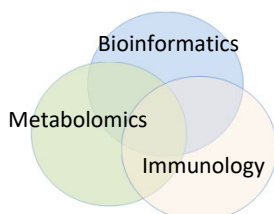




Mummichog, pathway and network analysis for Metabolomics



Shuzhao Li, Ph.D.
Assistant Professor, Dept. Medicine,
Emory University School of Medicine
E-mail: shuzhao.li@gmail.com
July 26, 2018



Where do my metabolites go?

What's common between *Fish and metabolites?*

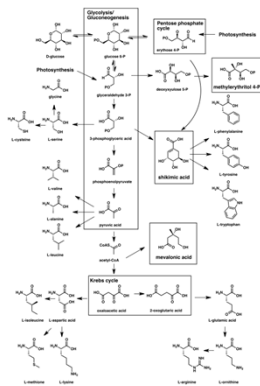
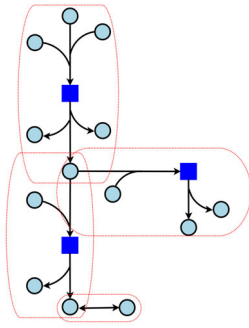


Photo by Joanna Penn.

Outline

- **Metabolomics pathway analysis and *mummichog***
- **Applications of *mummichog* to population studies and mechanistic investigations**
- **Integration of metabolomics with other data types**

Metabolic model



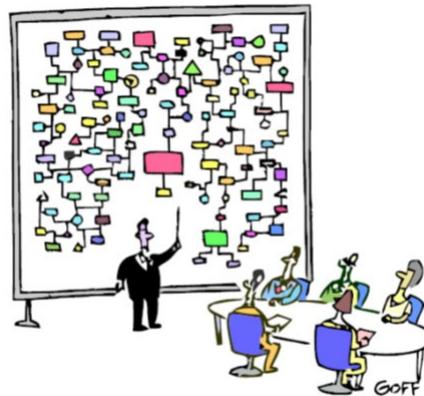
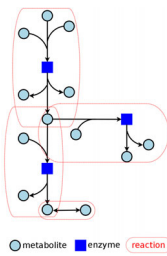
- A metabolic model consists of metabolites, enzymes, reactions, pathways
- Reactions can be described by differential equations (mathematical models)

$$\begin{array}{c}
 v_1 \quad v_2 \quad b_1 \quad b_2 \quad b_3 \quad b_4 \quad b_5 \\
 \begin{array}{l}
 A \begin{pmatrix} -1 & 0 & 1 & 1 & 0 & 0 & 0 \\
 B \begin{pmatrix} 1 & -1 & 0 & -1 & 0 & 0 & 0 \\
 C \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
 D \begin{pmatrix} -1 & 0 & 0 & 0 & 1 & 0 \\
 E \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1
 \end{array}
 \end{array}
 \end{array}
 \Rightarrow
 \begin{array}{c}
 \begin{pmatrix} A \\ B \\ C \\ D \\ E \end{pmatrix} = Sv = S \begin{pmatrix} v_1 \\ v_2 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{pmatrix} = 0 \\
 \text{Steady State Mass balance}
 \end{array}
 \end{array}$$

- We focus on statistical models
- Pathways and networks are mathematically indistinguishable

5

Metabolic model: genome-scale



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

MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

Please select a pathway library:

- Homo sapiens (human) [MFN]
- Homo sapiens (human) [BioCyc]
- Homo sapiens (human) [KEGG]
- Mus musculus (mouse) [BioCyc]
- Mus musculus (mouse) [KEGG]
- Rattus norvegicus (rat) [KEGG]
- Bos taurus (cow) [KEGG]
- Gallus gallus (chicken) [KEGG]
- Danio rerio (zebrafish) [KEGG]
- Danio rerio (zebrafish) [MTF]
- Drosophila melanogaster (fruit fly) [KEGG]
- Drosophila melanogaster (fruit fly) [BioCyc]
- Caenorhabditis elegans (nematode) [KEGG]
- Saccharomyces cerevisiae (yeast) [KEGG]
- Saccharomyces cerevisiae (yeast) [BioCyc]
- Oryza sativa japonica (Japanese rice) [KEGG]
- Arabidopsis thaliana (thale cress) [KEGG]
- Schistosoma mansoni [KEGG]

Metabolic models: where they come from

Pathway analysis for targeted and untargeted data

MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

Click a module to proceed, or scroll down for more details:

- Statistical Analysis
- Enrichment Analysis
- Pathway Analysis
- Joint Pathway Analysis
- Network Explorer
- MS Peaks to Pathways
- Other Utilities
- Spectral Analysis
- Biomarker Meta-analysis
- Power Analysis
- Time-series / Two-factor
- Biomarker Analysis

Legend:

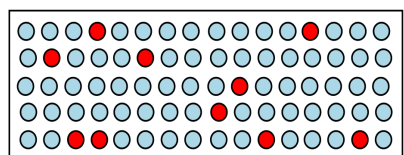
- Targeted or untargeted metabolomics
- Targeted or annotated metabolomics
- Untargeted metabolomics
- Multiple metabolomics data
- Integrating other omics

Targeted

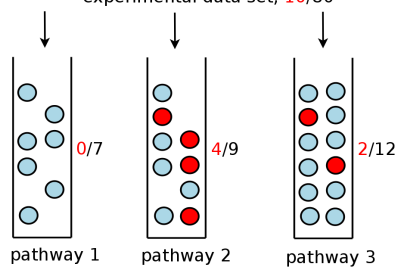
Untargeted (mummichog 1)

