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;

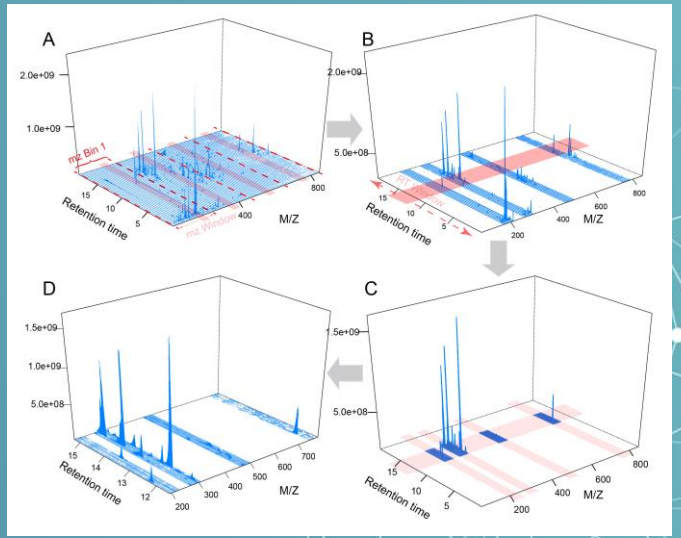
1.0 Knowledge & Background

- 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).

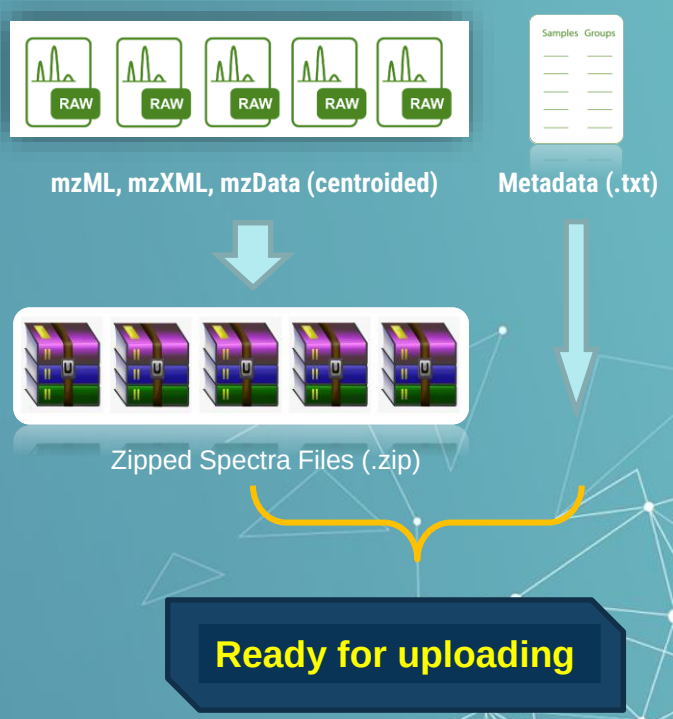
Then, a design-of-experiment (DoE) optimization will be executed to find out the combination of parameters with the most well-behaved shape and stable peak groups for the following whole spectrum detection.

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).



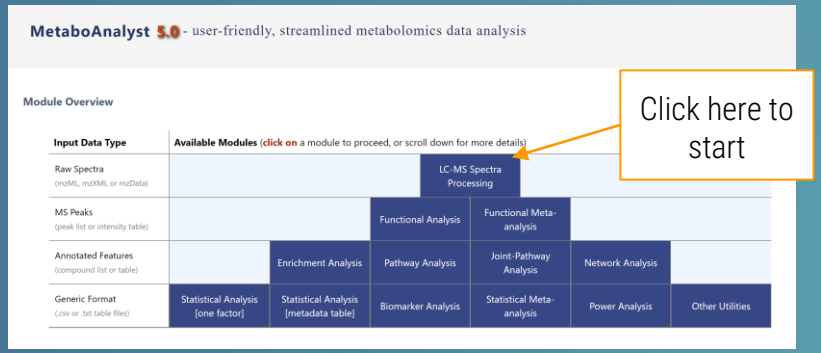
1.1 Preparation for MS Spectral Processing

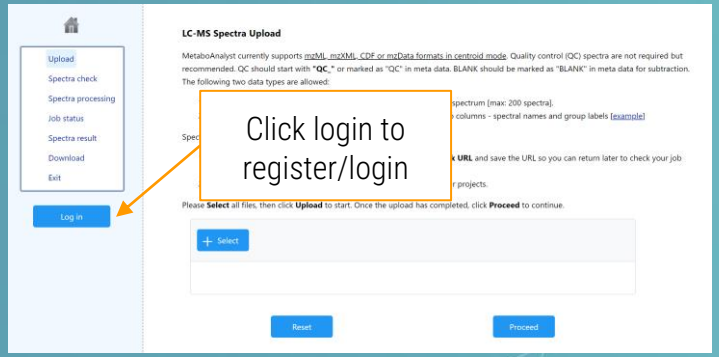
1. Users must upload their spectra as individual zip files - one zip (.zip) per spectrum [max: 200 spectra].
2. Optionally but strongly recommended, users can upload a metadata file uploaded as a plain text (.txt) file containing two columns - spectral names and group labels.
3. After their data is successfully 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.

