Mummichog: pathway and network analysis for metabolomics

Shuzhao Li, Ph.D
Assistant Professor, Department of Medicine,

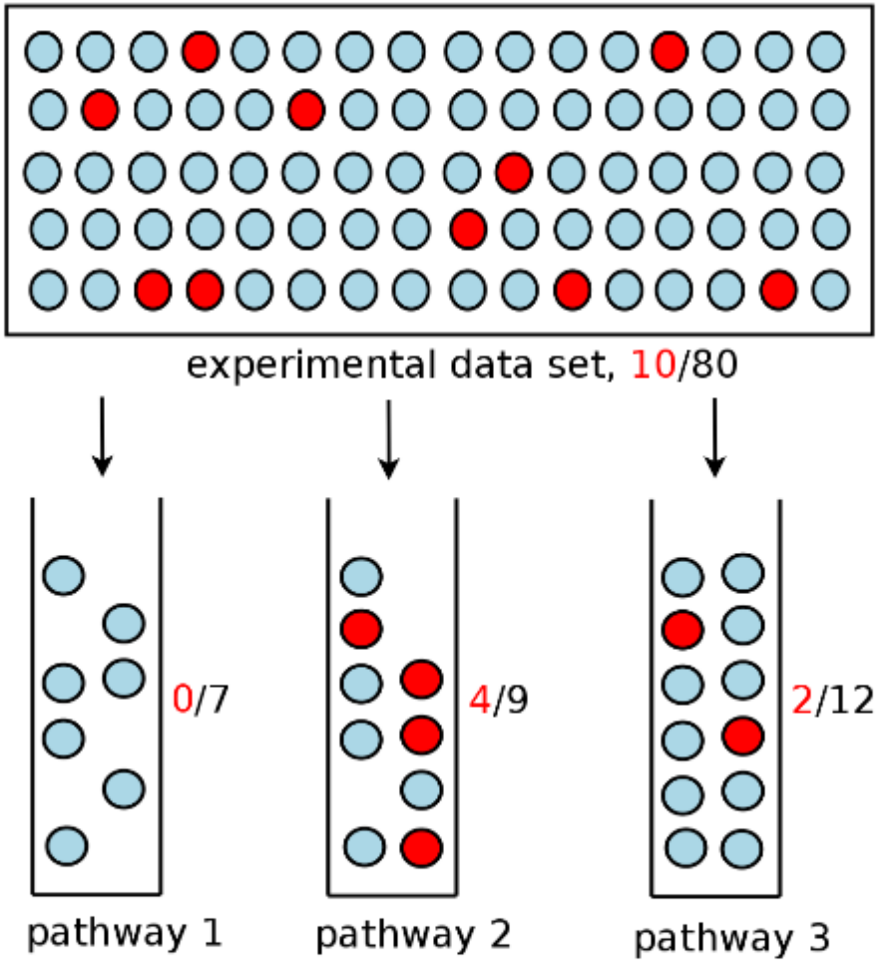
Division of Pulmonary, Allergy, Critical Care and Sleep Medicine
Emory University

June 18, 2015

EMORY UNIVERSITY SCHOOL OF MEDICINE

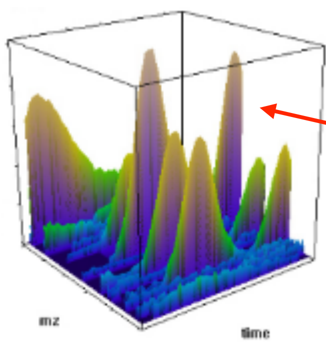
Pathway enrichment test

If metabolites are known; red are significant metabolites



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

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
722.705208561	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

Uncertainty in matching metabolites - features

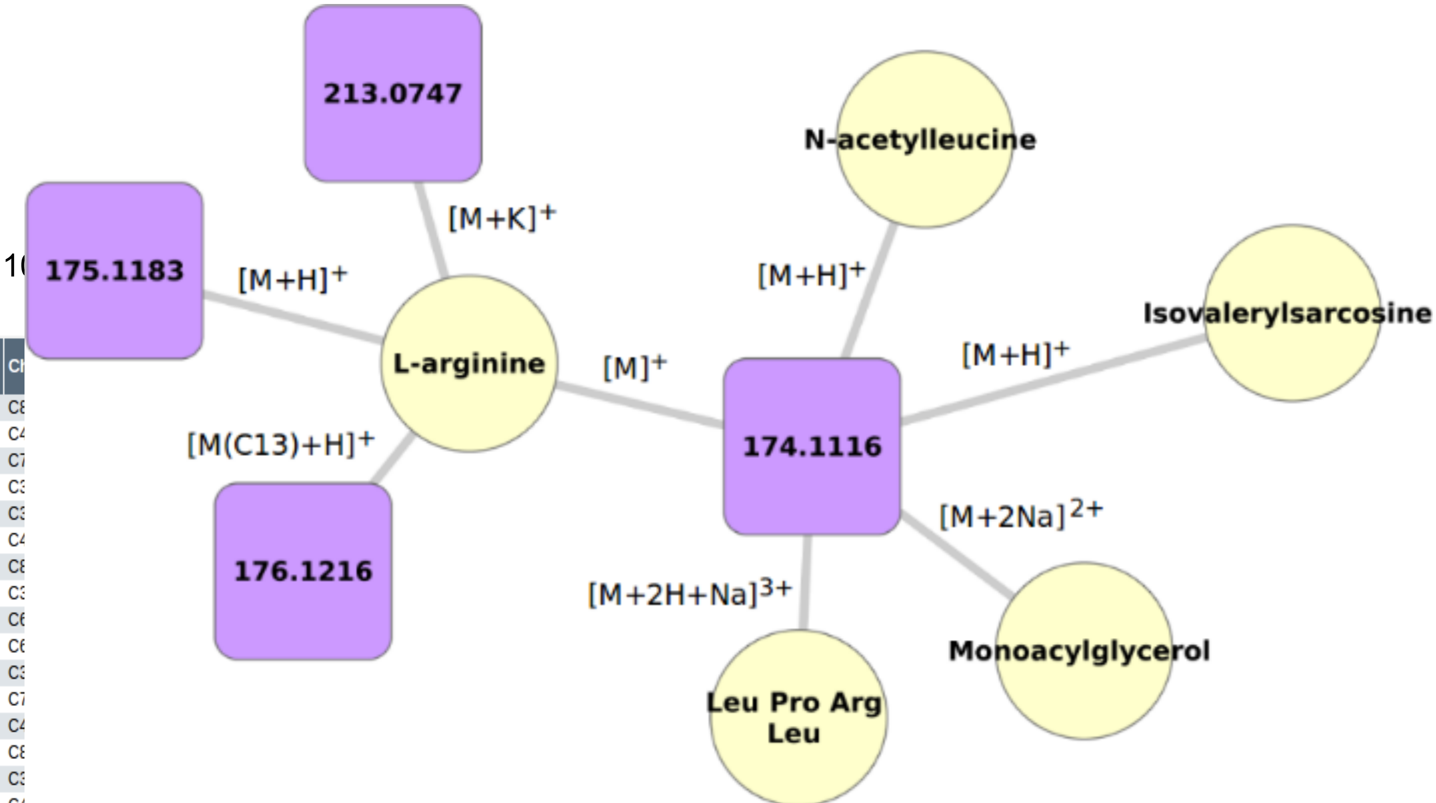
Search of m/z 190.1065 in HMDB with accurate matching

Common Name	Chemical Formula	Adduct MW (Da) [Matching HMDB MW]	MW Difference (Da) [QueryMass - AdductMass]	Adduct
cis-4-Octenedioic acid	C8H12O4	190.107391 [172.073563]	8.85E-4	M+NH4 [1+]
Gamma-Butyrolactone	C4H6O2	190.107391 [86.036781]	8.85E-4	2M+NH4 [1+]
3-Methylcrotonylglycine	C7H11NO3	190.107391 [157.073898]	8.85E-4	M+CH3OH+H [1+]
Lactaldehyde	C3H6O2	190.107391 [74.036781]	8.85E-4	2M+ACN+H [1+]
Hydroxyacetone	C3H6O2	190.107391 [74.036781]	8.85E-4	2M+ACN+H [1+]
Oxolan-3-one	C4H6O2	190.107391 [86.036781]	8.85E-4	2M+NH4 [1+]
2-Octenedioic acid	C8H12O4	190.107391 [172.073563]	8.85E-4	M+NH4 [1+]
Propionic acid	C3H6O2	190.107391 [74.036781]	8.85E-4	2M+ACN+H [1+]
Mevalonic acid	C6H12O4	190.107391 [148.073563]	8.85E-4	M+ACN+H [1+]
(R) 2,3-Dihydroxy-3-methylvalerate	C6H12O4	190.107391 [148.073563]	8.85E-4	M+ACN+H [1+]
D-Lactaldehyde	C3H6O2	190.107391 [74.036781]	8.85E-4	2M+ACN+H [1+]
Tiglylglycine	C7H11NO3	190.107391 [157.073898]	8.85E-4	M+CH3OH+H [1+]
But-2-enoic acid	C4H6O2	190.107391 [86.036781]	8.85E-4	2M+NH4 [1+]
trans-3-Octenedioic acid	C8H12O4	190.107391 [172.073563]	8.85E-4	M+NH4 [1+]
3-Hydroxypropanal	C3H6O2	190.107391 [74.036781]	8.85E-4	2M+ACN+H [1+]
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+]
Pyrroline 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+]

