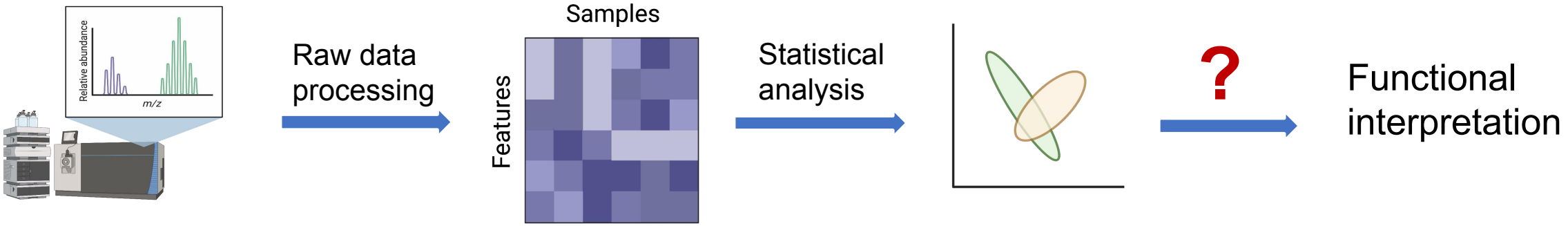
Functional interpretation of LC-MS peaks using MetaboAnalyst 5.0

Yao Lu, PhD Candidate

yao.lu5@mail.mcgill.ca

McGill University, Montreal, QC
Canada

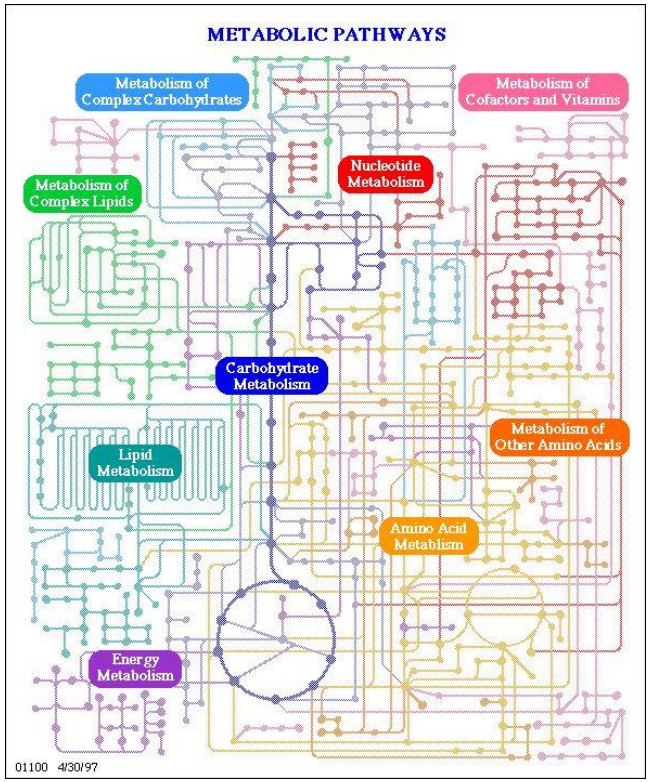
General untargeted metabolomics pipeline:



