Bioinformatics analysis for metabolomics dataset

Nattapon Simanon

PMRT, NOC, BIOTEC

2026-02-20

Outline

- Metabolomics data
- Identifier coversion
- Bioinformatics analysis for compound/ metabolite data
  - Enrichment analysis
  - Pathway analysis
  - Network analysis or compound interaction analysis

Metabolomics data

- What is metabolomics data?
- Analytical platforms:
  - MS data vs NMR data
- Targeted – untargeted metabolites:
  - Qualitative vs quantitative analysis
- Softwares

Personal

metaboanalyst.ca

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

Home
Data Formats
Tutorials
User Forum
MetaboAnalystR
Publications
Update History
Databases
User Stats
Data Policy
About
Contact

News & Updates

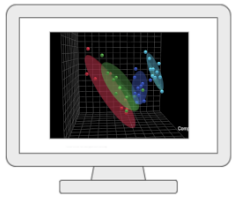
- NEW To help sustain our platform, we now offer self-paced Omics Data Science training coupled with AI assistant, Omics Data Science book and Pro version running on cloud or your server.
- Please use OmicsForum for discussion and troubleshooting requests so that more people can benefit (invitation code: xialab_omics!).
- Code refactoring for improved performance (01/23/2026);
- Enhanced PCA analysis to deal with large data and fixed issue with name conflict (01/08/2026);
- Upgraded R to the latest version 4.5.2 "[Not] Part in a Rumble" (12/15/2025);
- Added two metabolite set libraries to help identify enriched pathways in metabolomics studies of human gut microbiome (**Enrichment Analysis** module) (12/01/2025);
- Enrichment analysis now accepts input with both metabolites and common lipids names (11/19/2025);
- Added a new metabolite set library containing 817 entries for lipidomics functional analysis (**Enrichment Analysis** module) (11/18/2025);
- Updated RaMP-DB library to v3.0.7, featuring more complete metabolic and lipid pathways (11/18/2025);
- Fixed the issue with color and shape customization (09/25/2025);
- Enhanced joint pathway analysis based on user feedback (09/20/2025);

Read more .....

Click here to start

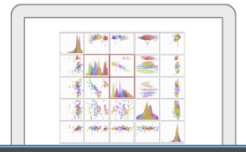
Overview

MetaboAnalyst is a web-based platform dedicated for comprehensive metabolomics data analysis, interpretation and integration with other omics data. Over the past decade, MetaboAnalyst has evolved from statistical and functional analysis for targeted metabolomics data, towards more streamlined analysis for both quantitative and untargeted metabolomics data. In addition to many feature enhancements, the version 6.0 contains three new modules - tandem MS spectral processing and **compound annotation**, dose response analysis for **chemical risk assessment**, and leveraging metabolite-genome wide association analysis and Mendelian randomization for **causal analysis**.



Statistical Analysis [single factor]

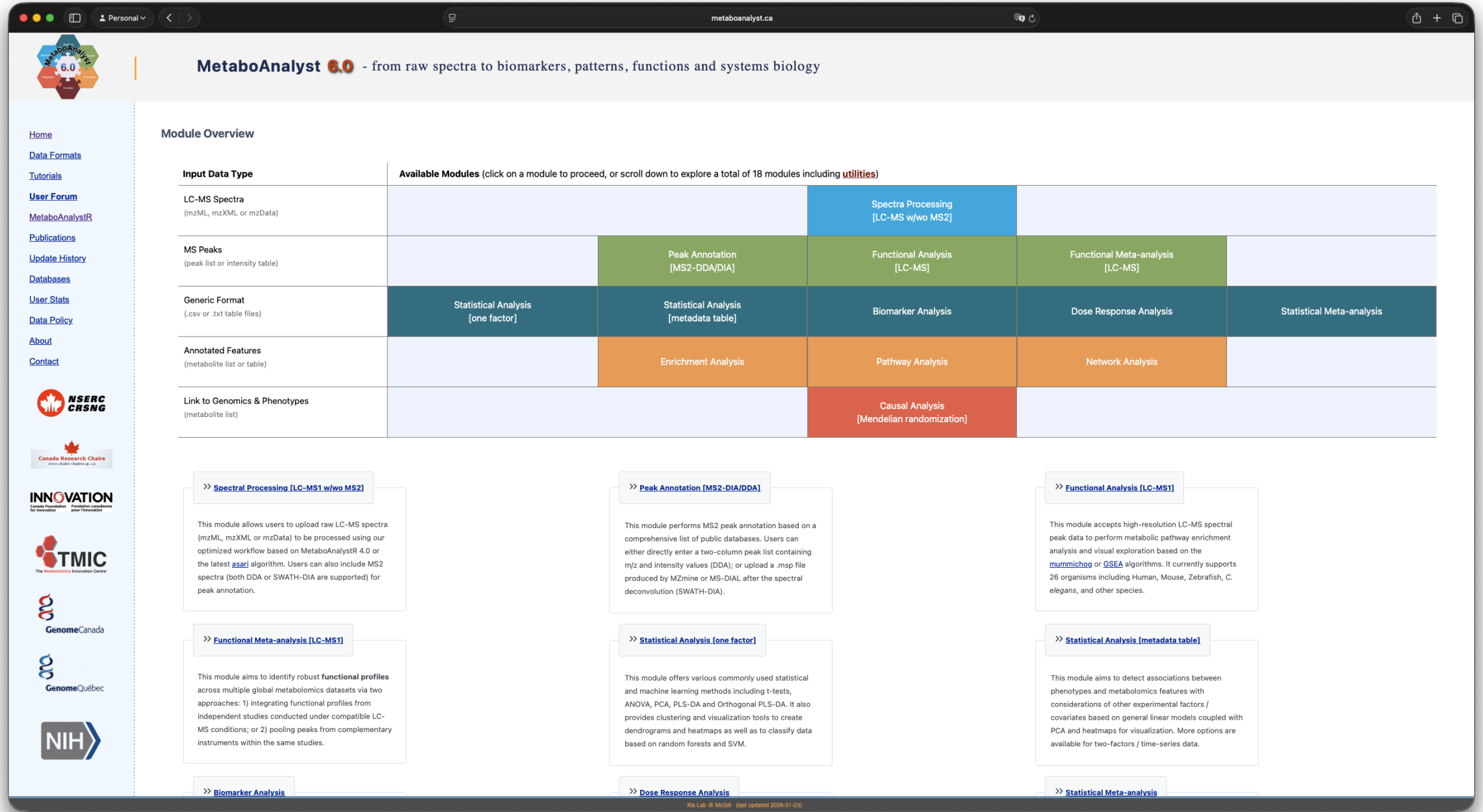
The module provides a wide array of commonly used statistical and machine learning methods: traditional univariate methods - fold change, t-test, volcano plot, ANOVA, correlation analysis; or more advanced methods designed for significance analysis of high-dimensional data, and empirical Bayesian analysis; multivariate statistics - principal component analysis (PCA), partial least squares-discriminant analysis (PLS-DA) or orthogonal partial least squares-discriminant analysis (OPLS-DA); clustering - dendrogram, heatmap, K-means, and self organizing map (SOM); as well as supervised classification - random forests and support vector machine (SVM).

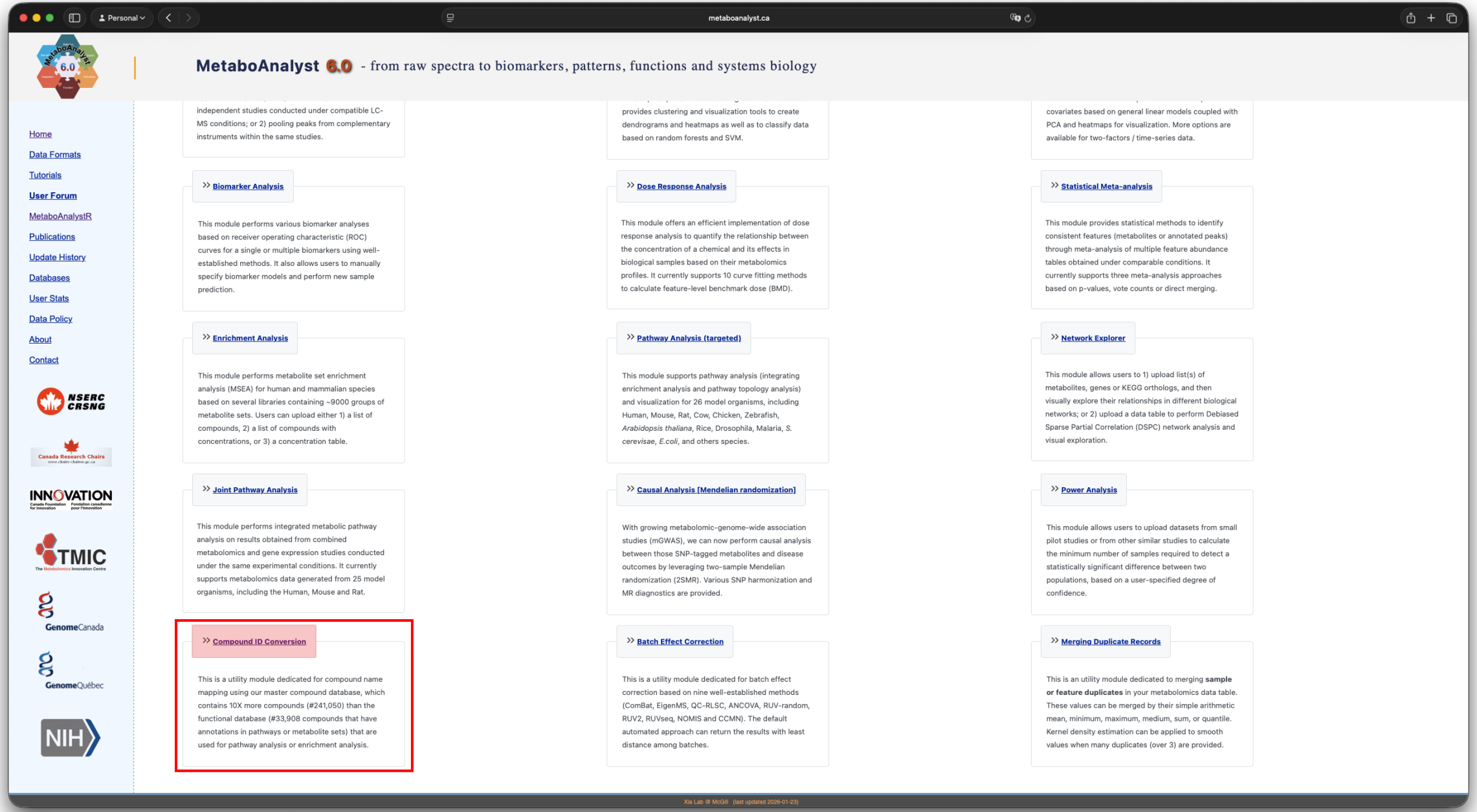


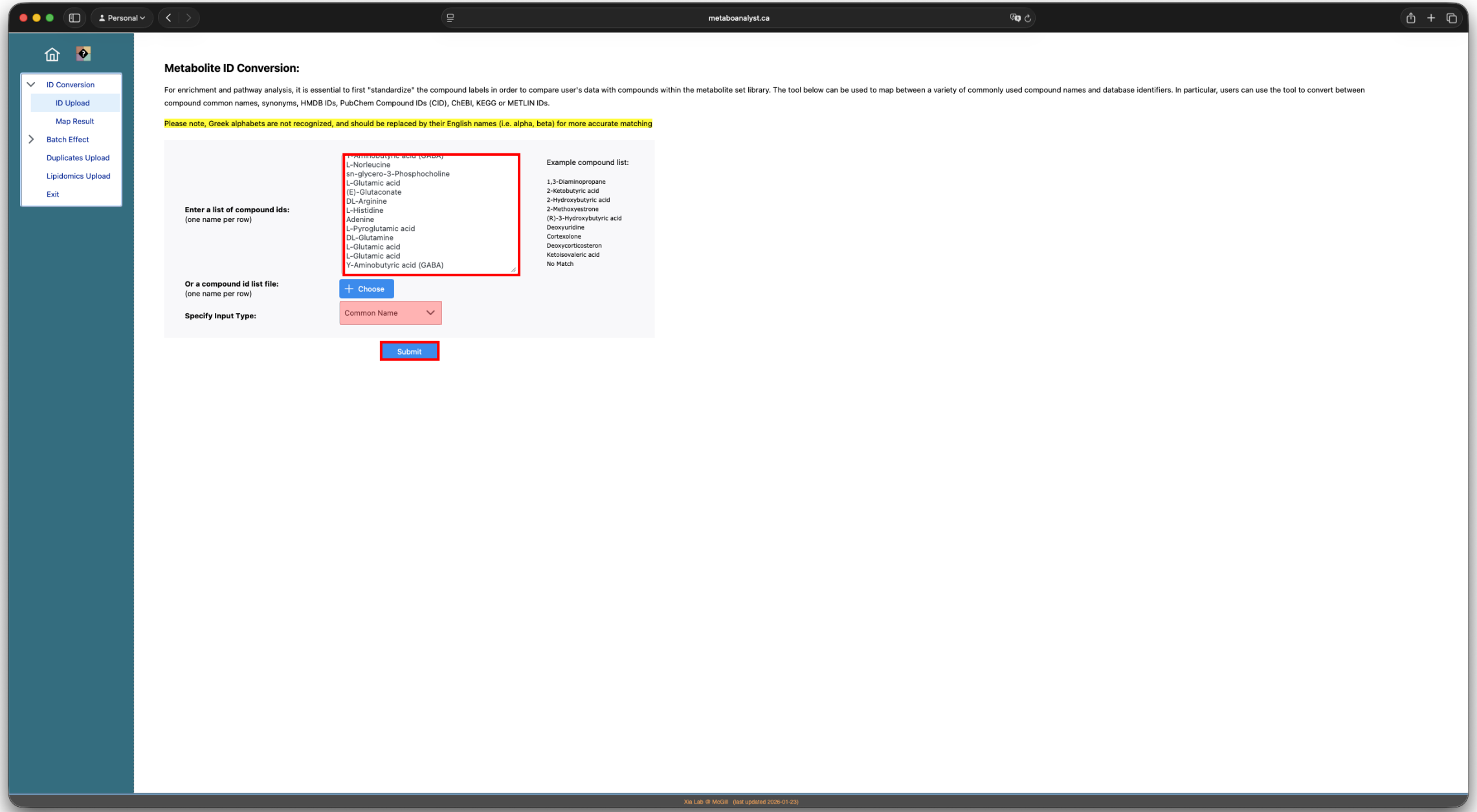
Statistical Analysis [metadata table]

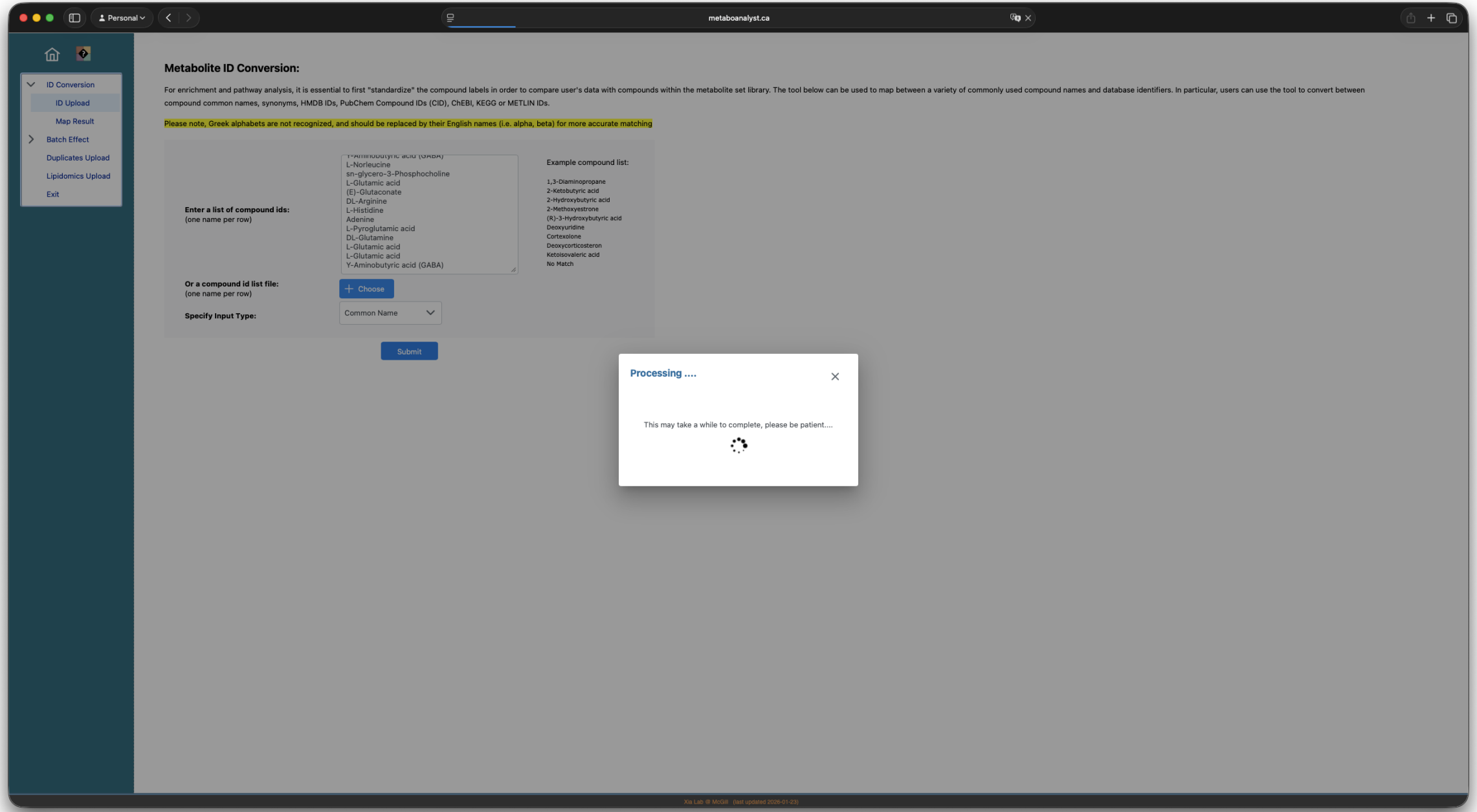
MetaboAnalyst now allows users to visualize and compute associations between phenotypes and metabolomics features with considerations of other experimental factors / covariates. It employs general linear models to accommodate modern epidemiological study, together with PCA and heatmaps for visual explorations. For two-factors / time-series data, users have more options including two-way ANOVA, multivariate empirical Bayes time-series analysis (MEBA), and ANOVA-simultaneous component analysis (ASCA).