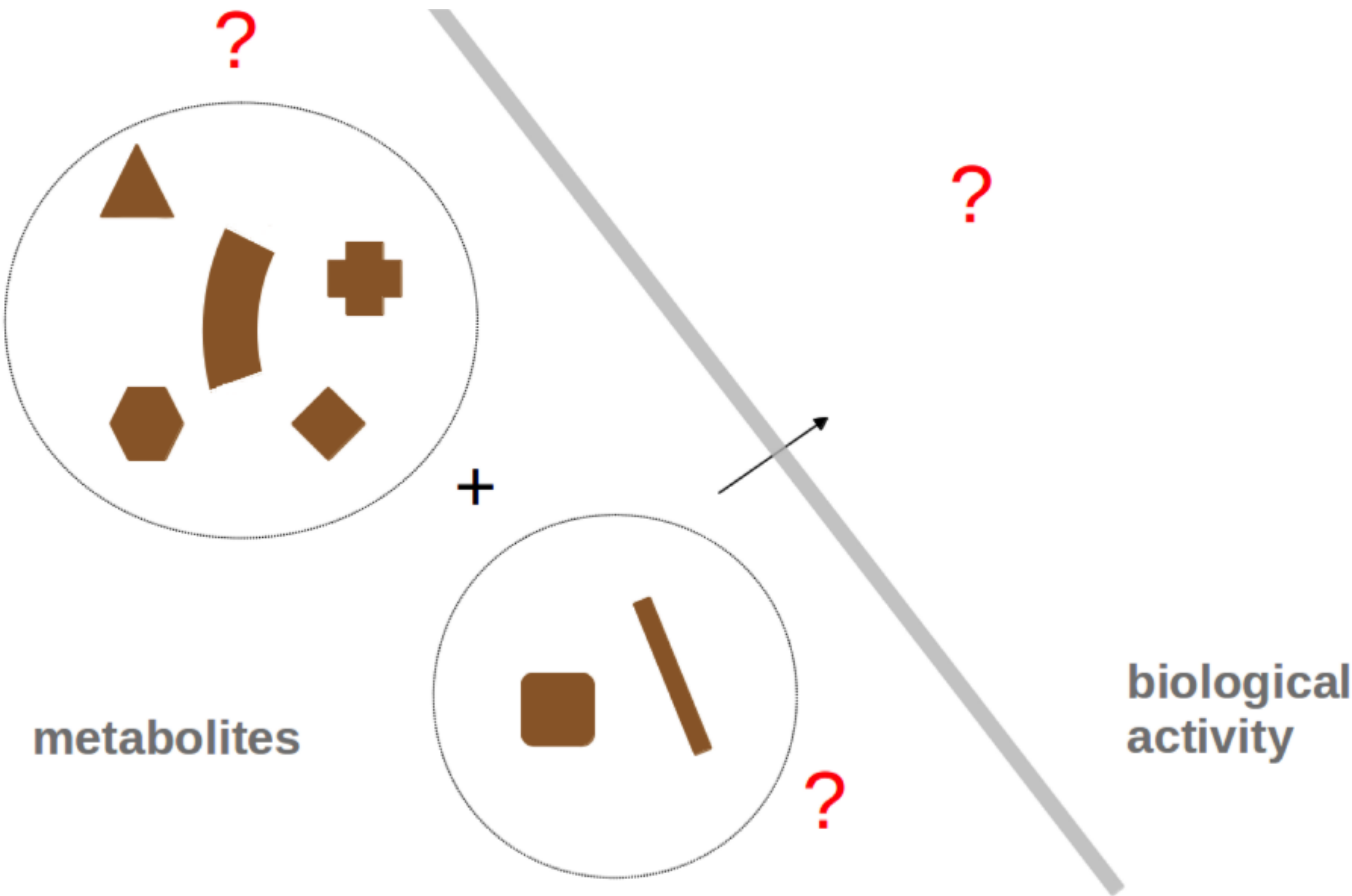
Uncertainty in matching metabolites - features

Search of m/z 190.1

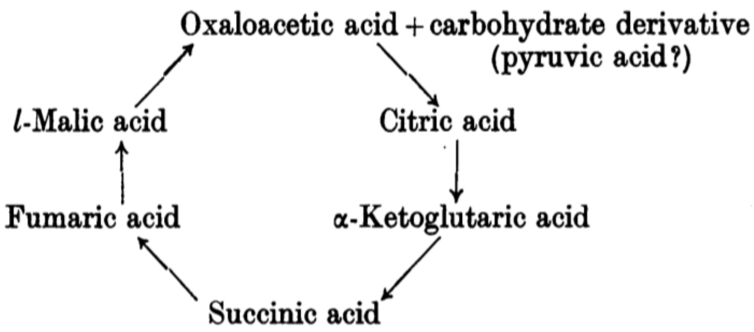
Common Name	Ch			
cis-4-Octenedioic acid	C8			
Gamma-Butyrolactone	C4			
3-Methylcrotonylglycine	C7			
Lactaldehyde	C3			
Hydroxyacetone	C3			
Oxolan-3-one	C4			
2-Octenedioic acid	C8			
Propionic acid	C3			
Mevalonic acid	C6			
(R) 2,3-Dihydroxy-3-methylvalerate	C6			
D-Lactaldehyde	C3			
Tiglylglycine	C7			
But-2-enoic acid	C4			
trans-3-Octenedioic acid	C8			
3-Hydroxypropanal	C3			
Diacetyl	[illegible]	[illegible]	[illegible]	[illegible]
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+]
Pyrroline 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+]



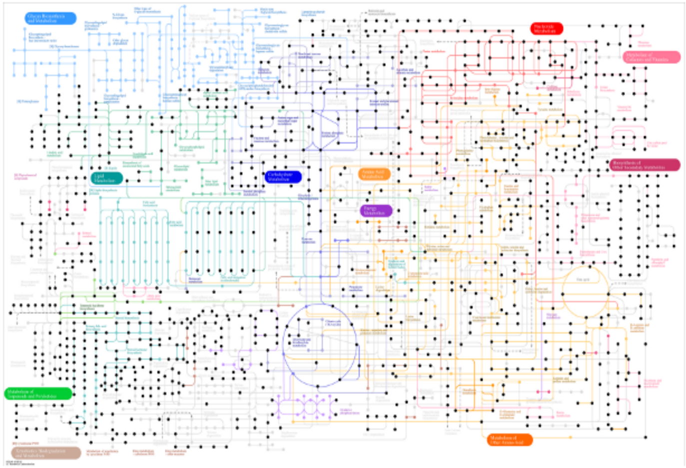
The puzzle of metabolomics



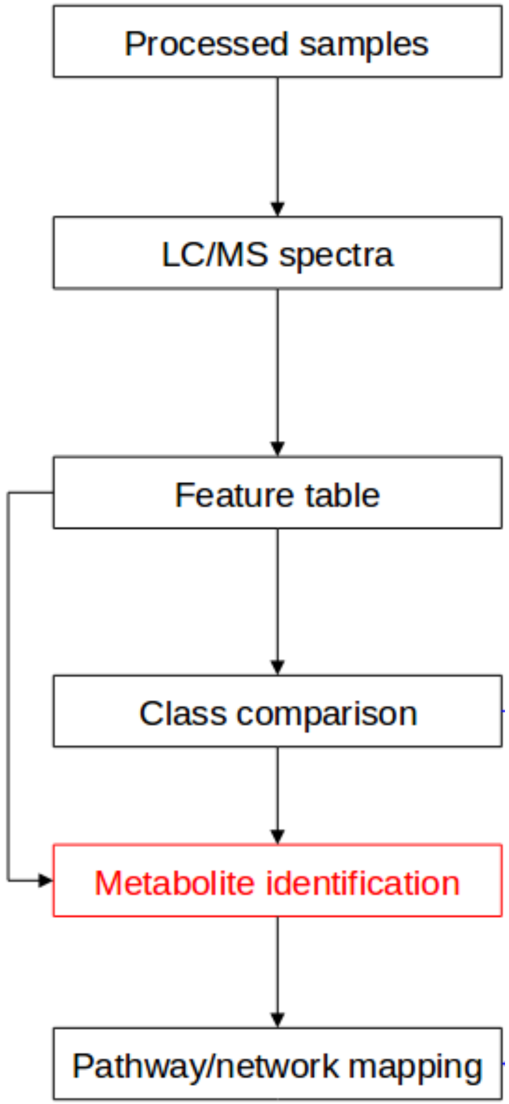
Genome-scale metabolic models as prior knowledge



