

Spectra processing, functional integration and covariate adjustment of global metabolomics data using MetaboAnalyst 5.0

Section I: LC-MS spectra data processing & functional analysis

TA: Zhiqiang Pang

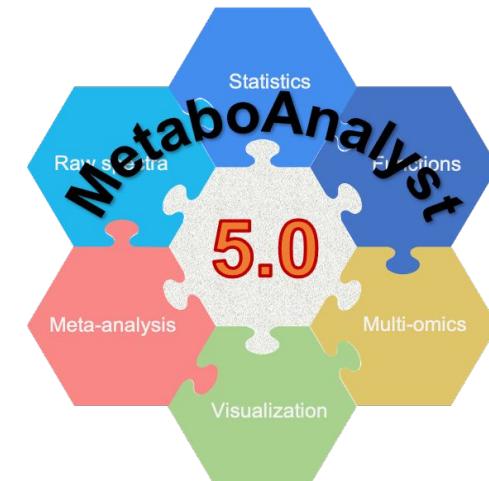
(zhiqiang.pang@mail.mcgill.ca)

18th Annual Conference of the Metabolomics Society

METABOLOMICS 2022

Valencia, Spain | JUNE 19-23

Pre-Conference Workshops



Schedule

Part I: 12:00 PM – 2:00 PM

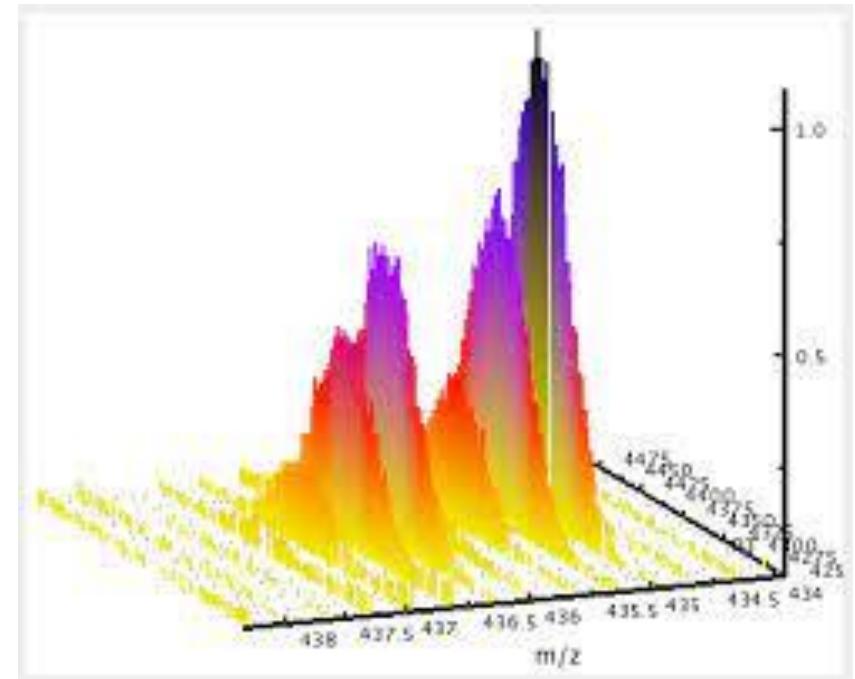
- **12:00 – 12:30:** Opening lecture (Jeff)
- **12:30 – 12:45:** Logistics
- • **12:50 – 1:10:** Section 1: LC-MS spectral processing and functional analysis (Qiang)
- **1:10 – 2:00:** Interactive protocol exercise

Part II: 2:15PM – 4:15PM

- **2:15 – 2:30:** Section 2: multi-omics integration using pathways and networks (Yao)
- **2:30 – 3:10:** Interactive protocol exercise
- **3:10 – 3:30:** Section 3: Complex meta-data lecture (Jessica)
- **3:30 – 4:00:** Interactive protocol exercise
- **4:00 – 4:15:** Summary (Jeff)

