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;

Spectra preparation

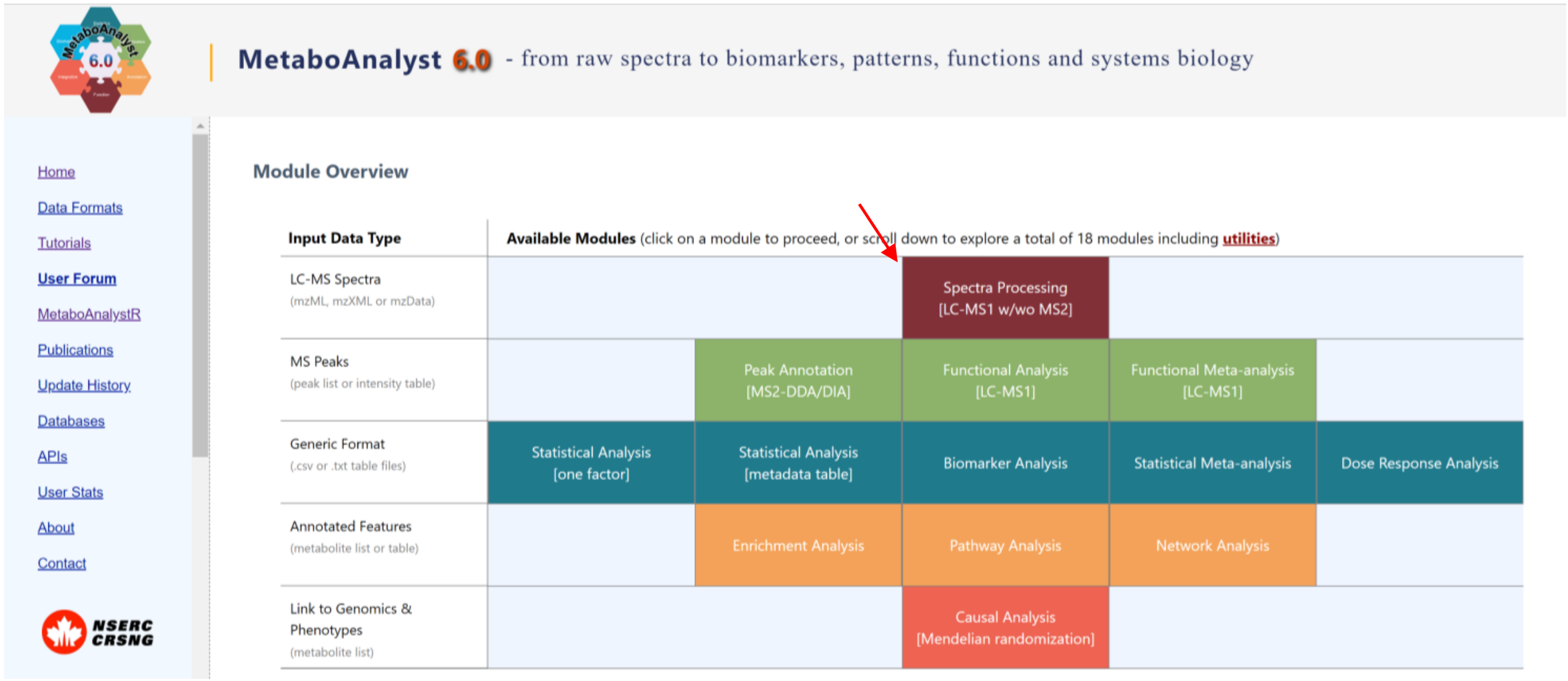
mzML, mzXML, mzData, CDF (centroided)

Zipped Spectra Files (.zip)

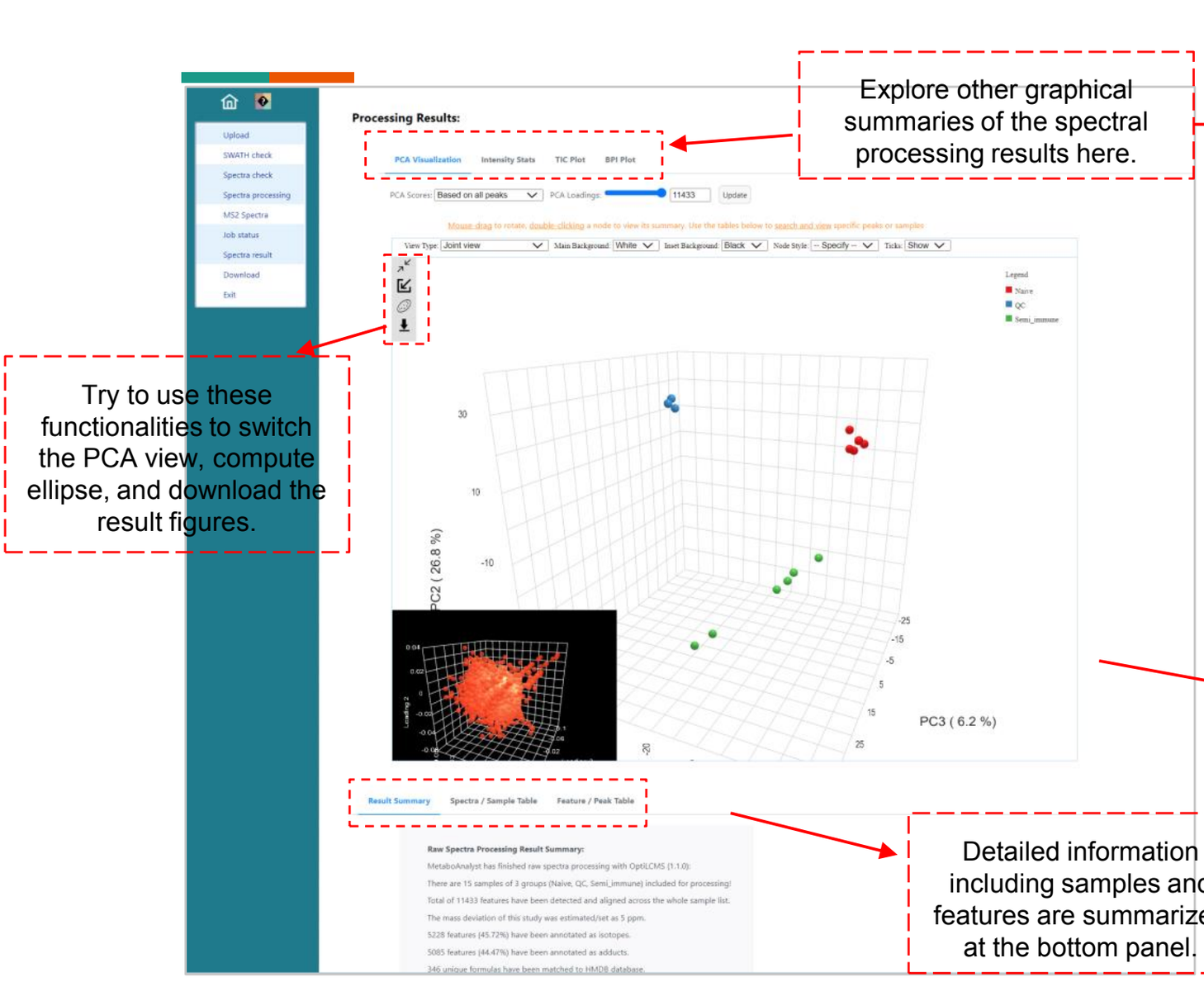
upload

Spectra processing and peak calling

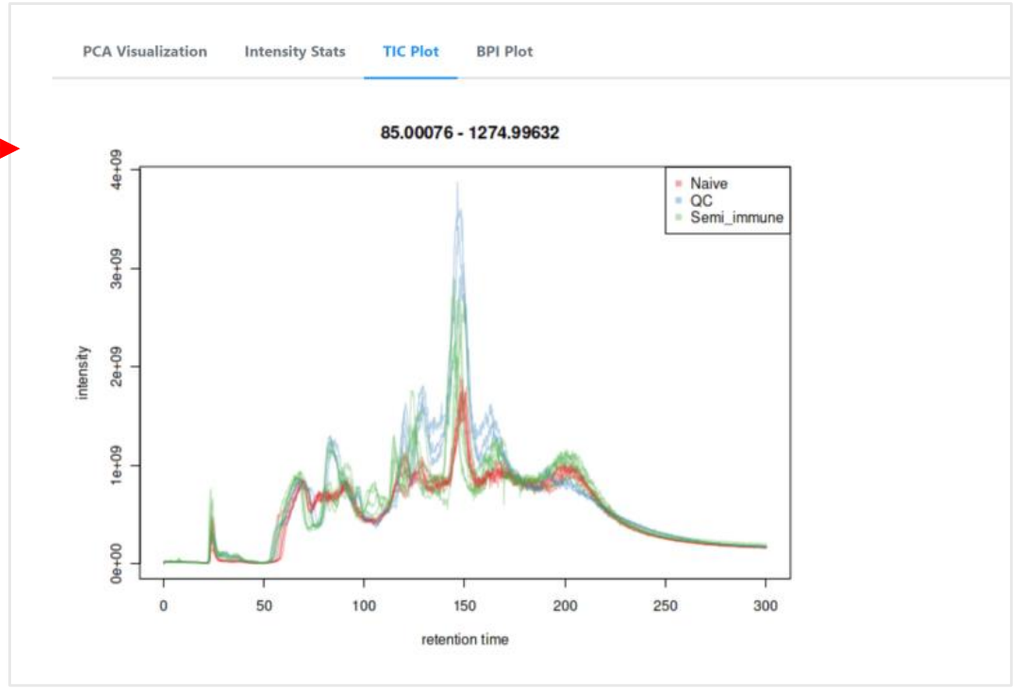
Go to MetaboAnalyst (https://www.metaboanalyst.ca), and select the module



Navigate results



Explore other graphical summaries of the spectral processing results here.



