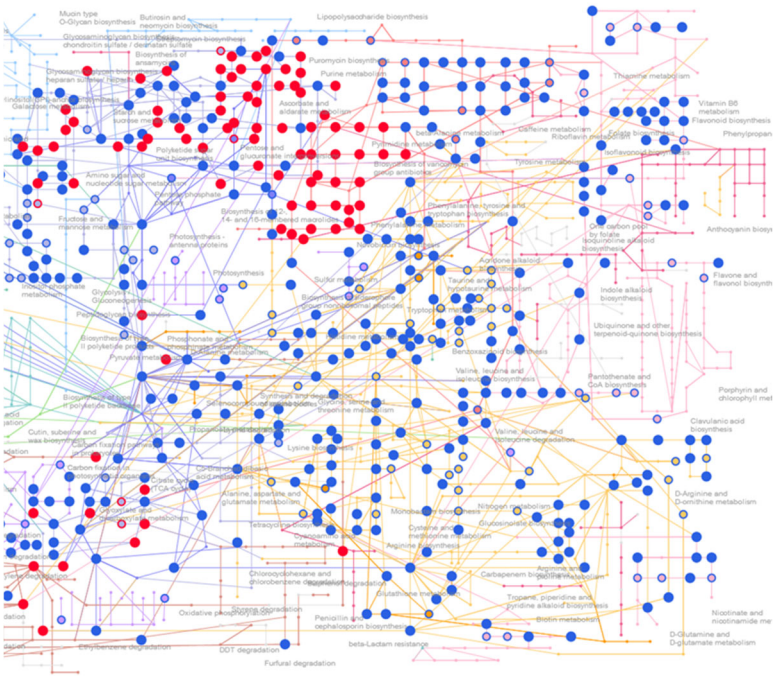
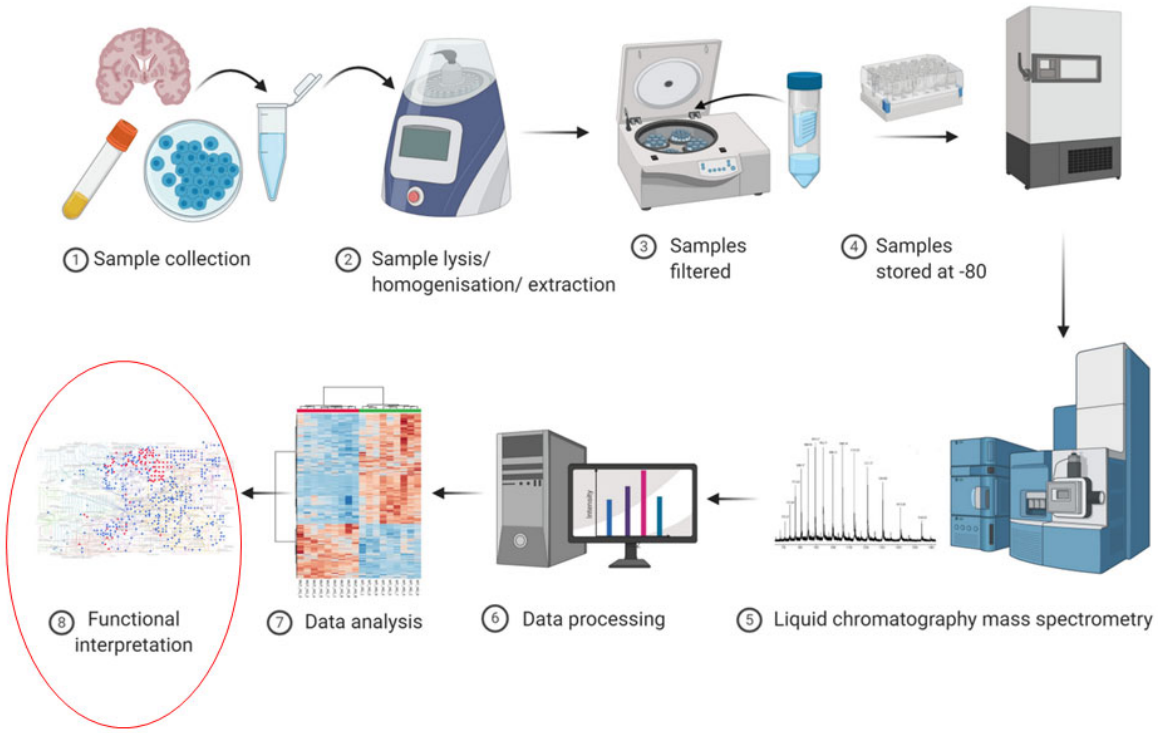
Lecture 2 Handout



