Spectra Processing, Compound Annotation, Functional Insight and Causal Analysis using MetaboAnalyst 6.0

Jianguo (Jeff) Xia, Associate Professor

Canada Research Chair in Bioinformatics & Big Data Analytics

jeff.xia@mcgill.ca | www.xialab.ca

McGill University, Canada

XiaLab.ca

Empowering researchers through trainings, tools, and AI

McGill UNIVERSITY

Schedule

Part I: 2:15 pm – 4:15 pm

- **2:15 – 2:30:** General introduction
- **2:30 - 3:00:** Untargeted metabolomics
  - ✓ LC-MS & MS/MS spectral processing
  - ✓ From peaks to functions
- **3:05 – 3:25**: Live demo
- **3:25 – 4:15:** Hands on practice

Part II: 4:30 p.m. – 6:30 p.m.

- **4:30 – 5:15**: Background
  - ✓ Statistical analysis
  - ✓ Causal analysis
- **5:15 – 5:35**: Live demo
- **5:40 – 6:20**: Hands on practice
- **6:20 – 6:30**: Summary & discussion

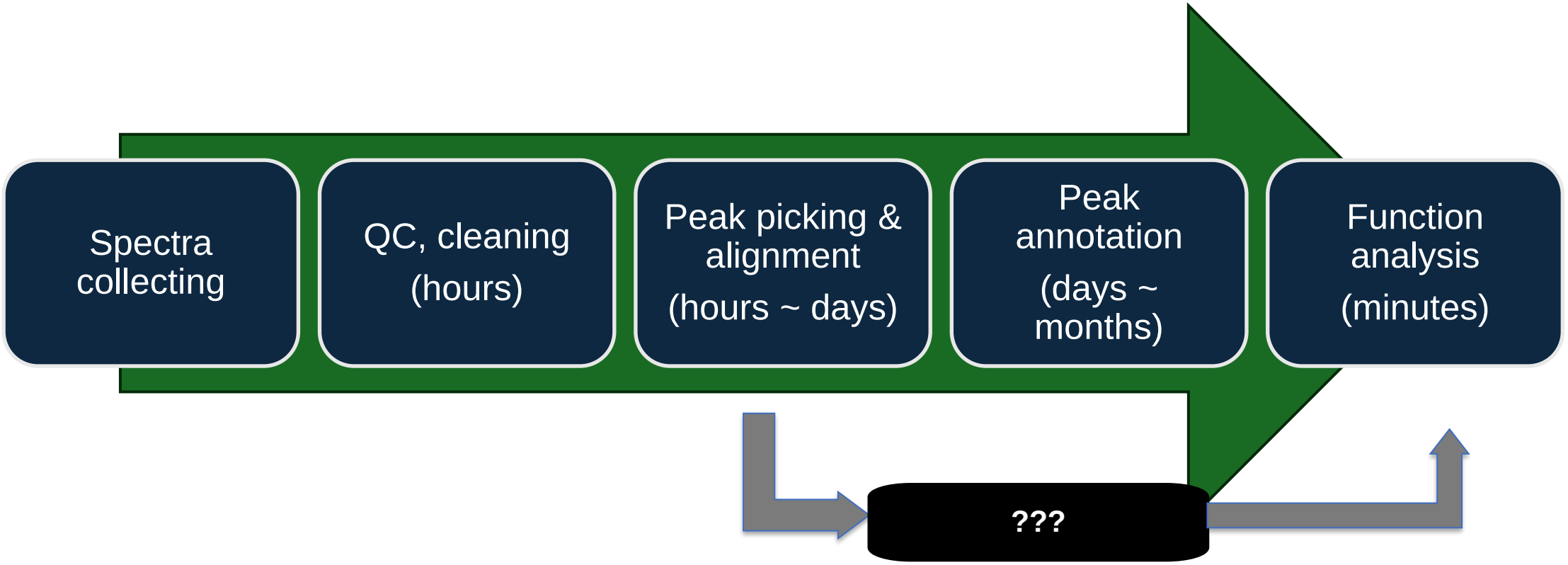
Github Repository

- https://github.com/xia-lab/Metabolomics_2024

- Slides (in PDF format);
- Example data;
- Reference literatures;
- Contact information.

