MetaboAnalyst 6.0

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

Spectra Processing [LC-MS w/wo MS2]

2024-03-10

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

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:

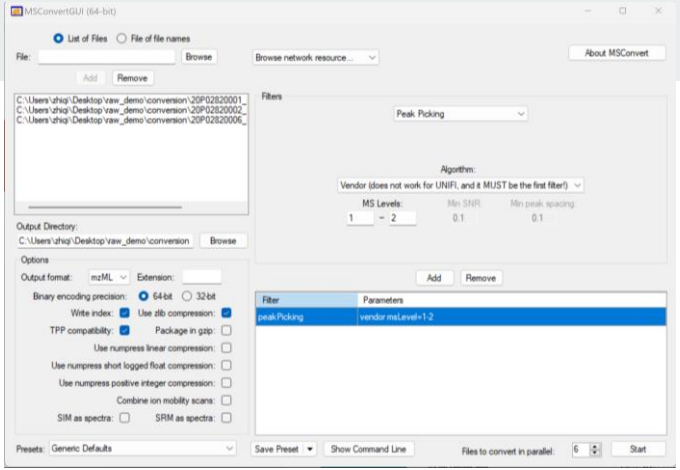
i. mzML (recommended);
ii. mzXML;
iii. cdf/CDF;
iv. mzData (phasing-out).

Expected Results

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

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

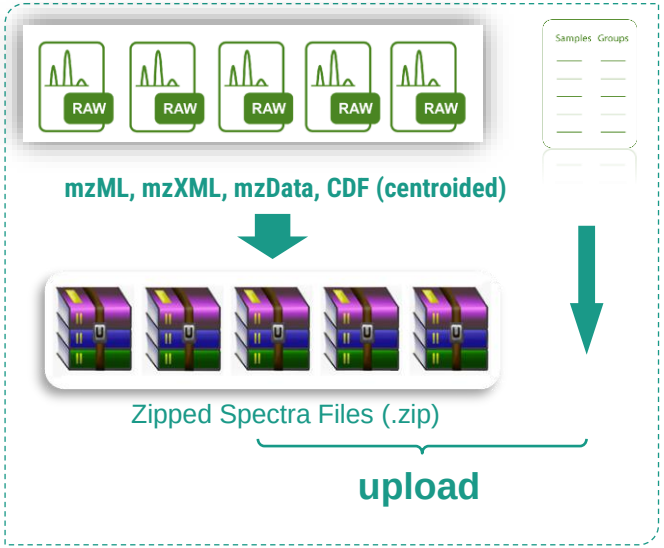
2. Spectra data preparation



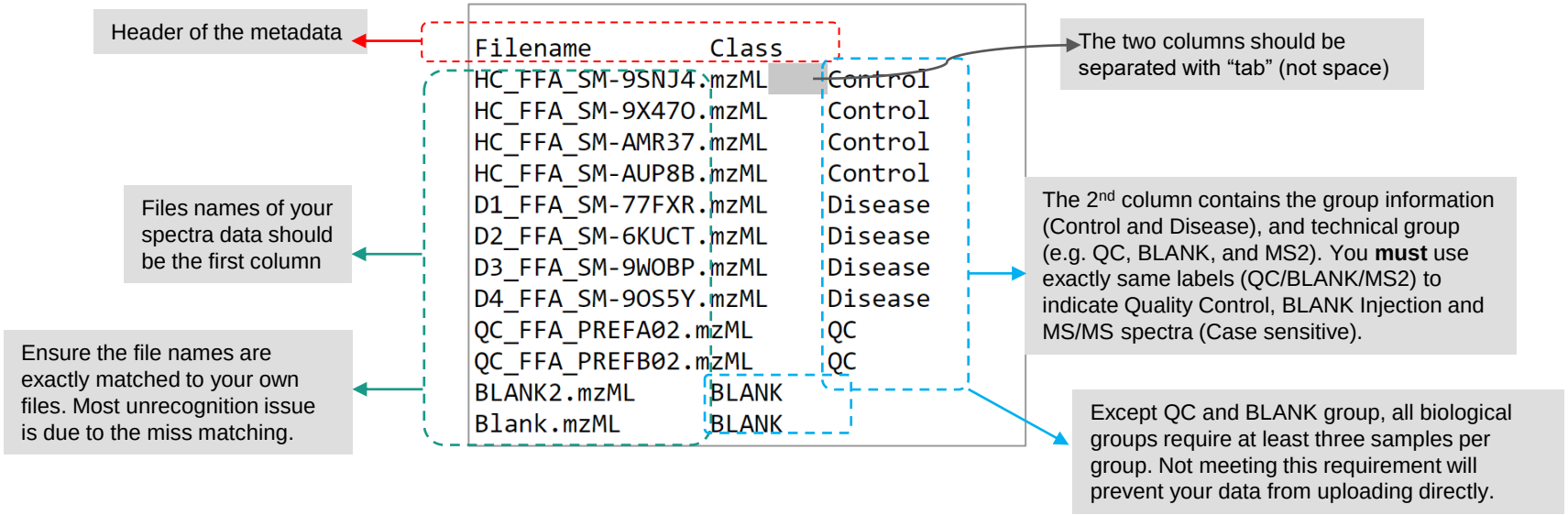
1. We highly recommend users to centroid MS files with **ProteoWizard**:

- **For GUI:** **Must** add 'Peak Picking' as the **1st 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.