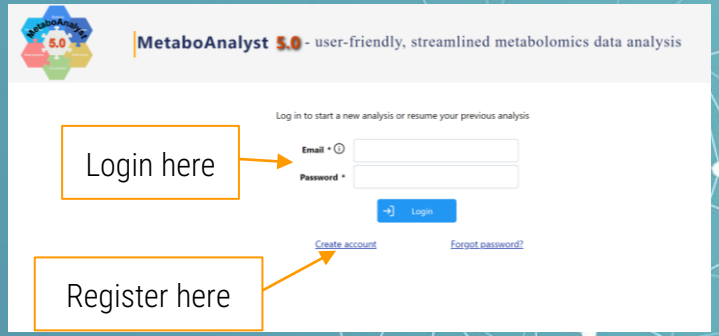


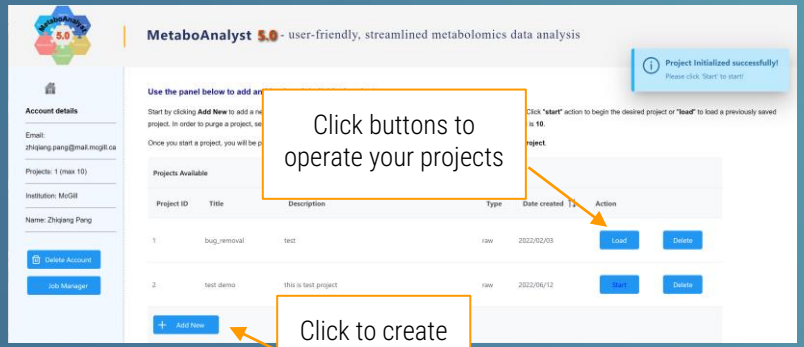
1.2 Register & Login (Optional)

