Knowledge that will change your world

Analyzing data with Mummichog

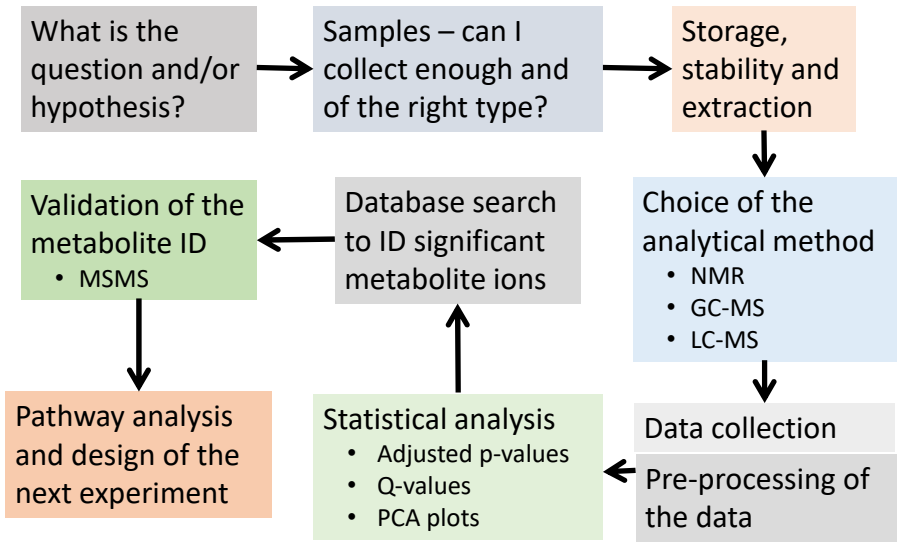
Stephen Barnes, PhD
University of Alabama at Birmingham
sbarnes@uab.edu

With acknowledgements to Shuzhao Li, PhD, Emory University

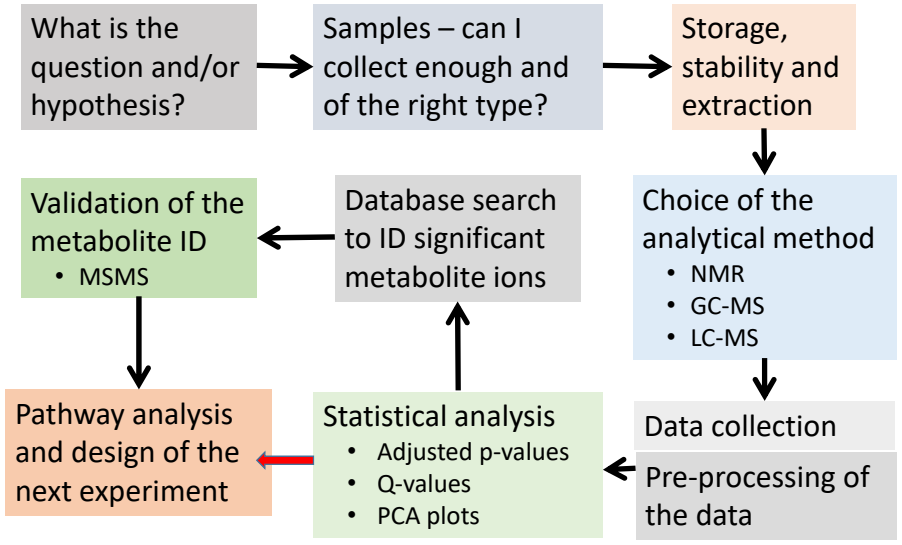
The biggest problem in metabolomics

- When a dataset has been processed to identify peaks and then retention time grouped, the resulting set of ions may exceed 3,000-4,000 (more if you use an FT-ICR instrument)
- The dataset is then subjected to statistical analysis and 300-400 ions pass criteria in mono- and multivariate statistics, causing rejection of the null hypothesis
- The significant ions are used to interrogate metabolite databases
- **Of these, less than 20% can be ascribed to known metabolites**

Metabolomics workflow



Metabolomics workflow



Crisis in -omics

- **In the paper by Prosser et al., the authors point out there is a serious issue of misannotation of the function of genes**
  - "*In silico sequence homology-based methods ….. are unable to identify the functions of novel gene sequences that have little to no homology with pre-existing database entries or may lead to the misannotation of gene products that share very high homology but catalyse fundamentally different reactions*."
  - "*the propagation of such misannotations is a serious and growing threat to the accuracy and reliability of genome and protein databases*."

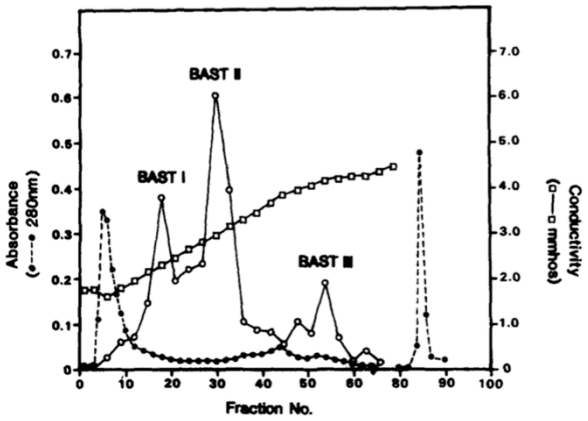
Prosser et al., EMBO Reports 2014

Role of metabolomics

- *"The metabolome can be perceived as the ultimate readout of the biochemical and physiological state of a cell*"
- (Using metabolomics) *new pathways and metabolites can be identified without the need for targeted genetic modification or recombinant protein studies, simplifying the workflow and allowing greater flexibility in the conditions and test organisms used*.

Prosser et al., EMBO Reports 2014

Why did this happen?