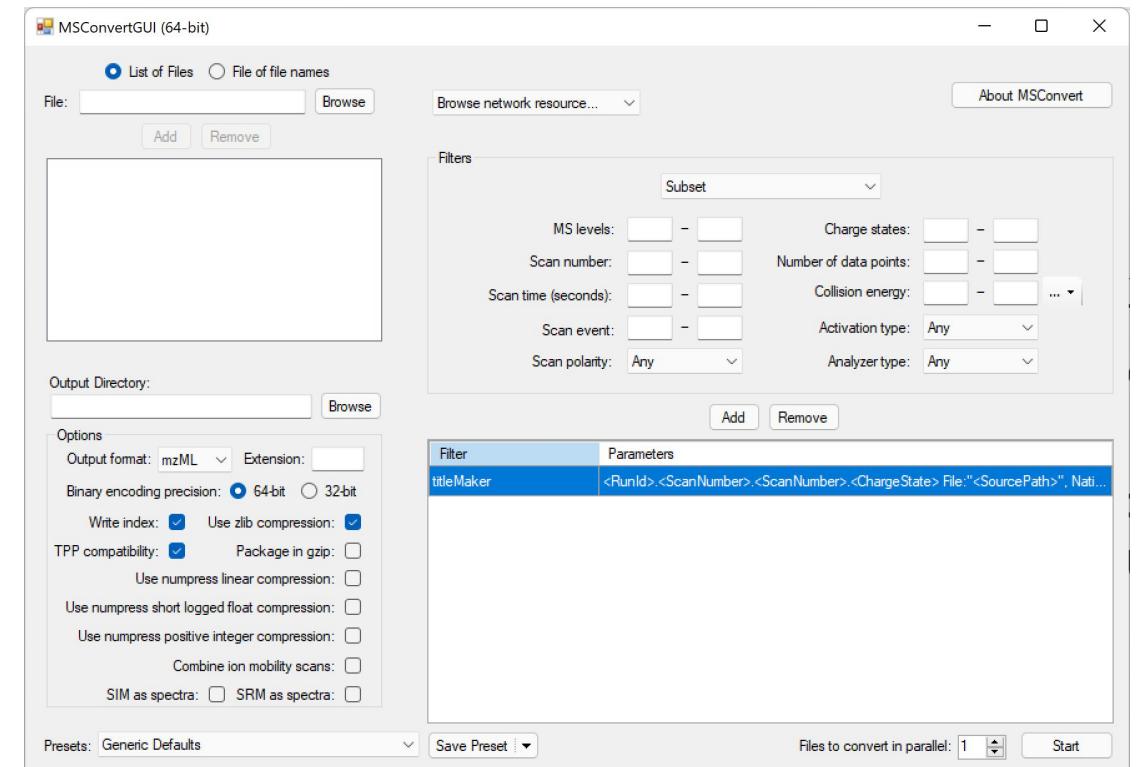
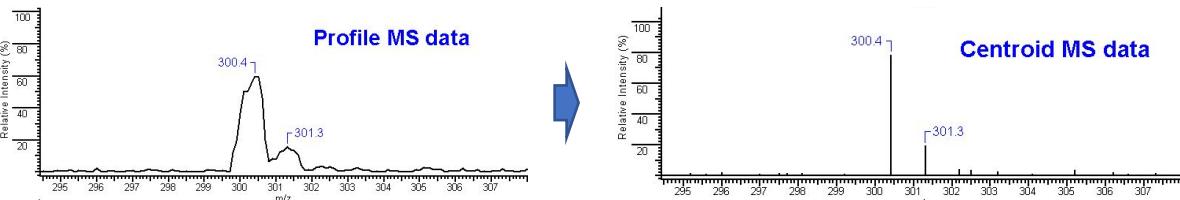
What is raw spectra (pre-)processing?

- Convert the raw spectra data from MS instrument into metabolic features (MS peaks);
- Usually contains 2~3 dimensions. For DI-MS, the raw spectra includes m/z and intensity (2D); while for the LC-MS, the raw spectra data includes m/z , retention time and intensity (3D);
- The most common vendor raw data file formats are .raw/.RAW/.wiff/.d/.D/ etc. They need to be converted into open-source format for further processing with open-source software.

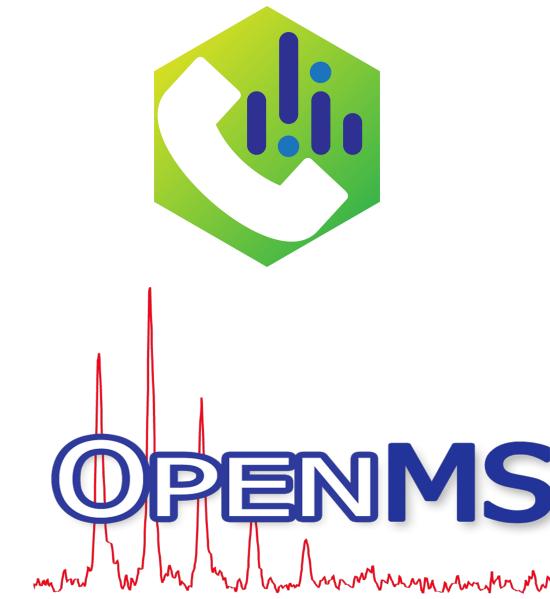
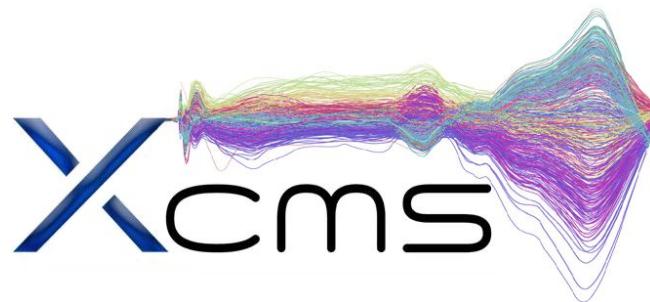


Profile or Centroid?