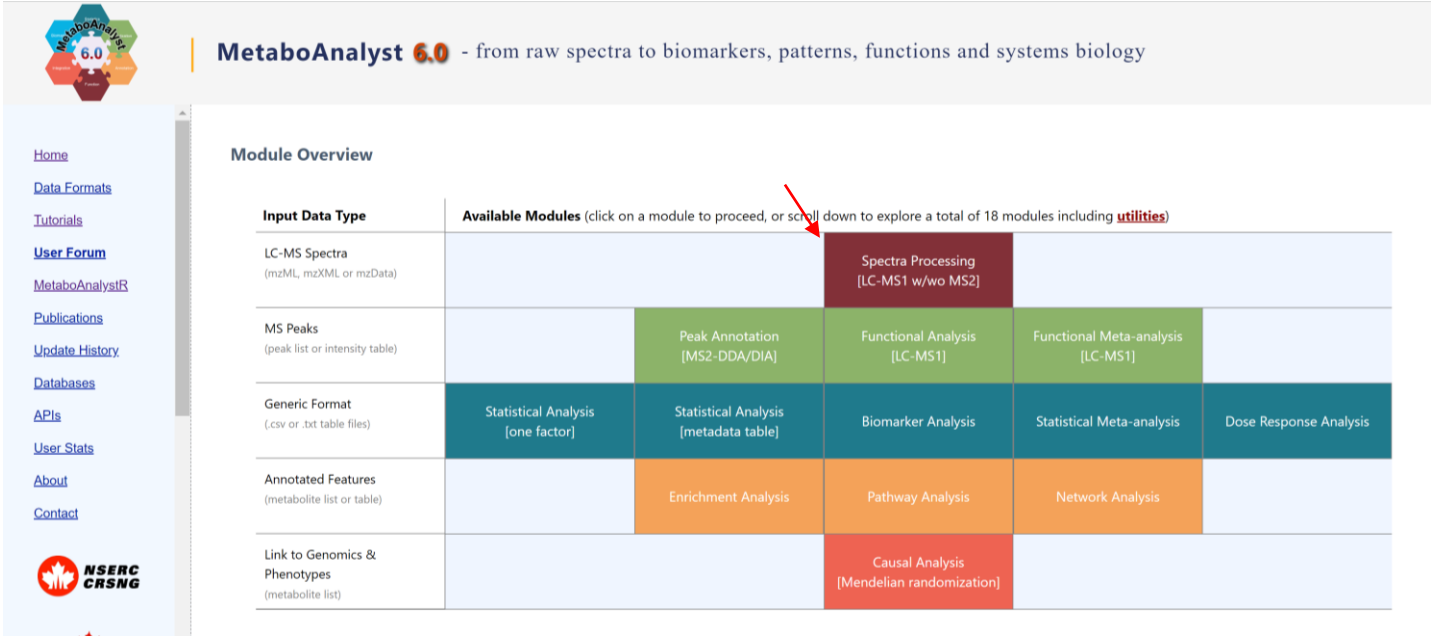


2. Metadata preparation



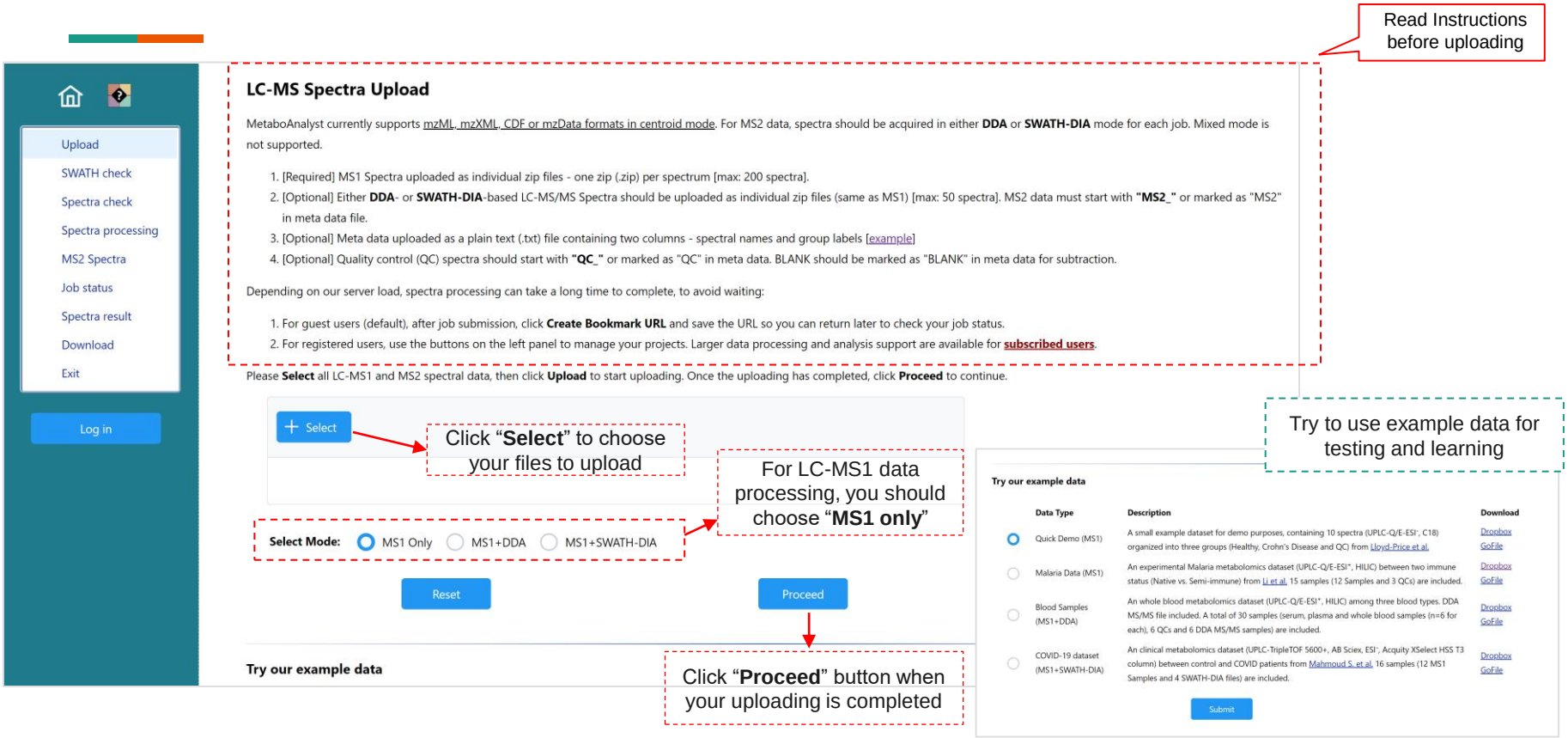
3. Choose the Module

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

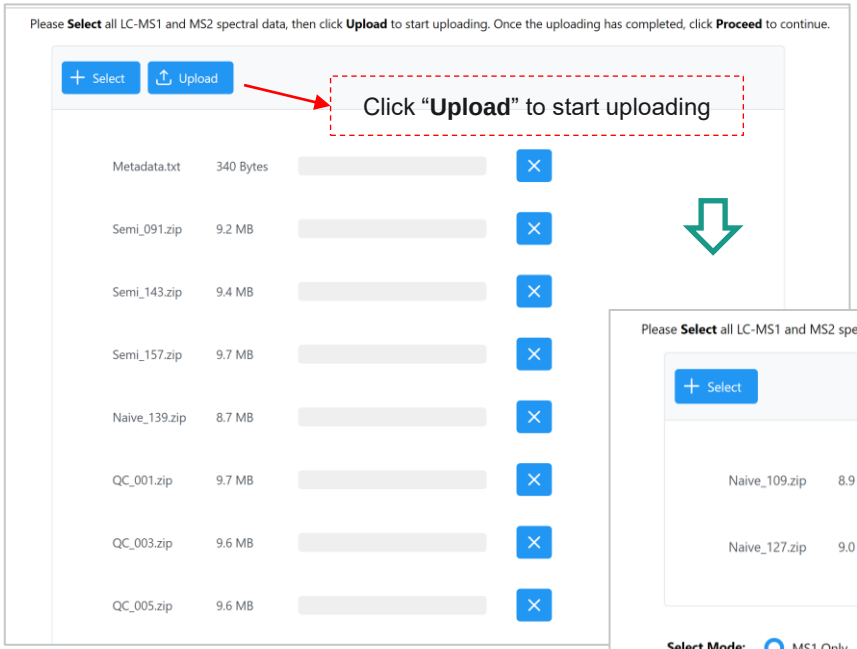


4. LC-MS1 Spectra Processing

4.1 LC-MS1 spectra upload

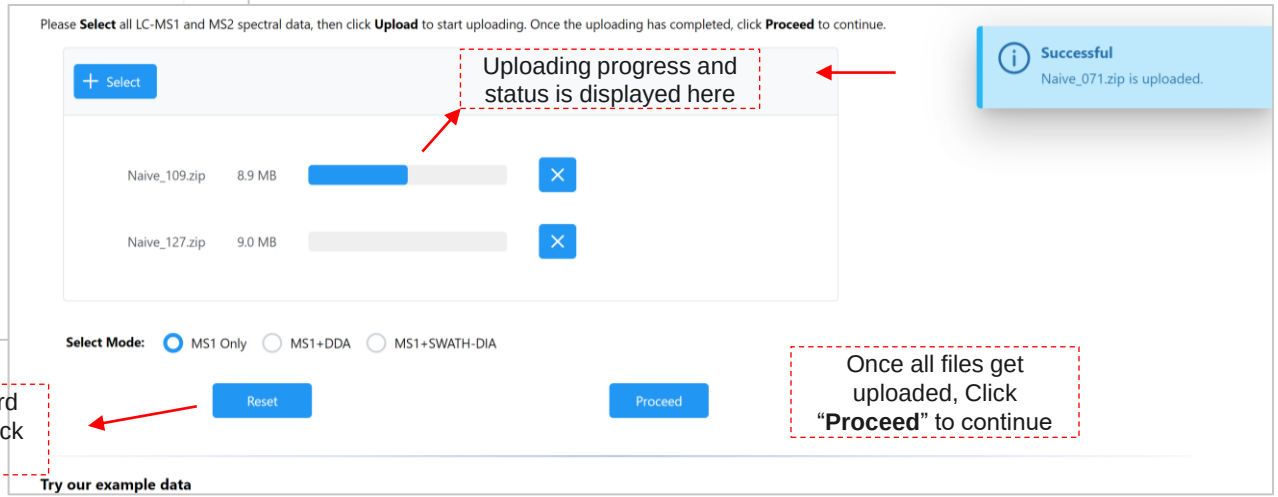


4.1 LC-MS1 spectra upload



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

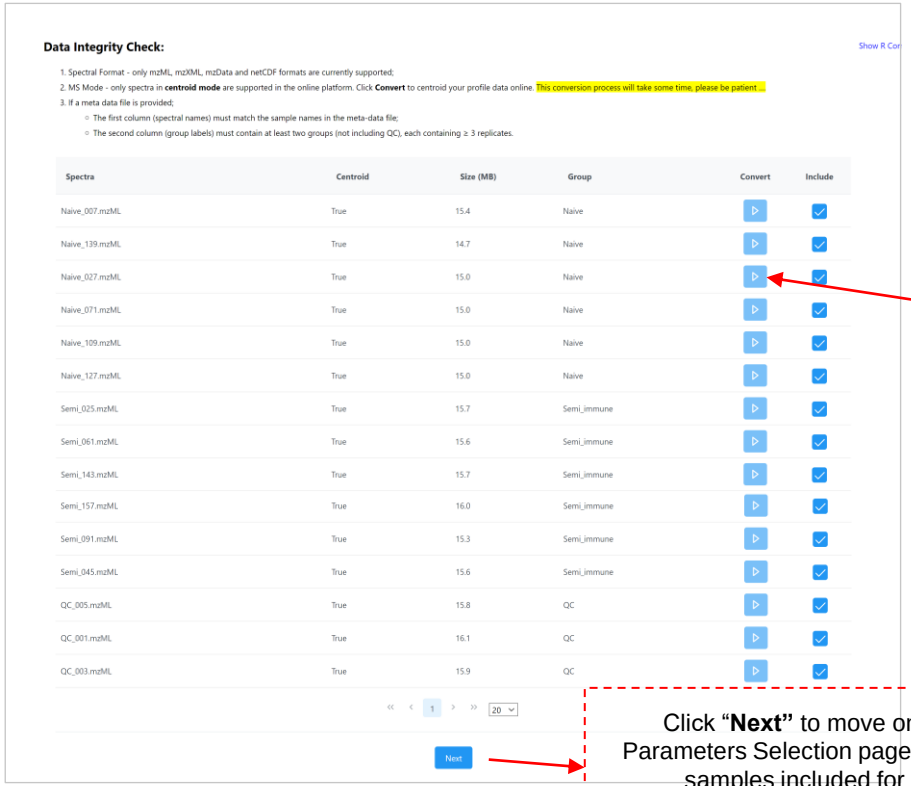
Data uploading may take long time due to the connection speed and high volumes of users. Please be patient or avoid the peak time (EST 8:00AM-11:00AM, Mon-Fri).



4.2 Integrity Check

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

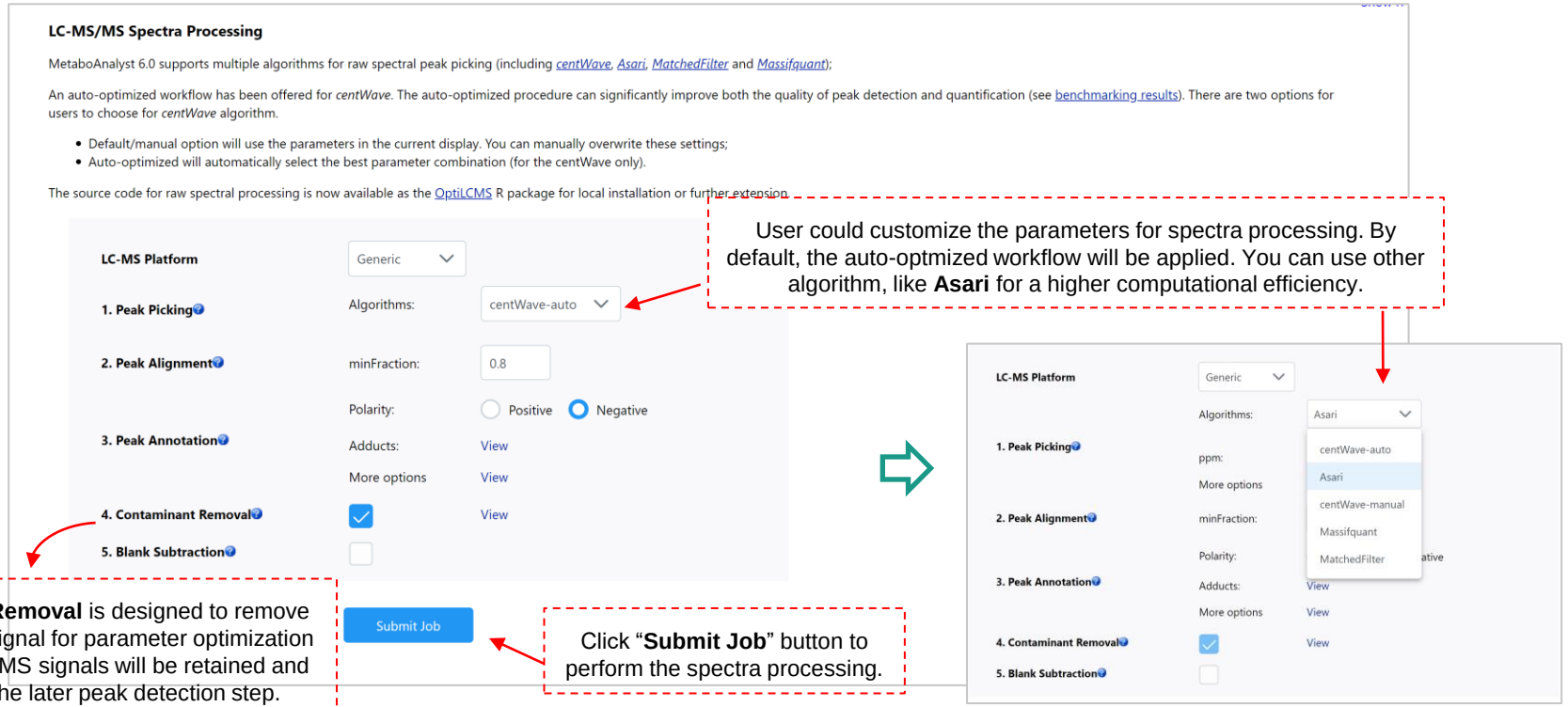
Results of the **Data Integrity Check** are shown here.