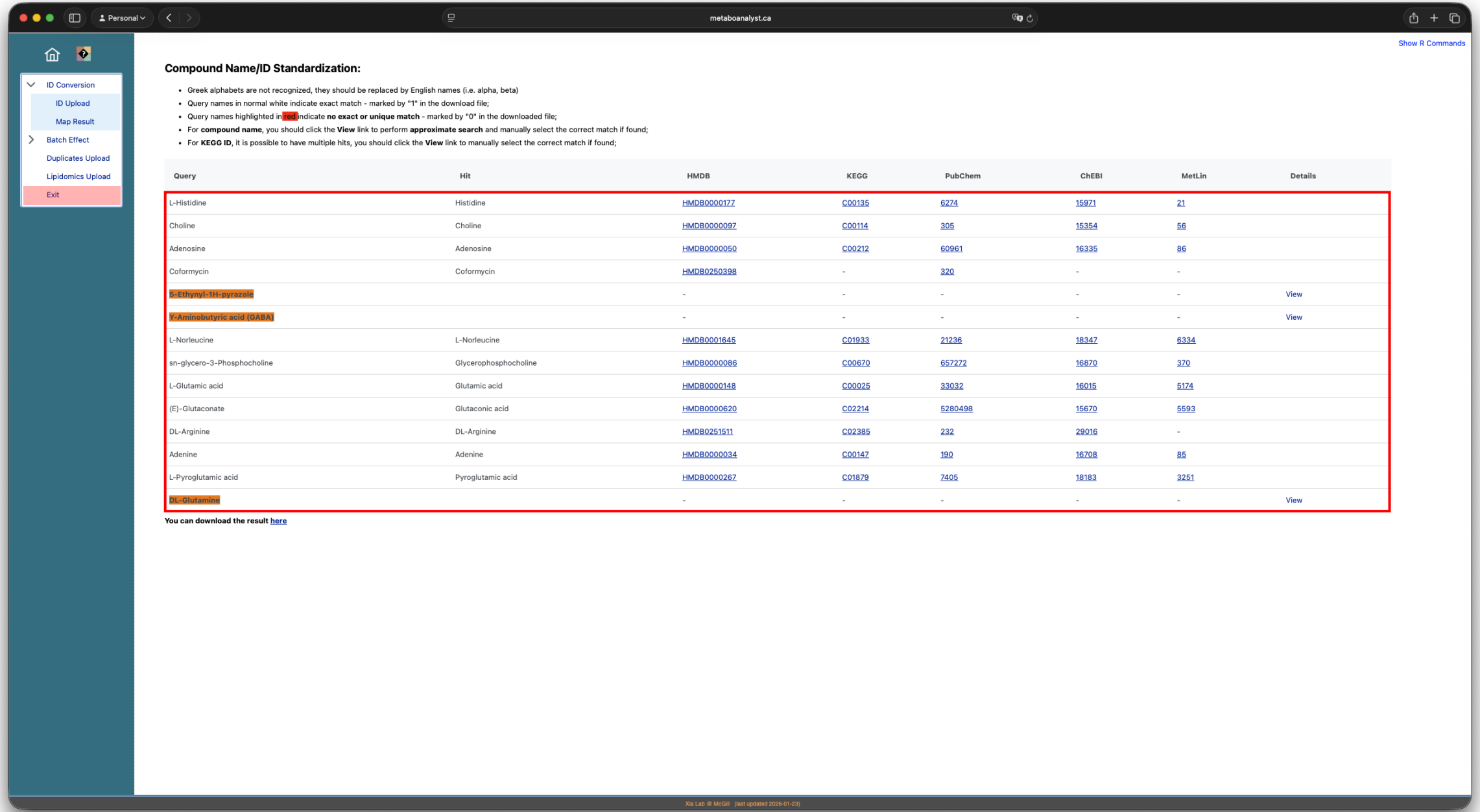
Xia Lab @ McGill (last updated 2026-01-23)

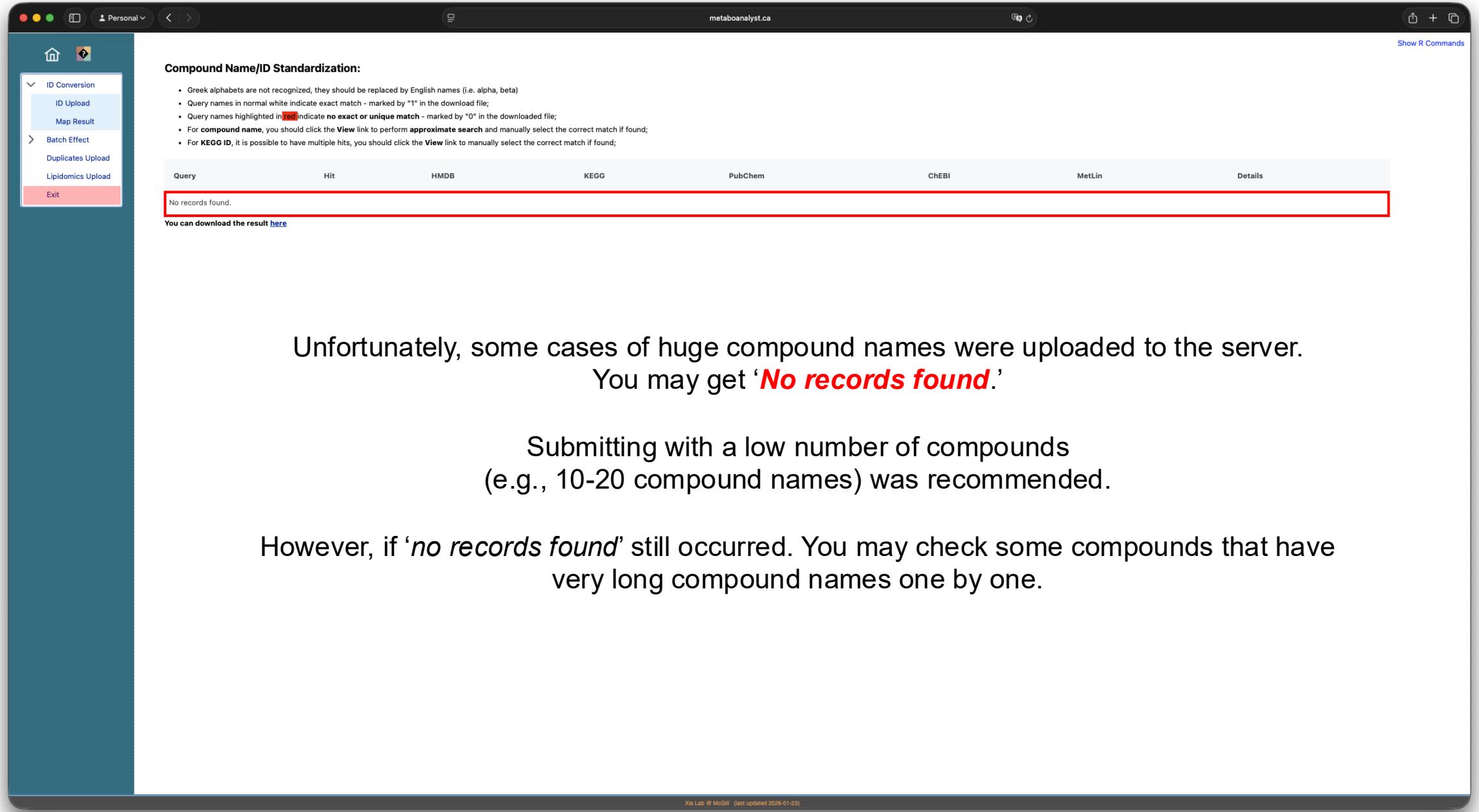

metaboanalyst.ca
MetaboAnalyst 6.0 - from raw spectra to biomarkers, patterns, functions and systems biology
Home
Data Formats
Tutorials
User Forum
MetaboAnalystR
Publications
Update History
Databases
User Stats
Data Policy
About
Contact
Module Overview
Input Data Type
Available Modules (click on a module to proceed, or scroll down to explore a total of 18 modules including utilities)
LC-MS Spectra
(mzML, mzXML or mzData)
MS Peaks
(peak list or intensity table)
Generic Format
(.csv or .txt table files)
Annotated Features
(metabolite list or table)
Link to Genomics & Phenotypes
(metabolite list)
Spectra Processing [LC-MS w/wo MS2]
Peak Annotation [MS2-DDA/DIA]
Functional Analysis [LC-MS]
Functional Meta-analysis [LC-MS]
Statistical Analysis [one factor]
Statistical Analysis [metadata table]
Biomarker Analysis
Dose Response Analysis
Statistical Meta-analysis
Enrichment Analysis
Pathway Analysis
Network Analysis
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 latest asari algorithm. Users can also include MS2 spectra (both DDA or SWATH-DIA are supported) for peak annotation.
Peak Annotation [MS2-DIA/DDA]
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 intensity values (DDA); or upload a .msp file produced by MZmine or MS-DIAL after the spectral deconvolution (SWATH-DIA).
Functional Analysis [LC-MS1]
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 algorithms. It currently supports 26 organisms including Human, Mouse, Zebrafish, C. elegans, and other species.
Functional Meta-analysis [LC-MS1]
This module aims to identify robust functional profiles across multiple global metabolomics datasets via two approaches: 1) integrating functional profiles from independent studies conducted under compatible LC-MS conditions; or 2) pooling peaks from complementary instruments within the same studies.
Statistical Analysis [one factor]
This module offers various commonly used statistical and machine learning methods including t-tests, ANOVA, PCA, PLS-DA and Orthogonal PLS-DA. It also provides clustering and visualization tools to create dendrograms and heatmaps as well as to classify data based on random forests and SVM.
Statistical Analysis [metadata table]
This module aims to detect associations between phenotypes and metabolomics features with considerations of other experimental factors / covariates based on general linear models coupled with PCA and heatmaps for visualization. More options are available for two-factors / time-series data.
Biomarker Analysis
Dose Response Analysis
Statistical Meta-analysis
Xia Lab @ McGill (last updated 2026-01-23)


metaboanalyst.ca
MetaboAnalyst 6.0 - from raw spectra to biomarkers, patterns, functions and systems biology
Home
Data Formats
Tutorials
User Forum
MetaboAnalystR
Publications
Update History
Databases
User Stats
Data Policy
About
Contact
independent studies conducted under compatible LC-MS conditions; or 2) pooling peaks from complementary instruments within the same studies.
provides clustering and visualization tools to create dendrograms and heatmaps as well as to classify data based on random forests and SVM.
covariates based on general linear models coupled with PCA and heatmaps for visualization. More options are available for two-factors / time-series data.
Biomarker Analysis
This module performs various biomarker analyses based on receiver operating characteristic (ROC) curves for a single or multiple biomarkers using well-established methods. It also allows users to manually specify biomarker models and perform new sample prediction.
Dose Response Analysis
This module offers an efficient implementation of dose response analysis to quantify the relationship between the concentration of a chemical and its effects in biological samples based on their metabolomics profiles. It currently supports 10 curve fitting methods to calculate feature-level benchmark dose (BMD).
Statistical Meta-analysis
This module provides statistical methods to identify consistent features (metabolites or annotated peaks) through meta-analysis of multiple feature abundance tables obtained under comparable conditions. It currently supports three meta-analysis approaches based on p-values, vote counts or direct merging.
Enrichment Analysis
This module performs metabolite set enrichment analysis (MSEA) for human and mammalian species based on several libraries containing ~9000 groups of metabolite sets. Users can upload either 1) a list of compounds, 2) a list of compounds with concentrations, or 3) a concentration table.
Pathway Analysis (targeted)
This module supports pathway analysis (integrating enrichment analysis and pathway topology analysis) and visualization for 26 model organisms, including Human, Mouse, Rat, Cow, Chicken, Zebrafish, Arabidopsis thaliana, Rice, Drosophila, Malaria, S. cerevisae, E.coli, and others species.
Network Explorer
This module allows users to 1) upload list(s) of metabolites, genes or KEGG orthologs, and then visually explore their relationships in different biological networks; or 2) upload a data table to perform Debiased Sparse Partial Correlation (DSPC) network analysis and visual exploration.
Joint Pathway Analysis
This module performs integrated metabolic pathway analysis on results obtained from combined metabolomics and gene expression studies conducted under the same experimental conditions. It currently supports metabolomics data generated from 25 model organisms, including the Human, Mouse and Rat.
Causal Analysis [Mendelian randomization]
With growing metabolomic-genome-wide association studies (mGWAS), we can now perform causal analysis between those SNP-tagged metabolites and disease outcomes by leveraging two-sample Mendelian randomization (2SMR). Various SNP harmonization and MR diagnostics are provided.
Power Analysis
This module allows users to upload datasets from small pilot studies or from other similar studies to calculate the minimum number of samples required to detect a statistically significant difference between two populations, based on a user-specified degree of confidence.
Compound ID Conversion
This is a utility module dedicated for compound name mapping using our master compound database, which contains 10X more compounds (#241,050) than the functional database (#33,908 compounds that have annotations in pathways or metabolite sets) that are used for pathway analysis or enrichment analysis.
Batch Effect Correction
This is a utility module dedicated for batch effect correction based on nine well-established methods (ComBat, EigenMS, QC-RLSC, ANCOVA, RUV-random, RUV2, RUVseq, NOMIS and CCMN). The default automated approach can return the results with least distance among batches.
Merging Duplicate Records
This is an utility module dedicated to merging sample or feature duplicates in your metabolomics data table. These values can be merged by their simple arithmetic mean, minimum, maximum, medium, sum, or quantile. Kernel density estimation can be applied to smooth values when many duplicates (over 3) are provided.
Xia Lab @ McGill (last updated 2026-01-23)


Personal
metaboanalyst.ca
ID Conversion
ID Upload
Map Result
Batch Effect
Duplicates Upload
Lipidomics Upload
Exit
Metabolite ID Conversion:
For enrichment and pathway analysis, it is essential to first "standardize" the compound labels in order to compare user's data with compounds within the metabolite set library. The tool below can be used to map between a variety of commonly used compound names and database identifiers. In particular, users can use the tool to convert between compound common names, synonyms, HMDB IDs, PubChem Compound IDs (CID), ChEBI, KEGG or METLIN IDs.
Please note, Greek alphabets are not recognized, and should be replaced by their English names (i.e. alpha, beta) for more accurate matching
Enter a list of compound ids:
(one name per row)
L-Norleucine
sn-glycero-3-Phosphocholine
L-Glutamic acid
(E)-Glutaconate
DL-Arginine
L-Histidine
Adenine
L-Pyroglutamic acid
DL-Glutamine
L-Glutamic acid
L-Glutamic acid
Y-Aminobutyric acid (GABA)
Example compound list:
1,3-Diaminopropane
2-Ketobutyric acid
2-Hydroxybutyric acid
2-Methoxyestrone
(R)-3-Hydroxybutyric acid
Deoxyuridine
Cortexolone
Deoxycorticosteron
Ketoisovaleric acid
No Match
Or a compound id list file:
(one name per row)
Choose
Specify Input Type:
Common Name
Submit
Xia Lab @ McGill (last updated 2026-01-23)


Personal
metaboanalyst.ca
ID Conversion
ID Upload
Map Result
Batch Effect
Duplicates Upload
Lipidomics Upload
Exit
Metabolite ID Conversion:
For enrichment and pathway analysis, it is essential to first "standardize" the compound labels in order to compare user's data with compounds within the metabolite set library. The tool below can be used to map between a variety of commonly used compound names and database identifiers. In particular, users can use the tool to convert between compound common names, synonyms, HMDB IDs, PubChem Compound IDs (CID), ChEBI, KEGG or METLIN IDs.
Please note, Greek alphabets are not recognized, and should be replaced by their English names (i.e. alpha, beta) for more accurate matching
Enter a list of compound ids:
(one name per row)
L-Norleucine
sn-glycero-3-Phosphocholine
L-Glutamic acid
(E)-Glutaconate
DL-Arginine
L-Histidine
Adenine
L-Pyroglutamic acid
DL-Glutamine
L-Glutamic acid
L-Glutamic acid
Y-Aminobutyric acid (GABA)
Example compound list:
1,3-Diaminopropane
2-Ketobutyric acid
2-Hydroxybutyric acid
2-Methoxyestrone
(R)-3-Hydroxybutyric acid
Deoxyuridine
Cortexolone
Deoxycorticosteron
Ketoisovaleric acid
No Match
Or a compound id list file:
(one name per row)
Choose
Specify Input Type:
Common Name
Submit
Processing ....
This may take a while to complete, please be patient....
Xia Lab @ McGill (last updated 2026-01-23)


Personal
metaboanalyst.ca
ID Conversion
ID Upload
Map Result
Batch Effect
Duplicates Upload
Lipidomics Upload
Exit
Show R Commands
Compound Name/ID Standardization:
Greek alphabets are not recognized, they should be replaced by English names (i.e. alpha, beta)
Query names in normal white indicate exact match - marked by "1" in the download file;
Query names highlighted in red indicate no exact or unique match - marked by "0" in the downloaded file;
For compound name, you should click the View link to perform approximate search and manually select the correct match if found;
For KEGG ID, it is possible to have multiple hits, you should click the View link to manually select the correct match if found;
Query | Hit | HMDB | KEGG | PubChem | ChEBI | MetLin | Details
L-Histidine | Histidine | HMDB0000177 | C00135 | 6274 | 15971 | 21
Choline | Choline | HMDB0000097 | C00114 | 305 | 15354 | 56
Adenosine | Adenosine | HMDB0000050 | C00212 | 60961 | 16335 | 86
Coformycin | Coformycin | HMDB0250398 | - | 320 | - | -
5-Ethynyl-1H-pyrazole | - | - | - | - | - | View
Y-Aminobutyric acid (GABA) | - | - | - | - | - | View
L-Norleucine | L-Norleucine | HMDB0001645 | C01933 | 21236 | 18347 | 6334
sn-glycero-3-Phosphocholine | Glycerophosphocholine | HMDB0000086 | C00670 | 657272 | 16870 | 370
L-Glutamic acid | Glutamic acid | HMDB0000148 | C00025 | 33032 | 16015 | 5174
(E)-Glutaconate | Glutaconic acid | HMDB0000620 | C02214 | 5280498 | 15670 | 5593
DL-Arginine | DL-Arginine | HMDB0251511 | C02385 | 232 | 29016 | -
Adenine | Adenine | HMDB0000034 | C00147 | 190 | 16708 | 85
L-Pyroglutamic acid | Pyroglutamic acid | HMDB0000267 | C01879 | 7405 | 18183 | 3251
DL-Glutamine | - | - | - | - | - | View
You can download the result here
Xia Lab @ McGill (last updated 2026-01-23)



Unfortunately, some cases of huge compound names were uploaded to the server.
You may get '***No records found***.'

Submitting with a low number of compounds
(e.g., 10-20 compound names) was recommended.

However, if '*no records found*' still occurred. You may check some compounds that have very long compound names one by one.

Checkpoint I

What do we have now?