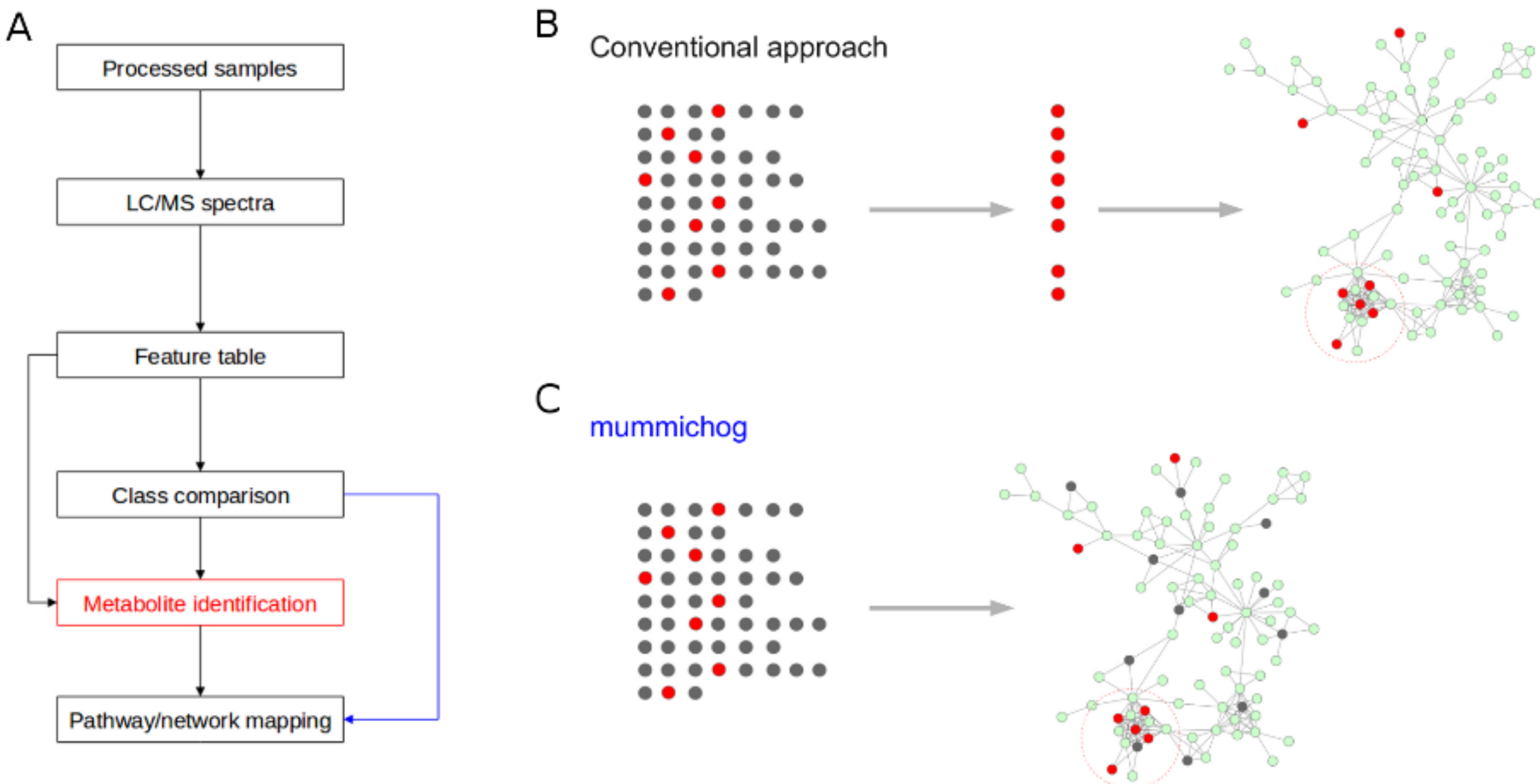
Krebs et al. 1938. Biochem Journal. 32:113