Functional Analysis

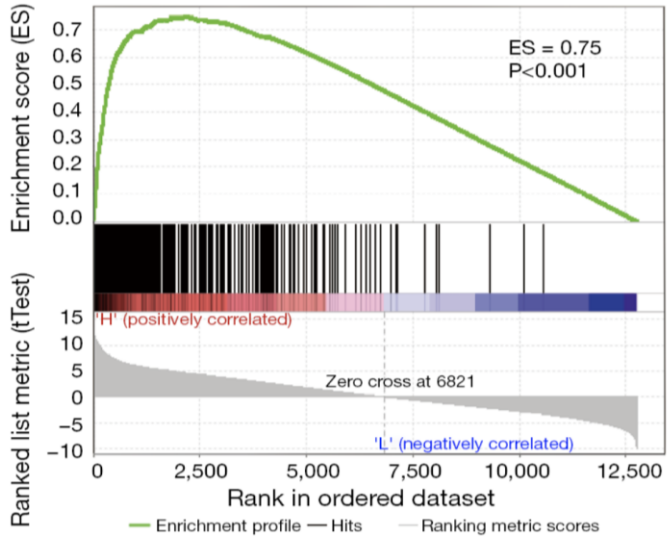
Conventional approaches is time consuming



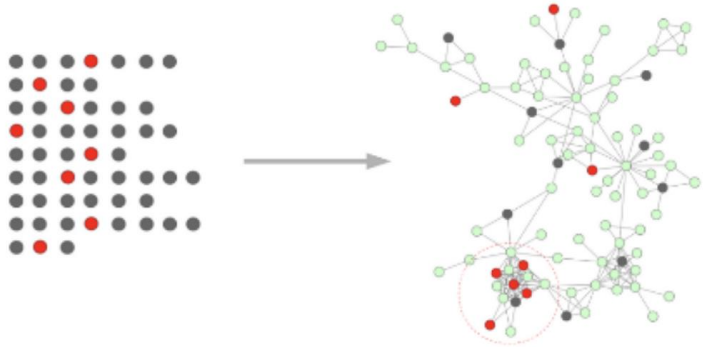
Can we perform enrichment analysis direct from peaks?

Key concept

- Biological systems showing coordinated changes or group behaviors
- Leveraging this collective power inherent in biological systems can tolerate the random errors/inaccuracies based on individual genes/metabolites
  - Gene set enrichment analysis (GSEA)
- Can we apply this concept on MS peaks?
  - ➔ The *mummichog* approach



A Subramanian *et al.*, *PNAS* (2005).



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

Critical: Input preparation

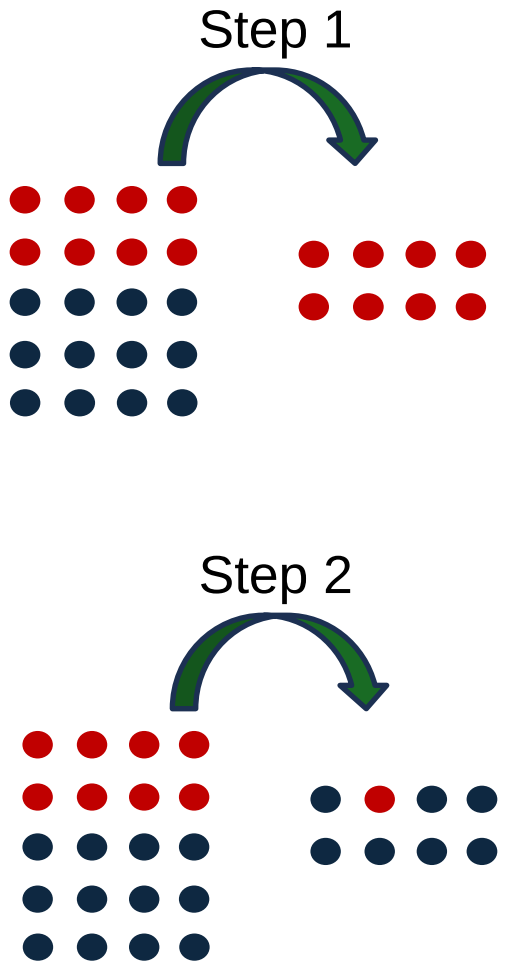
- LC - **high-resolution** MS (LCHR-MS)
  - Orbitrap, Q-TOF
  - Reason: putative annotation needs to be approximately correct (better guess leads to more accurate functional analysis)
- Needs to be **complete** peak list or peak intensity table
  - Not just significant peaks
  - Reason: mummichog using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

How does mummichog work?

