MetaboAnalyst 6.0

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

Functional Analysis [LC-MS]

2024-03-10

Module Overview

The module offers a comprehensive workflow for functional analysis on untargeted metabolomics dataset. Functional analysis based on LC-MS1 peak only (either peak list or peak table) is same as the previous version 5. In this tutorial, we focus on the enhanced features only

- ✓ Support for LC-MS1 feature based functional analysis;
- ✓ Support for LC-MS1 feature + LC-MS2 based compound information for functional analysis;
- ✓ Comprehensive knowledgebase have supported more than 130 species.

Content

In this tutorial, our focus lies solely on the newly incorporated functionalities pertaining to functional analysis involving LC-MS1 peaks in conjunction with MS2-based compound identification results. Any other functional analyses reliant only on MS1 features remain consistent with the previous version of the tutorial. For further details, please refer to the previous version here.

1. Introduction

Background

- Functional analysis of untargeted metabolomics was initially established based on *mummichog* and Gene Set Enrichment Analysis (GSEA) since MetaboAnalyst 4.0.
- It was further enhanced in MetaboAnalyst 5.0 by incorporating retention time data and m/z values into calculating empirical compounds.
- MetaboAnalyst 6.0 now allows users to upload an MS features list along with a corresponding MS2-based compound list to further filter out unrealistic empirical compounds to improve the accuracy in predicting pathway activity.

Data Formats

Functional analysis of untargeted metabolomics support multiple data formats as the input:

i. Peak List (including *m/z*, retention time, *p* values, t-scores, modes, etc.);
ii. Peak table (generic format);
iii. Peak List + Compound table.

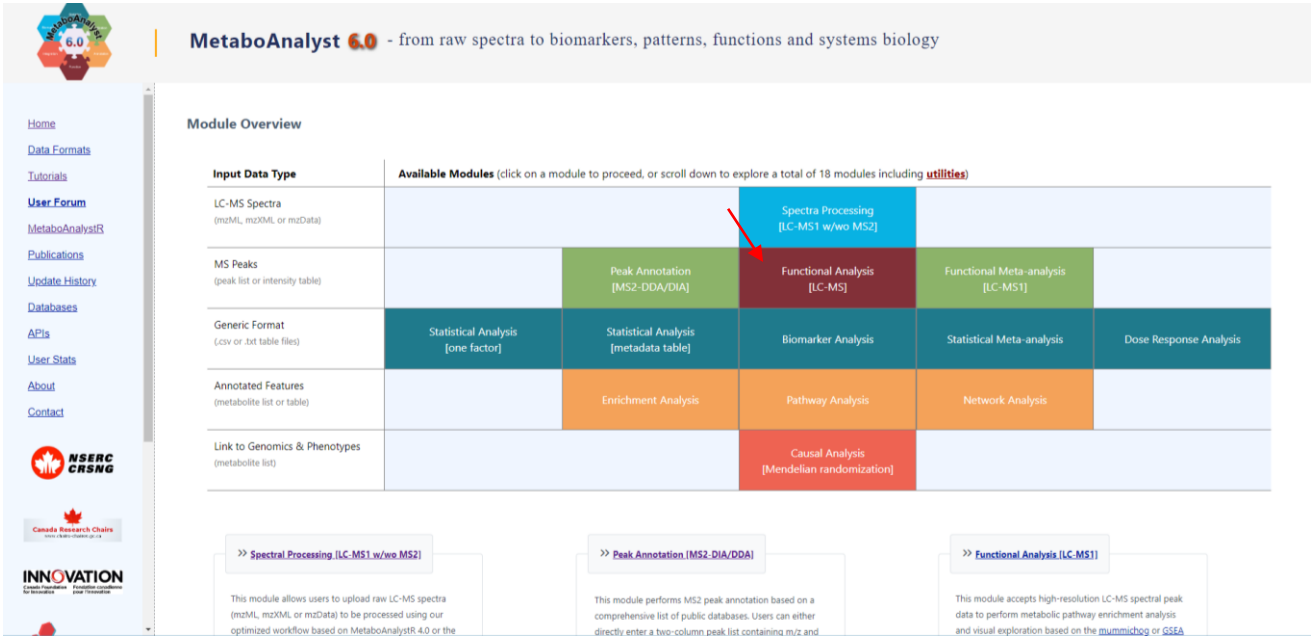
Expected Results

Functional Analysis [LC-MS] module provides user with the results on potential perturbed pathways and detailed potential chemical candidates:

i. Pathway enrichment results;
ii. Visualization on pathway analysis result in scatter plot or global network view;

2. Choose the Module

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



3. Data Preparation – format 1

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).