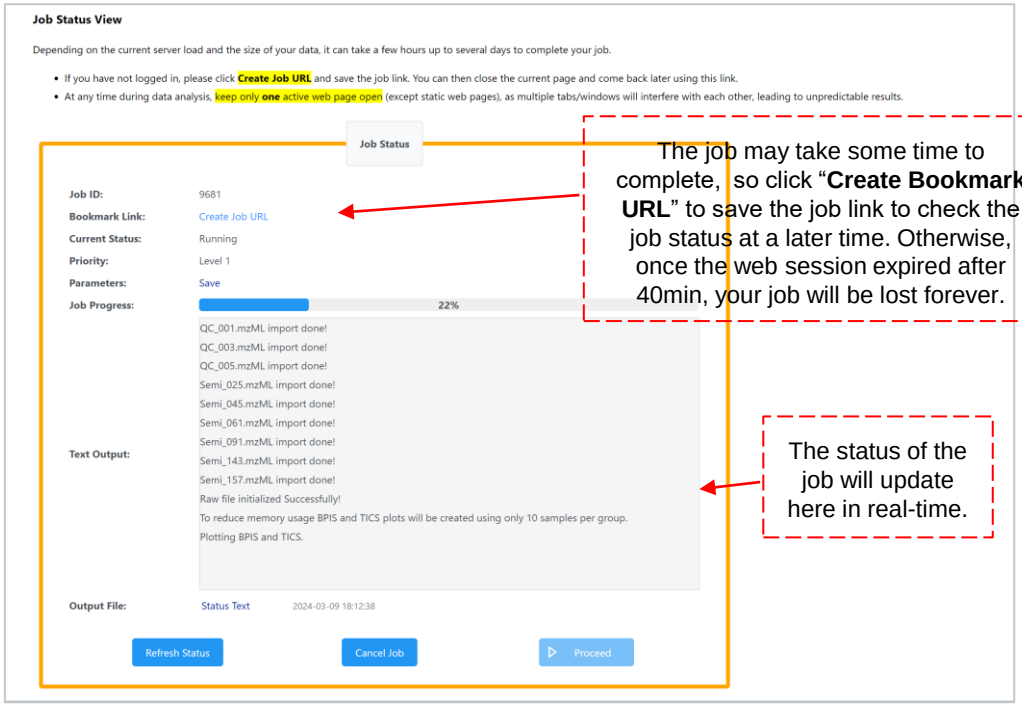
If your data is not in centroid mode, click **Convert** wrench button to convert it online.

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

4.3 Parameter Setting

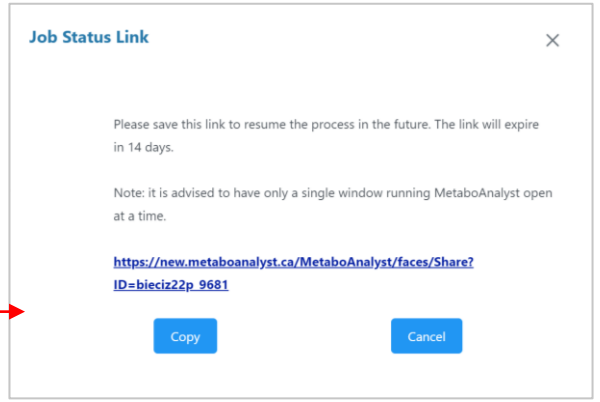


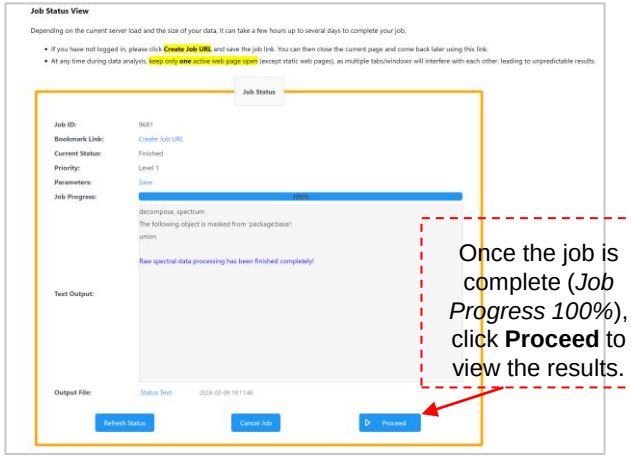
4.4 Jos Status



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.

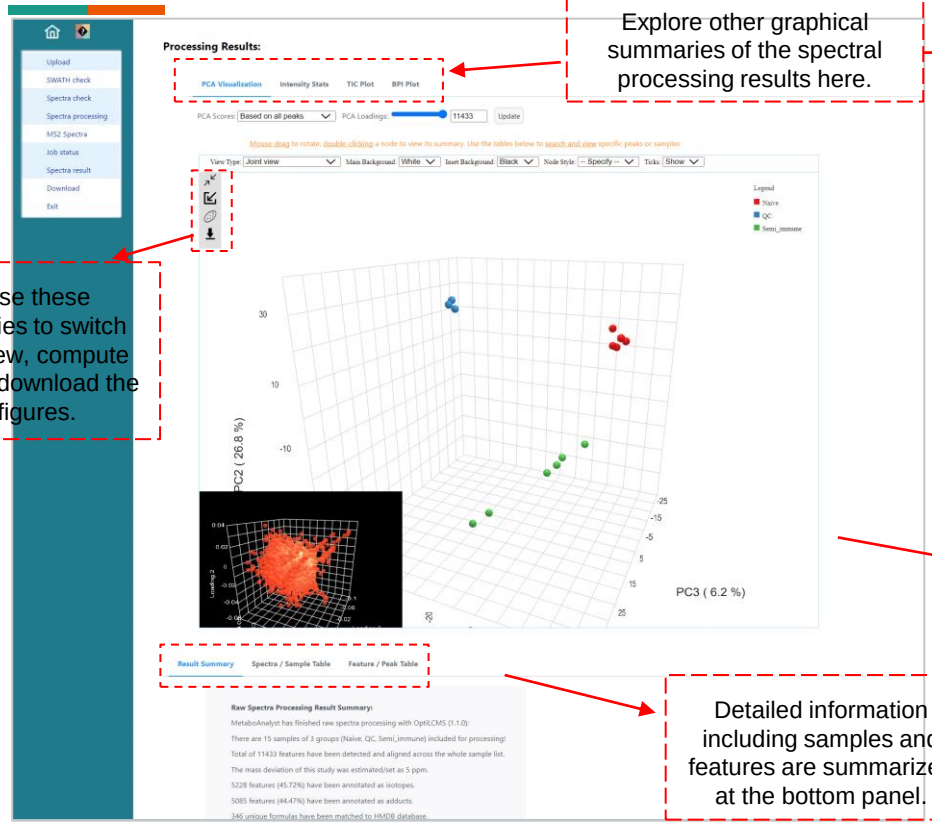
The status of the job will update here in real-time.



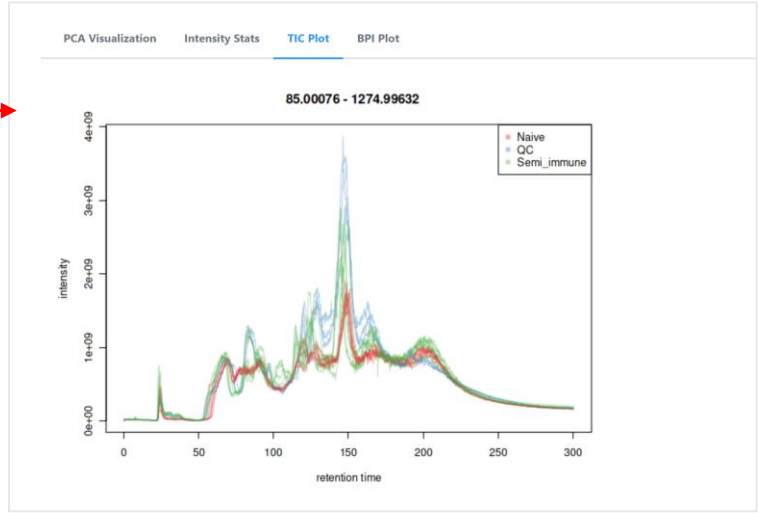


Once the job is complete (*Job Progress 100%*), click **Proceed** to view the results.

4.5 Result Exploration



Explore other graphical summaries of the spectral processing results here.



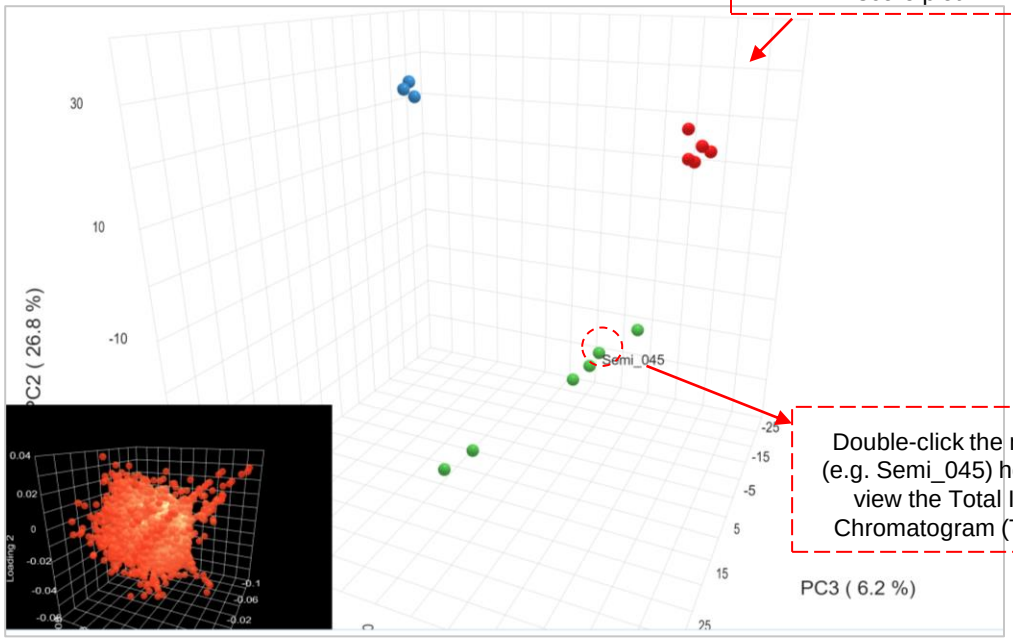
Try to use these functionalities to switch the PCA view, compute ellipse, and download the result figures.

