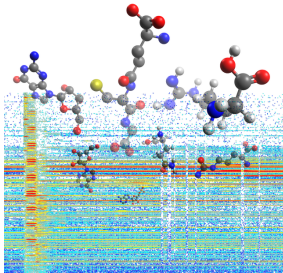




# Metabolomics pathway analysis and network integration

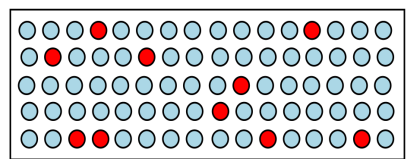


Shuzhao Li, Ph.D.  
Assistant Professor, Dept. Medicine,  
Emory University School of Medicine  
E-mail: [shuzhao.li@gmail.com](mailto:shuzhao.li@gmail.com)  
July 20, 2016

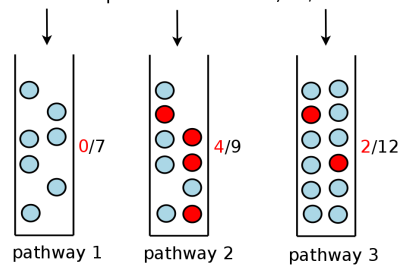
- **Metabolomics pathway analysis and *mummichog***
- **Examples**
  - Snyderome re-visited
  - MWAS of NAFLD
  - PCB Exposure
  - Cross-generation cancer risk
  - Memory T cells
  - VZV systems immunology

## Pathway enrichment test

If metabolites are known; red are significant metabolites

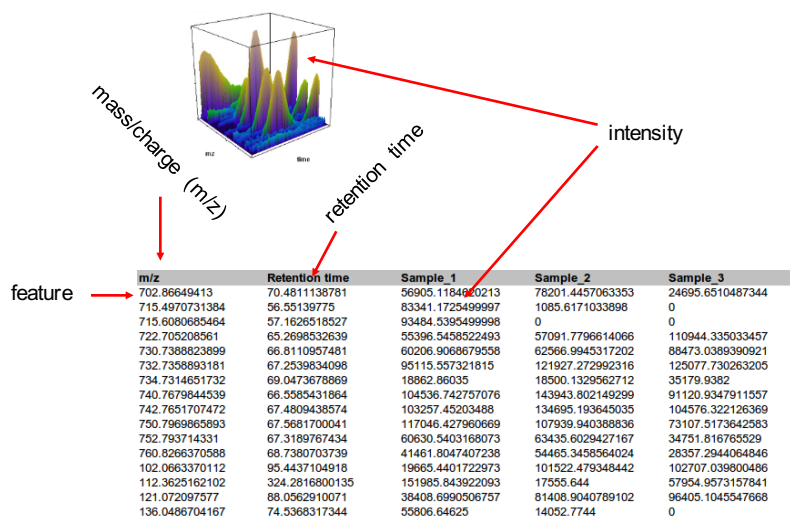


experimental data set, 10/80



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

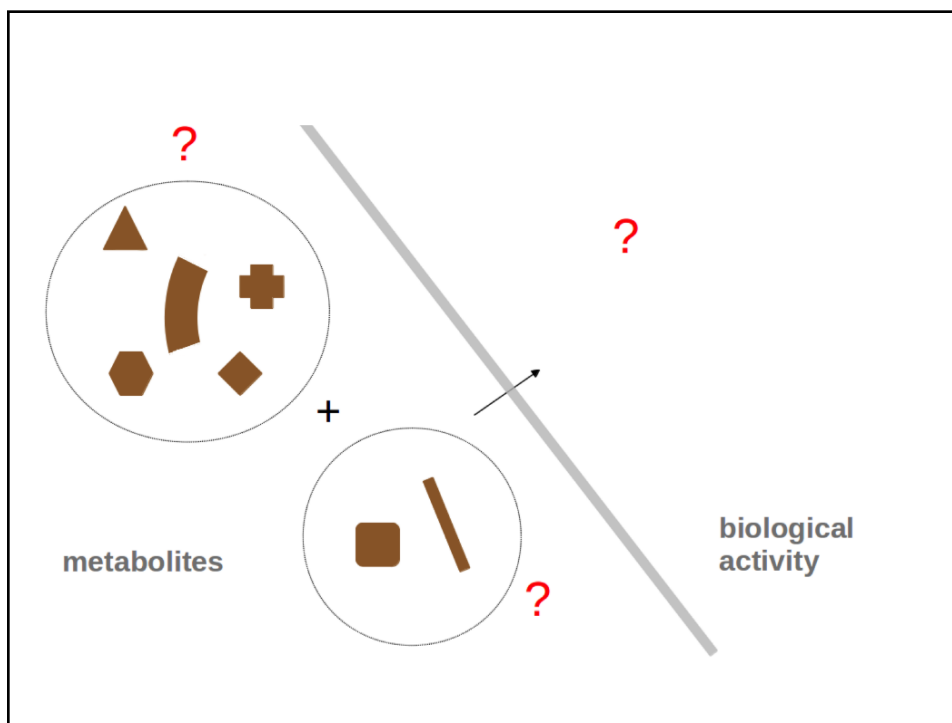
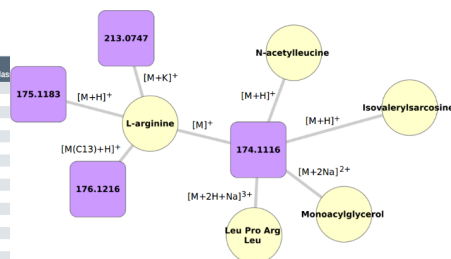
## Untargeted metabolomics data