This is from a paper I published in the Journal of Lipid Research in 1989. The enzyme being purified sulfates bile acids. I purified and chemically sequenced BAST I and then others cloned it. BAST II and III have not been purified. cDNA cloning and sequencing took over in place of purification.

We need a way to understand relationships between metabolites

- The answer is the mummichog approach
- Mummichog is a fish that swims in groups

- **Mummichog is a software program that finds metabolites that "swim" together**

A talk given by Shuzhao Li

- **Available on the UAB Metabolomics Workshop 2015 website**

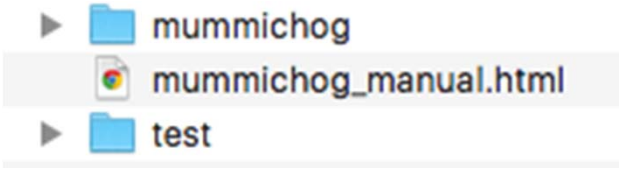
- http://www.uab.edu/proteomics/metabolomics/workshop/2015/Shuzhao_UAB_20150618_mummichog_v2.pdf
- http://www.uab.edu/proteomics/metabolomics/workshop/2015/videos/li_day4_1.html

Using mummichog

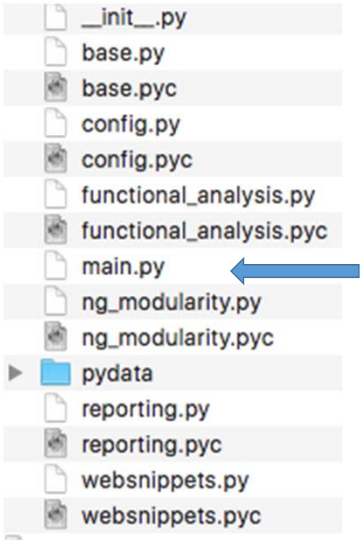
- **Two pieces of software are needed**
  - **Python and mummichog**
- **The recommended version of Python is Anaconda Python 2.7 (higher versions don't work)**
- **It is downloaded from** www.continuum.io/download
- **Unzip it – this can take a while since there are several hundred python scripts in the file**

Installing and running mummichog

- A new URL for mummichog will become available shortly
  - I have a mummichog.zip to distribute
  - Current version is mummichog-1.0.7
  - Unzipping it will create a mummichog-1.0.7 folder
  - Inside the mummichog-1.0.7 folder are the following:



Files in the mummichog folder



This is the script that is invoked

The test folder and preparing a data file for mummichog analysis

- The test folder contains testdata.txt for testing

m/z	retention_time	p-value	t-score	optional_id
304.2979	144.9853	1.0153771115e-10	14.7179316191	AEpos_304.2979_144
177.1024	64.70515	1.61647122234e-10	14.2666000207	AEpos_177.1024_64
345.0277	42.98387	1.71651483296e-10	-14.2091952724	AEpos_345.0277_42
491.0325	44.12778	1.83359804763e-10	-14.1463478332	AEpos_491.0325_44
258.0048	63.53025	2.16851438688e-10	-13.987636322	AEpos_258.0048_63
483.1205	337.3483	2.21510885538e-10	-13.9676335843	AEpos_483.1205_337
694.9937	43.82584	2.81091747637e-10	-13.7451720928	AEpos_694.9937_43
270.9767	101.3809	3.26786548614e-10	13.6060704804	AEpos_270.9767_101
371.6040	372.8965	3.53306845492e-10	-13.5344829919	AEpos_371.604_372
316.5773	327.6830	3.71195956728e-10	13.4893332694	AEpos_316.5773_327
451.0505	93.88615	4.03944158363e-10	-13.4123465751	AEpos_451.0505_93
257.0543	100.5867	4.08624036769e-10	-13.401886545	AEpos_257.0543_100
762.9787	42.77587	4.70637895081e-10	-13.2741409315	AEpos_762.9787_42
231.0422	297.2464	5.13599488249e-10	-13.1956772871	AEpos_231.0422_297
614.0797	449.5965	6.11088211834e-10	13.0407288286	AEpos_614.0797_449
213.0066	177.8462	6.79270965679e-10	12.9471714126	AEpos_213.0066_177
416.2122	374.8805	7.92384853978e-10	12.8119437544	AEpos_416.2122_374
310.0853	100.0514	8.54094386956e-10	-12.746529909	AEpos_310.0853_100
226.5313	93.51413	9.49577002679e-10	-12.6545669395	AEpos_226.5313_93
159.0492	70.53360	1.05661756669e-09	-12.5624353569	AEpos_159.0492_70
158.0440	300.0067	1.2234942889e-09	12.4368545377	AEpos_158.044_300
538.1246	295.1905	1.31061918452e-09	-12.3783014865	AEpos_538.1246_295
205.0280	115.5910	1.64760903079e-09	12.1851512666	AEpos_205.028_115
339.0594	163.8160	1.74282671318e-09	-12.138106566	AEpos_339.0594_163
131.0331	68.18234	1.99293957084e-09	12.026414889	AEpos_131.0331_68
110.0345	58.05023	2.4293029589e-09	11.8630366107	AEpos_110.0345_58

