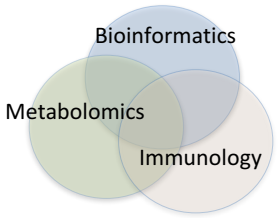
Pathway and network analysis for Metabolomics



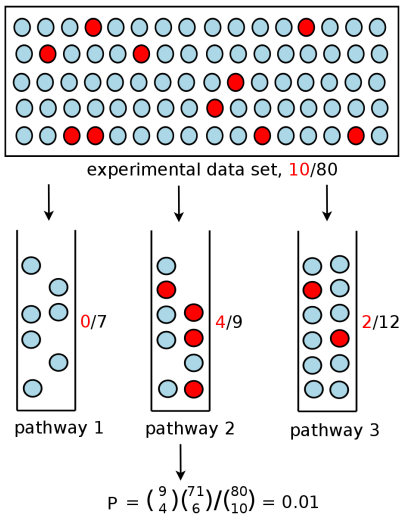
Shuzhao Li, Ph.D.
Assistant Professor, Dept. Medicine,
Emory University School of Medicine
E-mail: shuzhao.li@gmail.com
July 20, 2017

Outline

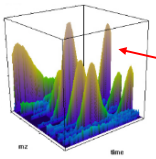
- **Metabolomics pathway analysis and *mummichog***
- **Applications of *mummichog* to population studies and mechanistic investigations**
- **Integration of metabolomics with other data types**

Pathway enrichment test

If metabolites are known; red are significant metabolites



Untargeted metabolomics data



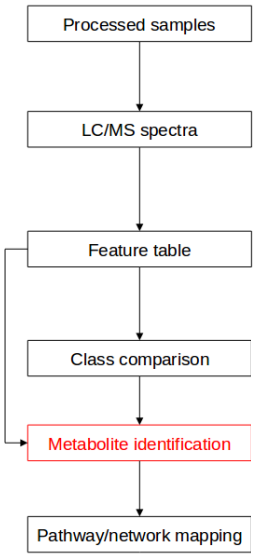
mass/charge (m/z)

retention time

intensity

feature

m/z	Retention time	Sample_1	Sample_2	Sample_3
702.86649413	70.4811138781	56905.1184620213	78201.4457063353	24695.6510487344
715.4970731384	56.55139775	83341.1725499997	1085.6171033898	0
715.6080685464	57.1626518527	93484.5395499998	0	0
722.7052085610	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

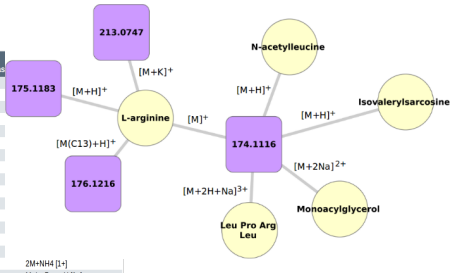


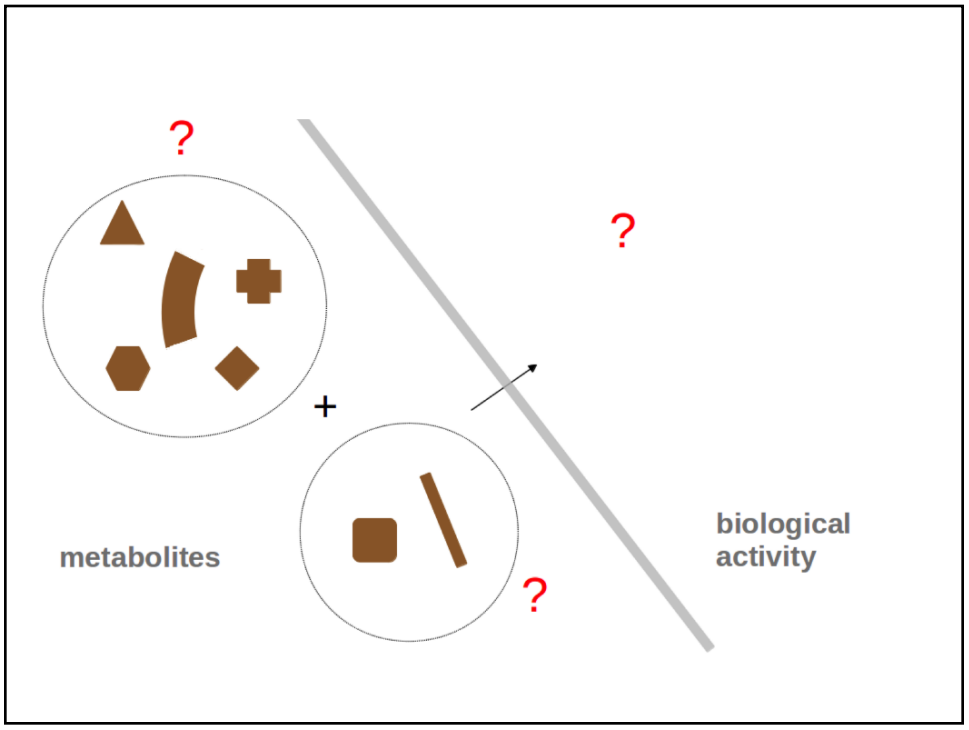
Metabolite identification is the bottleneck