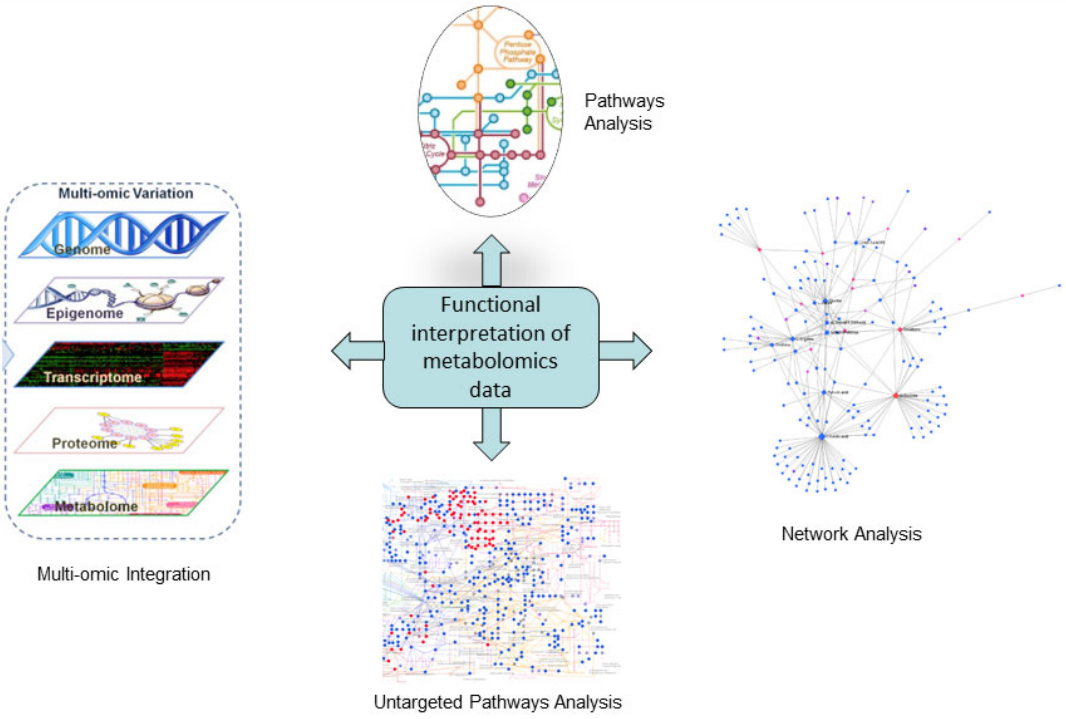
Functional Metabolic Pathways Analysis

Department of Chemistry, University of Oxford, 7th December 2022

OVERVIEW OF THE METABOLOMICS WORKFLOW



FUNCTIONAL INTERPRETATION OF METABOLOMICS DATA



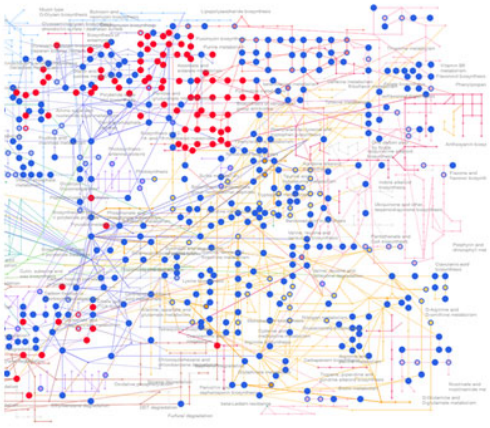
LECTURE 2: FUNCTIONAL INTERPRETATION OF UNTARGETED METABOLOMICS DATA

Part 1: Introduction the functional interpretation of metabolomics data

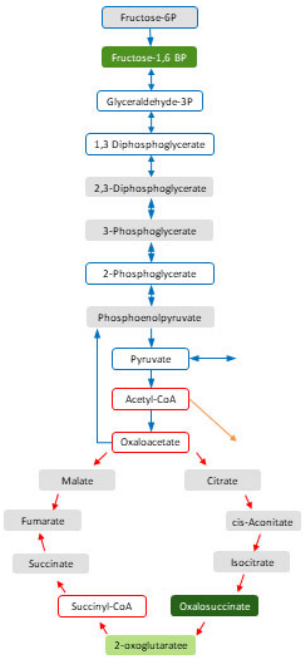
- The purpose of functional interpretation
- Approaches to functional interpretation
- Pathways Analysis
- Untargeted network analysis