Pathway enrichment test

If metabolites are known; red are significant metabolites



experimental data set, 10/80



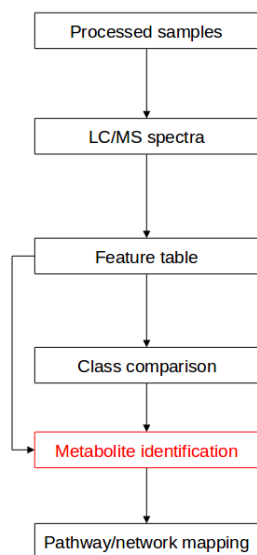
pathway 1

pathway 2

pathway 3

$$P = \binom{9}{4} \binom{71}{6} / \binom{80}{10} = 0.01$$

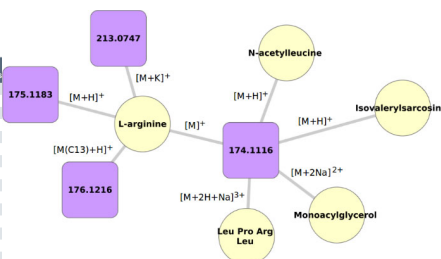
Metabolite identification is the bottleneck



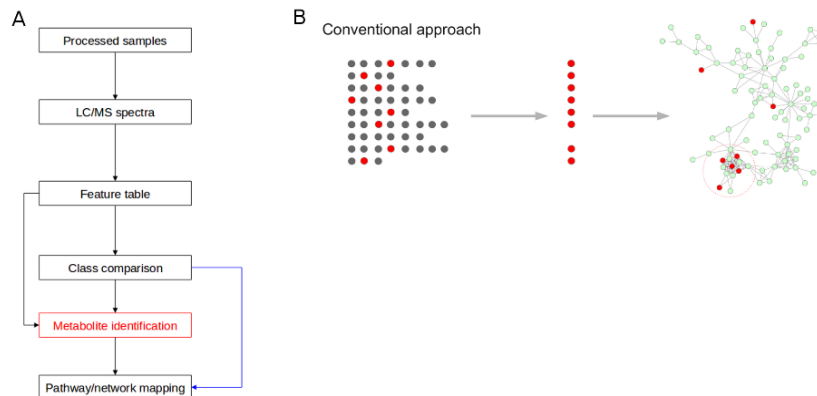
Uncertainty in matching metabolites - features

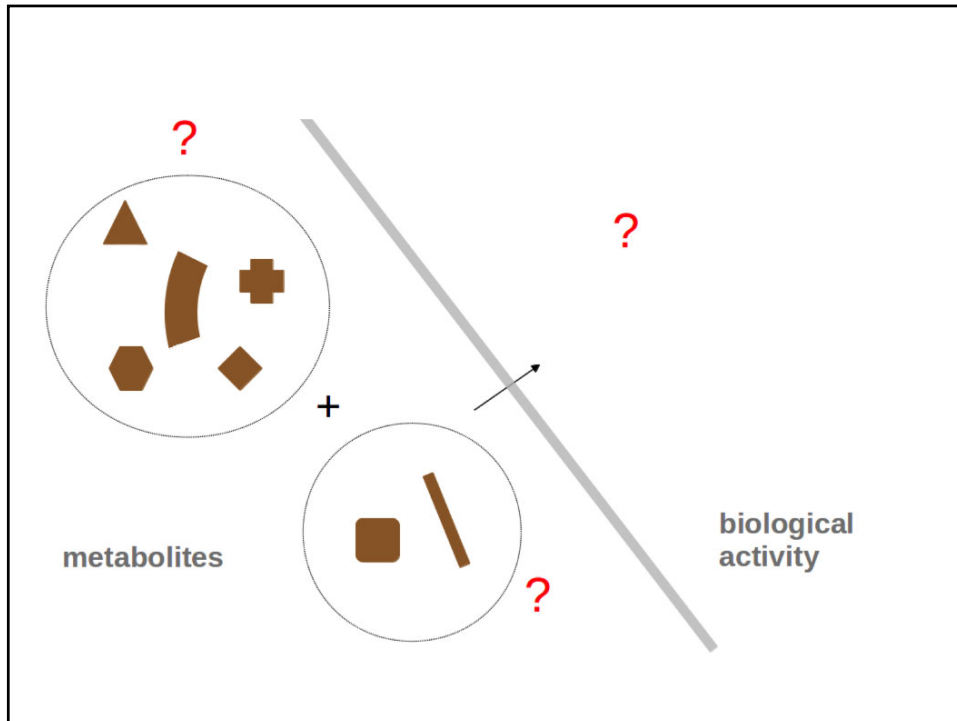
Search of m/z 190.1065 in HMDB with accurate matching