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

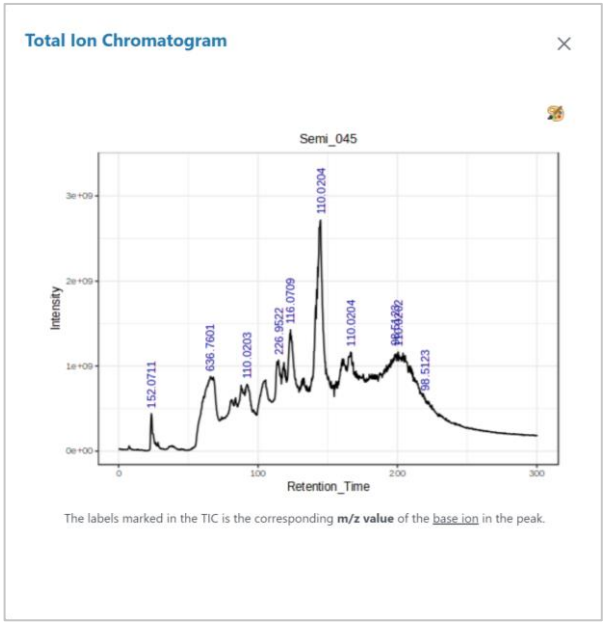
Detailed information including samples and features are summarized at the bottom panel.

4.5 Result Exploration

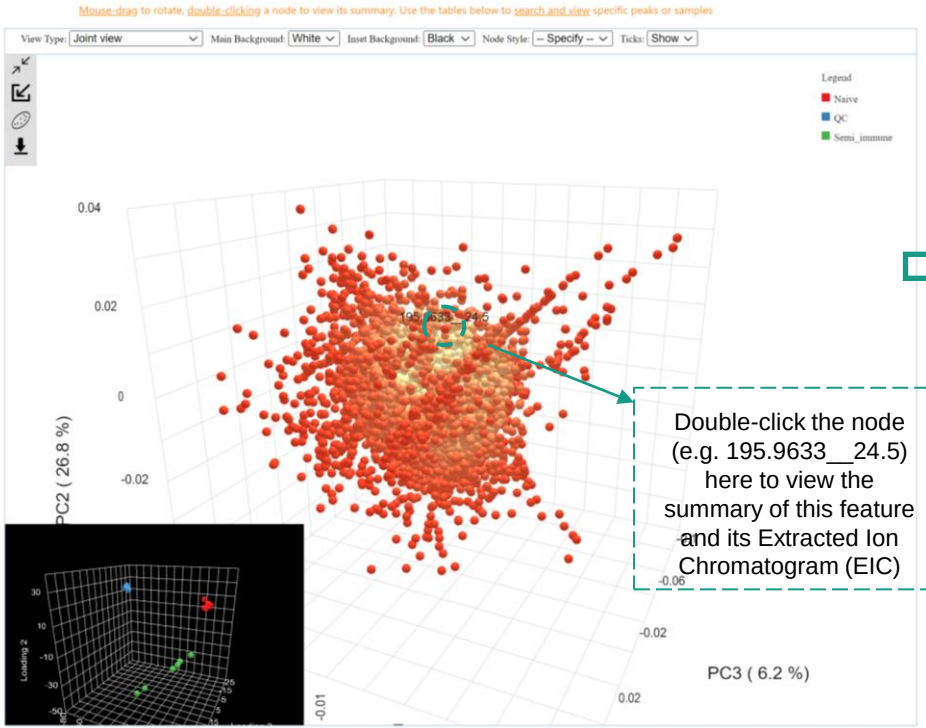
User could use their mouse to scroll and zoom in/out this score plot



Double-click the node (e.g. Semi_045) here to view the Total Ion Chromatogram (TIC).

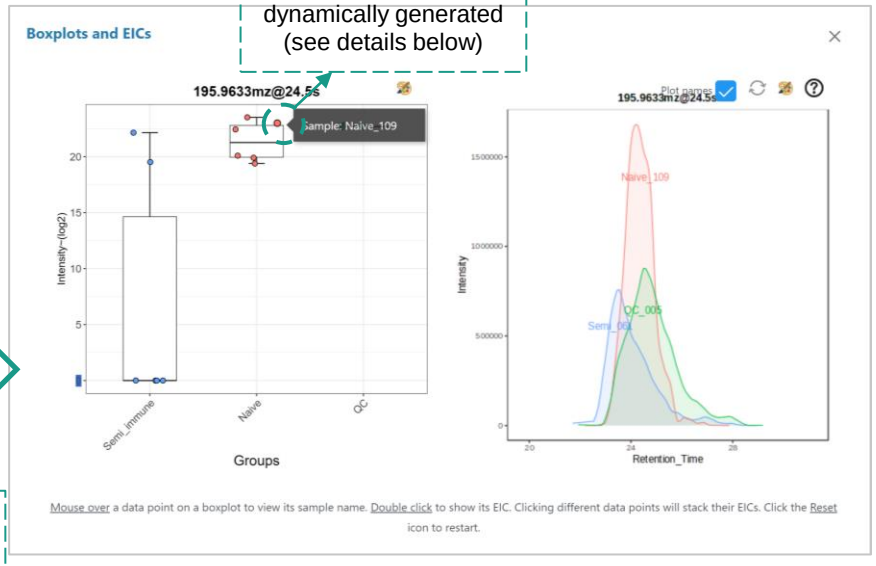


4.5 Result Exploration



Double-click the node (e.g. 195.9633__24.5) here to view the summary of this feature and its Extracted Ion Chromatogram (EIC)

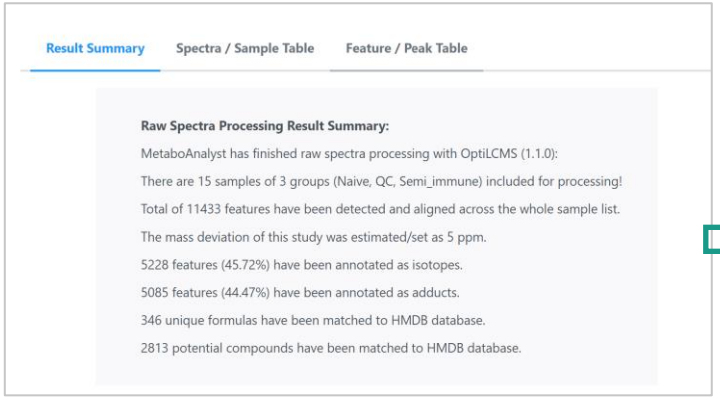
This EIC plot can be dynamically generated (see details below)

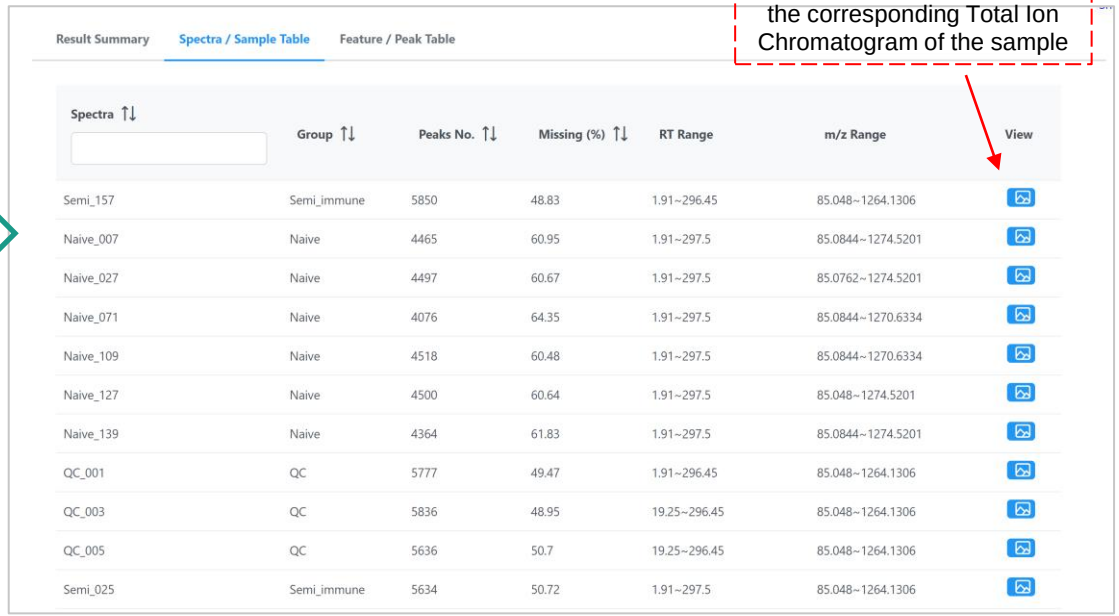


How to use the dynamic EIC generation,

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.

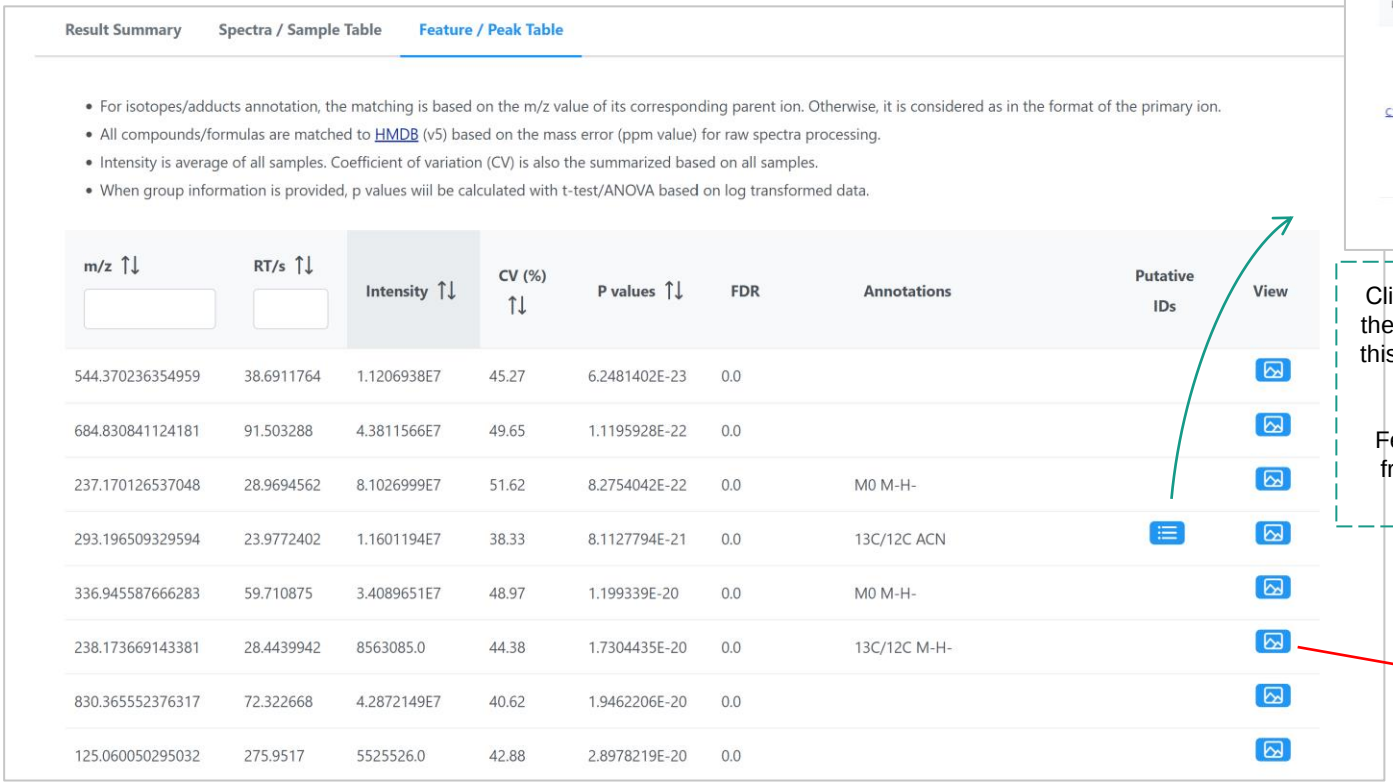
4.5 Result Exploration





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).