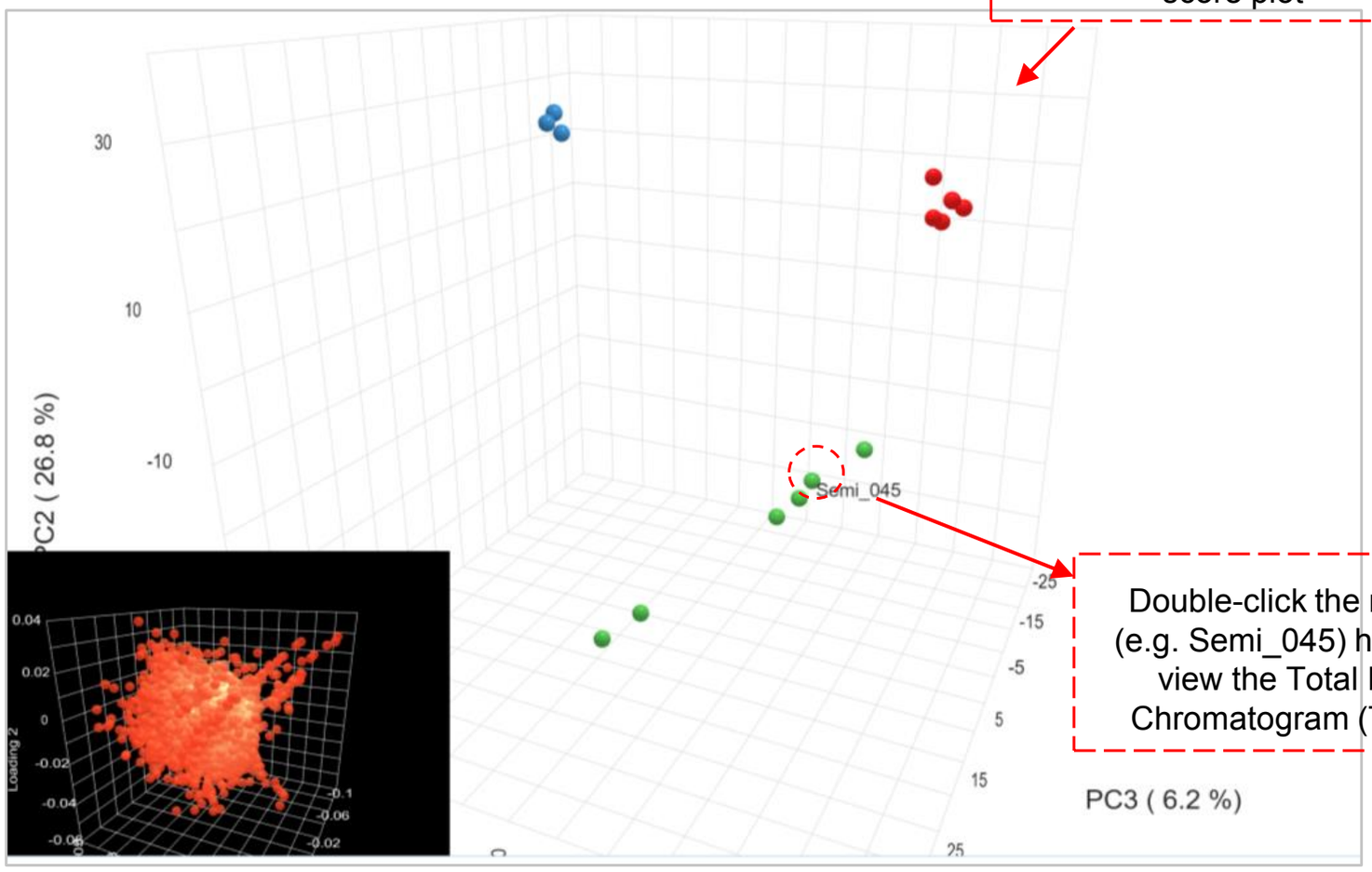
Try to use these functionalities to switch the PCA view, compute ellipse, and download the result figures.

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

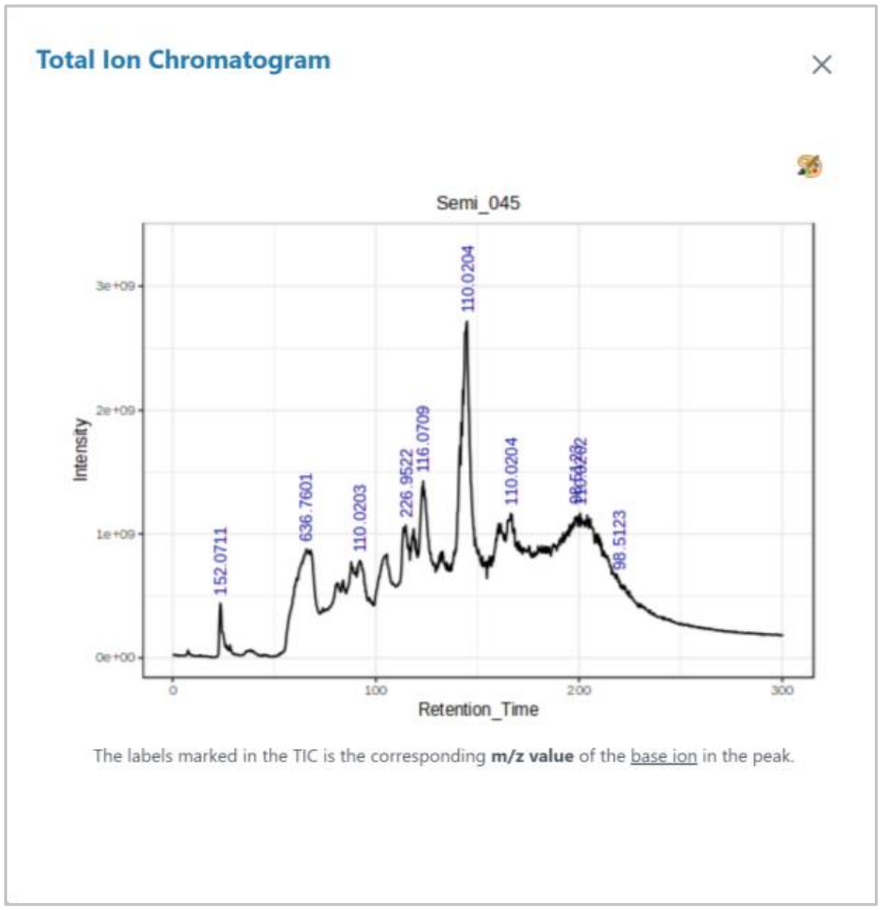
Detailed information including samples and features are summarized at the bottom panel.

Navigate results

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

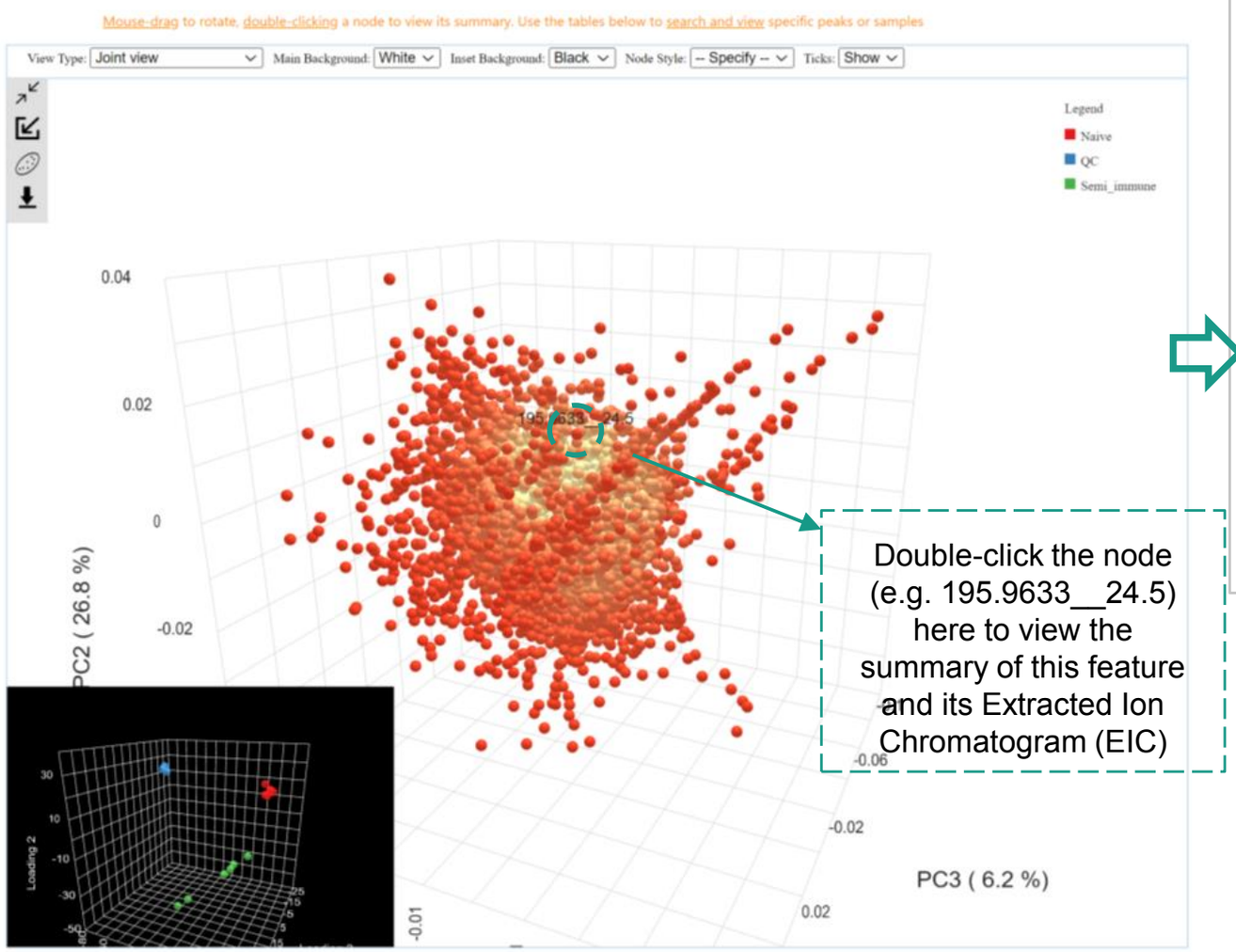


Double-click the node (e.g. Semi_045) here to view the Total Ion Chromatogram (TIC).