Common Name	Chemical Formula	Adduct MW (Da)	Matching [M+H] ⁺ [M+K] ⁺	MW Difference (Da)	QueryMass - AdductMass
cis-4-Octenedioic acid	C8H12O4	190.10791 [172.073563]		8.85E-4	
Gamma-Butyrolactone	C4H6O2	190.10791 [96.036781]		8.85E-4	
3-Methylcrotonylglycine	C7H11NO3	190.10791 [157.073896]		8.85E-4	
Lactaldehyde	C3H6O2	190.10791 [74.036781]		8.85E-4	
Hydroxyacetone	C3H6O2	190.10791 [74.036781]		8.85E-4	
Cocain-3-one	C8H6O2	190.10791 [95.036781]		8.85E-4	
2-Octenedioic acid	C8H12O4	190.10791 [172.073563]		8.85E-4	
Propionic acid	C3H6O2	190.10791 [74.036781]		8.85E-4	
Neopanic acid	C8H12O4	190.10791 [148.073563]		8.85E-4	
(R)-2,3-Dihydroxy-3-methylvaleric acid	C8H12O4	190.10791 [148.073563]		8.85E-4	
D-Lactaldehyde	C3H6O2	190.10791 [74.036781]		8.85E-4	
Tiglylglycine	C7H11NO3	190.10791 [157.073896]		8.85E-4	
Bul-2-enoic acid	C8H6O2	190.10791 [96.036781]		8.85E-4	
trans-3-Octenedioic acid	C8H12O4	190.10791 [172.073563]		8.85E-4	
3-Hydroxypropanal	C3H6O2	190.10791 [74.036781]		8.85E-4	
Diacetyl	C4H6O2	190.10791 [96.036781]		8.85E-4	2M+NH4 [1+]
Pyridolepic acid	C8H7NO3	190.107925 [129.042587]		0.001419	M+HisPropH [1+]
N-Acetyltyrosine	C8H7NO3	190.107925 [129.042587]		0.001419	M+HisPropH [1+]
1-Pyrroline-4-hydroxy-2-carboxylate	C8H7NO3	190.107925 [129.042587]		0.001419	M+HisPropH [1+]
Pyridone hydroxycarboxylic acid	C8H7NO3	190.107925 [129.042587]		0.001419	M+HisPropH [1+]
Pyridonecarboxylic acid	C8H7NO3	190.107925 [129.042587]		0.001419	M+HisPropH [1+]
Pentrosidine	C11H20N4O4	190.108041 [178.201339]		0.001341	M+2H [2+]
Kyotorphin	C15H23N5O4	190.108063 [177.17616]		0.001356	M+AcN+2H [2+]
18-Hydroxyortosol	C21H30O6	190.109406 [178.204254]		0.002899	M+2H [2+]
Ethepamil	C23H30N2	190.109665 [134.249906]		0.003199	M+2Na [2+]

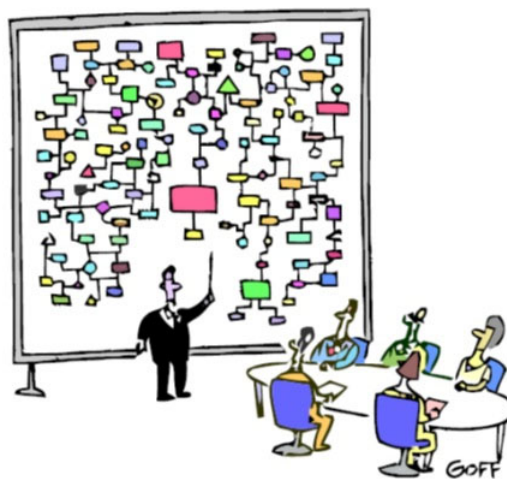


Conventional approach



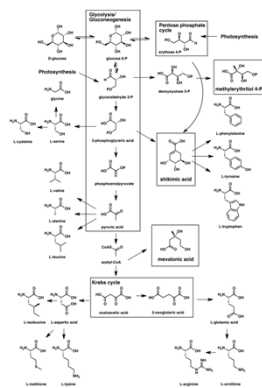


Known Metabolic reactions form a big network



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

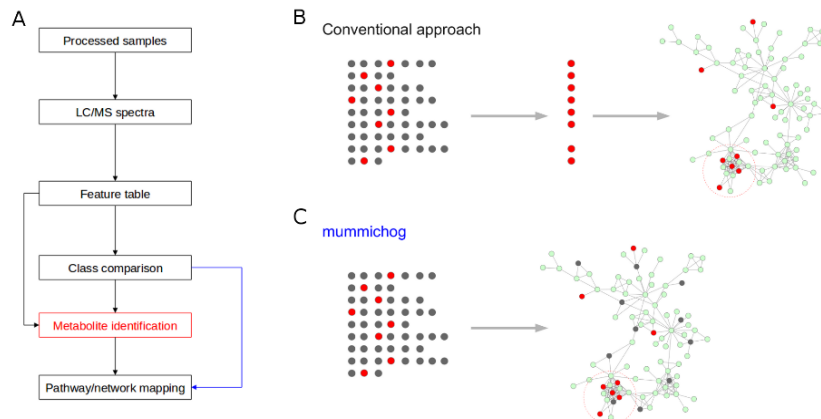
What's common between *Fish and metabolites?*



They swim in groups

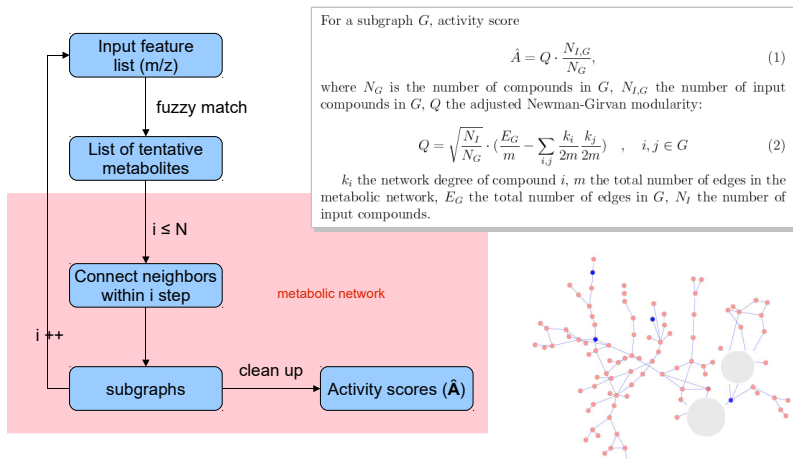
Photo by Joanna Penn. Slide inspired by Dr. Steve Barnes.

