



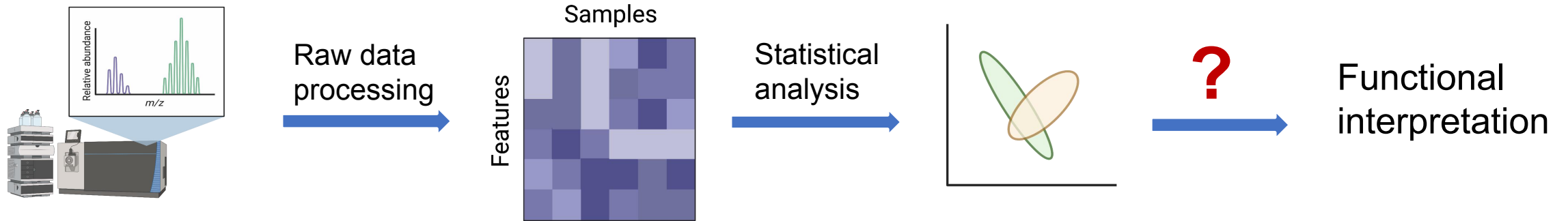
# Functional interpretation of LC-MS peaks using **MetaboAnalyst 5.0**

**Yao Lu, PhD Candidate**

[yao.lu5@mail.mcgill.ca](mailto:yao.lu5@mail.mcgill.ca)

McGill University, Montreal, QC  
Canada

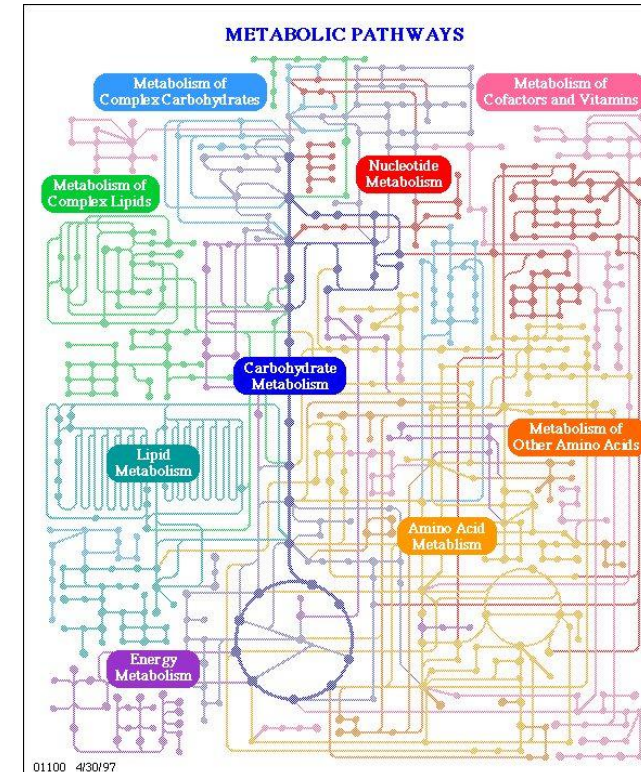
# General untargeted metabolomics pipeline:



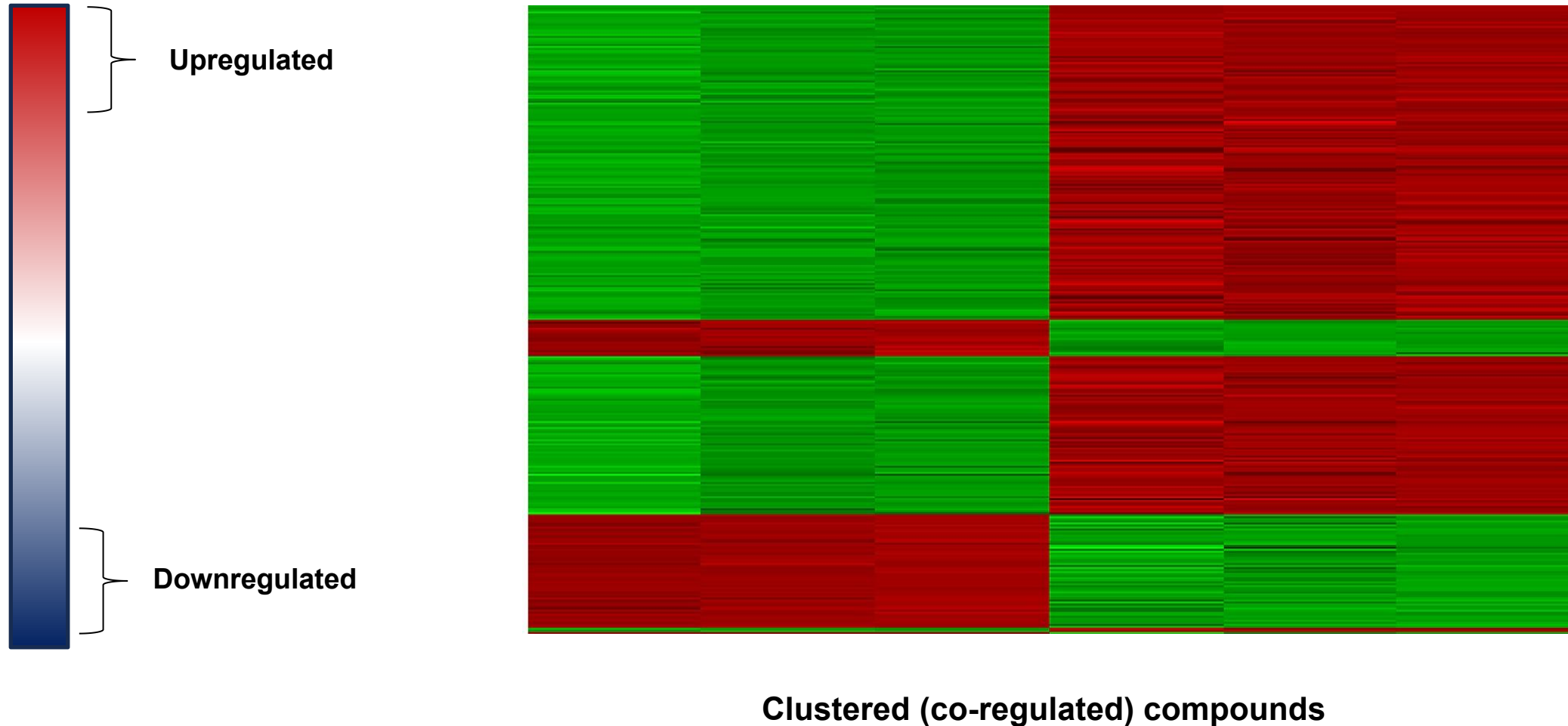
# From compounds / peaks to functions

Compound	Retention Time (min)	Conc. in Urine (μM)
Dns -o-phospho -L-serine	0.92	<D.L. *
Dns -o-phospho -L-tyrosine	0.95	<D.L.
Dns -adnosine monophosphate	0.99	<D.L.
<b>Dns -o-phosphoethanolamine</b>	1.06	16
<b>Dns -glucosamine</b>	1.06	22
Dns -o-phospho -L-threonine	1.09	<D.L.
Dns -6-dimet hylamine purine	1.20	<D.L.
<b>Dns -3-methyl -histidine</b>	1.22	80
<b>Dns -taurine</b>	1.25	834
<b>Dns -carnosine</b>	1.34	28
<b>Dns -Arg</b>	1.53	36
<b>Dns -Asn</b>	1.55	133
<b>Dns -hypotaurine</b>	1.58	10
<b>Dns -homocarnosine</b>	1.61	3.9
Dns -guanidine	1.62	<D.L.
<b>Dns -Gln</b>	1.72	633
<b>Dns -allantoin</b>	1.83	3.8
<b>Dns -L-citrulline</b>	1.87	2.9

m.z	p.value	t.score
304.2979	1.02E-10	14.7179316
177.1024	1.62E-10	14.2666
345.0277	1.72E-10	-14.209195
491.0325	1.83E-10	-14.146348
258.0048	2.17E-10	-13.987636
483.1205	2.22E-10	-13.967634
694.9937	2.81E-10	-13.745172
270.9767	3.27E-10	13.6060705
371.604	3.53E-10	-13.534483
316.5773	3.71E-10	13.4893333
451.0505	4.04E-10	-13.412347
257.0543	4.09E-10	-13.401887
762.9787	4.71E-10	-13.274141
231.0422	5.14E-10	-13.195677
614.0797	6.11E-10	13.0407288
213.0066	6.79E-10	12.9471714
416.2122	7.92E-10	12.8119438