Creating the data.txt file

- **From the Metaboanalyst download, open the peak_normalized_rt_mz.csv file**

	A	B	C	D	E	F	G	H	I
1		mz	rt	negmode_ir1	negmode_ir2	negmode_ir3	negmode_nr1	negmode_nr2	negmode_nr3
2	50.34162/15	50.34162	15.39	0.148667117	0.100607907	-0.08691448	0.005970895	-0.113363467	-0.054967974
3	53.54365/15	53.54365	15.44	0.022027737	0.143231789	0.071187108	-0.05565963	-0.117805979	-0.062981026
4	59.01358/5.(	59.01358	5.06	0.153929379	0.12793021	0.071772935	-0.11987708	-0.013593153	-0.22016229
5	59.01556/5.9	59.01556	5.95	0.094315769	0.137733715	0.108763696	-0.0380309	-0.062353297	-0.240428986
6	60.9946/5.8	60.99460	5.84	-0.353241484	-0.273543206	-0.18831904	0.116821606	0.37861277	0.319669357
7	75.00986/5.:	75.00986	5.1	-0.223289076	-0.103273807	-0.12848957	-0.25194841	0.191429949	0.515570921
8	80.96603/13	80.96603	13.56	-0.140827325	0.220678017	-0.25577116	0.018484612	0.010205157	0.147230701
9	88.99233/5.8	88.99233	5.84	-0.970356397	-0.647689336	-0.37012505	0.493293763	0.541574803	0.95330222
10	103.00362/5	103.00362	5.95	-0.172418644	0.871172426	-0.13779109	-0.4181252	0.363037141	-0.505874633
11	111.00859/6	111.00859	6.29	-0.457141712	-0.049171919	-0.11643688	0.171289507	0.628700463	-0.177239456
12	111.08148/1	111.08148	18.96	0.118844825	0.090871156	0.090055311	-0.0978834	-0.093741782	-0.108146113
13	111.08296/1	111.08296	13.96	0.106312506	-0.069272297	0.117487788	0.001445126	-0.137906612	-0.018066512
14	113.06305/1	113.06305	13.14	0.023266594	-0.116860337	-0.08681714	-0.11764614	0.064830988	0.233226034
15	113.09528/2	113.09528	20.4	-0.077110215	-0.091804843	-0.10446635	0.168524173	0.100138037	0.004719196
16	115.0039/5.:	115.00390	5.13	-0.226158249	0.192501491	0.153166618	0.127266637	-0.027255101	-0.219521397
17	115.00442/1	115.00442	15.56	0.155786852	0.132913393	0.048803682	-0.12857043	-0.032083058	-0.176850438
18	115.00418/7	115.00418	7.14	-0.180128099	-0.153507242	-0.16926756	0.230282477	0.132074925	0.140545498
19	116.05105/1	116.05105	10.01	-0.022235747	-0.212586788	-0.09163591	0.249938935	0.102788783	-0.026269273
20	117.01983/7	117.01983	7.22	-0.402906482	-0.320670354	-0.17946288	0.408740692	0.209864727	0.284434297

Open a new .txt file and transfer data as follows

	A	B	C	D	E	F	G	H	I	J	K	L
1	mz	rt			p-value	t-score	negmode_ir1	negmode_ir2	negmode_ir3	negmode_nr1	negmode_nr2	negmo ir3
2	50.34162	15.39					0.148667117	0.100607907	-0.08691448	0.005970895	-0.113363467	-0.054 74
3	53.54365	15.44					0.022027737	0.143231789	0.071187108	-0.05565963	-0.117805979	-0.062981026
4	59.01358	5.06					0.153929379	0.12793021	0.071772935	-0.11987708	-0.013593153	-0.22016229
5	59.01556	5.95					0.094315769	0.137733715	0.108763696	-0.0380309	-0.062353297	-0.240428986
6	60.99460	5.84					-0.353241484	-0.273543206	-0.18831904	0.116821606	0.37861277	0.319669357
7	75.00986	5.1					-0.223289076	-0.103273807	-0.12848957	-0.25194841	0.191429949	0.515570921
8	80.96603	13.56					-0.140827325	0.220678017	-0.25577116	0.018484612	0.010205157	0.147230701
9	88.99233	5.84					-0.970356397	-0.647689336	-0.37012505	0.493293763	0.541574803	0.95330222
10	103.00362	5.95					-0.172418644	0.871172426	-0.13779109	-0.4181252	0.363037141	-0.505874633
11	111.00859	6.29					-0.457141712	-0.049171919	-0.11643688	0.171289507	0.628700463	-0.177239456
12	111.08148	18.96					0.118844825	0.090871156	0.090055311	-0.0978834	-0.093741782	-0.108146113
13	111.08296	13.96					0.106312506	-0.069272297	0.117487788	0.001445126	-0.137906612	-0.018066512
14	113.06305	13.14					0.023266594	-0.116860337	-0.08681714	-0.11764614	0.064830988	0.233226034
15	113.09528	20.4					-0.077110215	-0.091804843	-0.10446635	0.168524173	0.100138037	0.004719196
16	115.00390	5.13					-0.226158249	0.192501491	0.153166618	0.127266637	-0.027255101	-0.219521397
17	115.00442	15.56					0.155786852	0.132913393	0.048803682	-0.12857043	-0.032083058	-0.176850438
18	115.00418	7.14					-0.180128099	-0.153507242	-0.16926756	0.230282477	0.132074925	0.140545498
19	116.05105	10.01					-0.022235747	-0.212586788	-0.09163591	0.249938935	0.102788783	-0.026269273
20	117.01983	7.22					-0.402906482	-0.320670354	-0.17946288	0.408740692	0.209864727	0.284434297
21	117.05527	10.79					-0.142277766	-0.144967533	-0.07749922	0.198289952	0.119644388	0.046810179
22	119.03942	14.01					-0.198873362	-0.137175564	-0.14729987	0.18923308	0.136582356	0.157533356
23	120.04327	14.09					-0.087783401	-0.119310224	-0.07474603	0.081720708	0.112809512	0.087309433
24	121.02912	13.69					0.604543077	0.578681881	0.496325628	-0.5159732	-0.620695751	-0.542881637
25	121.02927	15.56					0.426805581	0.342092211	0.44900921	-0.37906593	-0.495862687	-0.342978383

P-value = ttest(G2:I2,J2:L2,2,2)