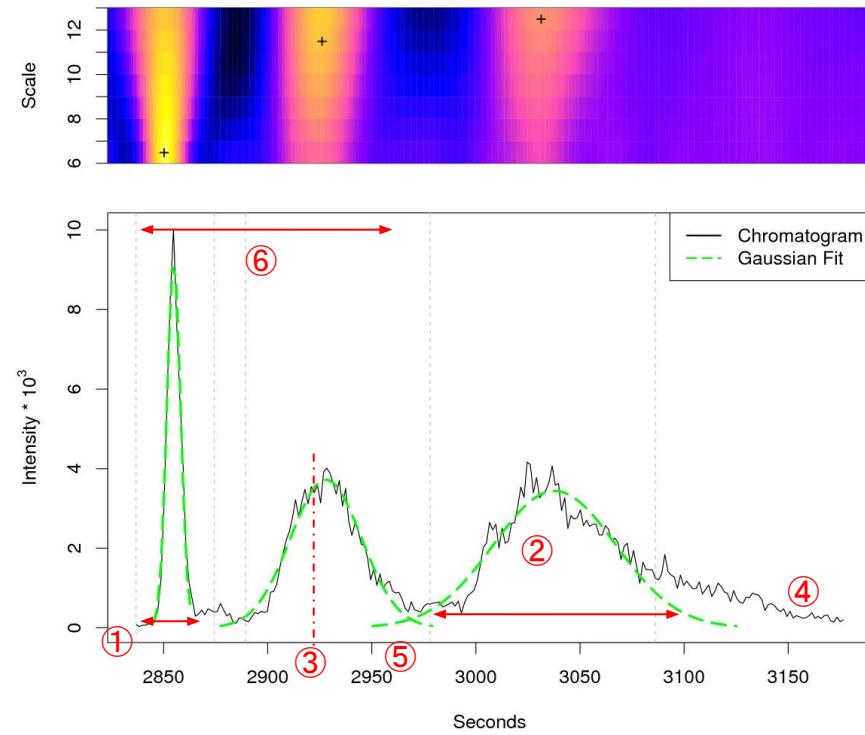
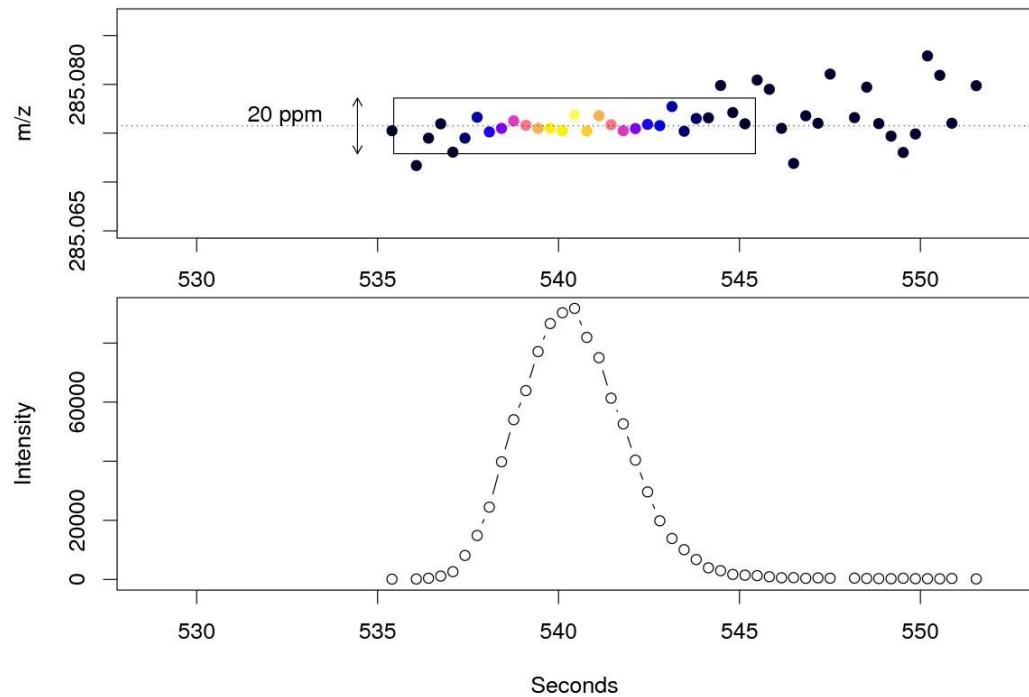
- The vendor raw spectra data is usually in profile format, which is redundant for regular LC-MS based metabolomics analysis;
- We need to convert the MS data into centroid mode to condense the Gaussian Profile peaks into centroids.
- Open-source formats (.mzML/ etc.)..