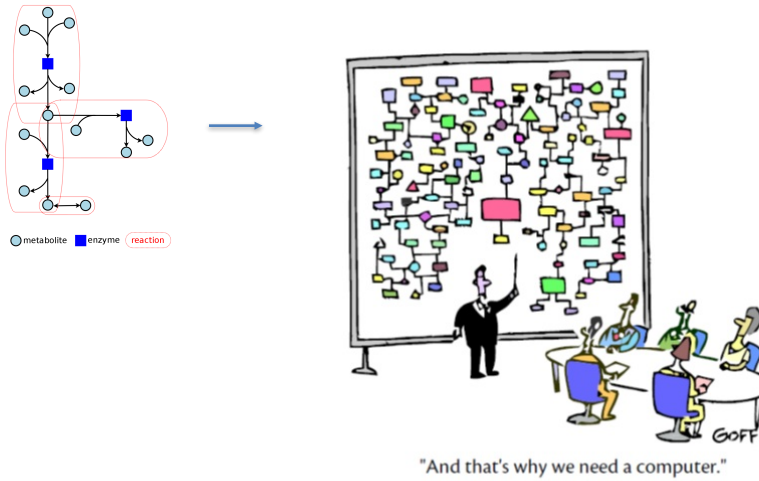
# Uncertainty in matching metabolites - features

Search of m/z 190.1065 in HMDB with accurate matching

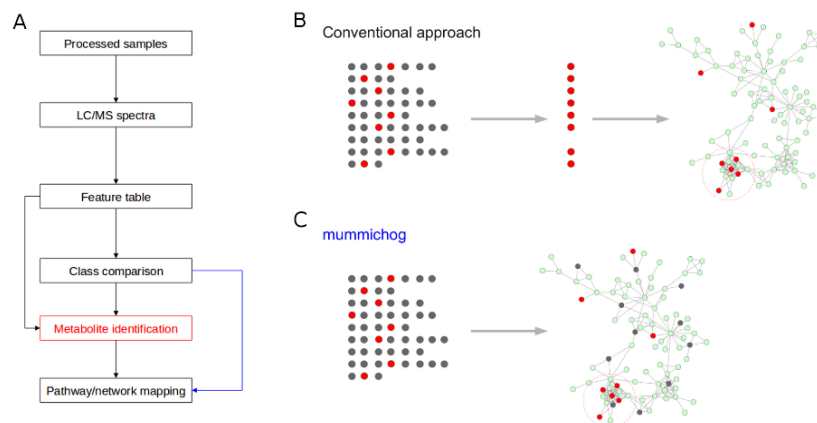
Common Name	Chemical Formula	Adduct MW (Da)	MW Difference (Da)	[Charge/Mass - Adduct/Mass]
cis-4-Octenedioic acid	C8H12O4	190.107391 [172.073563]	8.85E-4	
Gamma-Butyrolactone	C4H6O2	190.107391 [86.036781]	8.85E-4	
3-Methylcrotonylglycine	C7H11NO3	190.107391 [157.073989]	8.85E-4	
Lactaldehyde	C3H4O2	190.107391 [74.036781]	8.85E-4	
Hydroxyacetone	C3H6O2	190.107391 [74.036781]	8.85E-4	
Quotian-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	
Mivalonic acid	C8H12O4	190.107391 [148.073563]	8.85E-4	
(R)-2,3-Dihydroxy-3-methylvalerate	C8H12O4	190.107391 [148.073563]	8.85E-4	
D,Lactaldehyde	C3H4O2	190.107391 [74.036781]	8.85E-4	
Tyrosylglycine	C7H11NO3	190.107391 [157.073989]	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	
Diacyl	C4H6O2	190.107391 [86.036781]	8.85E-4	2M+NH4 [1+]
Pyrogulonic acid	C5H7NO3	190.107925 [129.042587]	0.001419	M+IsoProp+H [1+]
N-Acetylglycine	C5H7NO3	190.107925 [129.042587]	0.001419	M+IsoProp+H [1+]
3-Pyrroline-4-hydroxy-2-carboxylate	C5H7NO3	190.107925 [129.042587]	0.001419	M+IsoProp+H [1+]
Pyridine hydroxycarboxylic acid	C5H7NO3	190.107925 [129.042587]	0.001419	M+IsoProp+H [1+]
Pyridonecarboxylic acid	C5H7NO3	190.107925 [129.042587]	0.001419	M+IsoProp+H [1+]
Penicillidine	C17H26NO4	190.108047 [178.201538]	0.001541	M+2H [2-]
Kyotorphin	C15H23NO4	190.108063 [137.176018]	0.001556	M+ACN+2H [2-]
18-Hydroxy cortisol	C21H30O6	190.109406 [178.204254]	0.002099	M+2H [2-]
Etipramil	C23H30N2	190.109666 [134.249905]	0.003159	M+2Na [2-]



## Genome-scale Metabolic model



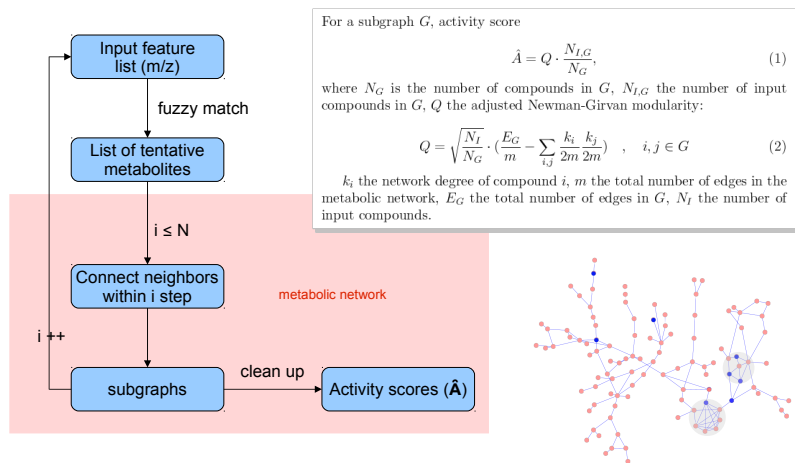
## *Mummichog* bridging metabolic models