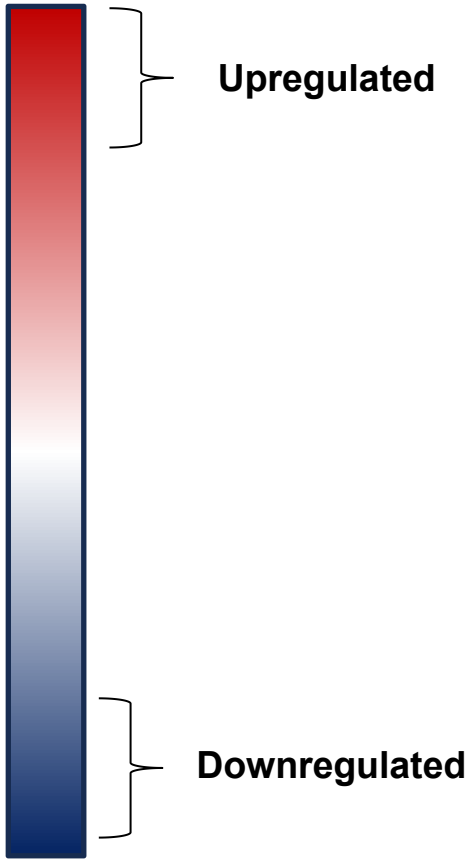
From compounds / peaks to functions

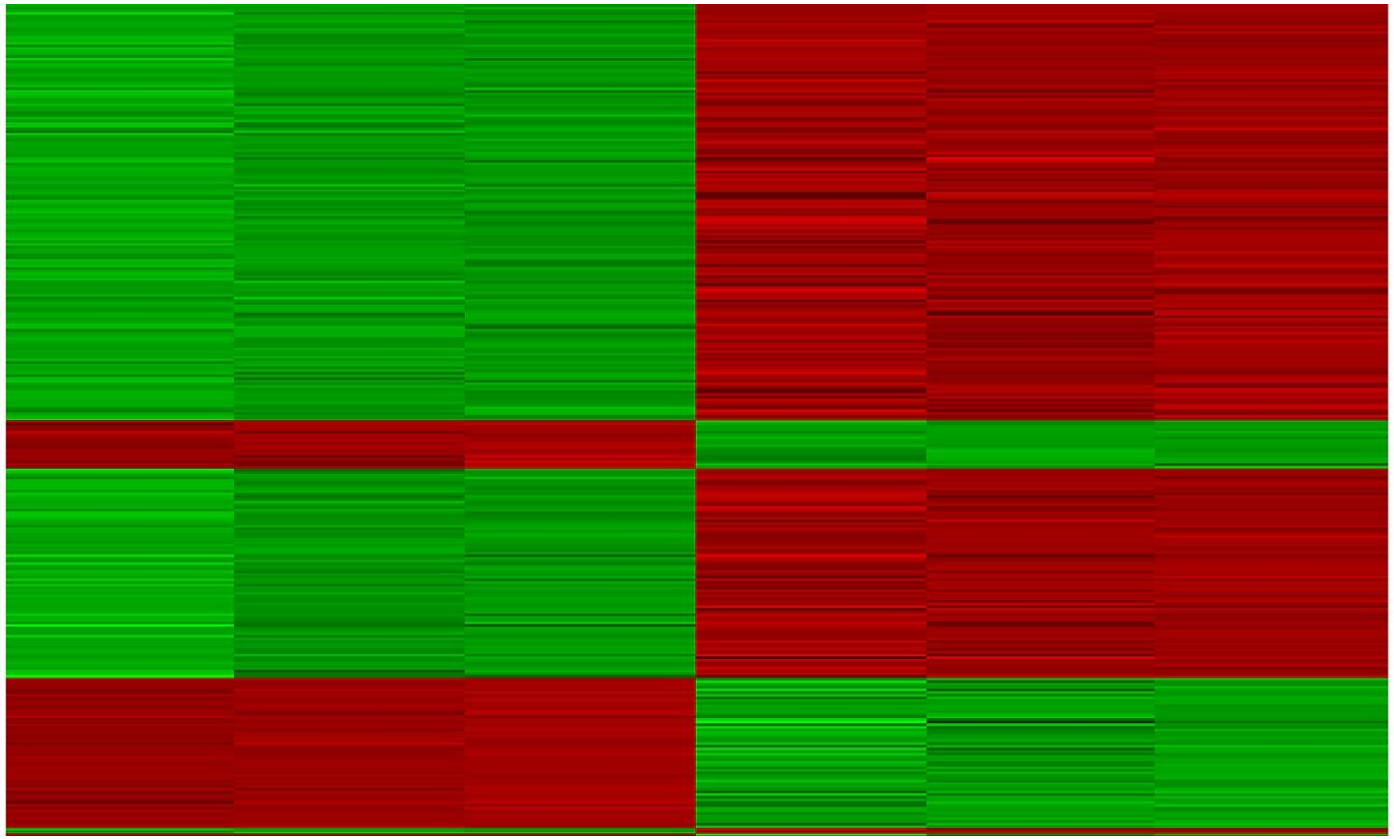
Compound	Retention Time (min)	Conc. in Urine (µM)
Dns-o-phospho -L-serine	0.92	<D.L. *
Dns-o-phospho -L-tyrosine	0.95	<D.L.
Dns-adnosine monophosphate	0.99	<D.L.
Dns-o-phosphoethanolamine	1.06	16
Dns-glucosamine	1.06	22
Dns-o-phospho -L-threonine	1.09	<D.L.
Dns-6-dimet hylamine purine	1.20	<D.L.
Dns-3-methyl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