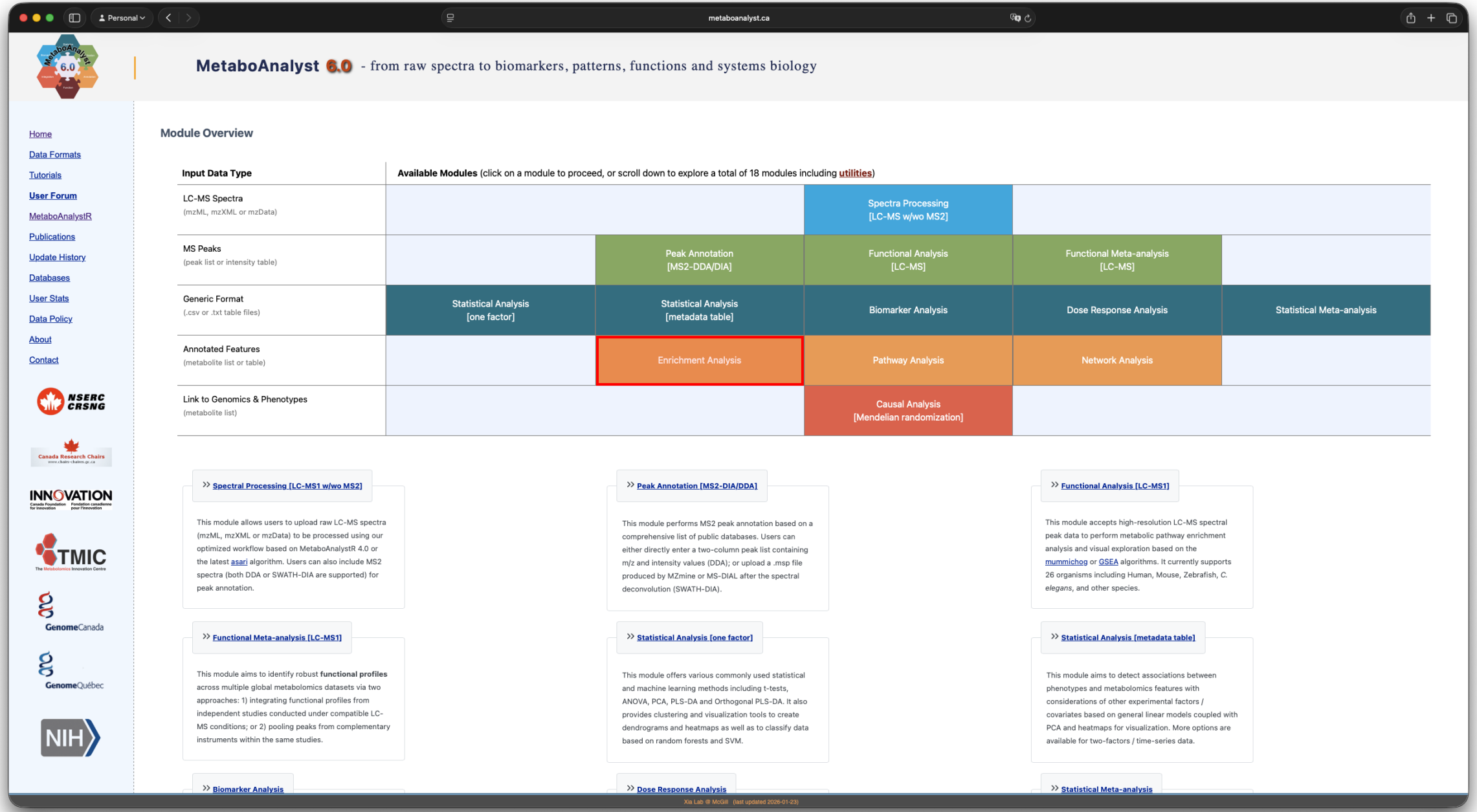
Metabolite list which converted to other database identifiers.

KEGG ID

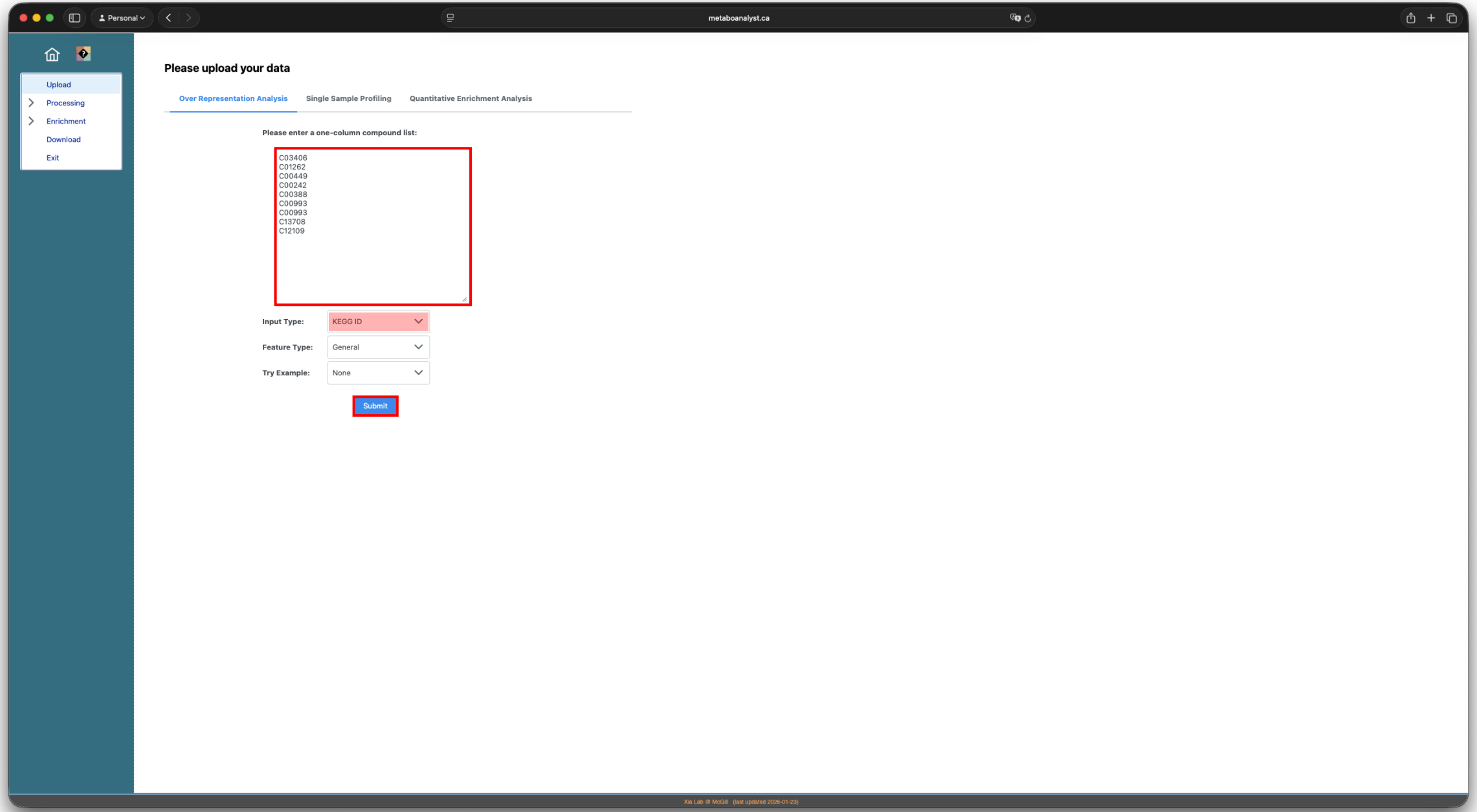
Ready for functional context

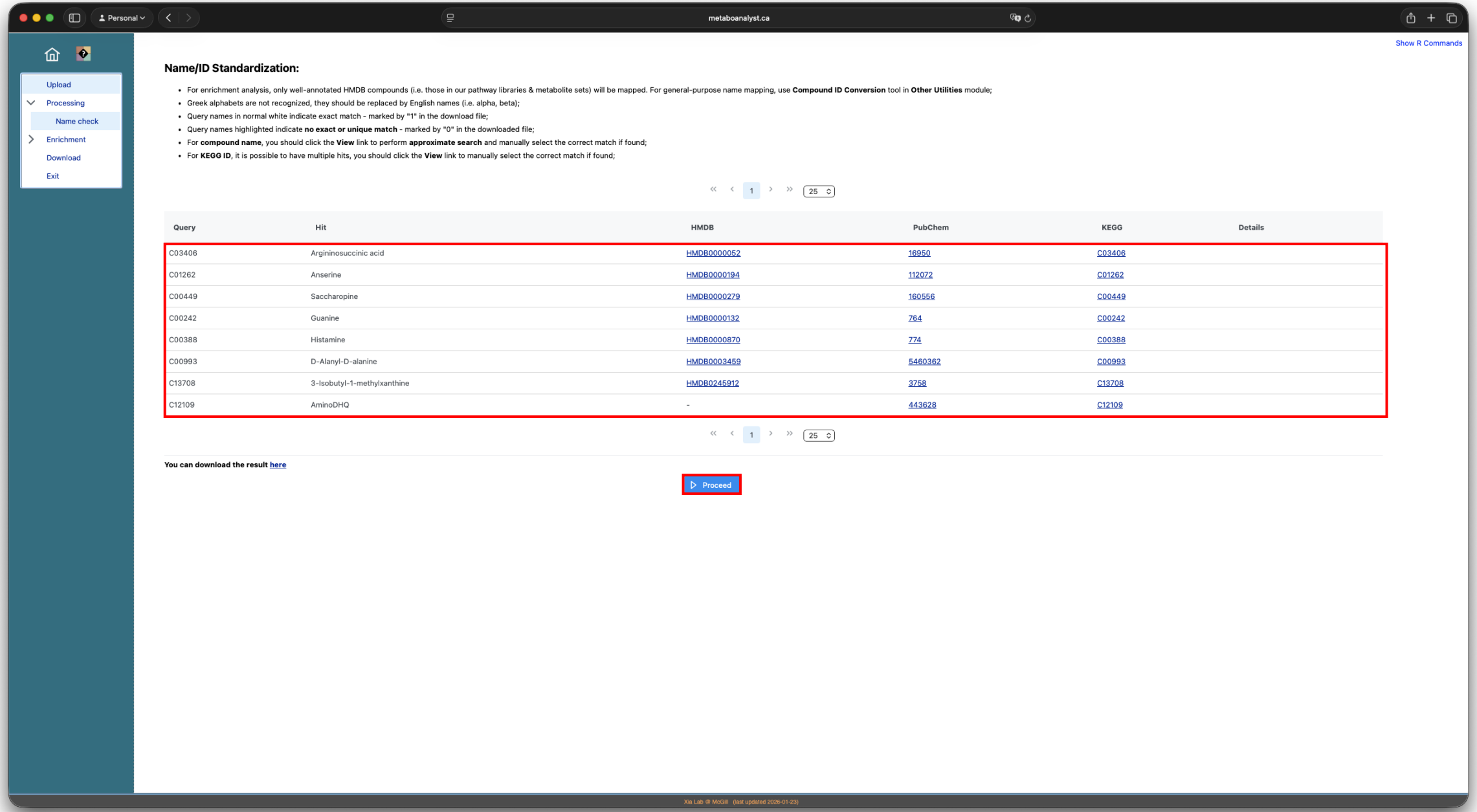
Enrichment analysis

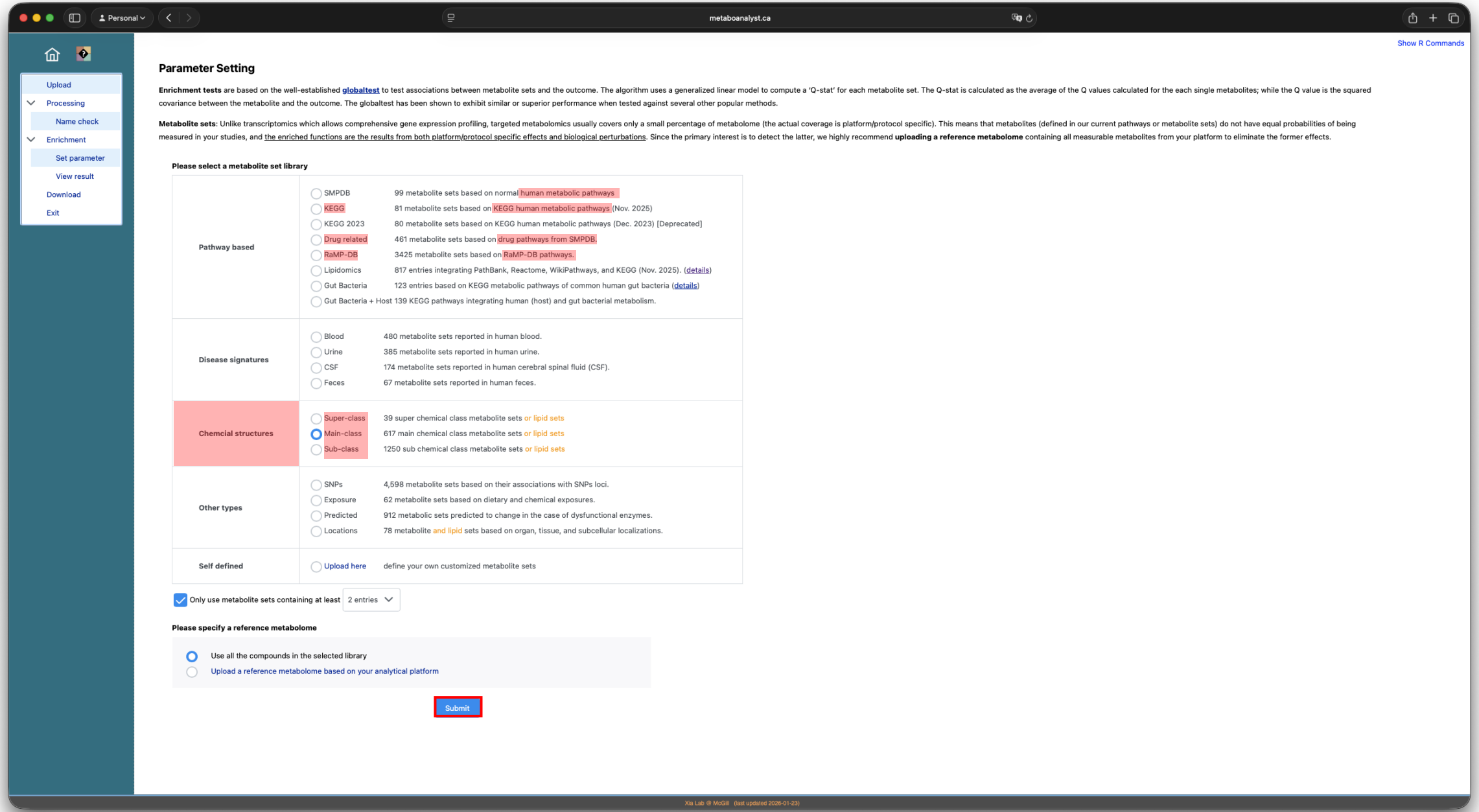
MetaboAnalyst 6.0

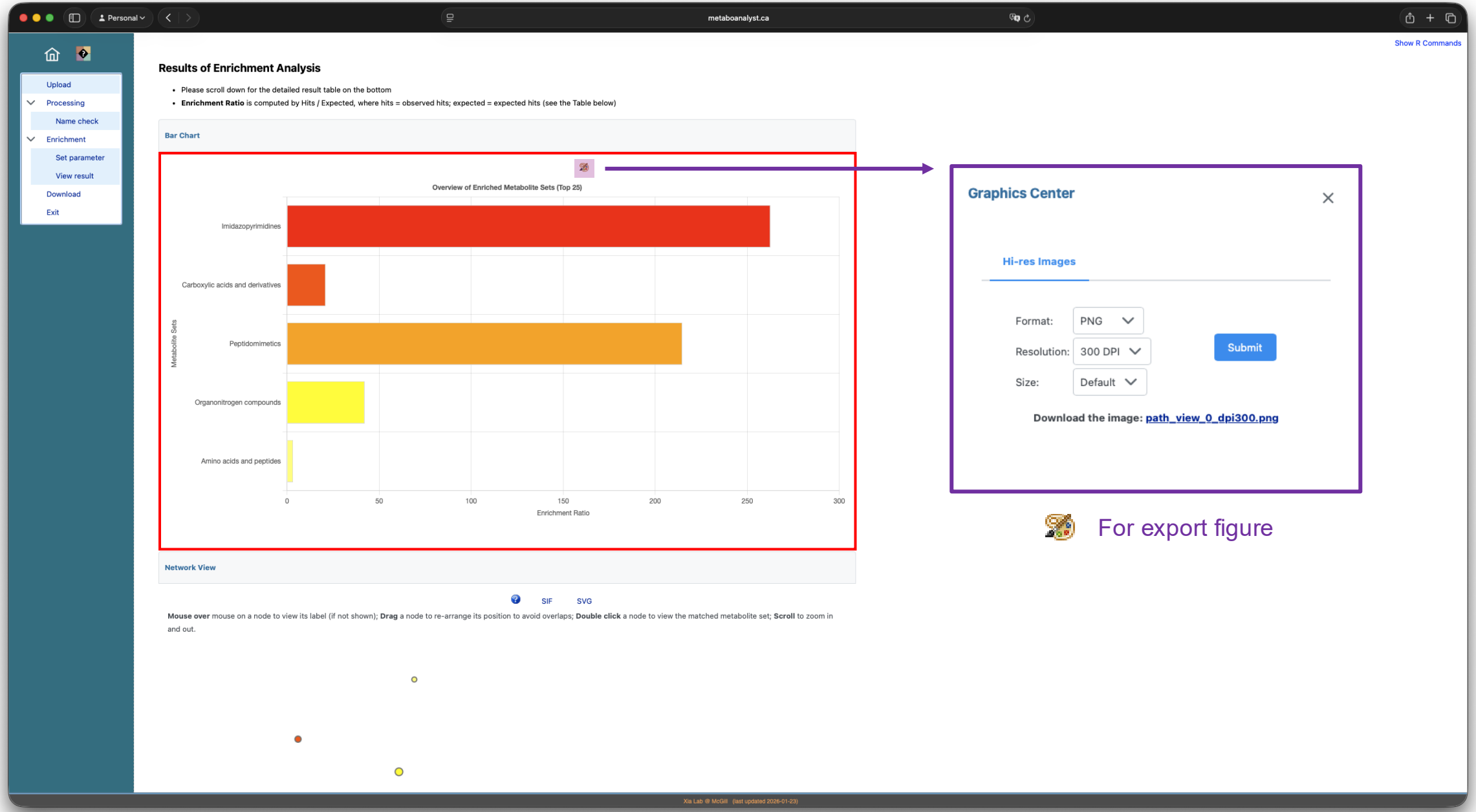

Personal
metaboanalyst.ca
MetaboAnalyst 6.0 - from raw spectra to biomarkers, patterns, functions and systems biology
Home
Data Formats
Tutorials
User Forum
MetaboAnalystR
Publications
Update History
Databases
User Stats
Data Policy
About
Contact
Module Overview
Input Data Type
Available Modules (click on a module to proceed, or scroll down to explore a total of 18 modules including utilities)
LC-MS Spectra
(mzML, mzXML or mzData)
MS Peaks
(peak list or intensity table)
Generic Format
(.csv or .txt table files)
Annotated Features
(metabolite list or table)
Link to Genomics & Phenotypes
(metabolite list)
Spectra Processing [LC-MS w/wo MS2]
Peak Annotation [MS2-DDA/DIA]
Functional Analysis [LC-MS]
Functional Meta-analysis [LC-MS]
Statistical Analysis [one factor]
Statistical Analysis [metadata table]
Biomarker Analysis
Dose Response Analysis
Statistical Meta-analysis
Enrichment Analysis
Pathway Analysis
Network Analysis
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 latest asari algorithm. Users can also include MS2 spectra (both DDA or SWATH-DIA are supported) for peak annotation.
Peak Annotation [MS2-DIA/DDA]
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 intensity values (DDA); or upload a .msp file produced by MZmine or MS-DIAL after the spectral deconvolution (SWATH-DIA).
Functional Analysis [LC-MS1]
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 algorithms. It currently supports 26 organisms including Human, Mouse, Zebrafish, C. elegans, and other species.
Functional Meta-analysis [LC-MS1]
This module aims to identify robust functional profiles across multiple global metabolomics datasets via two approaches: 1) integrating functional profiles from independent studies conducted under compatible LC-MS conditions; or 2) pooling peaks from complementary instruments within the same studies.
Statistical Analysis [one factor]
This module offers various commonly used statistical and machine learning methods including t-tests, ANOVA, PCA, PLS-DA and Orthogonal PLS-DA. It also provides clustering and visualization tools to create dendrograms and heatmaps as well as to classify data based on random forests and SVM.
Statistical Analysis [metadata table]
This module aims to detect associations between phenotypes and metabolomics features with considerations of other experimental factors / covariates based on general linear models coupled with PCA and heatmaps for visualization. More options are available for two-factors / time-series data.
Biomarker Analysis
Dose Response Analysis
Statistical Meta-analysis
Xia Lab @ McGill (last updated 2026-01-23)