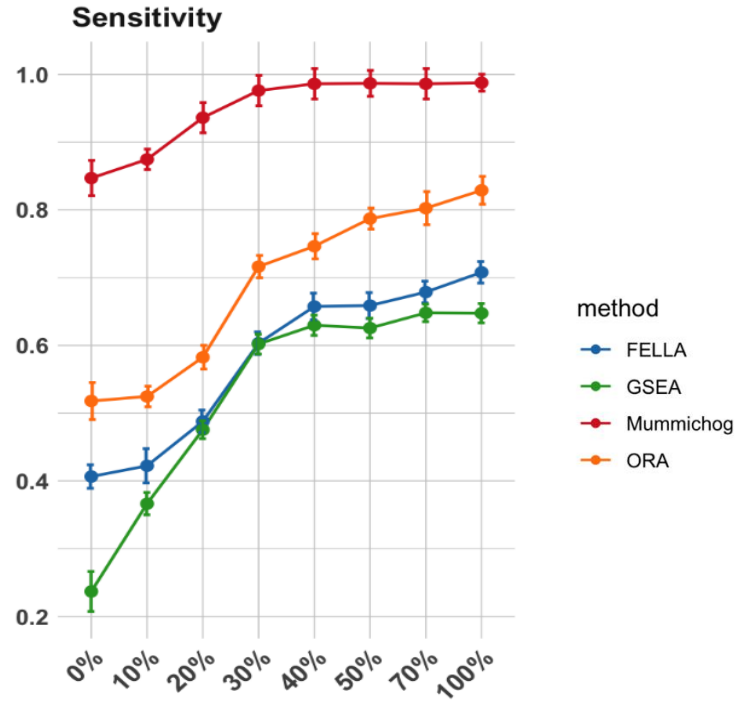
Step 1: Match the peaks to tentative metabolites. Looked up all the significant metabolites in each pathway and calculate the p-value using Fisher's exact test

Step 2: Randomly pull features with the same length as the significant ones, and repeat Step 1 for 100 ~ 1000 times