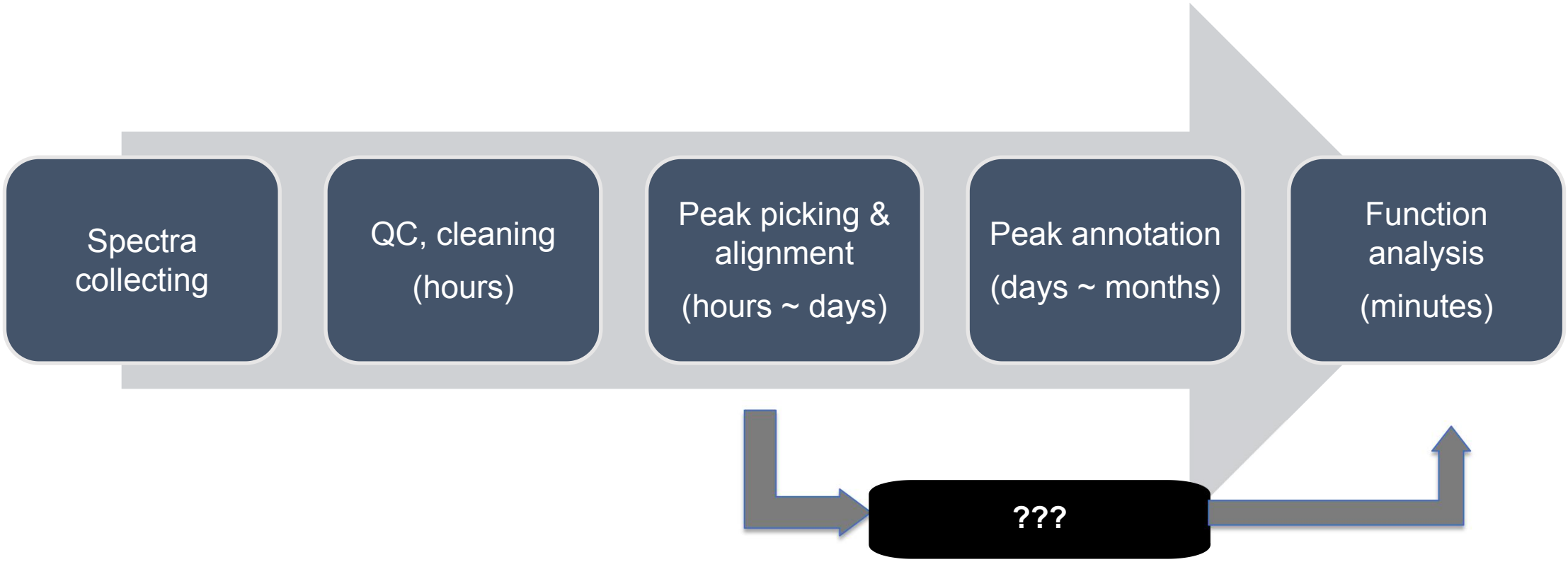


Clustered (co-regulated) compounds

How do we detect functional changes?

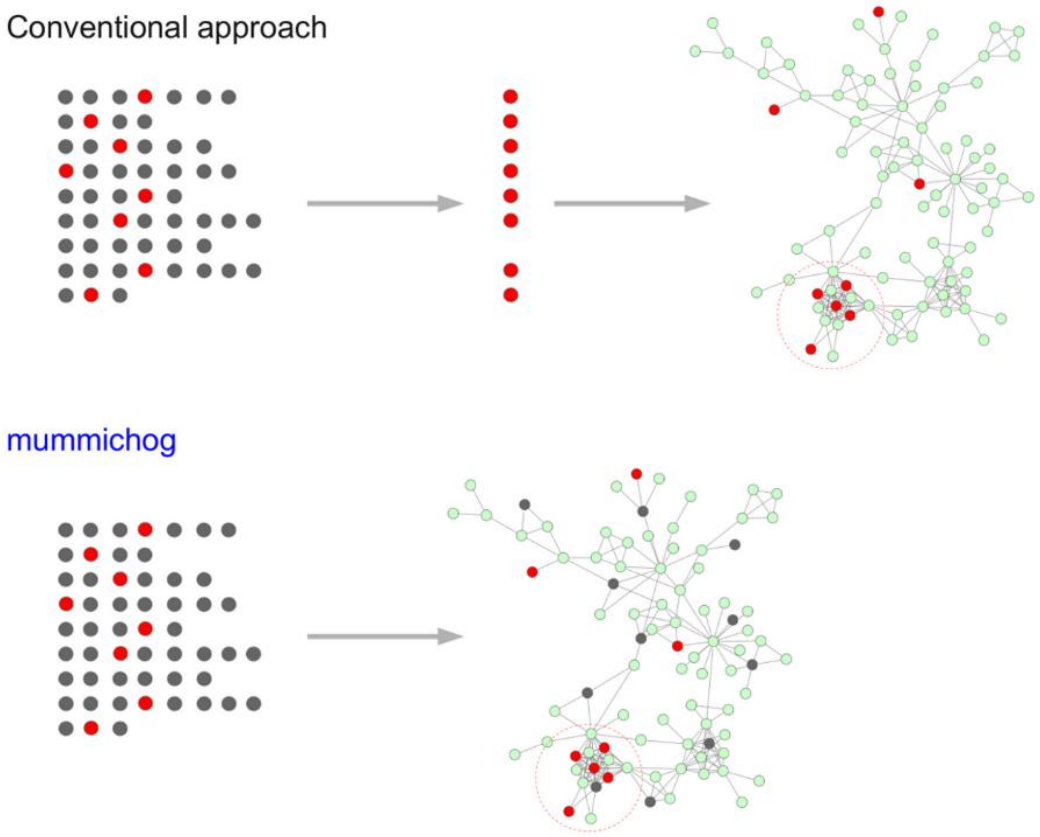
- Functions are group behaviors
  - Multiple molecules work together to finish a task
- To test whether members involved in a function have more consistent changes (i.e. most move in one direction) compared to random changes
- Enrichment analysis
  - Over representation analysis (ORA)
    - Starting 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