Uncertainty in matching metabolites - features

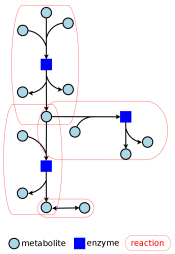
Search of m/z 190.1065 in HMDB with accurate matching

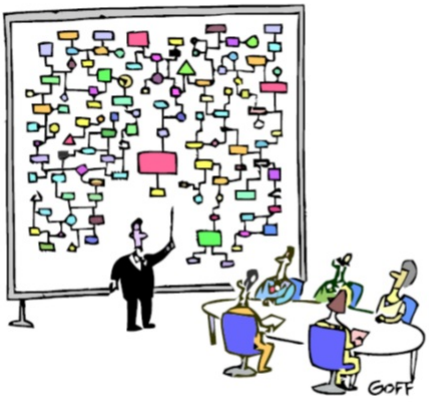
Common Name	Chemical Formula	Adduct MW (Da) [Matching HMDB MW]	MW Difference (Da) [QueryMass - AdductMas	
cis-4-Octenedioic acid	C8H12O4	190.107391 [172.073563]	8.85E-4	
Gamma-Butyrolactone	C4H6O2	190.107391 [86.036781]	8.85E-4	
3-Methylcrotonylglycine	C7H11NO3	190.107391 [157.073898]	8.85E-4	
Lactaldehyde	C3H6O2	190.107391 [74.036781]	8.85E-4	
Hydroxyacetone	C3H6O2	190.107391 [74.036781]	8.85E-4	
Oxolan-3-one	C4H6O2	190.107391 [86.036781]	8.85E-4	
2-Octenedioic acid	C8H12O4	190.107391 [172.073563]	8.85E-4	
Propionic acid	C3H6O2	190.107391 [74.036781]	8.85E-4	
Mevalonic acid	C6H12O4	190.107391 [148.073563]	8.85E-4	
(R) 2,3-Dihydroxy-3-methylvalerate	C6H12O4	190.107391 [148.073563]	8.85E-4	
D-Lactaldehyde	C3H6O2	190.107391 [74.036781]	8.85E-4	
Triglylglycine	C7H11NO3	190.107391 [157.073898]	8.85E-4	
But-2-enoic acid	C4H6O2	190.107391 [86.036781]	8.85E-4	
trans-3-Octenedioic acid	C8H12O4	190.107391 [172.073563]	8.85E-4	
3-Hydroxypropanal	C3H6O2	190.107391 [74.036781]	8.85E-4	
Diacetyl	C4H6O2	190.107391 [86.036781]	8.85E-4	2M+NH4 [1+]
Pyroglutamic acid	C5H7NO3	190.107925 [129.042587]	0.001419	M+IsoProp+H [1+]
N-Acryloylglycine	C5H7NO3	190.107925 [129.042587]	0.001419	M+IsoProp+H [1+]
1-Pyrroline-4-hydroxy-2-carboxylate	C5H7NO3	190.107925 [129.042587]	0.001419	M+IsoProp+H [1+]
Pyroline hydroxycarboxylic acid	C5H7NO3	190.107925 [129.042587]	0.001419	M+IsoProp+H [1+]
Pyrrolidonecarboxylic acid	C5H7NO3	190.107925 [129.042587]	0.001419	M+IsoProp+H [1+]
Pentosidine	C17H26N6O4	190.108047 [378.201538]	0.001541	M+2H [2+]
Kyotorphin	C15H23N5O4	190.108063 [337.175018]	0.001556	M+ACN+2H [2+]
18-Hydroxycortisol	C21H30O6	190.109406 [378.204254]	0.002899	M+2H [2+]
Emopamil	C23H30N2	190.109665 [334.240906]	0.003159	M+2Na [2+]