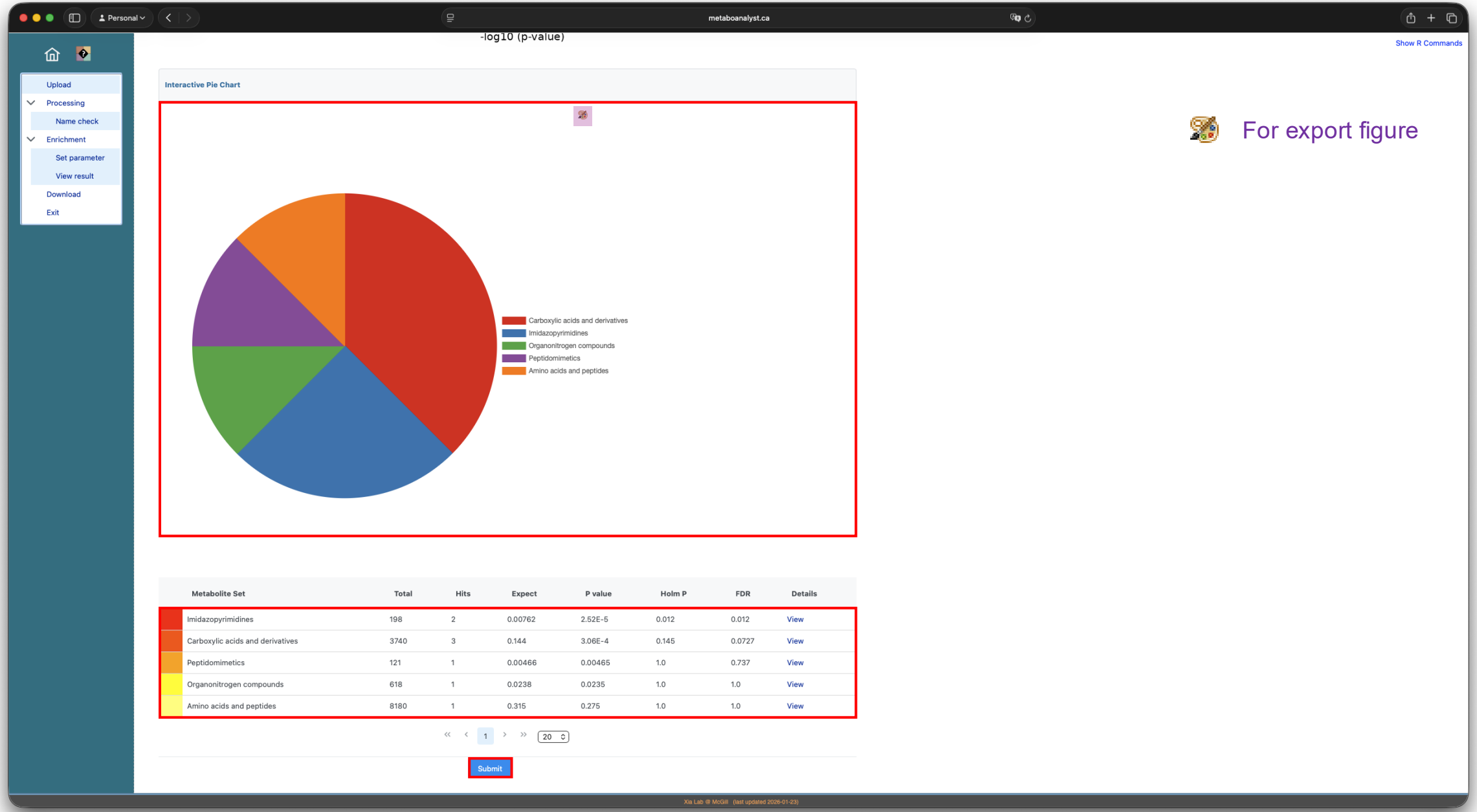
Enrichment analysis: upload metabolite IDs



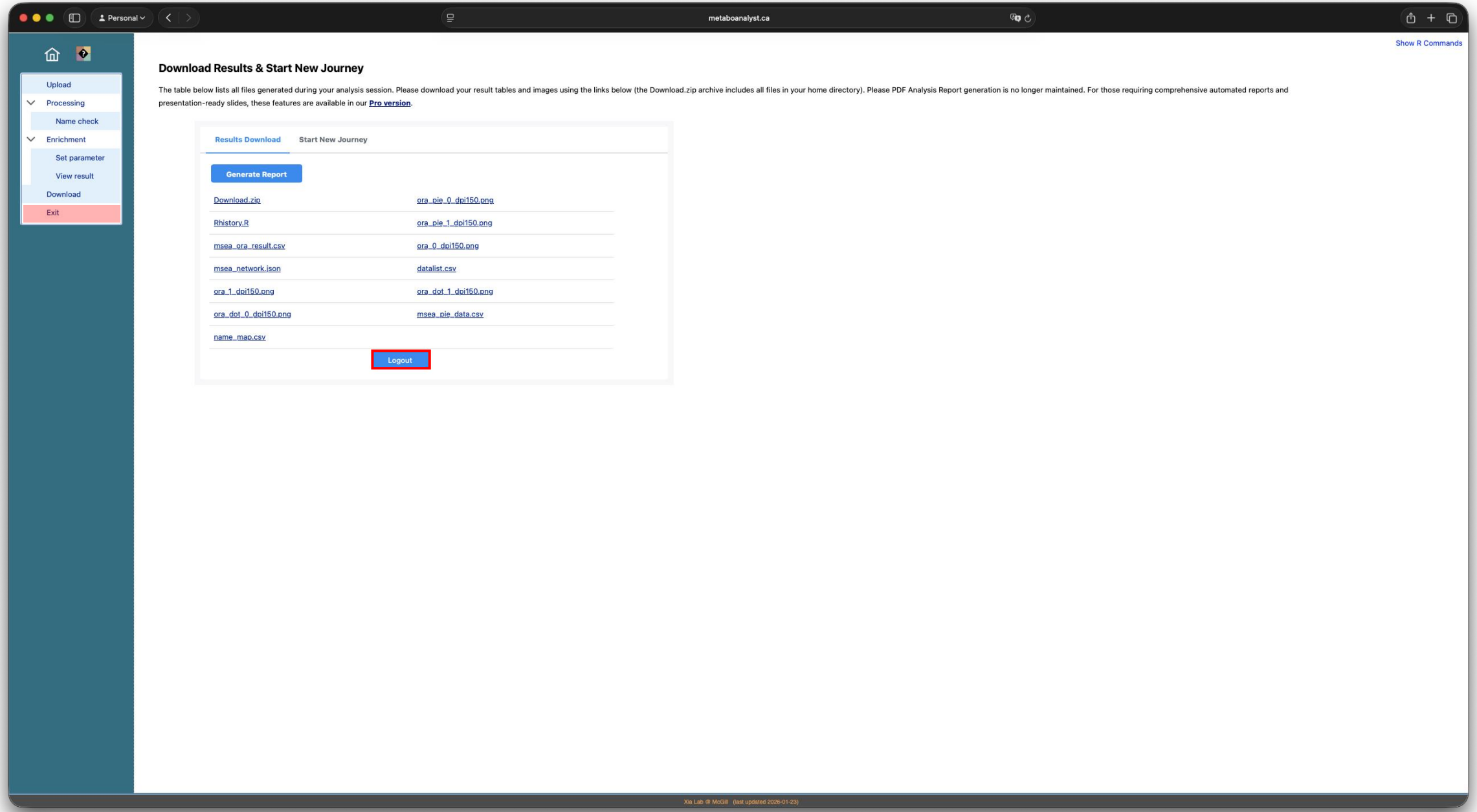

Personal
metaboanalyst.ca
Upload
Processing
Name check
Enrichment
Download
Exit
Show R Commands
Name/ID Standardization:
For enrichment analysis, only well-annotated HMDB compounds (i.e. those in our pathway libraries & metabolite sets) will be mapped. For general-purpose name mapping, use Compound ID Conversion tool in Other Utilities module;
Greek alphabets are not recognized, they should be replaced by English names (i.e. alpha, beta);
Query names in normal white indicate exact match - marked by "1" in the download file;
Query names highlighted indicate no exact or unique match - marked by "0" in the downloaded file;
For compound name, you should click the View link to perform approximate search and manually select the correct match if found;
For KEGG ID, it is possible to have multiple hits, you should click the View link to manually select the correct match if found;
1
25
Query
Hit
HMDB
PubChem
KEGG
Details
C03406 Argininosuccinic acid HMDB0000052 16950 C03406
C01262 Anserine HMDB0000194 112072 C01262
C00449 Saccharopine HMDB0000279 160556 C00449
C00242 Guanine HMDB0000132 764 C00242
C00388 Histamine HMDB0000870 774 C00388
C00993 D-Alanyl-D-alanine HMDB0003459 5460362 C00993
C13708 3-Isobutyl-1-methylxanthine HMDB0245912 3758 C13708
C12109 AminoDHQ - 443628 C12109
1
25
You can download the result here
Proceed
Xia Lab @ McGill (last updated 2026-01-23)


Personal
metaboanalyst.ca
Upload
Processing
Name check
Enrichment
Set parameter
View result
Download
Exit
Show R Commands
Parameter Setting
Enrichment tests are based on the well-established globaltest to test associations between metabolite sets and the outcome. The algorithm uses a generalized linear model to compute a 'Q-stat' for each metabolite set. The Q-stat is calculated as the average of the Q values calculated for the each single metabolites; while the Q value is the squared covariance between the metabolite and the outcome. The globaltest has been shown to exhibit similar or superior performance when tested against several other popular methods.
Metabolite sets: Unlike transcriptomics which allows comprehensive gene expression profiling, targeted metabolomics usually covers only a small percentage of metabolome (the actual coverage is platform/protocol specific). This means that metabolites (defined in our current pathways or metabolite sets) do not have equal probabilities of being measured in your studies, and the enriched functions are the results from both platform/protocol specific effects and biological perturbations. Since the primary interest is to detect the latter, we highly recommend uploading a reference metabolome containing all measurable metabolites from your platform to eliminate the former effects.
Please select a metabolite set library
Pathway based
SMPDB 99 metabolite sets based on normal human metabolic pathways
KEGG 81 metabolite sets based on KEGG human metabolic pathways (Nov. 2025)
KEGG 2023 80 metabolite sets based on KEGG human metabolic pathways (Dec. 2023) [Deprecated]
Drug related 461 metabolite sets based on drug pathways from SMPDB.
RaMP-DB 3425 metabolite sets based on RaMP-DB pathways.
Lipidomics 817 entries integrating PathBank, Reactome, WikiPathways, and KEGG (Nov. 2025). (details)
Gut Bacteria 123 entries based on KEGG metabolic pathways of common human gut bacteria (details)
Gut Bacteria + Host 139 KEGG pathways integrating human (host) and gut bacterial metabolism.
Disease signatures
Blood 480 metabolite sets reported in human blood.
Urine 385 metabolite sets reported in human urine.
CSF 174 metabolite sets reported in human cerebral spinal fluid (CSF).
Feces 67 metabolite sets reported in human feces.
Chemcial structures
Super-class 39 super chemical class metabolite sets or lipid sets
Main-class 617 main chemical class metabolite sets or lipid sets
Sub-class 1250 sub chemical class metabolite sets or lipid sets
Other types
SNPs 4,598 metabolite sets based on their associations with SNPs loci.
Exposure 62 metabolite sets based on dietary and chemical exposures.
Predicted 912 metabolic sets predicted to change in the case of dysfunctional enzymes.
Locations 78 metabolite and lipid sets based on organ, tissue, and subcellular localizations.
Self defined
Upload here define your own customized metabolite sets
Only use metabolite sets containing at least 2 entries
Please specify a reference metabolome
Use all the compounds in the selected library
Upload a reference metabolome based on your analytical platform
Submit
Xia Lab @ McGill (last updated 2026-01-23)


metaboanalyst.ca
Show R Commands
Upload
Processing
Name check
Enrichment
Set parameter
View result
Download
Exit
Results of Enrichment Analysis
Please scroll down for the detailed result table on the bottom
Enrichment Ratio is computed by Hits / Expected, where hits = observed hits; expected = expected hits (see the Table below)
Bar Chart
Overview of Enriched Metabolite Sets (Top 25)
Imidazopyrimidines
Carboxylic acids and derivatives
Peptidomimetics
Organonitrogen compounds
Amino acids and peptides
Metabolite Sets
0
50
100
150
200
250
300
Enrichment Ratio
Network View
SIF
SVG
Mouse over mouse on a node to view its label (if not shown); Drag a node to re-arrange its position to avoid overlaps; Double click a node to view the matched metabolite set; Scroll to zoom in and out.
Graphics Center
Hi-res Images
Format:
PNG
Resolution:
300 DPI
Size:
Default
Submit
Download the image: path_view_0_dpi300.png
For export figure
Xia Lab @ McGill (last updated 2026-01-23)



For export figure


Personal
metaboanalyst.ca
Upload
Processing
Name check
Enrichment
Set parameter
View result
Download
Exit
Show R Commands
Download Results & Start New Journey
The table below lists all files generated during your analysis session. Please download your result tables and images using the links below (the Download.zip archive includes all files in your home directory). Please PDF Analysis Report generation is no longer maintained. For those requiring comprehensive automated reports and presentation-ready slides, these features are available in our Pro version.
Results Download
Start New Journey
Generate Report
Download.zip
Rhistory.R
msea_ora_result.csv
msea_network.json
ora_1_dpi150.png
ora_dot_0_dpi150.png
name_map.csv
ora_pie_0_dpi150.png
ora_pie_1_dpi150.png
ora_0_dpi150.png
datalist.csv
ora_dot_1_dpi150.png
msea_pie_data.csv
Logout
Xia Lab @ McGill (last updated 2026-01-23)

Checkpoint II

What does this mean biologically?

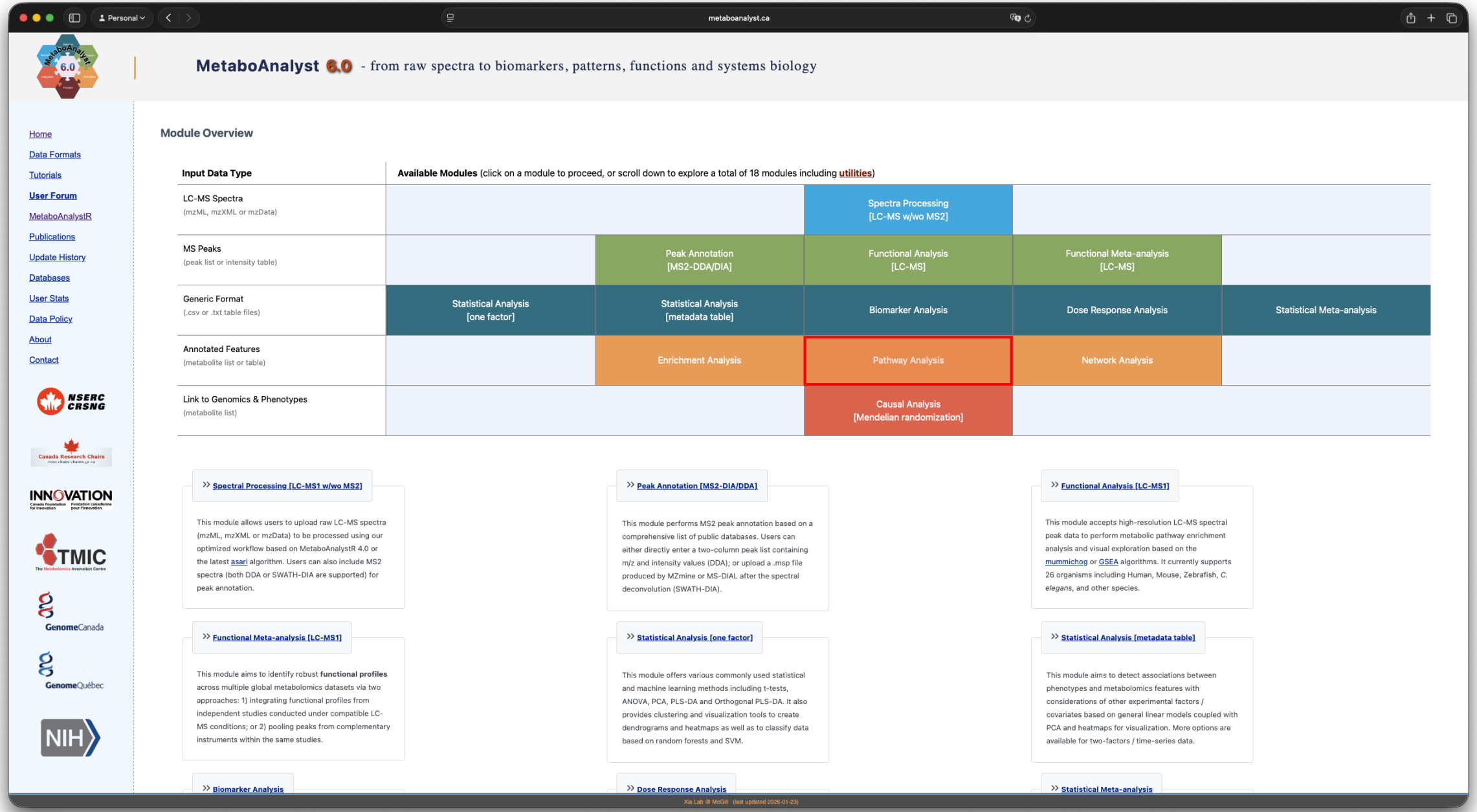
Enrichment analysis via ‘chemical class’