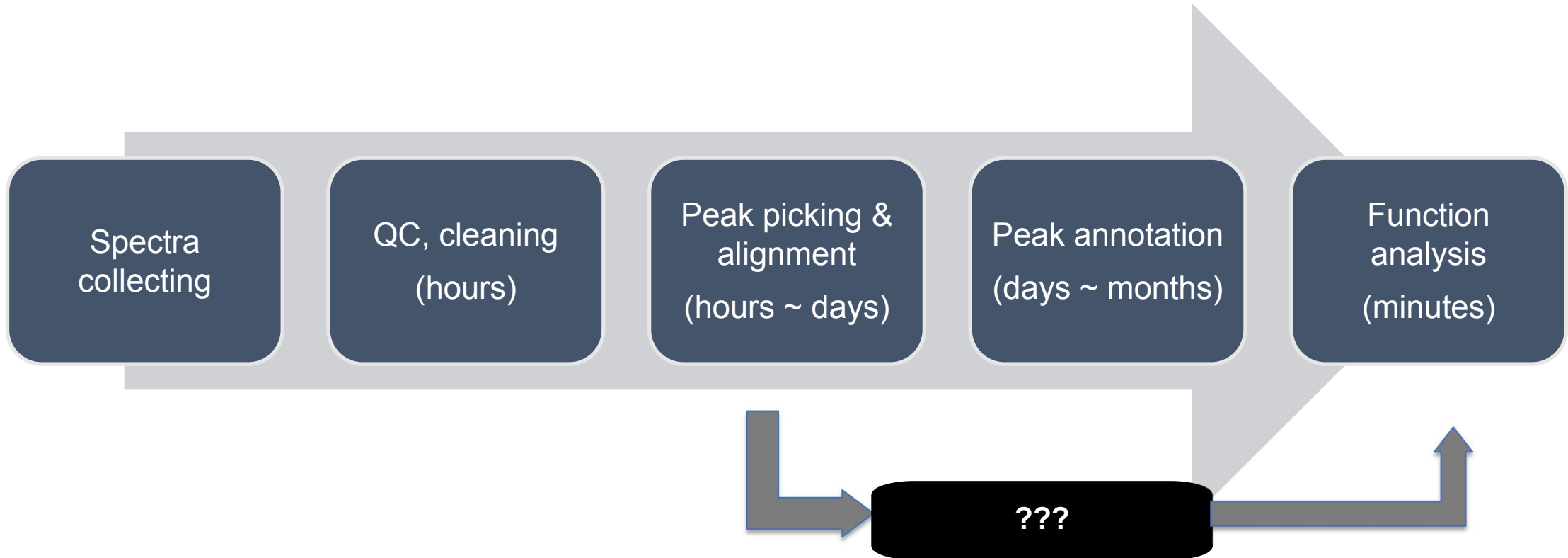
# Co-regulated metabolites are often related



# How do we detect functional changes?

- Functions are **group behaviors**
  - Multiple molecules work together to finish a task
- To test whether members involved in a function have more consistent changes (i.e. most move in one direction) compared to random changes
- Enrichment analysis
  - Over representation analysis (ORA)
    - Starting from significant compounds
  - Gene Set Enrichment Analysis (GSEA)
    - A complete ranked compound list

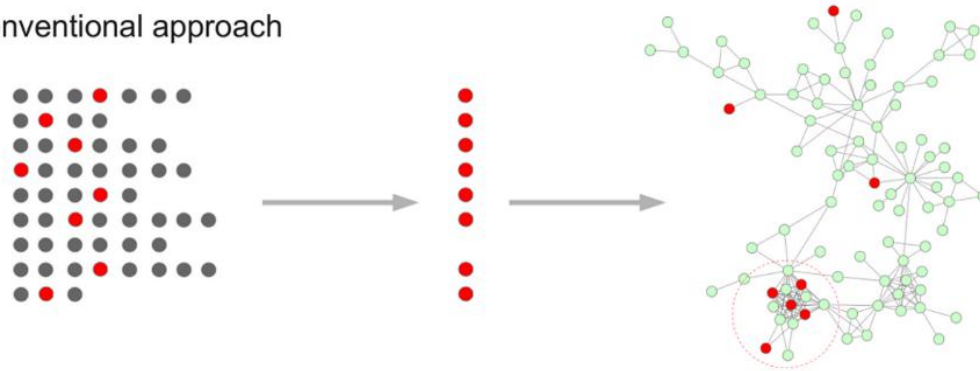
# Conventional approaches requires compound ID known



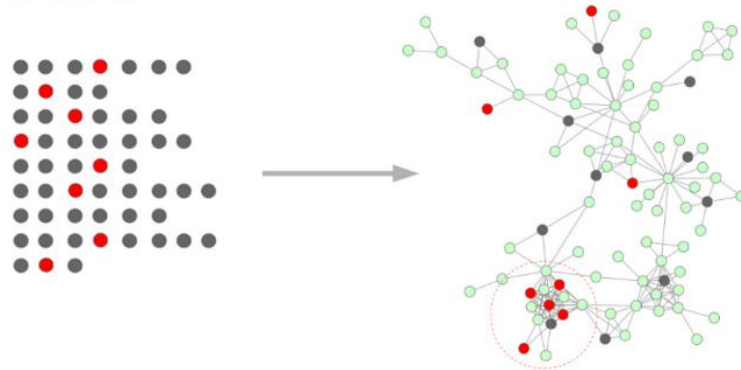
**Can we perform enrichment analysis direct from peaks?**

