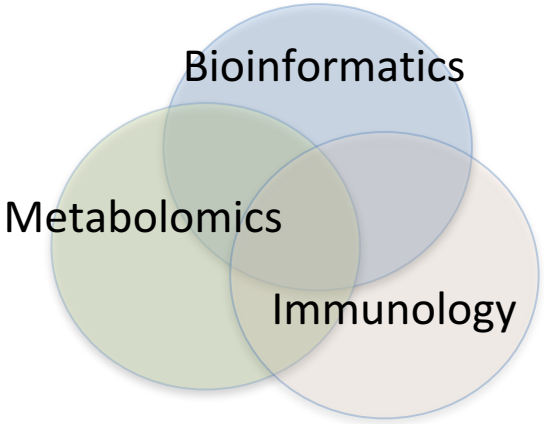
EMORY UNIVERSITY

Metabolomics Pathway Analysis



Shuzhao Li, Ph.D.
Assistant Professor, Dept. Medicine,
Emory University School of Medicine
E-mail: shuzhao.li@gmail.com
November 30, 2018

Outline

- **What are pathways and metabolic models**
- **Pathway analysis for targeted data**
- **Untargeted data analysis and *mummichog***
- **Application examples**
  - **Intracellular mechanisms**
  - **Population studies**
  - **Multi-omics integration**
- **Resources and issues**

Metabolic pathways we know today

https://www.genome.jp/kegg/pathway.html#global

1.0 Global and overview maps
- 01100 Metabolic pathways
- 01110 Biosynthesis of secondary metabolites
- 01120 Microbial metabolism in diverse environments
- 01130 Biosynthesis of antibiotics
- 01200 Carbon metabolism
- 01210 2-Oxocarboxylic acid metabolism
- 01212 Fatty acid metabolism
- 01230 Biosynthesis of amino acids
- 01220 Degradation of aromatic compounds

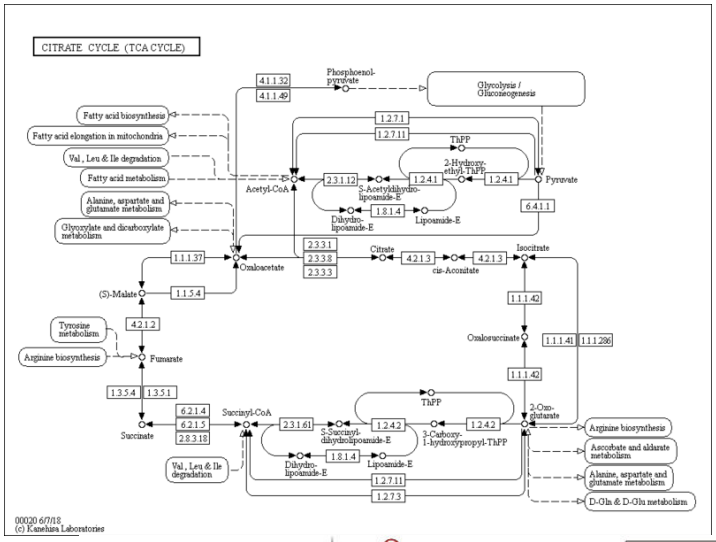
1.1 Carbohydrate metabolism
- 00010 Glycolysis / Gluconeogenesis
- 00020 Citrate cycle (TCA cycle)
- 00030 Pentose phosphate pathway
- 00040 Pentose and glucuronate interconversions
- 00051 Fructose and mannose metabolism
- 00052 Galactose metabolism
- 00053 Ascorbate and aldarate metabolism
- 00500 Starch and sucrose metabolism
- 00520 Amino sugar and nucleotide sugar metabolism
- 00620 Pyruvate metabolism
- 00630 Glyoxylate and dicarboxylate metabolism
- 00640 Propanoate metabolism
- 00650 Butanoate metabolism
- 00660 C5-Branched dibasic acid metabolism
- 00562 Inositol phosphate metabolism

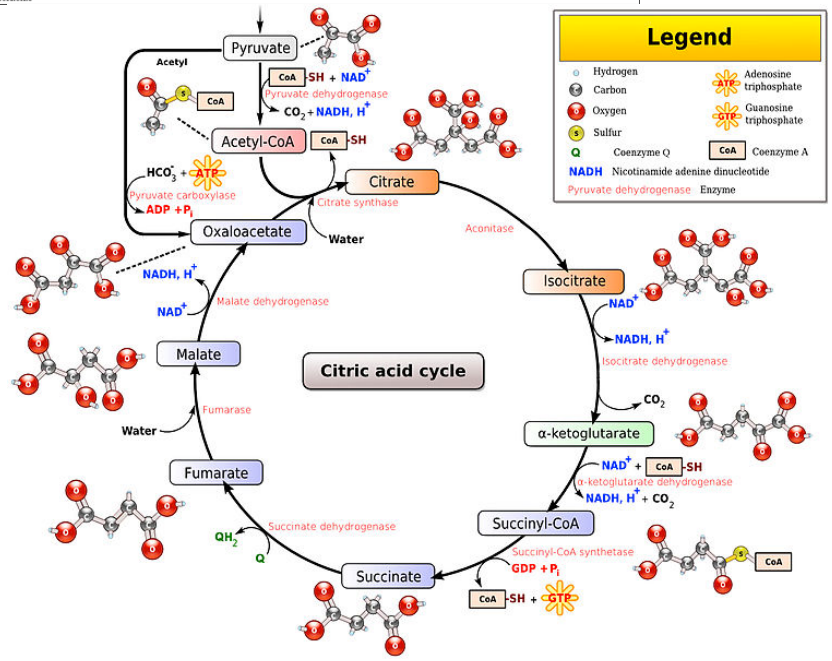
1.2 Energy metabolism
- 00190 Oxidative phosphorylation
- 00195 Photosynthesis
- 00196 Photosynthesis - antenna proteins
- 00710 Carbon fixation in photosynthetic organisms
- 00720 Carbon fixation pathways in prokaryotes
- 00680 Methane metabolism
- 00910 Nitrogen metabolism
- 00920 Sulfur metabolism

1.3 Lipid metabolism
- 00061 Fatty acid biosynthesis
- 00062 Fatty acid elongation
- 00071 Fatty acid degradation
- 00072 Synthesis and degradation of ketone bodies
- 00073 Cutin, suberine and wax biosynthesis
- 00100 Steroid biosynthesis
- 00120 Primary bile acid biosynthesis
- 00121 Secondary bile acid biosynthesis
- 00140 Steroid hormone biosynthesis
- 00561 Glycerolipid metabolism
- 00564 Glycerophospholipid metabolism
- 00565 Ether lipid metabolism
- 00600 Sphingolipid metabolism
- 00590 Arachidonic acid metabolism
- 00591 Linoleic acid metabolism
- 00592 alpha-Linolenic acid metabolism
- 01040 Biosynthesis of unsaturated fatty acids

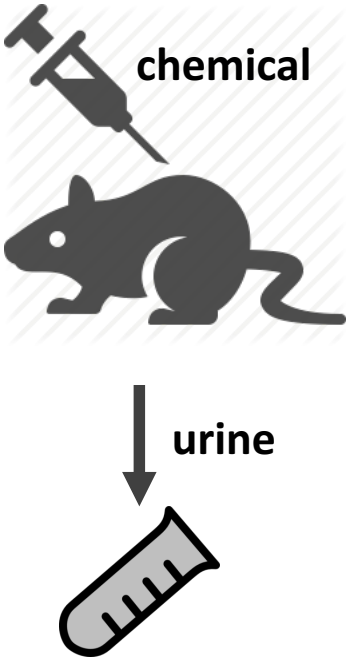
1.4 Nucleotide metabolism
- 00230 Purine metabolism
- 00240 Pyrimidine metabolism

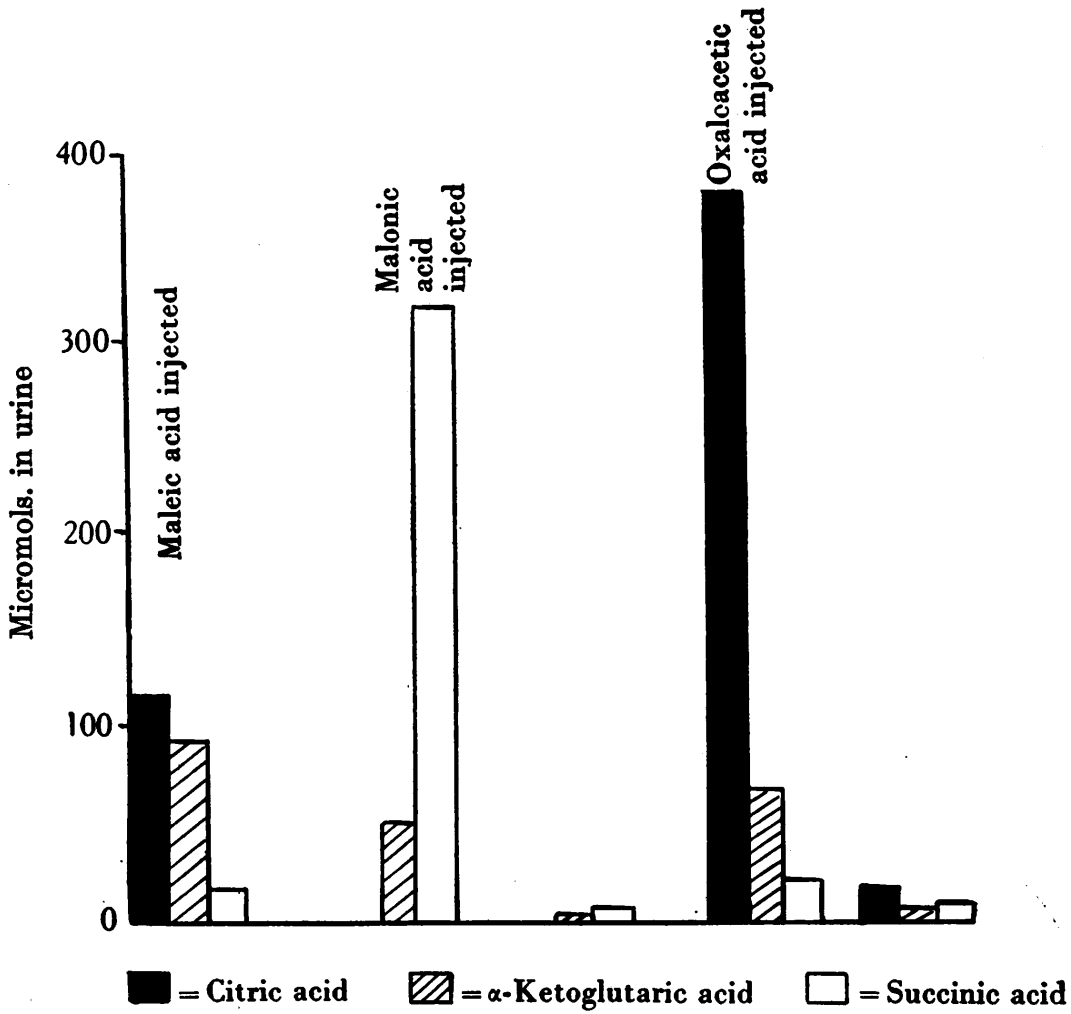
1.5 Amino acid metabolism
- 00250 Alanine, aspartate and glutamate metabolism
- 00260 Glycine, serine and threonine metabolism
- 00270 Cysteine and methionine metabolism
- 00280 Valine, leucine and isoleucine degradation