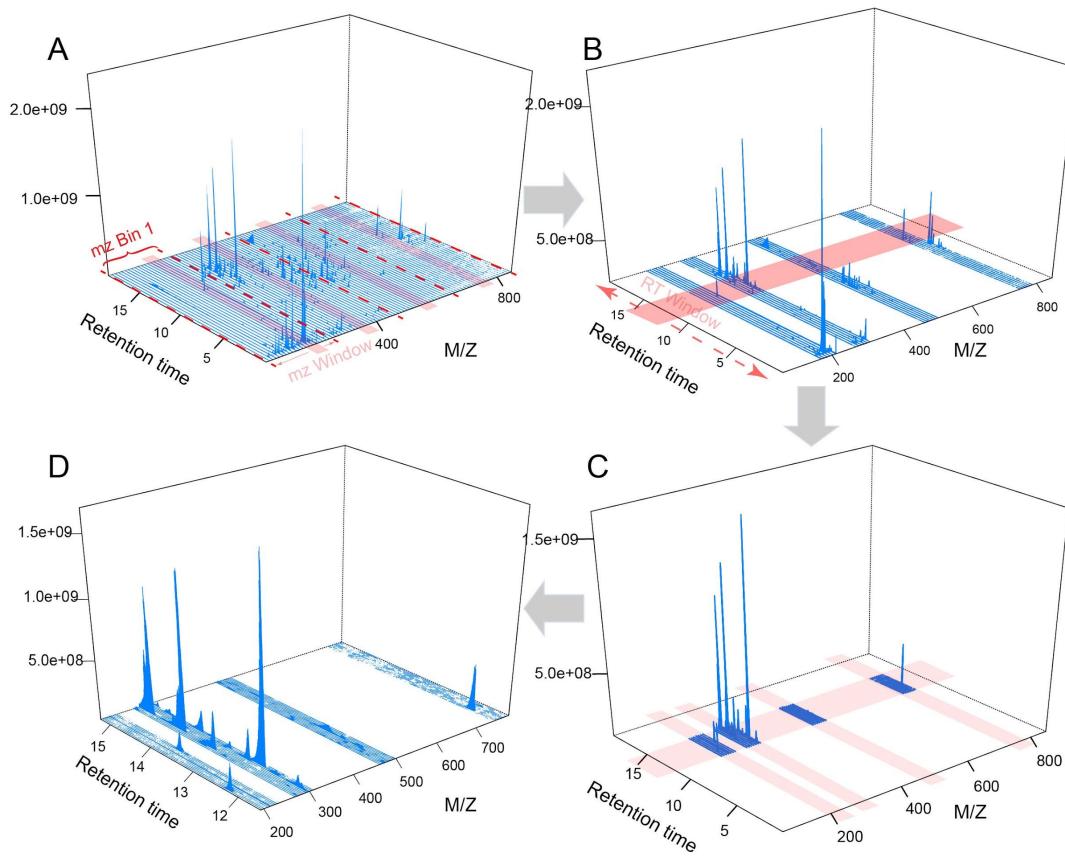
Open-source Software for raw spectra processing..



centWave



ROI Extraction



- Data-driven ROI extraction;
- Regions with high abundance of MS signals;
- Both low intensity peaks as well as high intensity peaks will be retained;

DoE-based Parameter Optimization

DoE -- Central composite design

Order	Peakwith_min	Peakwith_max	mzdiff	snthresh	bw
1	-1	-1	-1	-1	-1
2	1	-1	-1	-1	-1
3	-1	1	-1	-1	-1
...
43	0	0	0	0	1
44	0	0	0	0	0

44 runs

3 level for every parameters (-1, 0, 1)

Relative reliable peaks ratio
(identified by their isotopes)

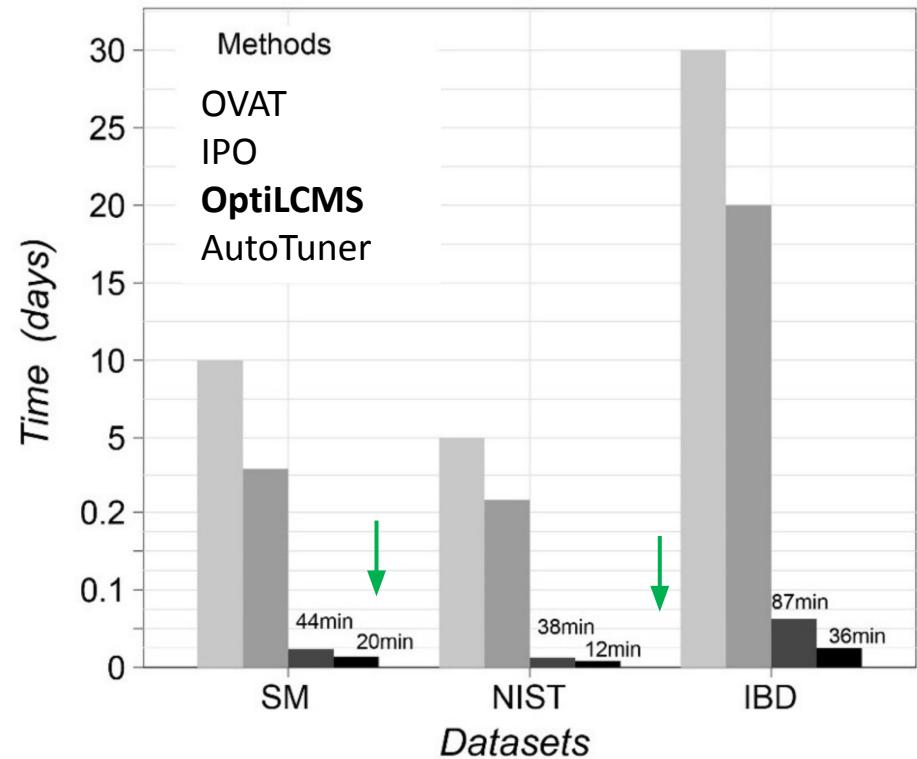
$$QS = \frac{RP^{3/2}}{'all\ peaks' - LIP}$$

A co-efficient describing the stability of
a grouped feature

Gaussian peaks ratio

- The most important parameters are evaluated with 44 DoE runs
- Instead of $3^8 = 6561$ one-variable-at-a-time runs.