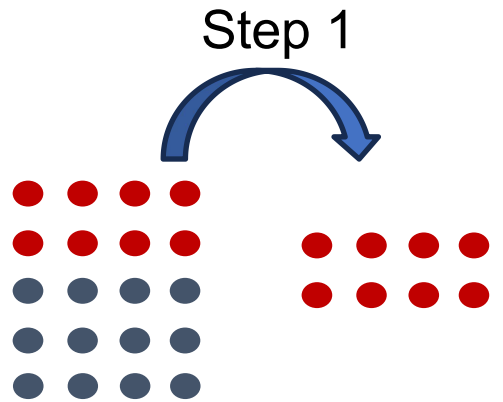
- We have a general idea what each peak is
- We have an idea of how each compound in a pathway could look like as peaks
- Mummichog: map many possible matches to a background pathway and look for consistent functional patterns
- Pathway-level results will be robust

S. Li *et al.*, *Plos Computational Biology* (2013).

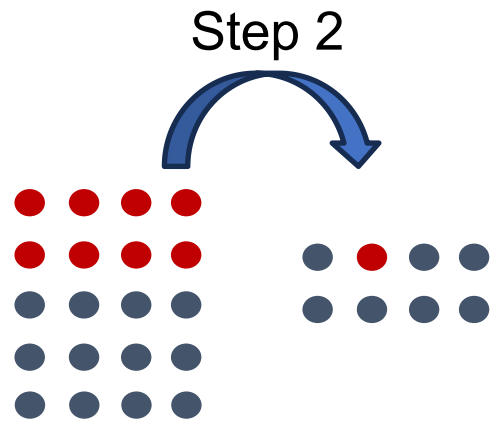
How does mummichog work?

Input requirement:

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



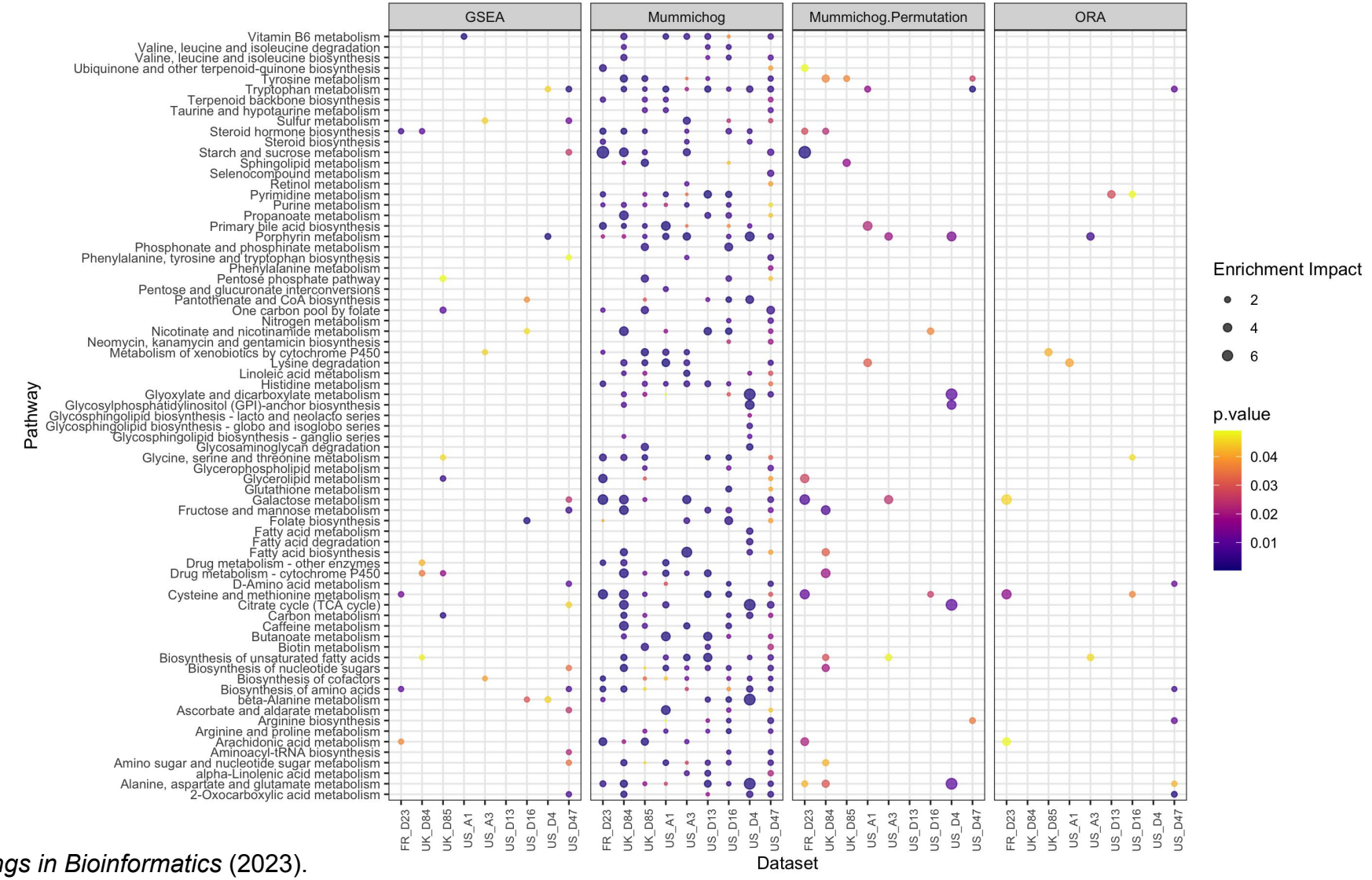
Step 1: Match the peaks to tentative metabolites. Looked up all the significant metabolites in each pathway and calculate the p-value using Fisher's exact test