4.5 Result Exploration



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

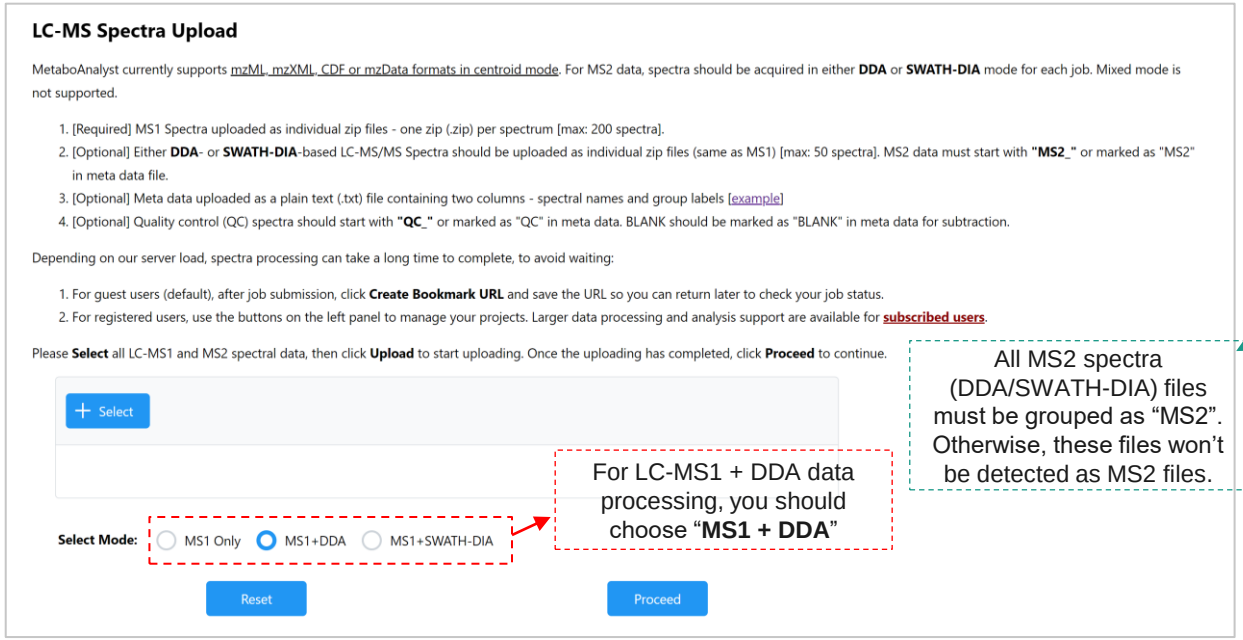
Features without the putative IDs from HMDB database would not show the Icon.

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.