Performance Evaluation - Speed



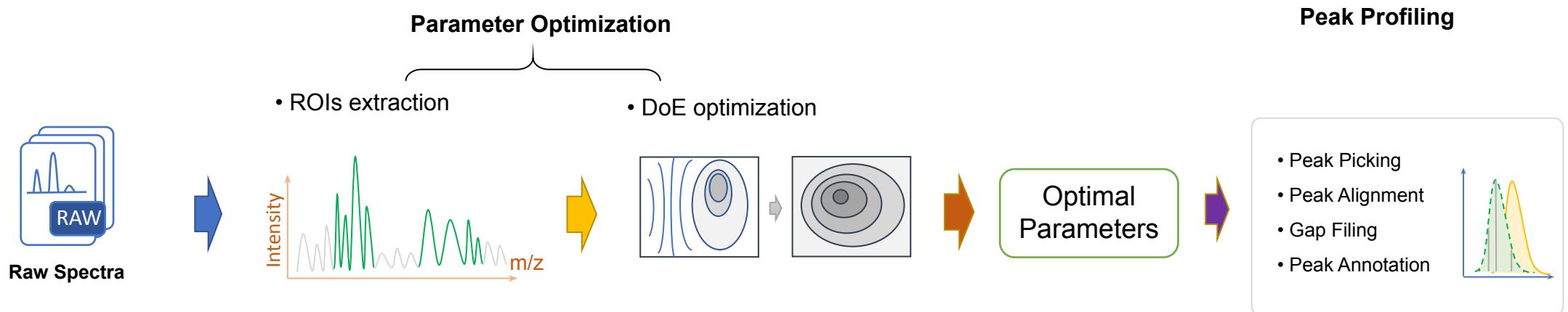
- + 3 datasets: Standard Mixture (SM), NIST-SRM 1950 and IBD data from iHMP2.

Bench markings

	Default	Optimized		Default	Optimized
Total peaks	2,492	2,423	Total peaks	4,344	5,113 (+ 17.7%)
Isotopes / Adducts	667 (26.8%)	1,112 (45.9%)	Isotopes	760	1,274 (+ 67.6%)
Formula Assigned	663	762	Adducts	927	1,132 (+ 22.1%)
Potential compounds	1,085	1,692	Formulas assigned	632	687 (+ 8.7%)
Variance (PC1 + PC2)	37%	50%	Potential compound matches	1,587	1,803 (+ 13.6%)
Significant peaks	855	1,091	Variance explained (PC1 + PC2)	76.5%	81.3% (+ 4.8%)

Workflow Overview

- Our optimization approach is designed to extract a region abundant with MS signals for a design of experiment (DoE)-based optimization.



← → C metaboanalyst.ca/MetaboAnalyst/ModuleView.xhtml



MetaboAnalyst 5.0 - user-friendly, streamlined metabolomics data analysis

[Home](#)

[Data Formats](#)

[Tutorials](#)

[OmicsForum](#)

[APIs](#)

[Update History](#)

[MetaboAnalystR](#)

[Contact](#)

[User Stats](#)

[Publications](#)

[COVID-19 Data](#)

[About](#)

 GenomeCanada

 GenomeQuébec



Module Overview

Input Data Type	Available Modules (click on a module to proceed, or scroll down for more details)					
Raw Spectra (mzML, mzXML or mzData)	LC-MS Spectra Processing					
MS Peaks (peak list or intensity table)			Functional Analysis	Functional Meta-analysis		
Annotated Features (compound list or table)	Enrichment Analysis	Pathway Analysis	Joint-Pathway Analysis	Network Analysis		
Generic Format (.csv or .txt table files)	Statistical Analysis [one factor]	Statistical Analysis [metadata table]	Biomarker Analysis	Statistical Meta-analysis	Power Analysis	Other Utilities

[>> 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 dendograms 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](#)