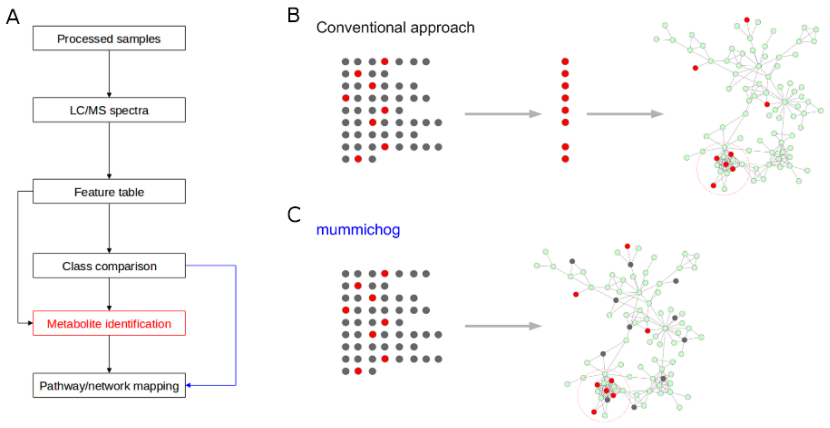
Genome-scale Metabolic model





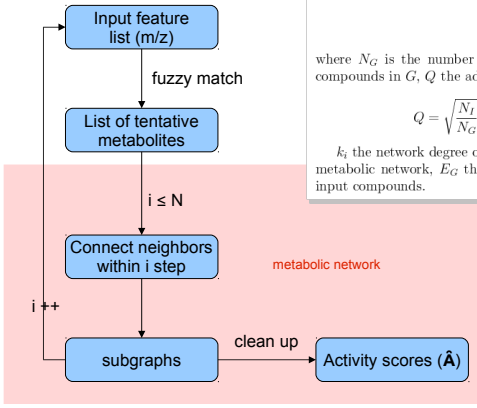
"And that's why we need a computer."

Mummichog bridging metabolic models



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

Module analysis in *mummichog*



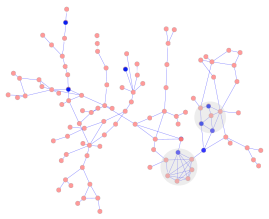
For a subgraph G, activity score

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

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:

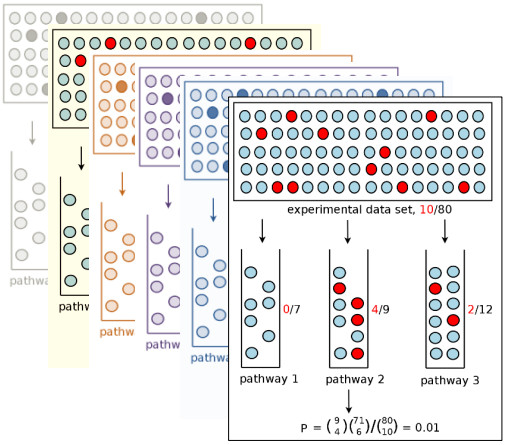
$$Q = \sqrt{\frac{N_I}{N_G}} \cdot \left(\frac{E_G}{m} - \sum_{i,j} \frac{k_i}{2m}\frac{k_j}{2m}\right) \quad , \quad i,j \in G \quad (2)$$

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.



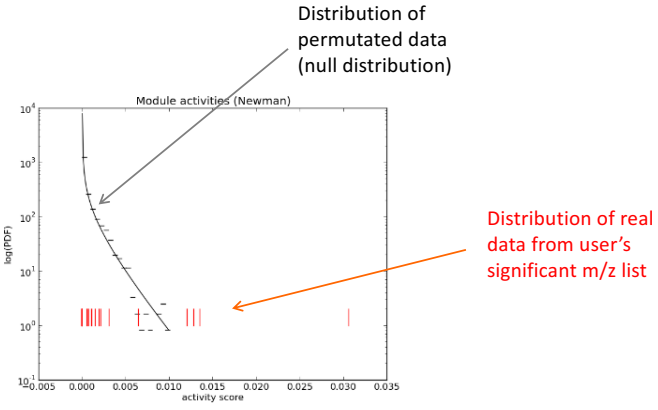
Li et al. 2013. PLoS Computational Biology. 9:e10031323