Bottleneck of untargeted metabolomics

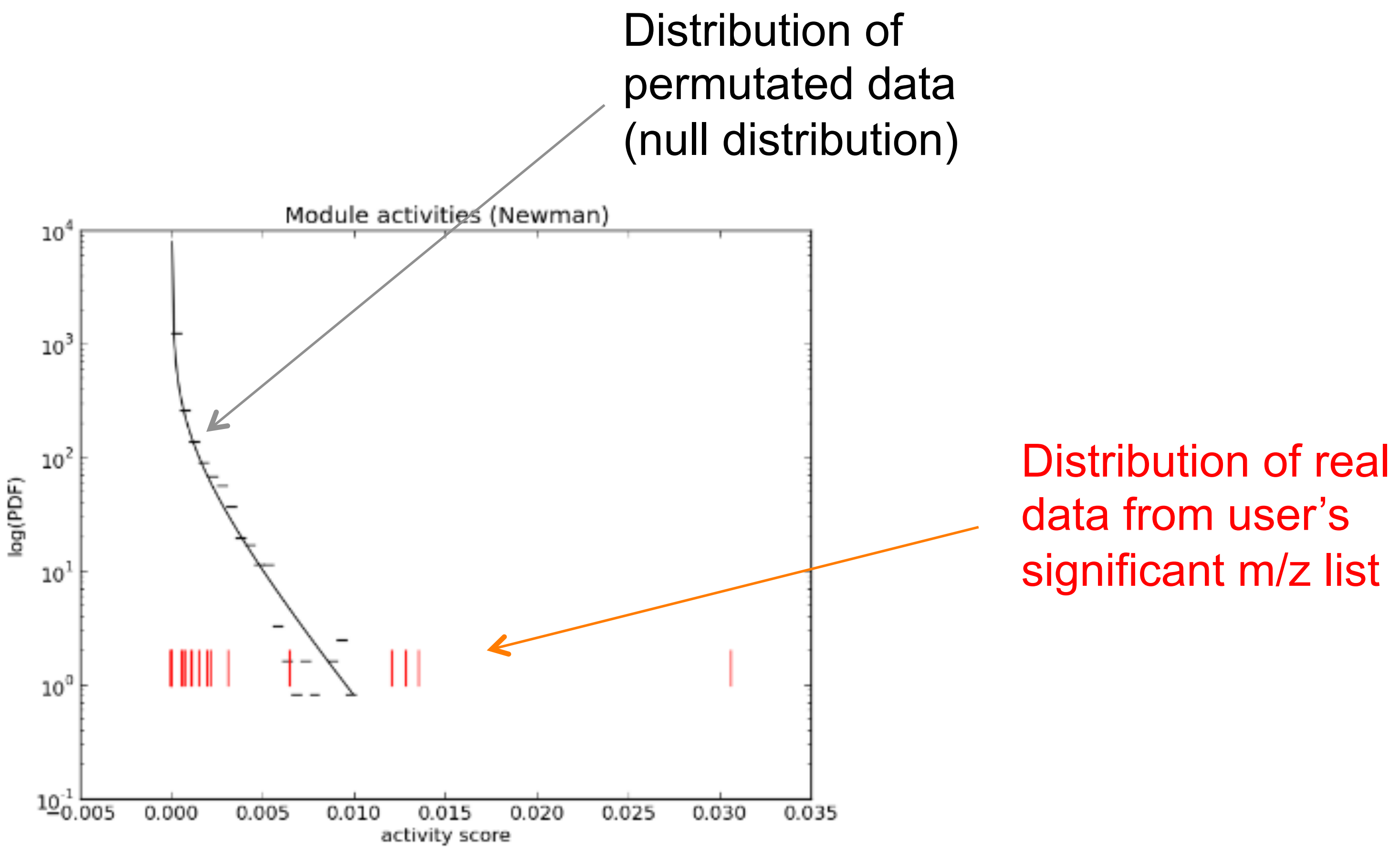


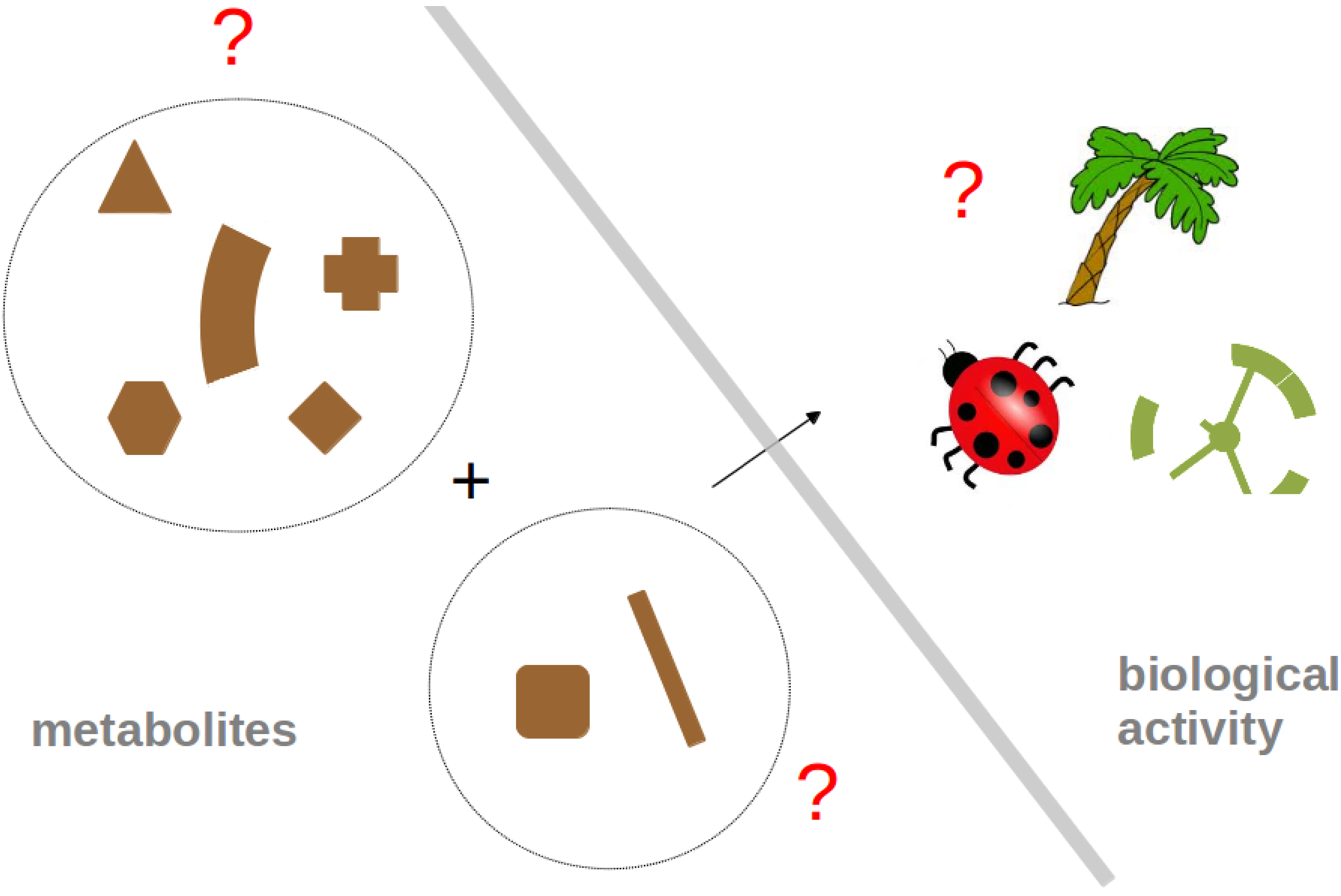
Mummichog for pathway/network analysis

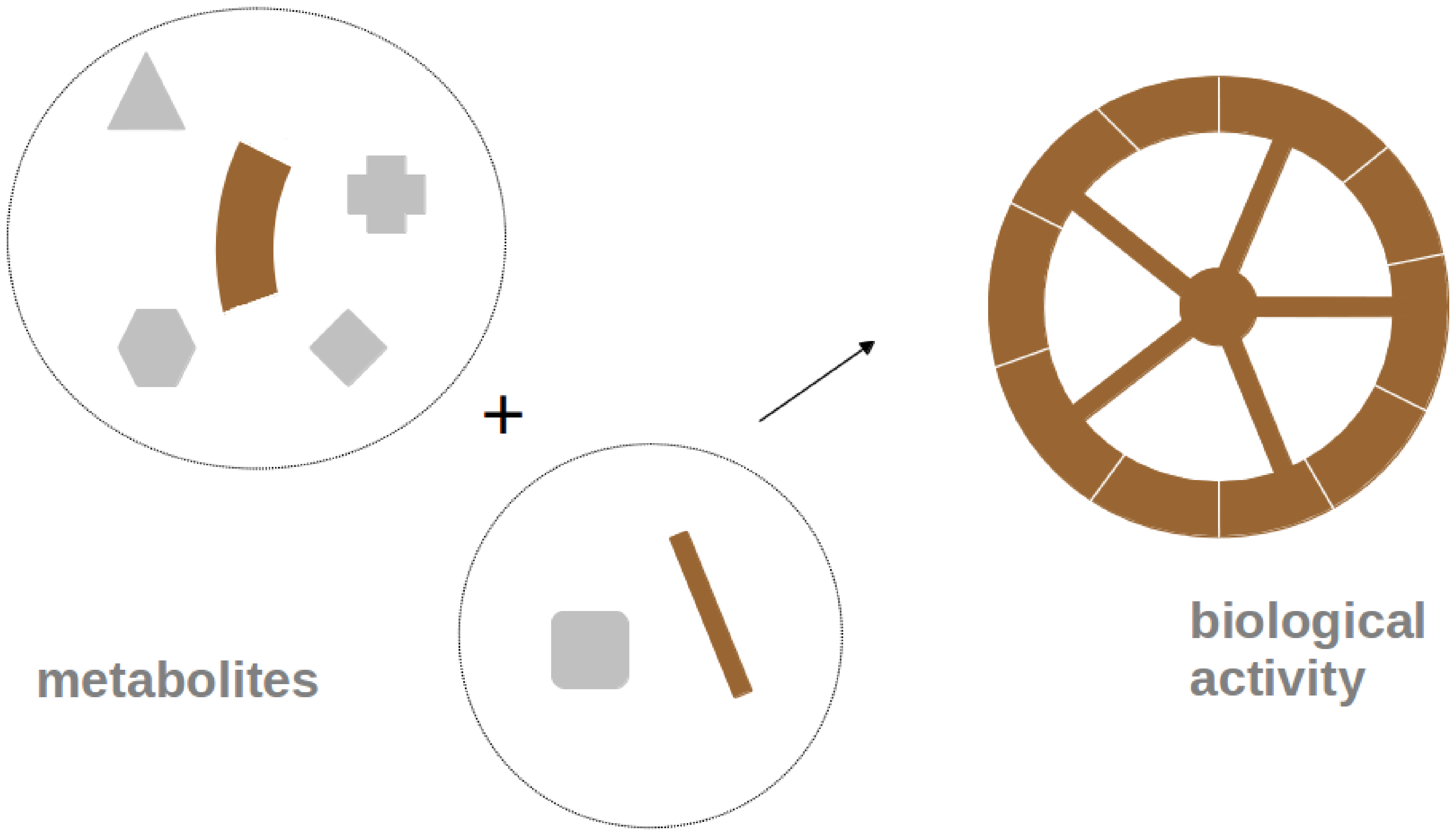


Li et al. 2013. PLoS Computational Biology. 9:e10031323

Testing statistical significance in *mummichog*




?
+
?
?
metabolites
biological
activity


metabolites
biological activity

Demo

Pathway vs module analysis in *mummichog*

- Pathways are predefined units with human knowledge. Network modules are less biased but data dependent. A module can be within a pathway or in between several pathways. The two approaches are rather complementary.

- Null distribution is estimated from permutation data, thus p-value for untargeted metabolomics data is empirically computed.

- Module statistics are based on a module activity score; pathway statistics are based on an enrichment test.

- Pathway definition may differ between databases.