Creating the p-values

	A	B	C	D	E	F	G	H	I	J	K	L
1	mz	rt			p-value	t-score	negmode_ir1	negmode_ir2	negmode_ir3	negmode_nr1	negmode_nr2	negmode_nr3
2	50.34162	15.39			**0.24598588**		0.148667117	0.100607907	-0.08691448	0.005970895	-0.113363467	-0.054967974
3	53.54365	15.44			**0.01735838**		0.022027737	0.143231789	0.071187108	-0.05565963	-0.117805979	-0.062981026
4	59.01358	5.06			**0.02154243**		0.153929379	0.12793021	0.071772935	-0.11987708	-0.013593153	-0.22016229
5	59.01556	5.95			**0.02507588**		0.094315769	0.137733715	0.108763696	-0.0380309	-0.062353297	-0.240428986
6	60.99460	5.84			**0.00419233**		-0.353241484	-0.273543206	-0.18831904	0.116821606	0.37861277	0.319669357
7	75.00986	5.1			**0.24961235**		-0.223289076	-0.103273807	-0.12848957	-0.25194841	0.191429949	0.515570921
8	80.96603	13.56			**0.47865298**		-0.140827325	0.220678017	-0.25577116	0.018484612	0.010205157	0.147230701
9	88.99233	5.84			**0.00426648**		-0.970356397	-0.647689336	-0.37012505	0.493293763	0.541574803	0.95330222
10	103.00362	5.95			**0.44302065**		-0.172418644	0.871172426	-0.13779109	-0.4181252	0.363037141	-0.505874633
11	111.00859	6.29			**0.19270996**		-0.457141712	-0.049171919	-0.11643688	0.171289507	0.628700463	-0.177239456
12	111.08148	18.96			**4.3005E-05**		0.118844825	0.090871156	0.090055311	-0.0978834	-0.093741782	-0.108146113
13	111.08296	13.96			**0.23908669**		0.106312506	-0.069272297	0.117487788	0.001445126	-0.137906612	-0.018066512
14	113.06305	13.14			**0.33527969**		0.023266594	-0.116860337	-0.08681714	-0.11764614	0.064830988	0.233226034
15	113.09528	20.4			**0.01935478**		-0.077110215	-0.091804843	-0.10446635	0.168524173	0.100138037	0.004719196
16	115.00390	5.13			**0.65813988**		-0.226158249	0.192501491	0.153166618	0.127266637	-0.027255101	-0.219521397
17	115.00442	15.56			**0.01368777**		0.155786852	0.132913393	0.048803682	-0.12857043	-0.032083058	-0.176850438
18	115.00418	7.14			**0.00048966**		-0.180128099	-0.153507242	-0.16926756	0.230282477	0.132074925	0.140545498
19	116.05105	10.01			**0.08885324**		-0.022235747	-0.212586788	-0.09163591	0.249938935	0.102788783	-0.026269273
20	117.01983	7.22			**0.00231874**		-0.402906482	-0.320670354	-0.17946288	0.408740692	0.209864727	0.284434297
21	117.05527	10.79			**0.00768404**		-0.142277766	-0.144967533	-0.07749922	0.198289952	0.119644388	0.046810179
22	119.03942	14.01			**0.00019232**		-0.198873362	-0.137175564	-0.14729987	0.18923308	0.136582356	0.157533356
23	120.04327	14.09			**0.00032547**		-0.087783401	-0.119310224	-0.07474603	0.081720708	0.112809512	0.087309433
24	121.02912	13.69			**1.5875E-05**		0.604543077	0.578681881	0.496325628	-0.5159732	-0.620695751	-0.542881637
25	121.02927	15.56			**0.00013606**		0.426805581	0.342092211	0.44900921	-0.37906593	-0.495862687	-0.342978383

Now move each p-value over to column C as a number (not a function)

Moved the p-values into place

	A	B	C	D	E	F	G	H	I	J	K
1	mz	rt	p-value		t-score	negmode_ir1	negmode_ir2	negmode_ir3	negmode_nr1	negmode_nr2	negmode_nr3
2	50.34162	15.39	0.24598588			0.148667117	0.100607907	-0.08691448	0.005970895	-0.113363467	-0.054967974
3	53.54365	15.44	0.01735838			0.022027737	0.143231789	0.071187108	-0.05565963	-0.117805979	-0.062981026
4	59.01358	5.06	0.02154243			0.153929379	0.12793021	0.071772935	-0.11987708	-0.013593153	-0.22016229
5	59.01556	5.95	0.02507588			0.094315769	0.137733715	0.108763696	-0.0380309	-0.062353297	-0.240428986
6	60.99460	5.84	0.00419233			-0.353241484	-0.273543206	-0.18831904	0.116821606	0.37861277	0.319669357
7	75.00986	5.1	0.24961235			-0.223289076	-0.103273807	-0.12848957	-0.25194841	0.191429949	0.515570921
8	80.96603	13.56	0.47865298			-0.140827325	0.220678017	-0.25577116	0.018484612	0.010205157	0.147230701
9	88.99233	5.84	0.00426648			-0.970356397	-0.647689336	-0.37012505	0.493293763	0.541574803	0.95330222
10	103.00362	5.95	0.44302065			-0.172418644	0.871172426	-0.13779109	-0.4181252	0.363037141	-0.505874633
11	111.00859	6.29	0.19270996			-0.457141712	-0.049171919	-0.11643688	0.171289507	0.628700463	-0.177239456
12	111.08148	18.96	4.3005E-05			0.118844825	0.090871156	0.090055311	-0.0978834	-0.093741782	-0.108146113
13	111.08296	13.96	0.23908669			0.106312506	-0.069272297	0.117487788	0.001445126	-0.137906612	-0.018066512
14	113.06305	13.14	0.33527969			0.023266594	-0.116860337	-0.08681714	-0.11764614	0.064830988	0.233226034
15	113.09528	20.4	0.01935478			-0.077110215	-0.091804843	-0.10446635	0.168524173	0.100138037	0.004719196
16	115.00390	5.13	0.65813988			-0.226158249	0.192501491	0.153166618	0.127266637	-0.027255101	-0.219521397
17	115.00442	15.56	0.01368777			0.155786852	0.132913393	0.048803682	-0.12857043	-0.032083058	-0.176850438
18	115.00418	7.14	0.00048966			-0.180128099	-0.153507242	-0.16926756	0.230282477	0.132074925	0.140545498
19	116.05105	10.01	0.08885324			-0.022235747	-0.212586788	-0.09163591	0.249938935	0.102788783	-0.026269273
20	117.01983	7.22	0.00231874			-0.402906482	-0.320670354	-0.17946288	0.408740692	0.209864727	0.284434297
21	117.05527	10.79	0.00768404			-0.142277766	-0.144967533	-0.07749922	0.198289952	0.119644388	0.046810179
22	119.03942	14.01	0.00019232			-0.198873362	-0.137175564	-0.14729987	0.18923308	0.136582356	0.157533356
23	120.04327	14.09	0.00032547			-0.087783401	-0.119310224	-0.07474603	0.081720708	0.112809512	0.087309433
24	121.02912	13.69	1.5875E-05			0.604543077	0.578681881	0.496325628	-0.5159732	-0.620695751	-0.542881637
25	121.02927	15.56	0.00013606			0.426805581	0.342092211	0.44900921	-0.37906593	-0.495862687	-0.342978383