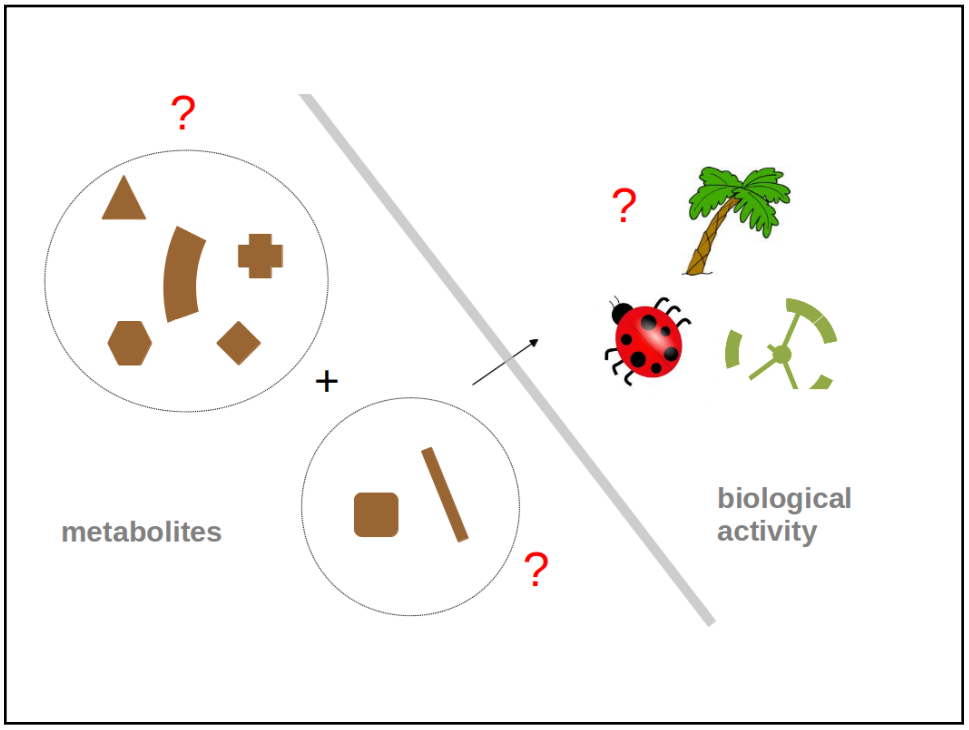
Pathway analysis in *mummichog*

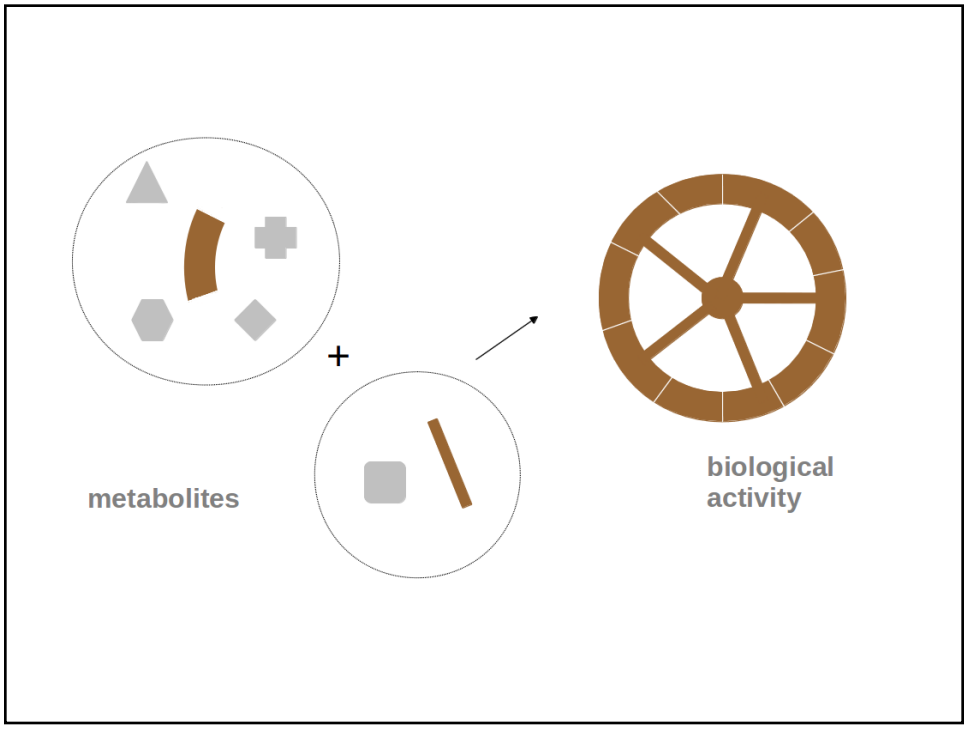


Li et al. 2013. PLoS Computational Biology. 9:e10031323

Testing module/pathway significance in *mummichog*

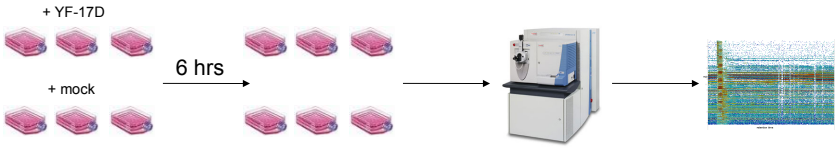


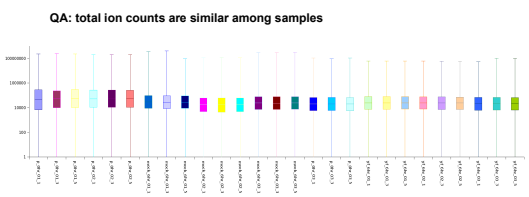

?
+
?
?
metabolites
biological
activity

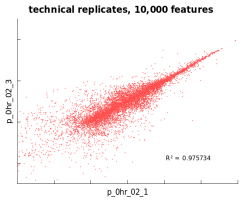

+
metabolites
biological
activity

Case study: viral activation of immune cells

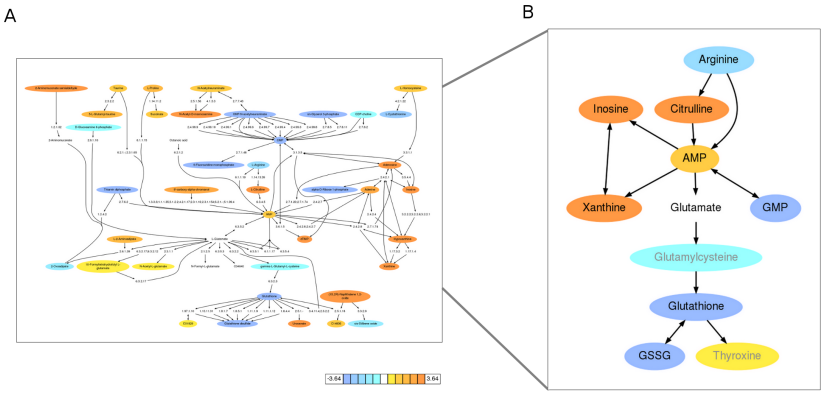
Monocyte derived dendritic cells (moDC)





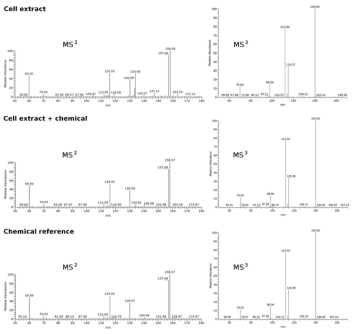


Mummichog: viral activation of immune cells

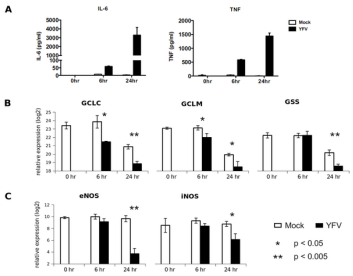


Li et al. 2013. PLoS Computational Biology. 9:e10031323

Experimental validation of *mummichog* prediction



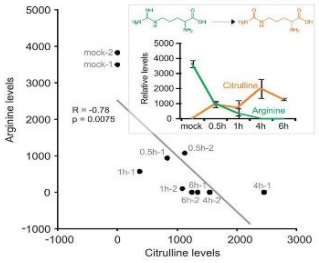
Tandem mass spectrometry confirmed 9/11 metabolites



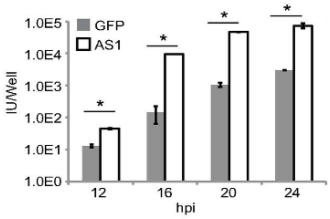
Gene expression supported GSH/GSSG depletion and Arg/Cit conversion

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

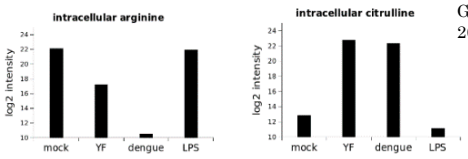
Arginine as master regulator of viral response



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 demo

```
sli49@Somits's MacBook Pro:~/play/mummichog-1.0.7/test$ python ../mummichog/main
.py -f testdata.txt -o demo

    ------------------------------------------------

            oO                              ooooooooo
          oOO   OOOOO  ooooo          ooo oooo
     oOO   O          ooooo  oooooo ooooo
    oooO               oooooo             oooo ooooo
        Oooo   o       OOOOOO    oooo    ooooooooo
             ooooo  oooo
                  o

    ------------------------------------------------

mummichog version 1.0.7

Pygraphviz is not found. Skipping...
Started @ Mon Jul 11 17:37:04 2016

Loading metabolic network MFN_1.10.2...
cpds with MW: 2016
Using 394 features (p < 0.001000) as significant list.
Got 394 significant features from 3878 references

Pathway Analysis...
query_set_size = 184 compounds
total_feature_num = 973 compounds
Resampling, 100 permutations to estimate background ...
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 3
0 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56
57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83
 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
Pathway background is estimated on 11900 random pathway values

Modular Analysis, using 100 permutations ...
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 3
```