Application cases

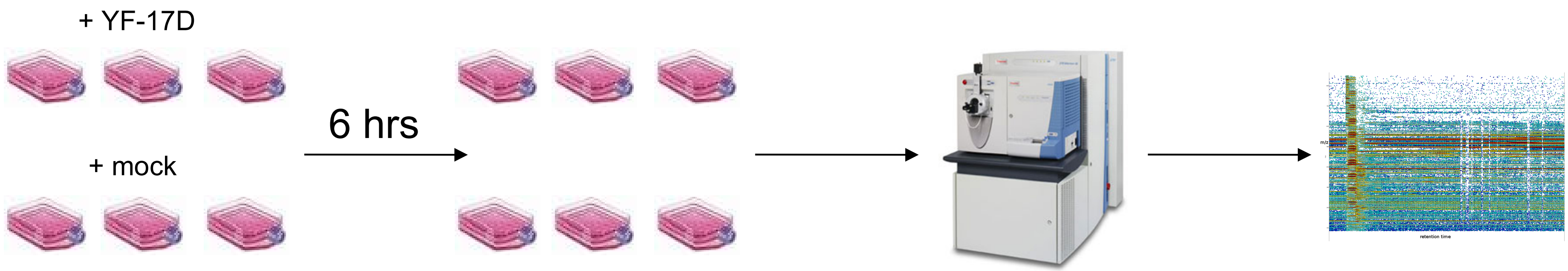
- Immune response to virus – arginine emerging as master regulator (Li et al, 2013, PLoS Computational Biology. 9:e10031323; Ravindran et al. 2014. Science 343:313)

- Combining *mummichog* with regression models – fly longevity (Hoffman et al, Aging Cell 13: 596-604)

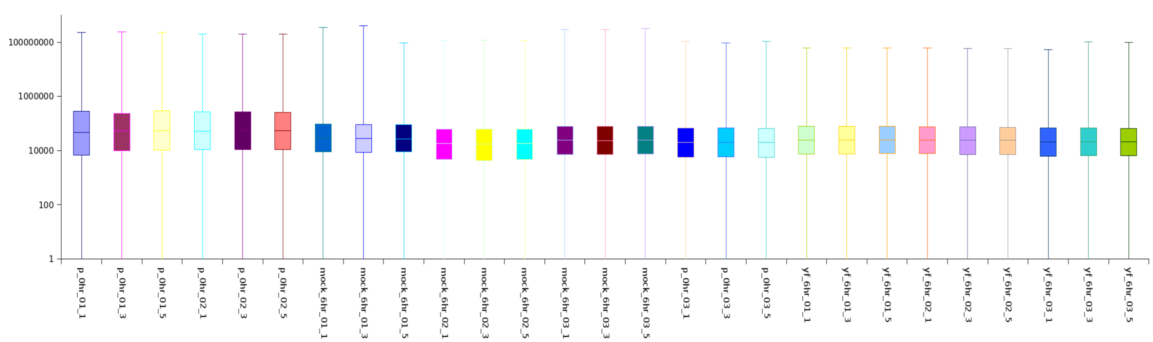
- Connecting transcriptomics – T cell autophagy and memory (Xu et al, Nature Immunology. 15:1152-1161)

Case study: viral activation of immune cells

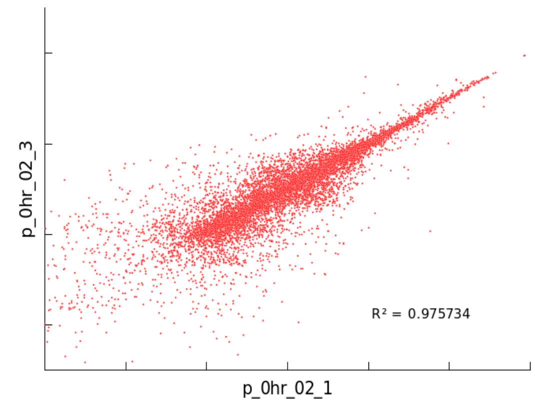
Monocyte derived dendritic cells (moDC)



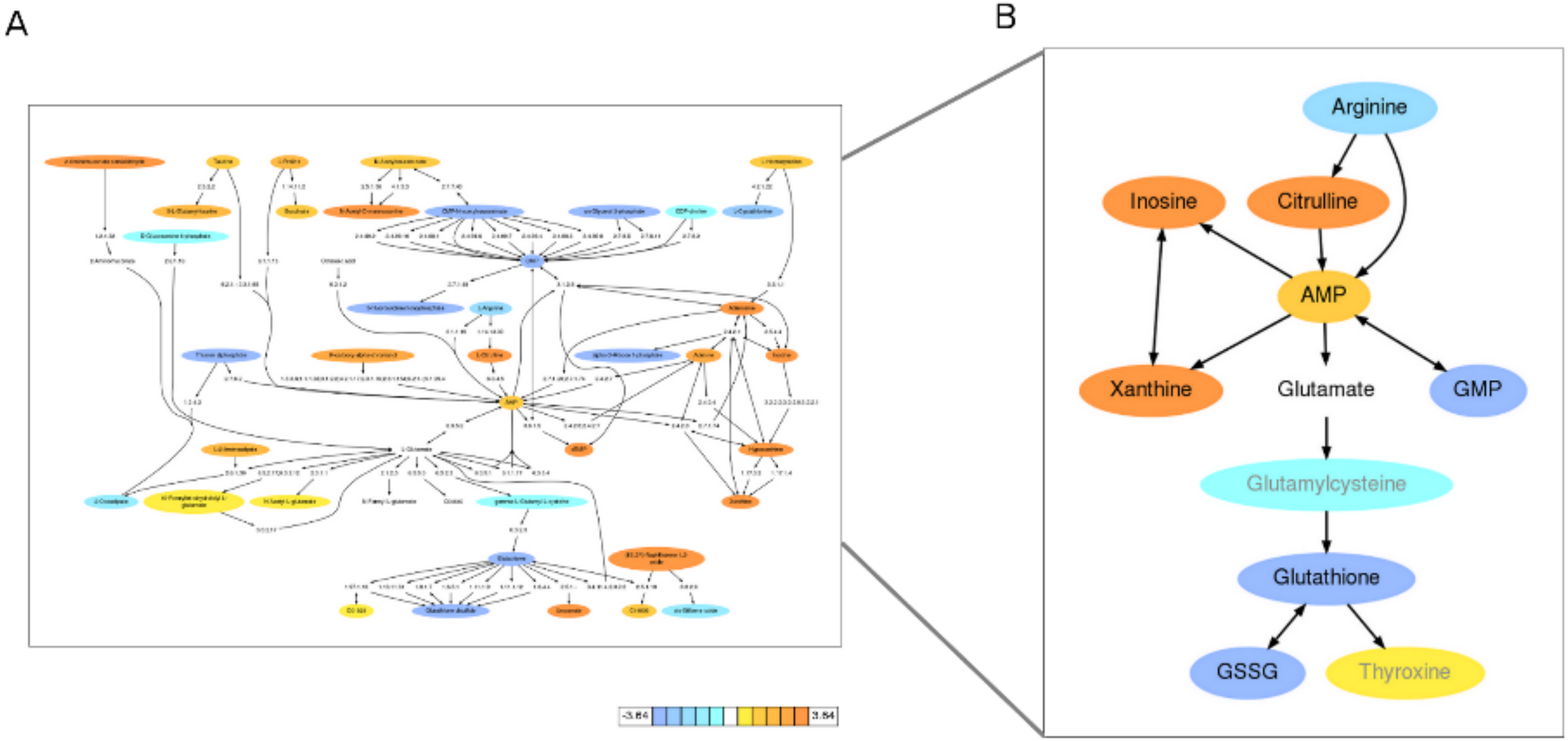
QA: total ion counts are similar among samples