List of compound in their class

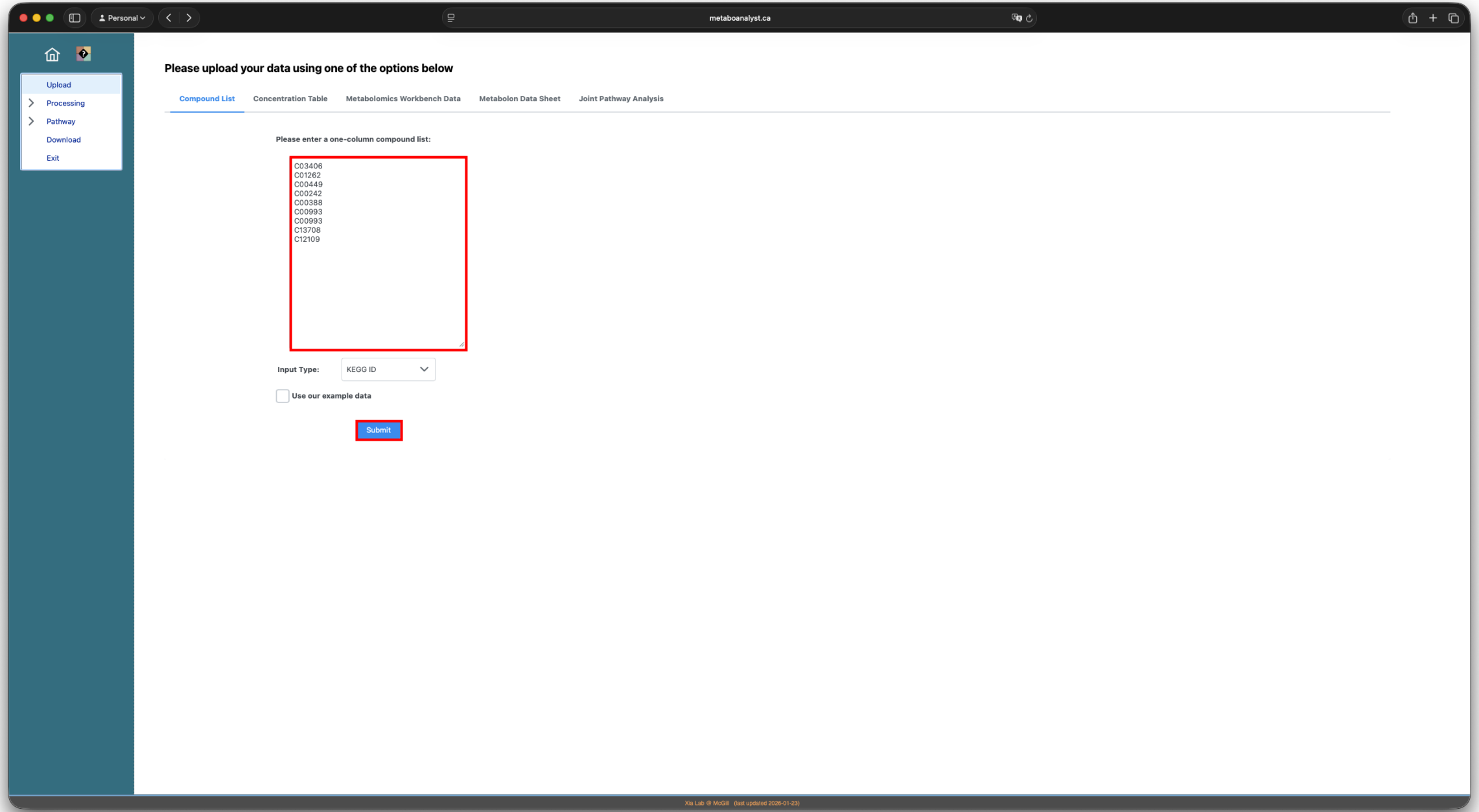
Biased chemical classes suggest …

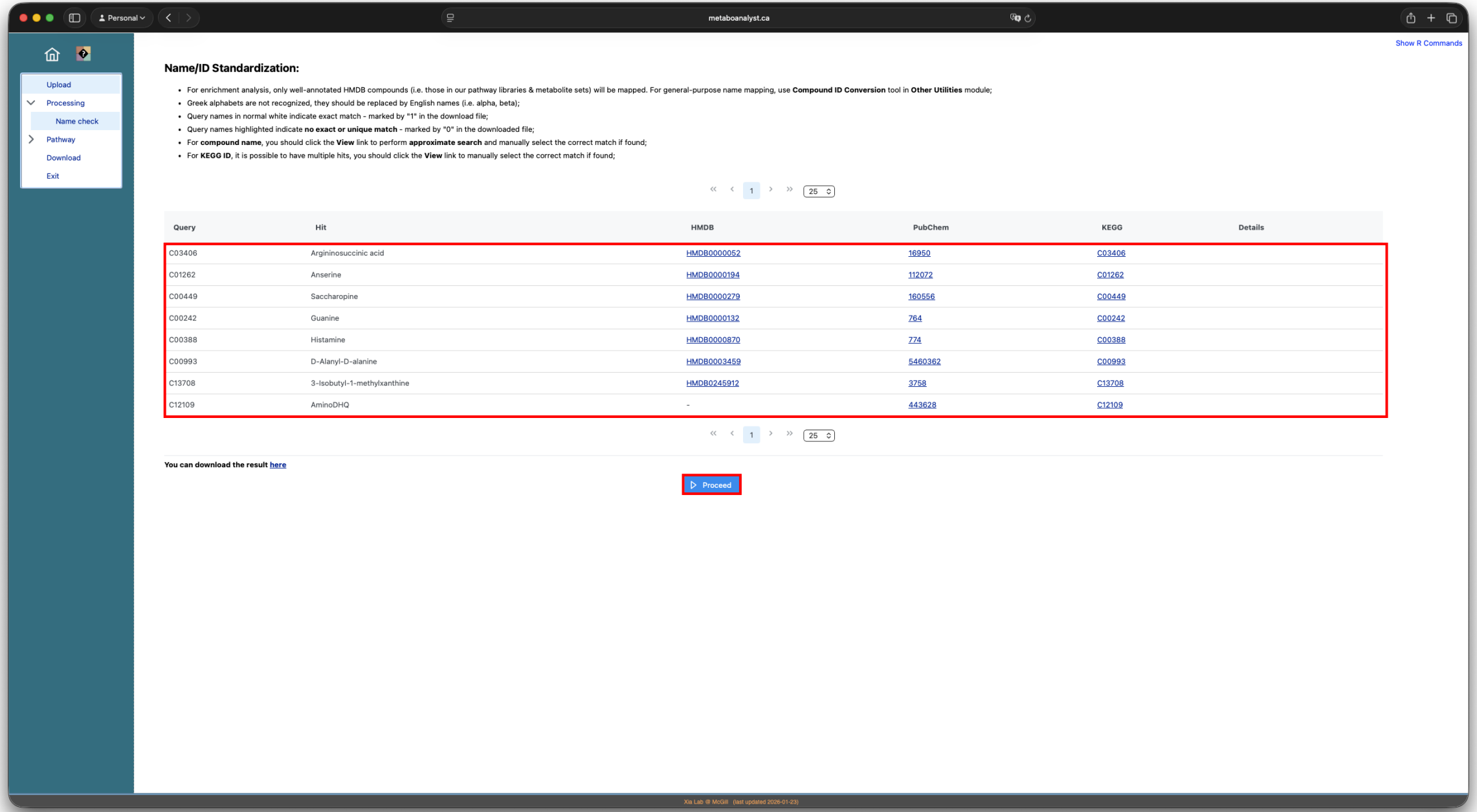
Pathway analysis

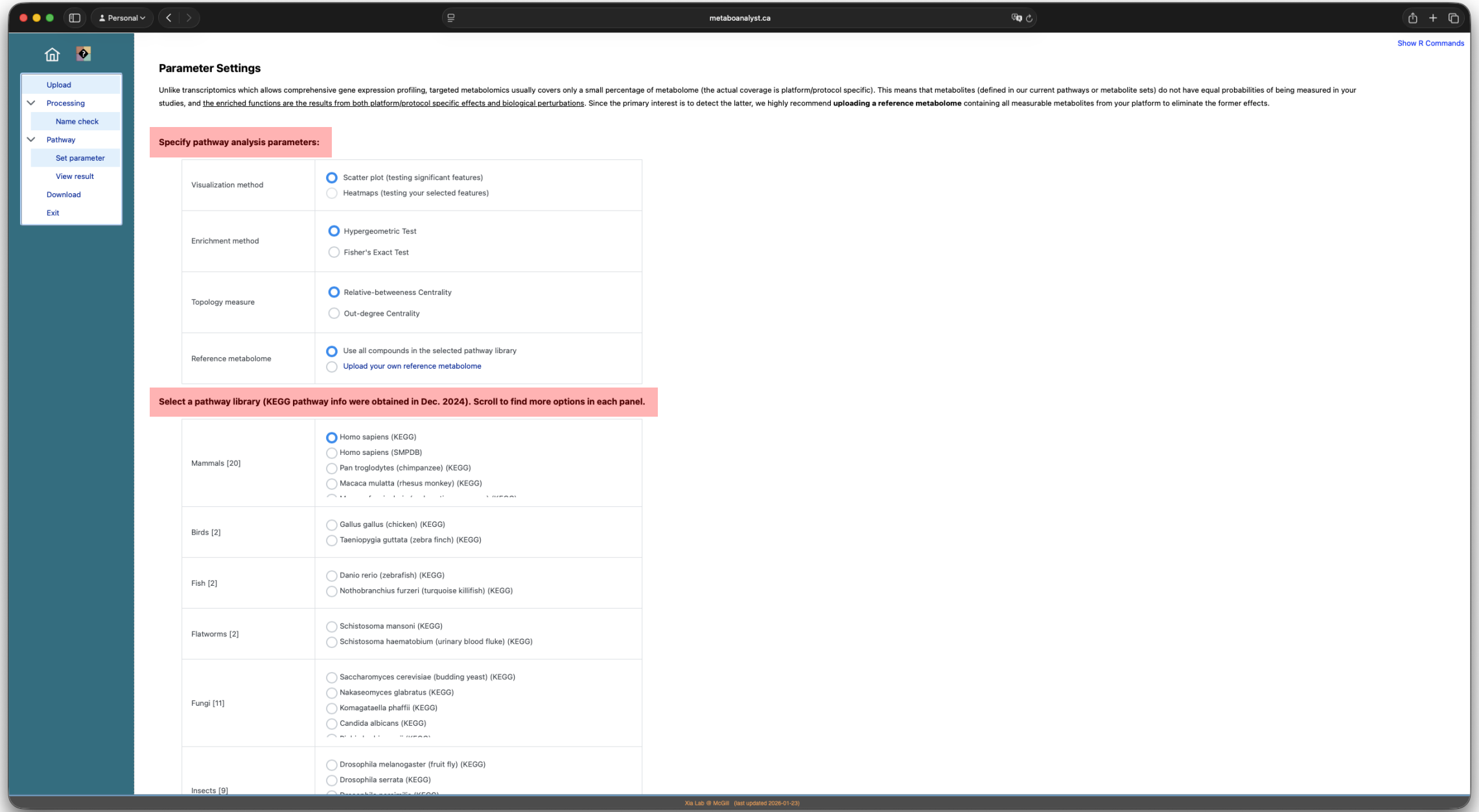
MetaboAnalyst 6.0

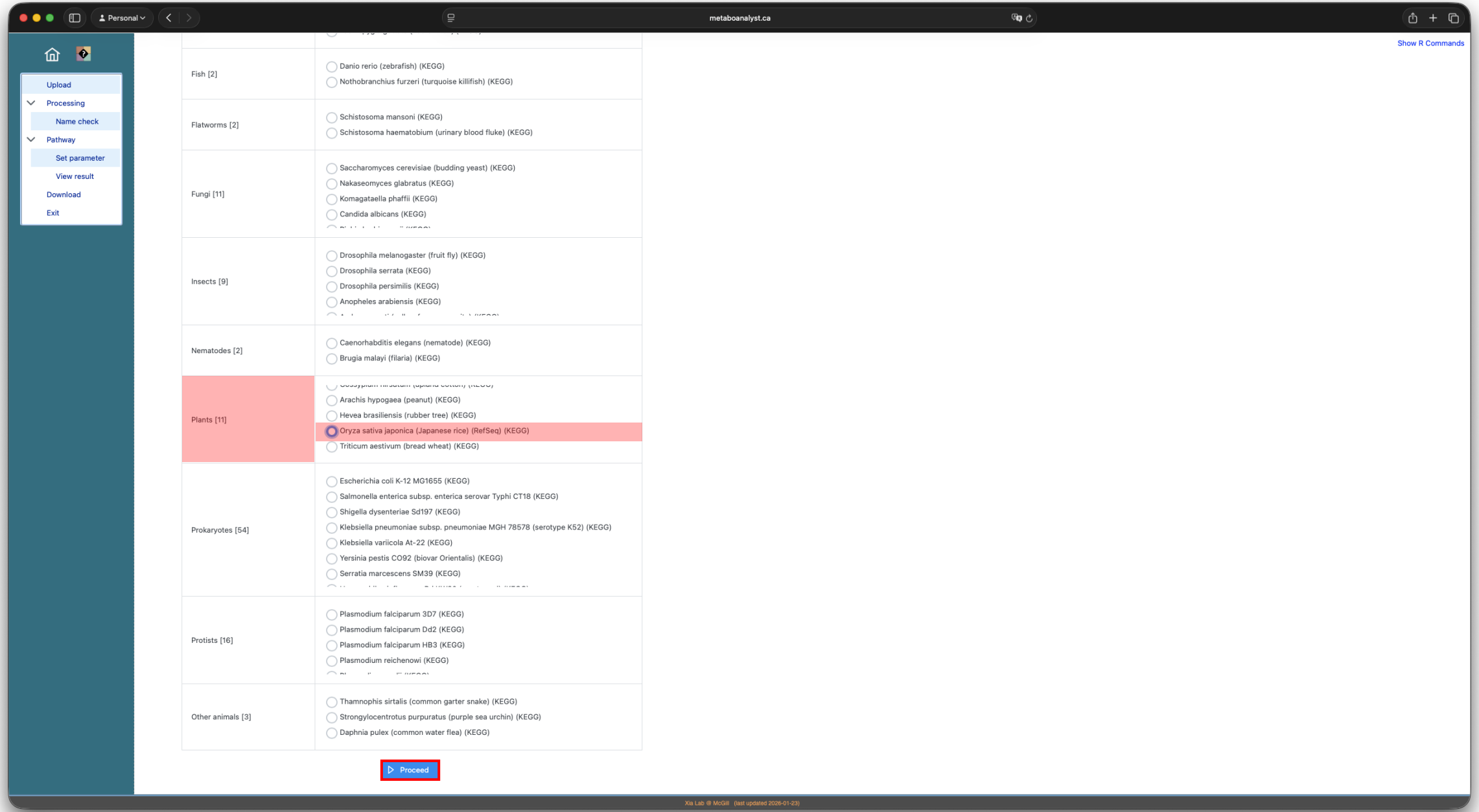

metaboanalyst.ca
MetaboAnalyst 6.0 - from raw spectra to biomarkers, patterns, functions and systems biology
Home
Data Formats
Tutorials
User Forum
MetaboAnalystR
Publications
Update History
Databases
User Stats
Data Policy
About
Contact
Module Overview
Input Data Type
Available Modules (click on a module to proceed, or scroll down to explore a total of 18 modules including utilities)
LC-MS Spectra
(mzML, mzXML or mzData)
MS Peaks
(peak list or intensity table)
Generic Format
(.csv or .txt table files)
Annotated Features
(metabolite list or table)
Link to Genomics & Phenotypes
(metabolite list)
Spectra Processing [LC-MS w/wo MS2]
Peak Annotation [MS2-DDA/DIA]
Functional Analysis [LC-MS]
Functional Meta-analysis [LC-MS]
Statistical Analysis [one factor]
Statistical Analysis [metadata table]
Biomarker Analysis
Dose Response Analysis
Statistical Meta-analysis
Enrichment Analysis
Pathway Analysis
Network Analysis
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 latest asari algorithm. Users can also include MS2 spectra (both DDA or SWATH-DIA are supported) for peak annotation.
Peak Annotation [MS2-DIA/DDA]
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 intensity values (DDA); or upload a .msp file produced by MZmine or MS-DIAL after the spectral deconvolution (SWATH-DIA).
Functional Analysis [LC-MS1]
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 algorithms. It currently supports 26 organisms including Human, Mouse, Zebrafish, C. elegans, and other species.
Functional Meta-analysis [LC-MS1]
This module aims to identify robust functional profiles across multiple global metabolomics datasets via two approaches: 1) integrating functional profiles from independent studies conducted under compatible LC-MS conditions; or 2) pooling peaks from complementary instruments within the same studies.
Statistical Analysis [one factor]
This module offers various commonly used statistical and machine learning methods including t-tests, ANOVA, PCA, PLS-DA and Orthogonal PLS-DA. It also provides clustering and visualization tools to create dendrograms and heatmaps as well as to classify data based on random forests and SVM.
Statistical Analysis [metadata table]
This module aims to detect associations between phenotypes and metabolomics features with considerations of other experimental factors / covariates based on general linear models coupled with PCA and heatmaps for visualization. More options are available for two-factors / time-series data.
Biomarker Analysis
Dose Response Analysis
Statistical Meta-analysis
Xia Lab @ McGill (last updated 2026-01-23)