Mummichog tests metabolite grouping patterns



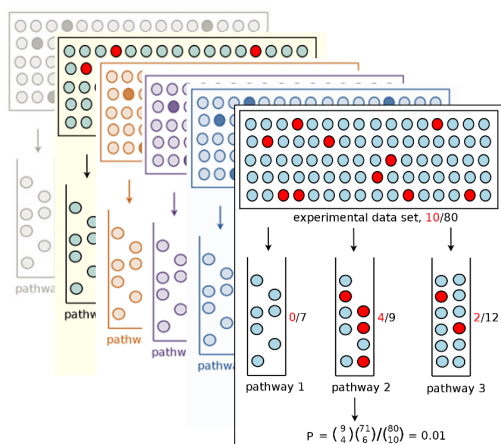
Li et al. 2013. PLoS Computational Biology. 9:e10031323

Module analysis in *mummichog*



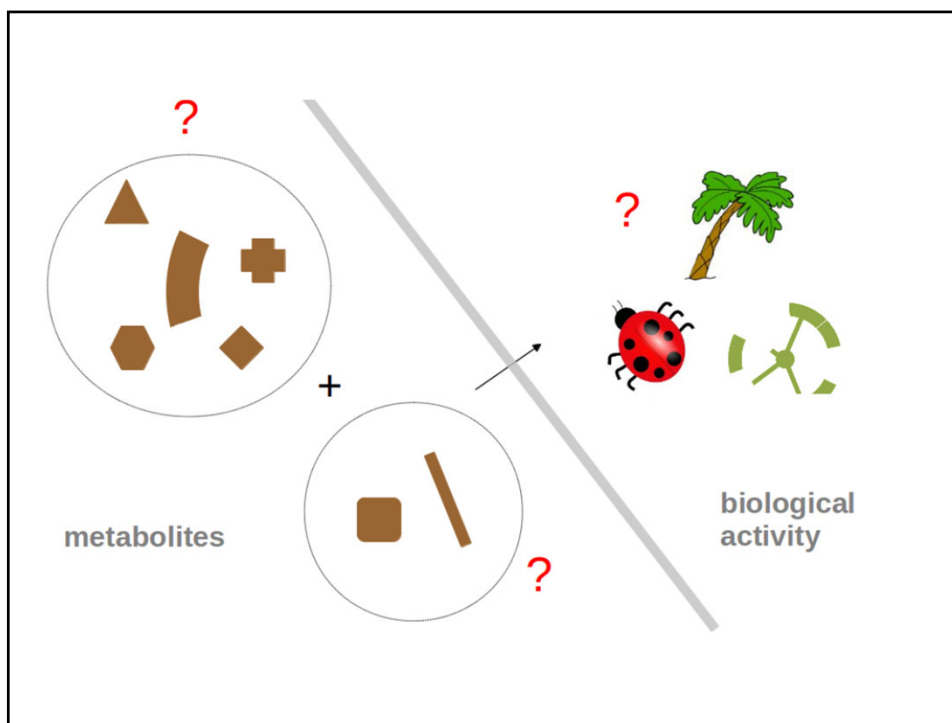
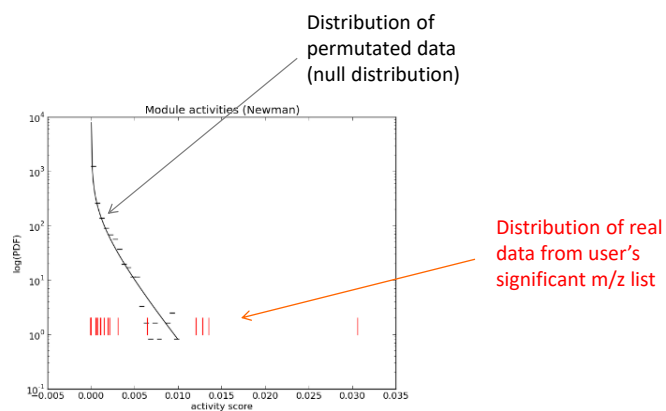
Li et al. 2013. PLoS Computational Biology. 9:e10031323

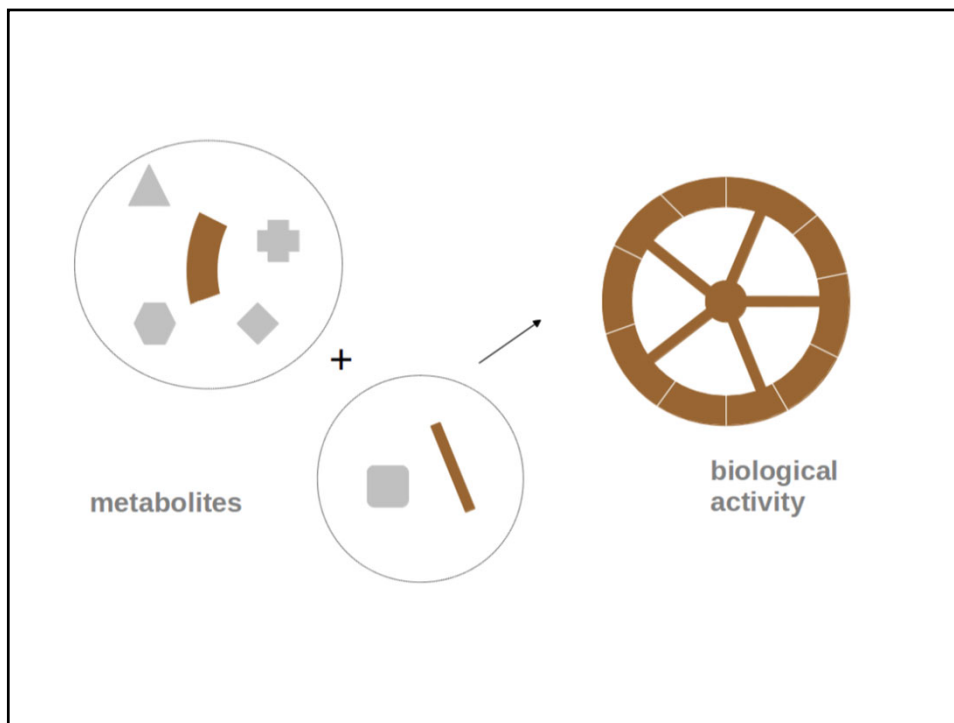
Pathway analysis in *mummichog*



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

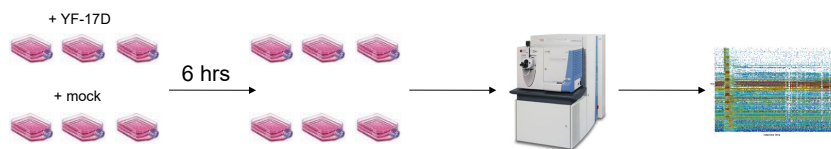
Testing module/pathway significance in *mummichog*