Part 2: How to use MetaboAnalyst functional interpretation

- Pathways Analysis module
- MS-Peaks to pathways module



METABOLIC PATHWAY HEATMAPS



Glycolysis and gluconeogenesis

Walsby-Tickle, J., Gannon, J., Hvinden, I. *et al.* Anion-exchange chromatography mass spectrometry provides extensive coverage of primary metabolic pathways revealing altered metabolism in IDH1 mutant cells. *Commun Biol* **3**, 247 (2020). https://doi.org/10.1038/s42003-020-0957-6

PATHWAY HEATMAPS

Pros

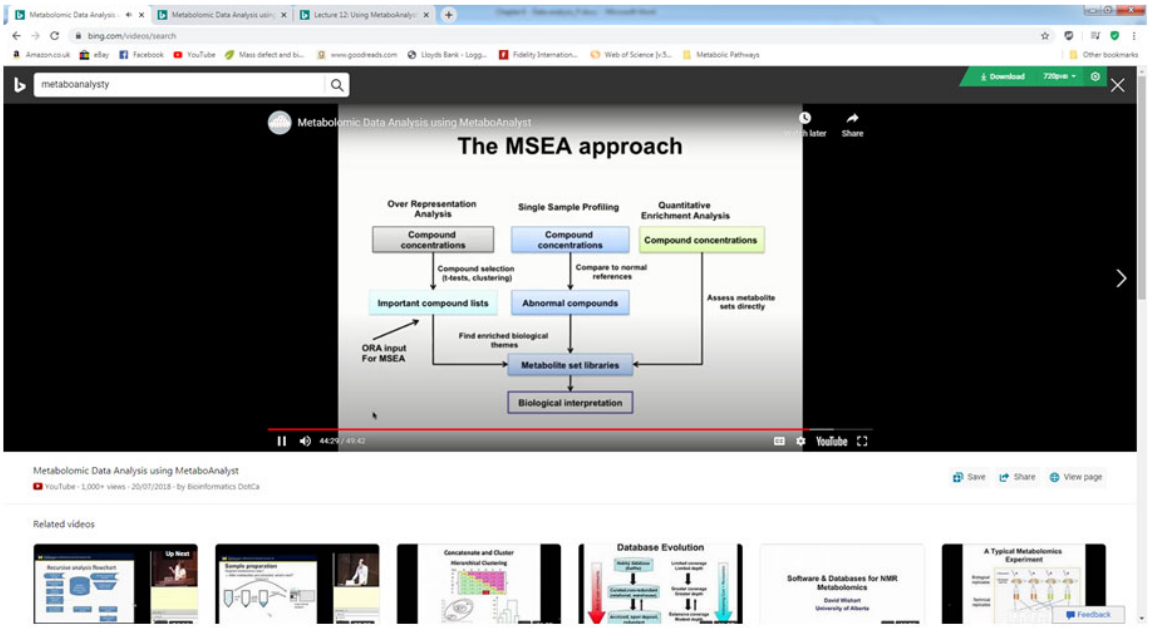
- Provide a biological context.
- Help identify which pathways may be affected by experimental changes.
- Link biomarkers to biological context.

Cons

- Does not evaluate **which pathway changes are most important.**
- **How to select 'pathways' which in themselves are** non-biological in isolation.

