS_72.png | annotated_peaklist.csv |
| TICS_72.png | metaboanalyst_input.csv |
| Peak_Intensity.png | peak_feature_summary.csv |
| Adjusted_RT.png | scores3D.png |
| Adjusted_BP1.png | loadings3D.png |
| spectra_3d_score.json | |

Logout



Results Download | **Start New Journey**

Statistical Analysis (one factor)

- Biomarker Analysis
- Statistical Analysis (metadata table)
- Power Analysis

General Statistics

- Enrichment Analysis
- Pathway Analysis

Targeted Metabolomics

- Functional Analysis

Global Metabolomics

GO!

1.12 PDF REPORT

Metabolomic Data Analysis with MetaboAnalyst 5.0

Name: guest5781233943536353632

January 13, 2021

1 Raw Spectra Processing

Global or untargeted metabolomics is increasingly used to investigate metabolic changes of various biological or environmental systems in an unbiased manner. Liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS) has become the main workhorse for global metabolomics. The typical LC-HRMS metabolomics workflow involves spectra collection, raw data processing, statistical and functional analysis.

MetaboAnalyst aims to provide an efficient pipeline to support end-to-end analysis of LC-HRMS metabolomics data in a high-throughput manner.

This module is designed to provide an automated workflow to process the raw spectra. 5 steps including parameters optimization/customization, peak picking, peak alignment, peak gap filling and peak annotation.

1.1 Reading and Processing the Raw Data

MetaboAnalyst MS Spectral Processing Module accepts several common MS formats including mzXML, mzML, mzData, CDF formats. Other vendor format will be supported soon. But all of them have to be centroided before processing. The Data Integrity Check is performed before the data processing starts. The basic information of all spectra is summarized in Table 1 shows the details of all spectra.

Table 1: Summary of data uploading results

| Spectrum | Centroided | Size (MB) | QC-pass |
|--------------------|------------|-----------|---------|
| 1 HPLC-DAD.mzML | True | 3.75 | Control |
| 2 CD-77FXR.mzML | True | 3.62 | Control |
| 3 CD-9059Y.mzML | True | 4.11 | Control |
| 4 CD-9059P.mzML | True | 3.7 | Control |
| 5 HC-9847D.mzML | True | 4.31 | Disease |
| 6 HC-9847D.mzML | True | 3.99 | Disease |
| 7 HC-AMR5T.mzML | True | 3.84 | Disease |
| 8 HC-ALP8B.mzML | True | 4.25 | Disease |
| 9 QC-PREFA02.mzML | True | 4.05 | QC |
| 10 QC-PREFB02.mzML | True | 4.03 | QC |

2.3 Peak Intensity Statistics

The general peaks' intensity is analyzed from different spectral files to show the peaks' intensity distribution. The statistics of all spectral peaks is displayed in Figure 3, as below.

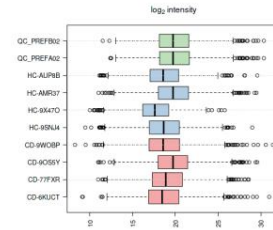


Figure 3: Peak Intensity Statistics of all spectral files.

TIPS: Raw spectral processing results will be reported as a PDF file from 'Generate Report' button in the previous page. Please try to switch to other modules and generate the corresponding report in different modules.

Thanks

*If you have any questions please read through the FAQs or contact us at
[Zhiqiang.pang\[at\]xialab.ca](mailto:Zhiqiang.pang@xialab.ca) or [Jeff.xia\[at\]xialab.ca](mailto:Jeff.xia@xialab.ca)*