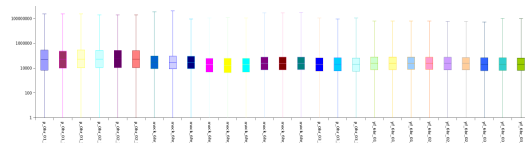


Case study: viral activation of immune cells

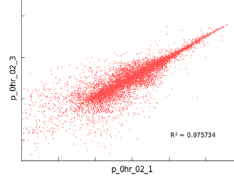
Monocyte derived dendritic cells (moDC)



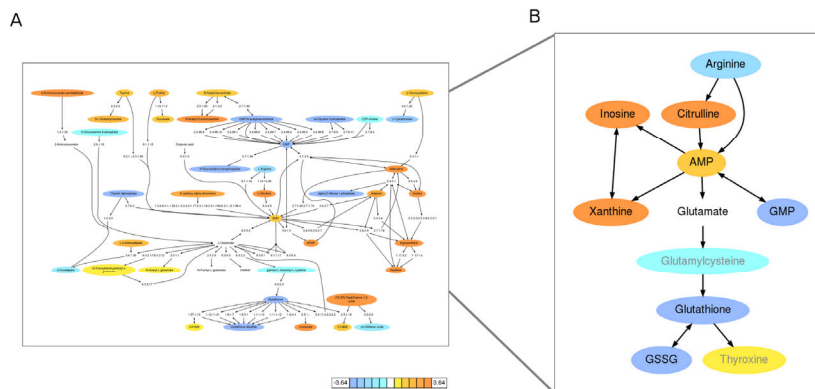
QA: total ion counts are similar among samples



technical replicates, 10,000 features

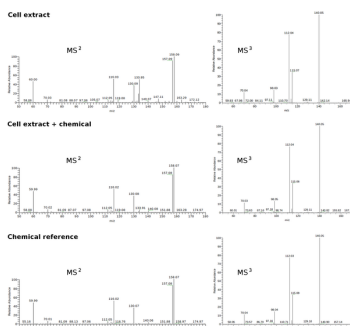


Mummichog: viral activation of immune cells

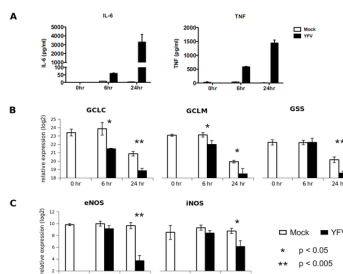


Li et al. 2013. PLoS Computational Biology. 9:e10031323

Experimental validation of *mummichog* prediction



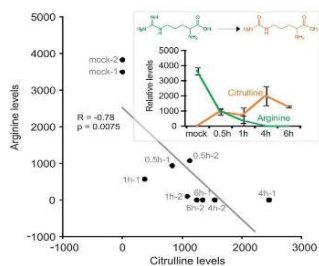
Tandem mass spectrometry confirmed 9/11 metabolites



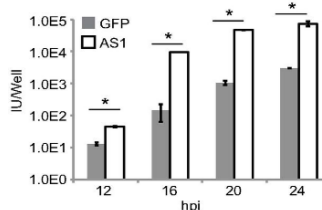
Gene expression supported GSH/GSSG depletion and Arg/Cit conversion

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

Arginine as master regulator of viral response

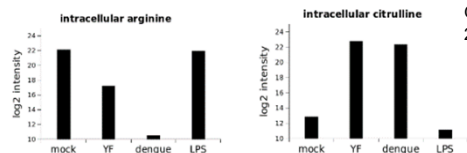


Ravindran et al. 2014. Science 343:313



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

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 increasingly 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

Version 2

- <http://mummichog.org>
- New data structures, using “Empirical compound” as intermediate concept
- Adducts and isotopes are computed based on chemical formula, and grouped by similar retention time
- Redesigned data structure, better user data tracking and web-oriented

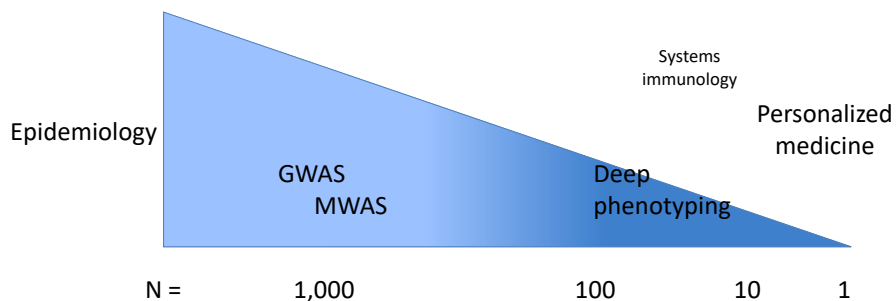
Future development

- Separation of mummichog algorithms from metabolic models
- The metabolic models are hosted in a new database, and update independently from the software
- 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
- Streamline applications to precision medicine