# Mummichog

Conventional approach



mummichog



- We have a general idea what each peak is
- We have an idea of how each compound in a pathway could look like as peaks
- Mummichog: map many possible matches to a background pathway and look for consistent functional patterns
- Pathway-level results will be robust

# How does mumichog work?

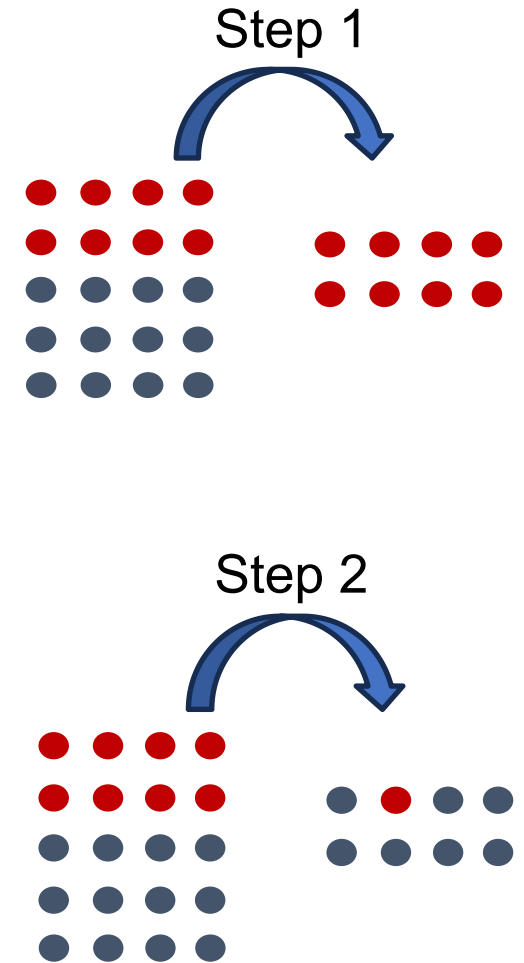
## Input requirement:

- Significant list: selected by t-test or fold change
- Reference list: all features detected

**Step 1:** Match the peaks to tentative metabolites. Looked up all the significant metabolites in each pathway and calculate the p-value using Fisher's exact test

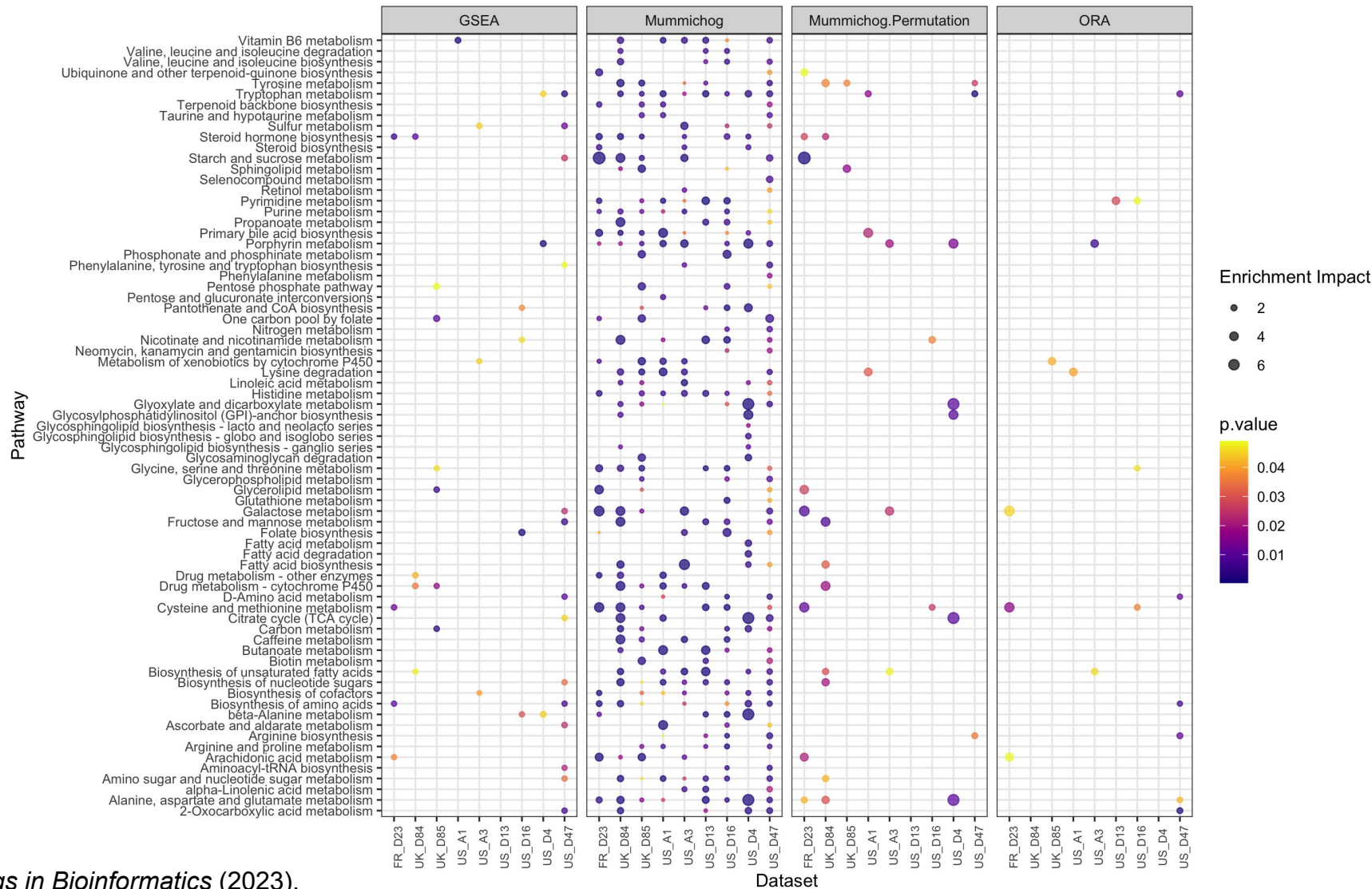
**Step 2:** Randomly pull features with the length same as the significant ones, and repeat Step 1 for thousand times