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

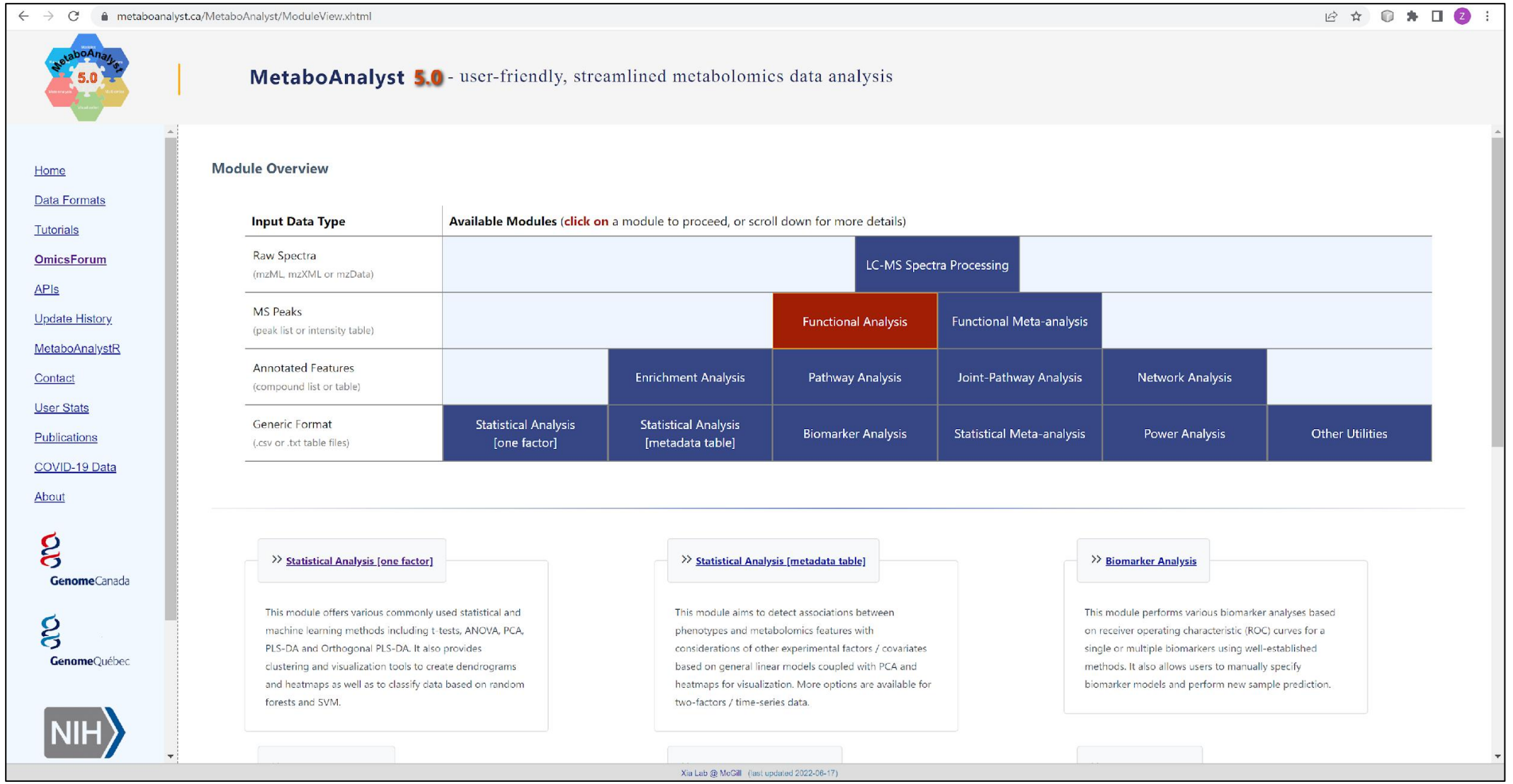
Step3: Test if certain pathways are enriched in the significant peaks as compared to null models (Gamma distribution)