Header of data is required

	A	B	C	D
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-I	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-I	XSXIVVZCUAHUJO-U	XUJWOMMOEOHPFP-U	TJQPWESSA-N	

LC-MS2-based compound identification results list

3. Data Preparation – format 2

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

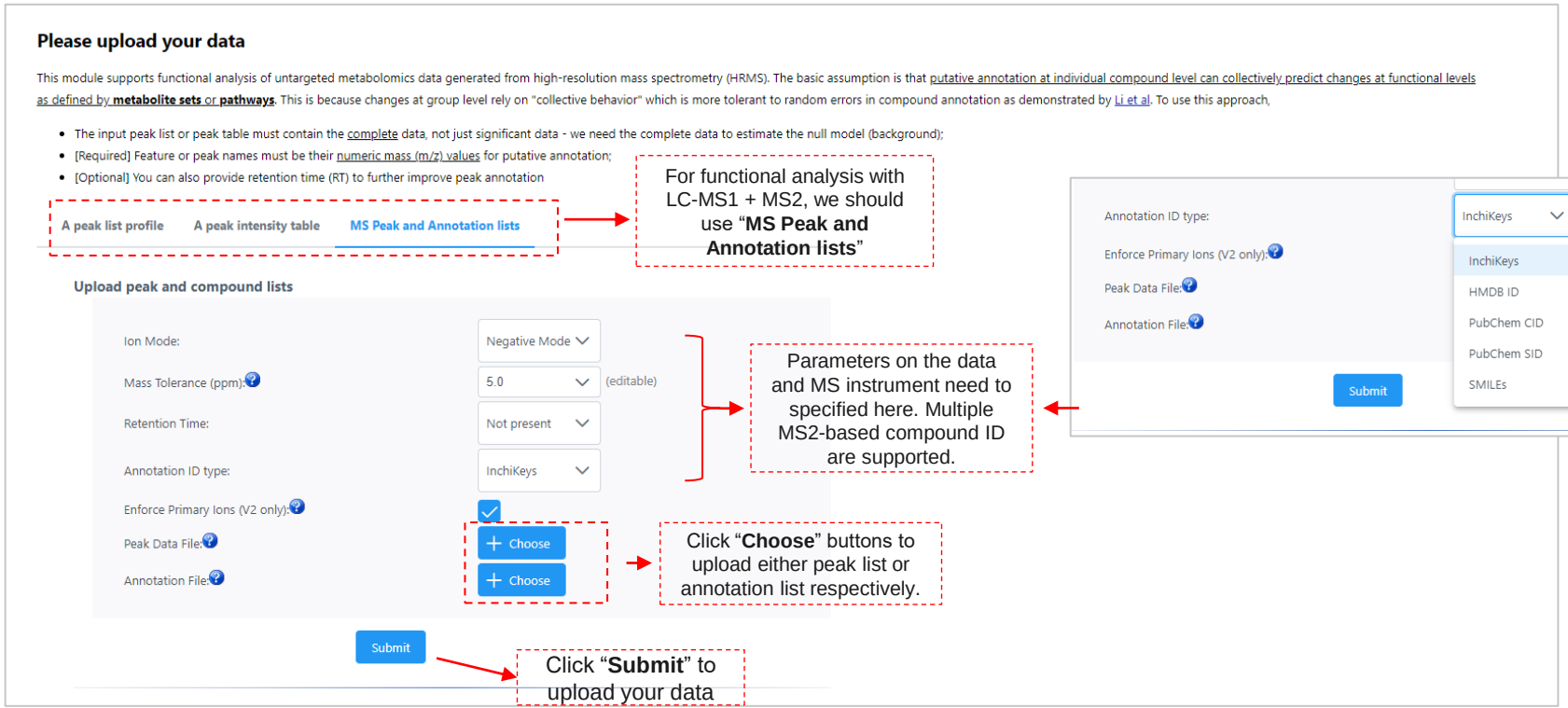
LC-MS1 peak list

	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-U	SECXISVLQFMRJM-U	NA	NA
9	XUWHAWMETYGRKE	NYEZZYQZRQDLEH-U	SECXISVLQFMRJM-U	NA	NA
10	XUWHAWMETYGRKE	NYEZZYQZRQDLEH-U	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-U	KFDVPJUYSDEJTH-U	KGIGUEBEKRSTEW-U	YAXKTBLXMTYWDQ-	NA
21	NUVWVUPJCXRIIW-U	KFDVPJUYSDEJTH-U	KGIGUEBEKRSTEW-U	YAXKTBLXMTYWDQ-	NA