**Step 3:** Test if certain pathways are enriched in the significant peaks as compared to null models (Gamma distribution)






# Mummichog – sensitive, robust enrichment analysis for metabolomics



# Implementation in MetaboAnalyst



## 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](#)

### 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 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](#)

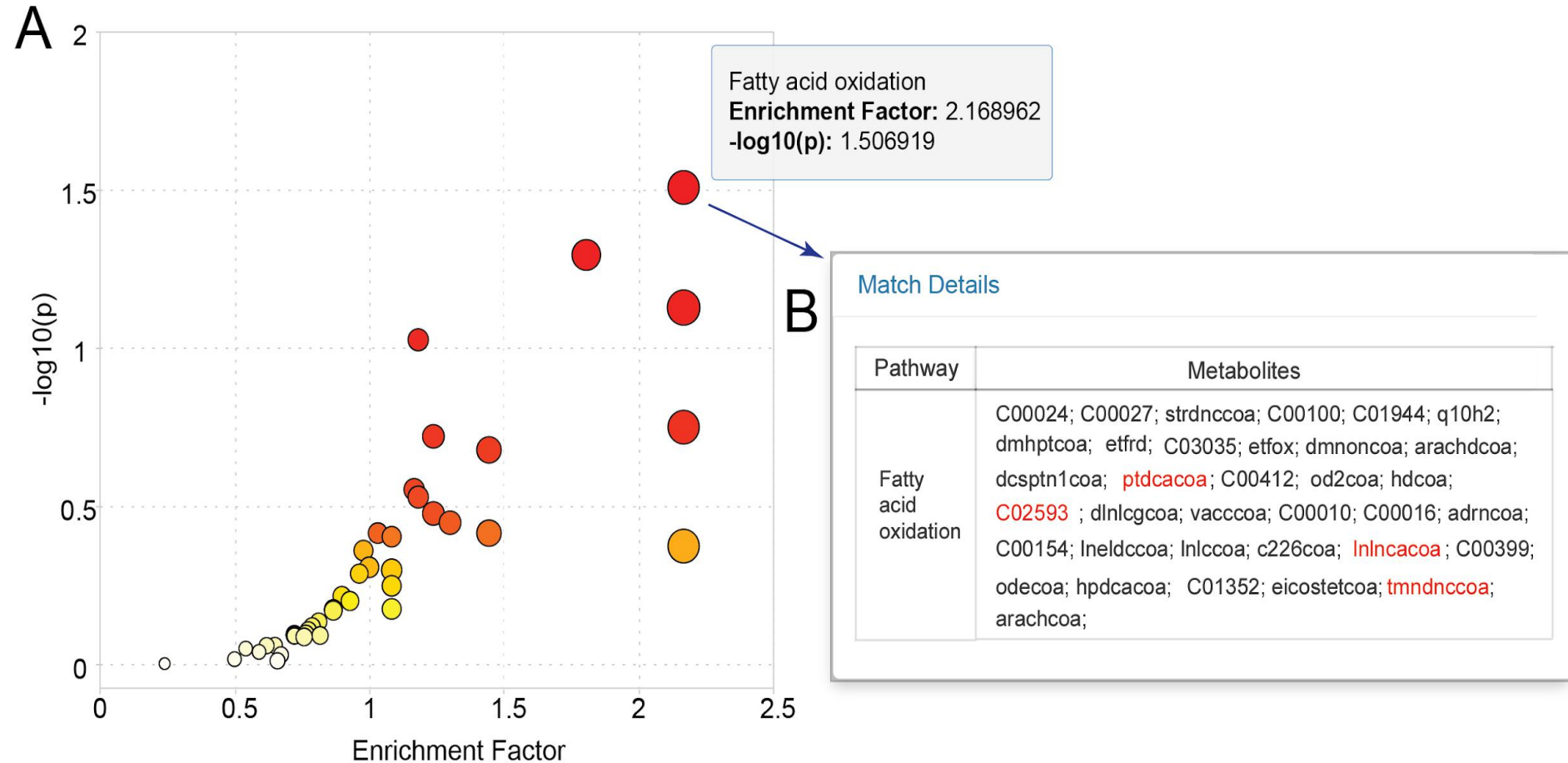
This module performs various biomarker analyses based on receiver operating characteristic (ROC) curves for a single or multiple biomarkers using well-established methods. It also allows users to manually specify biomarker models and perform new sample prediction.