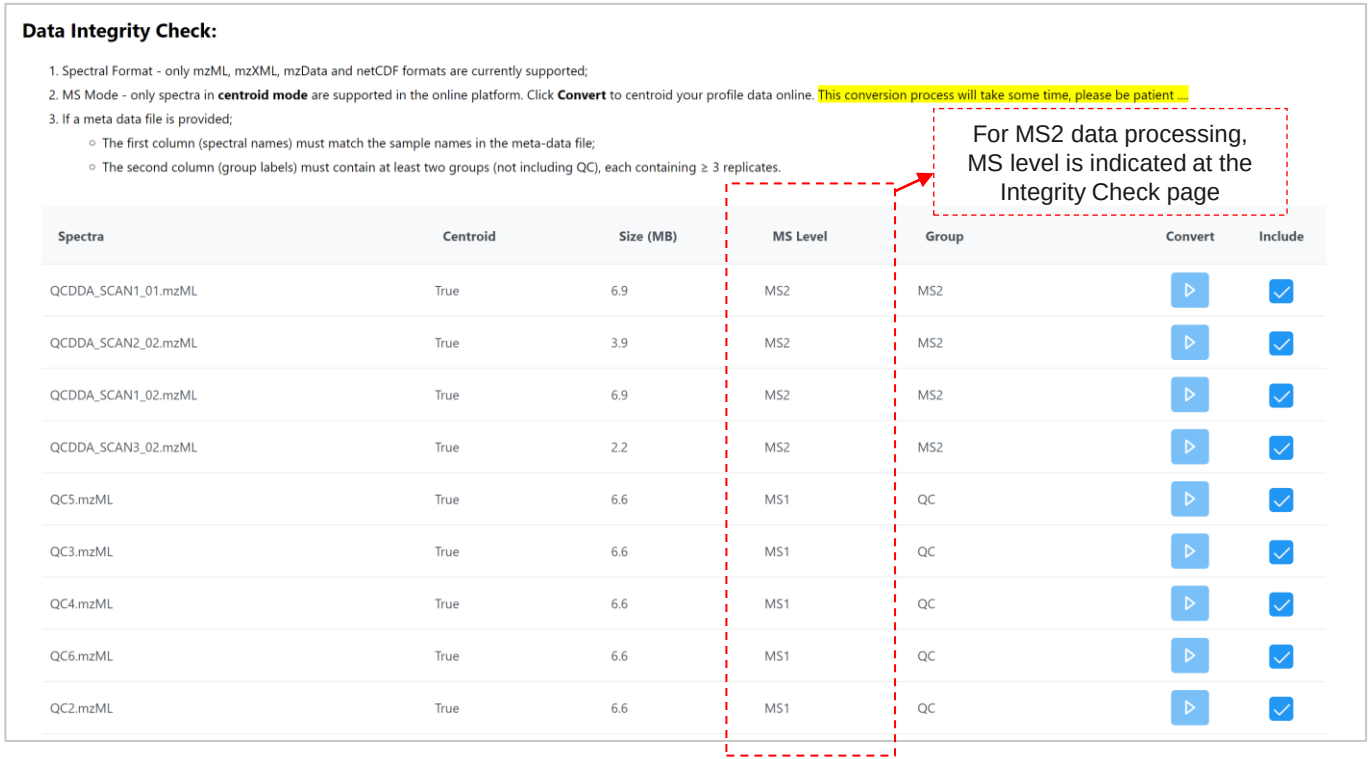
5.1 Spectra Files Upload



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

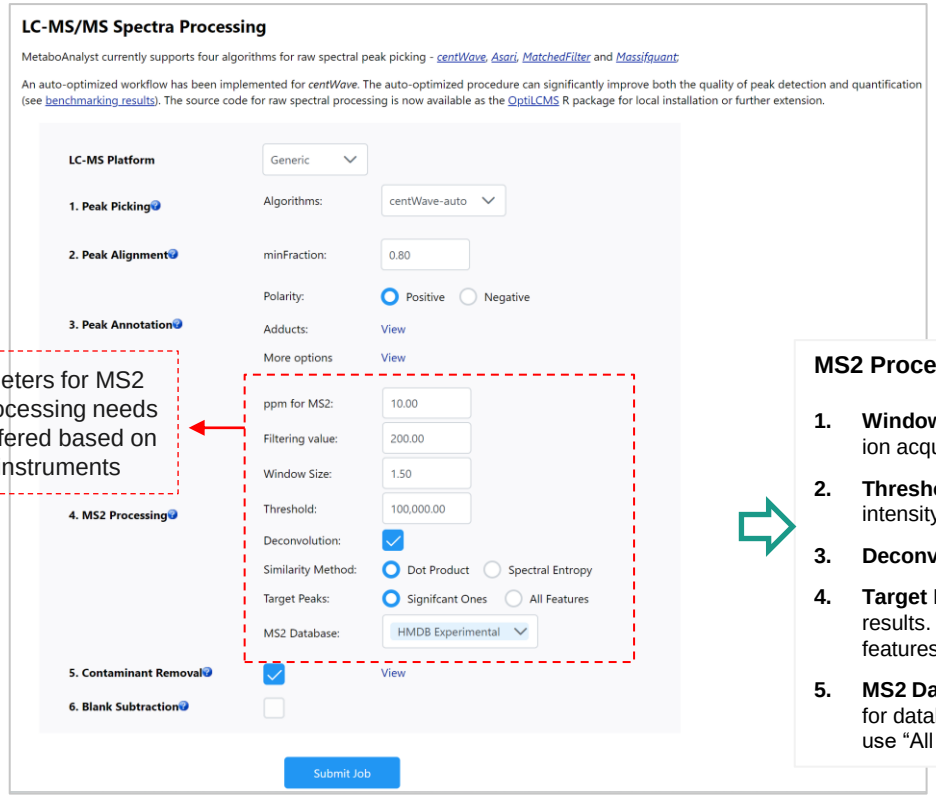
5.2 Integrity Check



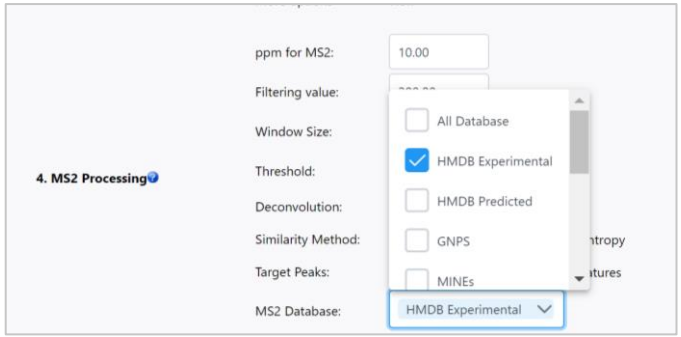
MS Levels,

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



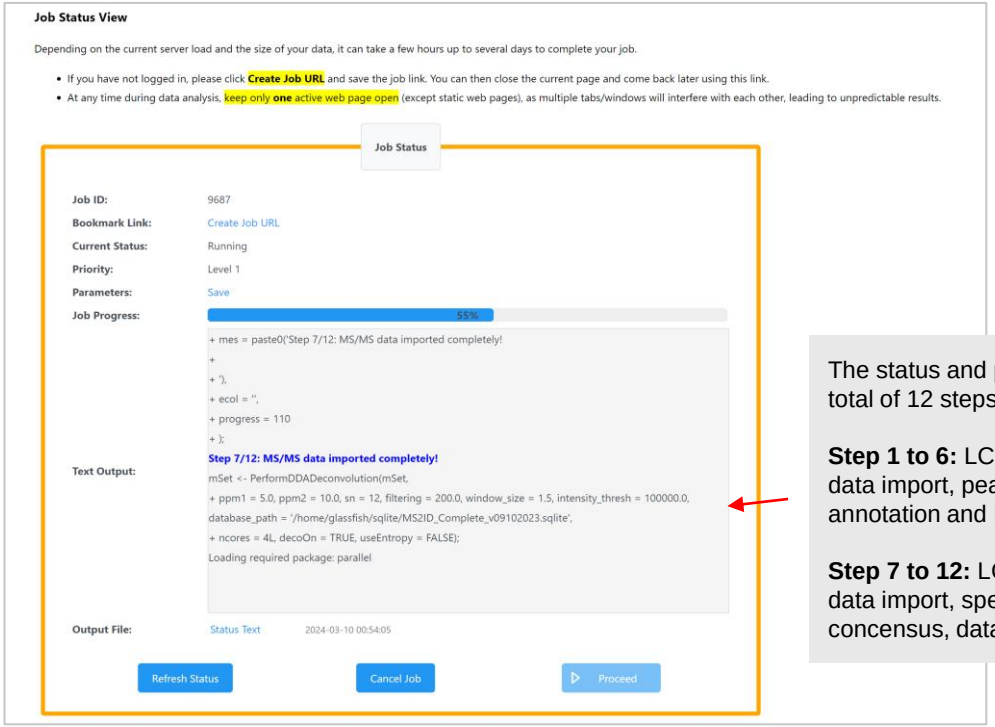
Parameters for MS2 data processing needs to be offered based on MS instruments



MS2 Processing Parameters,

1. **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;
2. **Threshold:** MS/MS acquisition usually requires a threshold to exclude the low-intensity ions to ensure the quality of acquired spectrum;
3. **Deconvolution:** Deconvolution on DDA spectra can be optionally disabled;
4. **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.
5. **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



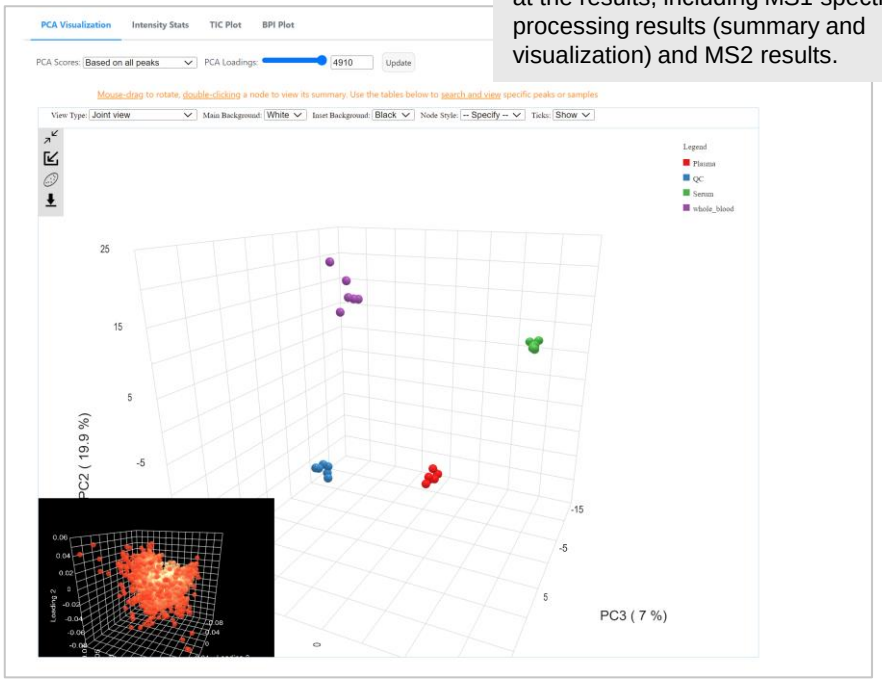
The status and processing logs are displayed here. A total of 12 steps will appear for MS1 + DDA:

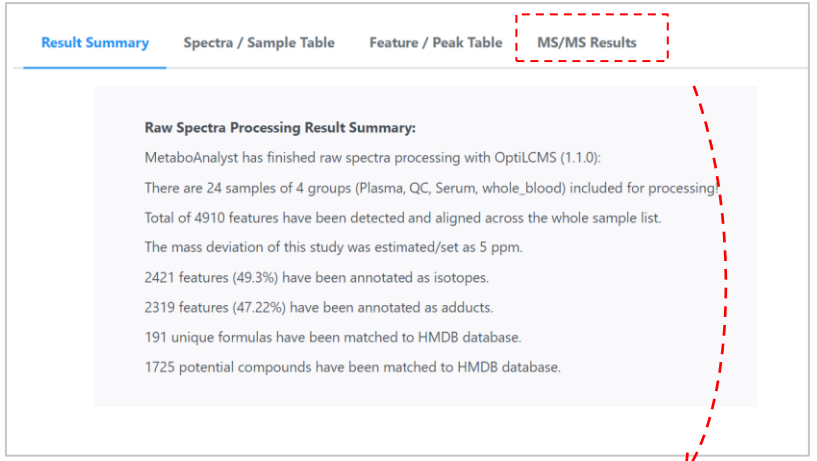
Step 1 to 6: LC-MS1 spectra processing, including data import, peak picking, alignment, gap filling, peak annotation and LC-MS1 results export;

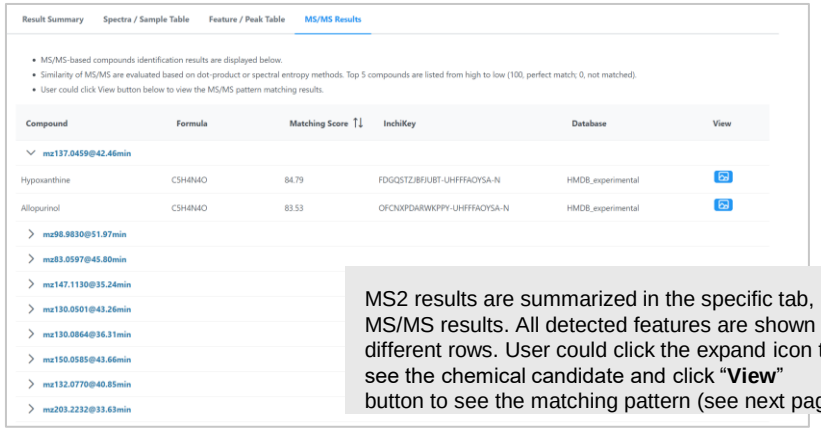
Step 7 to 12: LC-MS2 spectra processing, including data import, spectra deconvolution, spectra concensus, database searching and results export.

5.5 Result Exploration

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

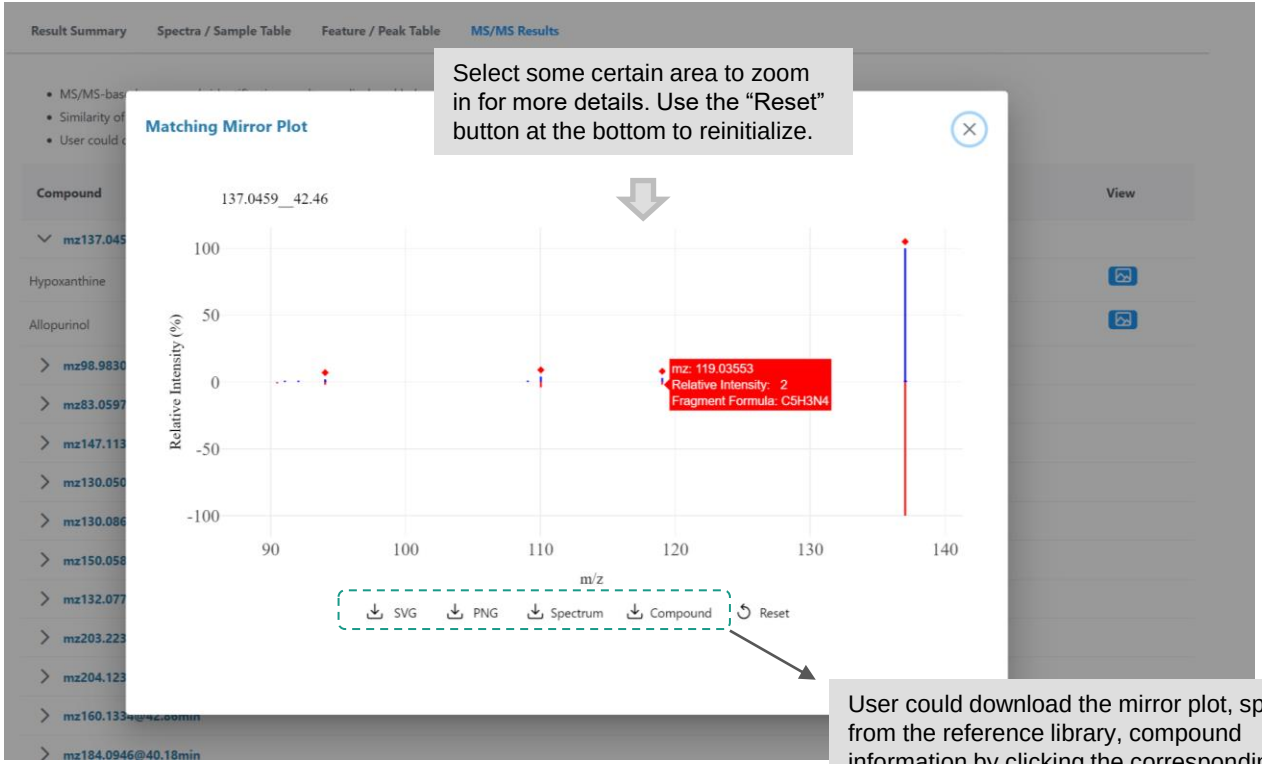






MS2 results are summarized in the specific tab, MS/MS results. All detected features are shown in different rows. User could click the expand icon to see the chemical candidate and click "**View**" button to see the matching pattern (see next page).