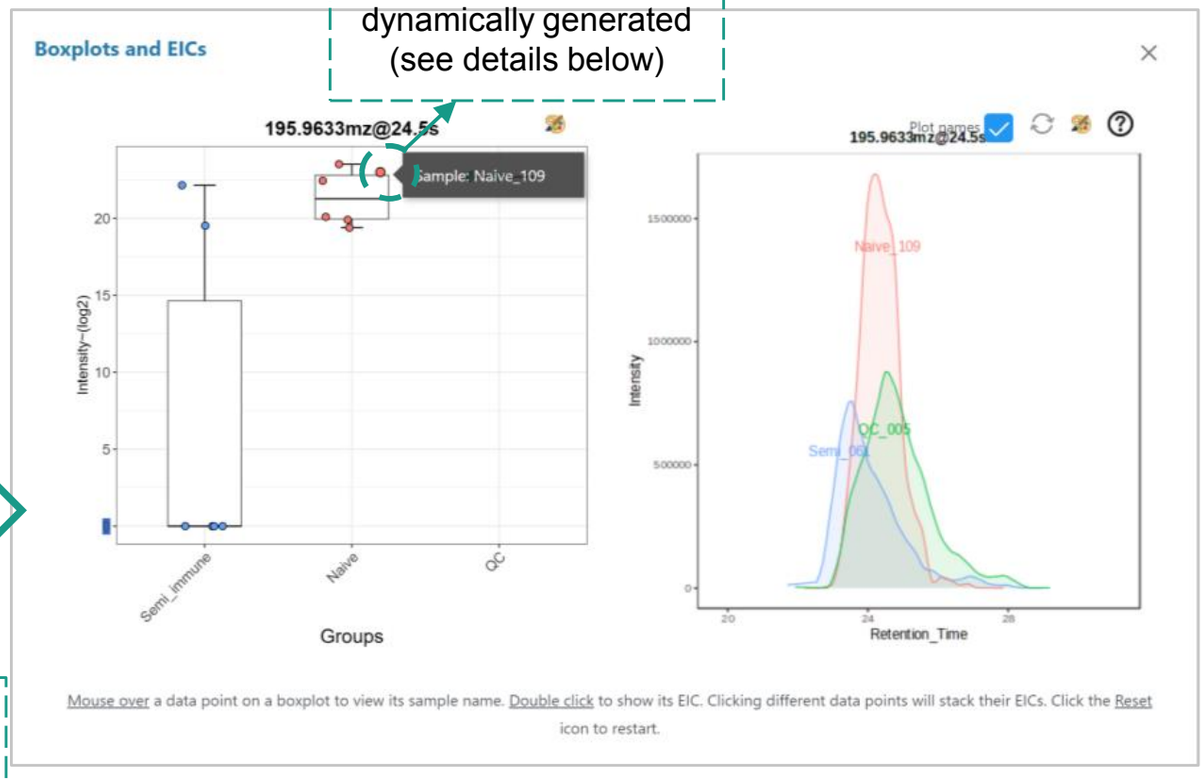


The labels marked in the TIC is the corresponding **m/z value** of the base ion in the peak.

Navigate results



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

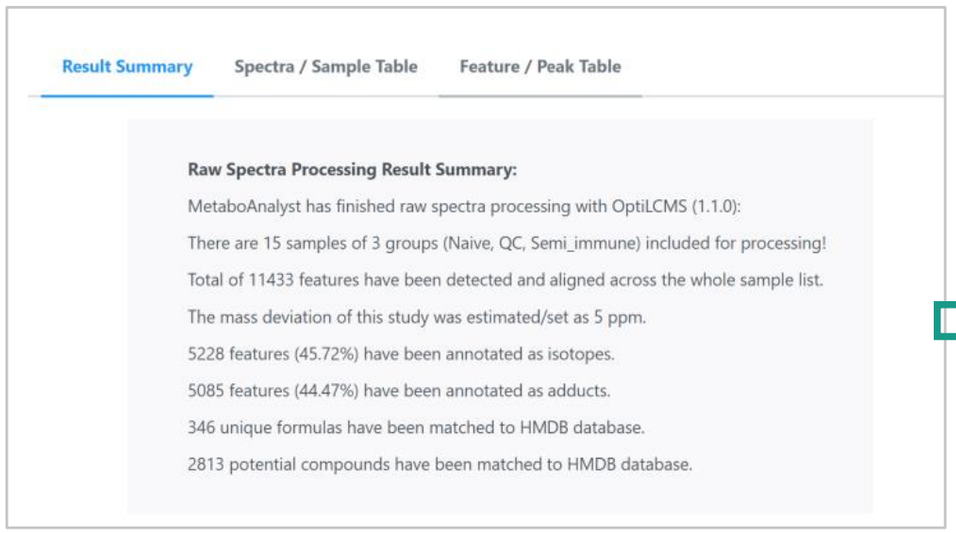


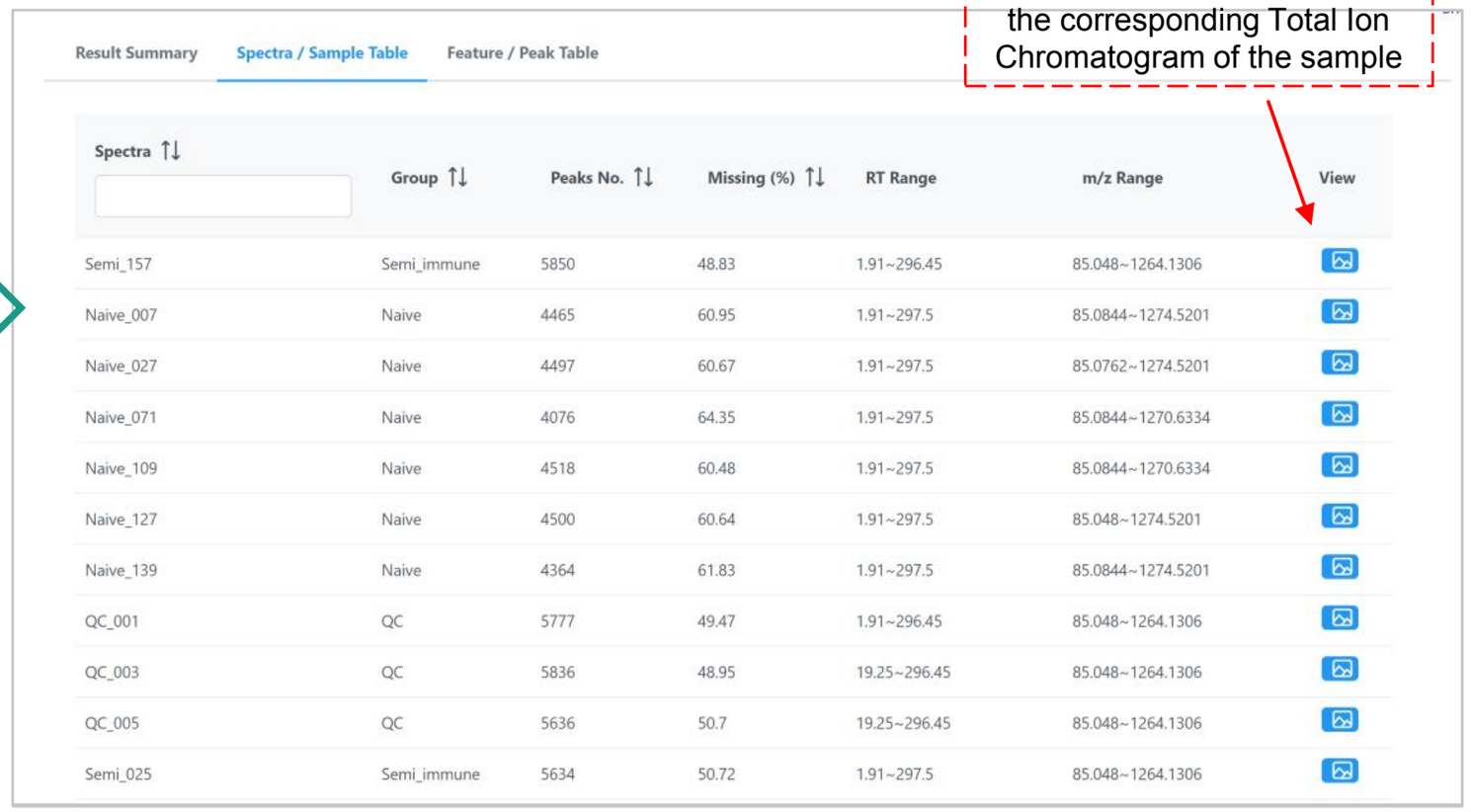
This EIC plot can be dynamically generated (see details below)

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.