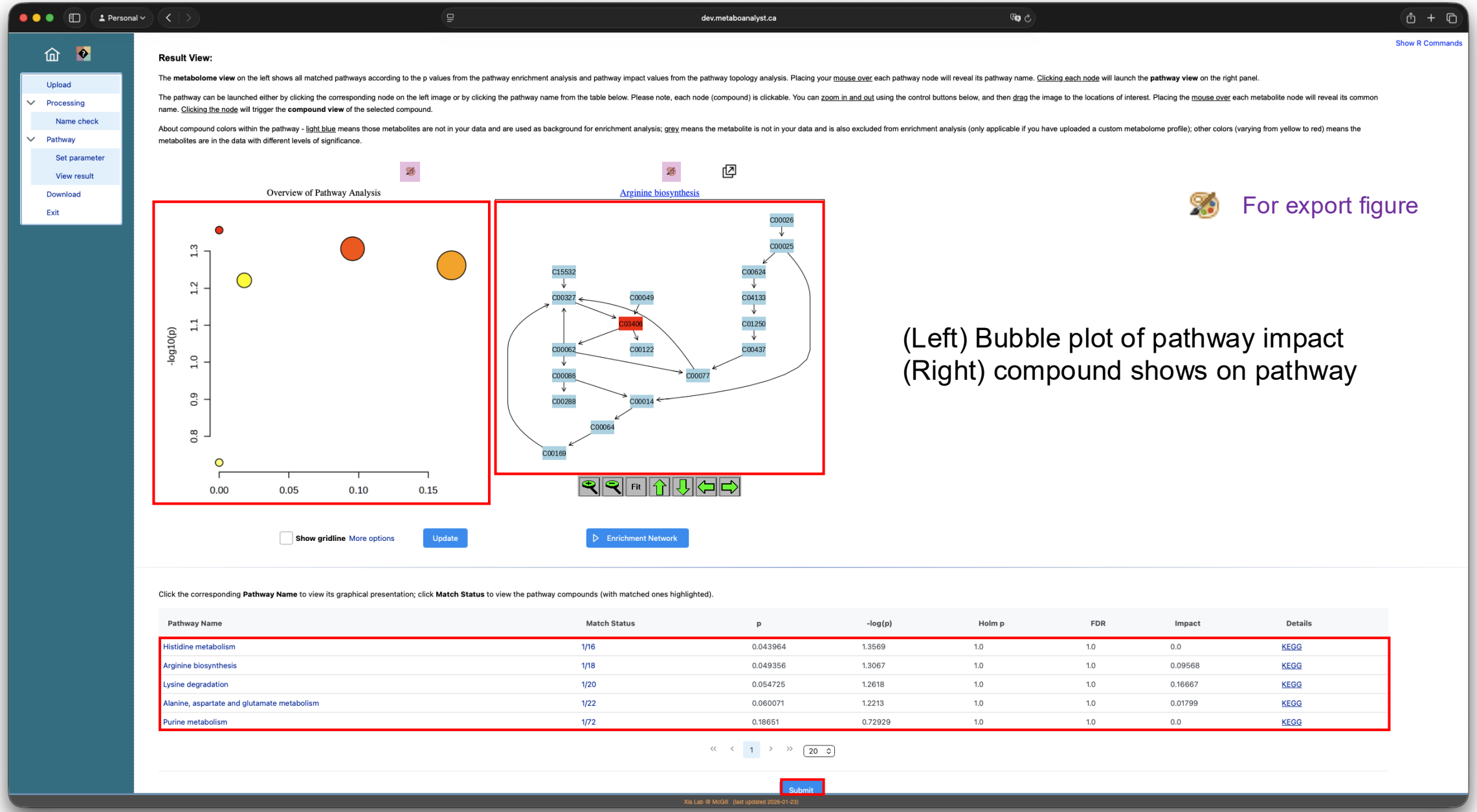
Pathway analysis: upload metabolite IDs

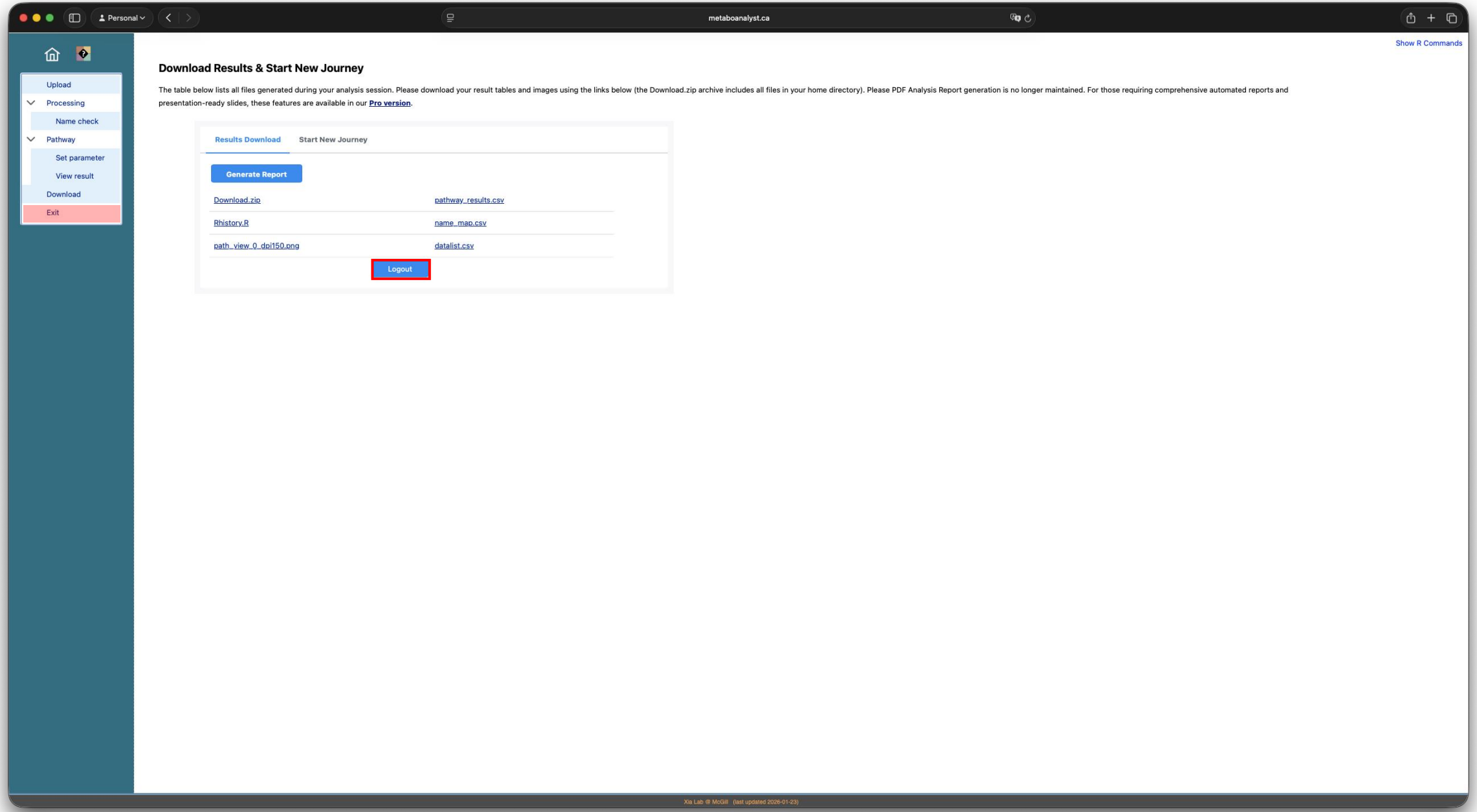



Personal
metaboanalyst.ca
Upload
Processing
Name check
Pathway
Download
Exit
Show R Commands
Name/ID Standardization:
For enrichment analysis, only well-annotated HMDB compounds (i.e. those in our pathway libraries & metabolite sets) will be mapped. For general-purpose name mapping, use Compound ID Conversion tool in Other Utilities module;
Greek alphabets are not recognized, they should be replaced by English names (i.e. alpha, beta);
Query names in normal white indicate exact match - marked by "1" in the download file;
Query names highlighted indicate no exact or unique match - marked by "0" in the downloaded file;
For compound name, you should click the View link to perform approximate search and manually select the correct match if found;
For KEGG ID, it is possible to have multiple hits, you should click the View link to manually select the correct match if found;
1
25
Query
Hit
HMDB
PubChem
KEGG
Details
C03406 Argininosuccinic acid HMDB0000052 16950 C03406
C01262 Anserine HMDB0000194 112072 C01262
C00449 Saccharopine HMDB0000279 160556 C00449
C00242 Guanine HMDB0000132 764 C00242
C00388 Histamine HMDB0000870 774 C00388
C00993 D-Alanyl-D-alanine HMDB0003459 5460362 C00993
C13708 3-Isobutyl-1-methylxanthine HMDB0245912 3758 C13708
C12109 AminoDHQ - 443628 C12109
1
25
You can download the result here
Proceed
Xia Lab @ McGill (last updated 2026-01-23)


Personal
metaboanalyst.ca
Upload
Processing
Name check
Pathway
Set parameter
View result
Download
Exit
Show R Commands
Parameter Settings
Unlike transcriptomics which allows comprehensive gene expression profiling, targeted metabolomics usually covers only a small percentage of metabolome (the actual coverage is platform/protocol specific). This means that metabolites (defined in our current pathways or metabolite sets) do not have equal probabilities of being measured in your studies, and the enriched functions are the results from both platform/protocol specific effects and biological perturbations. Since the primary interest is to detect the latter, we highly recommend uploading a reference metabolome containing all measurable metabolites from your platform to eliminate the former effects.
Specify pathway analysis parameters:
Visualization method
Scatter plot (testing significant features)
Heatmaps (testing your selected features)
Enrichment method
Hypergeometric Test
Fisher's Exact Test
Topology measure
Relative-betweeness Centrality
Out-degree Centrality
Reference metabolome
Use all compounds in the selected pathway library
Upload your own reference metabolome
Select a pathway library (KEGG pathway info were obtained in Dec. 2024). Scroll to find more options in each panel.
Mammals [20]
Homo sapiens (KEGG)
Homo sapiens (SMPDB)
Pan troglodytes (chimpanzee) (KEGG)
Macaca mulatta (rhesus monkey) (KEGG)
Birds [2]
Gallus gallus (chicken) (KEGG)
Taeniopygia guttata (zebra finch) (KEGG)
Fish [2]
Danio rerio (zebrafish) (KEGG)
Nothobranchius furzeri (turquoise killifish) (KEGG)
Flatworms [2]
Schistosoma mansoni (KEGG)
Schistosoma haematobium (urinary blood fluke) (KEGG)
Fungi [11]
Saccharomyces cerevisiae (budding yeast) (KEGG)
Nakaseomyces glabratus (KEGG)
Komagataella phaffii (KEGG)
Candida albicans (KEGG)
Insects [9]
Drosophila melanogaster (fruit fly) (KEGG)
Drosophila serrata (KEGG)
Xia Lab @ McGill (last updated 2026-01-23)


metaboanalyst.ca
Upload
Processing
Name check
Pathway
Set parameter
View result
Download
Exit
Show R Commands
Fish [2]
Danio rerio (zebrafish) (KEGG)
Nothobranchius furzeri (turquoise killifish) (KEGG)
Flatworms [2]
Schistosoma mansoni (KEGG)
Schistosoma haematobium (urinary blood fluke) (KEGG)
Fungi [11]
Saccharomyces cerevisiae (budding yeast) (KEGG)
Nakaseomyces glabratus (KEGG)
Komagataella phaffii (KEGG)
Candida albicans (KEGG)
Insects [9]
Drosophila melanogaster (fruit fly) (KEGG)
Drosophila serrata (KEGG)
Drosophila persimilis (KEGG)
Anopheles arabiensis (KEGG)
Nematodes [2]
Caenorhabditis elegans (nematode) (KEGG)
Brugia malayi (filaria) (KEGG)
Plants [11]
Arachis hypogaea (peanut) (KEGG)
Hevea brasiliensis (rubber tree) (KEGG)
Oryza sativa japonica (Japanese rice) (RefSeq) (KEGG)
Triticum aestivum (bread wheat) (KEGG)
Prokaryotes [54]
Escherichia coli K-12 MG1655 (KEGG)
Salmonella enterica subsp. enterica serovar Typhi CT18 (KEGG)
Shigella dysenteriae Sd197 (KEGG)
Klebsiella pneumoniae subsp. pneumoniae MGH 78578 (serotype K52) (KEGG)
Klebsiella variicola At-22 (KEGG)
Yersinia pestis CO92 (biovar Orientalis) (KEGG)
Serratia marcescens SM39 (KEGG)
Protists [16]
Plasmodium falciparum 3D7 (KEGG)
Plasmodium falciparum Dd2 (KEGG)
Plasmodium falciparum HB3 (KEGG)
Plasmodium reichenowi (KEGG)
Other animals [3]
Thamnophis sirtalis (common garter snake) (KEGG)
Strongylocentrotus purpuratus (purple sea urchin) (KEGG)
Daphnia pulex (common water flea) (KEGG)
Proceed
Xia Lab @ McGill (last updated 2026-01-23)


dev.metaboanalyst.ca
Show R Commands
Upload
Processing
Name check
Pathway
Set parameter
View result
Download
Exit
Result View:
The metabolome view on the left shows all matched pathways according to the p values from the pathway enrichment analysis and pathway impact values from the pathway topology analysis. Placing your mouse over each pathway node will reveal its pathway name. Clicking each node will launch the pathway view on the right panel.
The pathway can be launched either by clicking the corresponding node on the left image or by clicking the pathway name from the table below. Please note, each node (compound) is clickable. You can zoom in and out using the control buttons below, and then drag the image to the locations of interest. Placing the mouse over each metabolite node will reveal its common name. Clicking the node will trigger the compound view of the selected compound.
About compound colors within the pathway - light blue means those metabolites are not in your data and are used as background for enrichment analysis; grey means the metabolite is not in your data and is also excluded from enrichment analysis (only applicable if you have uploaded a custom metabolome profile); other colors (varying from yellow to red) means the metabolites are in the data with different levels of significance.
Overview of Pathway Analysis
-log10(p)
0.8
0.9
1.0
1.1
1.2
1.3
0.00
0.05
0.10
0.15
Arginine biosynthesis
C00026
C00025
C15532
C00624
C00327
C00049
C04133
C03406
C01250
C00062
C00122
C00437
C00086
C00077
C00288
C00014
C00064
C00169
Fit
Show gridline
More options
Update
Enrichment Network
For export figure
(Left) Bubble plot of pathway impact
(Right) compound shows on pathway
Click the corresponding Pathway Name to view its graphical presentation; click Match Status to view the pathway compounds (with matched ones highlighted).
Pathway Name
Match Status
p
-log(p)
Holm p
FDR
Impact
Details
Histidine metabolism
1/16
0.043964
1.3569
1.0
1.0
0.0
KEGG
Arginine biosynthesis
1/18
0.049356
1.3067
1.0
1.0
0.09568
KEGG
Lysine degradation
1/20
0.054725
1.2618
1.0
1.0
0.16667
KEGG
Alanine, aspartate and glutamate metabolism
1/22
0.060071
1.2213
1.0
1.0
0.01799
KEGG
Purine metabolism
1/72
0.18651
0.72929
1.0
1.0
0.0
KEGG
1
20
Submit
Xia Lab @ McGill (last updated 2026-01-23)


Personal
metaboanalyst.ca
Upload
Processing
Name check
Pathway
Set parameter
View result
Download
Exit
Show R Commands
Download Results & Start New Journey
The table below lists all files generated during your analysis session. Please download your result tables and images using the links below (the Download.zip archive includes all files in your home directory). Please PDF Analysis Report generation is no longer maintained. For those requiring comprehensive automated reports and presentation-ready slides, these features are available in our Pro version.
Results Download
Start New Journey
Generate Report
Download.zip
pathway_results.csv
Rhistory.R
name_map.csv
path_view_0_dpi150.png
datalist.csv
Logout
Xia Lab @ McGill (last updated 2026-01-23)

Checkpoint III

What does this mean biologically?

Pathway analysis result of the bubble plot

Various compounds regulate in molecular mechanism

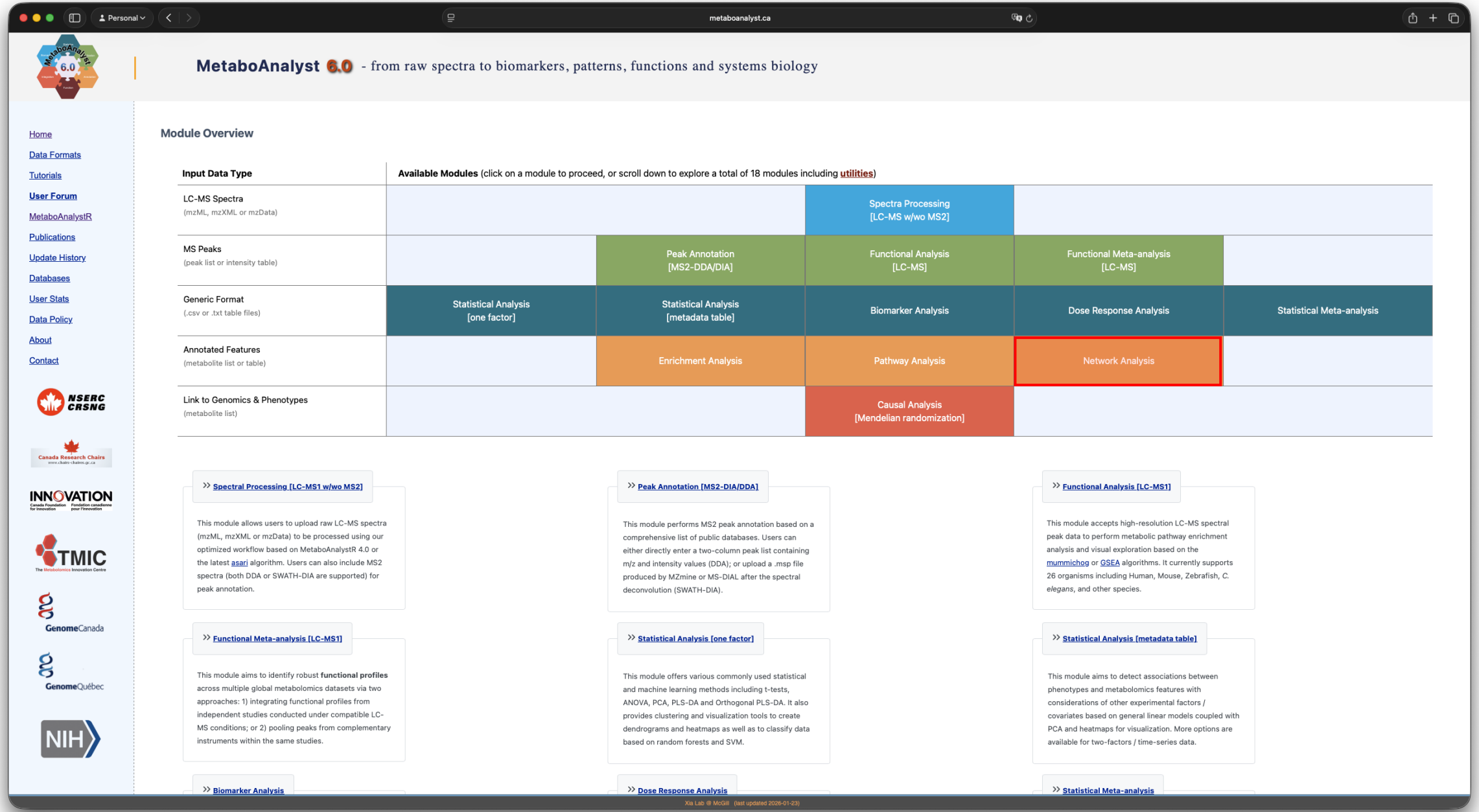
Topology ≠ Significance

(Optional)