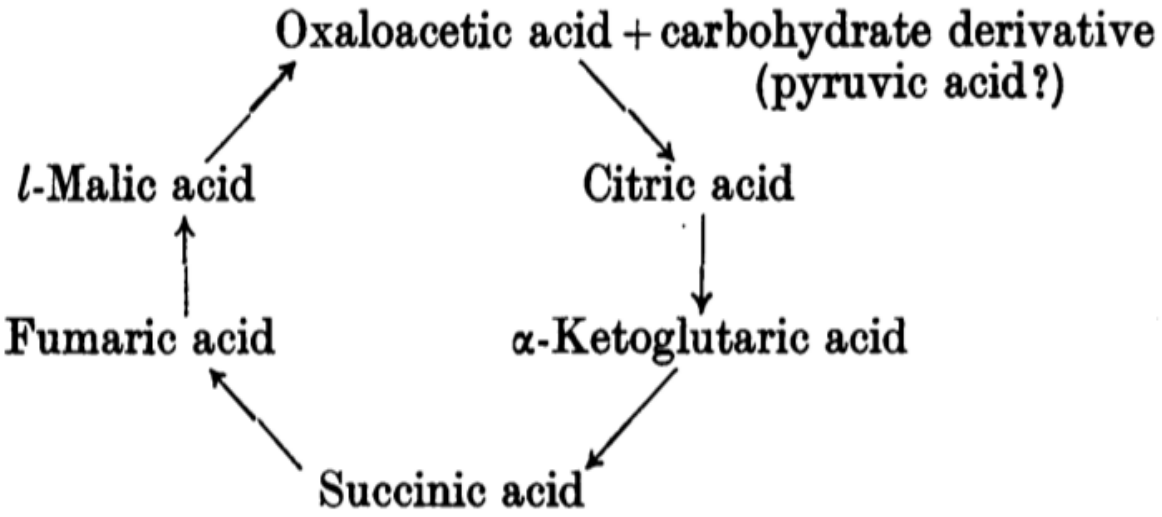


Krebs et al, 1937





Krebs Cycle, 1938



Krebs et al. 1938. Biochem Journal. 32:113

Beadle 1941

Genetic Control of Biochemical Reactions in Neurospora

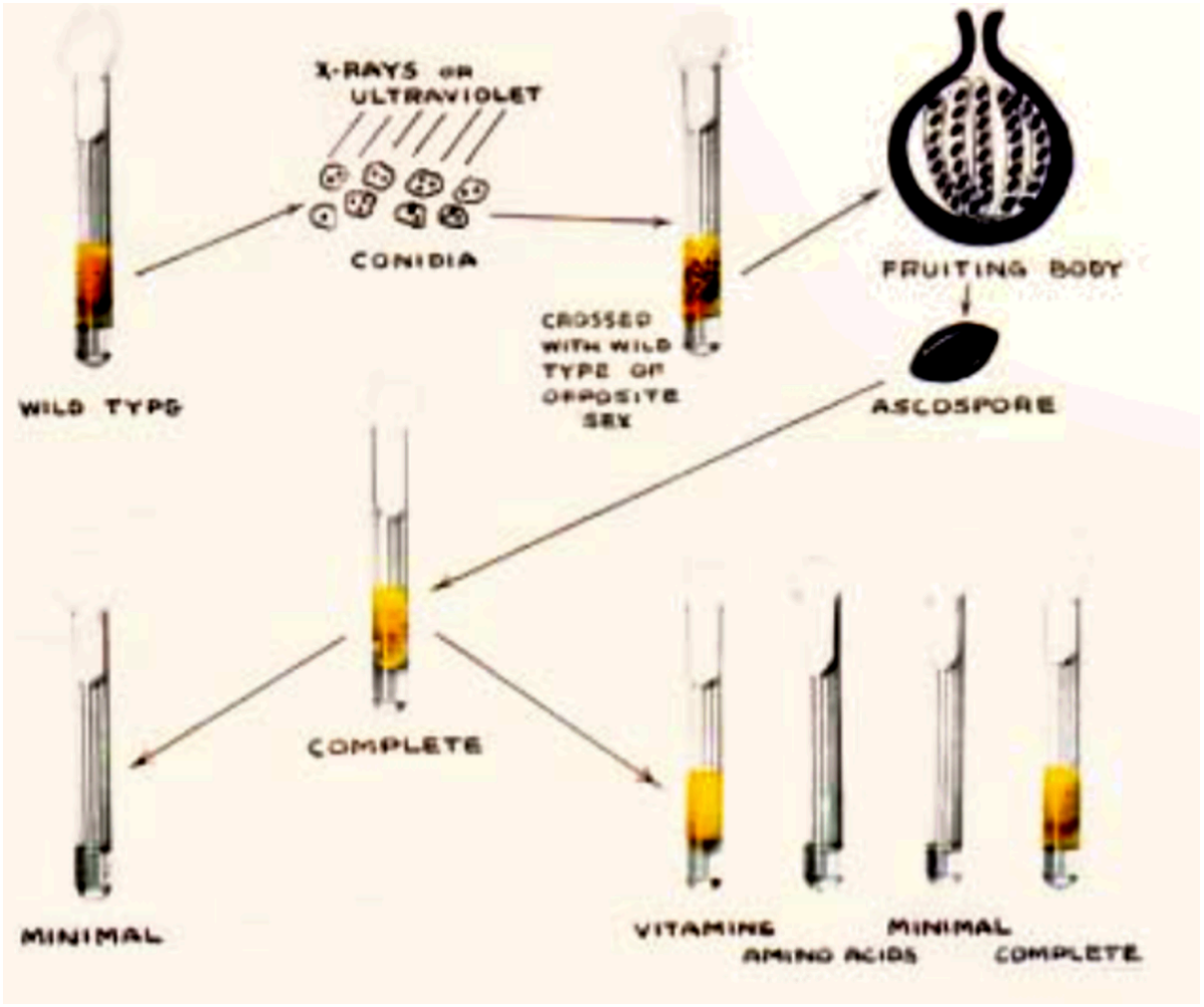
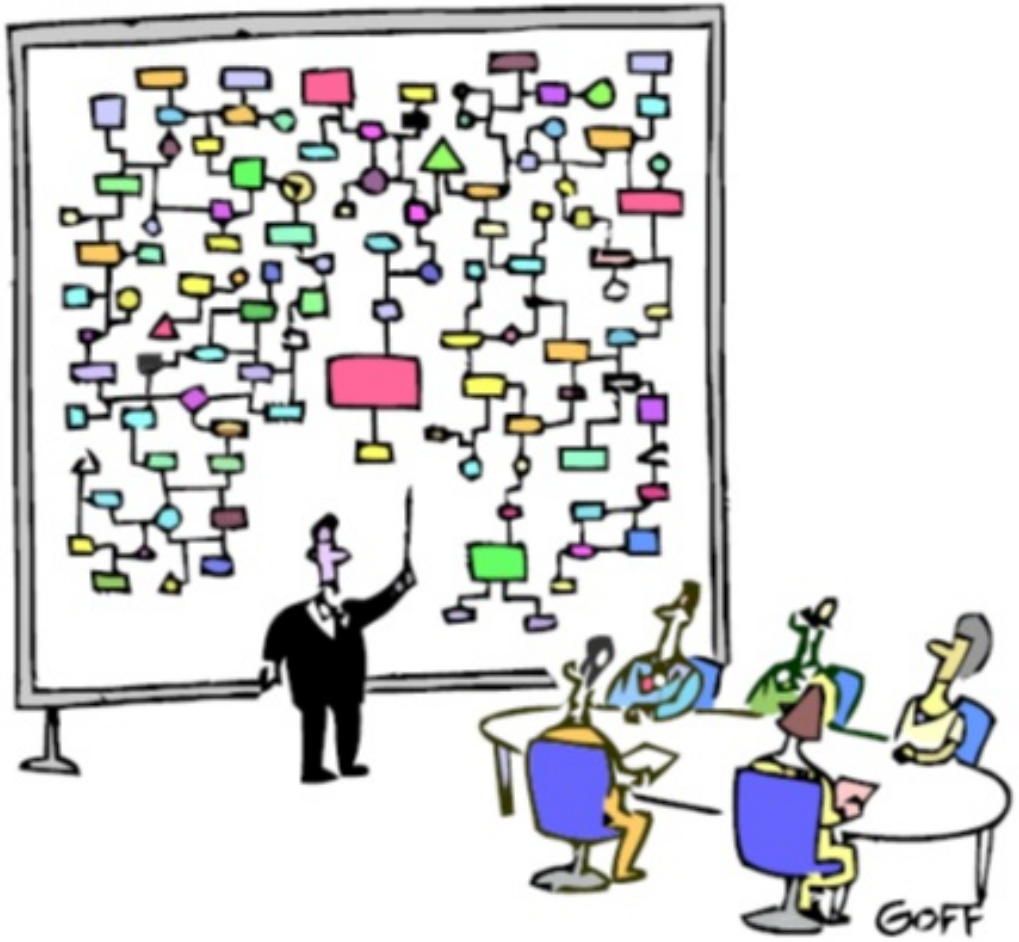


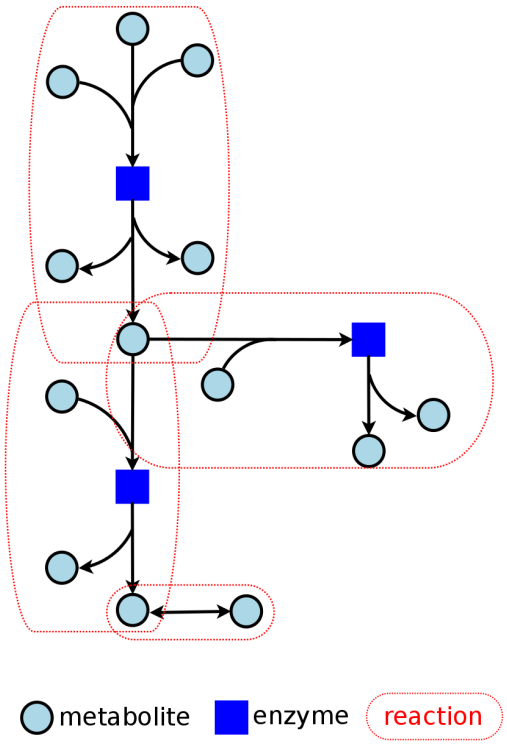
Figure from Nature Reviews Genetics **5**, 949-954, 2004

Metabolic pathways are curated human knowledge



"And that's why we need a computer."

Metabolic models



- A metabolic model consists of metabolites, enzymes, reactions, pathways
- Reactions can be described by differential equations (mathematical models)

$$\begin{array}{c|ccccccc} & v_1 & v_2 & b_1 & b_2 & b_3 & b_4 & b_5 \\ A & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ B & 1 & -1 & 0 & -1 & 0 & 0 & 0 \\ C & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ D & -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ E & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}$$

Stoichiometric matrix

$$\frac{d}{dt}\begin{pmatrix} A \\ B \\ C \\ D \\ E \end{pmatrix} = \mathbf{Sv} = \mathbf{S}\begin{pmatrix} v_1 \\ v_2 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{pmatrix} = \mathbf{0}$$

Steady State Mass balance

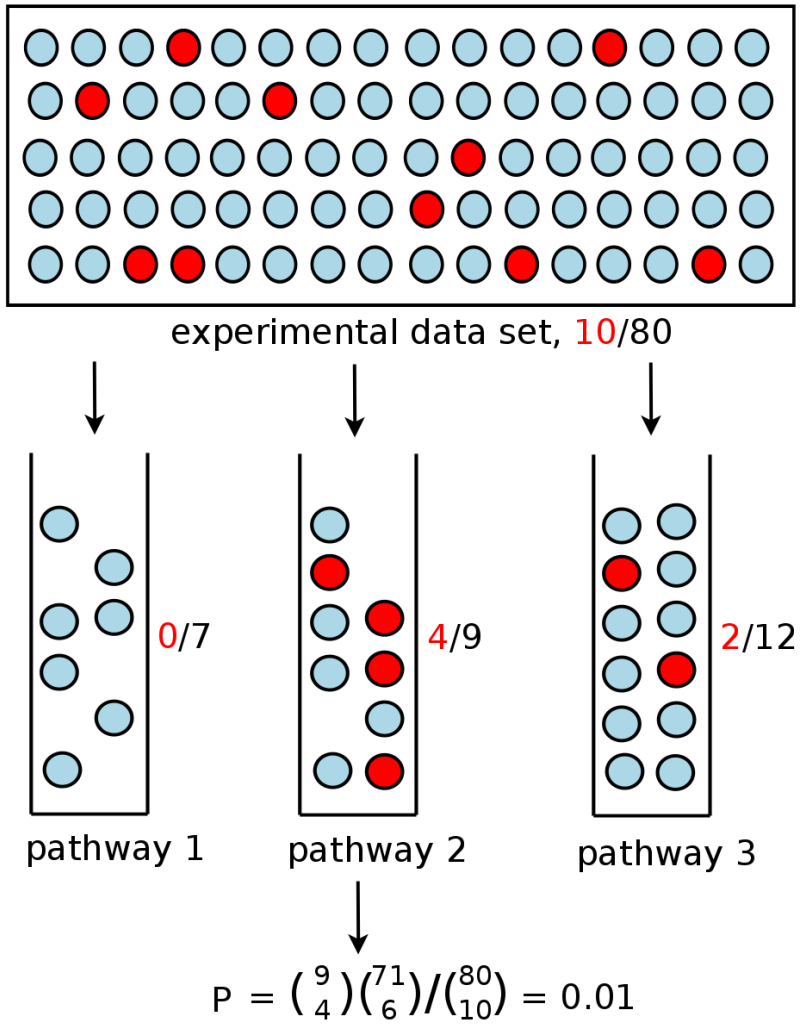
- We focus on statistical models; flux models are not covered here
- Pathways and networks are mathematically indistinguishable

Outline

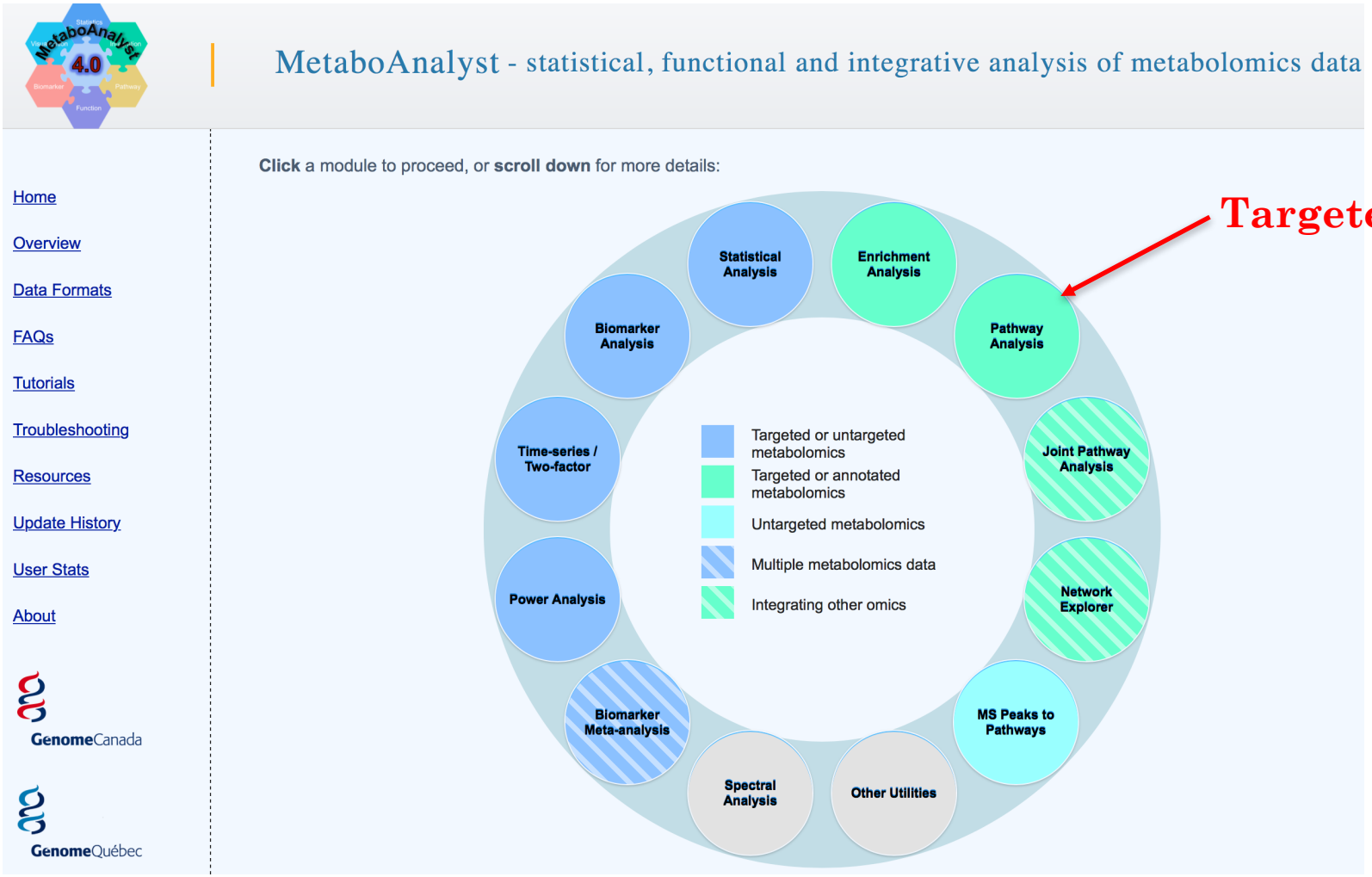
- What are pathways and metabolic models
- **Pathway analysis for targeted data**
- Untargeted data analysis and *mummichog*
- Application examples
  - Intracellular mechanisms
  - Population studies
  - Multi-omics integration
- Resources and issues