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

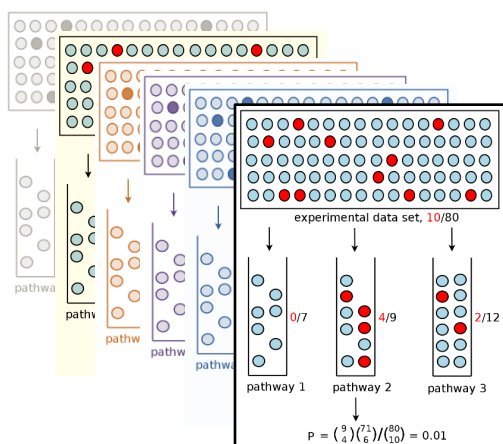


## Module analysis in *mummichog*



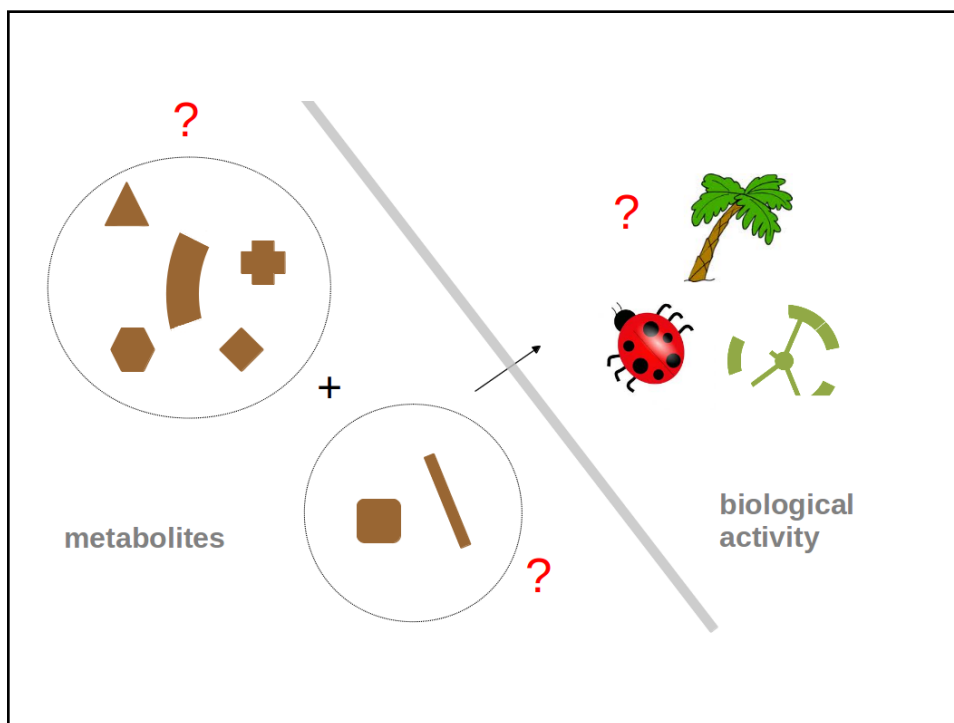
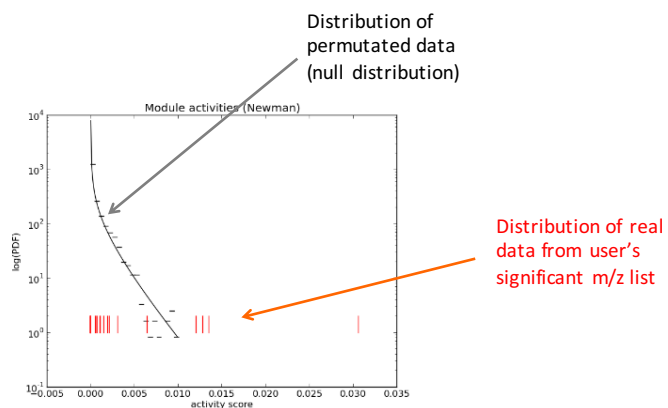
Li et al. 2013. PLoS Computational Biology. 9:e10031323