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

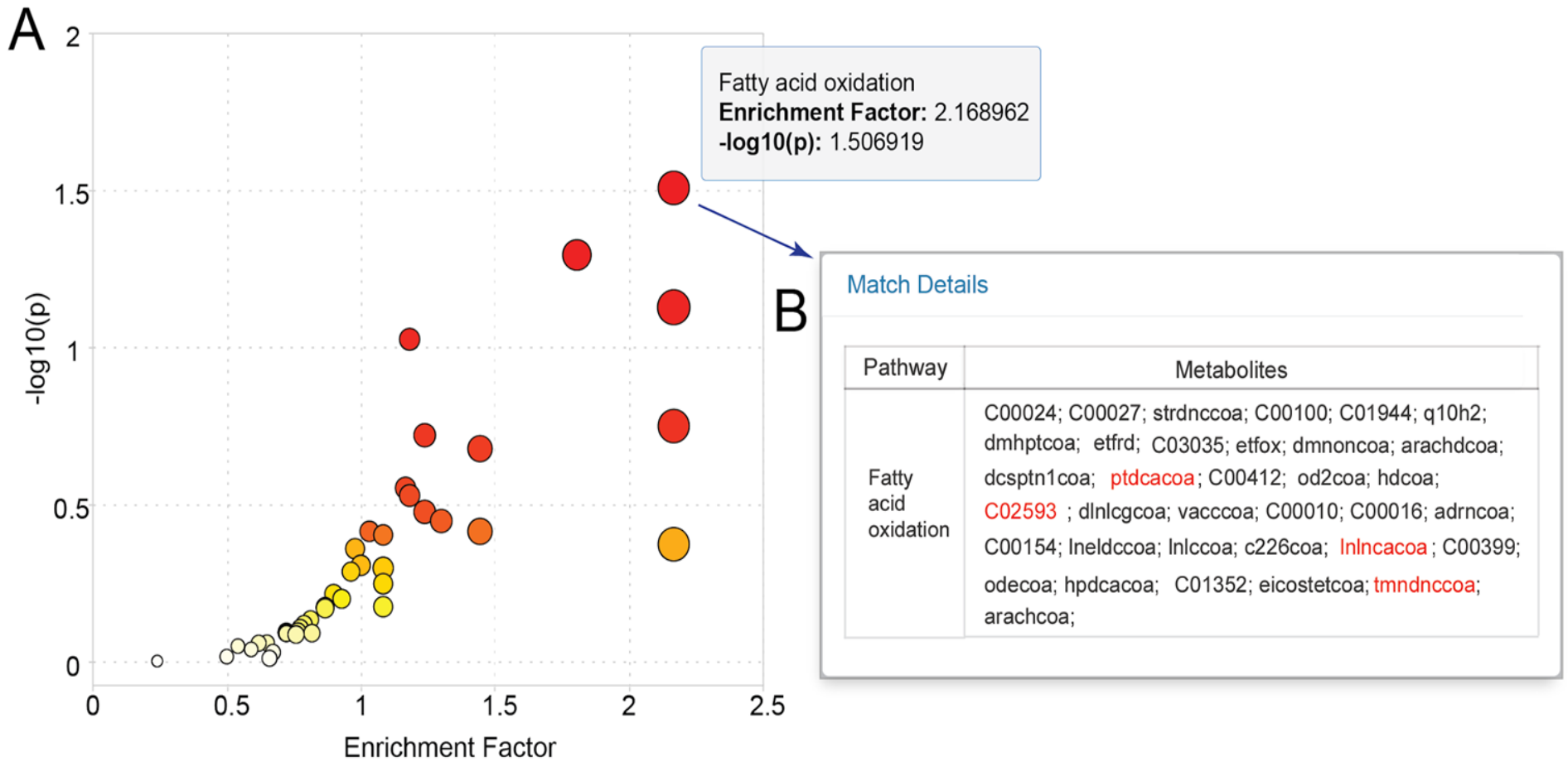


Mummichog is more sensitive, robust compared to GSEA



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

From ranked peak lists to functions

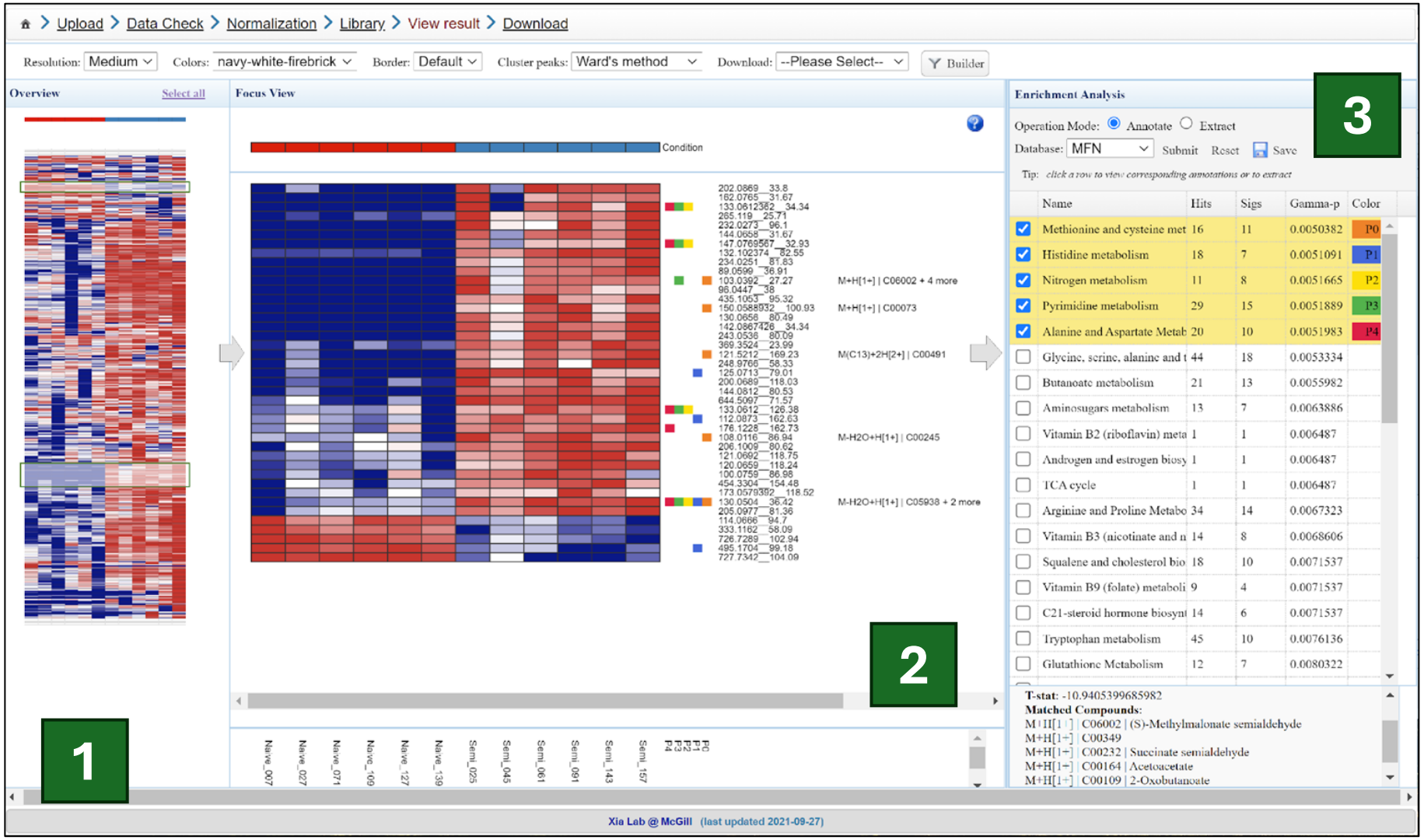