Pathway enrichment test

If metabolites are known; red are significant metabolites



MetaboAnalyst



Exercise 1-3

[Download link will be given at workshop]

1. Targeted pathway analysis using MetaboAnalyst
   https://www.metaboanalyst.ca

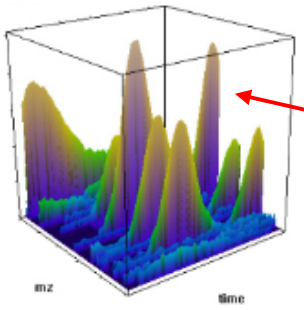
2. Enrichment analysis in command line
   [optional]

3. Pathway visualization – mapping, KEGG painting
   https://www.genome.jp/kegg/tool/map_pathway2.html

Outline

- What are pathways and metabolic models
- Pathway analysis for targeted data
- **Untargeted data analysis and *mummichog***
- Application examples
  - Intracellular mechanisms
  - Population studies
  - Multi-omics integration
- Resources and issues

Untargeted metabolomics data



mass/charge (m/z)

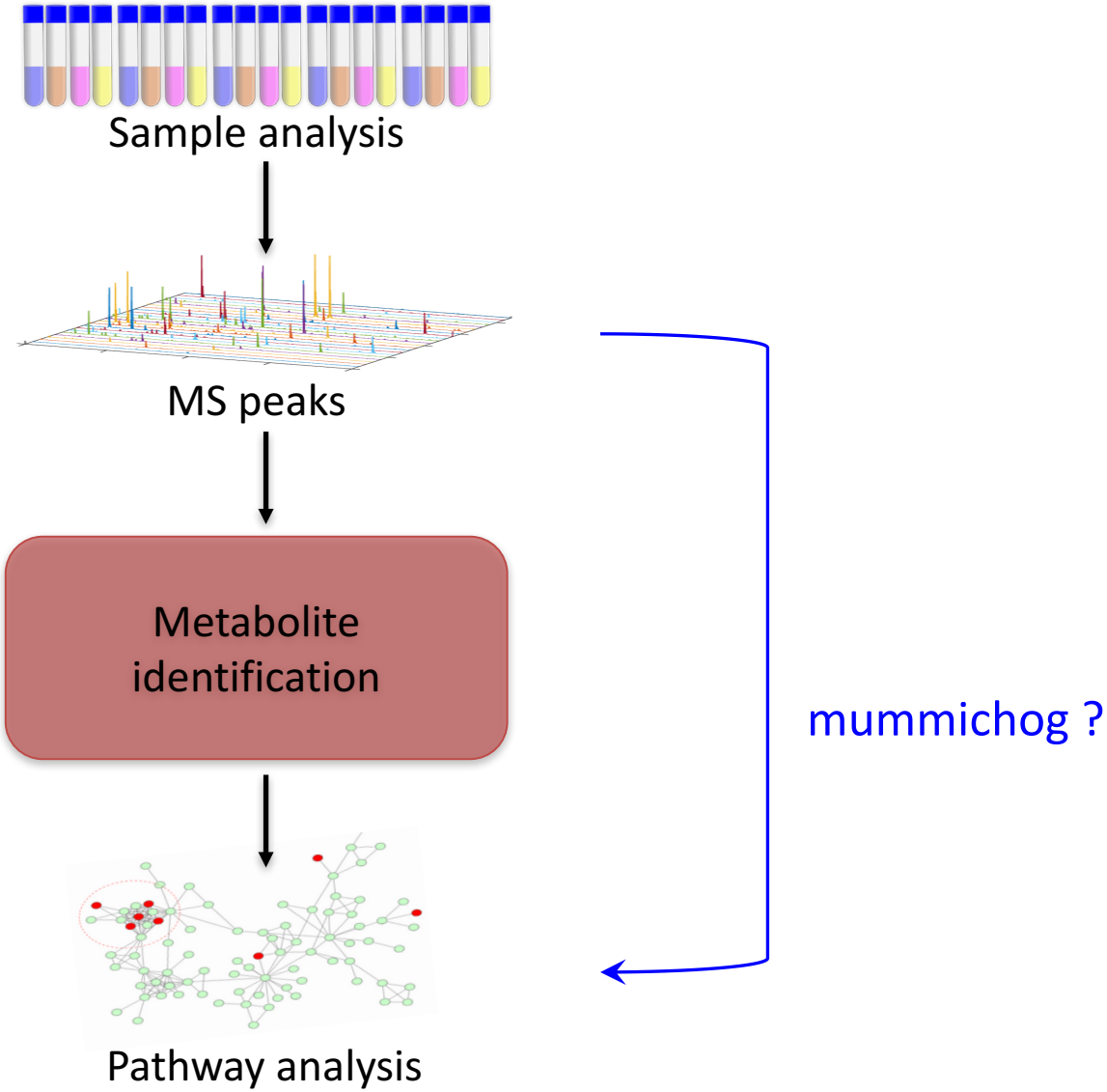
retention time

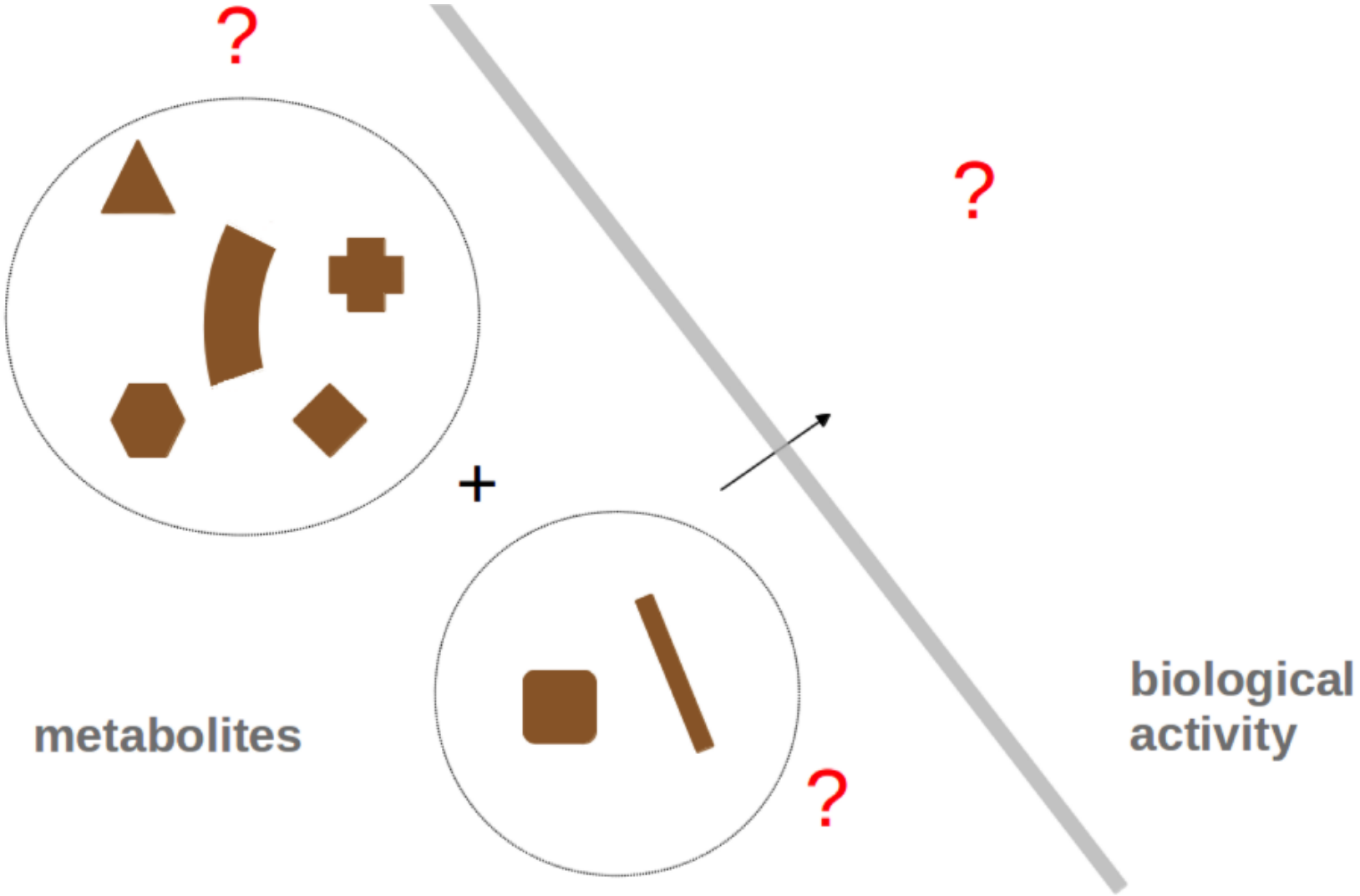
intensity

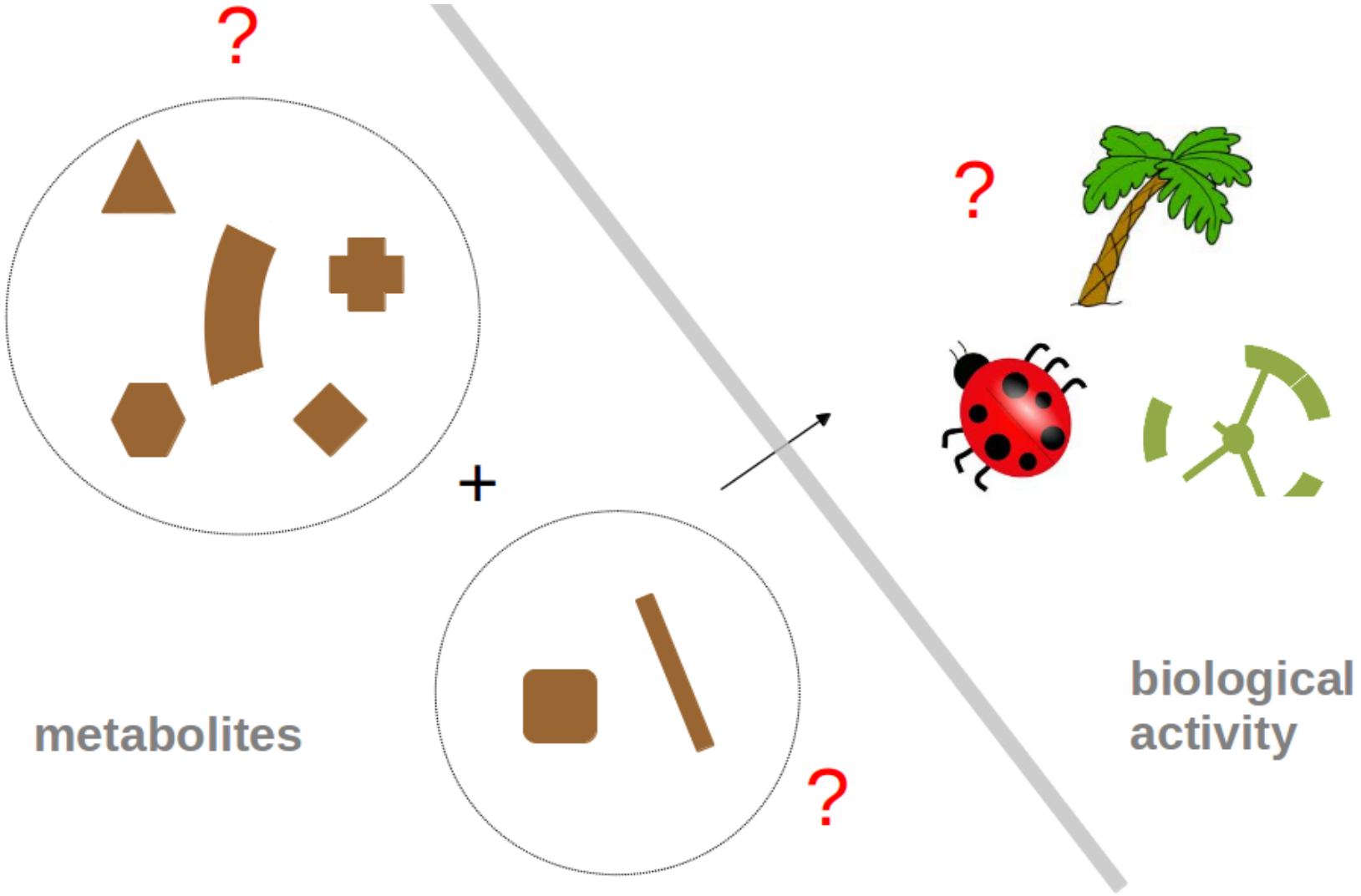
feature

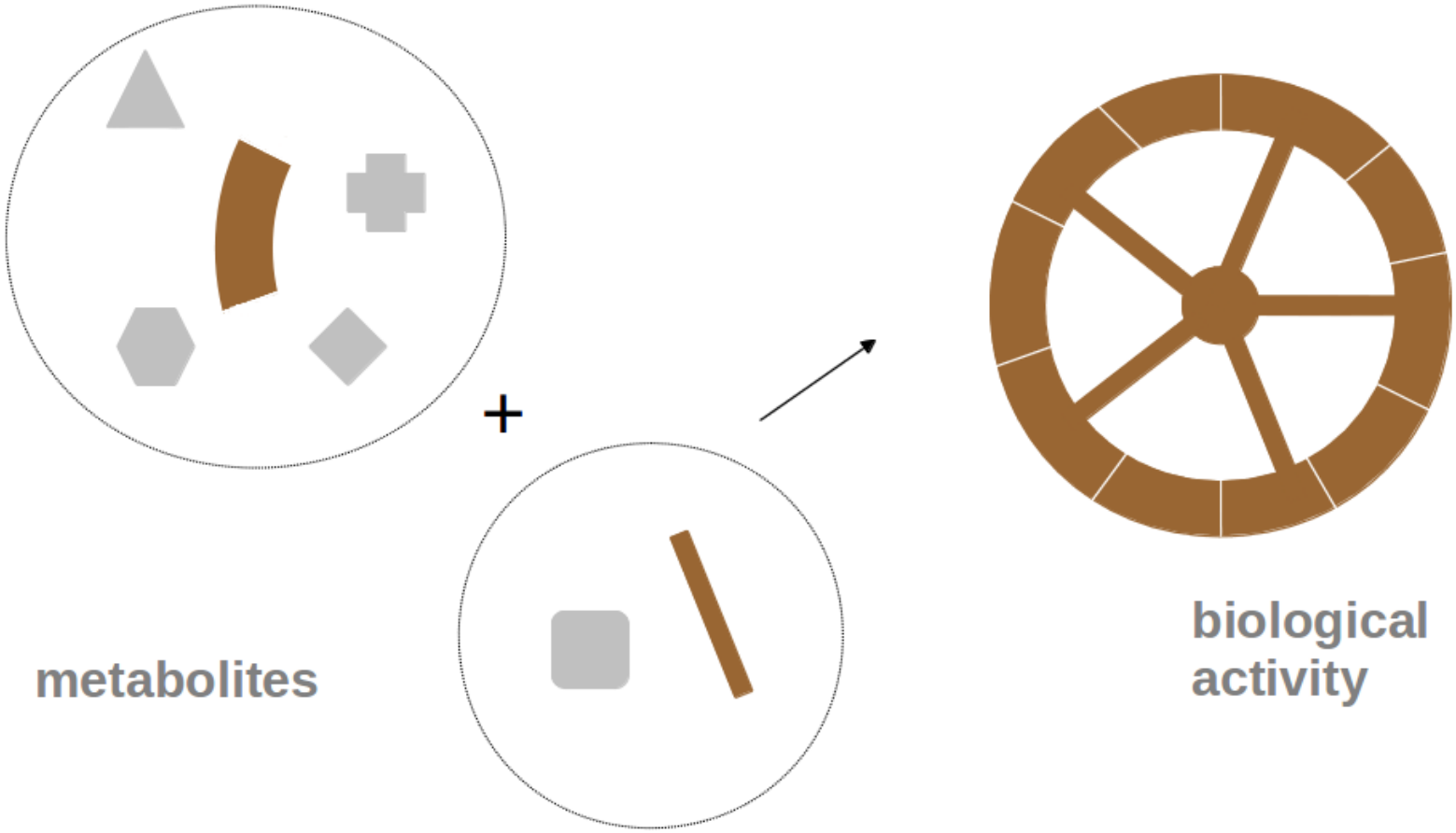
m/z	Retention time	Sample_1	Sample_2	Sample_3
702.86649413	70.4811138781	56905.1184820213	78201.4457063353	24695.6510487344
715.4970731384	56.55139775	83341.1725499997	1085.6171033898	0
715.6080685464	57.1626518527	93484.5395499998	0	0
722.705208561	65.2698532639	55396.5458522493	57091.7796614066	110944.335033457
730.7388823899	66.8110957481	60206.9068679558	62566.9945317202	88473.0389390921
732.7358893181	67.2539834098	95115.557321815	121927.272992316	125077.730263205
734.7314651732	69.0473678869	18862.86035	18500.1329562712	35179.9382
740.7679844539	66.5585431864	104536.742757076	143943.802149299	91120.9347911557