Navigate results





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).

Navigate results

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 wiil 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-		
830.365552376317	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

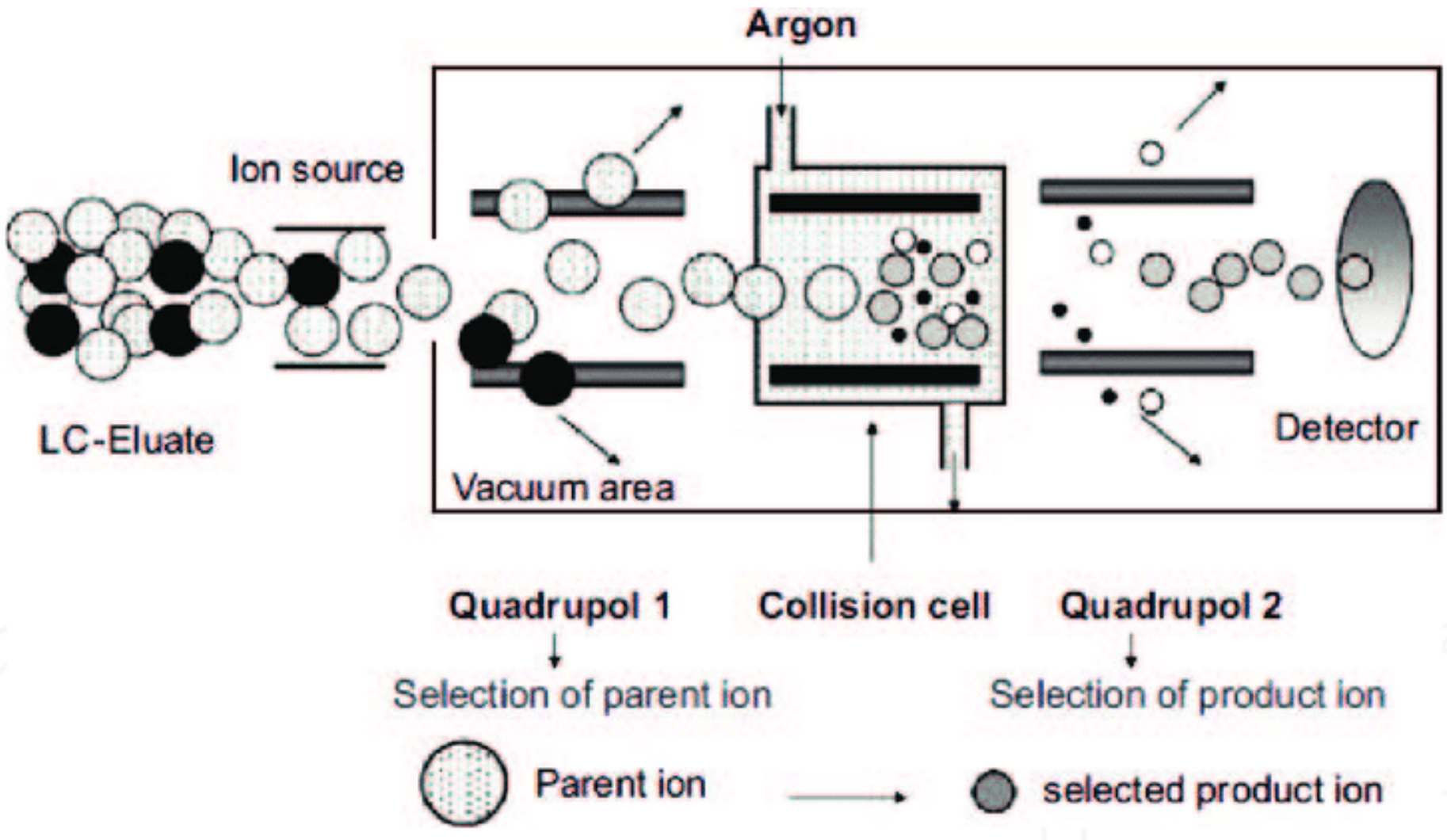
Formulas	Compounds
C15H24O3	1-Hydroxyepiacorone; Acorusdiol; 34->5-Abeo-4,11:4,12-diepoxy-3-eudesmanol; 4R,5S,7R,11R-11,12-Dihydroxy-110-spirovetiven-2-one; Apotrichodiol; 6alpha-Carissanol; alpha-Carissanol; Epioxylubimin; Dihydromyoporone; Piperalol; Zedoarondiol; Hydroxypelenolide; Toxin FS2; Urodiolenone; Bisacurone B; 2,3-Dihydroabscisic alcohol; 3-Methyl-5-pentyl-2-furanpentanoic acid; 3-Methyl-5-propyl-2-furanheptanoic acid;

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

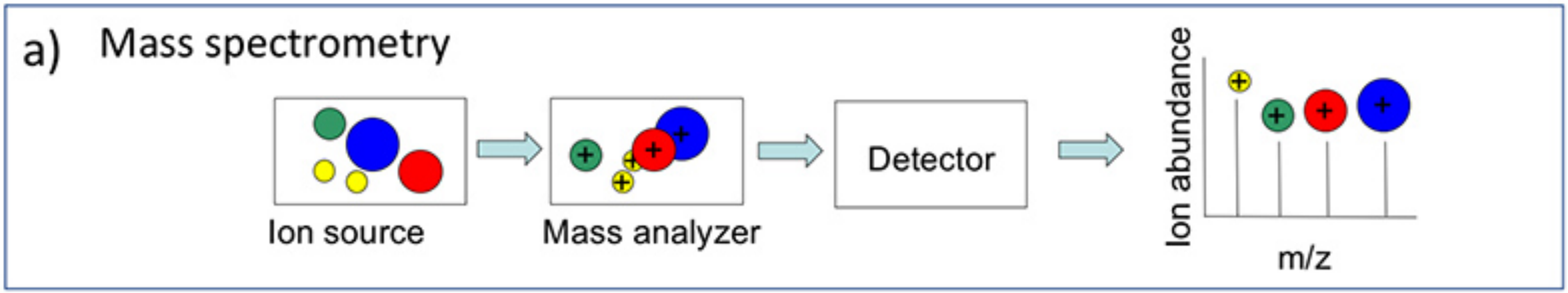
Features without the putative IDs from HMDB database would not show the Icon.

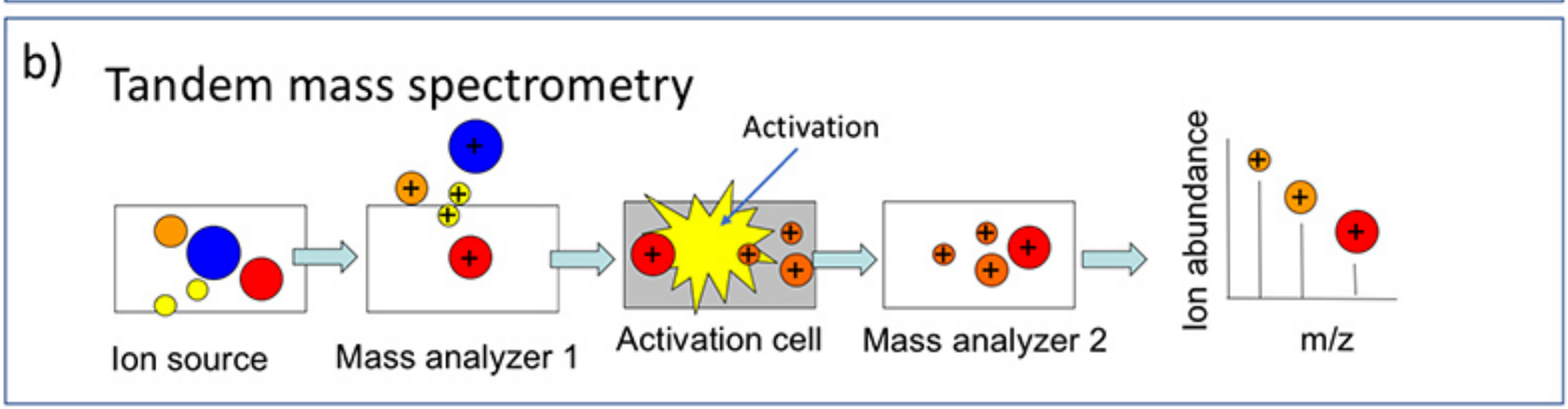
Click the “**View**” icon to start the corresponding dynamic EIC plot panel.