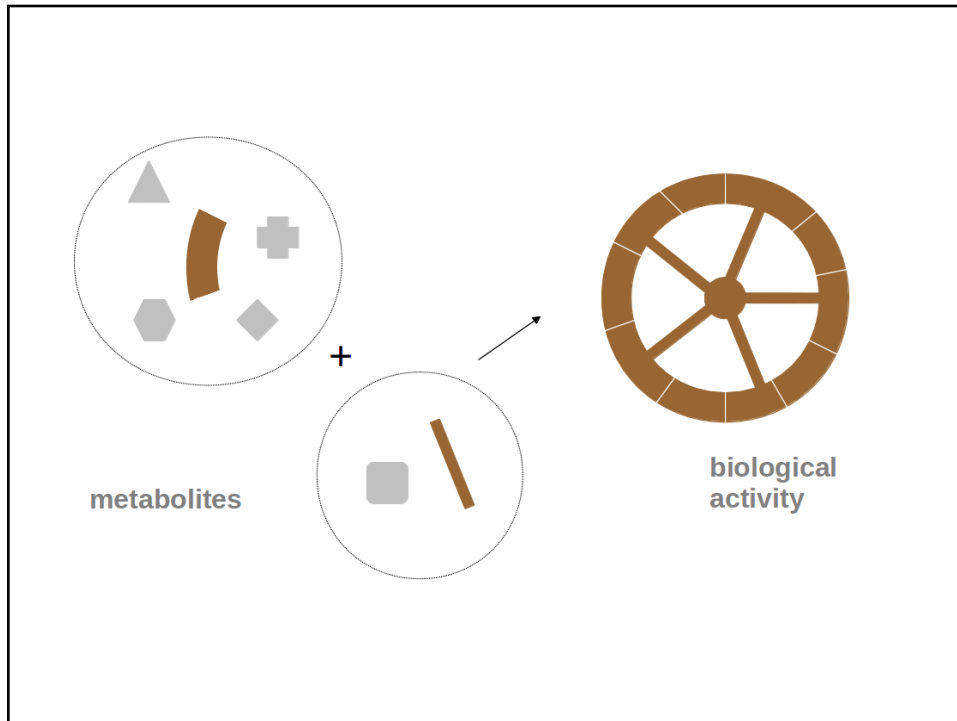
## Pathway analysis in *mummichog*



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

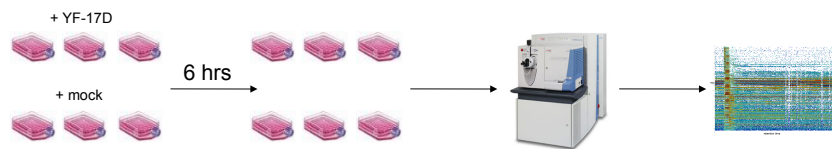
## Testing module/pathway significance in *mummichog*



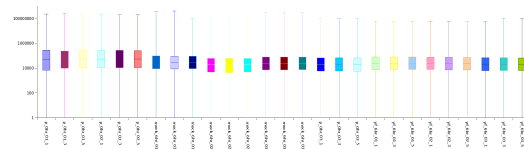


## Case study: viral activation of immune cells

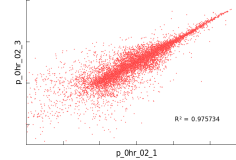
Monocyte derived dendritic cells (moDC)



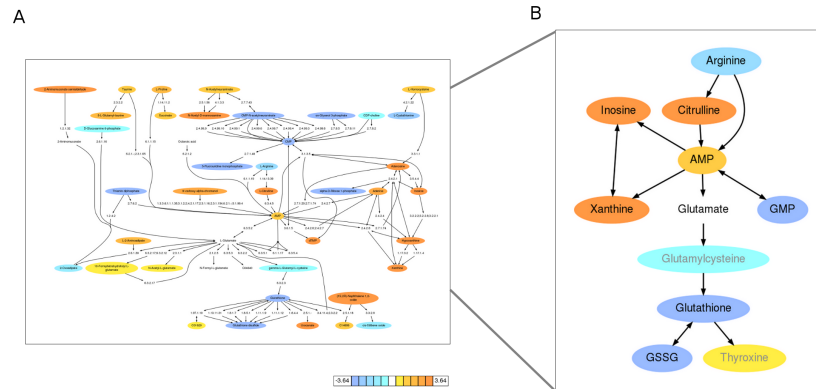
QA: total ion counts are similar among samples



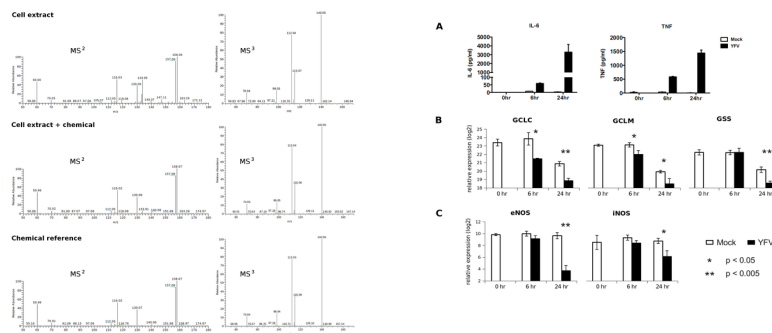
technical replicates, 10,000 features



## *Mummichog*: viral activation of immune cells



## Experimental validation of *mummichog* prediction

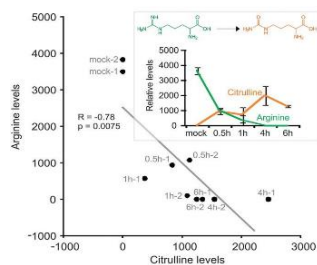


Tandem mass spectrometry confirmed 9/11 metabolites

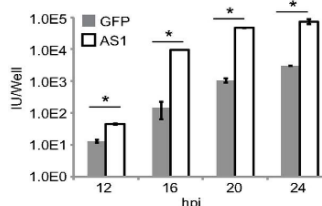
Gene expression supported GSH/GSSG depletion and Arg/Cit conversion

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

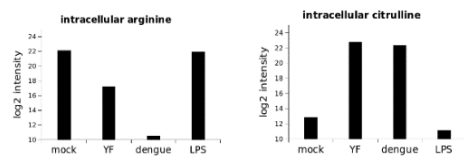
# Arginine as master regulator of viral response