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

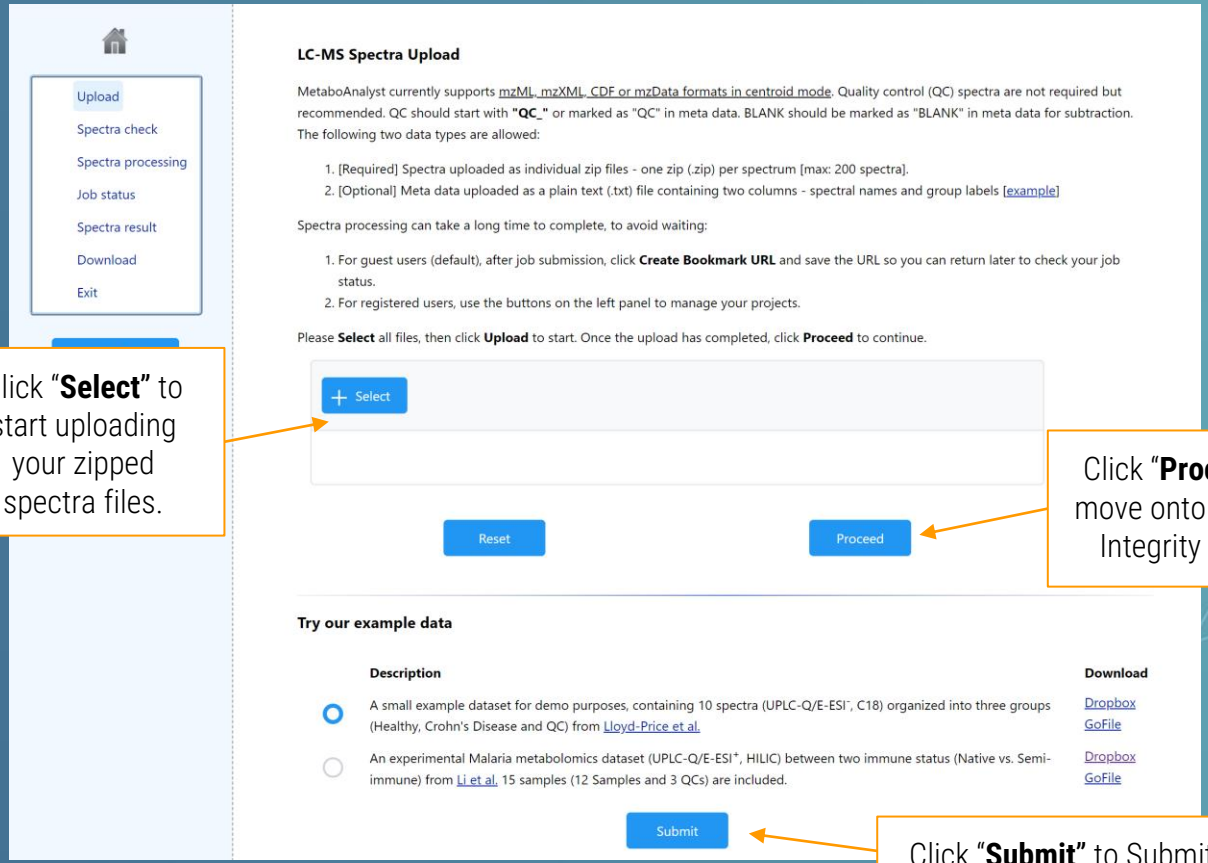








1.3 Spectra Upload



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

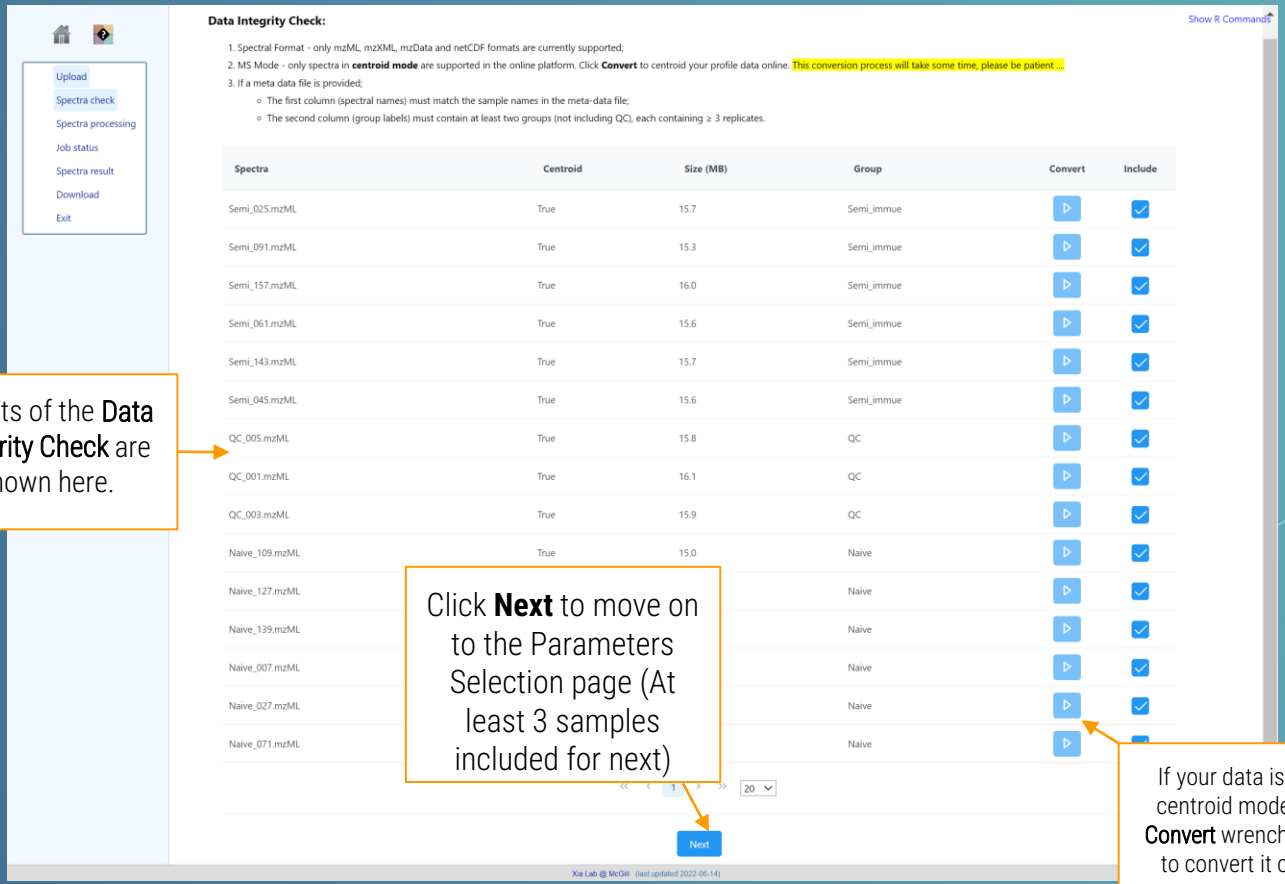
Click "**Proceed**" to move onto the Data Integrity Check.