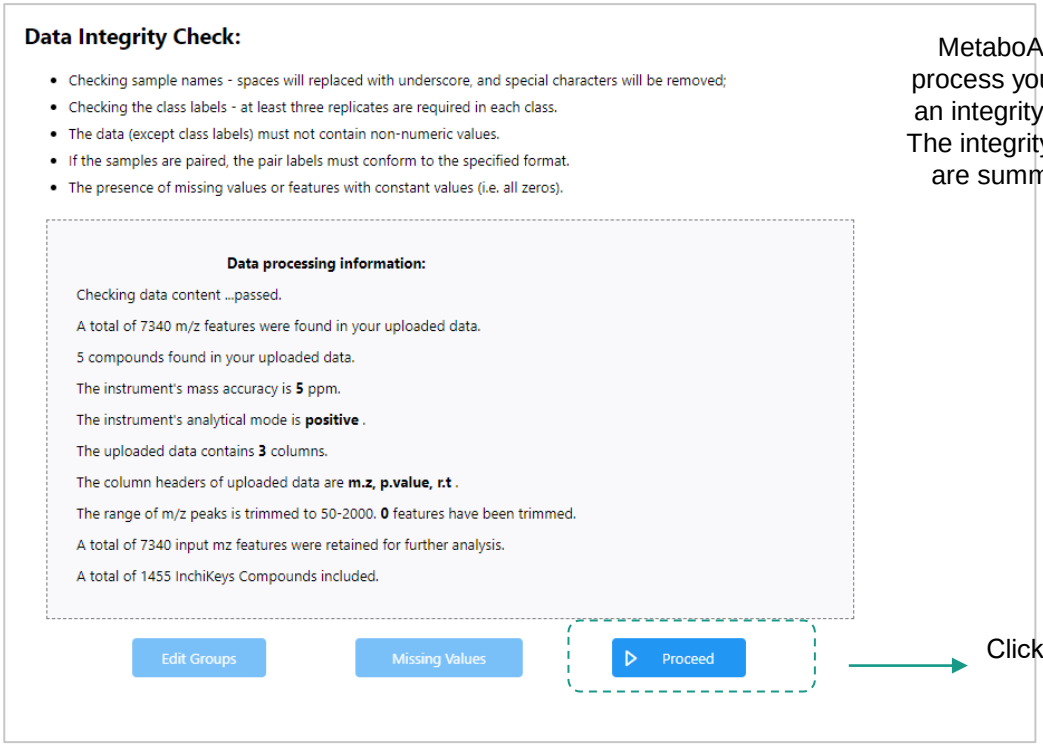
LC-MS2-based compound identification results list

