



MetaboAnalyst 6.0

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

Functional Analysis [LC-MS]

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

- ❑ 1. Introduction;
- ❑ 2. Choose the Module;
- ❑ 3. Data Preparation
- ❑ 4. Functional Analysis by using LC-MS1+MS2
 - 4.1 Data uploading
 - 4.2 Integrity Check
 - 4.3 Parameter Setting
 - 4.4 Result Exploration
- ❑ 5. Download Results.

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

MetaboAnalyst 6.0 - from raw spectra to biomarkers, patterns, functions and systems biology

Module Overview

Input Data Type	Available Modules (click on a module to proceed, or scroll down to explore a total of 18 modules including utilities)				
LC-MS Spectra (mzML, mzXML, or mzData)			Spectra Processing [LC-MS1 w/wo MS2]		
MS Peaks (peak list or intensity table)		Peak Annotation [MS2-DDA/DIA]	Functional Analysis [LC-MS]	Functional Meta-analysis [LC-MS1]	
Generic Format (csv or txt table files)	Statistical Analysis [one factor]	Statistical Analysis [metadata table]	Biomarker Analysis	Statistical Meta-analysis	Dose Response Analysis
Annotated Features (metabolite list or table)		Enrichment Analysis	Pathway Analysis	Network Analysis	
Link to Genomics & Phenotypes (metabolite list)			Causal Analysis [Mendelian randomization]		

>> Spectral Processing [LC-MS1 w/wo MS2]

This module allows users to upload raw LC-MS spectra (mzML, mzXML, or mzData) to be processed using our optimized workflow based on MetaboAnalystR 4.0 or the

>> Peak Annotation [MS2-DIA/DAA]

This module performs MS2 peak annotation based on a comprehensive list of public databases. Users can either directly enter a two-column peak list containing m/z and

>> Functional Analysis [LC-MS]

This module accepts high-resolution LC-MS spectral peak data to perform metabolic pathway enrichment analysis and visual exploration based on the [mummichog](#) or [GSEA](#)

3. Data Preparation – format 1

Two files need to be prepared:

- 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.
- 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	E	F	G	H
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	lnchKey 1	lnchKey 2	lnchKey 3	lnchKey 4	lnchKey 5		
2	2	MGSKVZWGBWPBTFTZJAEGLMLTGRJ-UQVRNDROMDRJTHS	NA			NA		
3	8	FWULQXY/OANGSS-JDVGAQPNXQDWBJTZATWAPRKXBV-U	LIWMQSWFLXEGMAVEQOALNAAJBPNY-UHFFFAOYSA-N					
4	10	LUALIOATIOESLM-U	NA		NA	NA		
5	42	WUYZYBFCWCCQJ-HGSOUPIFSDBJ-O	NA		NA	NA		
6	44	HOVAGTYPDGVJG-RYYVLZVUVJVGH-UI	NA		NA	NA		
7	56	FQZPXSRKCOWUEI-YRQOXUDKDCXMEYRQOXUDKDCXME	OTGQIQTPXJQRG-NA					
8	68	RYYVLZVUVJVGH-UIPXQPEWDEAKTCGB	KSEBMYQBYZTDHS-I	QURCVMIKCOAJU-NA				
9	70	HJMQDJPMQIHLBP-IAMDPNCEWKZEBQIRZVHDLBAYNPCT-U	NA		NA	NA		
10	74	YEJYLHKQOBOSCP-I	OKJCFMUGMSVJBG-NP	JICTMALKLTFW-C	NA	NA		
11	94	XJLSEXAGTICILF-UHJWYOAMQZLXDER-NILQLFBWTXNUOE	PJDFLNIOAUZSL-UH	DDSLGZOYEKPKSJ-UHFFFAOYSA-N				
12	112	DKLKMKYDWHYZTD-SSEBTPPFLLCUMN-I	MTFCPNHRBINLRQ-IGLHHSKNBDXCEY-	NA				
13	123	IEPGNWMPIDNSD-IEPGNWMPIDNSD-IEPGNWMPIDNSD	WECGLUPZRHILCT-NA					
14	129	HXFOXFJUNFFYMO-INA	NA		NA	NA		
15	141	OTCCIMWXLJLIA-B	NXQJVBMMRCKQC	NRNCYVBFDDJNE-POJWUDADGALRAB-NA				
16	160	IZYCZPAFZQFMCQ-L	NVEPPWDLBMNMEGHOKWGTUJZEAQD	WTDRLQBEARUVNC	JXXCENBLGFBQJM-UHFFFAOYSA-N			
17	161	LUINDDOUWHRIPW-DFQOXFIPAAFAU-I	NA		NA	NA		
18	186	VLSMHEGGTFMBBZ-NA	NA		NA	NA		
19	198	IRZTUXPRIUXMP-UI	RTIXKCRFFJGDFG-U	LCAWNFIIMLXZPQ-INA				
20	205	XUQWWIFROYJHCU-SEKYBDYVXDAPY-A	SEKYBDYVXDAPY-U	IDTCGADOYRIRKB-U	NA			
21	206	CMRNMZJAUFQXQF-FMMOOAVCKXGMF	XSXIVVZCUAHUJO-F	XSXIVVZCUAHUJO-L	XUJWOMMOEOHPPF-UTJQPWESSA-N			