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



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

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

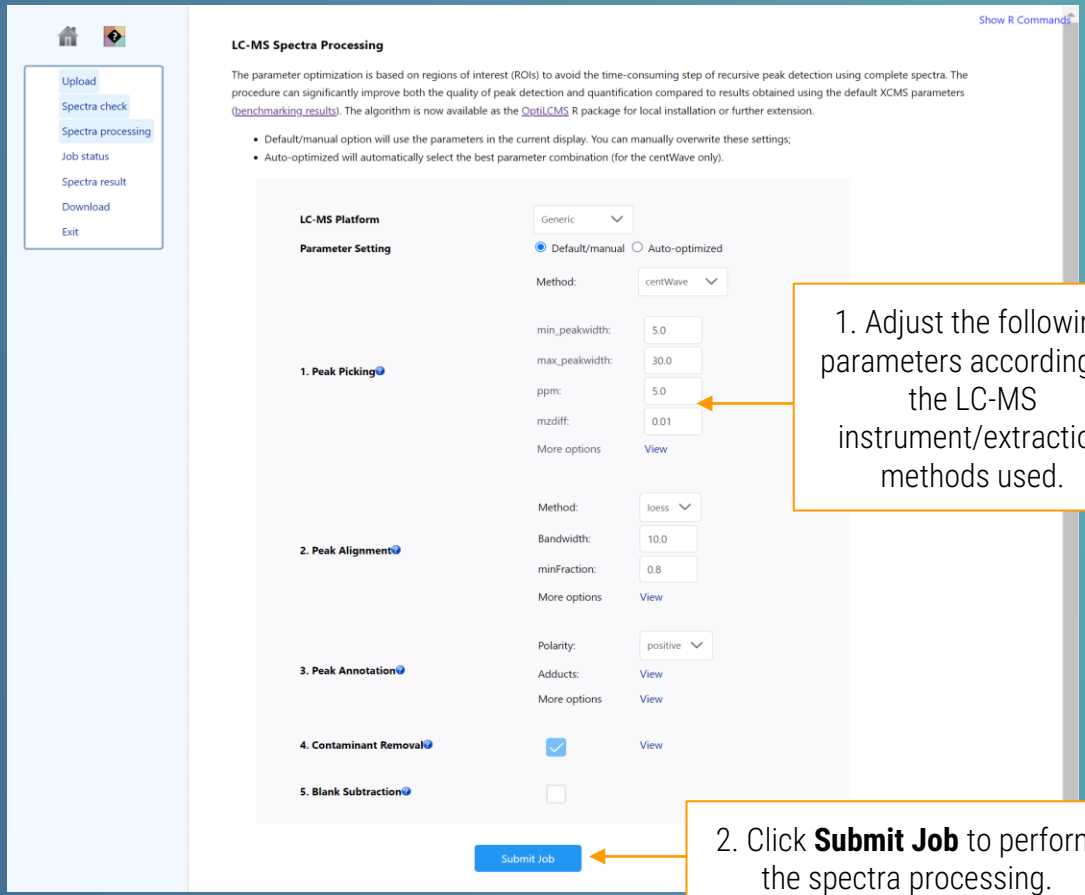
If your data is not in centroid mode, click **Convert** wrench button to convert it online.

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 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. OptiLCMS:
Install this R package from here and do the centroiding with function "***CentroidMSData***".

1.5 Parameter Selection



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

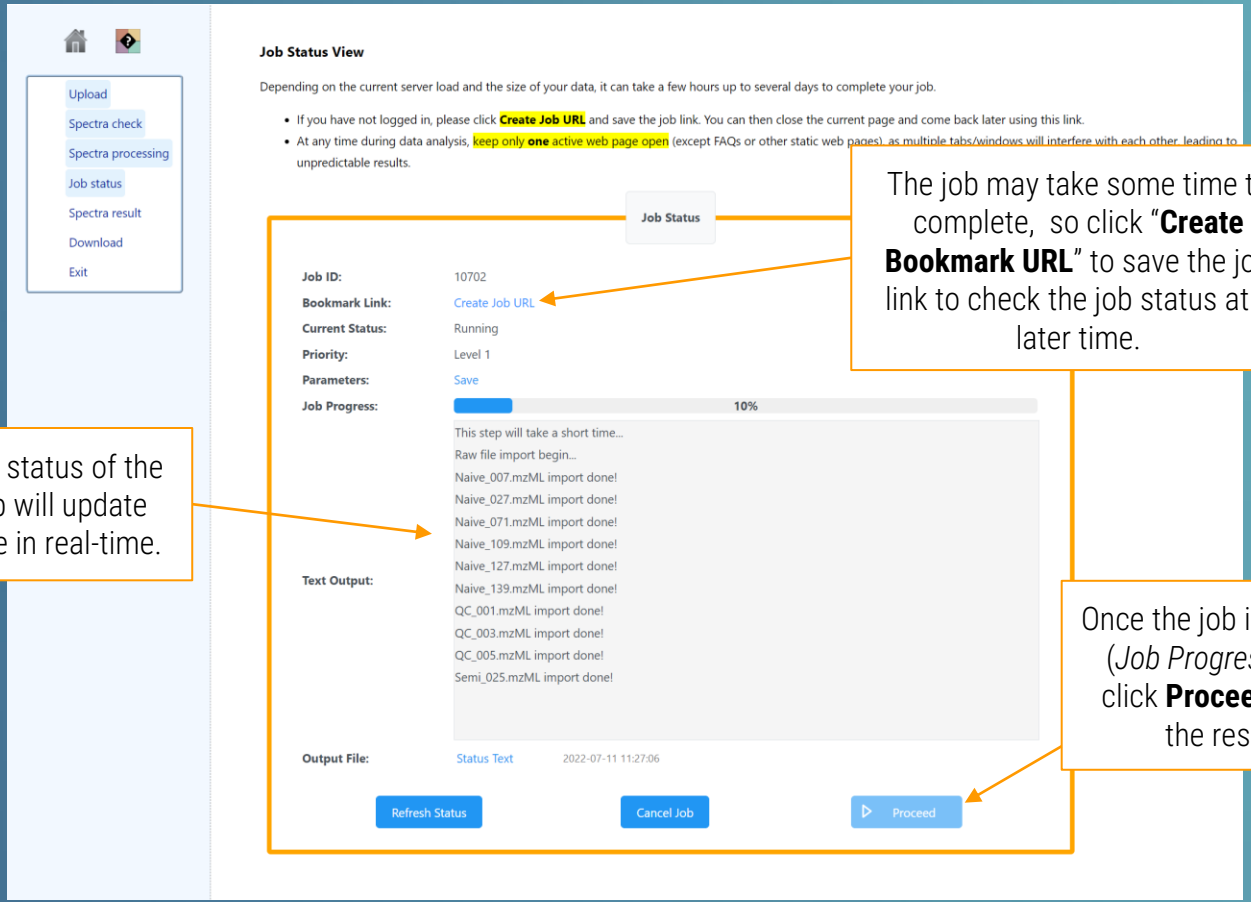
2. Click **Submit Job** to perform the spectra processing.