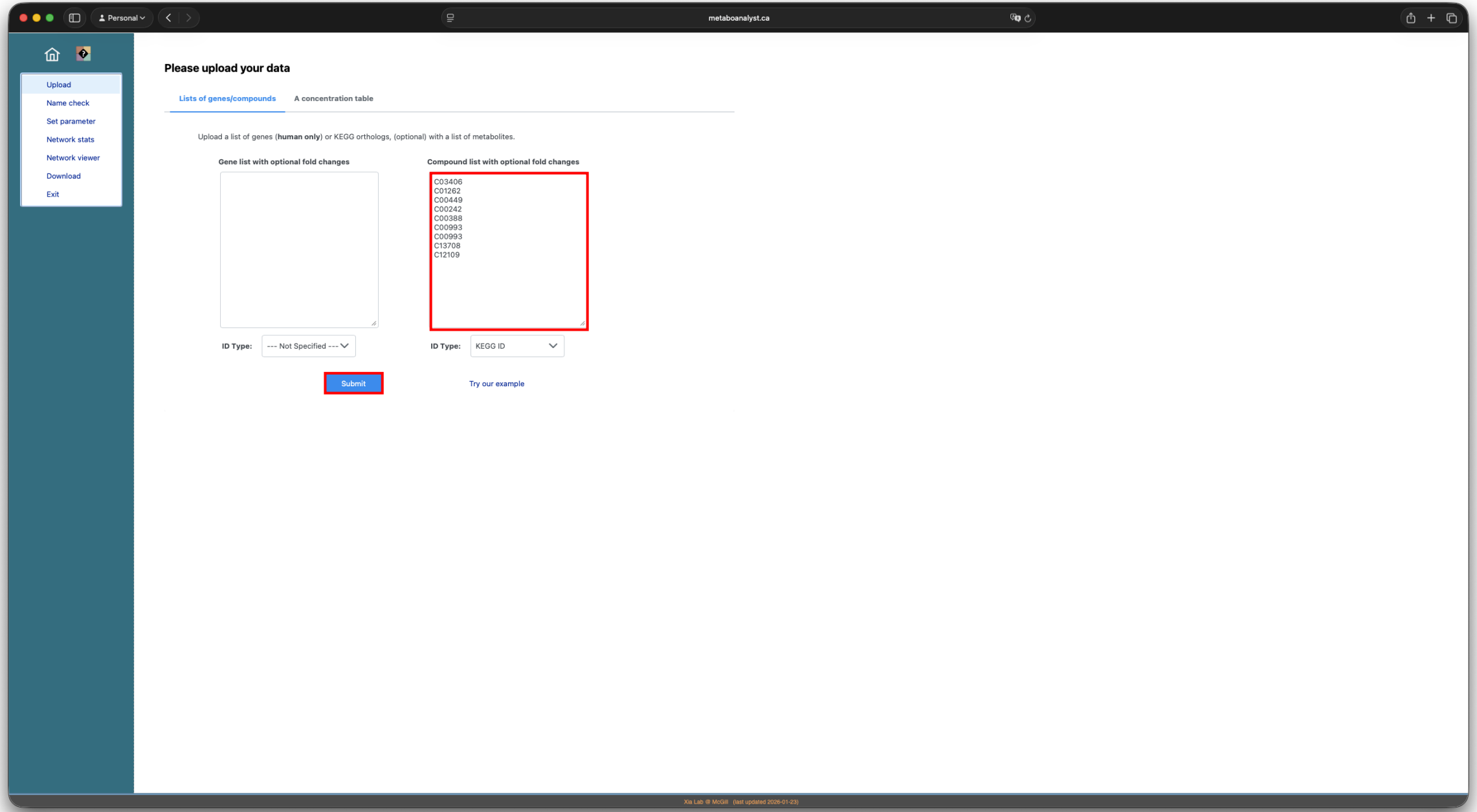
Network analysis

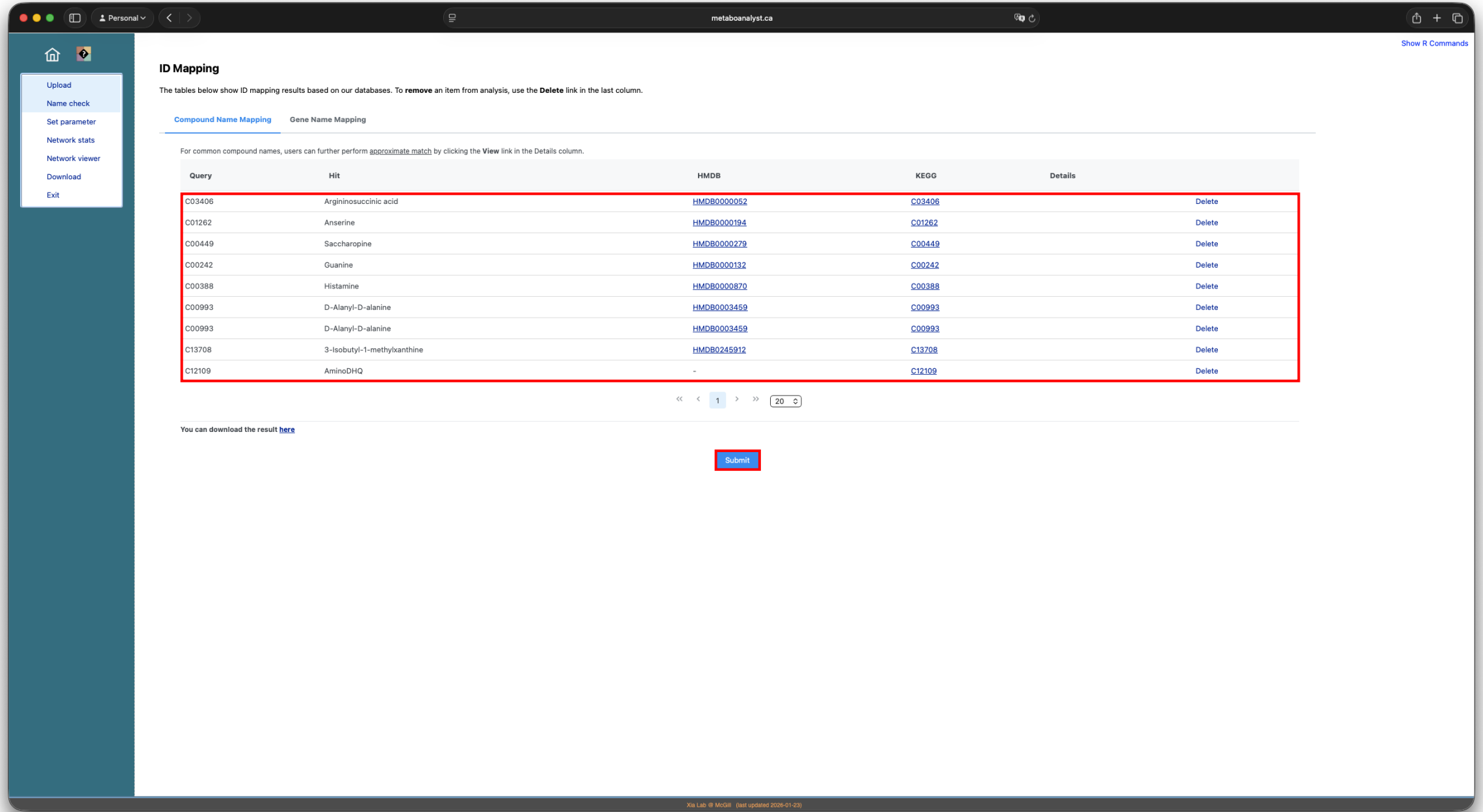
MetaboAnalyst 6.0

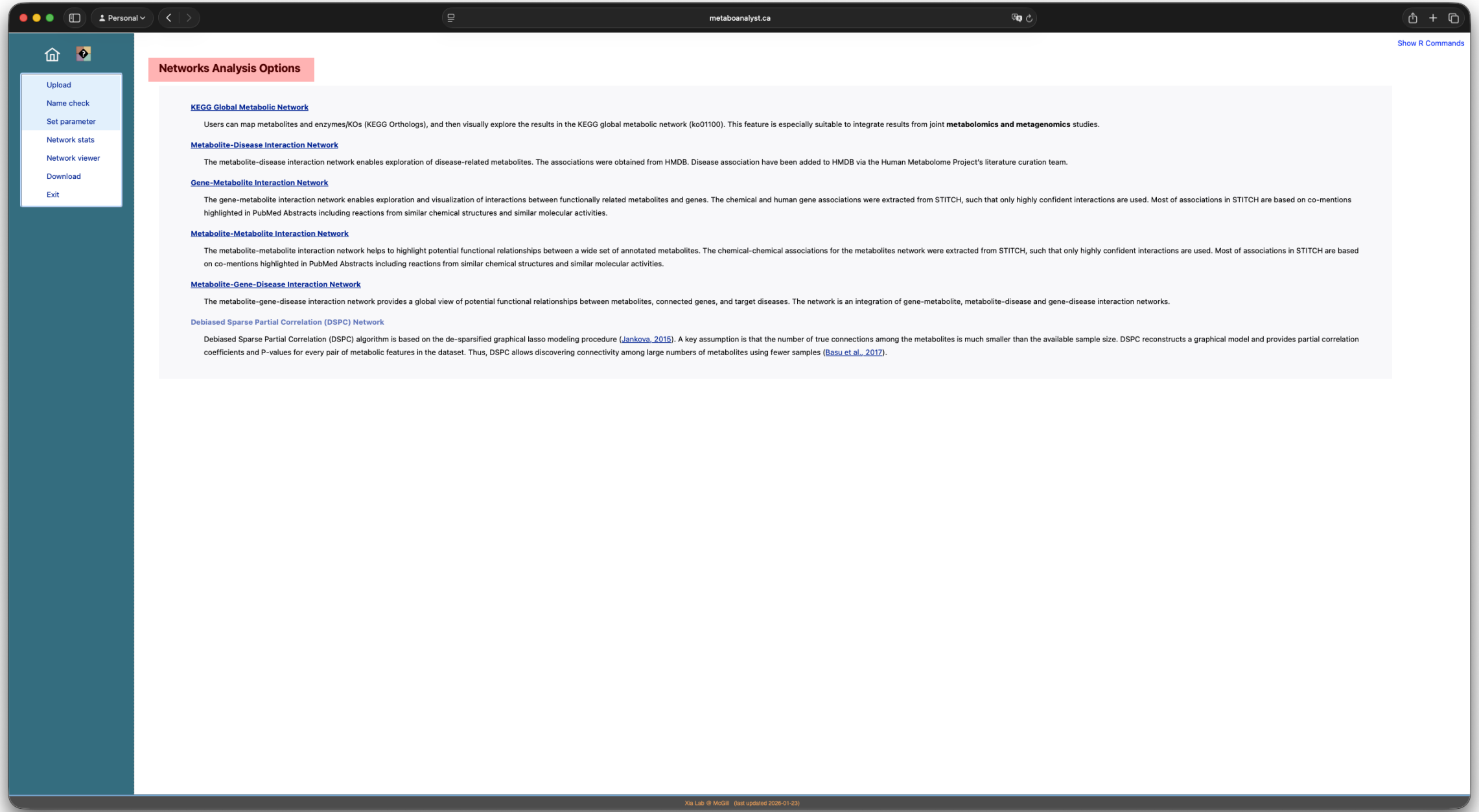
(Optional) Network analysis



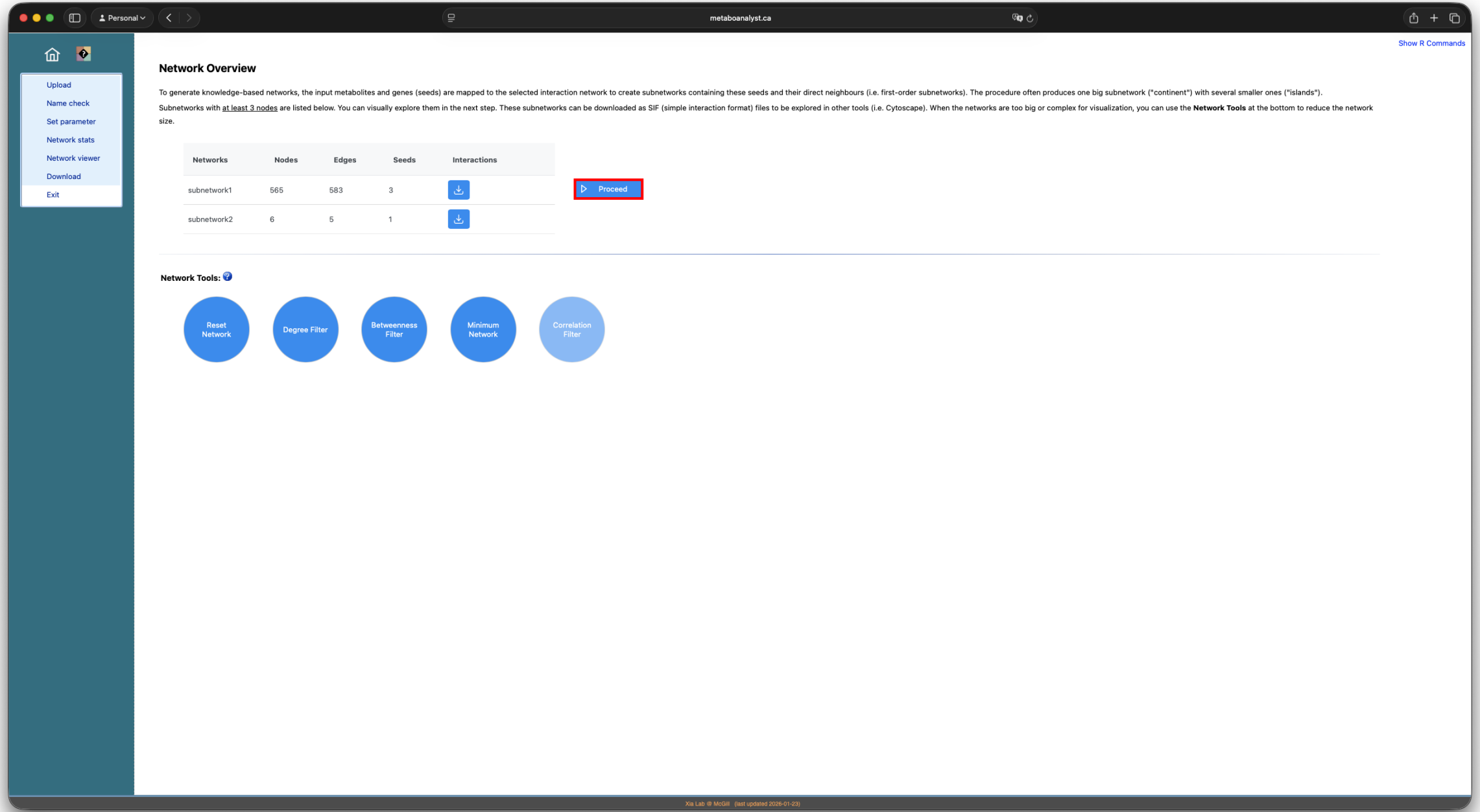
(Optional) Network analysis: upload metabolite IDs



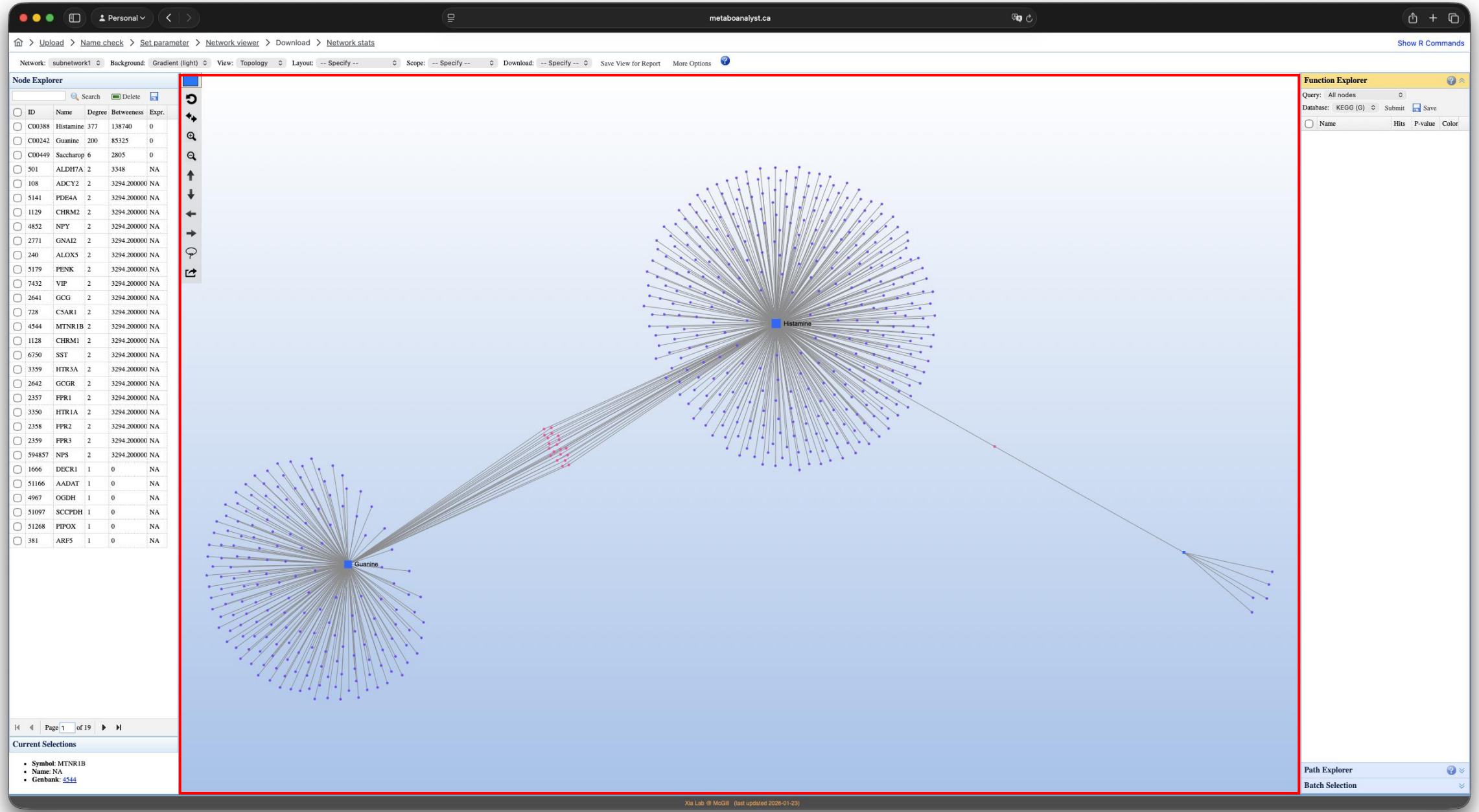

Personal
metaboanalyst.ca
Upload
Name check
Set parameter
Network stats
Network viewer
Download
Exit
Show R Commands
ID Mapping
The tables below show ID mapping results based on our databases. To remove an item from analysis, use the Delete link in the last column.
Compound Name Mapping
Gene Name Mapping
For common compound names, users can further perform approximate match by clicking the View link in the Details column.
Query
Hit
HMDB
KEGG
Details
C03406 Argininosuccinic acid HMDB0000052 C03406 Delete
C01262 Anserine HMDB0000194 C01262 Delete
C00449 Saccharopine HMDB0000279 C00449 Delete
C00242 Guanine HMDB0000132 C00242 Delete
C00388 Histamine HMDB0000870 C00388 Delete
C00993 D-Alanyl-D-alanine HMDB0003459 C00993 Delete
C00993 D-Alanyl-D-alanine HMDB0003459 C00993 Delete
C13708 3-Isobutyl-1-methylxanthine HMDB0245912 C13708 Delete
C12109 AminoDHQ - C12109 Delete
1
20
You can download the result here
Submit
Xia Lab @ McGill (last updated 2026-01-23)