mummichog.org

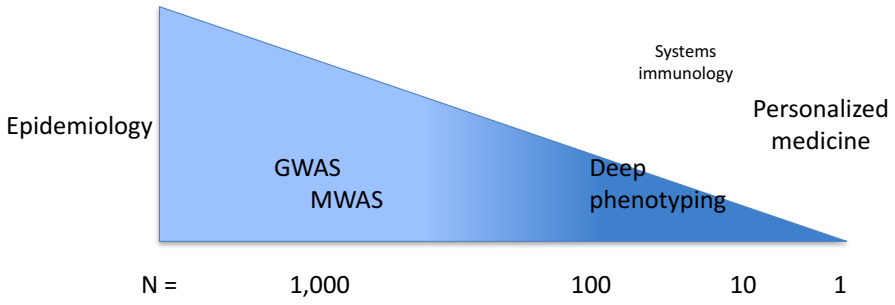
Mummichog version 2

- ❑ 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
- ❑ Better tracking of user input data
- ❑ Future: Modular web utilities;
  Updating metabolic models

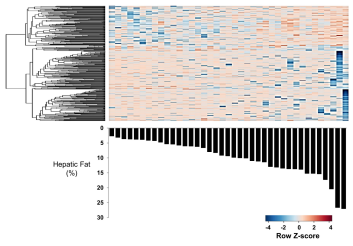
Outline

- 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)



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 × E ⟶ G × M × E

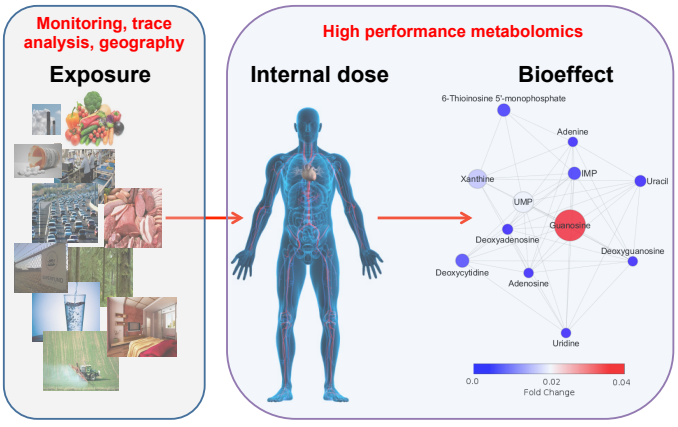
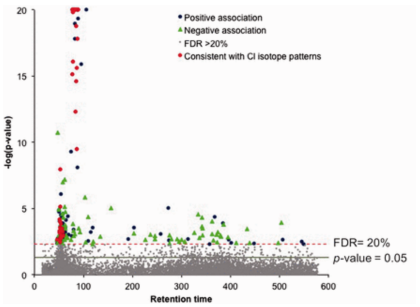
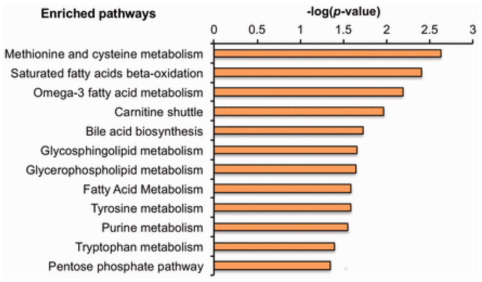
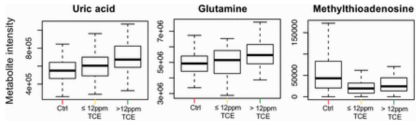


Figure Courtesy: Doug Walker

MWAS of occupational exposure to trichloroethylene

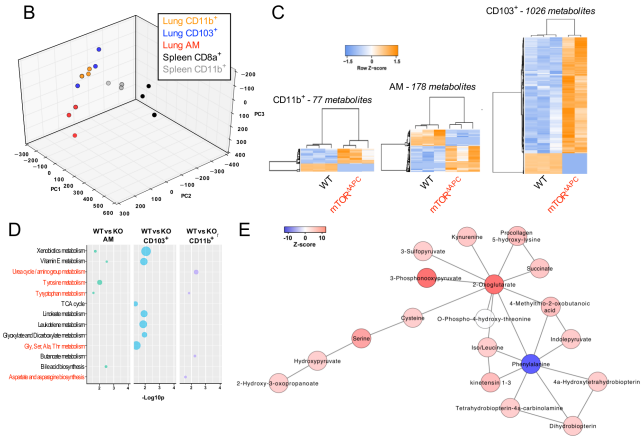






Walker DI. et al (2016), High-resolution metabolomics of occupational exposure to trichloroethylene. *Int J Epidemiol*. 45 (5): 1517-1527

mTOR regulates metabolic adaptation of APCs in the lung microenvironment



Sinclair et al (2017), *Science*. In press.