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

```
sl149@Soni's MacBook Pro:~/play/mummichog-1.0.7/tests$ python ../mummichog/main.py -f testdata.txt -o demo
```

```
-----
o0                                oooooooooo
o00 00000 00000 000 0000
o00 0      00000 00000 00000
o000 0000 000000 0000 000000
0000 0      000000 0000 00000000
00000 0000
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  
0 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 4  
57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75  
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 v

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

mummichog.org

mummichog.org

Mummichog

FOR VERSION 1.0.6  
Mummichog is a Python program for analyzing data from high throughput, untargeted metabolomics. It leverages the organization of metabolic networks to predict functional activity directly from feature labels, bypassing metabolic identification. The features include:  
1. identifying significantly enriched metabolic pathways;  
2. identifying significant modules in the metabolic network;  
3. visualization of top networks in web browser;  
4. visualization that also plugs into Cytoscape;  
5. metabolic enrichment;  
6. metabolic models for different species through plugins.  
Major updates since version 1.0.3  
1. improved network visualization options  
2. added support for single input files, where threshold of input significant list is automatically determined. Version 1.0.3 added -c option, where user can specify cutoff p-value.  
3. added Readme Converter tool (see forum).  
4. Version 1.0.7 bug fix on visualization and instrument parameter input.

Install

- Download mummichog from [here](#).
- Mummichog is a Python 2.x package, that requires a few libraries (numpy and scipy, networkx, obviously). While their dependencies (bracketes) are common and easy to install, we recommend to install Anaconda or Scientific Python (also better). They contain all dependencies, which will fulfill all dependencies.
- You can launch mummichog to where you like. No installation is necessary for mummichog. Go to the test directory to test mummichog on your computer.

python ../mummichog/main.py -f testdata.txt -o testdemo

Input format

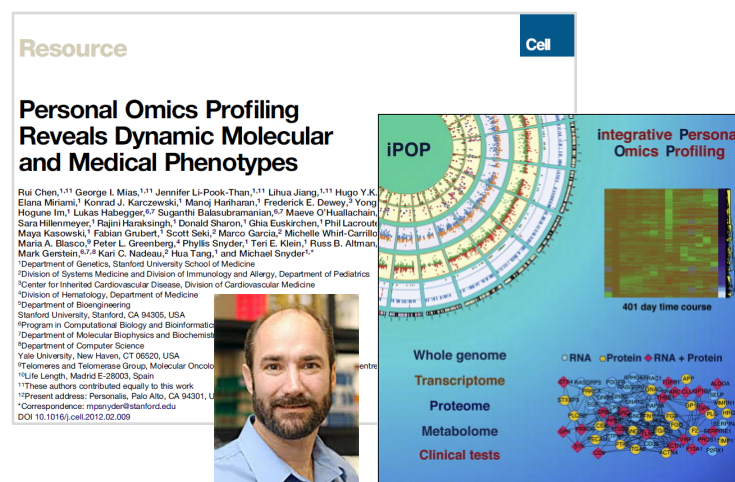