Now the t-score = (AVERAGE(F2:H2)-AVERAGE(I2:K2))/SQRT((STDEV(F2:H2)^2)/3+(STDEV(I2:K2)^2)/3)

Excel function to calculate t-score

- (AVERAGE(F2:H2)-AVERAGE(I2:K2))/SQRT((STDEV(F2:H2)^2)/3+(STDEV(I2:K2)^2)/3)
- The squared Standard Deviation for each group should be divided by the number of samples in the group. Adjust this for your experiment.

Completing file

	A	B	C	D	E
1	mz	rt	p-value	t-score	ID
2	50.34162	15.39	0.24598588	1.35809937	1
3	53.54365	15.44	0.01735838	3.91241657	2
4	59.01358	5.06	0.02154243	3.66199415	3
5	59.01556	5.95	0.02507588	3.49206298	4
6	60.99460	5.84	0.00419233	-5.875422	5
7	75.00986	5.1	0.24961235	-1.3457113	6
8	80.96603	13.56	0.47865298	-0.7805957	7
9	88.99233	5.84	0.00426648	-5.8472705	8
10	103.00362	5.95	0.44302065	0.8503815	9
11	111.00859	6.29	0.19270996	-1.5646701	10
1843	789.14983	13.52	0.00568379	-5.4022222	1842
1844	789.22709	14.97	2.4342E-05	22.2068568	1843
1845	790.15247	13.52	0.04920231	-2.7921825	1844
1846	790.21764	10.8	0.98894132	-0.0147456	1845
1847	790.22963	14.99	0.00119766	8.21329792	1846
1848	791.37662	17.74	0.00266658	6.64226723	1847
1849	795.32272	15.89	0.00349622	6.17363794	1848
1850	795.45448	21.29	0.11815283	-1.9847868	1849
1851	795.80482	15.85	0.03679348	3.08385241	1850
1852	799.23608	5.5	0.4724025	0.79253548	1851
1853	799.26262	10.44	0.00122052	-8.1726778	1852

```
Stephens-MacBook-Air-2:mummichog-1.0.5 stephenbarnes$ mummichog/main.py -c 0.05 -f test/d
iet_neg_test.txt -p 100 -m negative -o diet_neg_output

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

            o0                          ooooooooo
          oOO   OOOOO  ooooo        ooo oooo
     oOO   O        ooooo  oooooo oooo
    oooO             oooooo          oooo ooooo
        Oooo   o       OOOOOO   oooo   ooooooooo
            ooooo  oooo
                 o

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

mummichog version 1.0.5

Pygraphviz is not found. Skipping...
Started @ Sat Feb 27 22:18:10 2016

Loading metabolic network MFN_1.10.2...
cpds with MW: 2016
Got 964 significant features from 1846 references

Pathway Analysis...
query_set_size = 509 compounds
total_feature_num = 866 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 30 31 32 3
3 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
```

Command

```
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 30 31 32 3
3 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
Null distribution is estimated on 3055 random modules
User data yield 21 network modules

Got ActivityNetwork of 83 metabolites.

Annotation was written to
1456633089.87.diet_neg_output/tsv/_tentative_featurematch_diet_neg_output (.tsv and .xlsx
)
Pathway analysis report was written to
1456633089.87.diet_neg_output/tsv/mcg_pathwayanalysis_diet_neg_output (.tsv and .xlsx)

Modular analysis report was written to
1456633089.87.diet_neg_output/tsv/mcg_modularanalysis_diet_neg_output (.tsv and .xlsx)

Inspected network report was written to
1456633089.87.diet_neg_output/tsv/InspectedNodes_ActivityNetwork.tsv

Worksheet of top metabolites was written to
1456633089.87.diet_neg_output/tsv/mcg_metabolite_worksheet_diet_neg_output (.tsv and .xls
x)

Exporting top modules to 1456633089.87.diet_neg_output/sif/...

HTML report was written to
1456633089.87.diet_neg_output/result.html

Finished @ Sat Feb 27 22:21:23 2016
```