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-I	SECXISVLQFMRJM-U	NA	NA
9	XUWHAWMETYGRKE	NYEZZYQZRQDLEH-I	SECXISVLQFMRJM-U	NA	NA
10	XUWHAWMETYGRKE	NYEZZYQZRQDLEH-I	SECXISVLQFMRJM-U	NA	NA
11	PAFZNILMEXTMIY-UI	NA	NA	NA	NA
12	PAFZNILMEXTMIY-UI	NA	NA	NA	NA
13	NA	NA	NA	NA	NA
14	DYDCUQKUCUHJBH	KYCJNUIHWNJNCT-	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	NUVWWUPJCXRIIW-I	KFDVPJUYSDEJTH-U	KGIGUEBEKRSTEW-I	YAXKTBLXMTYWDQ-	NA
21	NUVWWUPJCXRIIW-I	KFDVPJUYSDEJTH-U	KGIGUEBEKRSTEW-I	YAXKTBLXMTYWDQ-	NA

LC-MS2-based compound identification results list



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

4.1 Data uploading

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

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

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

Submit

Click "Submit" to upload your data

Annotation ID type:

InchiKeys

Enforce Primary Ions (V2 only):

Peak Data File:

Annotation File:

InchiKeys

HMDB ID

PubChem CID

PubChem SID

SMILES

Submit

4.2 Integrity Check

Data Integrity Check:

- Checking sample names - spaces will be 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.

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.

Specify analysis parameters:

Algorithms	<input checked="" type="checkbox"/> Mummichog <input type="checkbox"/> GSEA	P-value cutoff: <input type="text" value="0.15"/> (default top 10% peaks) Which version: <input checked="" type="radio"/> 2.0 <input type="radio"/> 1.0 ? (using the overall rank based on t.score)
Visual analytics:	<input checked="" type="radio"/> Scatter plot - test significant peaks <input type="radio"/> Heatmaps - test peaks in a visual pattern (good for multiple groups)	
Advanced options ?	Edit Currency Metabolites Edit Adducts	

Select a pathway library: (KEGG pathway info were obtained in Dec. 2023. Scroll down to find more option in each panel.)

Mammals [22]	<input checked="" type="radio"/> Homo sapiens (human) [MFN] ? <input type="radio"/> Homo sapiens (human) [BioCyc] <input type="radio"/> Homo sapiens (human) [KEGG] <input type="radio"/> Pan troglodytes (chimpanzee) [KEGG] <input type="radio"/> Macaca mulatta (rhesus monkey) [KEGG]
Birds [2]	<input type="radio"/> Gallus gallus (chicken) [KEGG] <input type="radio"/> Taeniopygia guttata (zebra finch) [KEGG]
Fish [2]	<input type="radio"/> Danio rerio (zebrafish) [MTF] ? <input type="radio"/> Danio rerio (zebrafish) [KEGG] <input type="radio"/> Nothobranchius furzeri (turquoise killifish) [KEGG]