VARIOUS TYPES OF STATISTICAL PATHWAYS ANALYSIS APPROACHES

Pathway approach	Aim	Pros	Cons
Pathway heatmaps (TARGETED)	To provide a visualisation of the direction and magnitude of abundance changes mapped onto metabolic pathways.	Connects significantly altered metabolites with their *in vivo* pathways	Provides no way to determine the importance or impact of the changes in a metabolic context
Over enrichment analysis (ORA) (TARGETED)	Identifies if a group of significant compounds, related to specific pathways, that are over-represented within a list of identified metabolites.	Provides a **statistical measure** of pathway importance (Fishers exact test)	**Not quantitative** (does not consider the degree of fold-change.
Quantitative enrichment analysis (QEA) (TARGETED)	Identifies a list of **significant** metabolites from a peak list directly and links these to pathways	**Quantitative** – based on concentrations and more sensitive than ORA	Treats all pathways and **individual metabolites in pathways equally**.
Pathway topological analysis (TARGETED)	To provide a measure of the impact and significance of pathway changes bases on node centrality measures to estimate importance using *betweenness centrality* and *degree centrality*.	Is more sensitive to the **presence of specific metabolites** within a pathway and hence their impact.	Relies on **identified metabolites only** (relatively low and usually biased coverage)
Peaks to pathways analysis (UNTARGETED)	Untargeted pathways analysis tool representing metabolites as nodes accessing the full genome models and utilising untargeted datasets to their fullest extent.	Make use of **majority of untargeted compound-features** to create more comprehensive metabolic networks from the data	Provides a global metabolic context only, with no ability to combine **pre-identified metabolites** or use **chromatographic data**

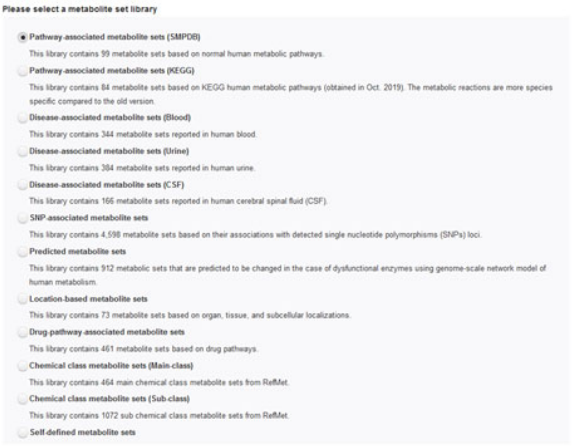


QEA PATHWAY VISUALISATION

Quantitative approach to metabolic pathways analysis based on **Gene Set Enrichment Analysis** (known as Metabolite Set Enrichment Analysis (MSEA)).

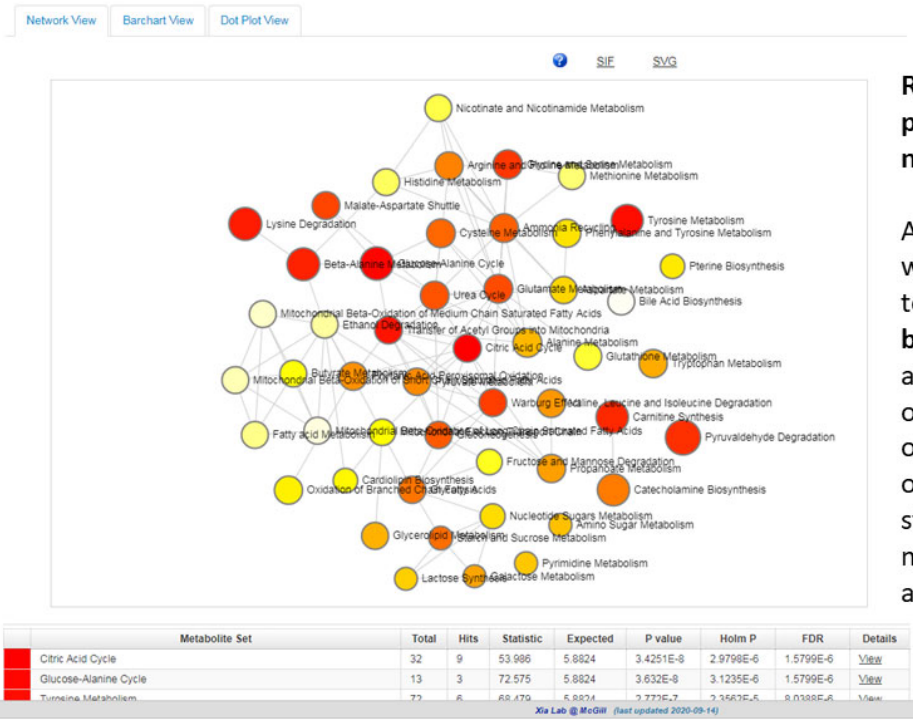
Aims to link changes in metabolite abundances quantitatively to **pathways diseases or localisation (e.g. specific tissues or organs).**

Requires metabolite libraries which provide a background from which significance is predicted.