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



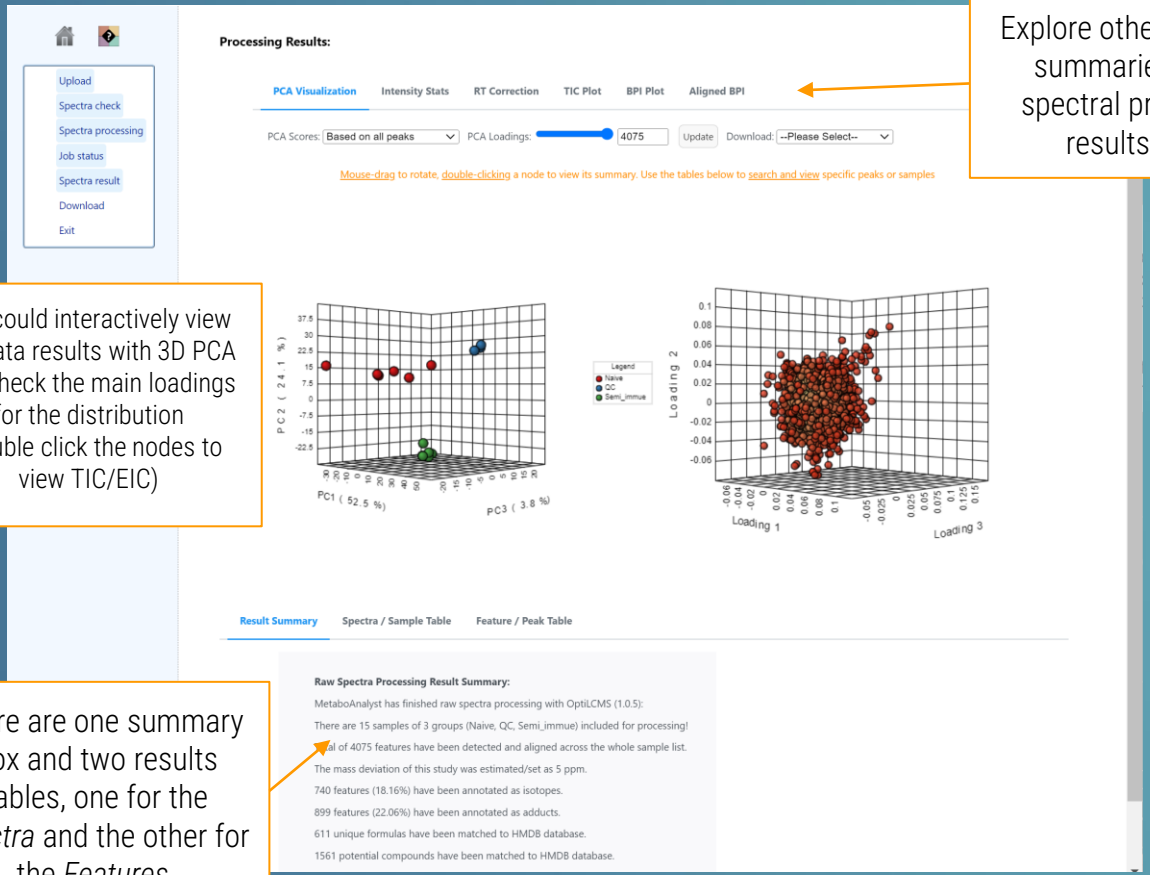
The job may take some time to complete, so click "**Create Bookmark URL**" to save the job link to check the job status at a later time.

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.