technical replicates, 10,000 features

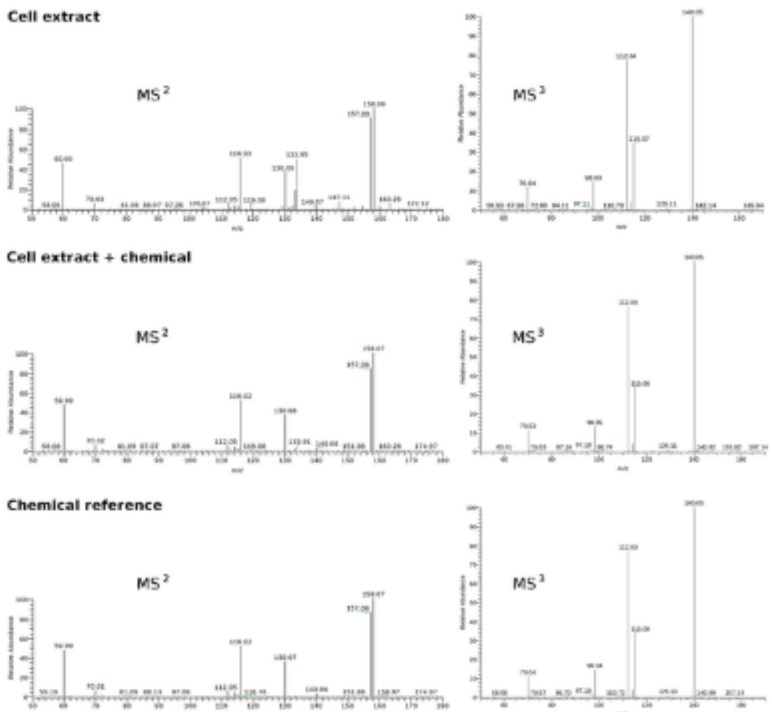


Metabolite network after viral activation

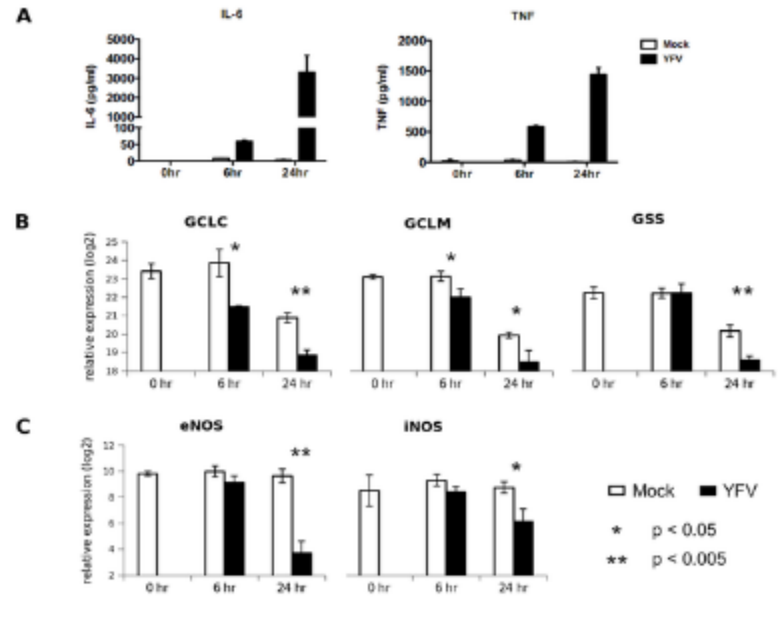


Li et al. 2013. PLoS Computational Biology. 9:e10031323

Experimental validation of *mummichog* prediction



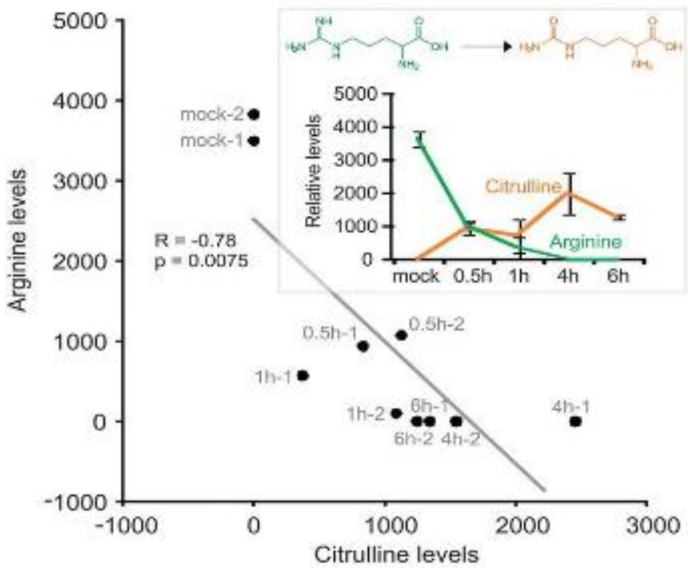
Tandem mass spectrometry confirmed 9/11 metabolites



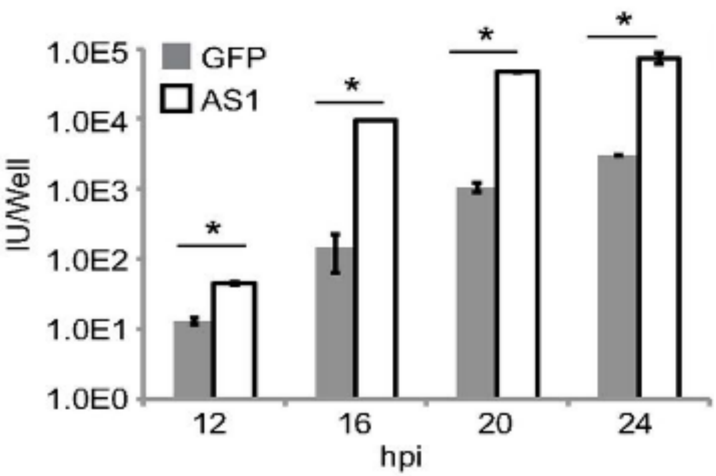
Gene expression supported GSH/GSSG depletion and Arg/Cit conversion

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

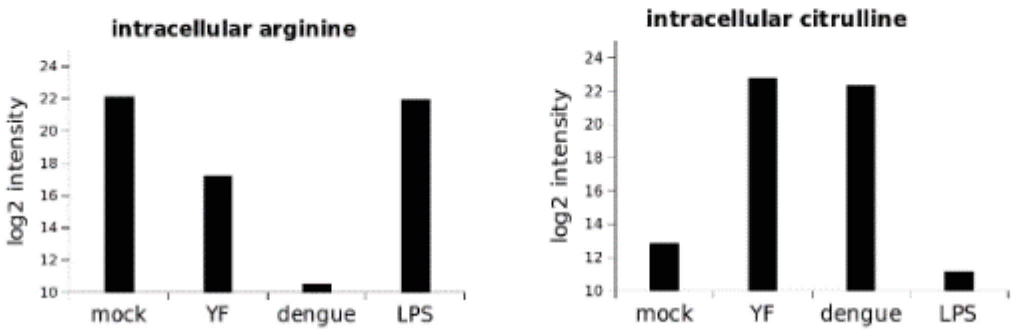
Arginine as master regulator of viral response



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

Aging in Drosophila

Aging Cell (2014) **13**, pp596–604 Doi: 10.1111/acel.12215

Effects of age, sex, and genotype on high-sensitivity metabolomic profiles in the fruit fly, *Drosophila melanogaster*

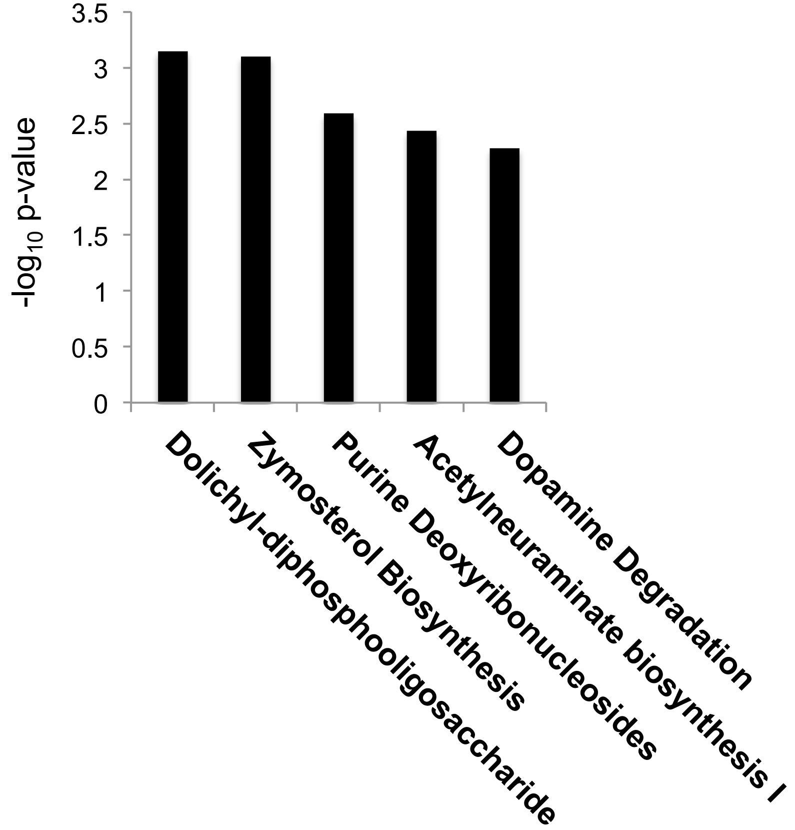
Jessica M. Hoffman,[1†] Quinlyn A. Soltow,[2,3,4†] Shuzhao Li,[2] Alfire Sidik,[1,6] Dean P. Jones[2,3,5]* and Daniel E. L. Promislow[1,7]*