21 libraries aviabale in MetaboAnalyst at present

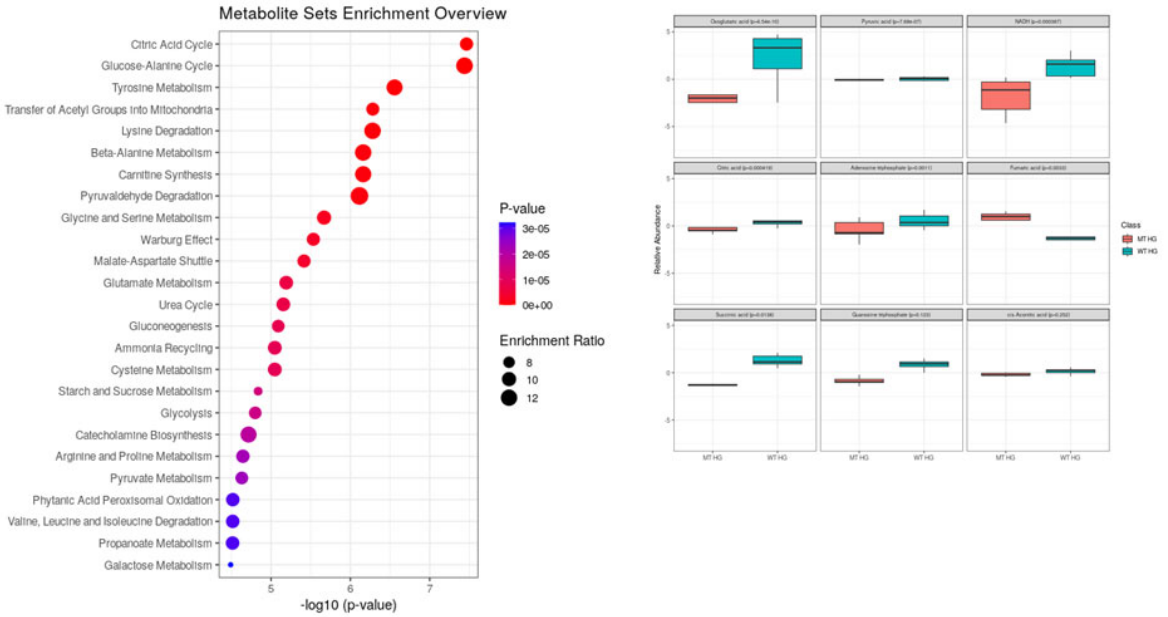
QEA PATHWAY VISUALISATION



Results are provided in a number of formats

As a **network view** where its possible to see the **links between pathways** and the **significance** of the modulation of pathways based on their colour (the stronger the red the more significantly altered the pathway

QUANTITATIVE ENRICHMENT ANALYSIS (QEA)



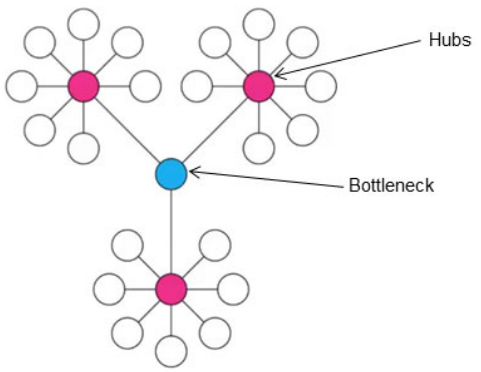
Uses a background metabolic pathways set to compute p-value significance for pathway enrichment using metabolite abundances. This puts the dataset into a **comprehensive metabolic context** and provides a **statistical measure of the relative importance** of altered pathways