Tandem MS


Argon
Ion source
LC-Eluate
Vacuum area
Detector
Quadrupol 1
Collision cell
Quadrupol 2
Selection of parent ion
Selection of product ion
Parent ion
selected product ion

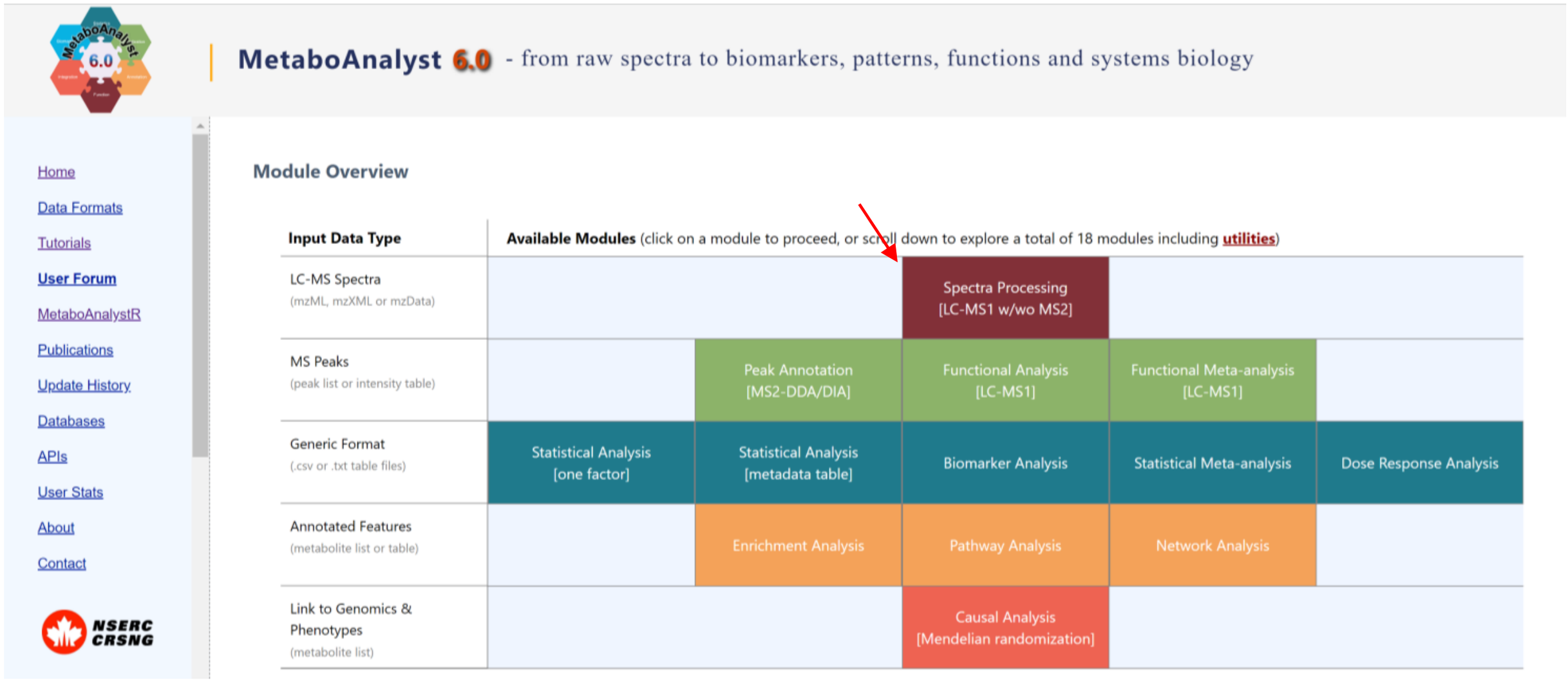
Tandem MS


a) Mass spectrometry
Ion source
Mass analyzer
Detector
Ion abundance
m/z

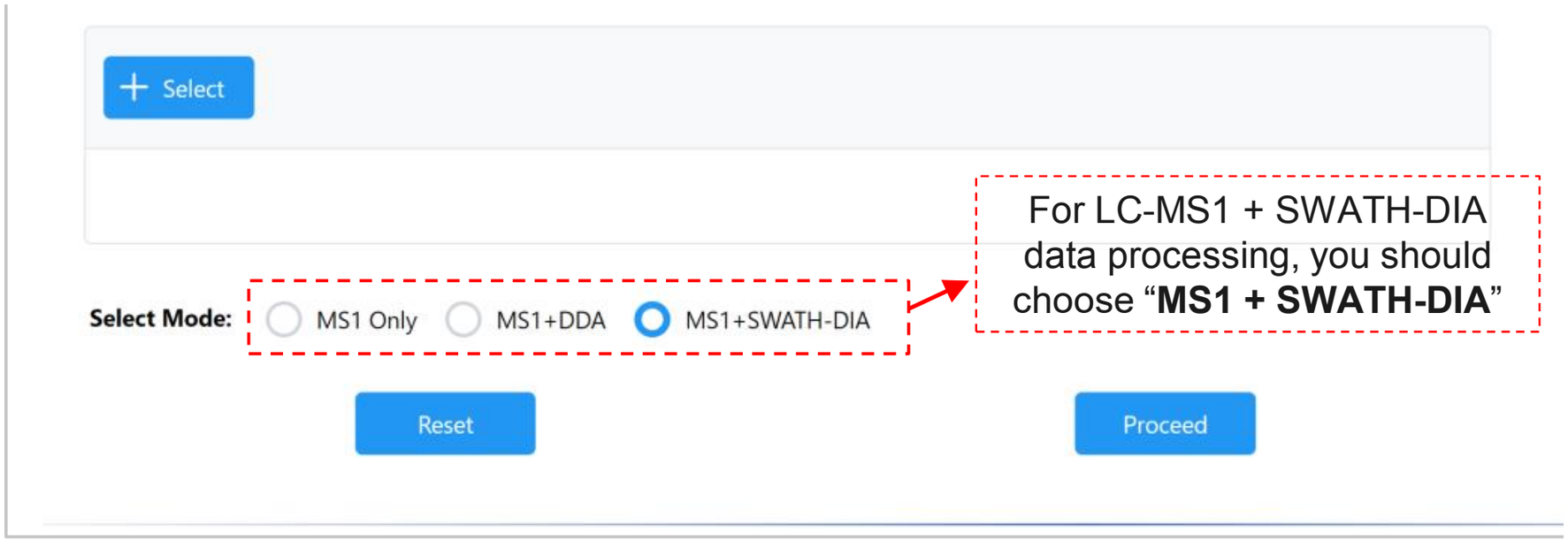

b) Tandem mass spectrometry
Activation
Ion source
Mass analyzer 1
Activation cell
Mass analyzer 2
Ion abundance
m/z