- Modeling effect on metabolite concentration (Y). A: age, S: sex, G: genotype

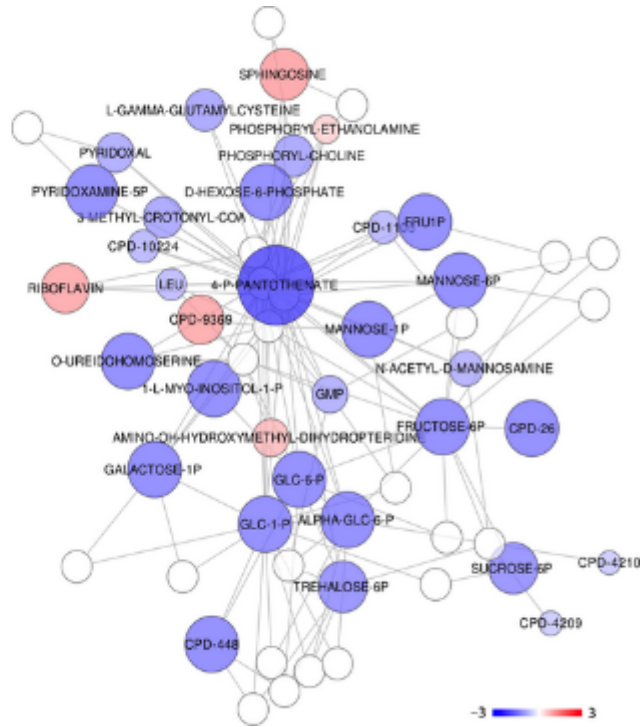
$$Y = \mu + A + S + G + A \times S + A \times G + G \times S + \varepsilon$$

- Using significant features from the model to test pathway/network enrichment in *mummichog*

Top features in regression model used for *mummichog* input



Pathways significantly associated with aging in this Drosophila model.

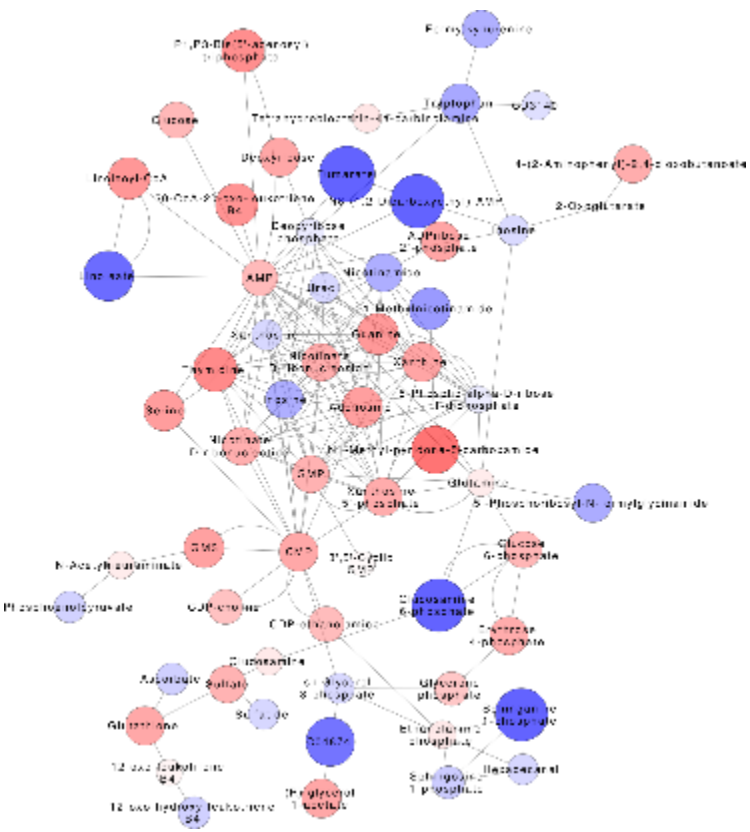


Example output module from *mummichog* analysis with color hue determined by the sign and size and color intensity determined by the magnitude of the regression coefficient in the age model (blue is negative, red is positive). The metabolites are putatively annotated based on *m/z* ratio. This particular module is enriched for metabolites associated with glycolysis, for metabolites that feed the glycolytic pathway, and for metabolites associated with glycophospholipid metabolism.

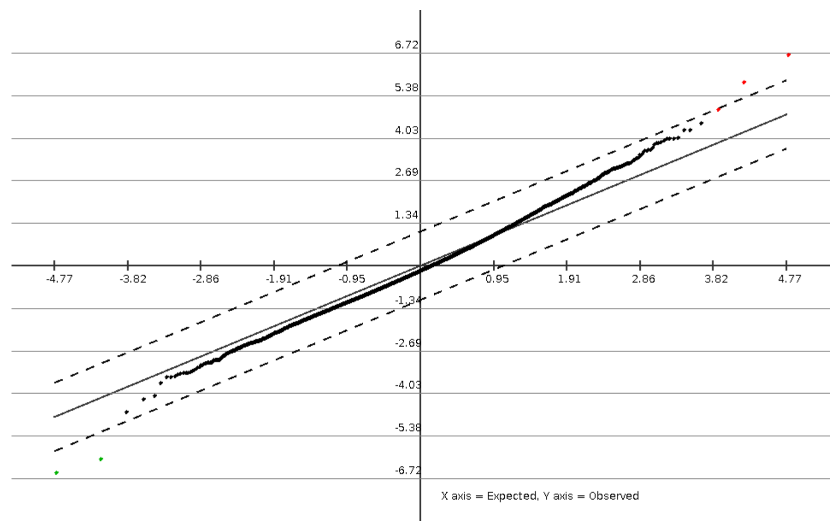
nature immunology

Autophagy is essential for effector CD8+ T cell survival and memory formation

Xiaojin Xu[1,5], Koichi Araki[1,5], Shuzhao Li[2], Jin-Hwan Han[1], Lilin Ye[1], Wendy G Tan[1], Bogumila T Konieczny[1], Monique W Bruinsma[3], Jennifer Martinez[4], Erika L Pearce[3], Douglas R Green[4], Dean P Jones[2], Herbert W Virgin[3] & Rafi Ahmed[1]