Mummichog options

```
-f, --infile: single file as input,
      containing all features with tab-delimited columns
      m/z, retention time, p-value, statistic score

-n, --network: network model to use (default human_mfn),
      [human, human_mfn, mouse, fly, yeast]

-o, --output: output file identification string (default 'mcgresult')
-k, --workdir: directory for all data files.
      Default is current directory.

-m, --mode: analytical mode of mass spec, [positive, negative, dpj].
      Default is dpj, a short version of positive.
-u, --instrument: [5, 10, 25, FTMS, ORBITRAP].
      Any integer is treated as ppm. Default is 10.
      Instrument specific functions may be implemented.

-p, --permutation: number of permutation to estimate null distributions.
      Default is 100.
-z,   --force_primary_ion: M+H[+] (M-H[-] for negative mode) must be
      present for a predicted metabolite, [True, False].
      Default is False.

-c, --cutoff: optional cutoff p-value in user supplied statistics,
      used to select significant list of features.
-e, --evidence: cutoff score for metabolite to be in activity network.
      Default is 3.
-d, --modeling: modeling permutation data, [no, gamma].
      Default is gamma.
```

Pathway output from mummichog

Top pathways

C = <0.05

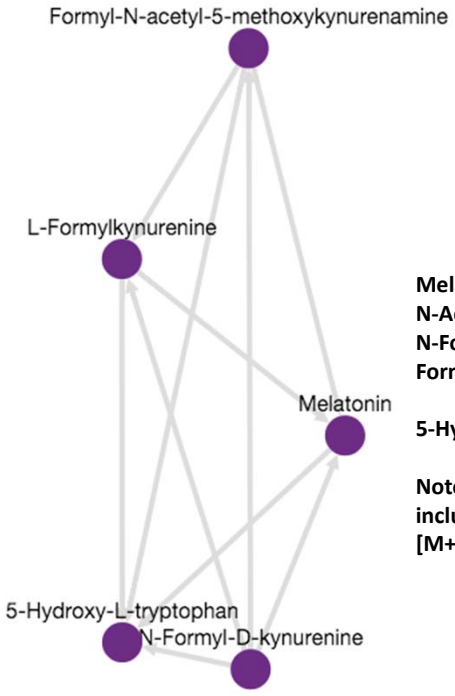
Pathways	overlap_size	pathway_size	p-value (raw)	p-value
Drug metabolism - cytochrome P450	26	30	0.00083	0.00216
Tryptophan metabolism	38	53	0.03192	0.00295
Linoleate metabolism	14	17	0.03611	0.0038
Porphyrin metabolism	10	13	0.14513	0.013
TCA cycle	10	13	0.14513	0.013
Tyrosine metabolism	43	69	0.31187	0.02118
Glycerophospholipid metabolism	13	19	0.26835	0.02671

C = <0.01

Pathways	overlap_size	pathway_size	p-value (raw)	p-value
Linoleate metabolism	14	17	0.00049	0.00122
Drug metabolism - other enzymes	6	8	0.05311	0.00368
TCA cycle	7	13	0.24237	0.01889
Drug metabolism - cytochrome P450	14	30	0.30801	0.0193
Glycerophospholipid metabolism	9	19	0.35269	0.03373

Limitation of digital pathways

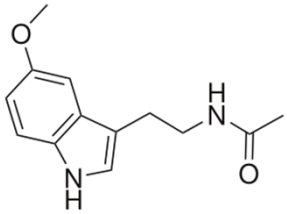
- **The traditional examination of pathways is intragenomic, i.e., within one organism**
- **In reality, life is intergenomic**
- **What you eat contains compounds that the body cannot make, e.g., vitamins, essential amino acids and lipids, and ?????**
- **Eaten food is exposed to the gut microbiome, either during initial ingestion (mostly in the small intestine) or after biliary excretion of phase II metabolites (now in the large intestine)**
- **The overall intergenomic pathways are not present in databases**
- **Better to look for chemical relationships (modules)**



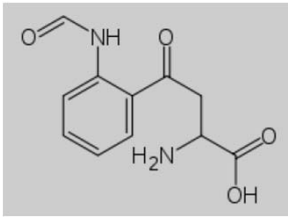
A module from the diet_neg data

	MW
Melatonin	232.1212
N-Acetyl-5-methoxytryptamine	232.1212
N-Formyl-D-kynurenine	236.0797
Formyl-N-acetyl-5-methoxykynurenamine	264.1110
5-Hydroxy-L-tryptophan	220.0848

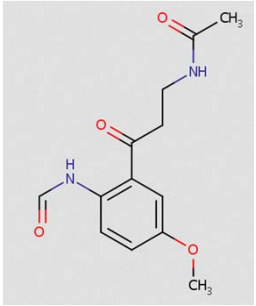
Note that the molecular ions of these metabolites include [M-H]-, [M(13C)-H]-, [M+ACN-H]-, [M+formate-H]-, [M+Na-H]-, [M+K-H]-



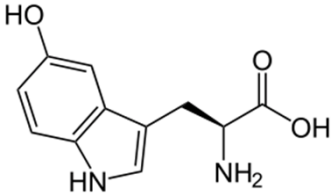
Melatonin
N-Acetyl-5-methoxytryptamine
$C_{13}H_{16}N_2O_2$