742.7651707472	67.4809438574	103257.45203488	134695.193645035	104576.322126369
750.7969865893	67.5681700041	117046.427960669	107939.940388836	73107.5173642583
752.793714331	67.3189767434	60630.5403168073	63435.6029427167	34751.816765529
760.8266370588	68.7380703739	41461.8047407238	54465.3458564024	28357.2944064846
102.0663370112	95.4437104918	19665.4401722973	101522.479348442	102707.039800486
112.3625162102	324.2816800135	151985.843922093	17555.644	57954.9573157841
121.072097577	88.0562910071	38408.6990506757	81408.9040789102	96405.1045547668
136.0486704167	74.5368317344	55806.64625	14052.7744	0

Fastlane from chemistry to biology?

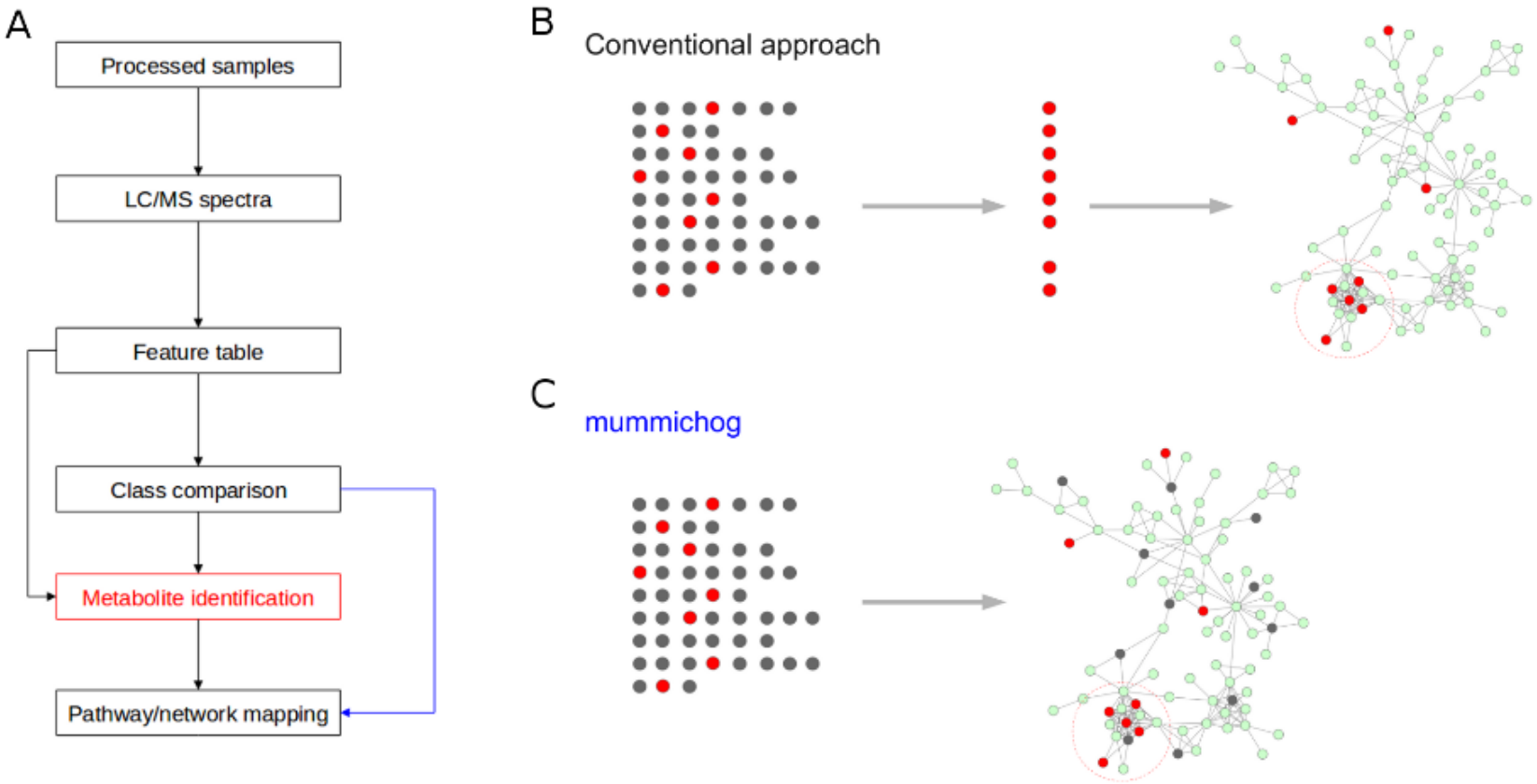



?
?
+
?
metabolites
biological
activity


?
+
?
?
metabolites
biological
activity

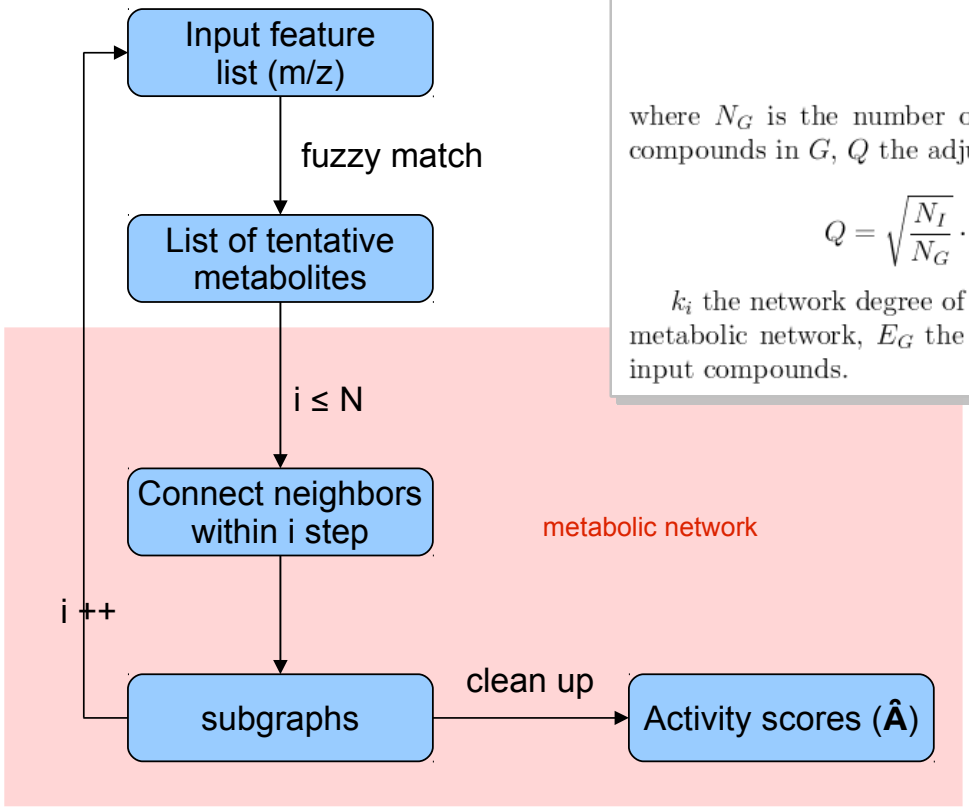

metabolites
biological
activity

Mummichog tests metabolite grouping patterns



Li et al. 2013. PLoS Computational Biology. 9:e10031323

Module analysis in *mummichog*



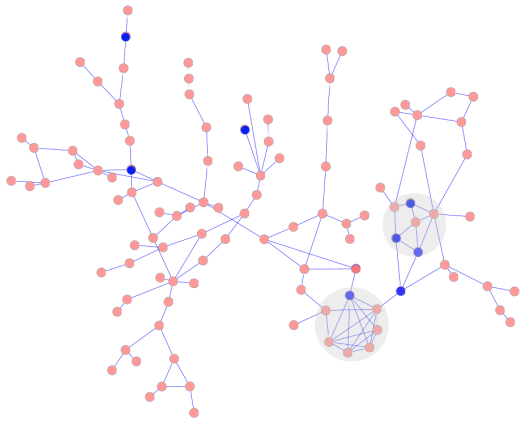
For a subgraph G, activity score

$$\hat{A} = Q \cdot \frac{N_{I,G}}{N_G}, \quad (1)$$

where N_G is the number of compounds in G, $N_{I,G}$ the number of input compounds in G, Q the adjusted Newman-Girvan modularity:

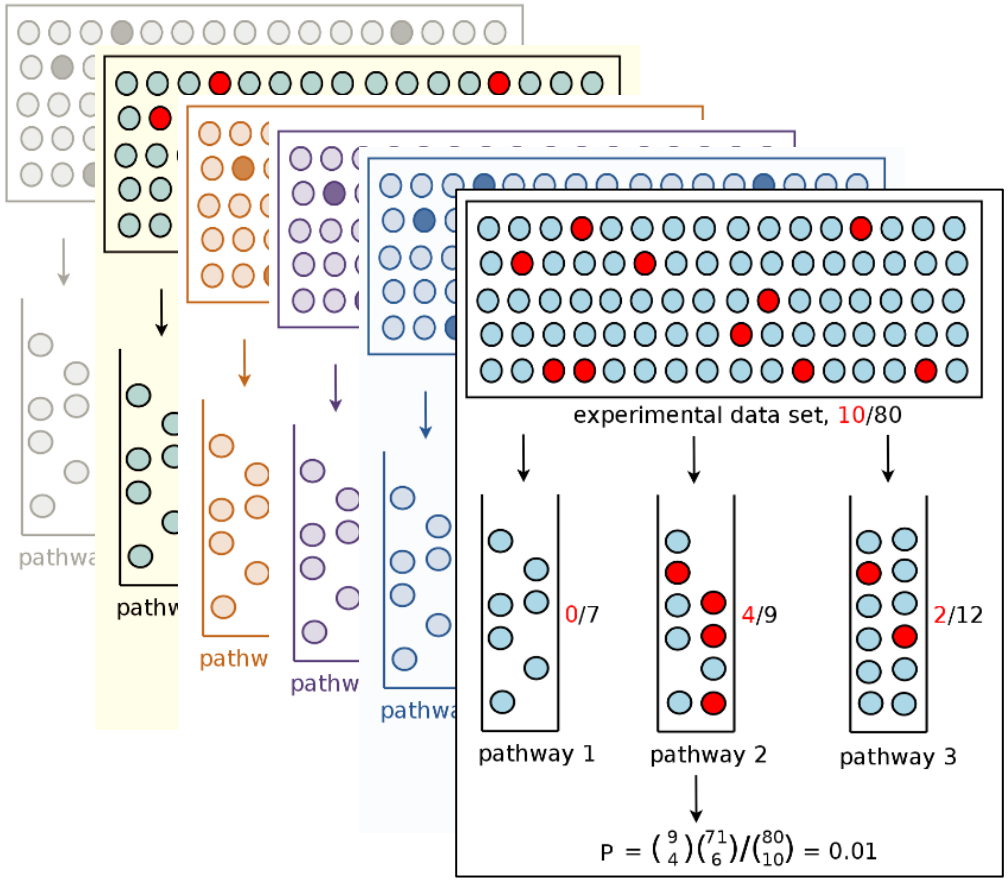
$$Q = \sqrt{\frac{N_I}{N_G}} \cdot \left(\frac{E_G}{m} - \sum_{i,j} \frac{k_i}{2m}\frac{k_j}{2m}\right) \quad , \quad i,j \in G \quad (2)$$

k_i the network degree of compound i, m the total number of edges in the metabolic network, E_G the total number of edges in G, N_I the number of input compounds.