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

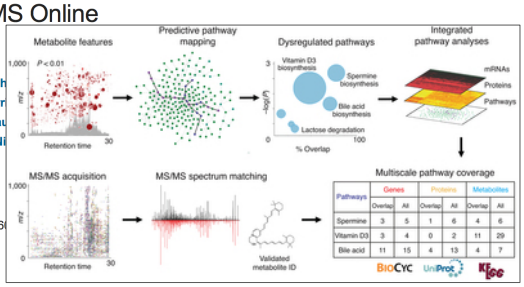
XCMS Online overlaps multiomics data to the same pathways

nature methods
Techniques for life scientists and chemists
Home | Current issue | Comment | Research | Archive | Authors & referees | About the journal
home ▸ current issue ▸ correspondence ▸ abstract
ARTICLE PREVIEW
view full access options ▸
NATURE METHODS | CORRESPONDENCE

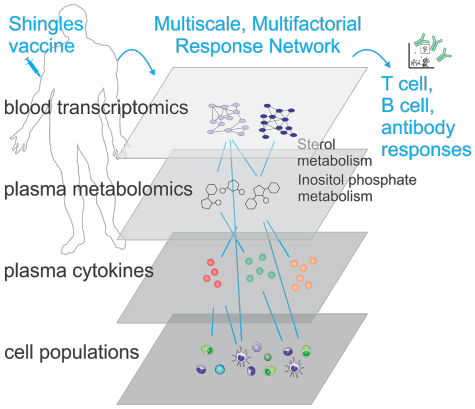
Systems biology guided by XCMS Online metabolomics

Tao Huan, Erica M Forsberg, Duane Rinehart, Caroline H Joh
Benton, Mingliang Fang, Aries Aisporna, Brian Hilmers, Farr
Michael W W Adams, Gregory Krantz, Matthew W Fields, Pau
Niedernhofer, Trey Ideker, Erica L Majumder, Judy D Wall, Ni
Goodacre, Luke L Lairson & Gary Siuzdak
Affiliations | Corresponding author

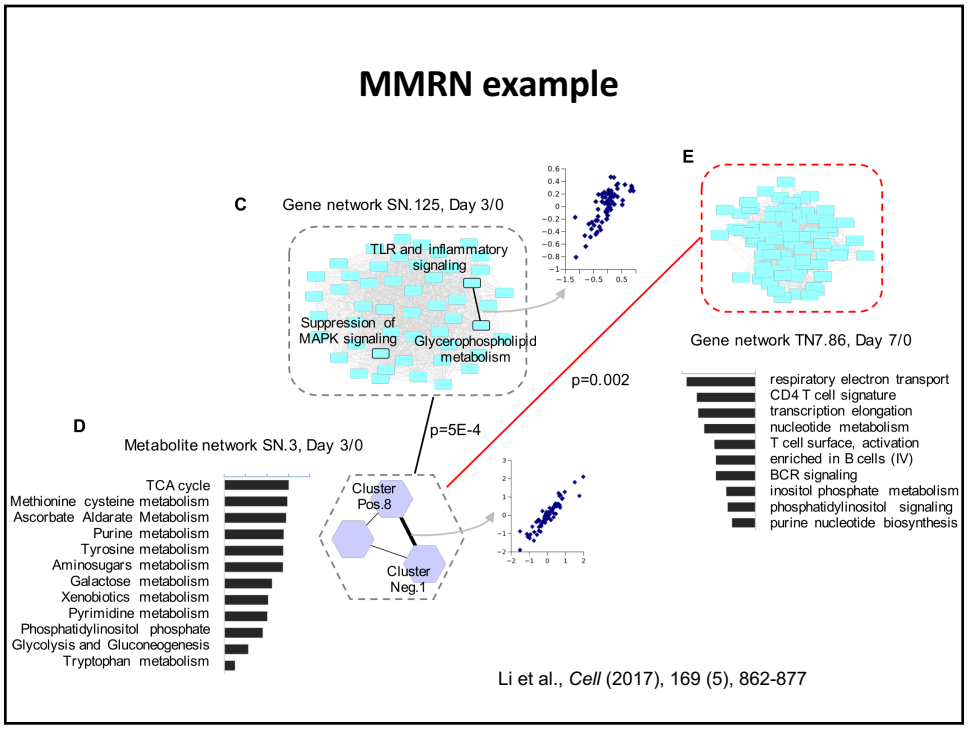
Nature Methods **14**, 461–462 (2017) | doi:10.1038/nmeth.426
Published online 27 April 2017