PATHWAY TOPOLOGICAL ANALYSIS: MORE SENSITIVE APPROACH

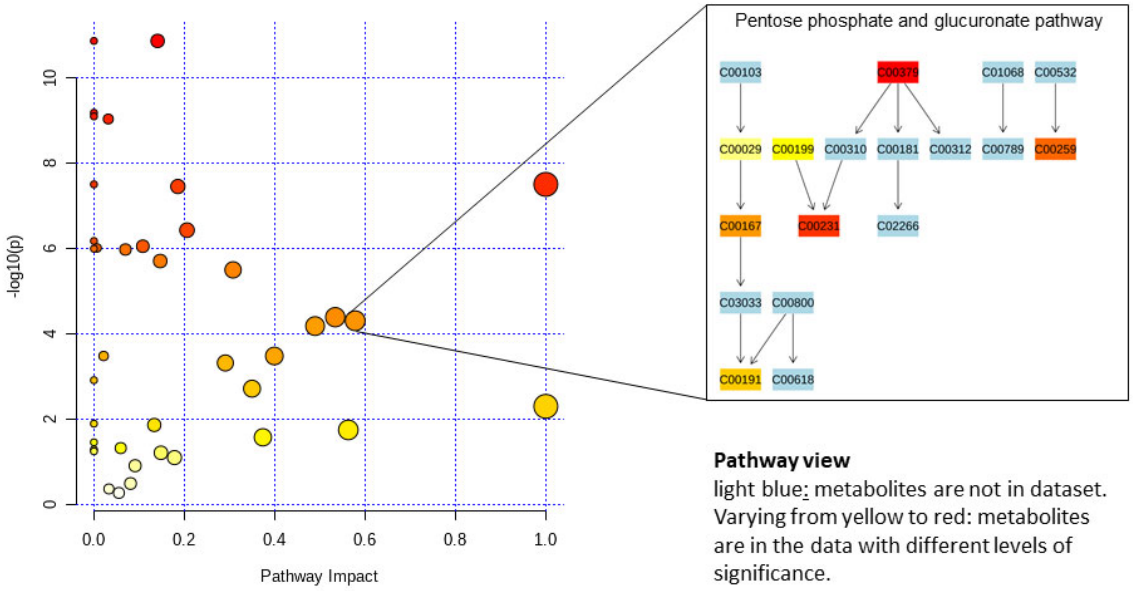
The **position of the metabolites in a pathway** has an impact on how their modulation affects metabolic processes. This is captured by the concepts of **'hubs'** and **'bottlenecks'** and **Pathways Topological Analysis** aims to model this into the pathways analysis process.

Pathway Position Matters

- **Which positions are important?**
  - **Hubs**
    - **Nodes that are highly connected (red ones)**
  - **Bottlenecks**
    - **Nodes on many shortest paths between other nodes (blue ones)**
- **Graph theory**
  - **Degree centrality**
  - **Betweenness centrality**



RESULTS FROM PATHWAY TOPOLOGICAL ANALYSIS



Metabolome view of the data: pathway enrichment analysis and pathway impact values from the pathway topology analysis.

Pathway view
light blue: metabolites are not in dataset. Varying from yellow to red: metabolites are in the data with different levels of significance.

PATHWAYS IMPACT

Pathway Impact

A complicated parameter, its calculation includes parameters such as:

- Log fold change of enriched metabolites
- Significance of pathway genes and the topology of signalling pathways
- Combines pathway topology with the over-enrichment data