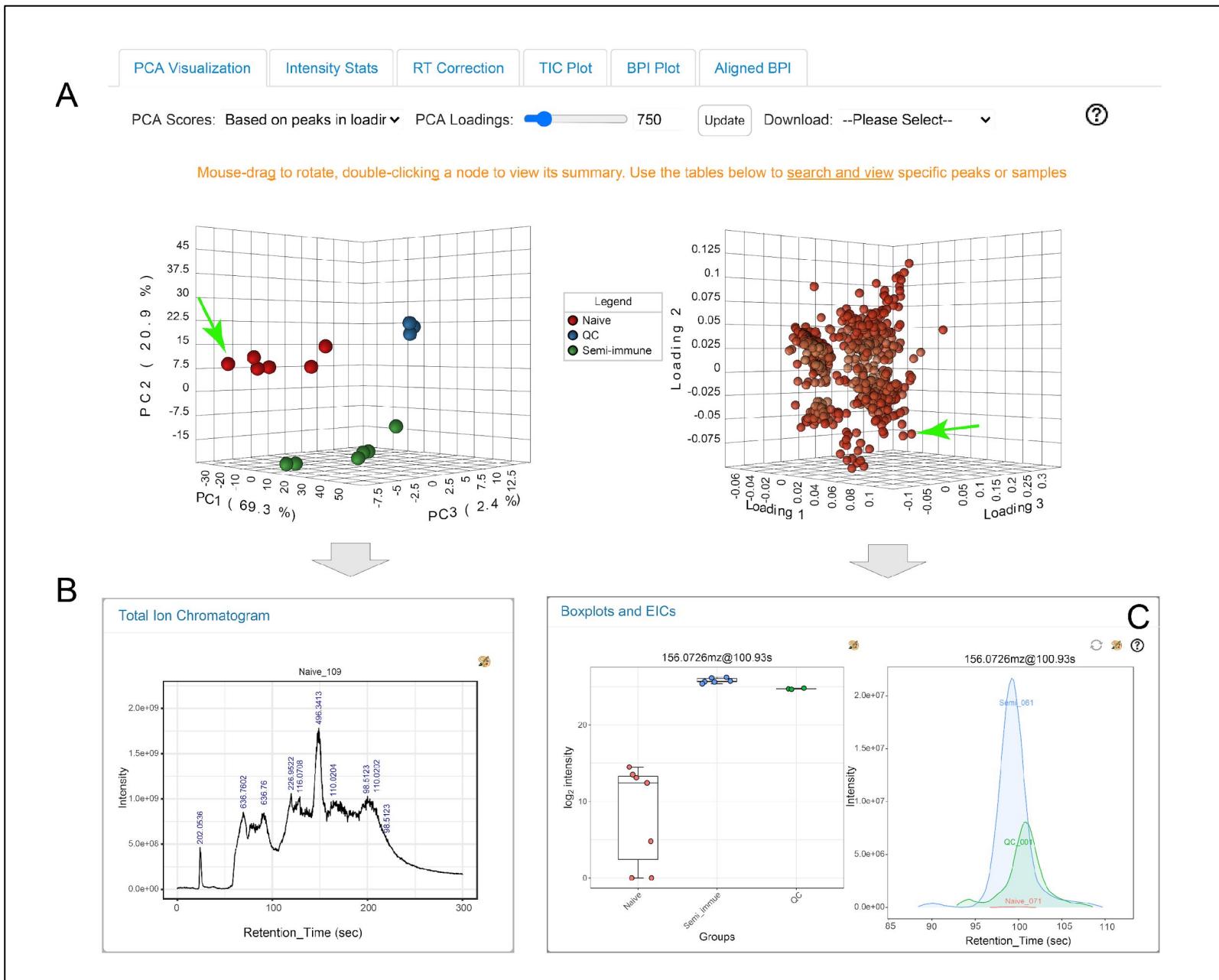
This module performs various biomarker analyses based on receiver operating characteristic (ROC) curves for a single or multiple biomarkers using well-established methods. It also allows users to manually specify biomarker models and perform new sample prediction.

Xia Lab @ McGill (last updated 2022-06-17)

Functional Utilities

- Raw Data Uploading (.mzML/.mzXML/.mzData/.cdf);
- Centroiding on the fly;
- Parameters Optimization (automatically);
- Peak Profiling (Peak Picking/Alignment/Gap filling);
- Peak Annotation (Adducts + Isotopes);
- Putative Compound Mapping;
- Result visualization...

Result Demo..



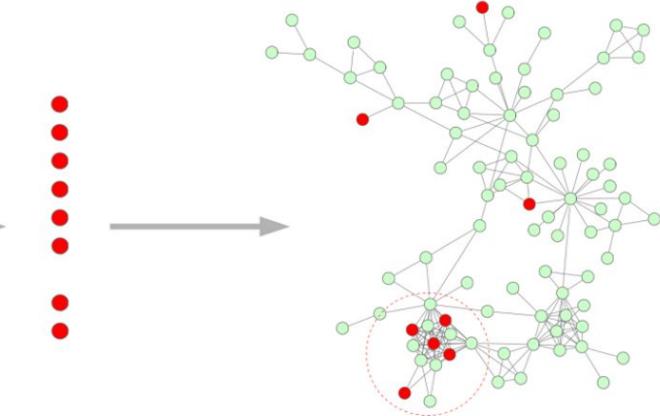
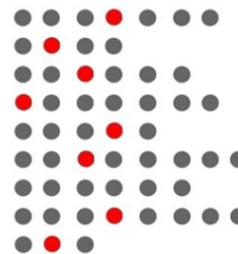
Functional Analysis

Functions describe the collective behavior of groups of molecules acting together to complete a task (e.g., lysine degradation).

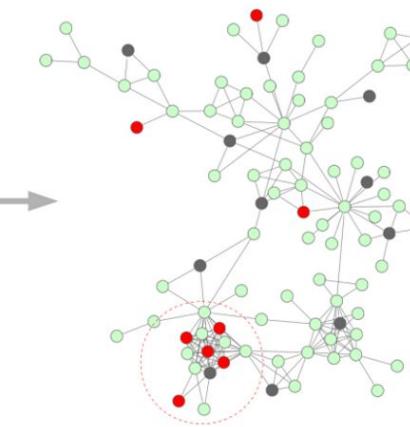
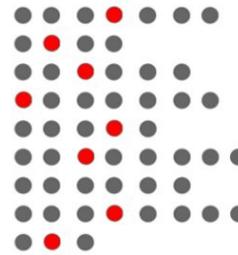
The goal of enrichment analysis is to evaluate whether the members involved in a particular tasks have more consistent behaviors (e.g., more changes larger than normal) compared with random variations.

Mummichog provides a practical solution of one-step functional analysis, bypassing the bottleneck of upfront metabolite identification.

Conventional approach



mummichog



Implementation in MetaboAnalyst

The screenshot shows the MetaboAnalyst 5.0 website at metaboanalyst.ca/MetaboAnalyst/ModuleView.xhtml. The page features a navigation bar with links to Home, Data Formats, Tutorials, OmicsForum, APIs, Update History, MetaboAnalystR, Contact, User Stats, Publications, COVID-19 Data, and About. It also includes logos for GenomeCanada and GenomeQuébec, and the NIH.