- A tab-delimited text file is used as input. One feature per line. Any rows starting with "f" will be skipped. Each line has to contain the following tab-delimited numeric fields: (ms, retention time, p-value, statistic score)

Example of input file:  
f1 123.456 100.000 0.0001  
f2 123.456 100.000 0.0001  
f3 123.456 100.000 0.0001  
f4 123.456 100.000 0.0001  
f5 123.456 100.000 0.0001

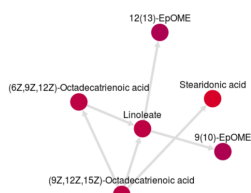
## A few more examples

- Snyderome re-visited
- MWAS of NAFLD
- PCB Exposure
- Cross-generation cancer risk
- Memory T cells
- VZV systems immunology

## Snyderome: personal omics (2012)

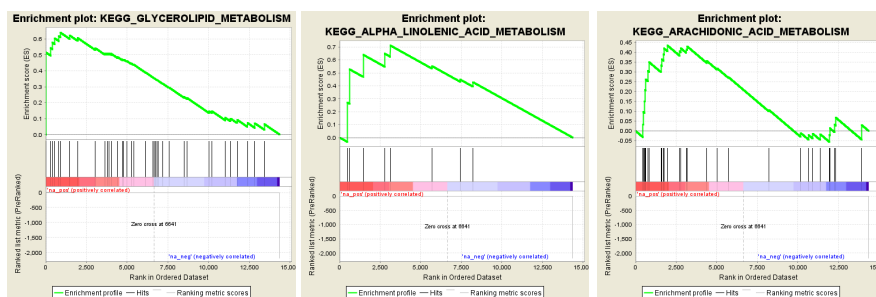


## Mummichog interpretation of Snyder metabolome



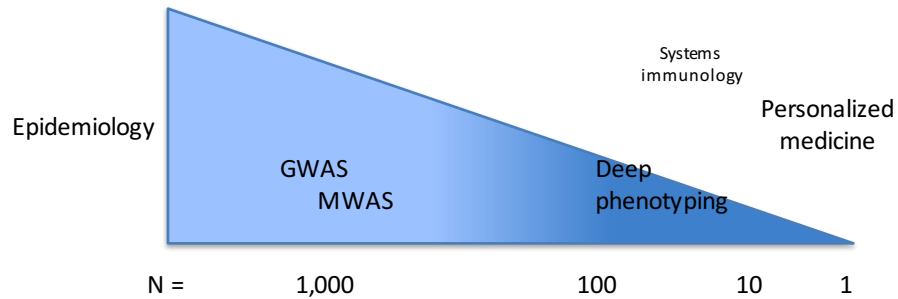
Pathways	overlap_size	pathway_size	p-value (raw)	p-value
Linoleate metabolism	8	10	3e-05	0.00635
Glycerophospholipid metabolism	7	19	0.04285	0.01085
Porphyryn metabolism	5	16	0.15234	0.03058

## Mummichog interpretation of Snyder metabolome

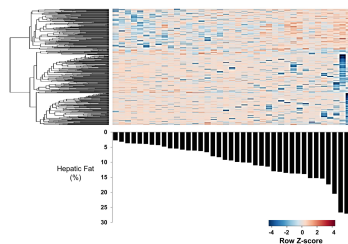


Pathways	overlap_size	pathway_size	p-value (raw)	p-value
Linoleate metabolism	8	10	3e-05	0.00635
Glycerophospholipid metabolism	7	19	0.04285	0.01085
Porphyryn metabolism	5	16	0.15234	0.03058

## 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
Biotin metabolism	3	13	0.02493
Di-saturated 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

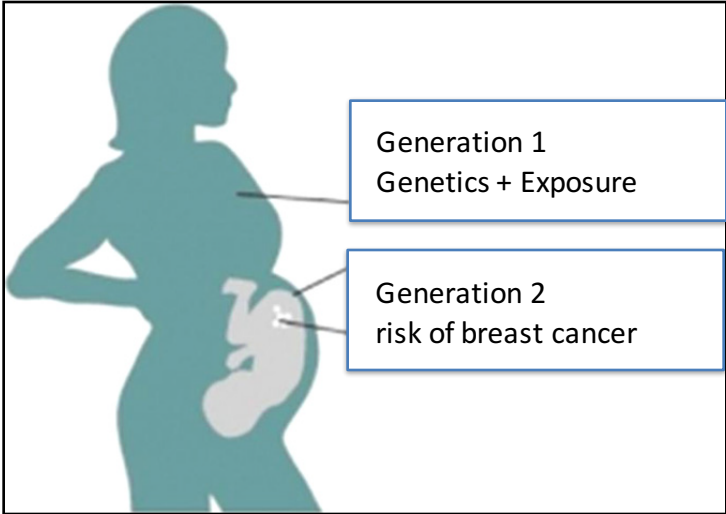
Jin, Banton, et al., 2016. *The Journal of pediatrics* 172: 14



# MWAS of Polychlorinated biophenyl

Unpublished data removed

# CHDS: exposure and risk of breast cancer



Generation 1  
Genetics + Exposure

Generation 2  
risk of breast cancer

Adapted from Perera F, Herbstman J, Reproductive Toxicology PMID: 21256208; Courtesy Barbara Cohn

## **Maternal metabolome associated with daughters' breast cancer (I)**

Unpublished data removed

## **Maternal metabolome associated with daughters' breast cancer (II)**

Unpublished data removed

$$G \times E \longrightarrow G \times M \times E$$

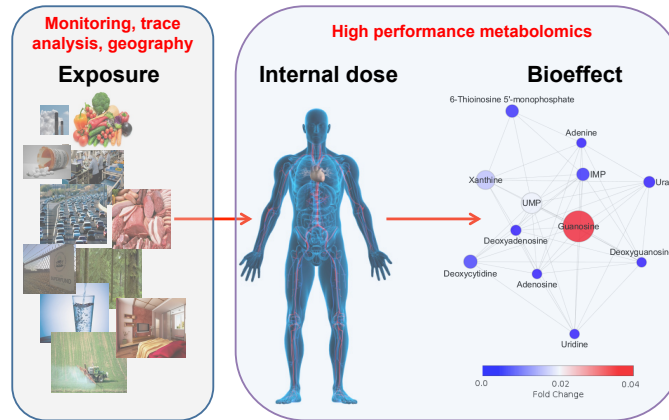
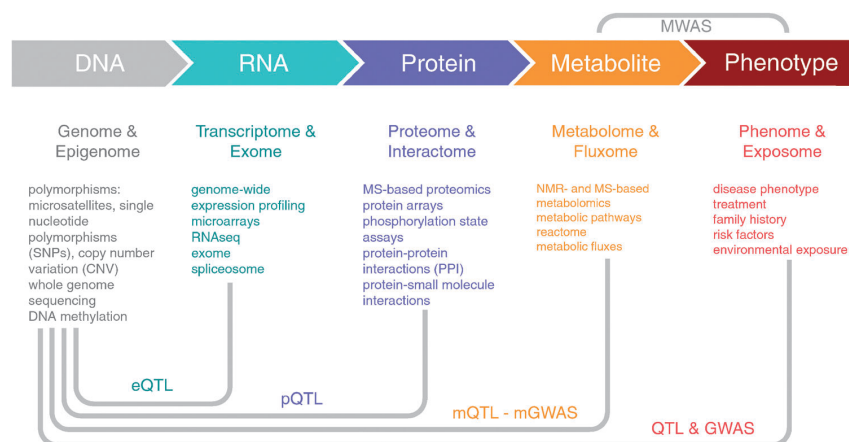


Figure Courtesy: Doug Walker

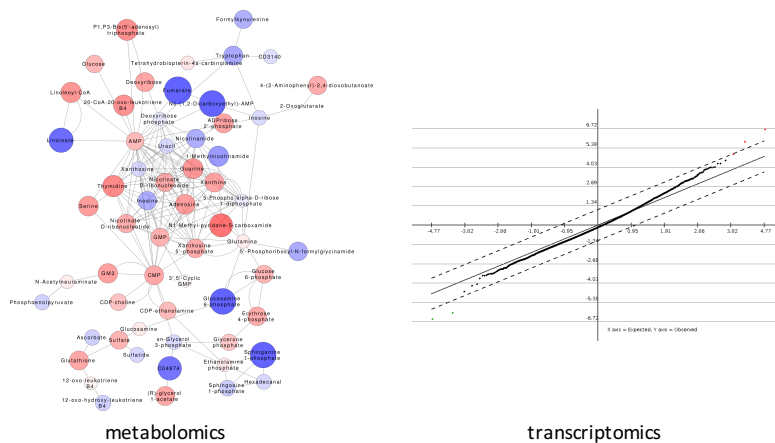
## Multi-omics integration



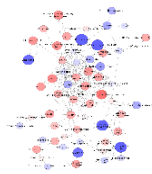
Dumas, 2012. Mol. BioSyst. 8:2494

## Autophagy is essential for effector CD8<sup>+</sup> T cell survival and memory formation

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



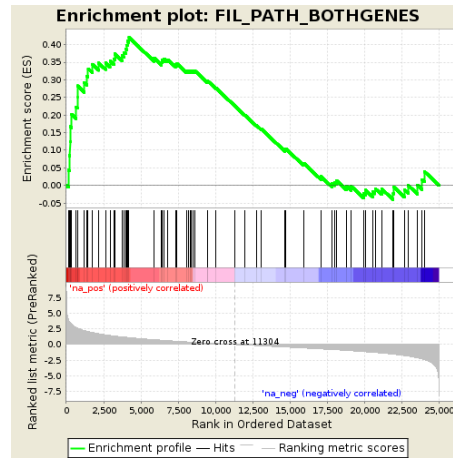