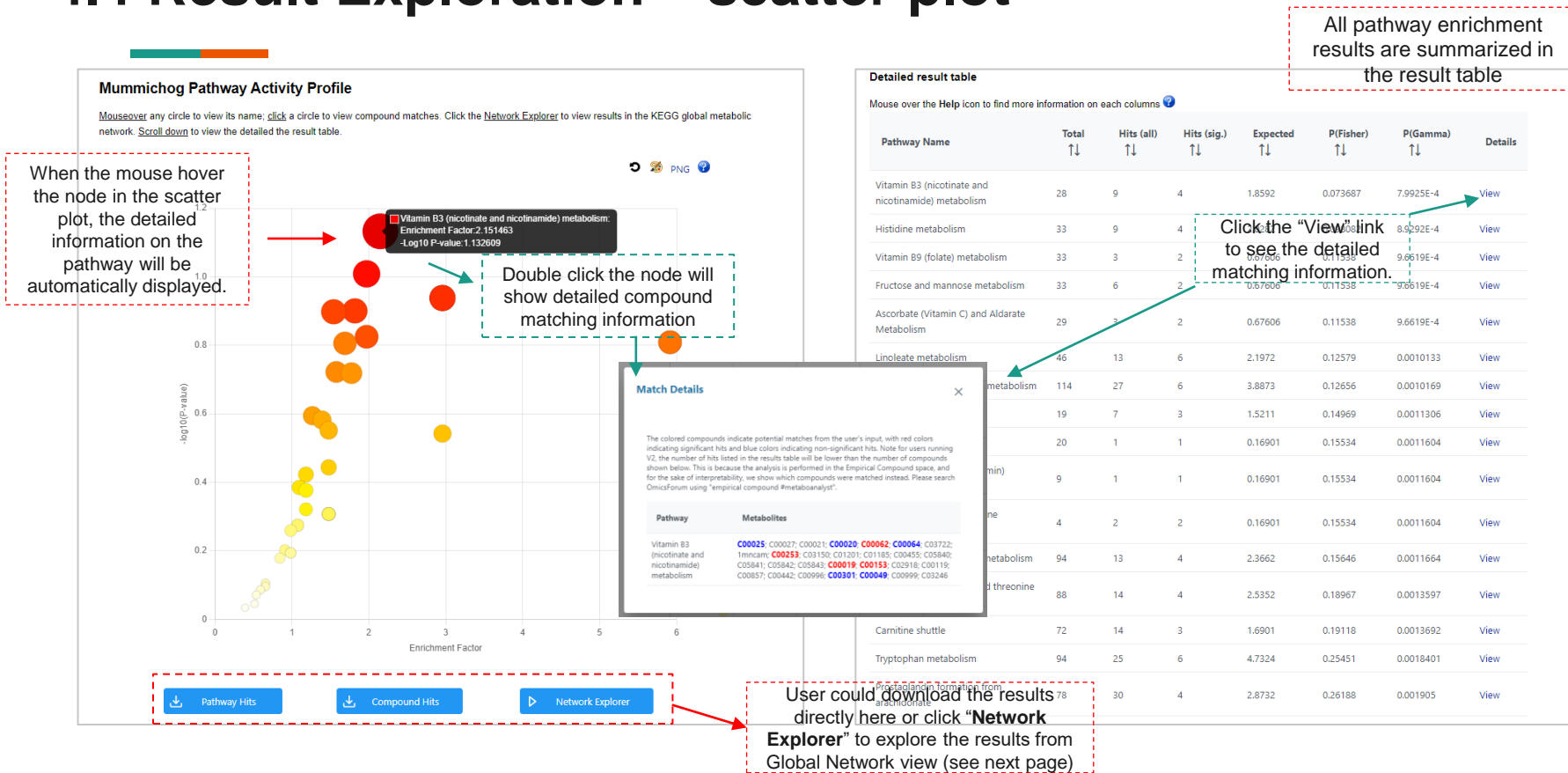
Protists [16]	<input type="radio"/> Plasmodium falciparum 3D7 [KEGG] <input type="radio"/> Plasmodium falciparum Dd2 [KEGG] <input type="radio"/> Plasmodium falciparum HB3 [KEGG] <input type="radio"/> Plasmodium reichenowi [KEGG] <input type="radio"/> Plasmodium vivax [KEGG]
Other animals [3]	<input type="radio"/> Thamnophis sirtalis (common garter snake) [KEGG] <input type="radio"/> Strongylocentrotus purpuratus (purple sea urchin) [KEGG] <input type="radio"/> Daphnia pulex (common water flea) [KEGG]
Metabolite Sets [10]	<input type="radio"/> Lipids - Main Chemical Class ? <input type="radio"/> Lipids - Sub Chemical Class ? <input type="radio"/> Non-Lipids - Main Chemical Class ? <input type="radio"/> Non-Lipids - Sub Chemical Class ? <input type="radio"/> Disease-associated Metabolite Sets (Blood) ? <input type="radio"/> Disease-associated Metabolite Sets (CSF) ? <input type="radio"/> Disease-associated Metabolite Sets (Urine) ? <input type="radio"/> SNP-associated Metabolite Sets ? <input type="radio"/> Location-based Metabolite Sets ? <input type="radio"/> Predicted Metabolite Sets ?