The main content area displays the "Module Overview" grid. The grid has "Input Data Type" columns (Raw Spectra, MS Peaks, Annotated Features, Generic Format) and "Available Modules" rows. The "Functional Analysis" row is highlighted in red, while other rows are light blue. The "Statistical Analysis [one factor]" and "Statistical Analysis [metadata table]" modules are shown under the Generic Format input type, and the "Biomarker Analysis" module is shown under the Annotated Features input type.

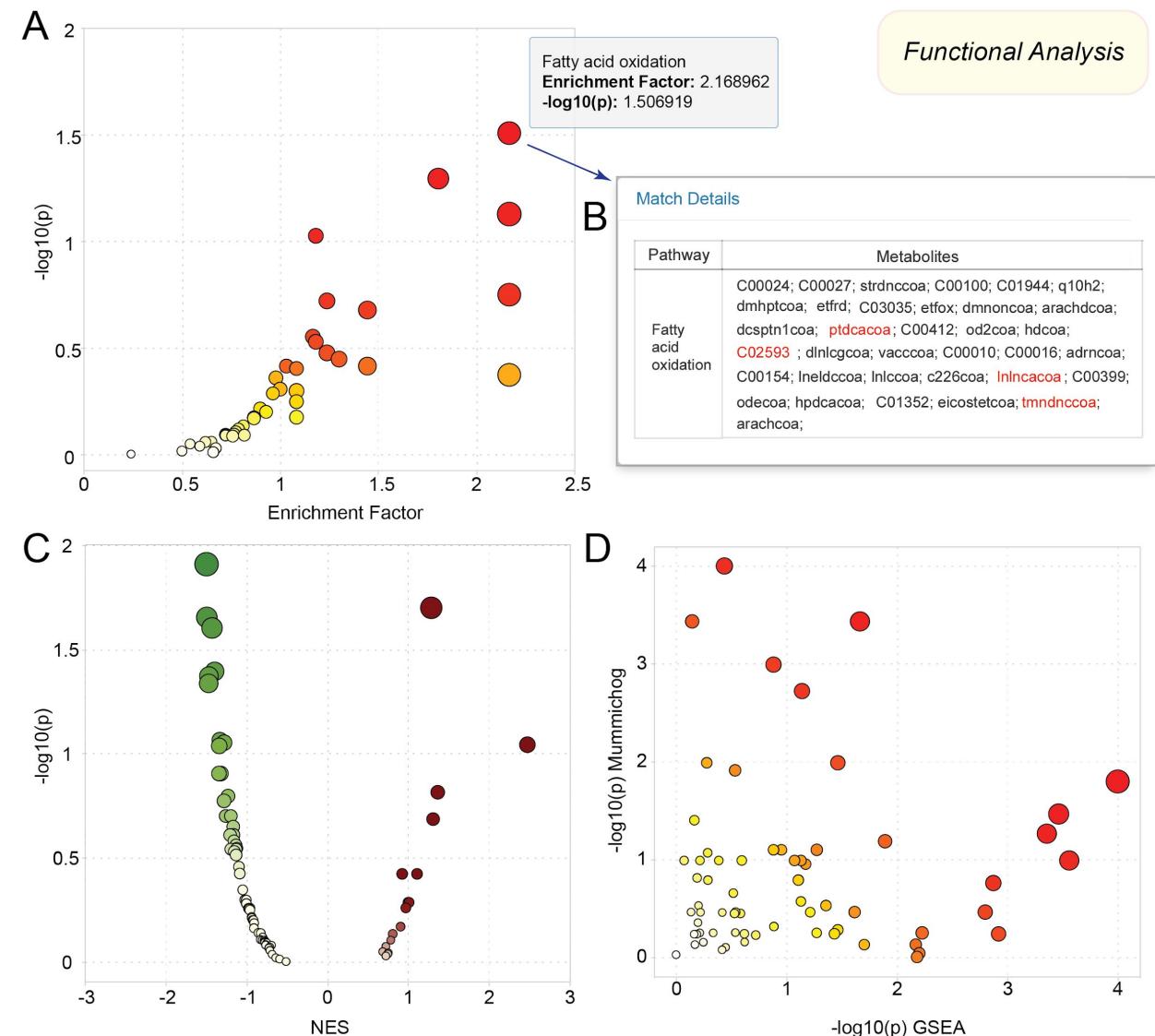
Below the grid, three boxes provide detailed descriptions of these modules:

- >> 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 dendograms 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**
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.