Spectra processing and peak calling

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



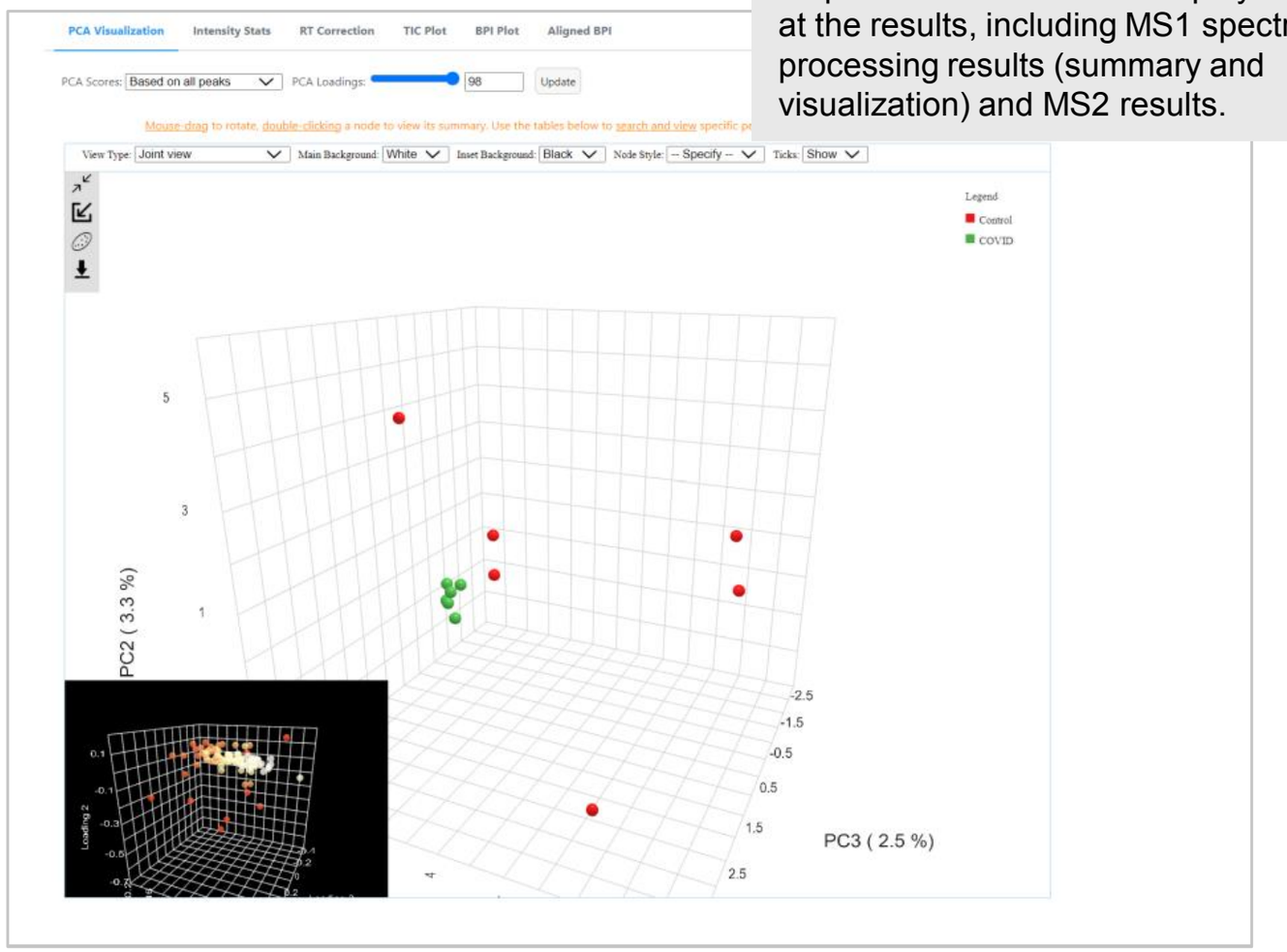
Spectra processing and peak calling

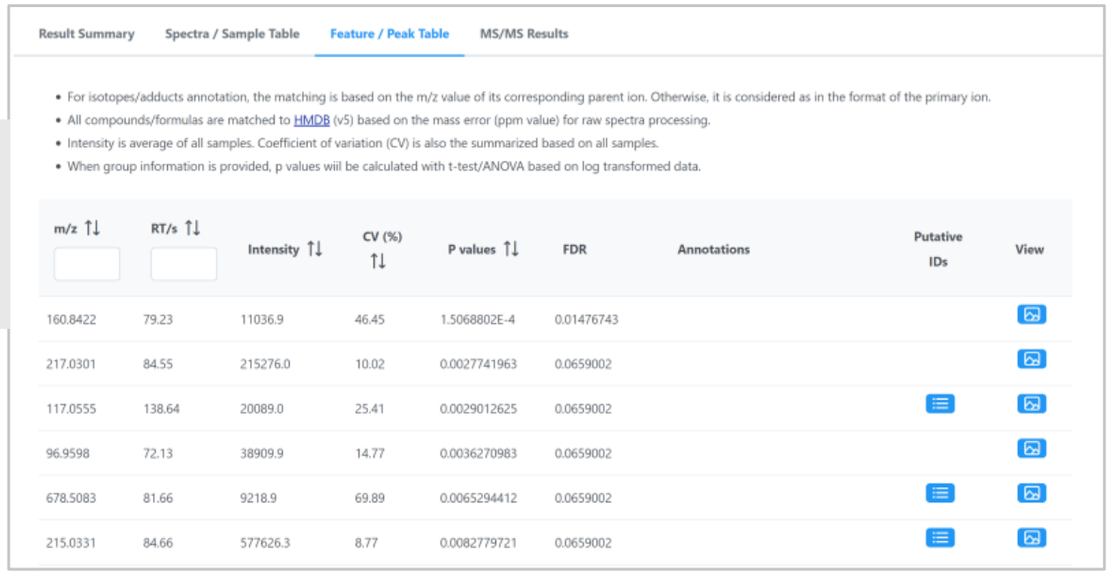


https://new.metaboanalyst.ca/MetaboAnalyst/faces/Share?ID=q2m2bo7wo_12873

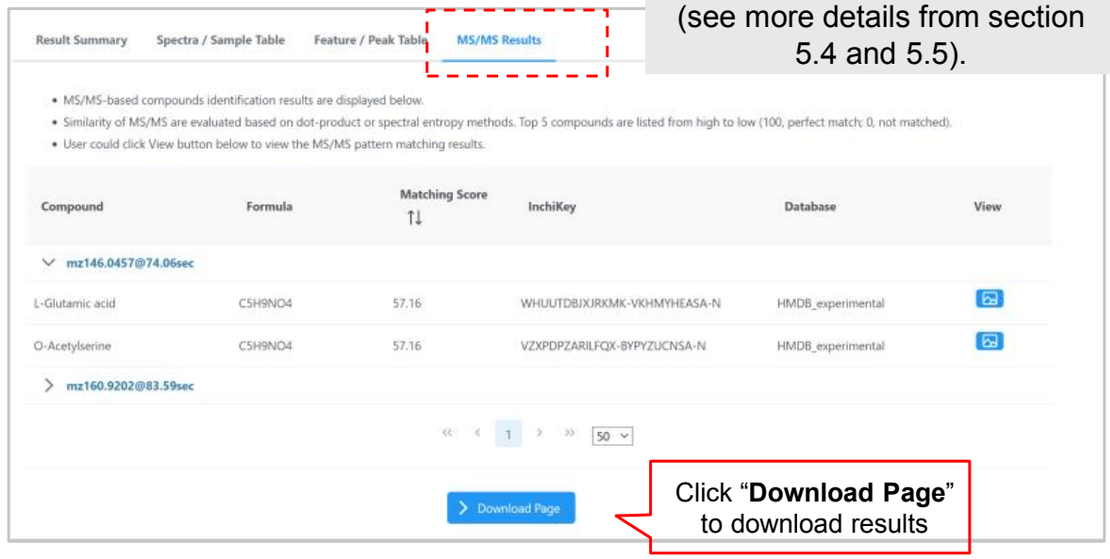
Navigate results

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

Summary peak calling

- 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.

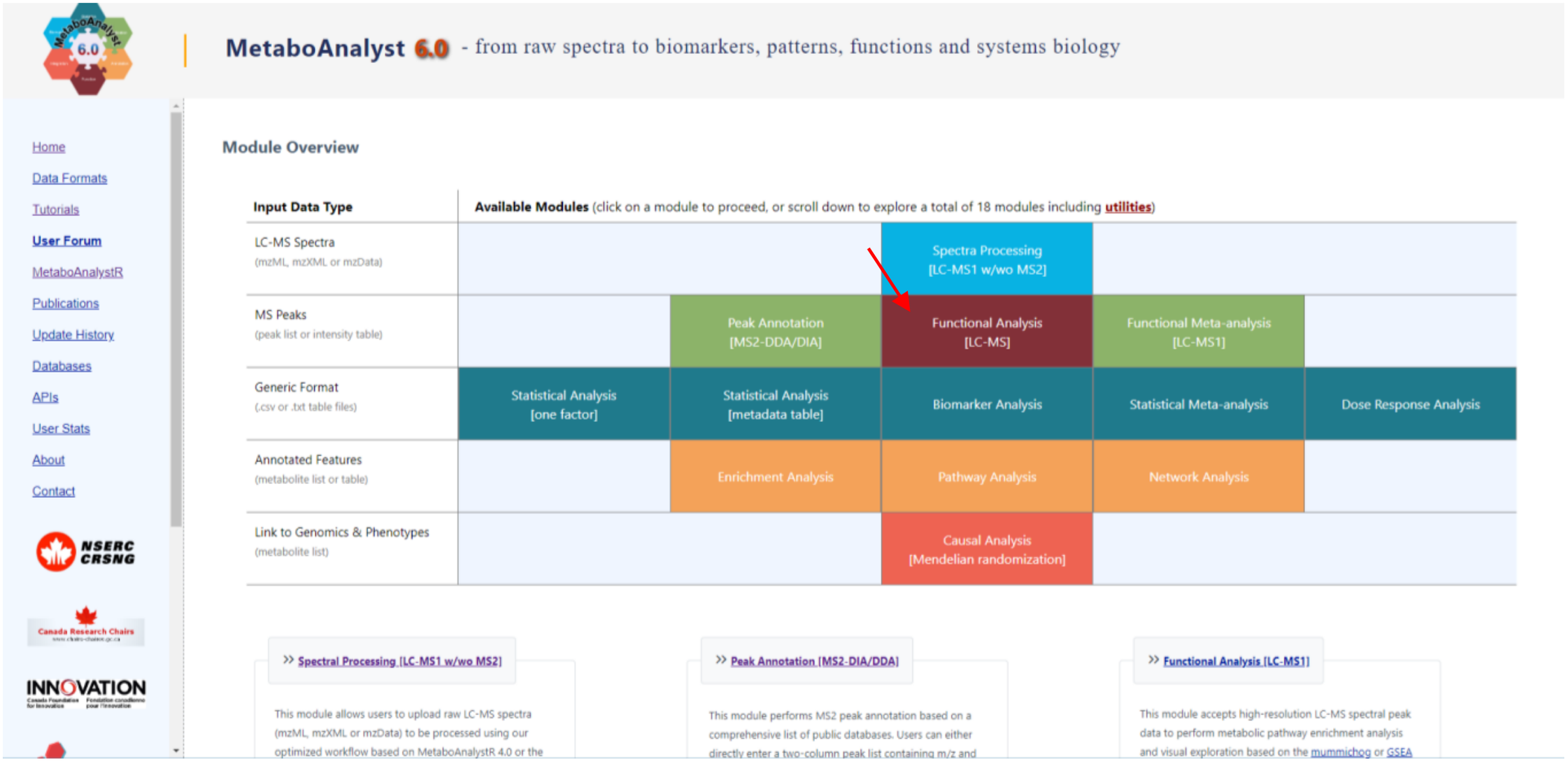
- For SWATH-DIA data, the SWATH window design is automatically extracted from the spectra. If the related information is missing, users will be prompted to enter the window design manually.

- 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.

Functional analysis

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



Functional analysis

File preparation

Two files need to prepared:

1. **LC-MS1 peak list:** this file should consist of multiple columns containing complete LC-MS1 features. m/z, retention time (rt), and p values are required for accurate functional analysis. Besides, users are recommended to provide t scores column. Please note that this peak list must contain all LC-MS1 features (no matter they are significant or not). Usually, for untargeted metabolomics on a biological sample, the complete features number is over 5,000.
2. **MS2-based compound candidate list:** This file should consist of the MS2-based compound identification results. This table can be in two formats:
   - **Format 1:** A specific column, named as “index” added before the compound candidate columns. The index refers to the corresponding number the LC-MS1 peak list (see example below). Users can provide 3-10 chemical candidate for each MS1 feature;
   - **Format 2:** The number of rows of the two data should be the same and corresponding (see next page).