Li et al. 2013. PLoS Computational Biology. 9:e10031323

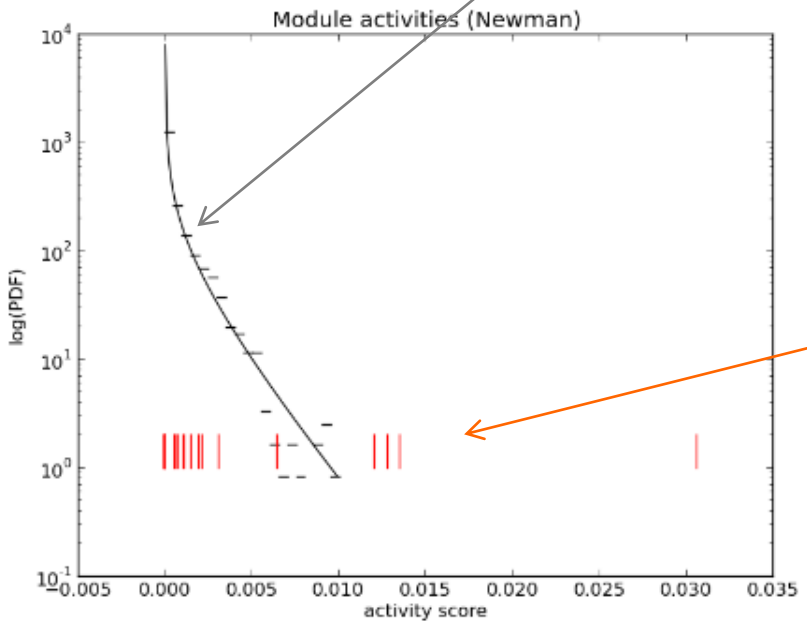
Pathway analysis in *mummichog*



Li et al. 2013. PLoS Computational Biology. 9:e10031323

Testing module/pathway significance in *mummichog*

Distribution of permutated data (null distribution)



Distribution of real data from user's significant m/z list

On the organization of pathways

- Pathway definition may be incomplete, different between databases and species
- Pathways may not be what you measured - compartmentalization, sampling, metabolite stability.
- Metabolic reactions connect to networks
- Modules are based on network topology, less biased by human definition but data dependent. A module can be within a pathway or in between several pathways.
- Other thoughts: metabolite sets, chemical similarity
- A major mission of metabolomics is to update biochemical pathways

mummichog.org

Mummichog pathway and network analysis for metabolomics

Home | Software | Publications | Notebooks | Version 1.0.9 | mummichog-server

Mummichog is a free Python program for analyzing data from high throughput, untargeted metabolomics. It leverages the organization of metabolic networks to predict functional activity directly from feature tables, bypassing metabolite identification. Thus, high-quality hypotheses can be quickly generated from a LC-MS data table.

Download and use mummichog. Now via standard Python package index.

OS independent install (version 1):
`pip install mummichog1`

Or mummichog 2 (beta test):
`pip install mummichog`

Run it in command line:
`mummichog -f myData -o myResult`

More to click "Software".

Example publications that were supported by using mummichog

Li et al. (2017) Metabolic Phenotypes of Response to Vaccination in Humans. *Cell* 169(5): p862-877.

Huan, Tao, et al. "Systems biology guided by XCMS Online metabolomics." *Nature methods* 14.5 (2017): 461.

Xu et al. "Autophagy is essential for effector CD8+ T cell survival and memory formation." *Nature immunology* 15.12 (2014): 1152-1161.

More to click "Publications".

Jupyter notebooks for data analysis in metabolomics and systems biology

Jupyter notebooks are to data people like lab notebooks to bench scientists. They keep code and result in the same web browser. It's increasinlgy popular for data analysis and collaboration. We continue posting notebooks as tutorials and for record keeping.
More to click "Notebooks".

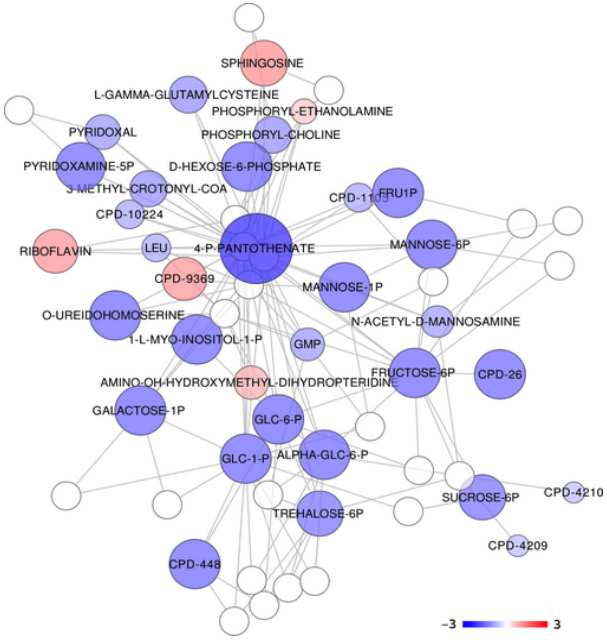
News

- Web version is now testing at here.
- Mummichog 2 is now on GitHub.
- Mummichog 1 with a web interface is now available on MetaboAnalyst 4.
- A common error, "AttributeError: 'NodeView' object has no attribute 'sort'", is caused by Networkx 2.x, which is not backward compatible. This can be fixed by installing Networkx 1.x via pip in your terminal, "sudo pip install networkx==1.10".
- Mummichog 2 test version is available at Pypi.
- Mummichog version 1.0.10 is available at Pypi.
- Mummichog helped decipher metabolic phenotypes in human vaccination - Li et al. (2017) at *Cell* 169(5): p862-877. Also see commentary "Orthogonal Data Integration to Define Immunometabolic-Phenotypes" at *Cell Systems*.
- Emory University Sys/Bio/Info Group runs a monthly meeting on Systems Biology and Bioinformatics. These are casual seminars and discussions, on every first Wednesday of the month, 12 pm, Whitehead Biomedical Research Building, Rm 200. Email Dr. Shuzhao Li to join the mailing list.

Copyright © 2016-2018

Exercise 4-6

4. Explore mummichog output
5. Using barplot to visualize pathway significance
6. Network visualization using Cytoscape: http://cytoscape.org [On your own]



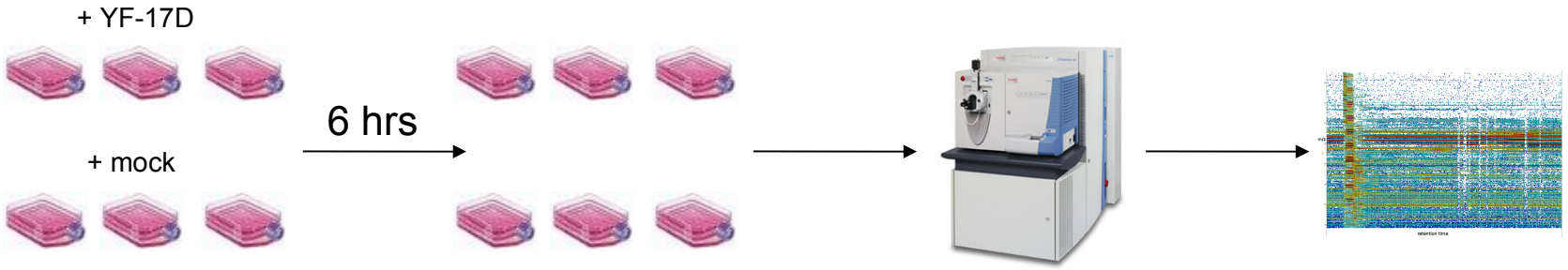
Example using Cytoscape to visualize mummichog metabolite module (Hoffman et al, 2014. Aging Cell, 13: 596)

Outline

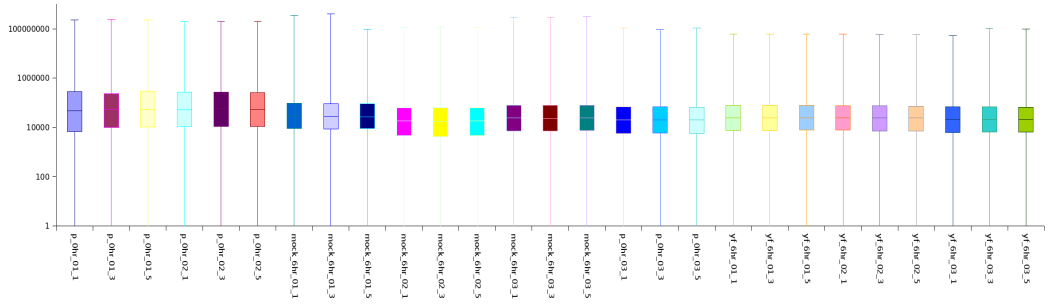
- What are pathways and metabolic models
- Pathway analysis for targeted data
- Untargeted data analysis and *mummichog*
- **Application examples**
  - **Intracellular mechanisms**
  - **Population studies**
  - Multi-omics integration
- Resources and issues

Case study: viral activation of immune cells

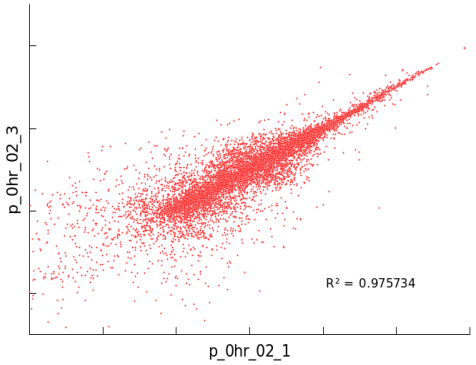
Monocyte derived dendritic cells (moDC)



QA: total ion counts are similar among samples

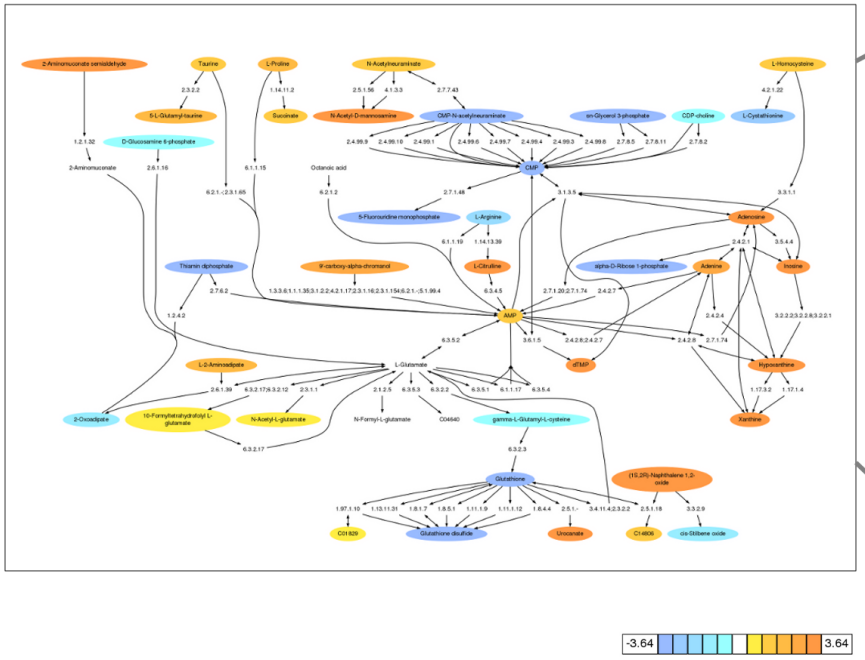


technical replicates, 10,000 features

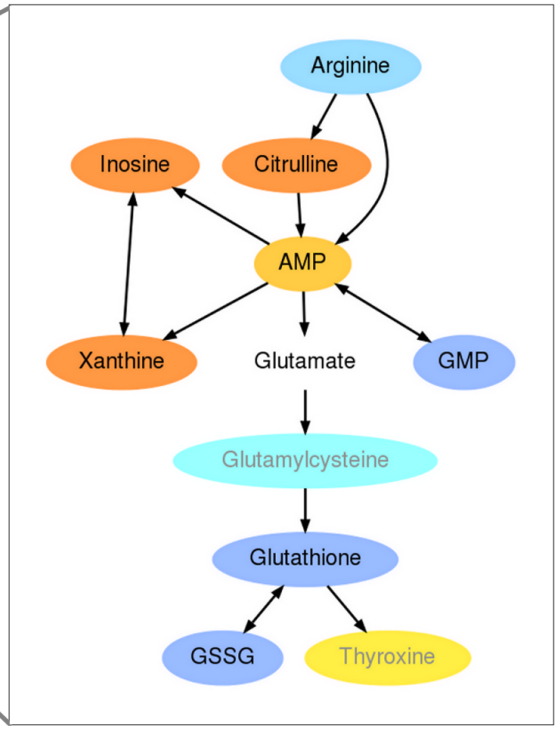


Mummichog: viral activation of immune cells

A

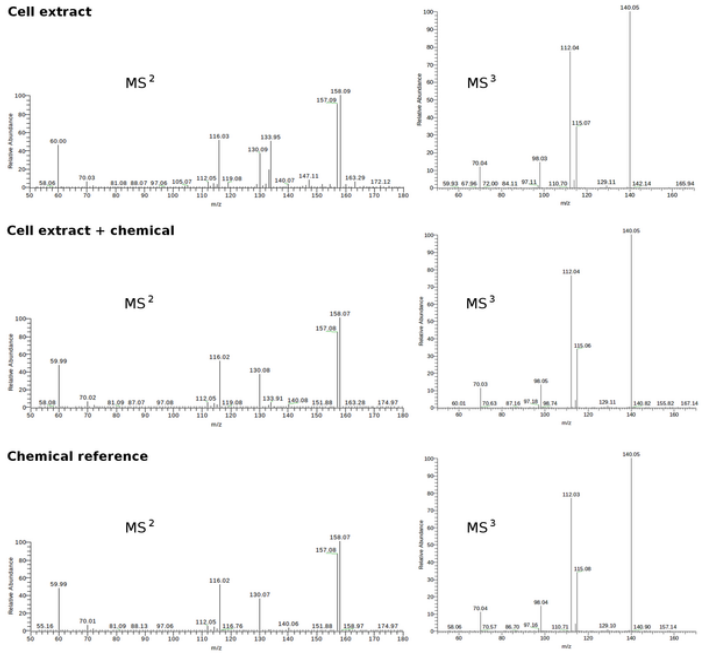


B

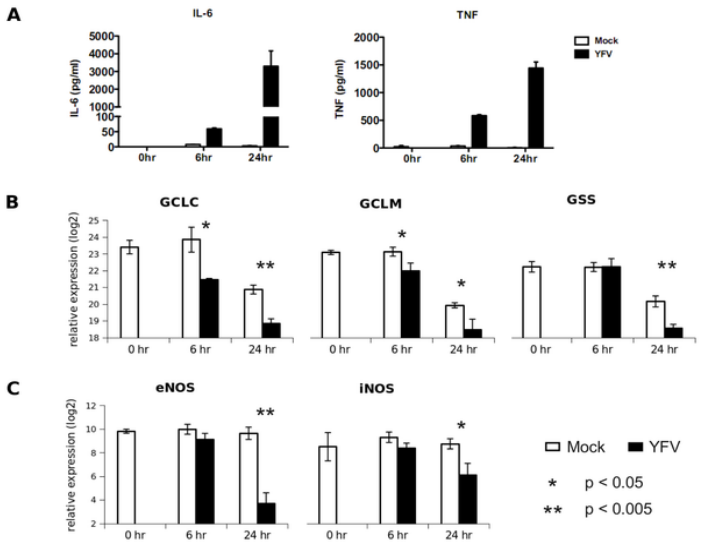


Li et al. 2013. PLoS Computational Biology. 9:e10031323

Experimental validation of *mummichog* prediction



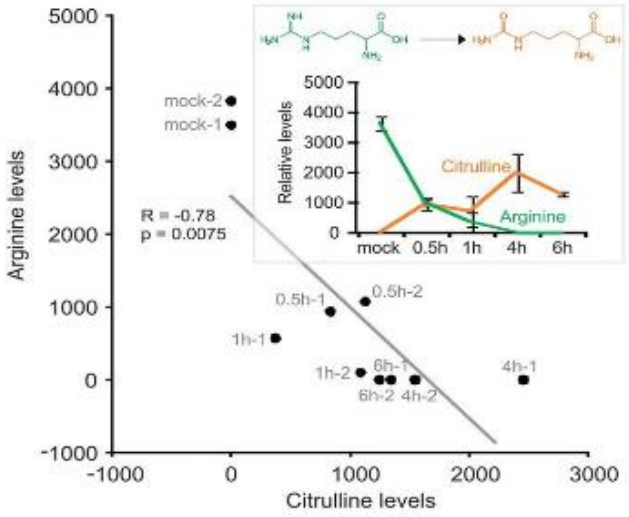
Tandem mass spectrometry confirmed 9/11 metabolites