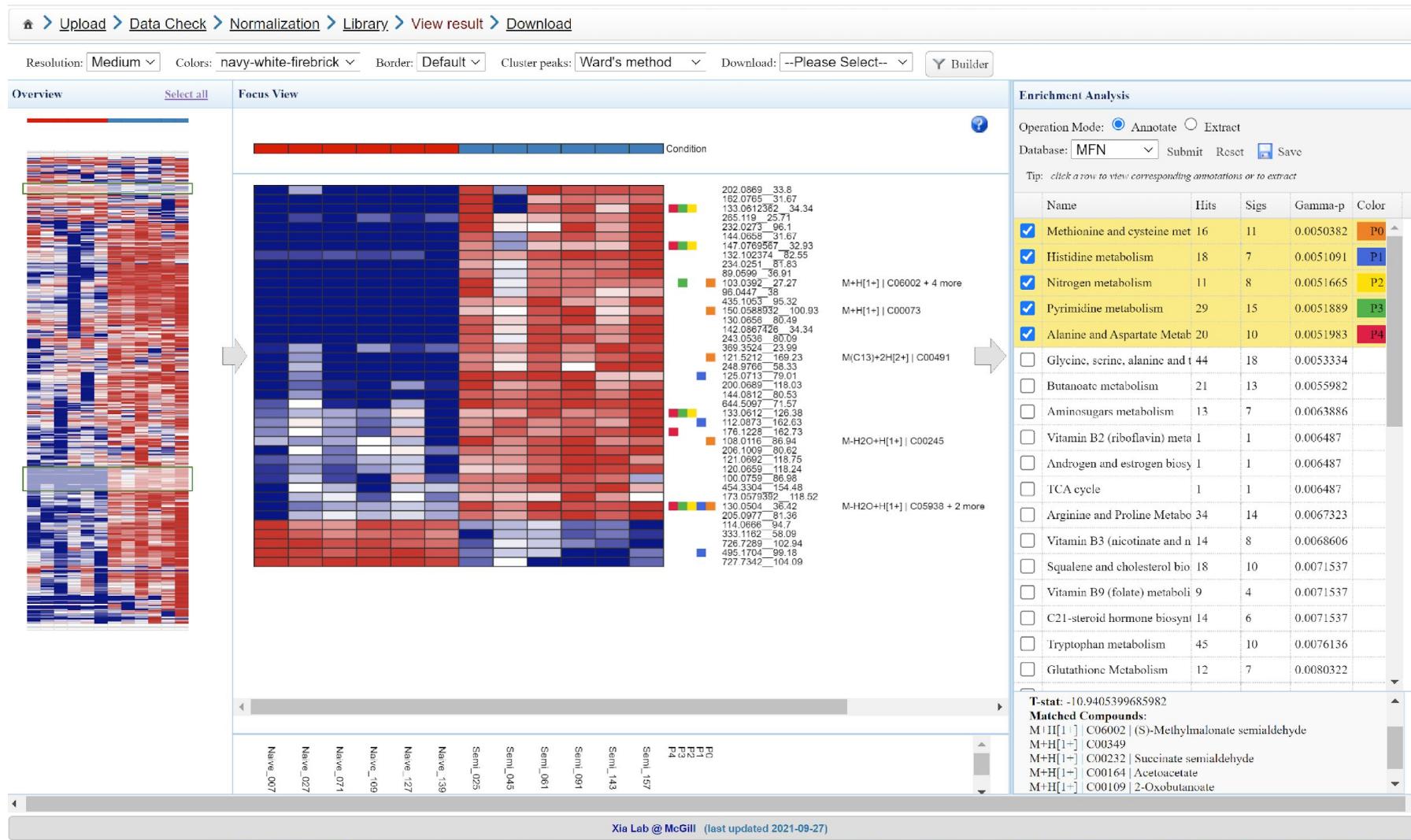
At the bottom, a footer note states: "Xia Lab @ McGill (last updated 2022-08-17)".

Functional Analysis Results

- **Mummichog**: implements an over-representation analysis (ORA) method to evaluate pathway-level enrichment based on significant peaks. Users need to specify a pre-defined cutoff.
- **GSEA**: considers the overall ranks of features without using a significance cutoff, and is claimed to be able to detect subtle and consistent changes.
- **Integrated**: combines both Mummichog and GSEA together with Fisher's Method.



Metabolic Pattern-based functional analysis



From raw spectra into biological insight

The screenshot shows a web browser window for the MetaboAnalyst platform. The URL is dev.metaboanalyst.ca/MetaboAnalyst/Secure/details/DownloadView.xhtml. The interface includes a navigation bar with back, forward, and search icons, as well as a "Show R Commands" button.

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**

General Statistics

- Statistical Analysis [one factor]
- Biomarker Analysis
- Statistical Analysis [metadata table]
- Power Analysis
- Enrichment Analysis
- Pathway Analysis

Targeted Metabolomics

- Functional Analysis

Global Metabolomics

GO!

The sidebar on the left contains links: Home, Upload, Spectra check, Spectra processing, Job status, Spectra result, Download, and Exit. The "Upload" link is highlighted.

Learning Questions..

- Generate an EIC plot (overlay of at least one sample from the three groups) for the most significant peak?
- What is the most significantly perturbed pathway for Native vs. Semi-immune subjects when using Malaria Example?

Tutorials

- Publication: <https://www.nature.com/articles/s41596-022-00710-w>
- Or our manuscript: <https://www.dropbox.com/s/7184c4dheeiiz2p/NP-MetaboAnalyst-2022.pdf?dl=0>
 - Stage 1: LC–HRMS raw spectra processing
 - Stage 2: functional analysis of LC–HRMS peaks

Questions?

- <https://www.omicsforum.ca/>
- If your question is not covered, please create a new topic – we will try to answer them in the coming days.

Caution:

1. For raw spectra processing, you are strongly encouraged to use 1st example rather than the 2nd one to avoid waiting in queue for learning purpose;
2. Avoid downloading and uploading any example raw spectra data due to the limited bandwidth.
3. Default MetaboAnalyst includes www.metaboanalyst.ca and dev.metaboanalyst.ca, please use backup node genap.metaboanalyst.ca ONLY if the defaults are not accessible.