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

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

1.7 Exploring the Results -1



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*.

Inroduce more about these Result tables from the next pages.

1.8 Exploring the Results -2

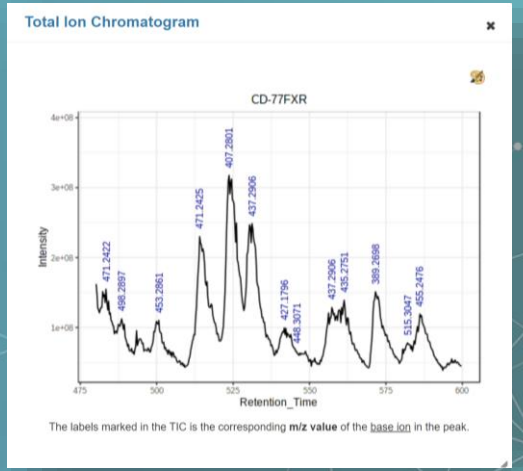
Result Summary | Spectra / Sample Table | Feature / Peak Table

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	

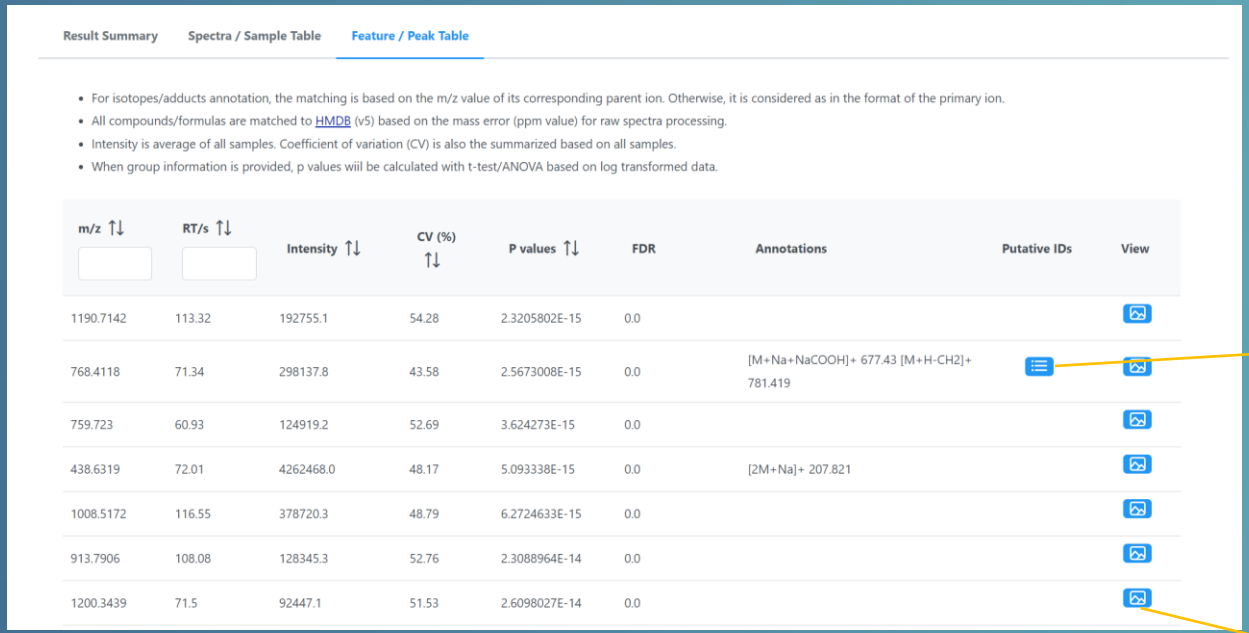
« ‹ 1 › » 20

Download Page

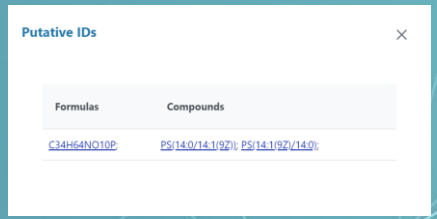
Click the ***View*** button to seeTIC of the corresponding spectra.