1 2 3

Pathway Name	Total	Hits	p	-log(p)	Holm p	FDR	Impact	Details
Valine, leucine and isoleucine degradation	40	2	1.1954E-4	9.0319	0.0059769	0.0031356	0.02232	KEGG SMP
Valine, leucine and isoleucine biosynthesis	27	4	1.2542E-4	8.9838	0.0061458	0.0031356	0.04823	KEGG SMP
Glycine, serine and threonine metabolism	48	8	2.4586E-4	8.3107	0.011801	0.0040977	0.48394	KEGG SMP
Methane metabolism	34	6	3.8485E-4	7.8626	0.018088	0.0043833	0.16466	KEGG
Sulfur metabolism	18	2	4.755E-4	7.6512	0.021873	0.0043833	0.03307	KEGG SMP
Arginine and proline metabolism	77	6	6.578E-4	7.3266	0.029601	0.0043833	0.06203	KEGG SMP
Aminoacyl-tRNA biosynthesis	75	10	6.6275E-4	7.3191	0.029601	0.0043833	0.11268	KEGG
Nicotinate and nicotinamide metabolism	44	5	7.0133E-4	7.2625	0.030157	0.0043833	0.04113	KEGG SMP
Glutathione metabolism	38	2	0.0011587	6.7605	0.048664	0.0063514	0.0019	KEGG SMP
Propanoate metabolism	35	4	0.0013934	6.576	0.057129	0.0063514	0.01603	KEGG SMP
Galactose metabolism	41	3	0.001486	6.5116	0.059441	0.0063514	0.01992	KEGG SMP
Taurine and hypotaurine metabolism	20	3	0.0015243	6.4862	0.059449	0.0063514	0.35252	KEGG SMP
Cyanoamino acid metabolism	16	4	0.0016826	6.3874	0.06394	0.0064716	0.0	KEGG
Nitrogen metabolism	39	7	0.0021434	6.1454	0.079305	0.0070701	0.00763	KEGG SMP
Inositol phosphate metabolism	39	1	0.002215	6.1125	0.079741	0.0070701	0.13703	KEGG SMP
Pyruvate metabolism	32	4	0.0022624	6.0913	0.079741	0.0070701	0.41957	KEGG SMP
Cysteine and methionine metabolism	56	2	0.0026796	5.9221	0.091106	0.0078811	0.02846	KEGG SMP SMP
Alanine, aspartate and glutamate metabolism	24	6	0.0029727	5.8183	0.0981	0.0082576	0.25546	KEGG SMP SMP SMP
Pantothenate and CoA biosynthesis	27	4	0.0034143	5.6798	0.10926	0.0089486	0.18014	KEGG SMP
Phenylalanine metabolism	45	6	0.0036884	5.6026	0.11434	0.0089486	0.0315	KEGG SMP

Results: a ranking of pathways based on p-value.

MS PEAKS TO PATHWAYS (MUMMICHOG)

Untargeted metabolomics generates large amounts of data **>90% of which is not utilised** when it comes to one of the most important aspects in metabolomics – biological interpretation of the data. The MS peaks to Pathways analysis **bypasses the typical bottleneck of metabolite identification** by directly linking compounds features with metabolic pathways

Predicting Network Activity from High Throughput Metabolomics

Shuzhao Li[1,2]*, Youngja Park[3,4], Sai Duraisingham[1,2], Frederick H. Strobel[5], Nooruddin Khan[1,2], Quinlyn A. Soltow[3], Dean P. Jones[3], Bali Pulendran[1,2]

1 Emory Vaccine Center, Emory University, Atlanta, Georgia, United States of America, **2** Yerkes National Primate Research Center, Emory University, Atlanta, Georgia, United States of America, **3** Department of Medicine, Emory University, Atlanta, Georgia, United States of America, **4** College of Pharmacy, Korea University, Seoul, South Korea, **5** Mass Spectrometry Center, Emory University, Atlanta, Georgia, United States of America

Abstract

The functional interpretation of high throughput metabolomics by mass spectrometry is hindered by the identification of metabolites, a tedious and challenging task. We present a set of computational algorithms which, by leveraging the collective power of metabolic pathways and networks, predict functional activity directly from spectral feature tables without a priori identification of metabolites. The algorithms were experimentally validated on the activation of innate immune cells.

Citation: Li S, Park Y, Duraisingham S, Strobel FH, Khan N, et al. (2013) Predicting Network Activity from High Throughput Metabolomics. PLoS Comput Biol 9(7): e1003123. doi:10.1371/journal.pcbi.1003123

Editor: Christos A. Ouzounis, The Centre for Research and Technology, Hellas, Greece

Received November 7, 2012; **Accepted** May 15, 2013; **Published** July 4, 2013

Funding: Work supported by grants from US National Institutes of Health AG038746, ES016731 (to DP); U19AI090023, U54AI057157, R37AI48638, R37DK057665, U19AI057266, PO1A1096187 (to BP); Scripps CHAVI-ID Award (UM1AI100663 to BP); and the Bill and Melinda Gates Foundation (to BP). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Li et al., 2013

The **mummichog** (*Fundulus heteroclitus*) is a small fish found in the Atlantic – well known for living in large groups.