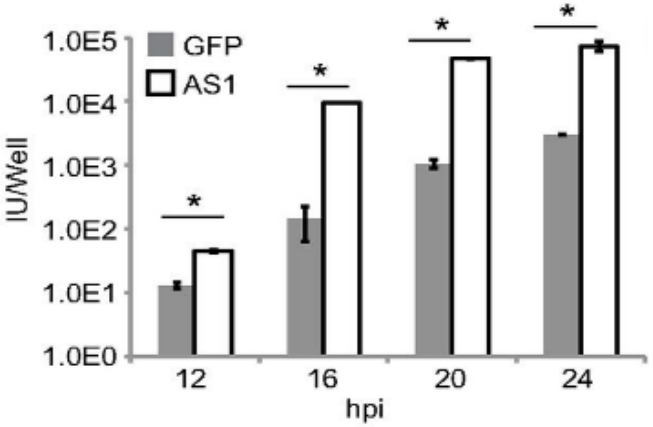
Gene expression supported GSH/GSSG depletion and Arg/Cit conversion

Li et al. 2013. PLoS Computational Biology. 9:e10031323

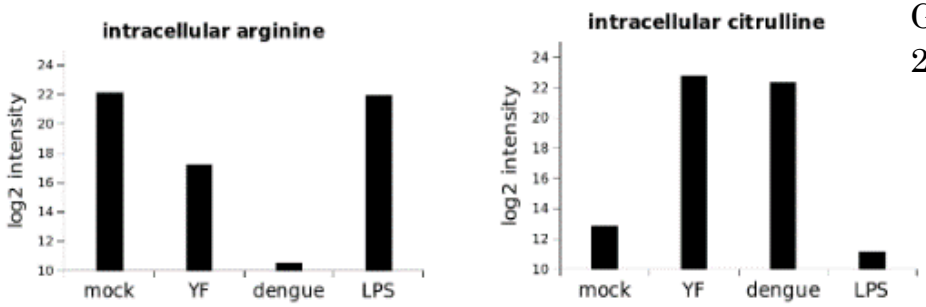
Arginine as master regulator of viral response



Ravindran et al. 2014. Science 343:313

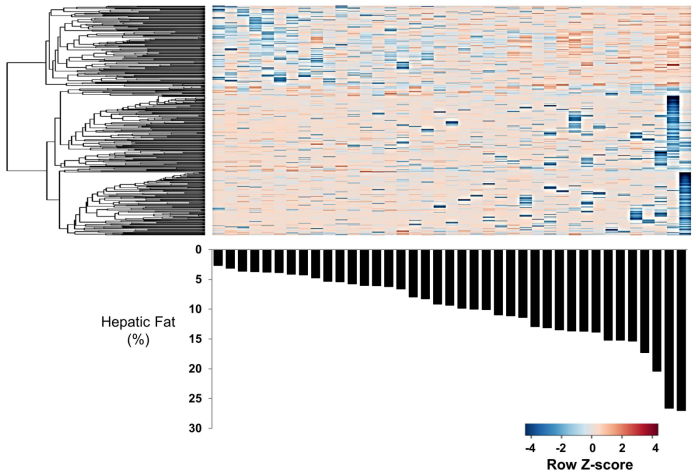


Argininosuccinate synthetase 1 knockdown led to increased replication of HSV-1.
Grady, Purdy, Rabinowitz & Shenk. 2013. PNAS 110:E5006.



Li et al. 2013. PLoS Computational Biology. 9:e10031323

MWAS + *mummichog* (NAFLD)

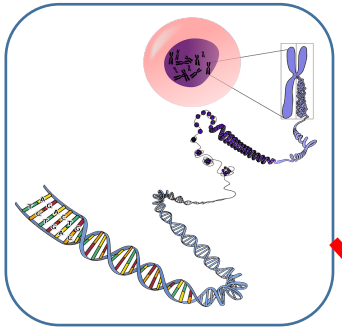


Pathway	Overlap_size	Pathway size	Model p-value
Vitamin E metabolism	9	32	0.00095
Drug metabolism - cytochrome P450	8	34	0.00196
Tyrosine metabolism	15	79	0.00202
Vitamin B2 (riboflavin) metabolism	3	6	0.00229
Purine metabolism	10	51	0.00332
Ascorbate (Vitamin C) and Aldarate Metabolism	4	16	0.00773
Vitamin B9 (folate) metabolism	4	18	0.01307
Glutamate metabolism	3	12	0.01834
Methionine and cysteine metabolism	7	42	0.02026
Alanine and Aspartate Metabolism	4	20	0.02159
Biopterin metabolism	3	13	0.02493
Di-unsaturated fatty acid beta-oxidation	3	13	0.02493
Histidine metabolism	4	22	0.03449
Glycine, serine, alanine and threonine metabolism	8	53	0.03499
Valine, leucine and isoleucine degradation	7	46	0.03894

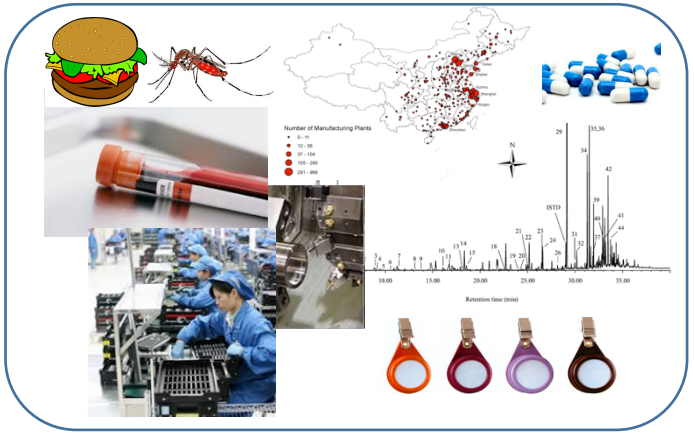
Jin, Banton, et al., 2016.
Amino Acid Metabolism is Altered in Adolescents with Nonalcoholic Fatty Liver Disease - An Untargeted, High Resolution Metabolomics Study.
The Journal of pediatrics 172: 14

G × M × E

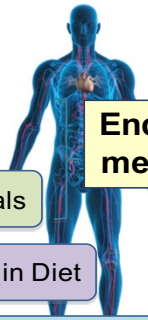
Genome



Environment



Gene function



- Core Biological Metabolome
- Microbiome-related Chemicals
- Non-nutritive Chemicals in Diet
- Supplements and Pharmaceuticals

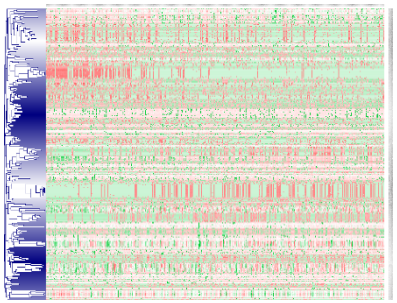
Endogenous metabolome

Environmental metabolome

- Commercial Products
- Environmental Chemicals

Body burden

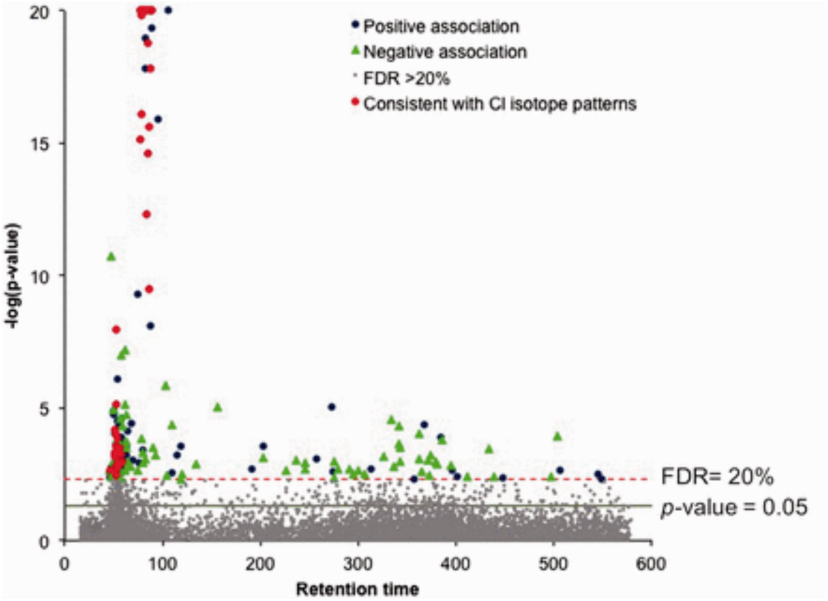
Molecular response

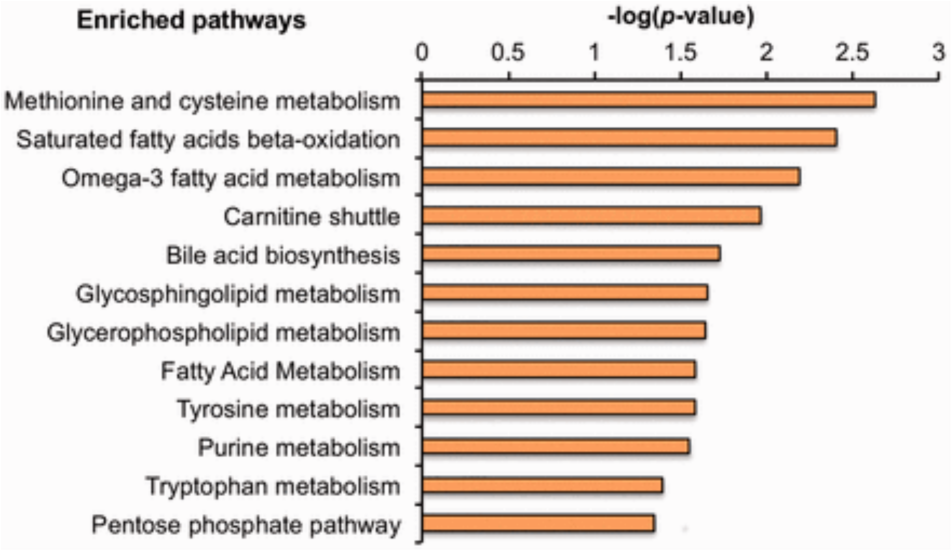


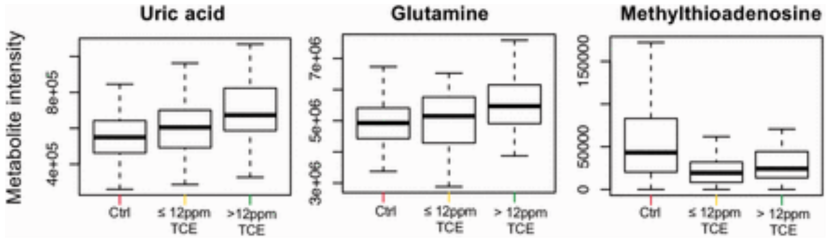
Metabolome

Modified based on figures by Doug Walker & Dean Jones

MWAS of occupational exposure to trichloroethylene







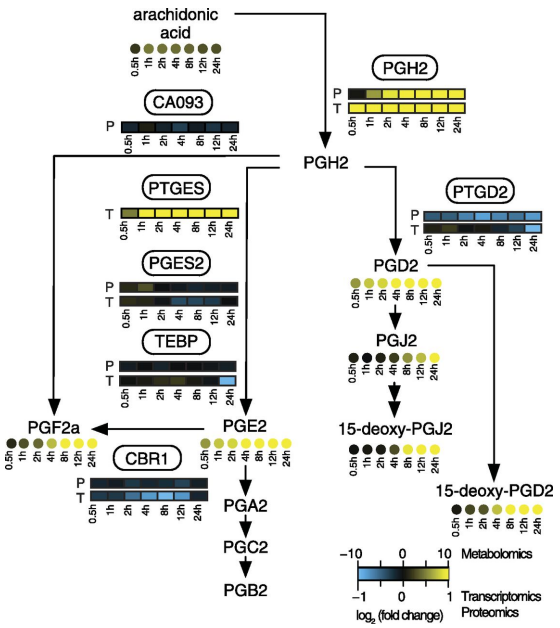
Walker DI. et al (2016), High-resolution metabolomics of occupational exposure to trichloroethylene.
Int J Epidemiol. 45 (5): 1517-1527