☒ Only use pathways / metabolite sets containing at least entries

[Proceed](#)

Click "**Proceed**" to start analysis

4.4 Result Exploration – scatter plot



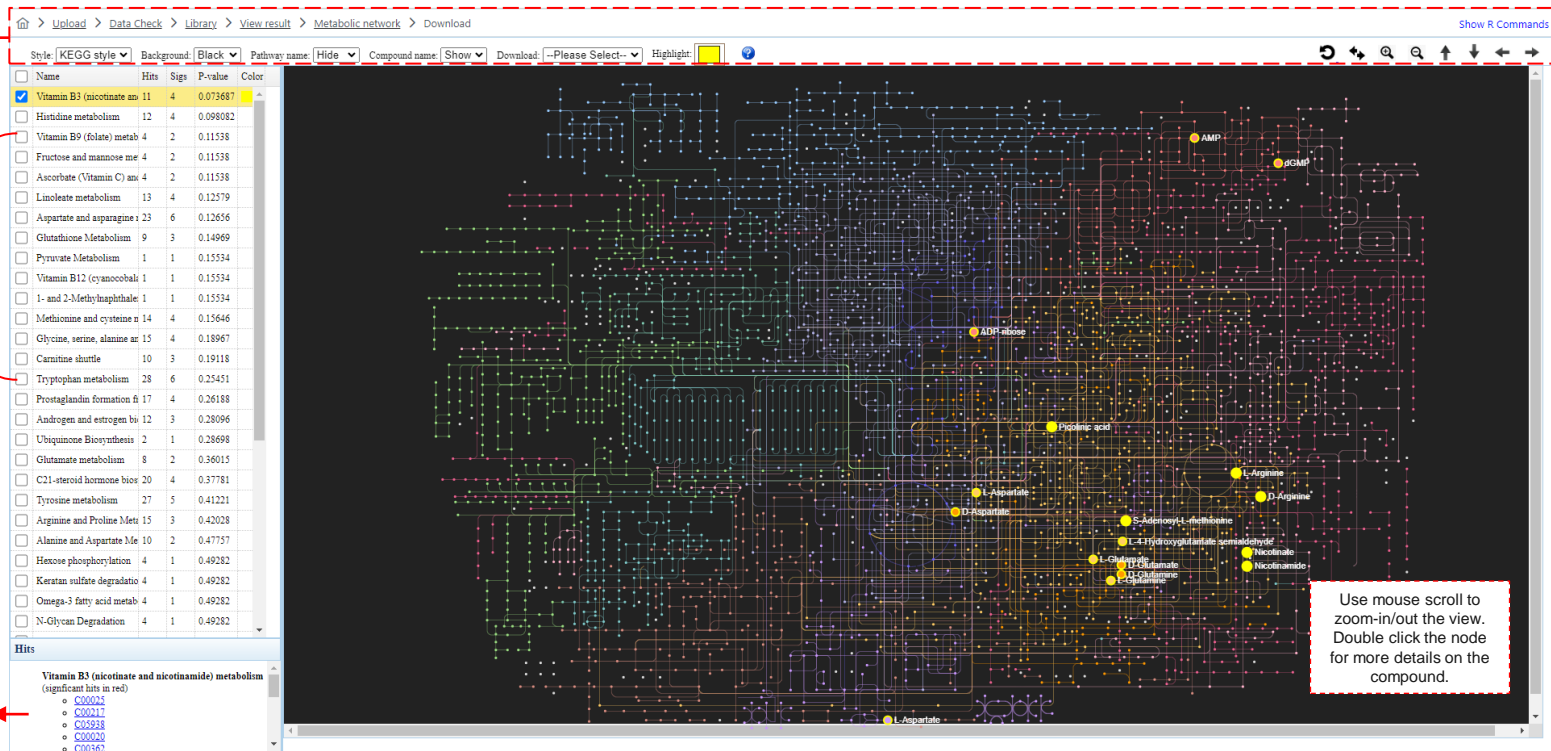
4.4 Result Exploration – global network view

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



Download Results & Start New Journey

Please download the results (tables and images) from the **Results Download** tab below. The **Download.zip** contains all the files in your home directory. You can also generate a **PDF analysis report** using the button. Finally, you can continue to explore other compatible modules using the **Start New Journey** tab.

[Results Download](#) [Start New Journey](#)

[Generate Report](#)

Download.zip	data_original.csv
Rhistory.R	peaks_to_paths_0_dpi72.png
scattermum.json	mummichog_pathway_enrichment_mummichog.csv
raw_dataview.csv	mummichog_query_mummichog.json
data_processed.csv	mummichog_matched_compound_all.csv

[Logout](#)



All results can be
downloaded here.



In summary

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

Or contact us:

zhiqiang.pang@xialab.ca

jeff.xia@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