Mummichog – sensitive, robust enrichment analysis for metabolomics



Y. Lu *et al.*, *Briefings in Bioinformatics* (2023).

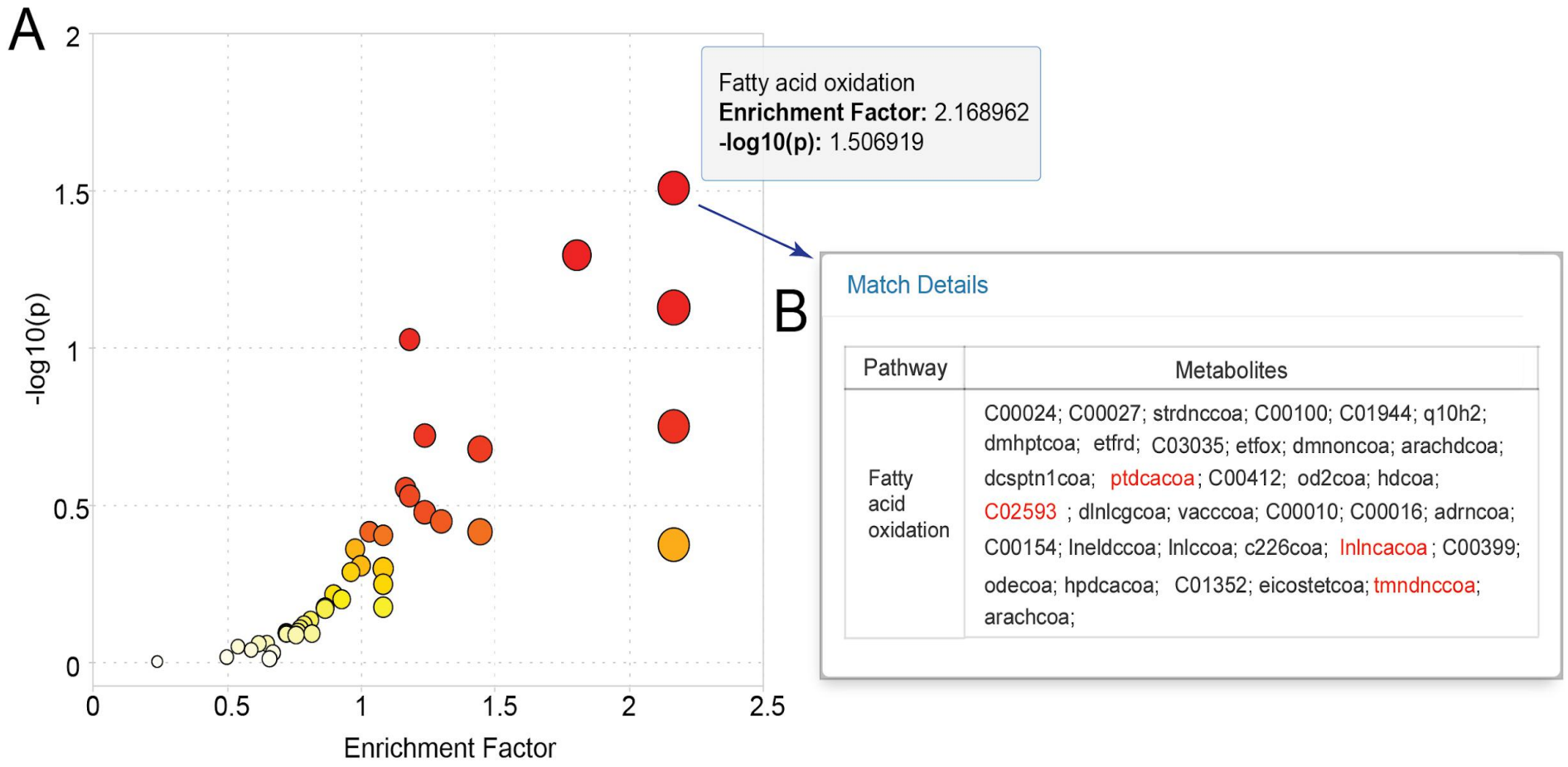
Implementation in MetaboAnalyst



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

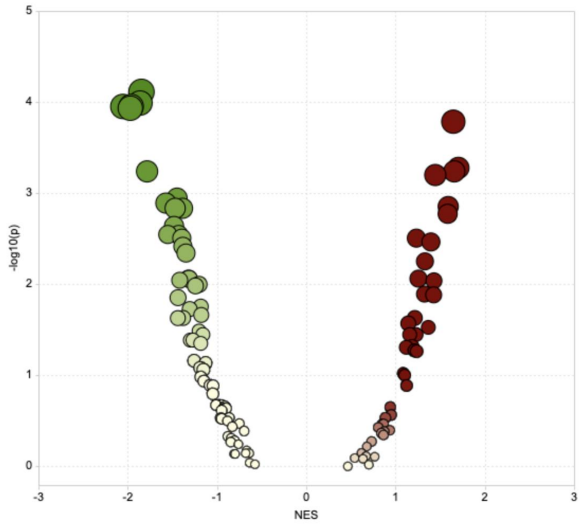


How to interpret result table (key parameters)

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	View
Carnitine shuttle	72	25	10	6.7418	0.06554	0.028334	View

- **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.