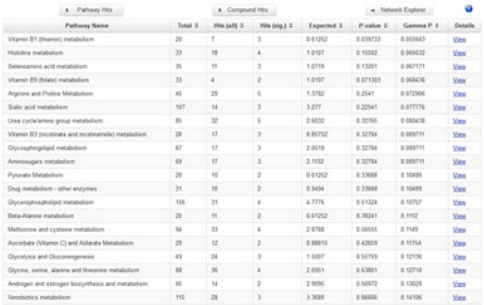
Mummichog is a software approach to **predict biological activity directly from mass spectrometry data without identifying metabolites formally.** This is approached by unifying network analysis and the annotation of pathways (using accurate *m/z* values) by combining both into the same computational framework.

MS PEAKS TO PATHWAYS

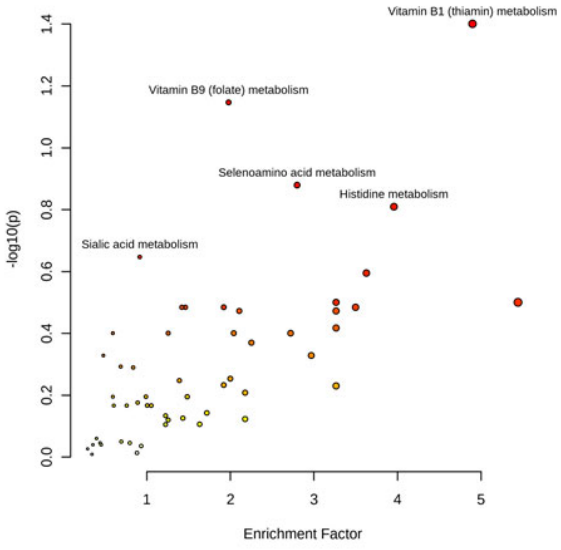
A **comprehensive metabolic network/pathways** are used to tailored to a particular organism (a range of libraries can be selected).

Possible ***m/z* values are predicted** based on the model organism (including isotopes and adducts) These are then matched with the experimental data. This step is repeated multiple times to calculate the null distribution and modelled.

Significant ***m/z* features are then used to calculate a p-value** for each pathway.



Pathway Name	Total	Hits (all)	Hits (sig.)	Expected	P-value	Gamma P	Details
Vitamin B1 (thiamin) metabolism	20	7	3	0.61252	0.03733	0.055643	View
Histidine metabolism	33	10	4	1.0107	0.15502	0.066632	View
Selenoamino acid metabolism	35	11	3	1.0719	0.13201	0.067171	View
Vitamin B9 (folate) metabolism	33	4	2	1.0107	0.071303	0.068436	View
Arginine and Proline Metabolism	45	29	5	1.3782	0.2541	0.072906	View
Sialic acid metabolism	107	14	3	3.277	0.22541	0.077776	View
Urea cycle/amino group metabolism	85	32	5	2.6032	0.32765	0.080436	View
Vitamin B3 (nicotinate and nicotinamide) metabolism	28	17	3	0.85752	0.32784	0.089711	View
Glycosphingolipid metabolism	67	17	3	2.0519	0.32784	0.089711	View
Aminosugars metabolism	69	17	3	2.1132	0.32784	0.089711	View
Pyruvate Metabolism	20	10	2	0.61252	0.33688	0.10489	View
Drug metabolism - other enzymes	31	10	2	0.9494	0.33688	0.10489	View
Glycerophospholipid metabolism	156	31	4	4.7776	0.51324	0.10707	View
Beta-Alanine metabolism	20	11	2	0.61252	0.38241	0.1112	View
Methionine and cysteine metabolism	94	33	4	2.8788	0.56555	0.1149	View
Ascorbate (Vitamin C) and Aldarate Metabolism	29	12	2	0.88815	0.42659	0.11754	View
Glycolysis and Gluconeogenesis	49	24	3	1.5007	0.55769	0.12136	View
Glycine, serine, alanine and threonine metabolism	88	36	4	2.6951	0.63801	0.12718	View
Androgen and estrogen biosynthesis and metabolism	95	14	2	2.9095	0.50972	0.13029	View
Xenobiotics metabolism	110	28	3	3.3688	0.96666	0.14106	View



Results: a ranking based on p-value for the pathways, in a similar way to targeted pathways analysis

RESULTS FROM MS PEAKS TO PATHWAYS: NETWORK VIEW