1.9 Exploring the Results -3



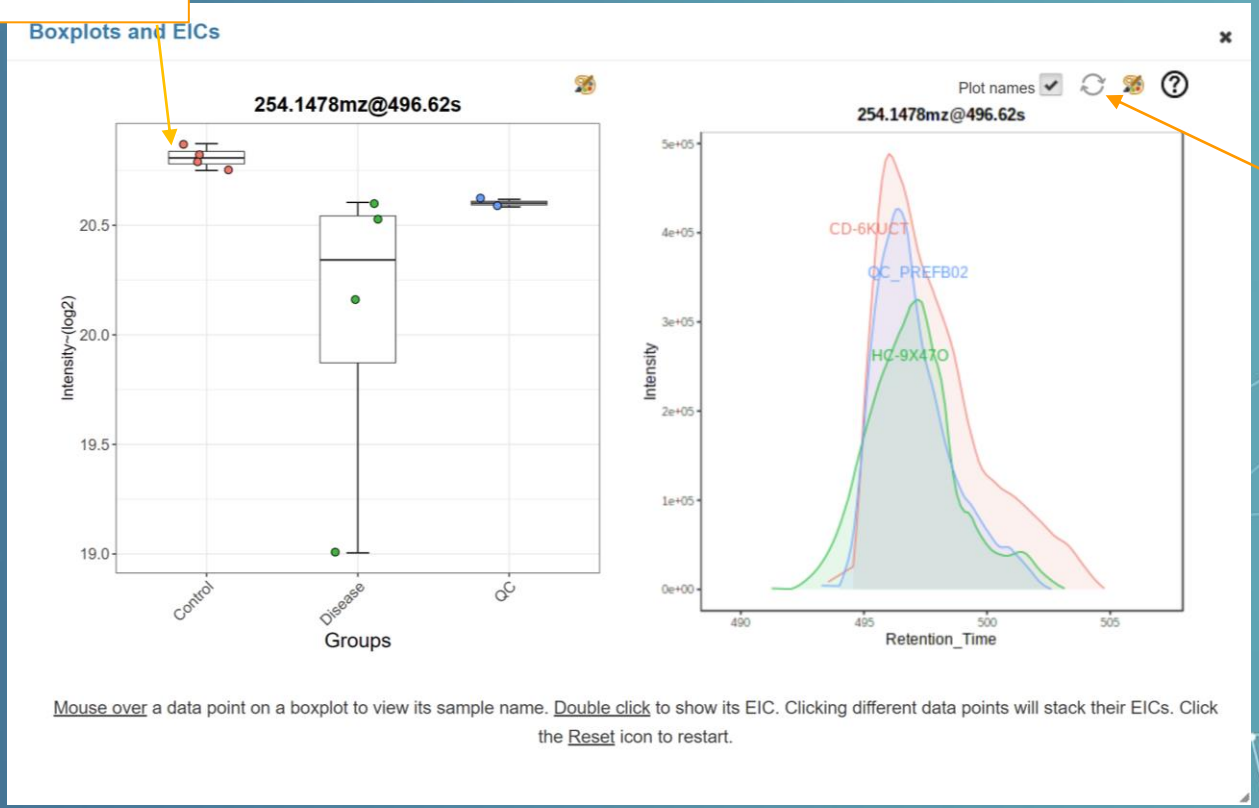
This table is showing all MS features. Click the button of Putative IDs show the potential Chemical IDs of the features towards HMDB.



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.

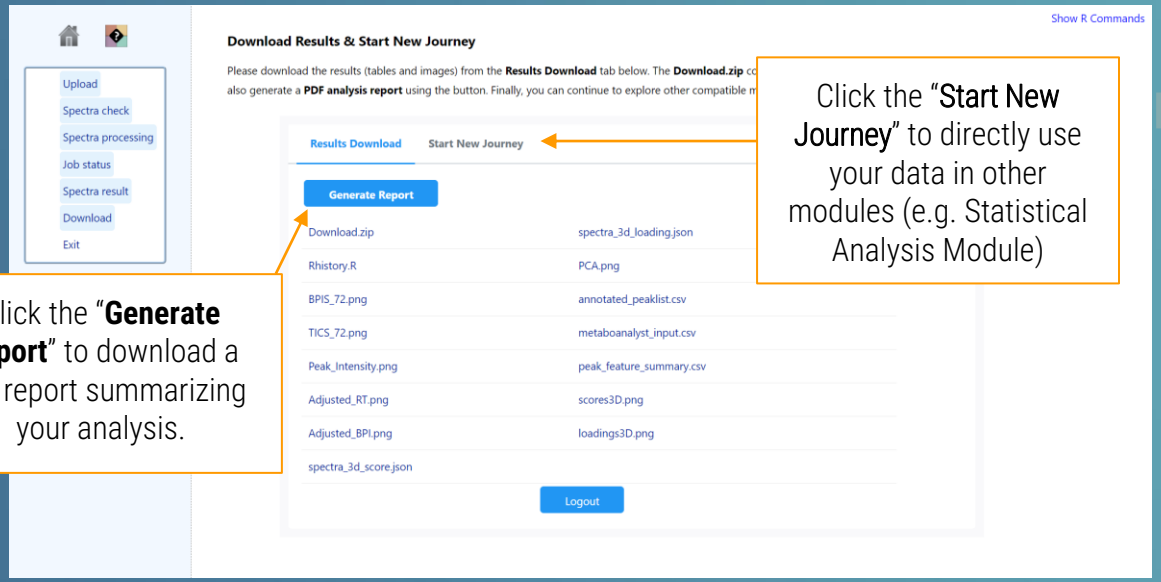


TIP1: If the plotting failed, please clean the cache of your browser or use another browser.