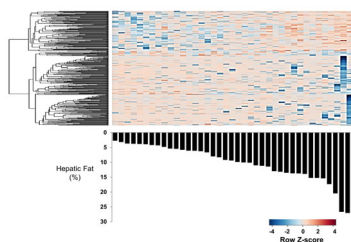
Outline

- **Metabolomics pathway analysis and *mummichog***
- **Applications of *mummichog* to population studies and mechanistic investigations**
- **Integration of metabolomics with other data types**

The “N” in systems medicine

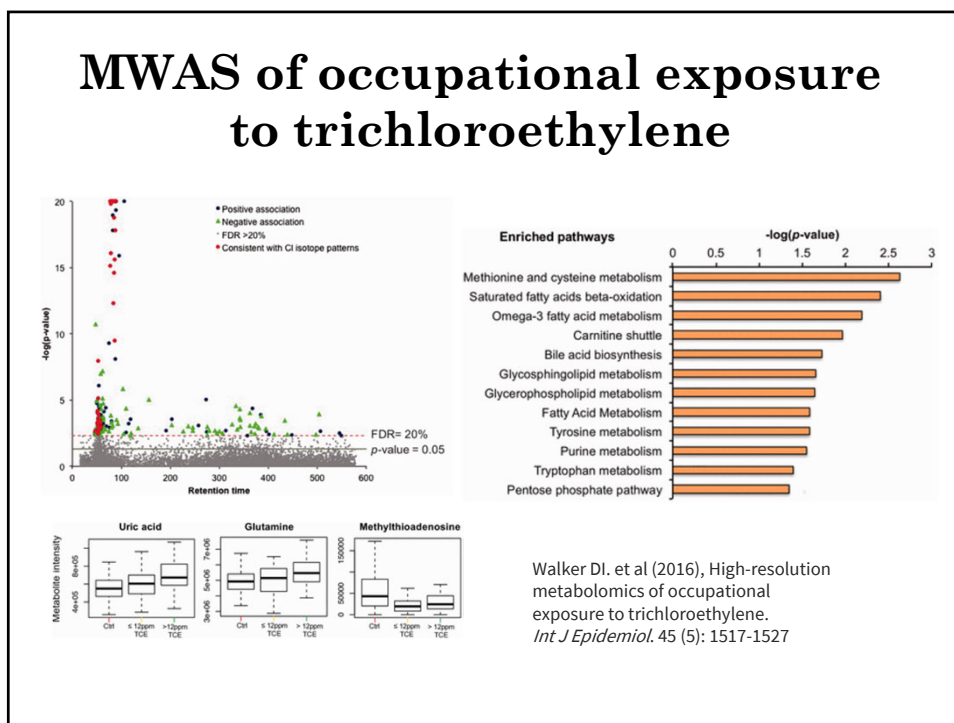
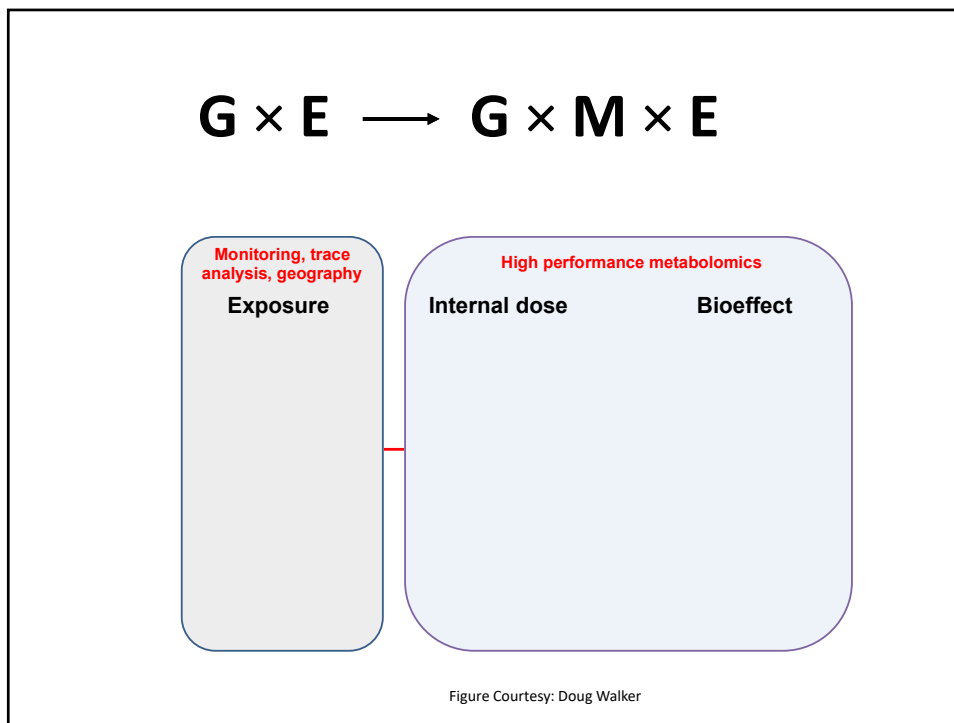


MWAS + *mummichog* (NAFLD)

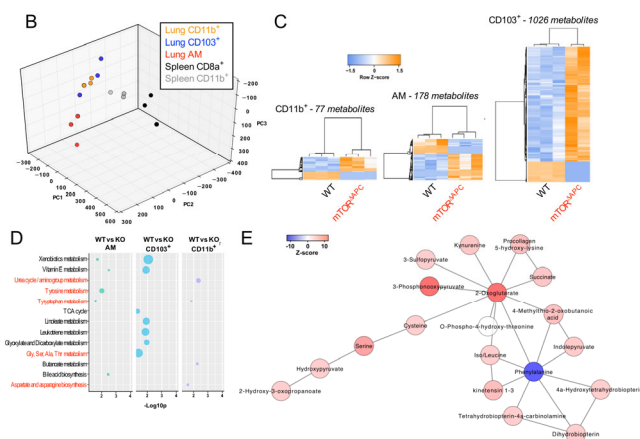


Jin, Banton, et al., 2016.
 Amino Acid Metabolism is Altered in Adolescents with Nonalcoholic Fatty Liver Disease - An Untargeted, High Resolution Metabolomics Study. *The Journal of pediatrics* 172: 14

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



mTOR regulates metabolic adaptation of APCs in the lung microenvironment



Sinclair et al (2017), *Science*. 357:1014.