Personal
metaboanalyst.ca
Upload
Name check
Set parameter
Network stats
Network viewer
Download
Exit
Show R Commands
Networks Analysis Options
KEGG Global Metabolic Network
Users can map metabolites and enzymes/KOs (KEGG Orthologs), and then visually explore the results in the KEGG global metabolic network (ko01100). This feature is especially suitable to integrate results from joint metabolomics and metagenomics studies.
Metabolite-Disease Interaction Network
The metabolite-disease interaction network enables exploration of disease-related metabolites. The associations were obtained from HMDB. Disease association have been added to HMDB via the Human Metabolome Project's literature curation team.
Gene-Metabolite Interaction Network
The gene-metabolite interaction network enables exploration and visualization of interactions between functionally related metabolites and genes. The chemical and human gene associations were extracted from STITCH, such that only highly confident interactions are used. Most of associations in STITCH are based on co-mentions highlighted in PubMed Abstracts including reactions from similar chemical structures and similar molecular activities.
Metabolite-Metabolite Interaction Network
The metabolite-metabolite interaction network helps to highlight potential functional relationships between a wide set of annotated metabolites. The chemical-chemical associations for the metabolites network were extracted from STITCH, such that only highly confident interactions are used. Most of associations in STITCH are based on co-mentions highlighted in PubMed Abstracts including reactions from similar chemical structures and similar molecular activities.
Metabolite-Gene-Disease Interaction Network
The metabolite-gene-disease interaction network provides a global view of potential functional relationships between metabolites, connected genes, and target diseases. The network is an integration of gene-metabolite, metabolite-disease and gene-disease interaction networks.
Debiased Sparse Partial Correlation (DSPC) Network
Debiased Sparse Partial Correlation (DSPC) algorithm is based on the de-sparsified graphical lasso modeling procedure (Jankova, 2015). A key assumption is that the number of true connections among the metabolites is much smaller than the available sample size. DSPC reconstructs a graphical model and provides partial correlation coefficients and P-values for every pair of metabolic features in the dataset. Thus, DSPC allows discovering connectivity among large numbers of metabolites using fewer samples (Basu et al., 2017).
Xia Lab @ McGill (last updated 2026-01-23)

(Optional) Network analysis: setup parameters II



(Optional) Network analysis: result



Checkpoint IV

What does this mean biologically?

Network of compound interaction

Figure of interaction

How each compound do together?

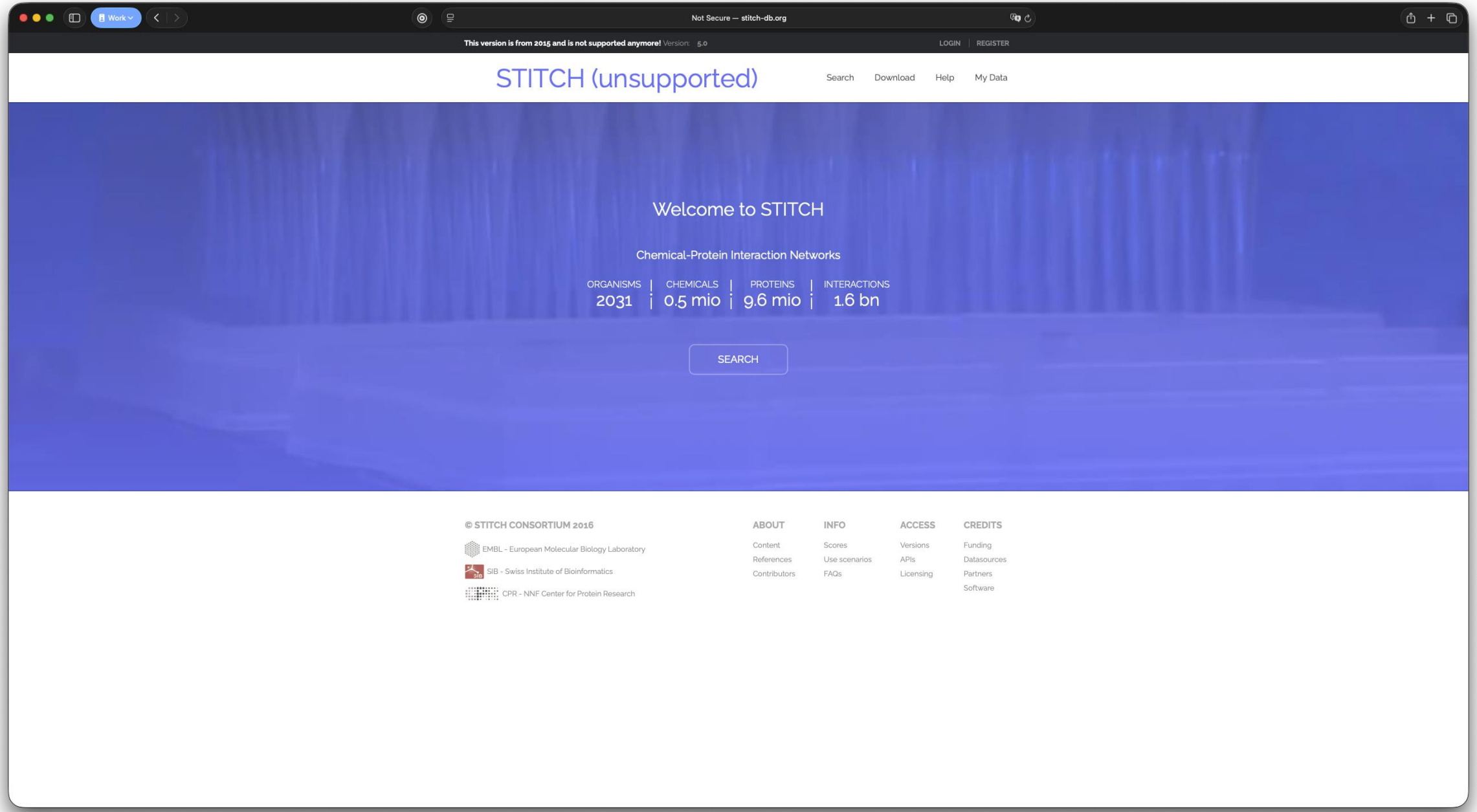
It’s all connected!?

(Optional)

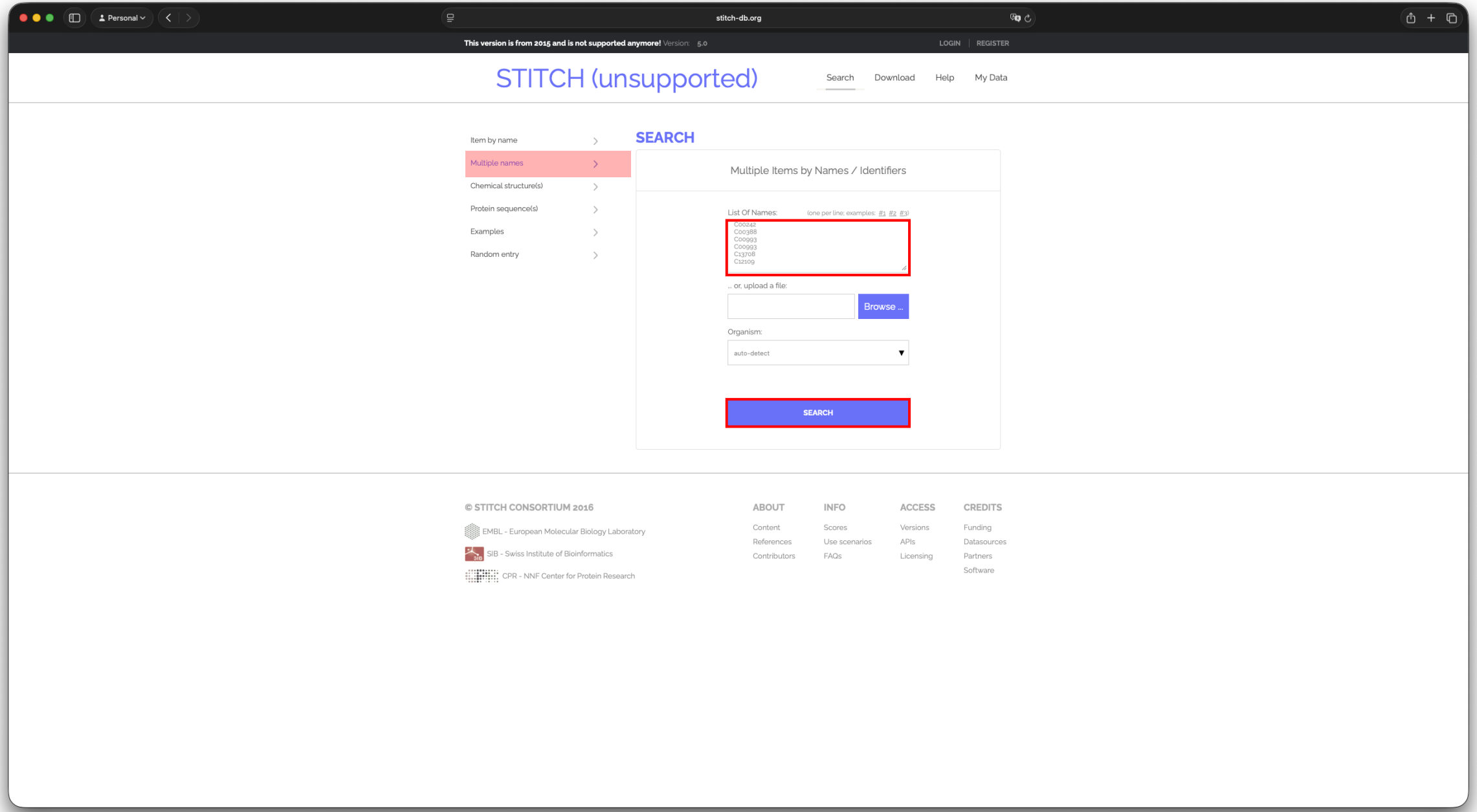
Compounds interaction analysis

STITCH database (Unsupported)

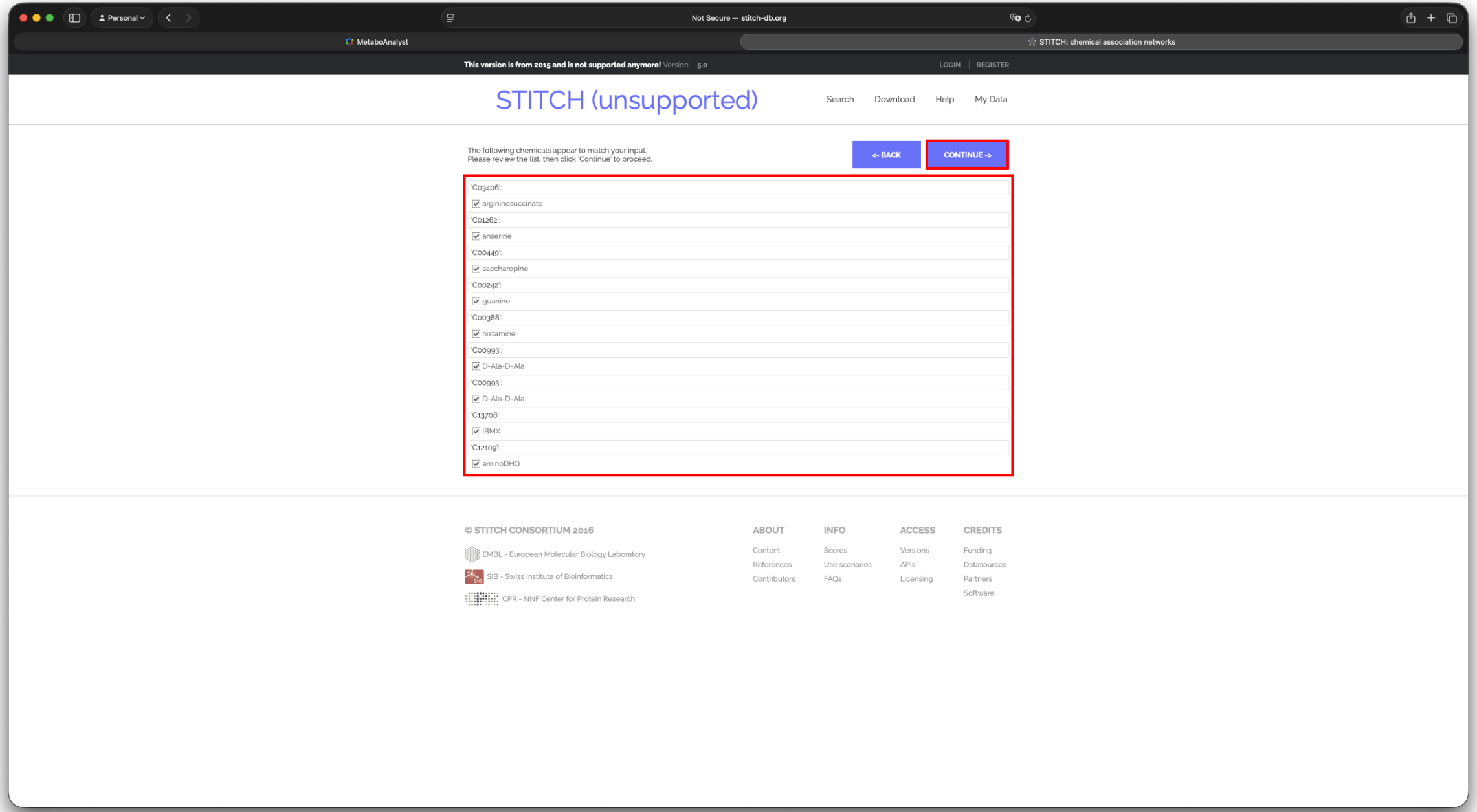
(Optional) STITCH analysis



(Optional) STITCH analysis



(Optional) STITCH analysis



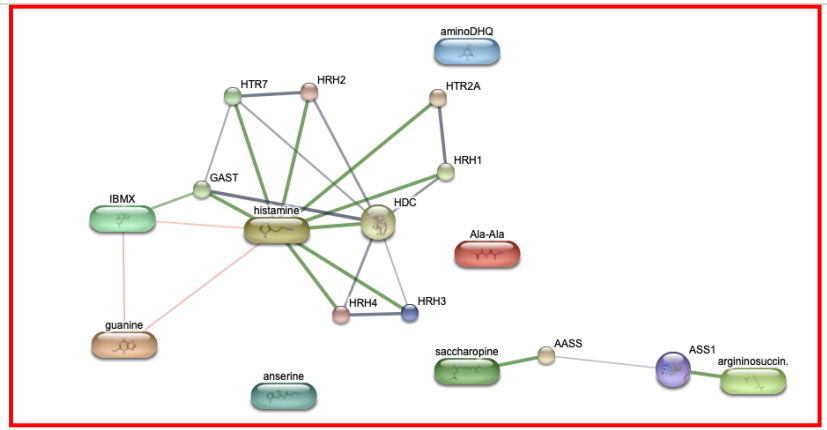
Personal — Not Secure — stitch-db.org

MetaboAnalyst — 8 items (Homo sapiens) - STITCH network view

This version is from 2015 and is not supported anymore! Version: 5.0 — LOGIN | REGISTER

STITCH (unsupported)

Search Download Help My Data



This is the **confidence view**. Stronger associations are represented by thicker lines.
Protein-protein interactions are shown in grey, chemical-protein interactions in green and interactions between chemicals in red.

Viewers › | Legend | Settings › | Analysis › | Tables / Exports › | More | Less

Nodes:

Network nodes represent proteins

splice isoforms or post-translational modifications are collapsed, i.e. each node represents all the proteins produced by a single, protein-coding gene locus.

Node Size

small nodes: protein of unknown 3D structure

large nodes: some 3D structure is known or predicted

Node Color

colored nodes: query proteins and first shell of interactors

white nodes: second shell of interactors

Edges:

Edges represent protein-protein associations

associations are meant to be specific and meaningful, i.e. proteins jointly contribute to a shared function; this does not necessarily mean they are physically binding each other.

Edge Confidence

low (0.150) *high (0.700)*

medium (0.400) *highest (0.900)*

Your Input:

Ala-Ala	*Ala-Ala (160.2 g/mol)*
guanine	*Guanine (G, Gua) is one of the four main nucleobases found in the nucleic acids DNA and RNA, the others being adenine, cytosine, and thymine (uracil in RNA). In DNA, guanine is paired with cytosine. With the formula $C_5H_5N_5O$, guanine is a derivative of purine, consisting of a fused pyrimidine-imidazole ring system with conjugated double bonds. Being unsaturated, the bicyclic molecule is planar. The guanine nucleoside is called guanosine. (151.1 g/mol)*
histamine	*HD-O is an amine derived by enzymatic decarboxylation of histidine. It is a powerful stimulant of gastric secretion, a constrictor of bronchial smooth muscle, a vasodilator, and also a centrally acting neurotransmitter. (111.1 g/mol)*
argininosuccin.	*argininosuccinate; This compound belongs to the class of organic compounds known as l-alpha-amino acids. These are alpha amino acids which have the L-configuration of the alpha-carbon atom. (290.3 g/mol)*
saccharopine	*This compound belongs to the class of organic compounds known as l-alpha-amino acids. These are alpha amino acids which have the L-configuration of the alpha-carbon atom. (276.3 g/mol)*
IBMX	*IBMX, 3-isobutyl-1-methylxanthine, like other methylated xanthine derivatives, is both a # competitive nonselective phosphodiesterase inhibitor which raises intracellular cAMP, activates PKA, inhibits TNF-alpha and leukotriene synthesis, and reduces inflammation and innate immunity As a phosphodiesterase inhibitor, IBMX has IC50 = 2-50 ÎĽM and does not inhibit PDE8 or PDEg. (222.2 g/mol)*
anserine	*A dipeptide comprised of a beta-alanine and a methylhistidine that is found in dietary red meat and has antioxidant activity. (240.3 g/mol)*

Checkpoint V

What does this mean biologically?

Compound interaction

STITCH figure

The relationship between compounds and other molecules

Enhance biological information

Thank you