N-Formyl-D-kynurenine
$C_{11}H_{12}N_2O_4$



Formyl-N-acetyl-5-methoxykynurenamine
$C_{13}H_{16}N_2O_4$



5-hydroxy-tryptophan
$C_{11}H_{12}N_2O_3$

Mummichog in XCMSonline
It's called Connections

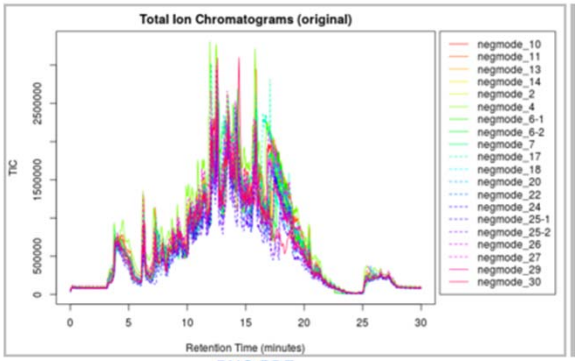
View Results Table

View Interactive Cloud Plot

View Interactive Heatmap

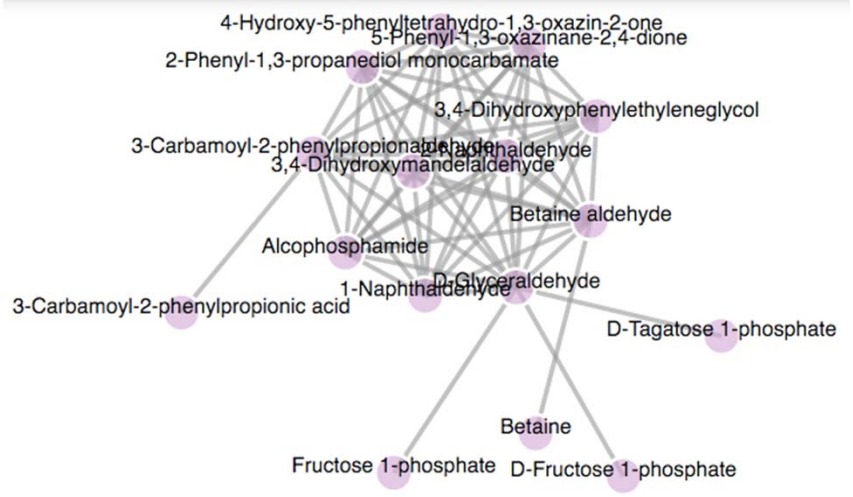
View iPCA

Connections

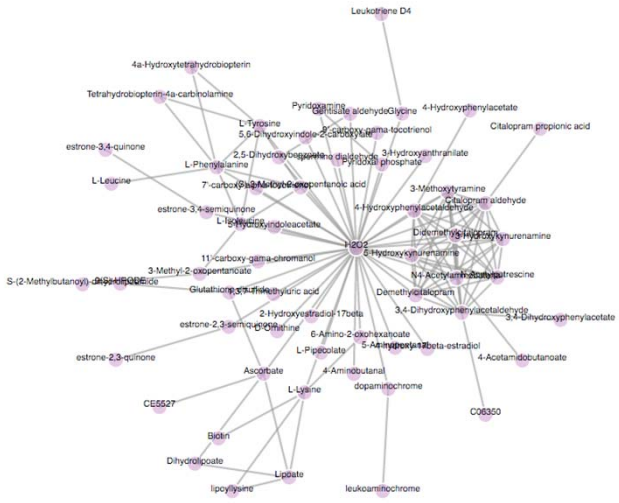


PNG PDF

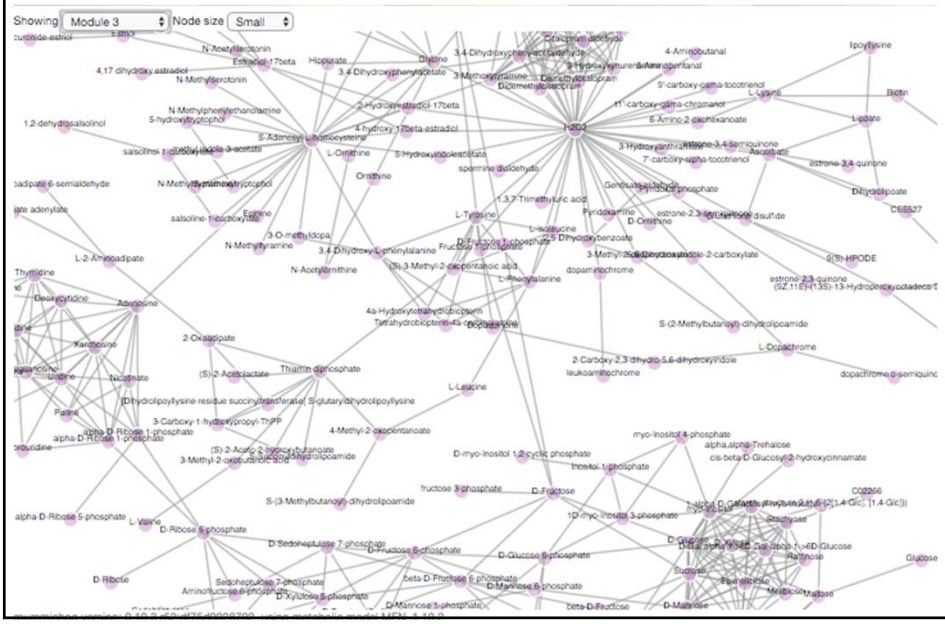
Mummichog – module 1



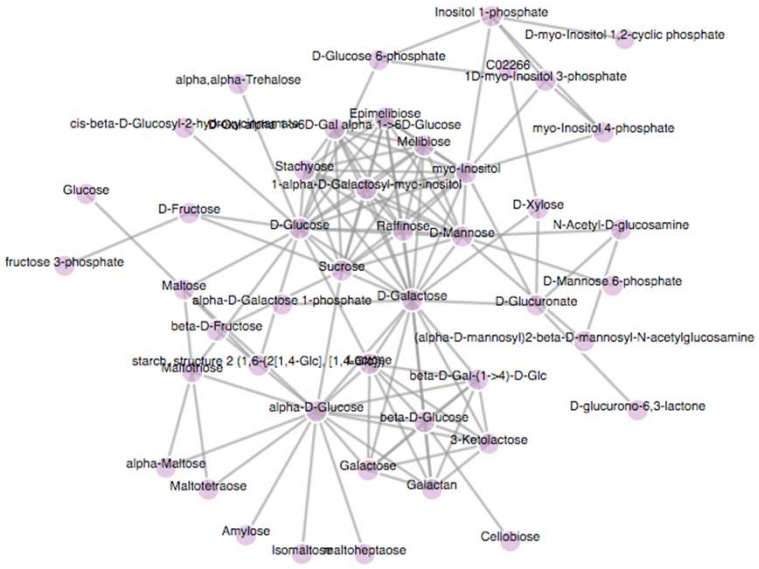
Mummichog – module 2



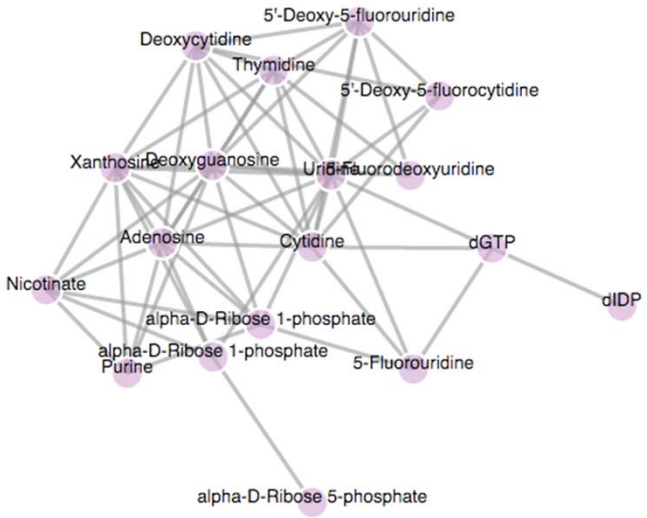
Mummichog – module 3



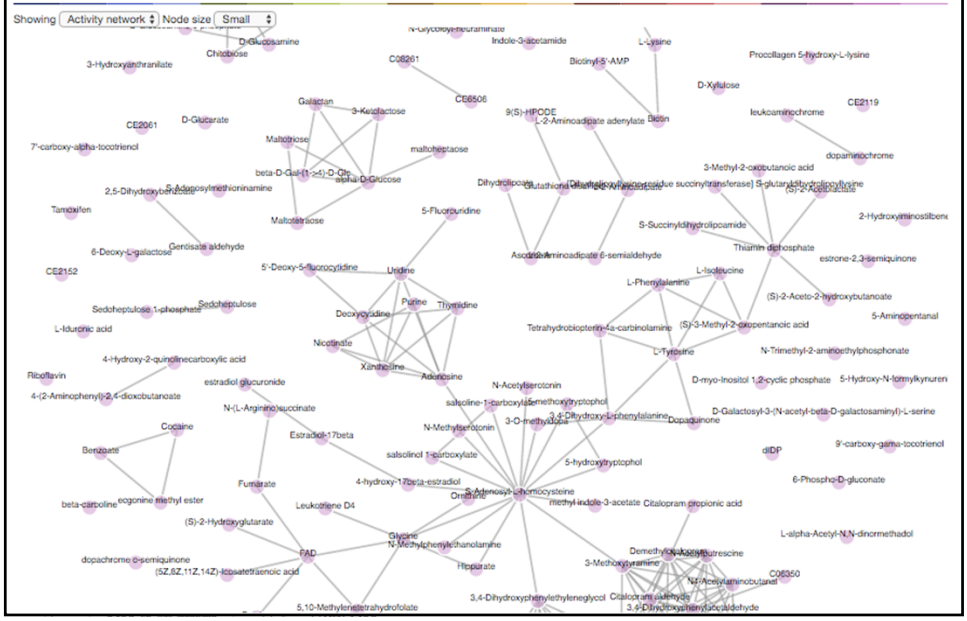
Mummichog – module 4



Mummichog – module 5



Mummichog – the network



The pathways by mummichog

Pathways	overlap_size	pathway_size	p-value (raw)	p-value
Starch and Sucrose Metabolism	13	16	0.0007	0.00049
Drug metabolism - cytochrome P450	22	35	0.00355	0.00051
Linoleate metabolism	12	16	0.00385	0.00052
Sialic acid metabolism	18	29	0.01	0.00056
N-Glycan Degradation	6	7	0.01671	0.00081