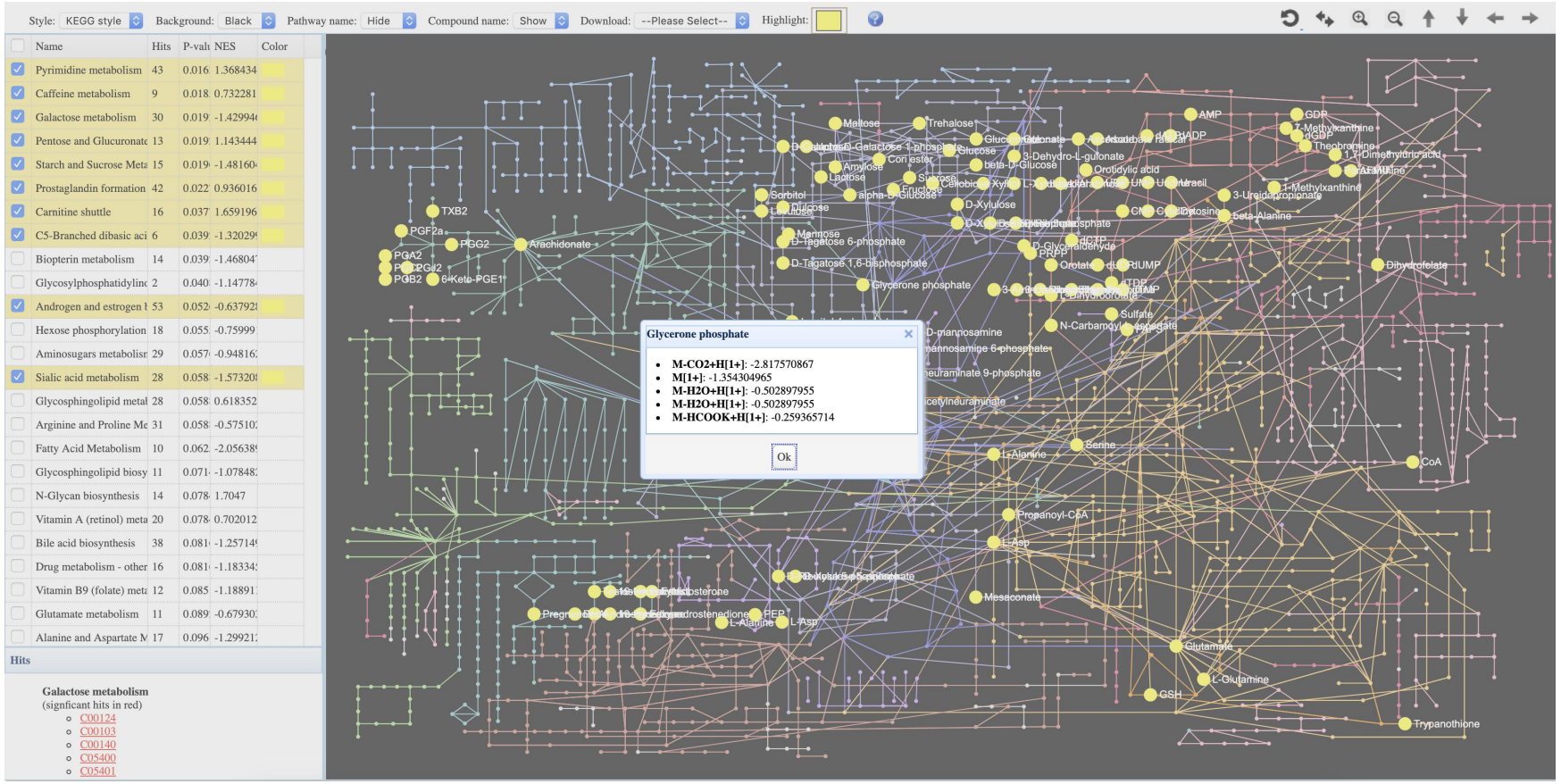
From pathways to “candidate” compounds



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

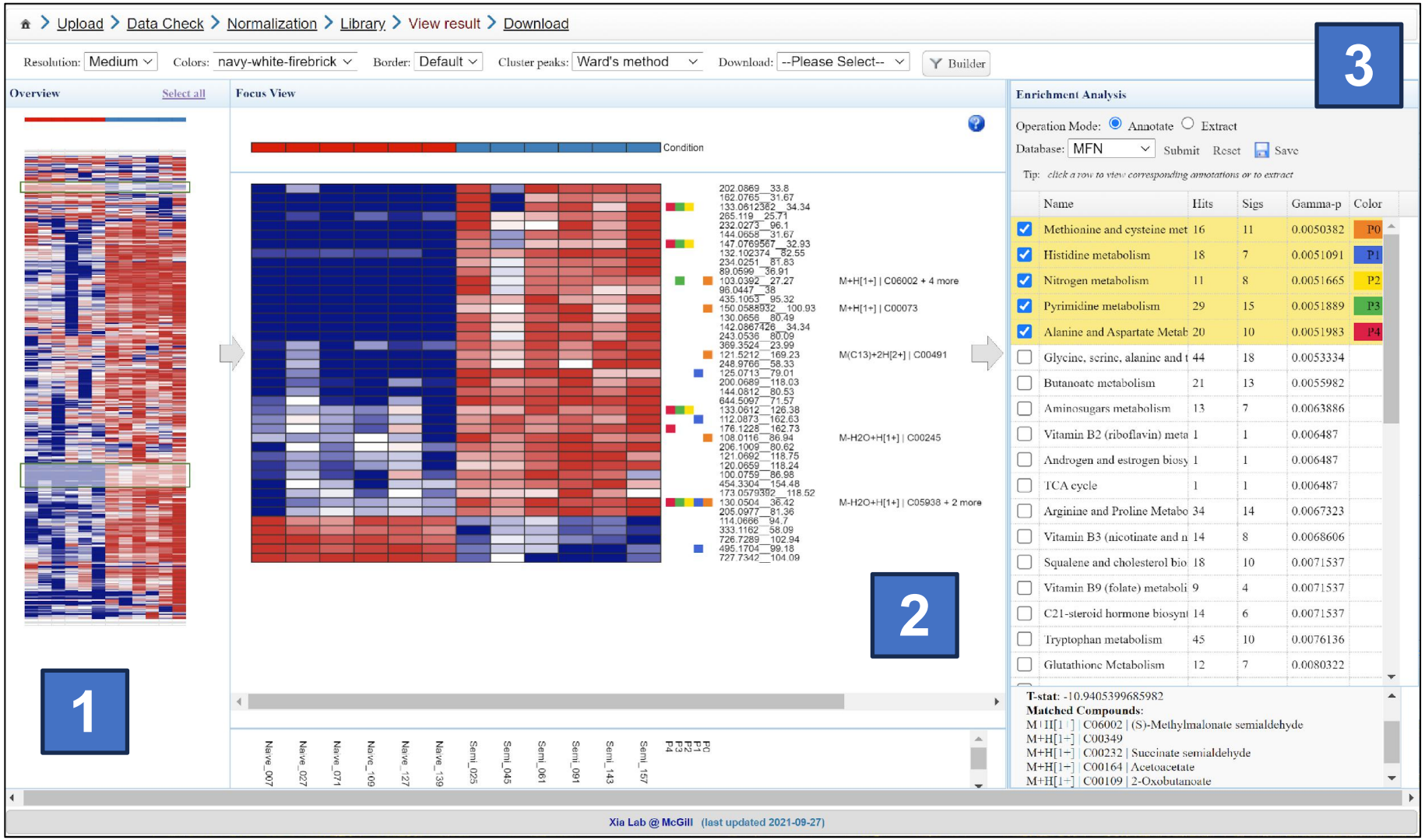
Pathway	Metabolites
Prostaglandin formation from arachidonate	**CE5703**; CE5700; CE6238; **C00425**; CE5707; **CE5704**; **CE5705**; CE5724; CE5708; CE1243; CE6236; C00584; **CE5535**; **C00959**; **CE5537**; **CE5925**; **CE5533**; C13856; **CE0955**; **CE4980**; C01312; CE5304; **C04686**; CE4938;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

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/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