4. Functional Analysis with LC-MS1 Peaks + MS2-based annotation

4.1 Data uploading



4.2 Integrity Check

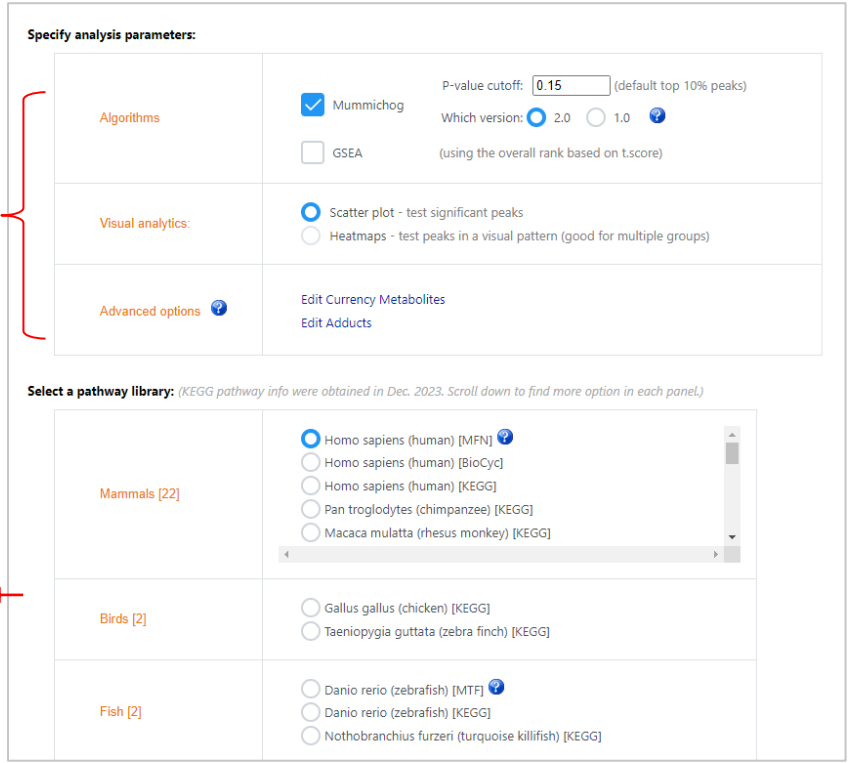


MetaboAnalyst could process your data and do an integrity check at first. The integrity check results are summarized here.

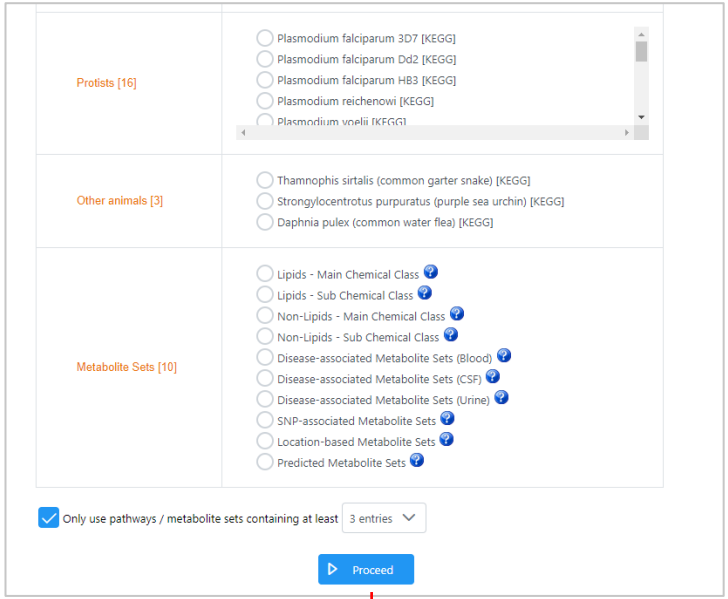
Click “***Proceed***” button to continue.

4.3 Parameter Setting

Multiple parameters can be customized for functional analysis. Click the question mark to read more details on each option.



Over 140 pathway libraries, supporting over 130 species have been provided here. Please select one based on your case.