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

# Critical: Input preparation

- LC - **high-resolution** MS (HR-MS)
  - Orbitrap, Q-TOF
  - Reason: putative annotation needs to be approximately correct (better guess leads to more accurate functional analysis)
- Needs to be **complete** peak list or peak intensity table
  - Not just significant peaks
  - Reason: mummichog using permutation to estimate the null/background distribution
- In general, the algorithm works well for **> 3000** peaks (assuming human plasma samples).

# From ranked peak lists to functions



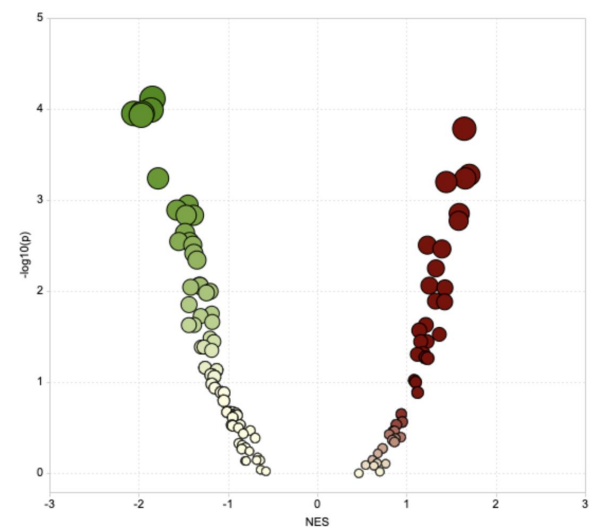
# How to interpret result table (key parameters)

Pathway Name	Total ↑↓	Hits (all) ↑↓	Hits (sig.) ↑↓	Expected ↑↓	P(Fisher) ↑↓	P(Gamma) ↑↓	Details
Vitamin E metabolism	54	38	15	5.0563	0.030024	0.025523	<a href="#">View</a>
Carnitine shuttle	72	25	10	6.7418	0.06554	0.028334	<a href="#">View</a>

- **Total:** the total number of the given pathway
- **Hits (all):** all the peaks mapped to the pathway
- **Hits (sig):** all the significant peaks mapped to the pathway
- **Expected:** The expected number of metabolite hits in the pathway.
- **P(Fischer):** The Fisher's exact p-value for the pathway
- **P(Gamma):** P-values derived from Gamma distribution based on permutation tests for the pathway.