How to interpret result table (key parameters)

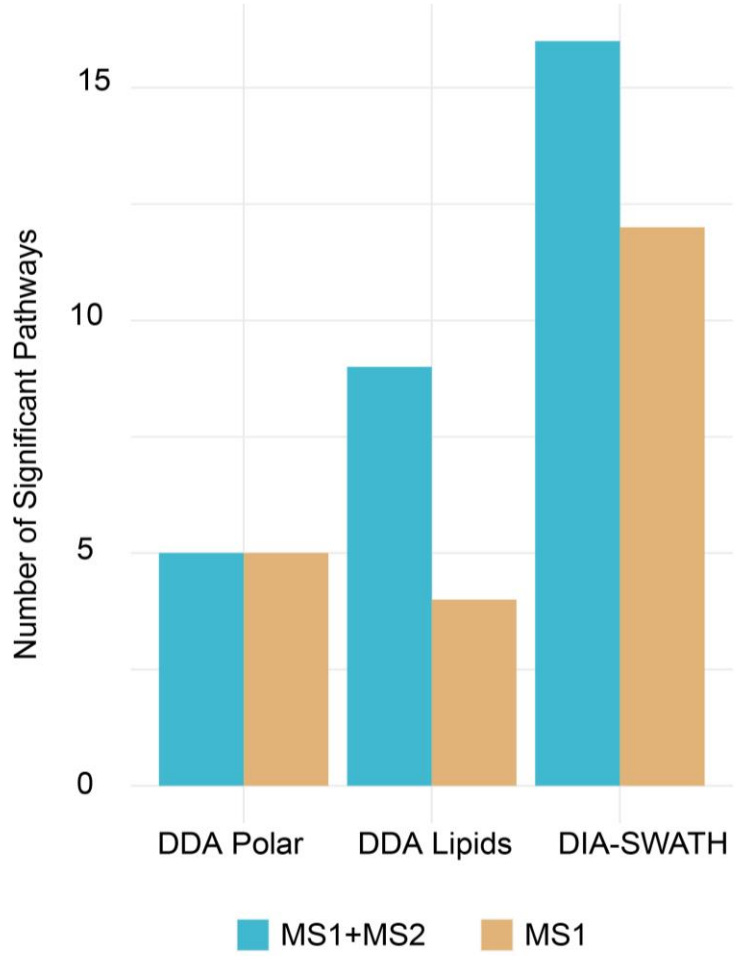
Pathway Name	Total	Hits (all)	Hits (sig.)	Expected	P(Fisher)	P(Gamma)	Details
Vitamin E metabolism	54	38	15	5.0563	0.030024	0.025523	View
Carnitine shuttle	72	25	10	6.7418	0.06554	0.028334	View