Functional analysis

File preparation

Header of data is required

	A	B	C
1	mz	rt	p.value
2	139.5311	44.87	5.38E-06
3	204.056	42.26	2.31E-05
4	203.1279	784.65	7.5E-05
5	521.3151	304.9	0.001754
6	345.1516	577.72	0.002652
7	250.1777	481.76	0.004474
8	714.5077	762.3	0.004555
9	189.0737	53.16	0.004971
10	307.5487	461.67	0.005336
11	366.1607	190.66	0.005795
12	776.3663	782.33	0.005948
13	486.2957	714.81	0.006051
14	599.4561	768.32	0.006113
15	344.2124	512.66	0.006171
16	492.2939	319.1	0.006195
17	678.3317	782.33	0.007085
18	713.422	283.4	0.00759
19	238.5637	44.38	0.008047

LC-MS1 peak list

	A	B	C	D	E	F	G	H
1	index	InchiKey_1	InchiKey_2	InchiKey_3	InchiKey_4	InchiKey_5		
2	2	MGSKVZWGBWPBTF	TZJAEGCLMLTGRJ-U	OVRNDRQMDRJTHS-	NA	NA		
3	8	FWULQXYJOANGSS-	JDVVGAQPNNXQDW	BJTZATWAPRKXBV-U	LIWMQSWFLXEGMA	VEQOALNAAJBPNY-U	HFFFAOYSA-N	
4	10	LUALIOATIOESLM-U	NA	NA	NA	NA		
5	42	WIYUZYBFCWCCQJ-	HGSOUJPIFSDBKJ-O	NA	NA	NA		
6	44	HOVAGTYPODGVJG-	RYYVLZVUVIJVGH-UI	NA	NA	NA		
7	56	FQZPXSRKCOWUEI-	YRXOQXUDKDCXME	YRXOQXUDKDCXME	OTGQIQQTPXJQRG-	NA		
8	68	RYYVLZVUVIJVGH-UI	PXQPEWDEAKTCGB	KSEBMYQBYZTDHS-I	QURCVMIEKCOAJU-	NA		
9	70	HJMQDJPMQIHLPB-I	AMDPNECWKZZEBQ	IRZVHDLBAYNPCT-U	NA	NA		
10	74	YEJYLHKQOBOSCP-(	OKJCFMUGMSVJBG-	NPJICTMALKLTFW-C	NA	NA		
11	94	XJLSEXAGTJCILF-UH	JWYOAMOZLZXDER-	NILQLFBWTXNUOE-	PJDFLNIOAUIZSL-UH	DDSLGZOYEPKPSJ-U	HFFFAOYSA-N	
12	112	DKLKMKYDWHYZTD-	SSEBTPPFLLCUMN-I	MTFCPNHRBINLRQ-	IGLHHSKNBDXCEY-	NA		
13	123	IEPGNWMPIFDNSD-	IEPGNWMPIFDNSD-	WECGLUPZRHILCT-	NA	NA		
14	129	HXFOXFJUNFFYMO-I	NA	NA	NA	NA		
15	141	OTCCIMWXFLJLIA-B	NXQJDVBMMRCKQC	NRNCYVBFPDDJNE-	POJWUDADGALRAB-	NA		
16	160	IZYCZPAFZQFMCQ-U	NVEPPWDVLBMNME	GHOKWGTUZJEAQD	WTDRDQBEARUVNC	JXXCENBLGFBQJM-U	HFFFAOYSA-N	
17	161	LUINDDOUWHRIPW-	DFQOXFIPAAMFAU-U	NA	NA	NA		
18	186	VLSMHEGGTFMBBZ-	NA	NA	NA	NA		
19	198	IRZTUXPRIUZXMP-UI	RTIXKCRFFJGDFG-U	LCAWNFIFMLXZPQ-I	NA	NA		
20	205	XUQWWIFROYJHCU-	SEKYBDYVXDAYPY-A	SEKYBDYVXDAYPY-U	IDTCGADOYRIRKB-U	NA		
21	206	CMRNMZJAUFXOQF-	FMMOOAYVCKXGMF	XSXIVVZCUAHUJO-H	XSXIVVZCUAHUJO-U	XUJWOMMOEOHPFP-U	TJQPWESSA-N	

LC-MS2-based compound identification results list

Functional analysis

File preparation

The number of rows of the two data should be the same and corresponding to each other;
If there are no MS2-based compound identification results, please fill **NA** in the rows.
You can provide 3-10 chemical candidate for each MS1 feature.

	A	B	C	D	E
1	mz	rt	t.score	p.value	mode
2	52.99813	1202.7	1.4439	0.165	positive
3	53.00038	1295.16	2.474	0.0193	positive
4	53.00228	1291.68	2.7635	0.0094	positive
5	53.00296	1273.08	3.1435	0.0037	positive
6	53.00432	1264.08	2.7164	0.0106	positive
7	59.04659	1076.76	2.1416	0.042	positive
8	59.04682	1106.76	0.9601	0.3477	positive
9	59.04705	992.94	1.3661	0.1827	positive
10	59.04716	871.5	1.6124	0.1168	positive
11	59.04739	1259.94	0.2004	0.8432	positive
12	59.04752	1129.98	0.5856	0.5635	positive
13	59.04757	677.64	0.2032	0.8408	positive
14	59.04778	1002.6	0.1681	0.8682	positive
15	59.048	1302.06	2.0595	0.0483	positive
16	59.04802	968.1	0.2093	0.8358	positive
17	59.04805	1184.76	0.852	0.402	positive
18	59.04829	1122.66	0.2977	0.7684	positive
19	59.04842	865.98	0.387	0.7017	positive
20	59.04854	1069.5	0.8101	0.428	positive
21	59.04856	1033.62	1.0973	0.2836	positive

LC-MS1 peak list

	A	B	C	D	E
1	Inchikey1	Inchikey2	Inchikey3	Inchikey4	Inchikey5
2	NA	NA	NA	NA	NA
3	NA	NA	NA	NA	NA
4	NA	NA	NA	NA	NA
5	NA	NA	NA	NA	NA
6	NA	NA	NA	NA	NA
7	NA	NA	NA	NA	NA
8	XUWHAWMETYGRKE	NYEZZYQZRQDLEH-I	SECXISVLQFMRJM-U	NA	NA
9	XUWHAWMETYGRKE	NYEZZYQZRQDLEH-I	SECXISVLQFMRJM-U	NA	NA
10	XUWHAWMETYGRKE	NYEZZYQZRQDLEH-I	SECXISVLQFMRJM-U	NA	NA
11	PAFZNILMFXTMIY-UI	NA	NA	NA	NA
12	PAFZNILMFXTMIY-UI	NA	NA	NA	NA
13	NA	NA	NA	NA	NA
14	DYDCUQKUCUHJBH	KYCJNIUHWNJNCT-	KHIQJCVGWNEQMI-	DYDCUQKUCUHJBH	NA
15	NA	NA	NA	NA	NA
16	NA	NA	NA	NA	NA
17	NA	NA	NA	NA	NA
18	NA	NA	NA	NA	NA
19	NA	NA	NA	NA	NA
20	NUVWVUPJCXRIIW-I	KFDVPJUYSDEJTH-U	KGIGUEBEKRSTEW-I	YAXKTBLXMTYWDQ-	NA
21	NUVWVUPJCXRIIW-I	KFDVPJUYSDEJTH-U	KGIGUEBEKRSTEW-I	YAXKTBLXMTYWDQ-	NA

LC-MS2-based compound identification results list

Functional analysis

Please upload your data

This module supports functional analysis of untargeted metabolomics data generated from high-resolution mass spectrometry (HRMS). The basic assumption is that putative annotation at individual compound level can collectively predict changes at functional levels as defined by **metabolite sets** or **pathways**. This is because changes at group level rely on "collective behavior" which is more tolerant to random errors in compound annotation as demonstrated by Li et al. To use this approach,

- The input peak list or peak table must contain the complete data, not just significant data - we need the complete data to estimate the null model (background);
- [Required] Feature or peak names must be their numeric mass (m/z) values for putative annotation;
- [Optional] You can also provide retention time (RT) to further improve peak annotation

A peak list profile | A peak intensity table | MS Peak and Annotation lists

For functional analysis with LC-MS1 + MS2, we should use "**MS Peak and Annotation lists**"

Upload peak and compound lists

Ion Mode: Negative Mode

Mass Tolerance (ppm): 5.0 (editable)

Retention Time: Not present

Annotation ID type: InchiKeys

Enforce Primary Ions (V2 only):

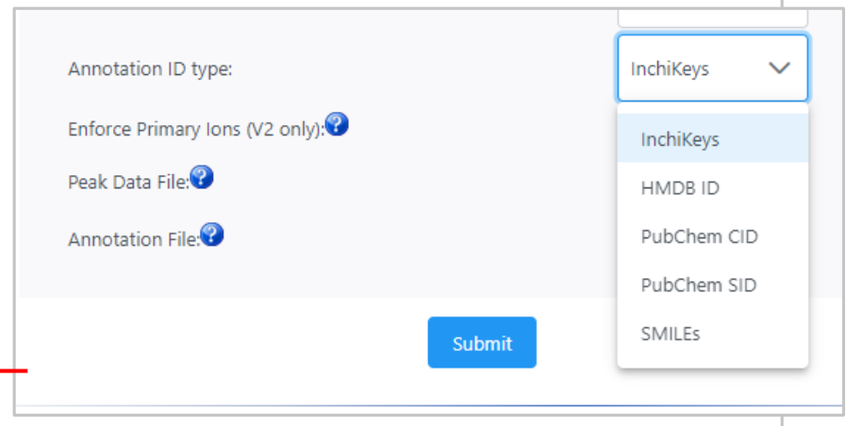
Peak Data File: + Choose

Annotation File: + Choose

Submit

Click "**Choose**" buttons to upload either peak list or annotation list respectively.