5.5 Result Exploration



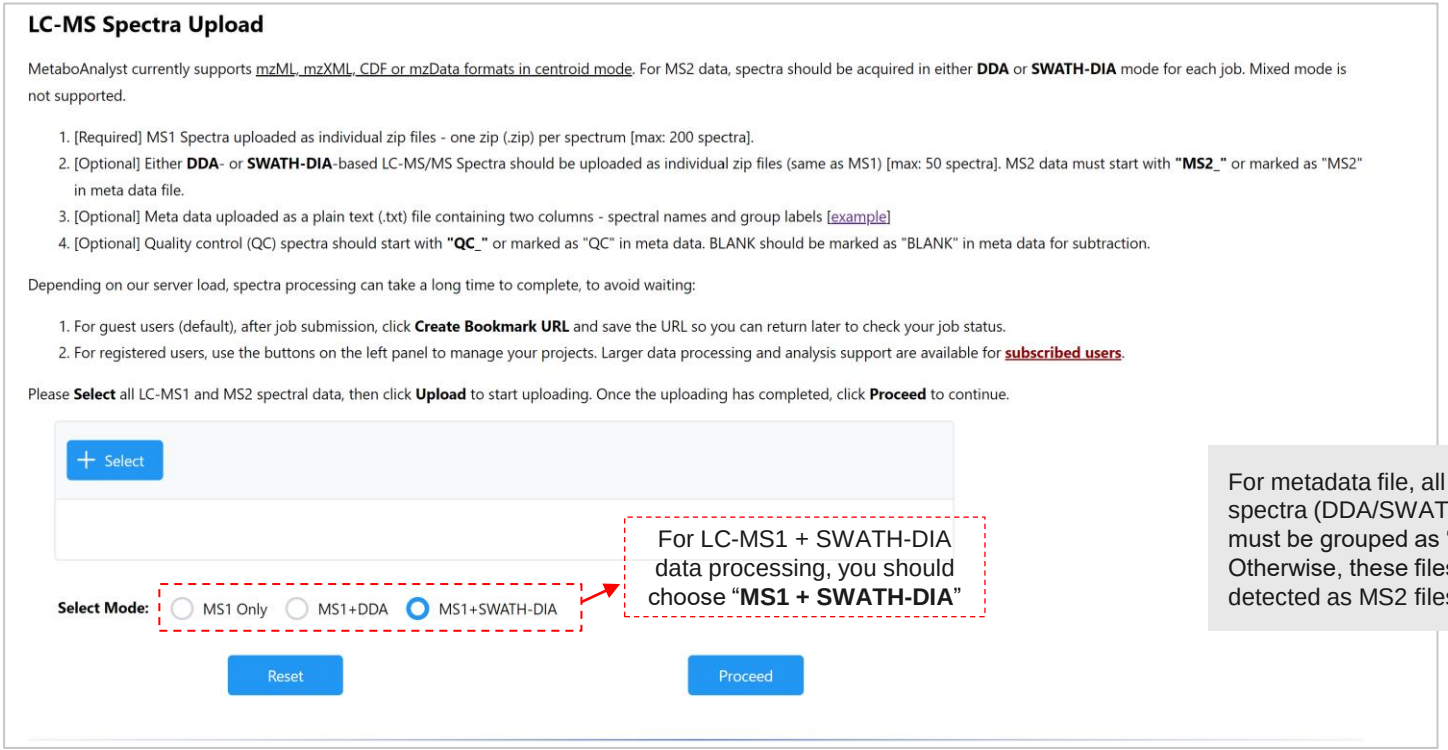
MS2 results visualization,

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

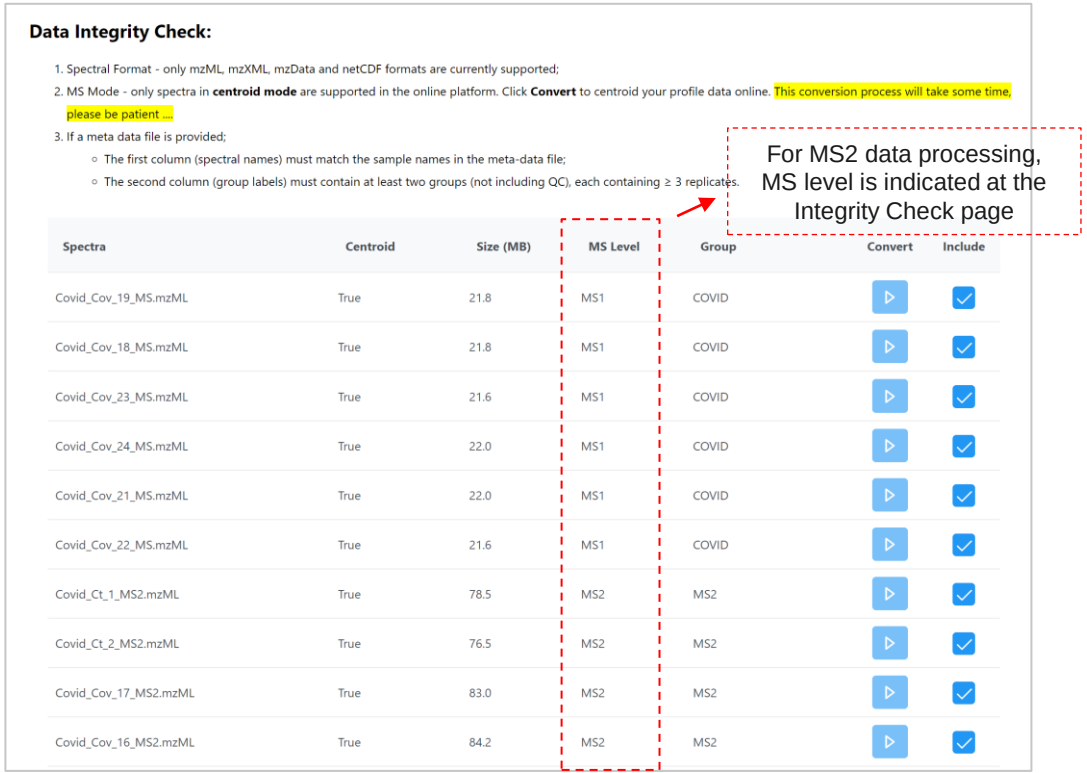
In this section, we only emphasize the functionalities specifically related to the SWATH-DIA data processing. All other MS1 associated features have been included with details in the previous section 4, therefore they won’t be repeated here. Please read the section 3 and 4 at first if you are not familiar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

6.1 Spectra Files Upload

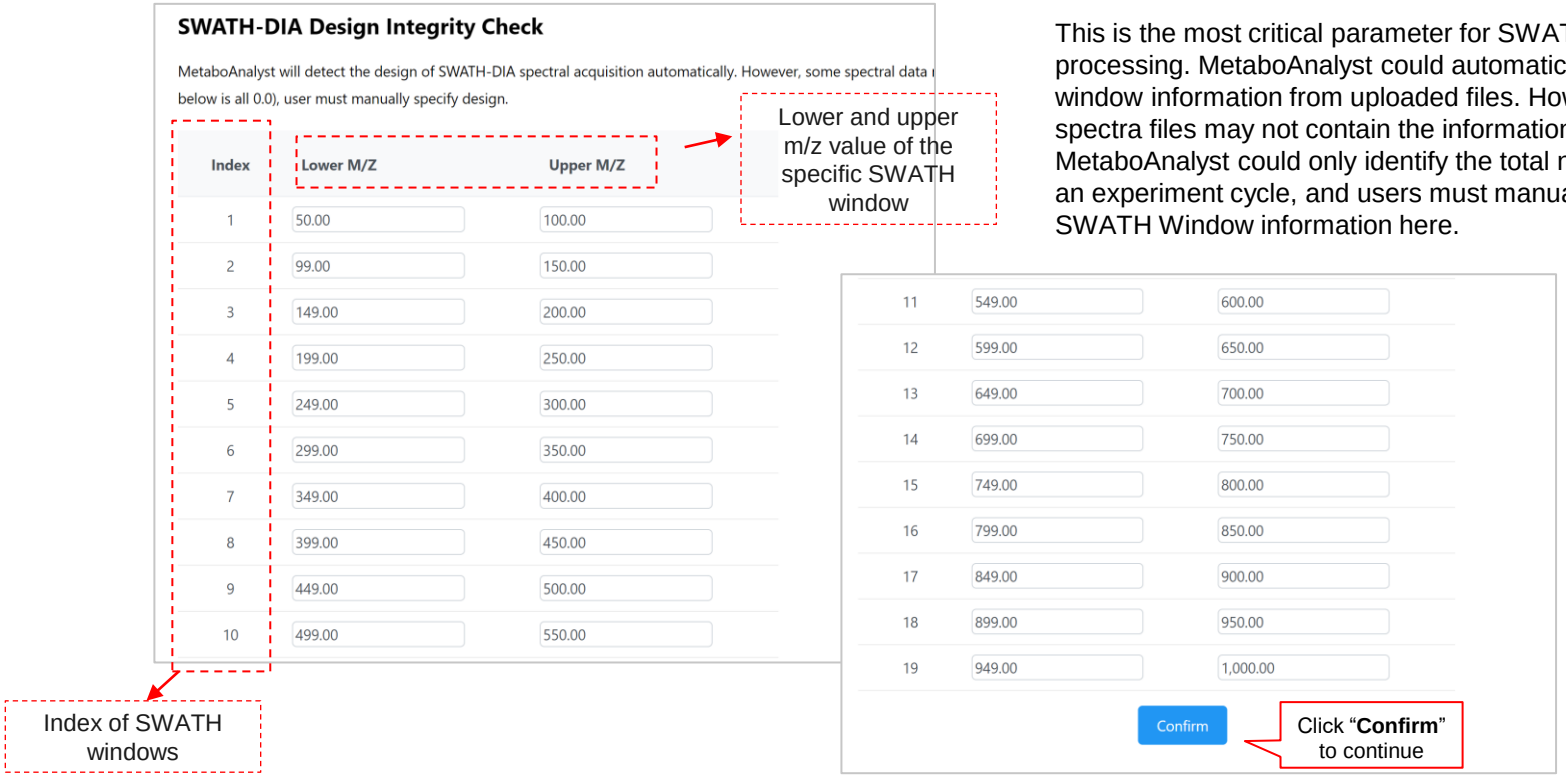


For metadata file, all MS2 spectra (DDA/SWATH-DIA) files must be grouped as “MS2”. Otherwise, these files won’t be detected as MS2 files.