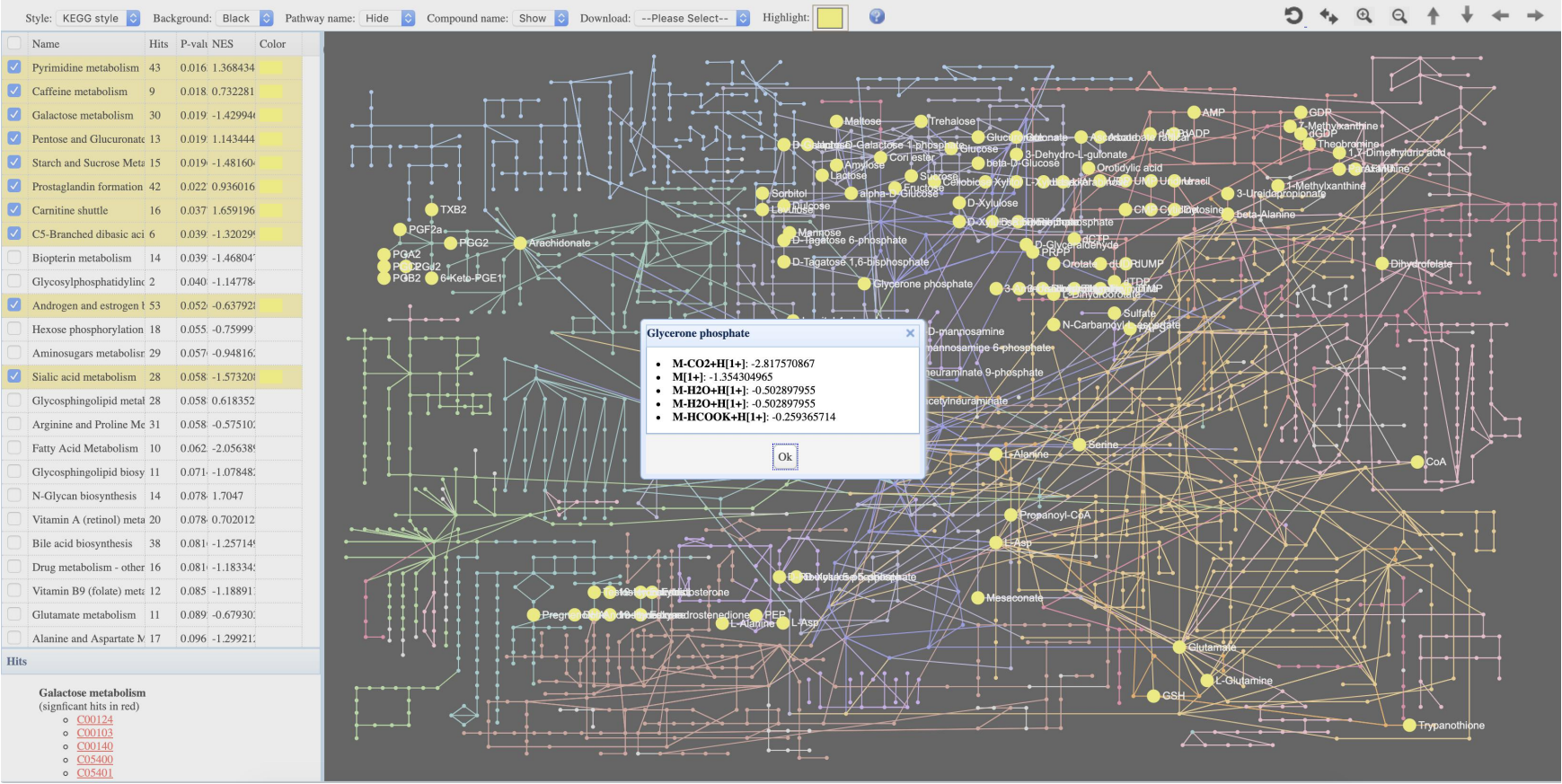
# From pathways to “candidate” compounds



The red compounds indicate all potential matches from the user's input to the selected pathway.