Click “**Proceed**” to start analysis

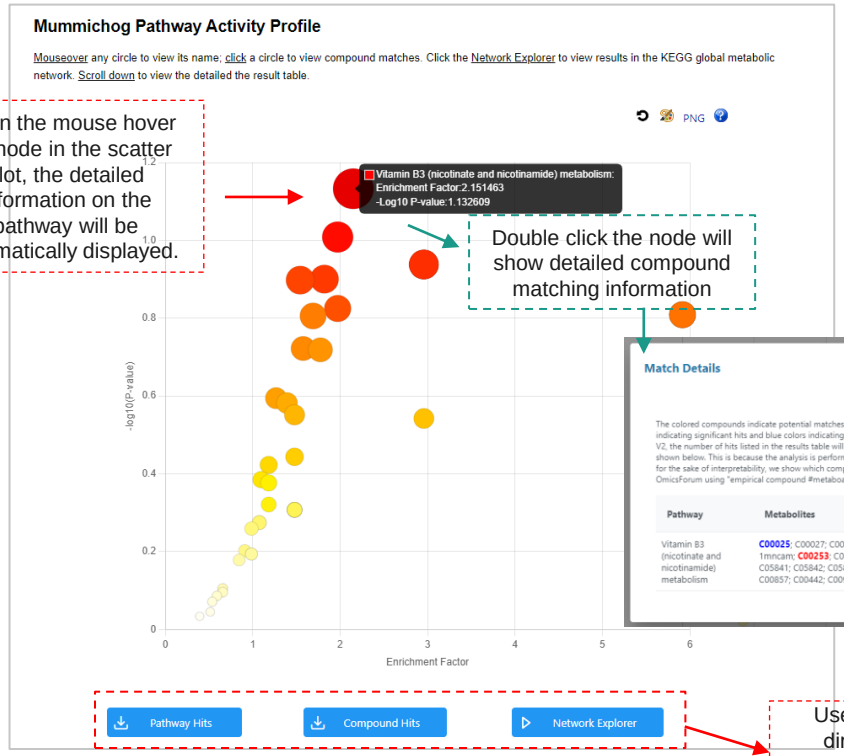
4.4 Result Exploration – scatter plot

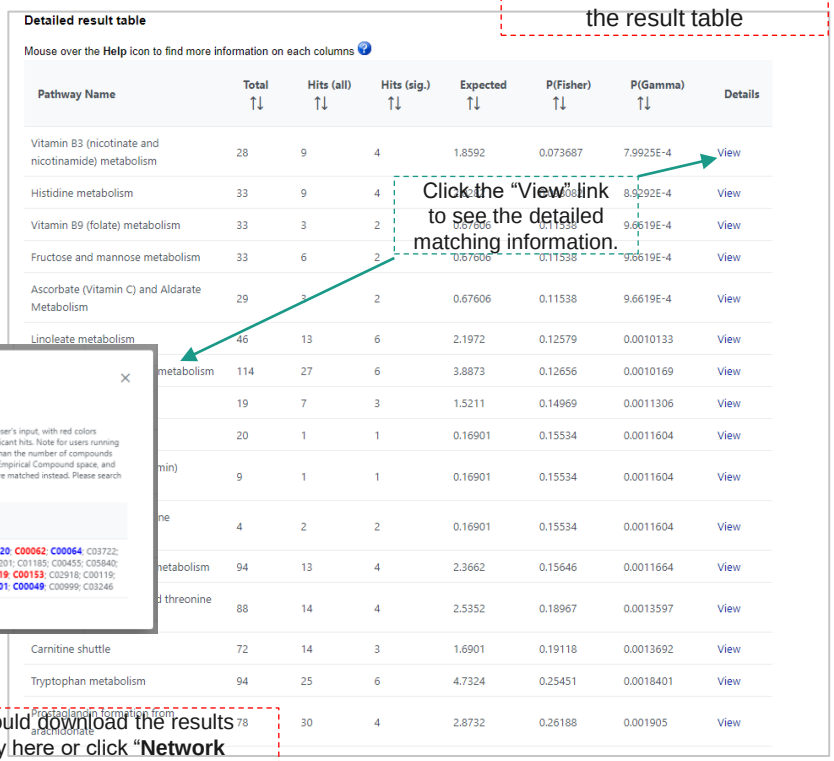
All pathway enrichment results are summarized in the result table

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)

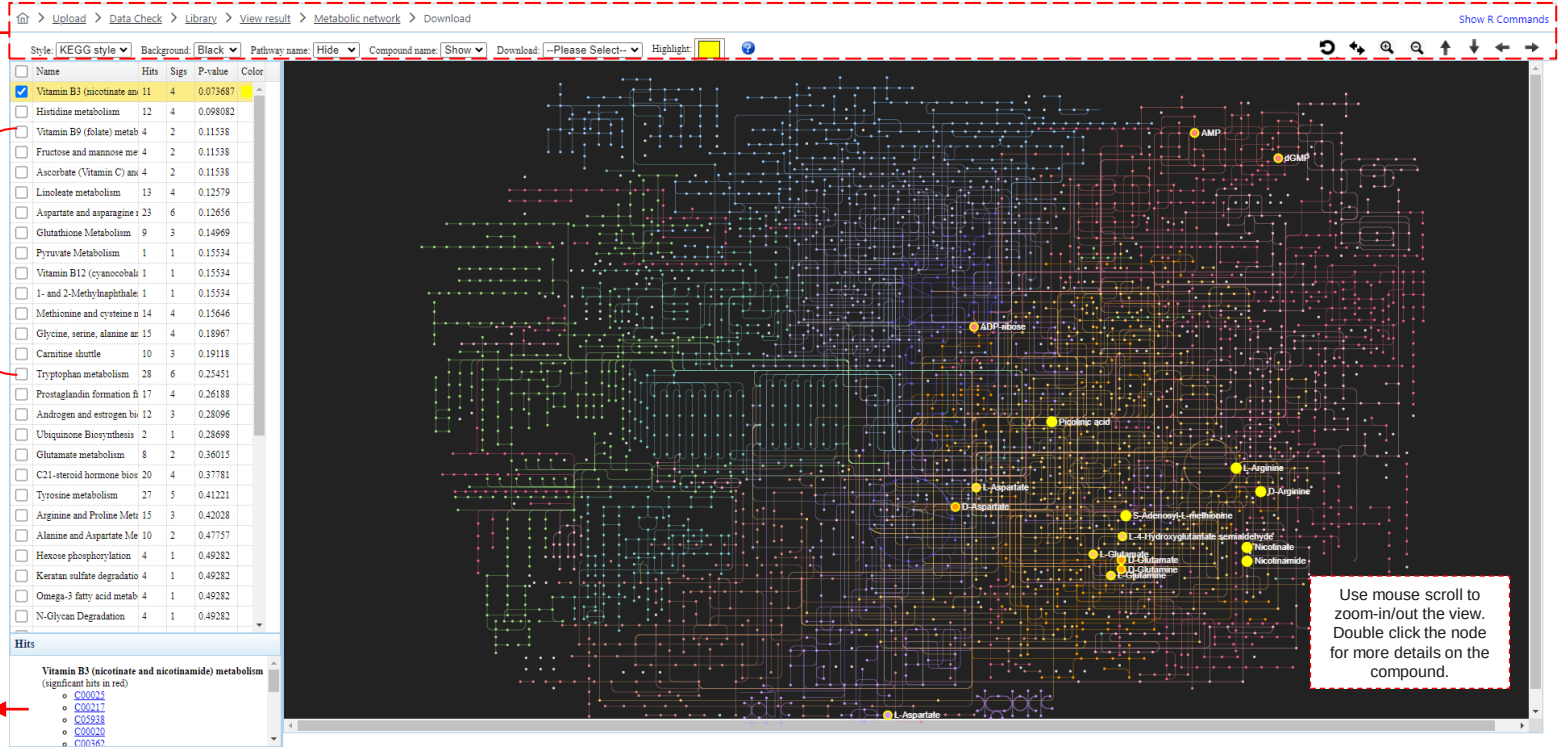
4.4 Result Exploration – global network view