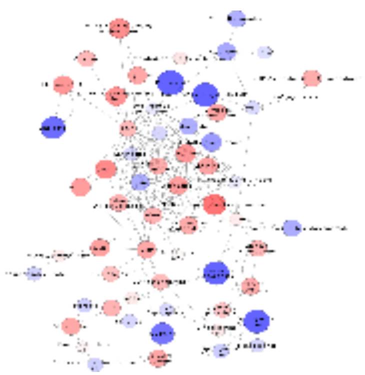


metabolomics



transcriptomics

Enzymes associated with significant metabolites

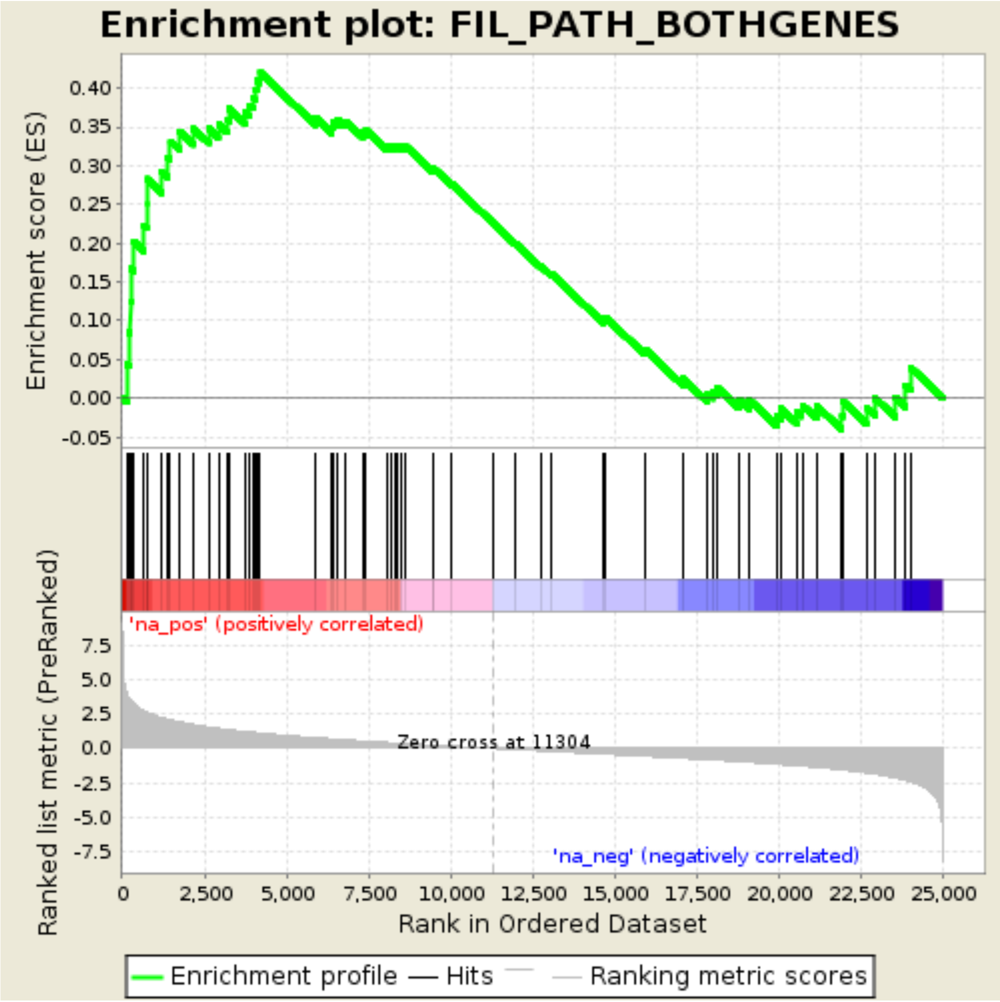


2.5.1.56, 2.7.1.91, 2.4.2.9, 2.4.2.8, 1.14.16.4, 1.14.16.5, 3.6.1.22, 2.4.2.1, 2.4.1.80, 3.1.4.35, 2.4.2.4, 2.4.2.7, 2.4.2.14, 2.4.2.11, 2.4.2.12, 3.5.4.17, 2.4.2.19, 1.1.1.94, 3.1.6.8, 3.1.6.1, 4.3.2.2, 1.14.14.1, 3.1.3.4, 3.1.3.5, 3.1.4.46, 2.4.1.141, 1.3.99.13, 3.6.1.5, 3.6.1.6, 3.6.1.9, 3.6.1.8, 2.1.1.1, 3.5.1.9, 2.7.1.1, 2.7.1.8, 3.1.1.4, 2.7.8.-, 3.2.1.18, 2.7.8.2, 2.7.8.5, 2.7.8.8, 1.1.99.4, 1.1.99.5, 2.7.1.74, 2.7.7.14, 3.6.1.29, 3.6.1.19, 3.6.1.17, 2.7.1.138, 2.4.1.47, 6.3.5.1, 6.3.5.3, 6.3.5.2, 6.2.1.3, 1.1.1.102, 4.1.3.3, 1.14.13.30, 3.2.2.1, 2.5.1.18, 3.5.1.23, 1.13.11.11, 2.6.1.7, 2.7.1.59, 4.1.2.13, 2.4.99.8, 2.4.99.9, 1.3.3.6, 3.1.3.10, 3.2.1.46, 3.2.1.45, 6.3.4.4, 2.2.1.1, 2.2.1.2, 6.3.4.1, 2.7.8.1, 2.7.1.20, 1.7.1.7, 2.4.2.22, 2.3.1.24, 2.7.8.11, 2.7.8.15, 3.5.4.3, 3.1.4.2, 3.5.4.6, 2.7.6.1, 2.6.1.16, 3.1.4.12, 3.1.4.17, 2.4.1.117, 1.2.3.1, 3.5.4.4, 1.4.3.2, 4.1.2.27, 3.1.4.3, 6.1.1.2, 4.2.1.17, 3.2.2.2, 3.1.2.2, 3.2.2.6, 3.2.2.5, 3.5.99.6, 3.2.2.8, 1.1.1.8, 3.7.1.3, 1.13.11.34

Gpd1l, Kdsr, Ado, Acox1, Gmpr2, Tkt, Alg5, Alg13, Hprt, Nampt, Gsta4, Gstk1, Gstm1, Gstm4, Gsto1, Gstp1, Gstp2, Gstt2, Hpgds, Gfpt1, Adk, Nagk, Dck, Sphk1, Sphk2, Prps1, Prps2, Cept1, Ept1, Cept1, Cdipt, Plb1, Acot2, Lpin1, Lpin2, Pde1b, Pde2a, Pde3b, Pde4a, Pde4d, Pde7a, Pde8a, Pde5a, Arsa, Gba2, Galc, Bst1, Cd38, Asah1, Asah2, Ada, Ampd1, Ampd2, Ampd3, Cant1, Enpp1, Itpa, Enpp4, Aldoa, Aldoc, Sgpl1, Npl, Acsl1, Acsl3, Acsl4, Acsl5

Gpda, Acox3, Oxla, Gmpr1, T23o, Lox5, Cp4f3, Cp4fe, Cp19a, Cp1a1, Cp1a2, Cp1b1, Cp237, Cp238, Cp239, Cp240, Cp254, Cp255, Cp270, Cp2a4, Cp2a5, Cp2ac, Cp2b9, Cp2ba, Cp2bj, Cp2ct, Cp2d9, Cp2da, Cp2db, Cp2dq, Cp2j5, Cp2j6, Cp2s1, Cp2u1, Cp341, Cp3ab, Cp3ad, Cp3ag, Cp3ap, Cp4b1, Cp4ca, Cp4x1, Cy250, Tph1, Tph2, Alkmo, Nnmt, Tktl1, Tktl2, Taldo, Cegt, Pnph, Typh, Apt, Nadc, Sia8a, Siat9, Gsta1, Gsta2, Gsta3, Gstm2, Gstm5, Gstm6, Gstm7, Gsto2, Gstt1, Gstt4, Maai, Mgst1, Mgst3, Aadat, Aatm, Kat1, Kat3, Gfpt2, Hkdc1, Hxk1, Hxk2, Hxk3, Cerk1, Pcy2, Chpt1, Pgps1, Gpt, Hrsl3, Pa21b, Pa24a, Pa24b, Pa24d, Pa24e, Pa24f, Pa2g5, Pa2ga, Pa2gc, Pa2gd, Pa2ge, Pa2gf, Pa2gx, Pg12a, Plpl9, Aco15, Acot1, Acot3, Acot4, Acot5, Baat, Bach, Them4, Lpin3, Lpp1, Lpp2, Lpp3, Lppr2, Lppr3, Lppr4, Ppc1a, Ppc1b, 5nt1a, 5nt1b, 5nt3a, 5nt3b, 5ntc, 5ntd, Ppap, Gpcp1, Asm, Nsma2, Nsma3, Nsma, Pde10, Pde11, Pde1a, Pde1c, Pde3a, Pde4c, Pde7b, Pde8b, Cncg, Cnrg, Pde10, Pde11, Pde6a, Pde6b, Pde6c, Pde9a, Neur1, Neur2, Neur3, Neur4, Glcm, Kfa, Acer1, Acer2, Guad, Gnpi1, Gnpi2, Entp1, Entp8, Entp4, Entp5, Enpp3, Ap4a, Nud12, Fhit, Kynu, Aldob, Echa, Echm, Echp, Pur8, Sywc, Sywm, Acbg1, Acbg2, Acsl6, S27a2, Pura1, Pura2, Nade, Guaa, Pur4

Enzymes associated with significant metabolites, significantly up in KO by GSEA analysis



Expression of genes corresponding to related enzymes are enriched for KO cells, DNA microarray data, GSEA (Gene Set Enrichment Analysis). Nominal p = 0, FWER p = 0.024.

Summary and future directions

- *Mummichog* rewrites the workflow of untargeted metabolomics, enabling rapid generation of quality hypotheses
- Limited by metabolic models, i.e. known metabolic knowledge
- Download sites: http://clinicalmetabolomics.org (phasing out http://atcg.googlecode.com)
- Version 1.0 is out. Version 2 is in the making

Acknowledgement of research supports from US National Institute of Health (NIEHS P30 ES019776, NIAID HHSN272201200031C, NIA 1R01 AG038746-02, NHLBI 1P20 HL113451-01), Department of Defense (HT9404-13-1-003) and California Breast Cancer Research Program (21UB-8002).