Exercise 7-9

7. Importance of experiment design and biological controls

8. MWAS + mummichog

9. More data visualization – heatmap, boxplot

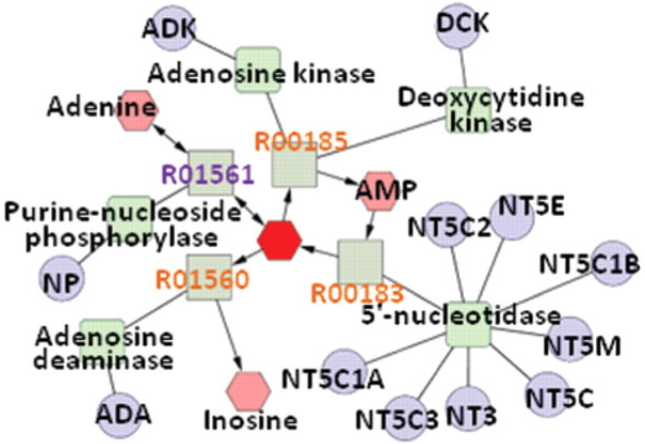


Example of manual layout of multiple data types
Sabido et al, 2012. PMID: 22361236

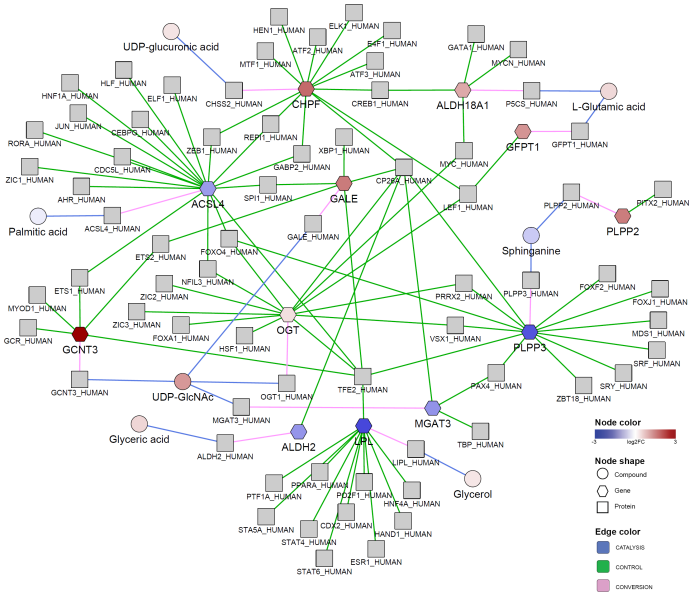
Outline

- What are pathways and metabolic models
- Pathway analysis for targeted data
- Untargeted data analysis and *mummichog*
- Application examples
  - Intracellular mechanisms
  - Population studies
  - **Multi-omics integration**
- Resources and issues

Combining metabolite data with gene expression

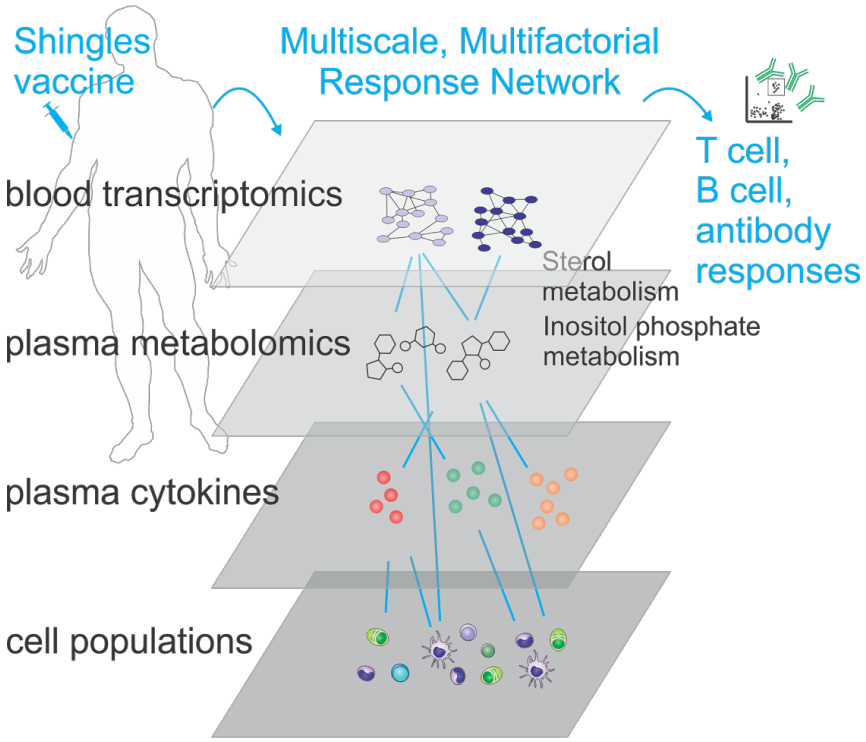


Metscape
Karnovsky et al. (2012)
Bioinformatics. 28:373



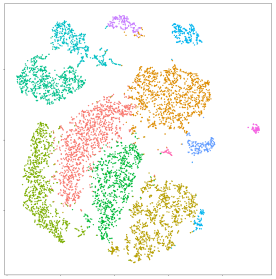
Metabox
Wanichthanarak et al. (2017)
PloS ONE 12(1):e0171046

Multi-omics in immune response

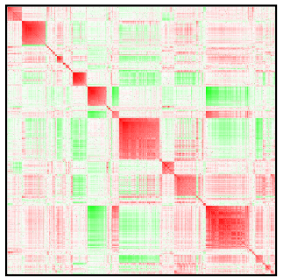


Li et al., (2017) Metabolic Phenotypes of Response to Vaccination in Humans, *Cell* 169 (5), 862-877