All pathway enrichment results can be viewed from global metabolic network explorer, like below.

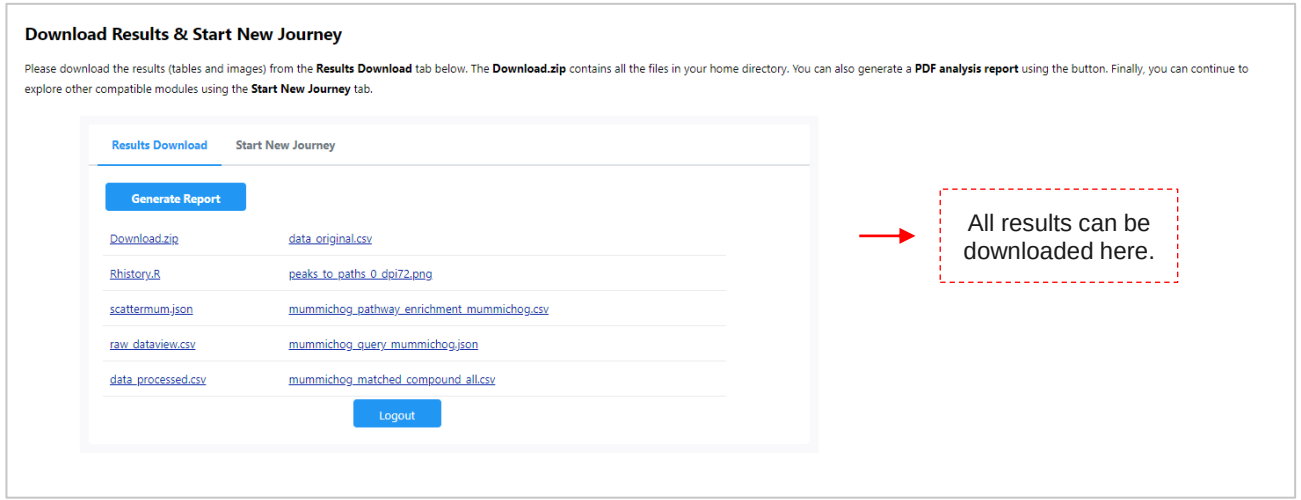
Use these functionalities to customize the network view.

All pathways are listed here. By selecting one or multiple pathways, the matches in the global metabolic network will be displayed and labelled with colors directly.

Compound hits are listed here



5. Download Results



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

- User could provide MS1 features for functional prediction or optionally together with MS2-based to remove the impractical compounds to improve the accuracy.
- MS2-based compounds list can be formatted in two formats.
- MS2-based compound identification results can be from DDA or DIA.
- The parameters setting page offers over 130 pathway libraries which basically covered all common model and non-model organisms.
- Users can interactively explore the results from pathway levels to underlying individual compounds