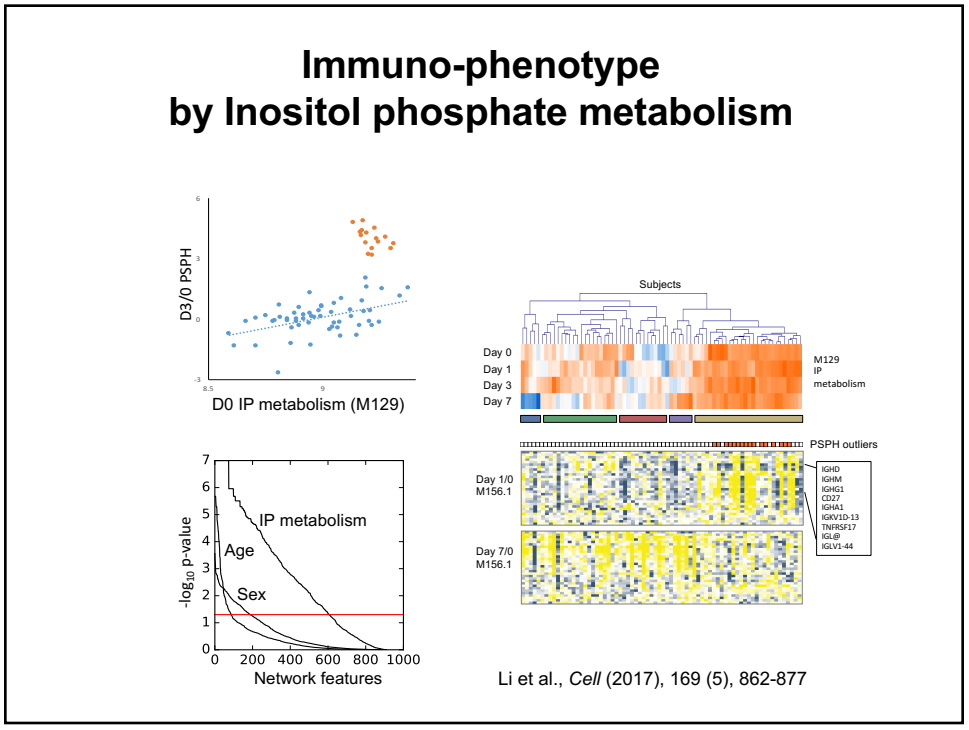


MMRN integrating multiple data types



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

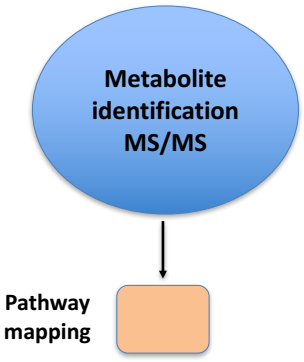

MMRN example
C
Gene network SN.125, Day 3/0
TLR and inflammatory signaling
Suppression of MAPK signaling
Glycerophospholipid metabolism
E
Gene network TN7.86, Day 7/0
p=0.002
p=5E-4
D
Metabolite network SN.3, Day 3/0
TCA cycle
Methionine cysteine metabolism
Ascorbate Aldarate Metabolism
Purine metabolism
Tyrosine metabolism
Aminosugars metabolism
Galactose metabolism
Xenobiotics metabolism
Pyrimidine metabolism
Phosphatidylinositol phosphate
Glycolysis and Gluconeogenesis
Tryptophan metabolism
Cluster Pos.8
Cluster Neg.1
respiratory electron transport
CD4 T cell signature
transcription elongation
nucleotide metabolism
T cell surface, activation
enriched in B cells (IV)
BCR signaling
inositol phosphate metabolism
phosphatidylinositol signaling
purine nucleotide biosynthesis
Li et al., Cell (2017), 169 (5), 862-877


Immuno-phenotype
by Inositol phosphate metabolism
D3/0 PSPH
D0 IP metabolism (M129)
Subjects
Day 0
Day 1
Day 3
Day 7
M129
IP
metabolism
PSPH outliers
Day 1/0
M156.1
Day 7/0
M156.1
IGHD
IGHM
IGHG1
CD27
IGHA1
IGKV1D-13
TNFRSF17
IGL@
IGLV1-44
-log10 p-value
IP metabolism
Age
Sex
Network features
Li et al., Cell (2017), 169 (5), 862-877

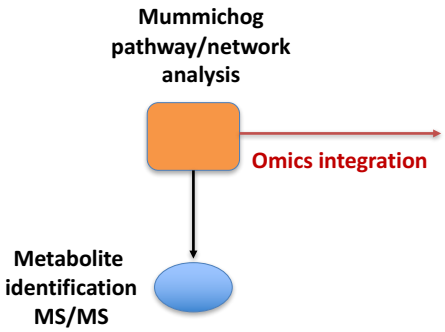
Summary

- 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.

Old workflow



New workflow



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 (NIAID HHSN272201200031C, U19AI090023, HHSN272201300018I, 1UH2AI132345, NIEHS P30 ES019116, U2CES026560, P50ES026071, NIDDK R010DK107900, NIAID R01AI121252), DoD DARPA (W911NF-16-C-0008), EPA (83615301) and California Breast Cancer Research Program (21UB-8002).