## Enzymes associated with significant metabolites



2.5.156, 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, 2.4.1.180, 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,  
1.3.9.13, 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, 2.7.1.188,  
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

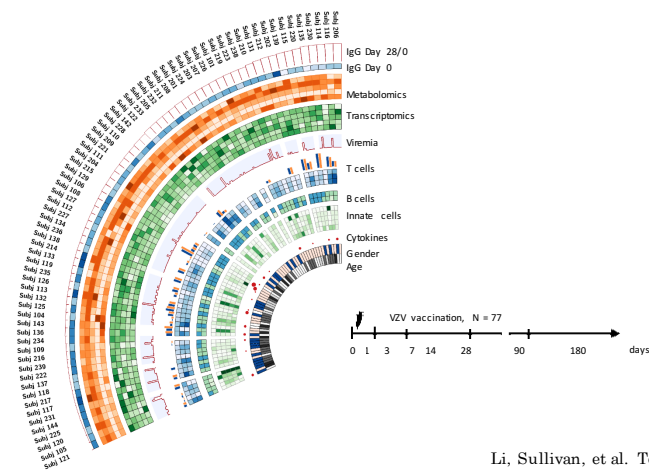
Gpd1l, Kdsr, Ado, Acox1, Gmp2, Tkt, Alg5,  
Alg13, Hprt, Nampt, Gsta4, Gstk1, Gstm1,  
Gstm2, Gstm3, Gstm4, Gstm5, Hpgds,  
Glt1, Adk, Nagk, Dck, Spnk1, Spnk2, Prps1,  
Prps2, Cept1, Ept1, Cept1, Cdpt, Plcb1, Aco2,  
Lpin1, Lpin2, Pde7b, Pde7a, Pde8b, Pde7a,  
Pde7d, Pde7a, Pde7a, Pde7a, Ansa, Gta2,  
Galc, Bst1, Cd38, Asah1, Asah2, Ada, Ampd1,  
Ampd2, Ampd3, Cant1, Enpp1, Ilpa, Enpp4,  
Aldoa, Aldoc, Sgpl1, Npl, Acsf1, Acsf3, Acsf4,  
Acsf5  
Gpda, Acox3, Oxla, Gmpr1, Tz3o, Lox5, Cp4f3,  
Cp4f6, Cp19a, Cp1a1, Cp1a2, Cp1b1, Cp237,  
Cp238, Cp23a, Cp24a, Cp24b, Cp25a, Cp25b,  
Cp24f, Cp2a6, Cp2a6, Cp2b9, Cp2ba, Cp2b,  
Cp2d1, Cp2f9, Cp2ba, Cp2bb, Cp2d9, Cp2f5,  
Cp2f6, Cp2a1, Cp2d1, Cp341, Cp3ab, Cp3d,  
Cp3ag, Cp3ap, Cp4b1, Cp4ca, Cp4x1, Cy250,  
Tph1, Tph2, Alkno, Nnmt, Tkl1, Tkl2, Taldo,  
Cegl, Prgh, Typh, Apt, Nadc, Slaiba, Slaf9,  
Gsta1, Gsta2, Gsta3, Gstm2, Gstm3, Gstm6,  
Gstm7, Gsto2, Gstm1, Gsta4, Maa1, Mgst1,  
Mgst3, Aadat, Aatm, Kat1, Kat3, Glp2, Hkdc1,  
Hxk1, Hxk2, Hxk3, Cekt1, Pcg2, Chpt1, Pgsf,  
Gpi, Hmst, P20r1, Pa2aa, Pa2ab, Pa2ba,  
Pa2be, Pa2d1, Pa2d5, Pa2ga, Pa2gc, Pa2gd,  
Pa2ge, Pa2gf, Pa2gh, Pg12a, Pg19, Aco15,  
Aco11, Aco3, Aco4, Aco5, Bcat, Bcat,  
Them4, Lpin3, Lpp1, Lpp2, Lpp3, Lpp4,  
Lppf, Ppca1, Ppca1b, Snt1a, Snt1b, Snt3a,  
Snt3b, Sntc, Sntd, Ppao, Gpcp1, Asm, Nma2,  
Nma3, Nma4, Pde10, Pde11, Pde1a, Pde1c,  
Pde3a, Pde4c, Pde7b, Pde8b, Cncc, Cncc,  
Pde10, Pde11, Pde6a, Pde6b, Pde6c, Pde6a,  
Neur1, Neur2, Neur3, Neur4, Glcm, Kfa, Aco1,  
Aco2, Gsta1, Gsta1, Gsta2, Entp1, Entp8,  
Entp4, Entp5, Enpp3, Apla, Nudt12, Pht, Kynj,  
Aldob, Echa, Echm, Echp, Par6, Sywc, Sywm,  
Acgl1, Acgl2, Acsb6, S27a2, Para1, Para2,  
Nada, Gaa, Puri

## Enzyme genes significantly enriched towards KO



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

## Comprehensive profiling of VZV immunization

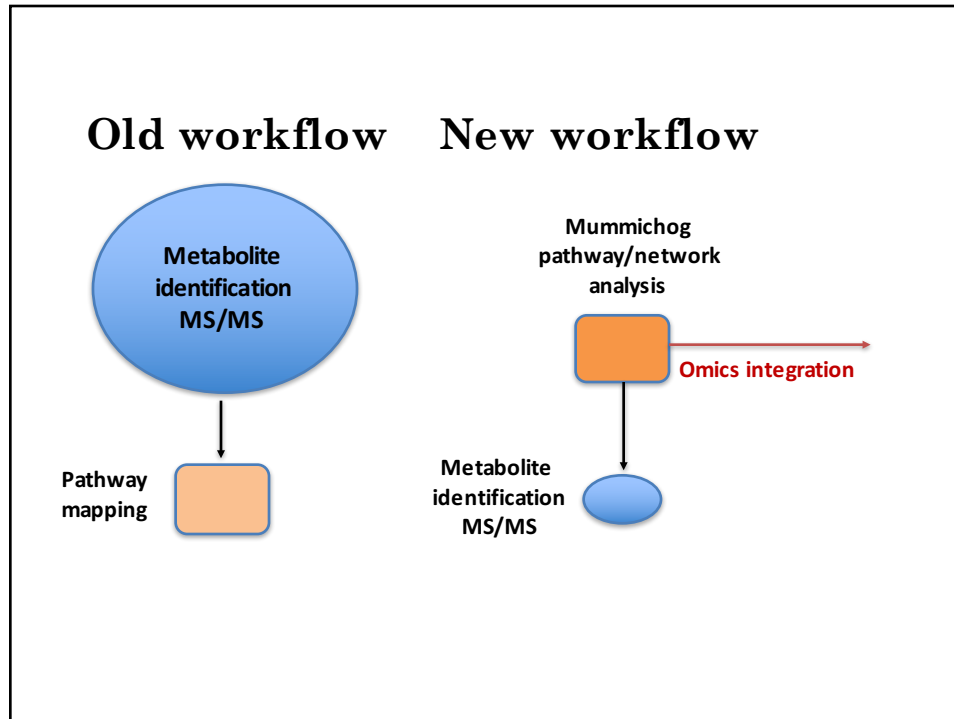


Li, Sullivan, et al. To be submitted.

Unpublished data removed

## Summary and future directions

- ❑ 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. Download at <http://mummichog.org>. Version 2 and server in the work.