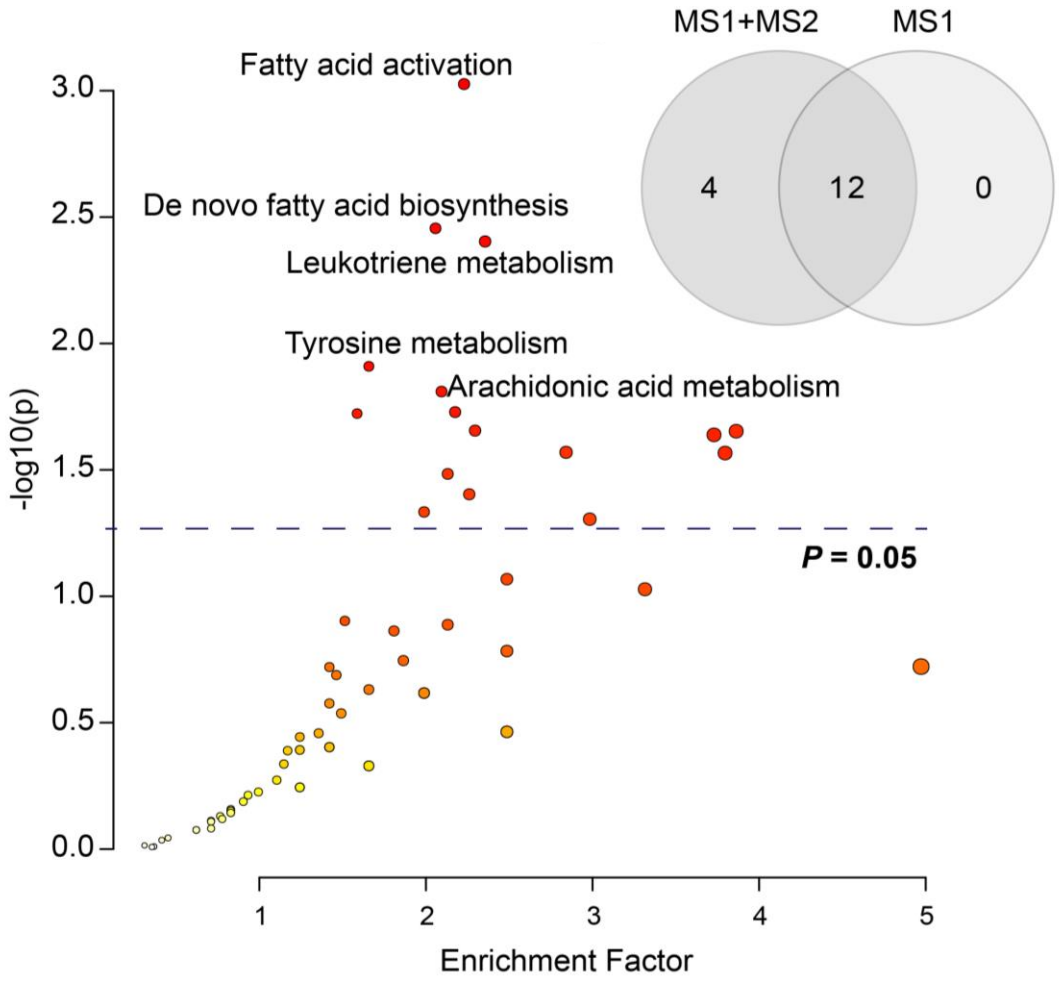
- **Total**: the total number of the given pathway
- **Hits (all)**: all the peaks mapped to the pathway
- **Hits (sig)**: all the significant peaks mapped to the pathway
- **Expected:** The expected number of metabolite hits in the pathway.
- **P(Fischer)**: The Fisher's exact p-value for the pathway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

Not just significant peaks



Integrating LC-MS and MS/MS improves results





Recap: a unified workflow for global metabolomics