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

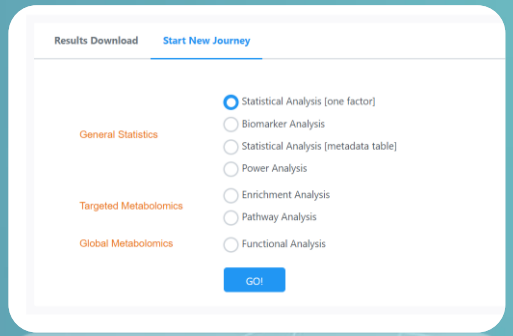
3. Scroll down to the bottom of page and click "**Proceed**" to view the Downloads page.

1.11 Result Downloading & New Journey

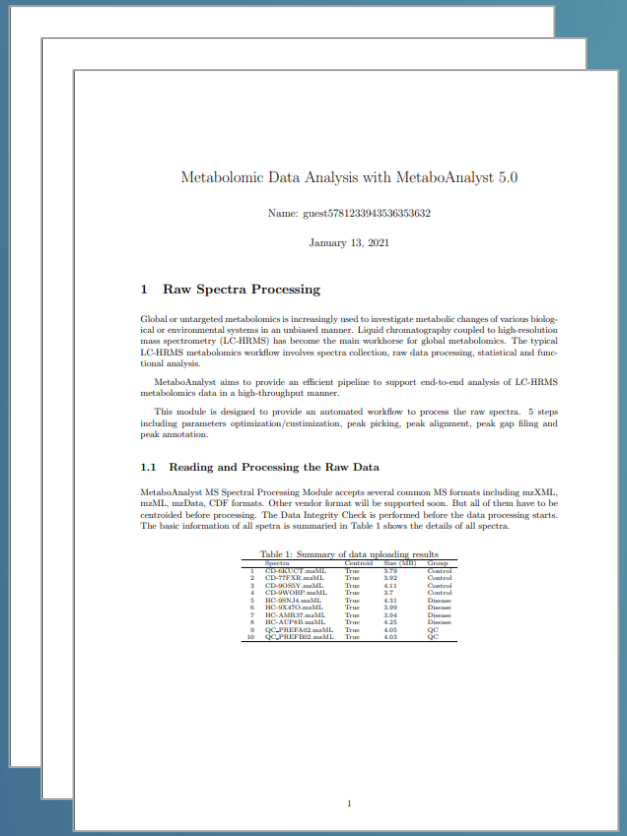


Click the "**Generate Report**" to download a pdf report summarizing your analysis.

Click the "**Start New Journey**" to directly use your data in other modules (e.g. Statistical Analysis Module)



1.12 PDF REPORT



Metabolomic Data Analysis with MetaboAnalyst 5.0

Name: guest57812339435363536 32

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/custimization, peak picking, peak alignment, peak gap filing and peak annotation.

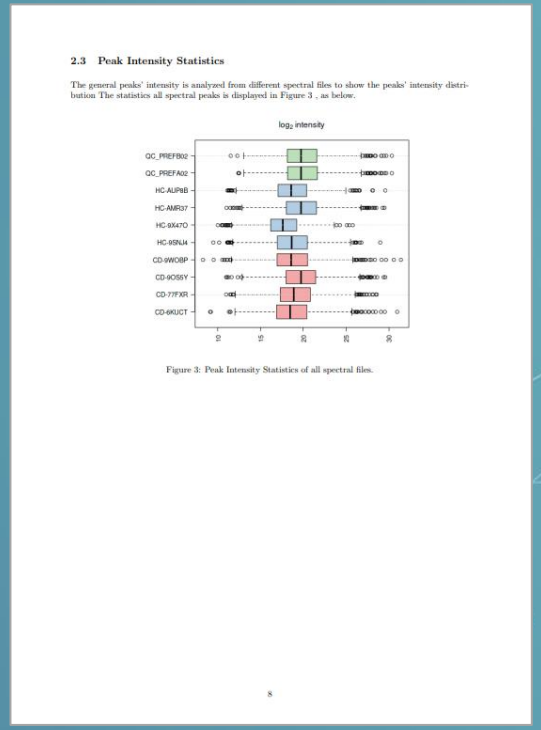
1.1 Reading and Processing the Raw Data

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 spetra is summaried in Table 1 shows the details of all spectra.

Table 1: Summary of data uploading results

	Spectra	Centroid	Size (MB)	Group
1	CD-6KUCT.mzML	True	3.79	Control
2	CD-77FXR.mzML	True	3.92	Control
3	CD-9OS5Y.mzML	True	4.11	Control
4	CD-9WOBP.mzML	True	3.7	Control
5	HC-9SNJ4.mzML	True	4.31	Disease
6	HC-9X47O.mzML	True	3.99	Disease
7	HC-AMR37.mzML	True	3.94	Disease
8	HC-AUP8B.mzML	True	4.25	Disease
9	QC_PREFA02.mzML	True	4.05	QC
10	QC_PREFB02.mzML	True	4.03	QC

1



2.3 Peak Intensity Statistics

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

Figure 3: Peak Intensity Statistics of all spectral files.

8

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 or Jeff.xia[at]xialab.ca