Click "**Submit**" to upload your data



Parameters on the data and MS instrument need to specified here. Multiple MS2-based compound ID are supported.

Functional analysis

Data Integrity Check:

- Checking sample names - spaces will replaced with underscore, and special characters will be removed;
- Checking the class labels - at least three replicates are required in each class.
- The data (except class labels) must not contain non-numeric values.
- If the samples are paired, the pair labels must conform to the specified format.
- The presence of missing values or features with constant values (i.e. all zeros).

Data processing information:

Checking data content ...passed.

A total of 7340 m/z features were found in your uploaded data.

5 compounds found in your uploaded data.

The instrument's mass accuracy is **5** ppm.

The instrument's analytical mode is **positive** .

The uploaded data contains **3** columns.

The column headers of uploaded data are **m.z, p.value, r.t** .

The range of m/z peaks is trimmed to 50-2000. **0** features have been trimmed.

A total of 7340 input mz features were retained for further analysis.

A total of 1455 InchiKeys Compounds included.

Edit Groups | Missing Values | Proceed

MetaboAnalyst could process your data and do an integrity check at first. The integrity check results are summarized here.

Click "***Proceed***" button to continue.

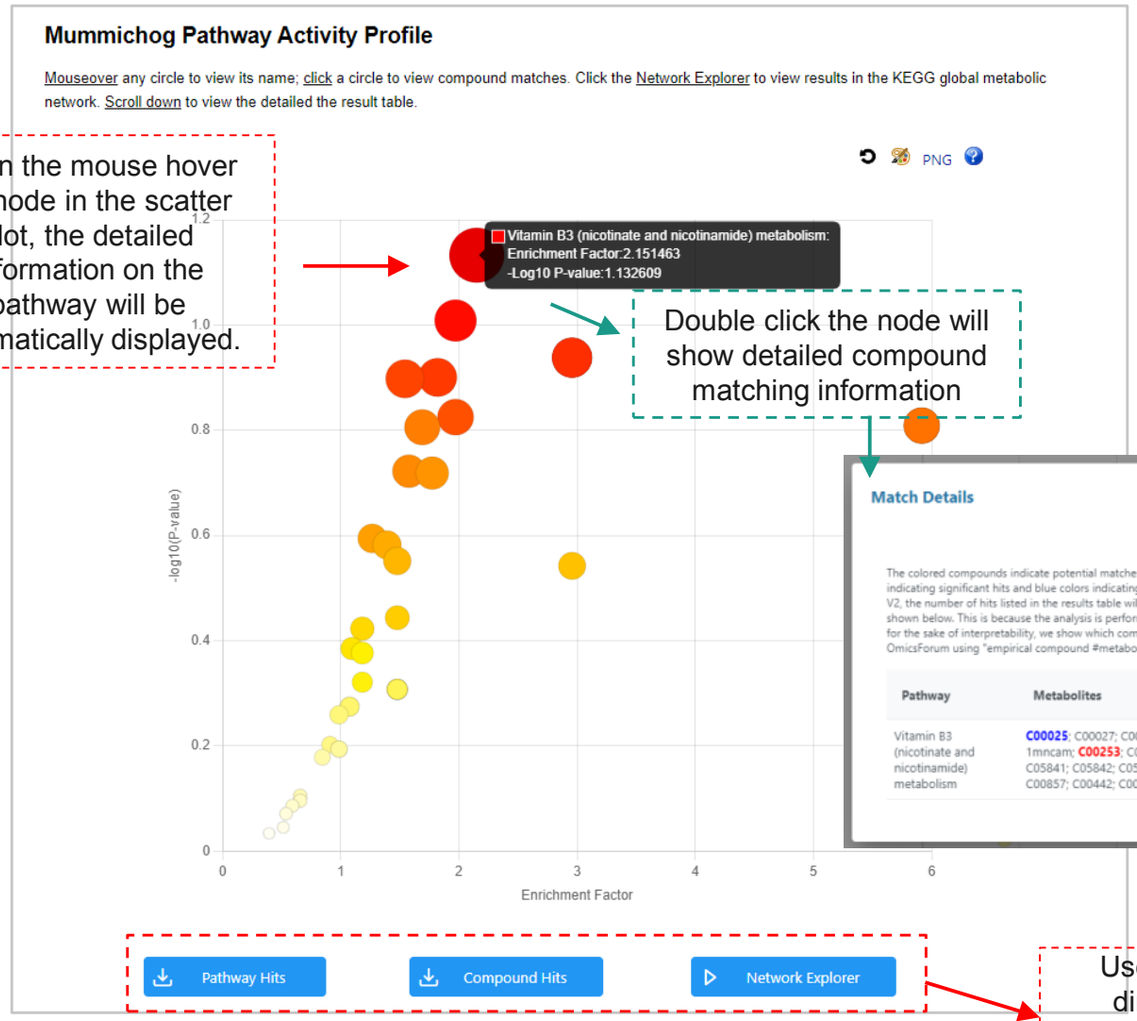
Navigate results

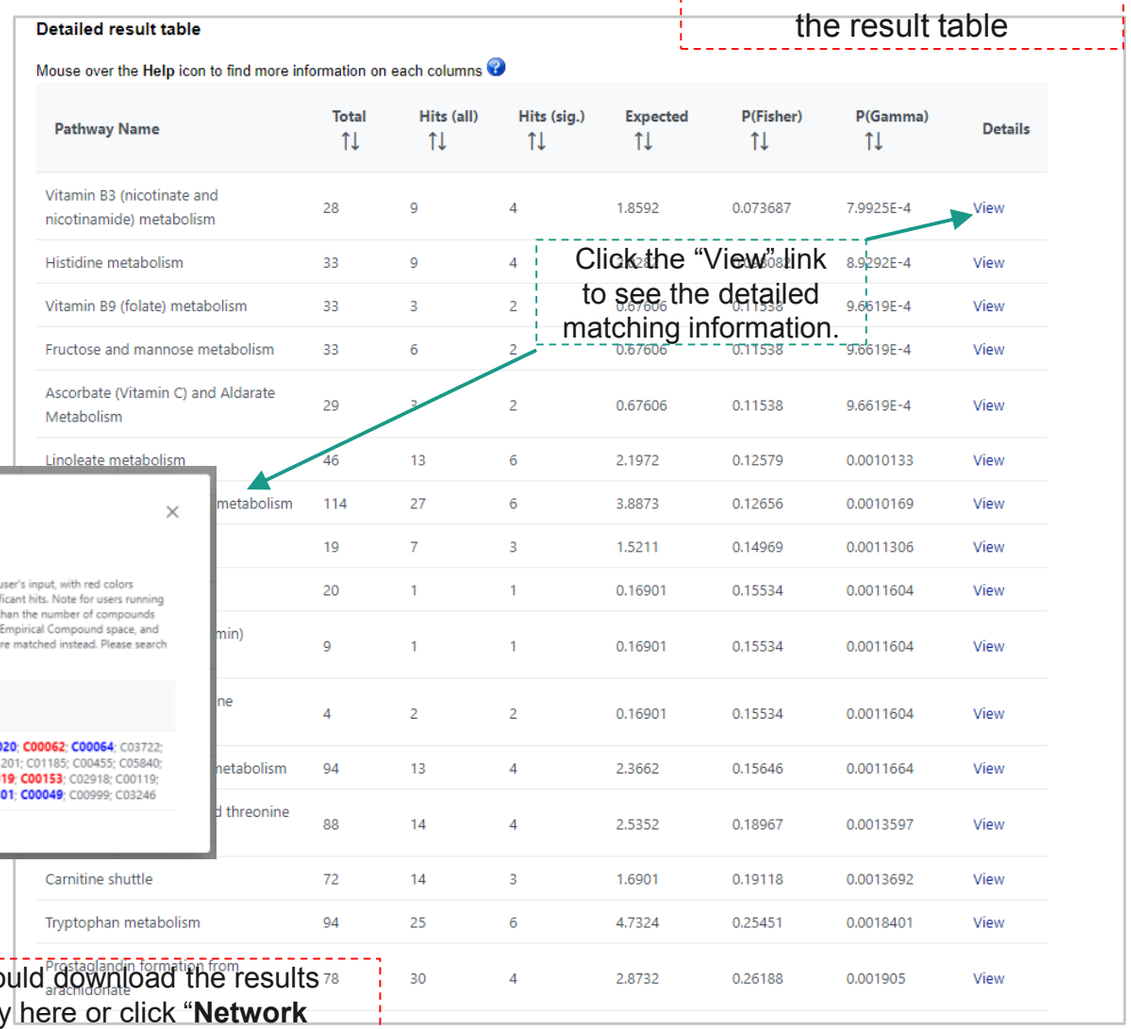
All pathway enrichment results are summarized in the result table

When the mouse hover the node in the scatter plot, the detailed information on the pathway will be automatically displayed.

Double click the node will show detailed compound matching information

Click the “View” link to see the detailed matching information.





User could download the results directly here or click “**Network Explorer**” to explore the results from Global Network view (see next page)