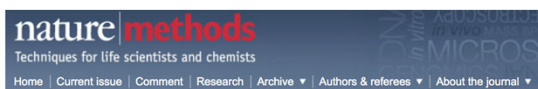
Outline

- Metabolomics pathway analysis and *mummichog*
- Applications of *mummichog* to population studies and mechanistic investigations
- Integration of metabolomics with other data types

Multi-omics integration

- Knowledge driven
 - Proteins and genes can be linked to metabolites via enzymatic reactions
 - Multiple data types can be overlaid to same pathways, given prior pathway definition
 - Prior knowledge can be coded into network statistics and topology
 - MWAS still in early days
- Data driven
 - Statistical association via CCA, PLS, etc
 - Evidence propagation in various forms
 - Machine learning and artificial intelligence

XCMS Online overlaps multiomics data to the same pathways



home > current issue > correspondence > abstract

ARTICLE PREVIEW
view full access options >

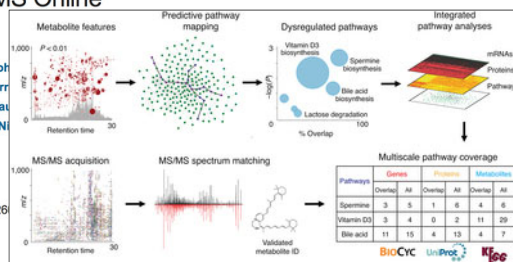
NATURE METHODS | CORRESPONDENCE

Systems biology guided by XCMS Online metabolomics

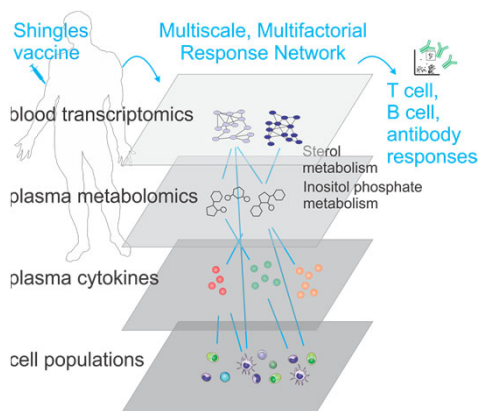
Tao Huan, Erica M Forsberg, Duane Rinehart, Caroline H Johnson, Mingliang Fang, Aries Aisporna, Brian Hilmer, Farr Michael W W Adams, Gregory Krantz, Matthew W Fields, Pat Niedernhofer, Trey Ideker, Erica L Majumder, Judy D Wall, Ni Goodacre, Luke L Lairson & Gary Szuzdak

Affiliations | Corresponding author

Nature Methods 14, 461–462 (2017) | doi:10.1038/nmeth.426
Published online 27 April 2017