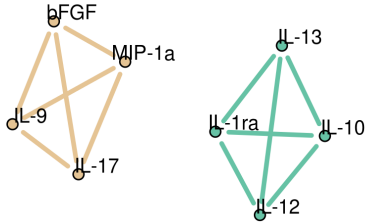
Cell communities



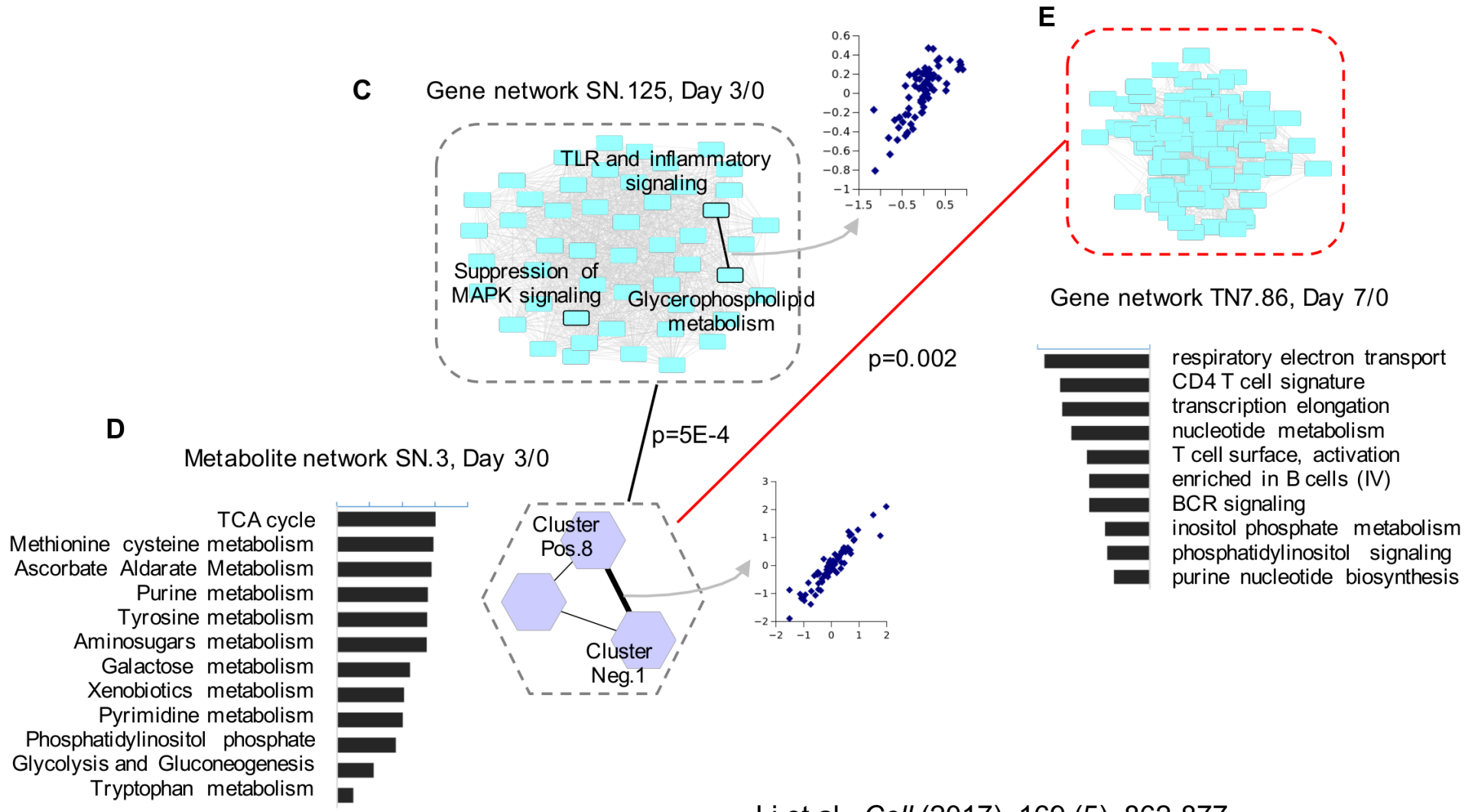
Metabolite communities



Cytokine communities

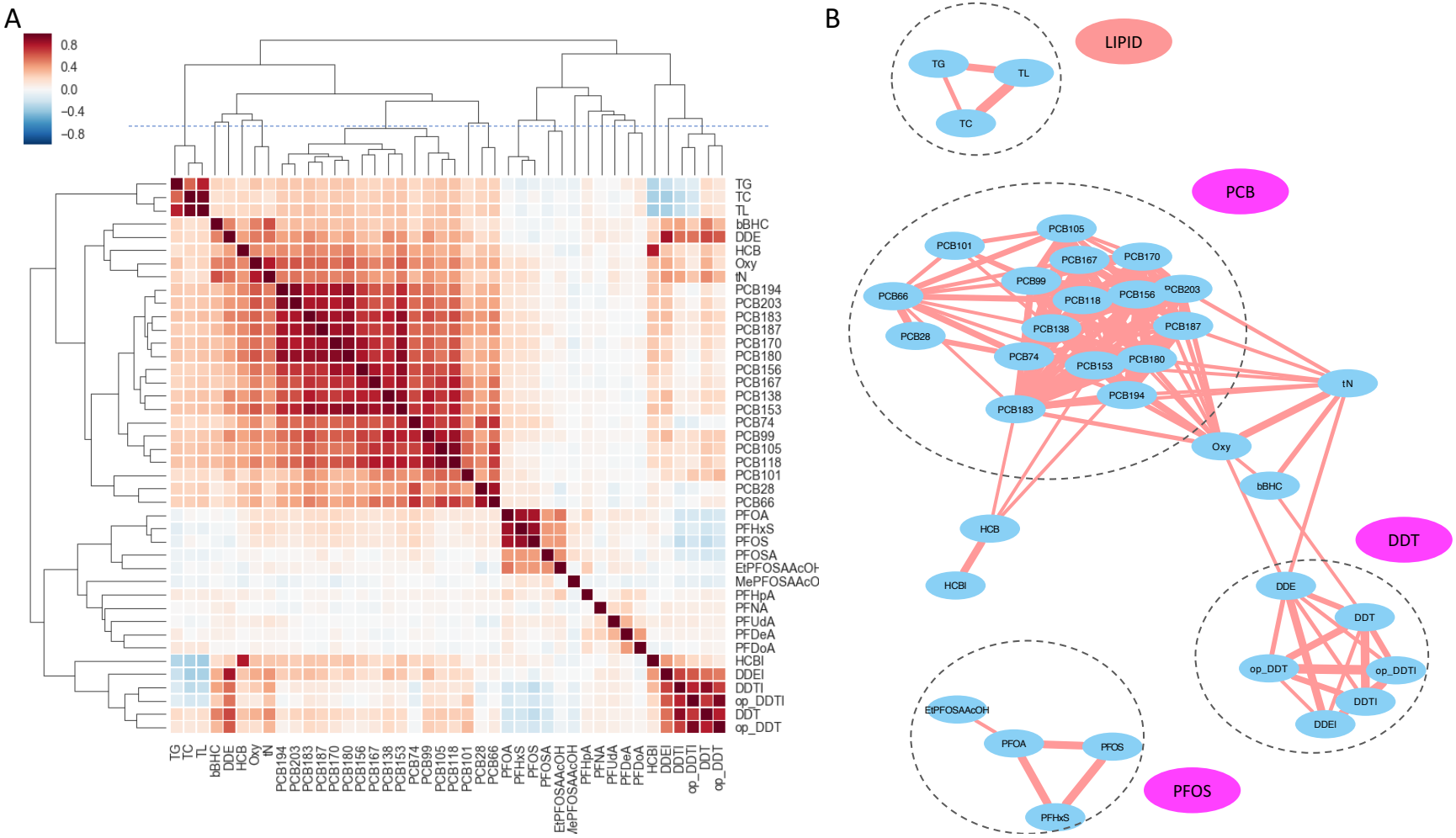


Multi-omics in immune response



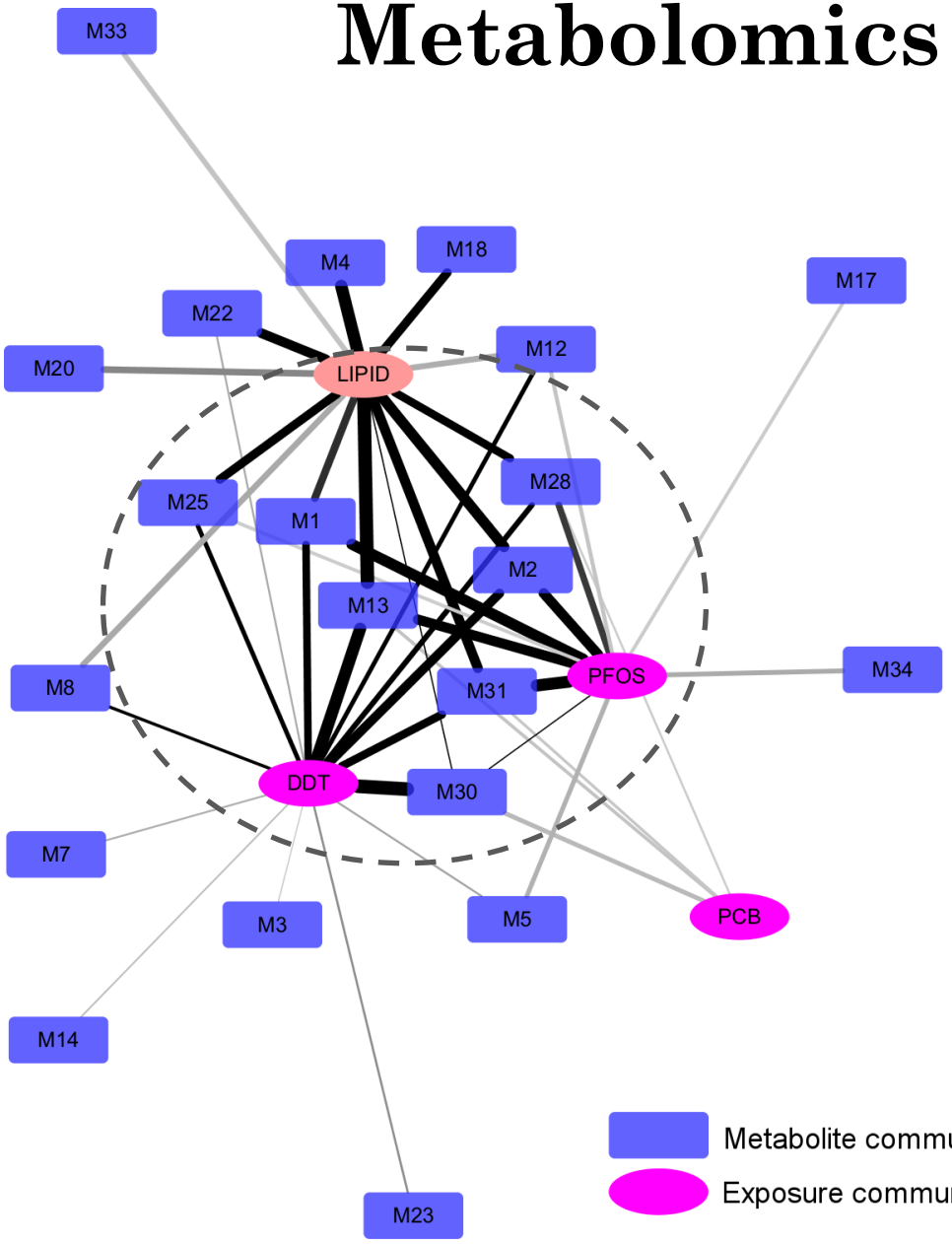
Li et al., *Cell* (2017), 169 (5), 862-877

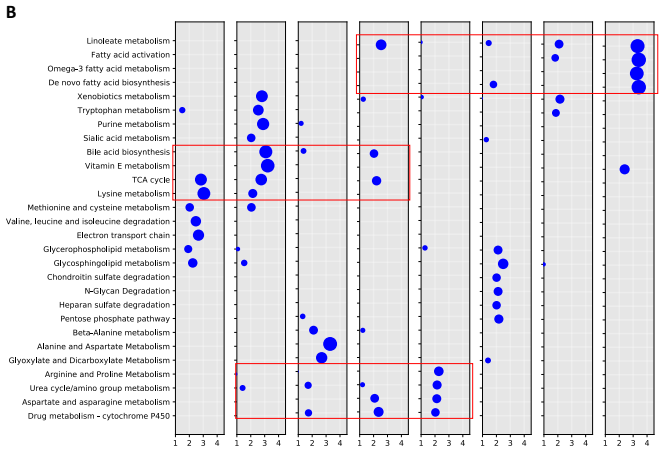
Exposure communities in CHDS data



Li et al., *Under review*

Metabolomics × Exposures





Li et al., *Under review*

Outline

- What are pathways and metabolic models
- Pathway analysis for targeted data
- Untargeted data analysis and *mummichog*
- Application examples
  - Intracellular mechanisms
  - Population studies
  - Multi-omics integration
- **Resources and issues**

Resources

Metabolic pathways and models

KEGG: http://www.genome.jp/kegg/kegg2.html
BioCyc: https://biocyc.org
Recon: https://www.vmh.life/#human/all

Metabolite ID conversion

https://cts.fiehnlab.ucdavis.edu
https://www.metaboanalyst.ca/faces/upload/ConvertView.xhtml

Metabolomics data analysis tools

MetaboAnalyst: https://www.metaboanalyst.ca
MetScape: http://metscape.ncibi.org/
MetExplore: https://metexplore.toulouse.inra.fr/
Metabox/Met-DA: http://metda.fiehnlab.ucdavis.edu
XCMS Online: https://xcmsonline.scripps.edu
Mummichog: http://mummichog.org
Commercial: MetaCore, Ingenuity Pathway Analysis

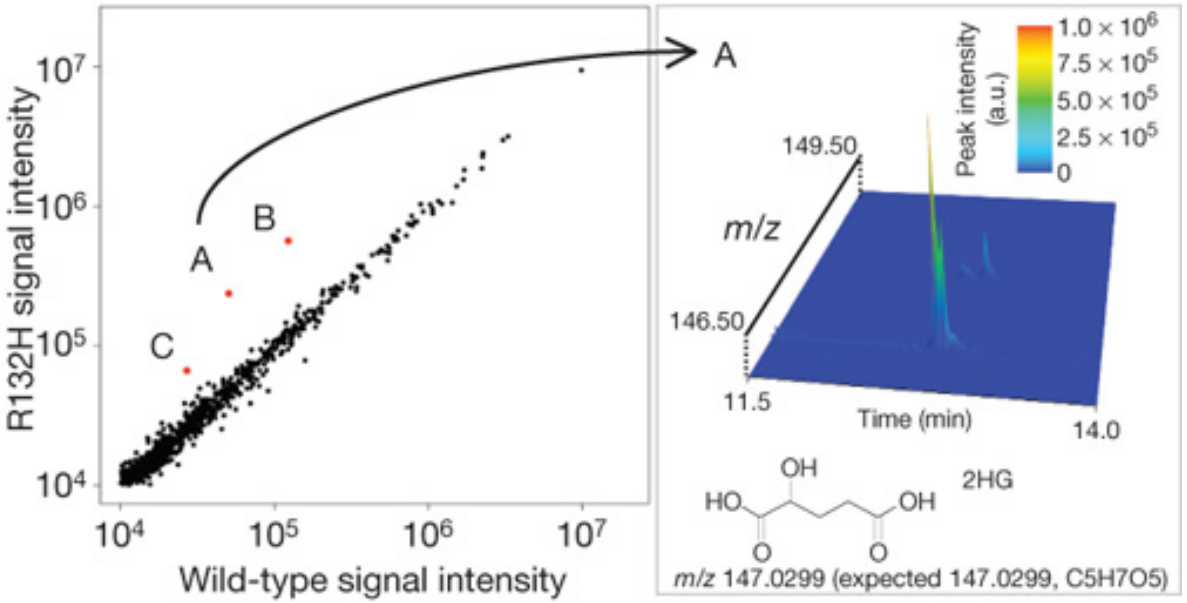
Network visualization tools

Cytoscape: http://cytoscape.org
Gehpi: https://gephi.org

General programming and Data analytics

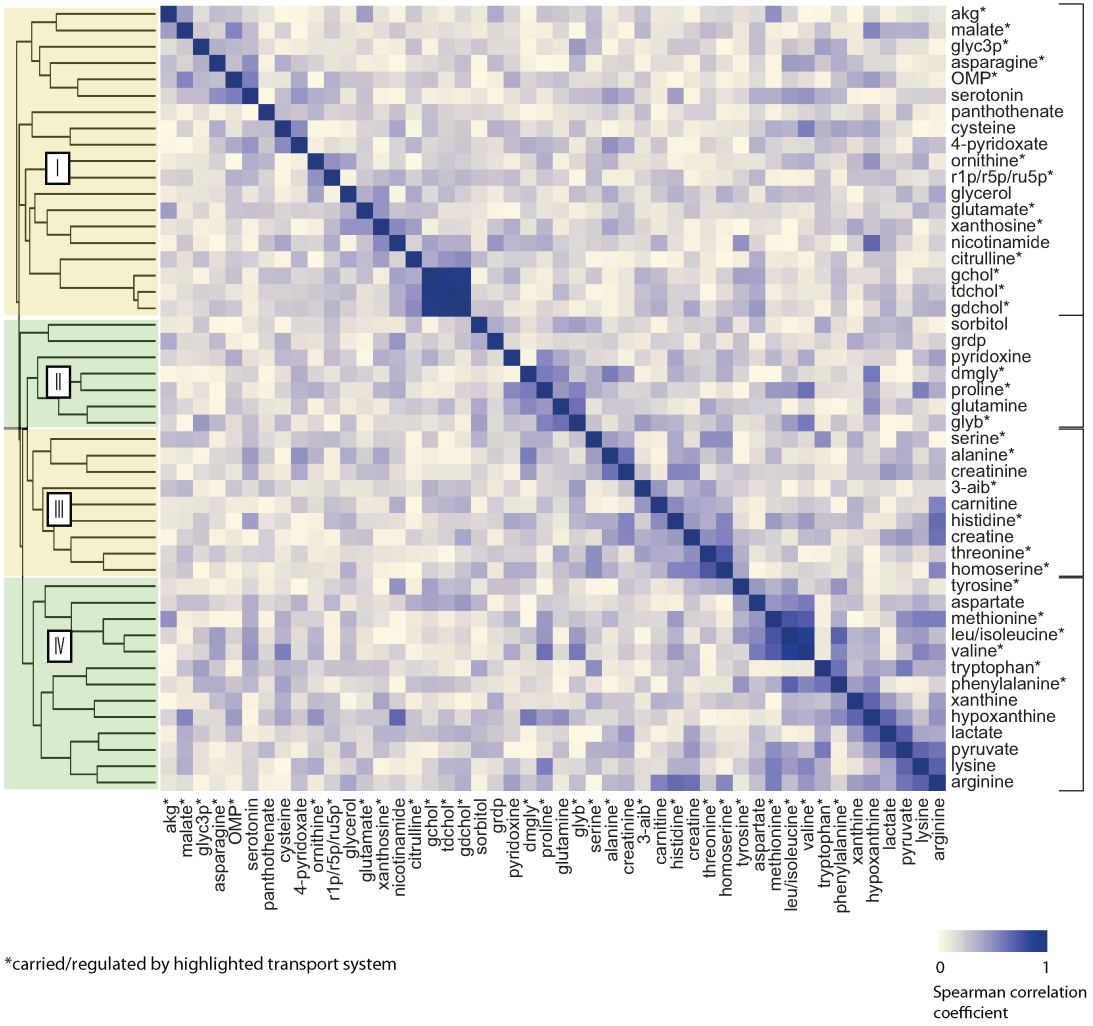
Python – Conda – Pandas – Seaborn
R – Bioconductor – Rstudio
Jupyter Notebooks

Limits of Pathway analysis (I)



Dang *et al. Nature* **462**, 739 (2009) doi:10.1038/nature08617
Cancer-associated IDH1 mutations produce 2-hydroxyglutarate

Limits of Pathway analysis (II)



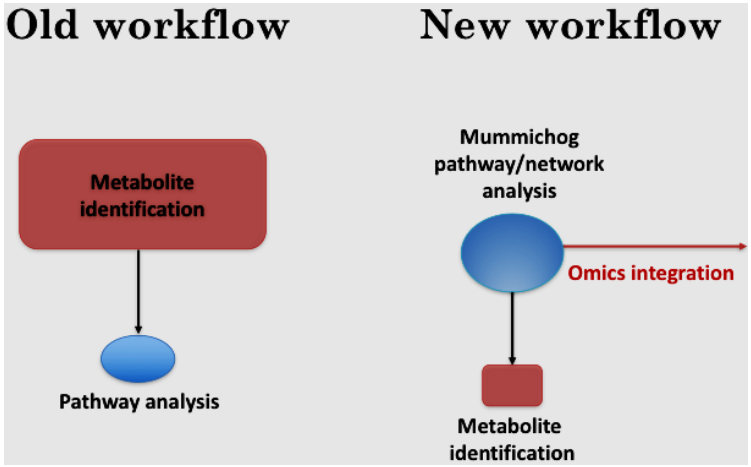
Deo et al. (2010) PLOS Computational Biology 6(2): e1000692.
Glucose challenge induces many transporter reactions.

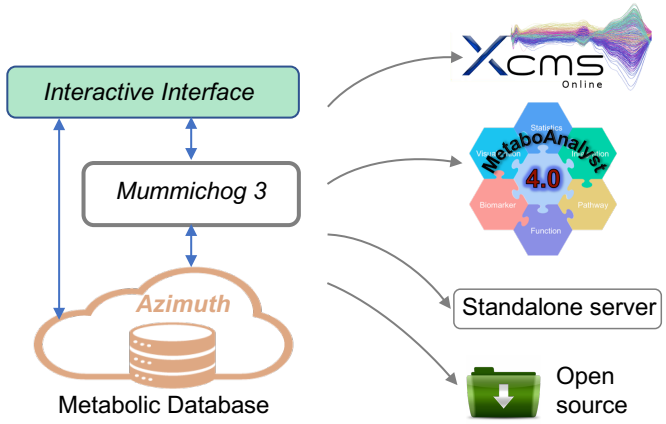
Summary

- Metabolic pathways are part of metabolic models
- Pathway analysis for targeted metabolomics data typically involve over-representation test and mapping
- *Mummichog* identifies pathway patterns from untargeted metabolomics
- Statistical significance ≠ Biological significance
- Need validation of the hypotheses generated from bioinformatics tools
- Pay attention to metabolites and confirmation levels

Mummichog & future development

- Active development of version 2 and server
- Separating metabolic models to a new database, updated independently from the software
- Send in your feature requests





- Coupling visual analytics
- Adding compatibility with upstream data processing
- Incorporating research in metabolic network reconstruction, especially using high-resolution mass spectrometry data
- Incorporating network alignment methods for data integration

Acknowledgement

Emory University
Dean P. Jones
Xin Hu
Douglas Walker (MSSM)
ViLinh Tran
Bill Liang
Fred Strobel
Miriam Vos
Tianwei Yu

CHDS
Barbara Cohn
Piera Cirrilo
Nickilou Krigbaum

Stanford University
Bali Pulendran

Scripps Research Institute
Gary Siuzdak
Paul Benton

McGill University
Jeff Xia

Thanks for funding from US National Institutes of Health U19AI090023, HHSN272201300018I, UH2AI132345, U01OD026489, U2COD026485, P30ES019776, U2CES026560, P50ES026071, R010DK107900, R01AI121252, R01GM124061, EPA (83615301) and California Breast Cancer Research Program (21UB-8002).