Heparan sulfate degradation	5	6	0.03723	0.00131
Chondroitin sulfate degradation	5	6	0.03723	0.00131
Lysine metabolism	13	24	0.09777	0.00147
Hexose phosphorylation	10	18	0.11895	0.00202
Alkaloid biosynthesis II	4	5	0.08095	0.003
Galactose metabolism	17	36	0.20466	0.00315
Vitamin H (biotin) metabolism	3	3	0.06022	0.00378
Vitamin B2 (riboflavin) metabolism	3	3	0.06022	0.00378
Keratan sulfate degradation	5	8	0.16146	0.00506
Urea cycle/amino group metabolism	16	36	0.31442	0.00711
Ascorbate (Vitamin C) and Aldarate Metabolism	10	21	0.28217	0.00731
Tryptophan metabolism	25	59	0.35394	0.00775
Lipoate metabolism	3	4	0.17026	0.01116
Valine, leucine and isoleucine degradation	11	25	0.38319	0.01371
Fatty Acid Metabolism	4	7	0.27487	0.01563
Omega-6 fatty acid metabolism	2	2	0.15385	0.02459
Saturated fatty acids beta-oxidation	2	2	0.15385	0.02459
Fatty acid oxidation	2	2	0.15385	0.02459
Butanoate metabolism	9	22	0.51743	0.03555
Vitamin B9 (folate) metabolism	6	14	0.49095	0.03977
Aminosugars metabolism	11	28	0.5707	0.04378