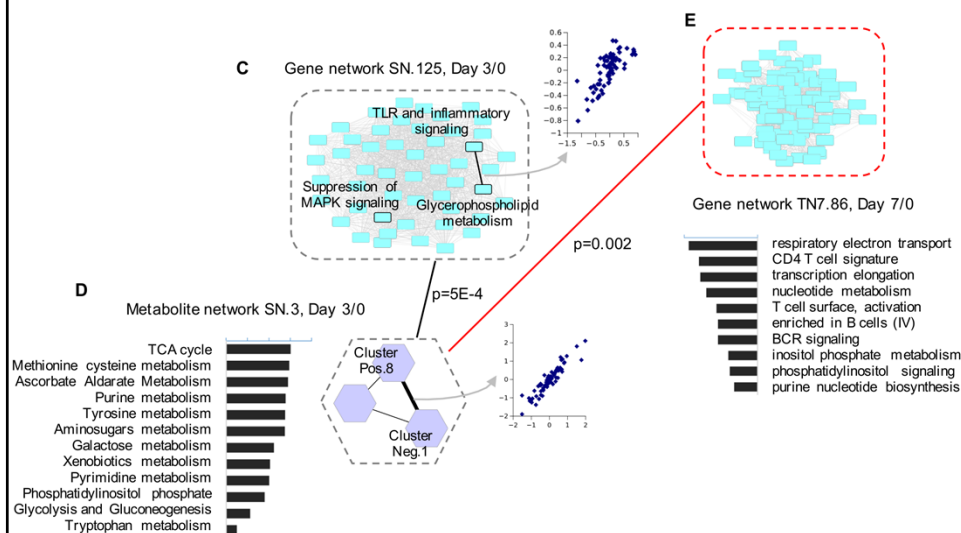


MMRN integrating multiple data types

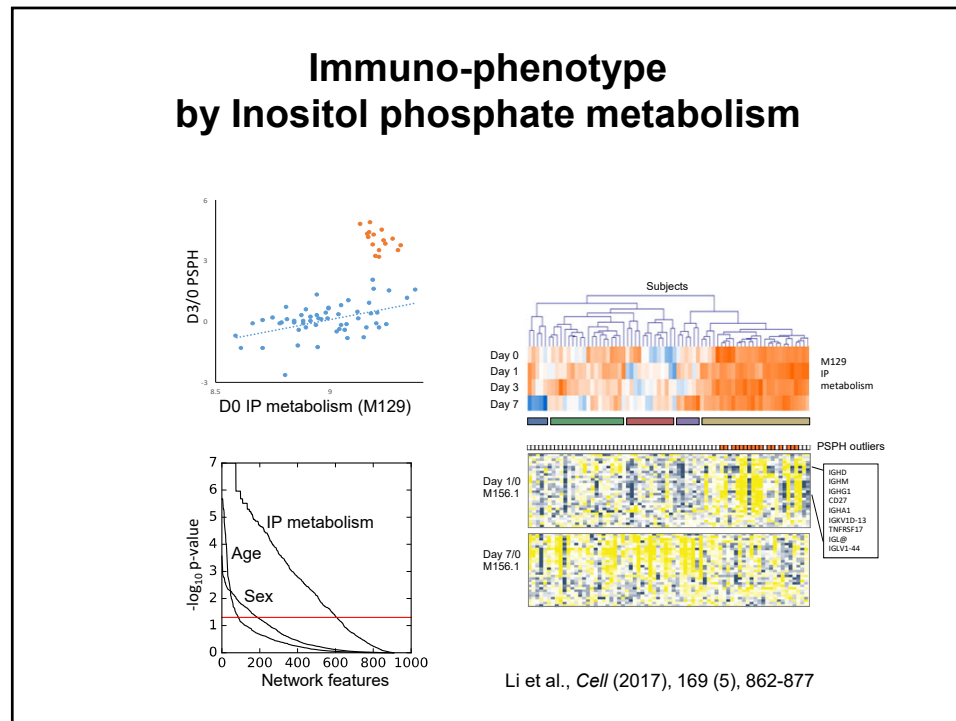


Li et al., Metabolic Phenotypes of Response to Vaccination in Humans, *Cell* (2017), 169 (5), 862-877

MMRN example

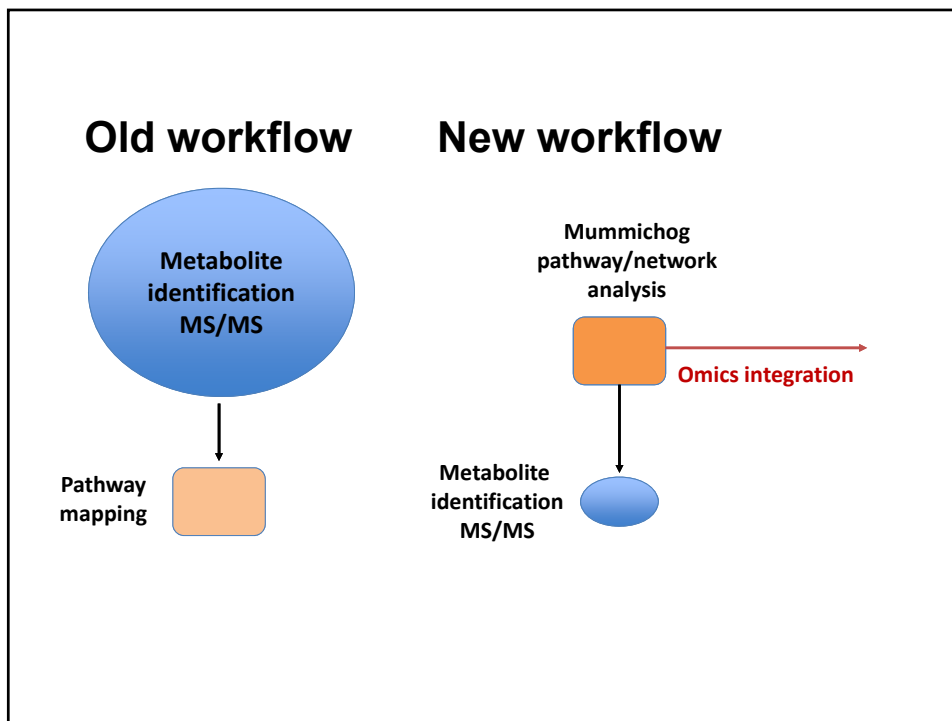


Li et al., *Cell* (2017), 169 (5), 862-877



Summary

- ❑ Advancing of mass spectrometry enables deep sequencing of metabolome and exposome; filling gap for G x E
- ❑ *Mummichog* rewrites the workflow of high-throughput metabolomics, bridging genome-scale metabolic models and untargeted metabolomics.
<http://mummichog.org>.
- ❑ MWAS + *mummichog* is a powerful approach to understand health and disease
- ❑ Combining multiple omics is critical to small “N”, human studies. Their integration can be driven by data mining or by knowledge models.



Thank you!

Emory University

Dept. Medicine

Dean P. Jones
Young-Mi Go
Douglas Walker
ViLinh Tran
Bill Liang
Karan Uppal
Ken Liu
Sophia Banton
Andrei Todor
Luiz Gardinassi

Mark Mulligan
Nadine Rouphael
Aneesh Mehta
Jennifer Whitaker

Emory Vaccine Center

Bali Pulendran
Helder Nakaya
Mohan Maddur
Sathyanarayana
Sai Duraisingham

Rafi Ahmed
Nicole Sullivan
Andrea Wieland
Megan McCausland
Chris Chiu

Dept. Pediatrics

Miriam B. Vos
Ran Jin

School of Public Health

Tianwei Yu

University of Colorado

Adriana Weinberg
Myron Levin
Jennifer Canniff

Acknowledgement of research supports from US National Institutes of Health (HHSN272201200031C, U19AI090023, HHSN272201300018I, UH2AI132345, NIEHS P30ES019776, U2CES026560, P50ES026071, R010DK107900, R01AI121252, R21HD089056, R01GM124061), DoD DARPA (W911NF-16-C-0008), EPA (83615301) and California Breast Cancer Research Program (21UB-8002).



Untargeted metabolomics data