- ❑ 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.



<p><b>Emory University</b></p> <p><u>Dept. Medicine</u>            Dean P. Jones            Young-Mi Go            Douglas Walker            ViLinh Tran            Bill Liang            Karan Uppal            Ken Liu            Loukia Lili-Williams            Sophia Banton            Andrei Todor</p> <p>Michael Hart            David Green            Annette Esper            Sushma Cribbs</p> <p>Mark Mulligan            Nadine Roupheal            Aneesh Mehta            Mary Galinski</p> <p><u>Dept. Genetics</u>            Judy Fridovich-Keil</p>		<p><u>Emory Vaccine Center</u>            Bali Pulendran            Helder Nakaya            Mohan Maddur            Sathyanarayana            Sai Duraisingham</p> <p>Rafi Ahmed            Nicole Sullivan            Xiaojin Xu            Koichi Araki            Megan McCausland</p> <p><u>Dept. Pediatrics</u>            Miriam B. Vos            Ran Jin</p> <p><u>Dept. Neurology</u>            William Hu</p> <p><u>School of Public Health</u>            Tianwei Yu            Qi Long            Sandra Sato</p>		<p><b>Thank you!</b></p> <p><b>Vanderbilt University</b>            Milam Brantley</p> <p><b>Georgia State University</b>            Ruiyan Luo</p> <p><b>University of Colorado</b>            Adriana Weinberg            Myron Levin            Nancy Lang</p> <p><b>CHDS, Public Health Institute</b>            Barbara Cohn            Piera Cirillo            Nickilou Krigbaum            Lauren Zimmermann</p> <p><b>National Jewish</b>            Russ Bowler</p>		<p><b>La Jolla Institute for Allergy and Immunology</b>            Alex Sette            Bjoern Peters            John Sidney</p> <p><b>Merck Inc.</b>            Kalpit Vora            Sheri Dubey            Janet Sei            Joseph Antonello            Eberhard Durr            Kelly Collins</p> <p><b>UCSD</b>            Shankar Subramaniam            Thomas Hagen</p> <p><b>Dana Farber Cancer Institute</b>            Nick Haining</p>	
<p>Acknowledgement of research supports from US National Institute of Health (NIEHS P30 ES019116, 1U2CES026560-01, 1P50ES026071-01, NIAID HHSN 272201200031C, 2U19AI090023-06, HHSN272201300018, NIDDK 1R01DK107900-01), Department of Defense (HT9404-13-1-0030) and California Breast Cancer Research Program (21UB-8002).</p>							