6.2 Integrity Check



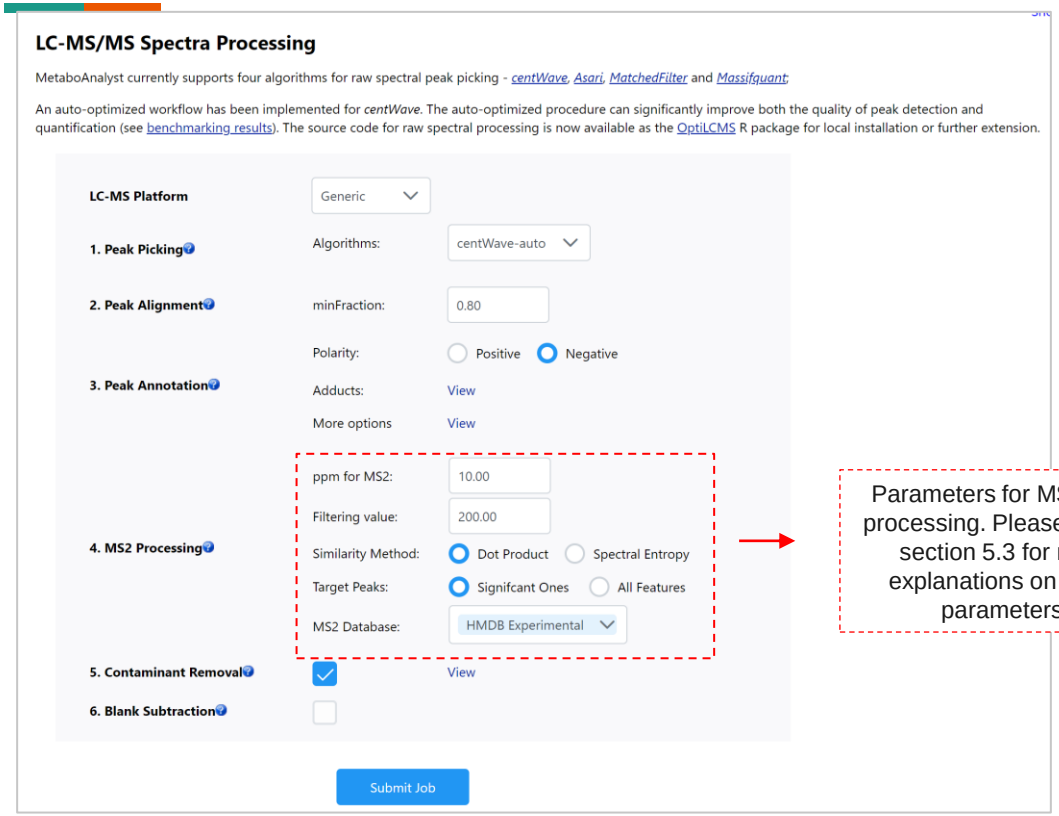
6.3 SWATH window check



SWATH-DIA Design window:

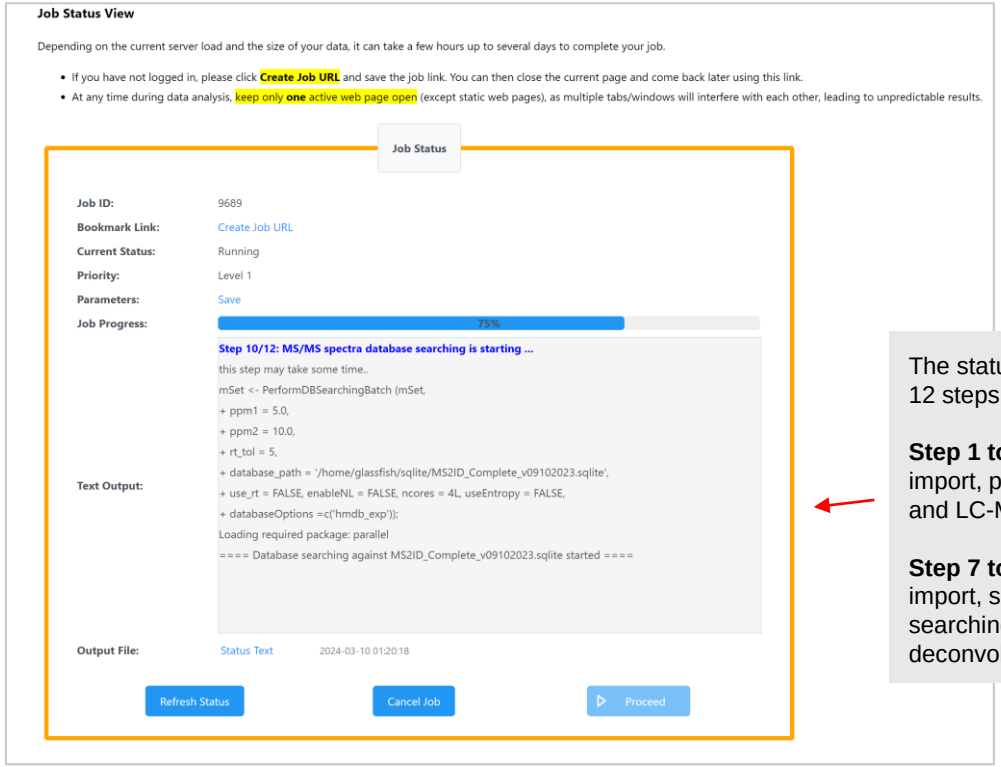
This is the most critical parameter for SWATH-DIA spectra processing. MetaboAnalyst could automatically detect the SWATH window information from uploaded files. However, some raw spectra files may not contain the information. In this case, MetaboAnalyst could only identify the total number of windows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

6.4 Parameter Setting



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

6.5 Jos Status



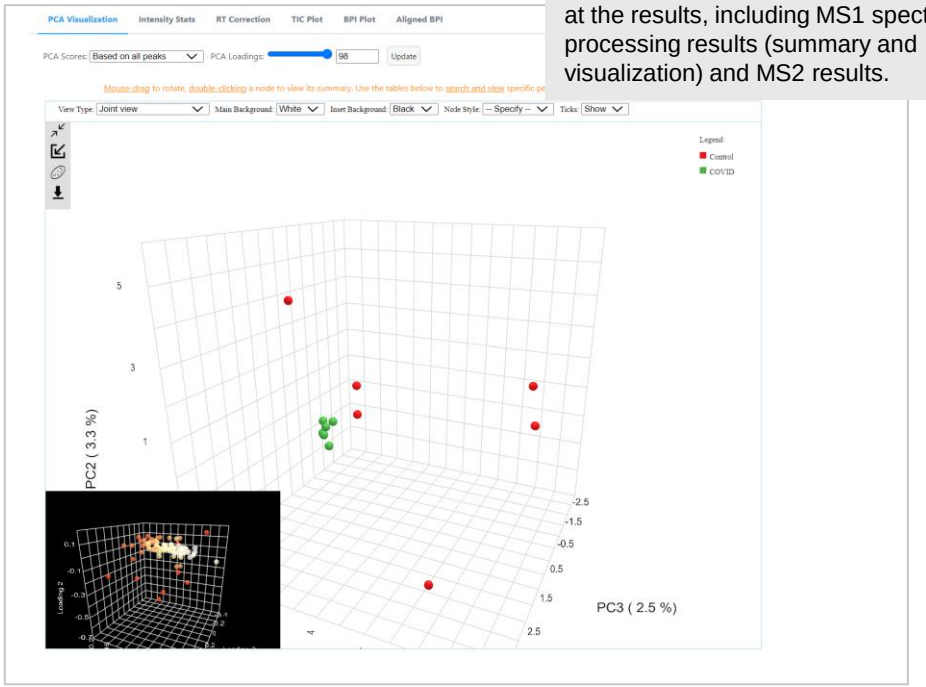
The status and processing logs are displayed here. A total of 12 steps will appear for MS1 + SWATH-DIA:

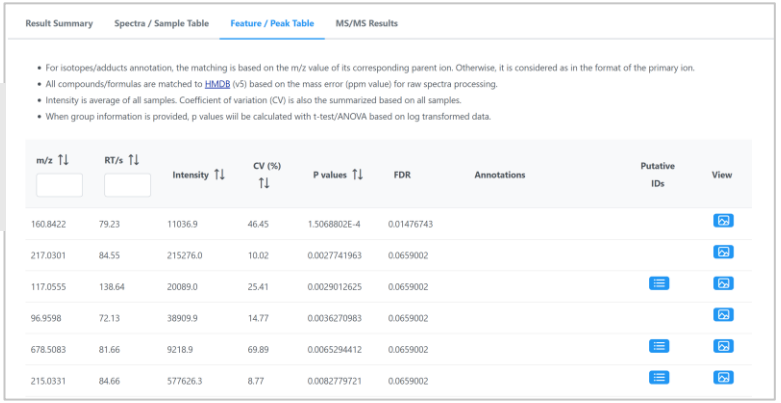
Step 1 to 6: LC-MS1 spectra processing, including data import, peak picking, alignment, gap filling, peak annotation and LC-MS1 results export;

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

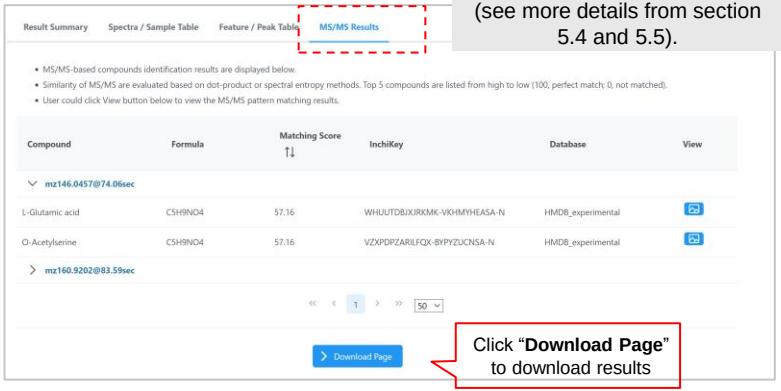
6.6 Result Exploration

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



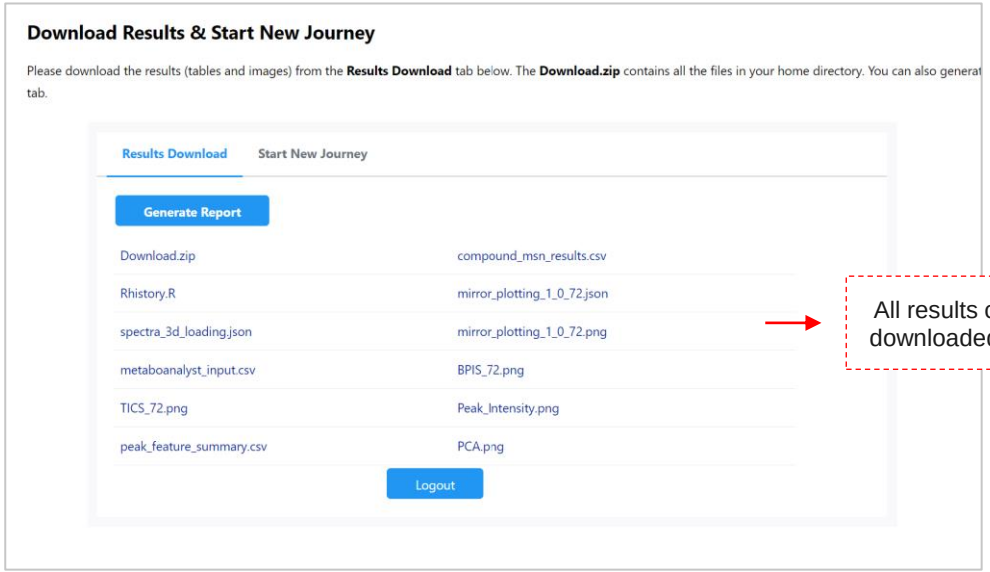


MS2 results are summarized in the tab, MS/MS results. (see more details from section 5.4 and 5.5).



Click “**Download Page**” to download results

7 Result Downloading



In summary

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

Or contact us:

zhiqiang.pang[at]xialab.ca

jeff.xia[at]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.