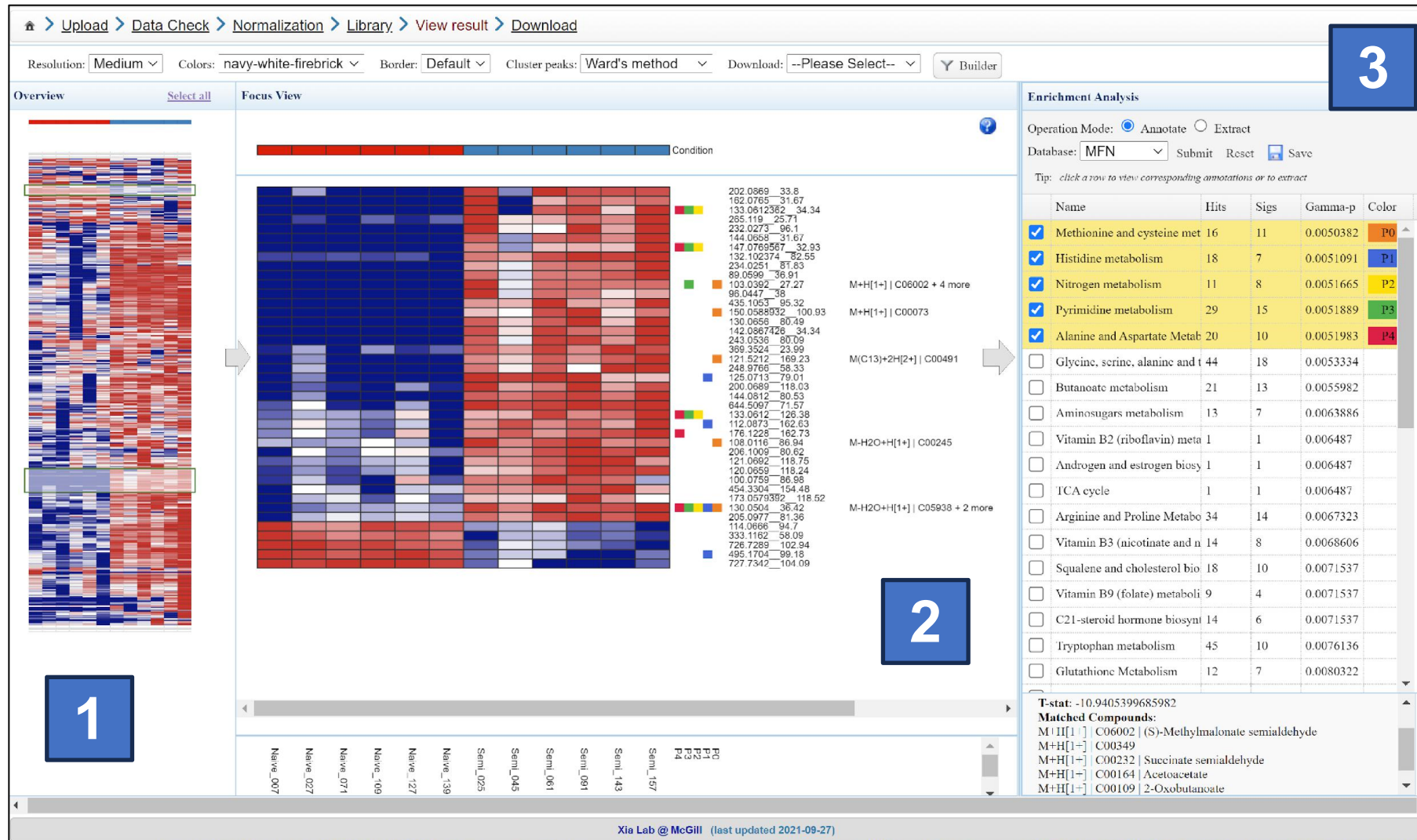
Pathway	Metabolites
Prostaglandin formation from arachidonate	<b>CE5703</b> ; CE5700; CE6238; <b>C00425</b> ; CE5707; <b>CE5704</b> ; <b>CE5705</b> ; CE5724; CE5708; CE1243; CE6236; C00584; <b>CE5535</b> ; <b>C00959</b> ; <b>CE5537</b> ; <b>CE5925</b> ; <b>CE5533</b> ; C13856; <b>CE0955</b> ; <b>CE4980</b> ; C01312; CE5304; <b>C04686</b> ; CE4938; <b>C05963</b> ; <b>C04741</b> ; <b>CE5931</b> ; <b>CE0737</b> ; <b>C01041</b> ; CE4878; <b>CE5926</b> ; <b>C00051</b> ; <b>CE5924</b> ; <b>C00219</b> ; C00704; <b>CE5928</b> ; <b>CE5929</b> ; <b>C00020</b> ; CE4876; <b>C00010</b> ; <b>C05954</b> ; C00024; <b>C05962</b> ; C00030; <b>C00072</b> ; CE7107; CE7105; C00028; C00696; <b>CE1447</b> ; <b>C04685</b> ; <b>C11695</b> ; CE6244; CE6245; CE6242; CE6243; <b>CE6240</b> ; CE6241; <b>C05953</b> ; C00027; <b>C05956</b> ; <b>C05957</b> ; C00427; <b>C05955</b> ; CE4877; CE6235; <b>C05959</b> ; <b>C00639</b> ; CE5730; <b>CE5534</b> ; <b>CE5930</b> ; C00282; <b>CE3481</b> ; C00189; CE7054; C11304; C02198; <b>CE5828</b>

OK



Both Mummichog and GSEA are available in MetaboAnalyst 5.0

# From patterns to functions



**Demo**



# What's coming

- Integrating topology
- Integrating MS/MS

# Tips:

- If you want to learn about the functional analysis in MetaboAnalyst, please refer to:

[https://www.xialab.ca/api/download/metaboanalyst/2\\_Functional\\_Analysis.pdf](https://www.xialab.ca/api/download/metaboanalyst/2_Functional_Analysis.pdf)

[https://www.xialab.ca/api/download/metaboanalyst/8\\_Video\\_Tutorial.pdf](https://www.xialab.ca/api/download/metaboanalyst/8_Video_Tutorial.pdf)

<https://www.nature.com/articles/s41596-022-00710-w.pdf>

- If you want to know more about the performance of enrichment methods and parameters selection, please refer to:

<https://academic.oup.com/bib/article-abstract/24/1/bbac553/6960976?redirectedFrom=fulltext>