Knowledge that will change your world

Analyzing data with Mummichog

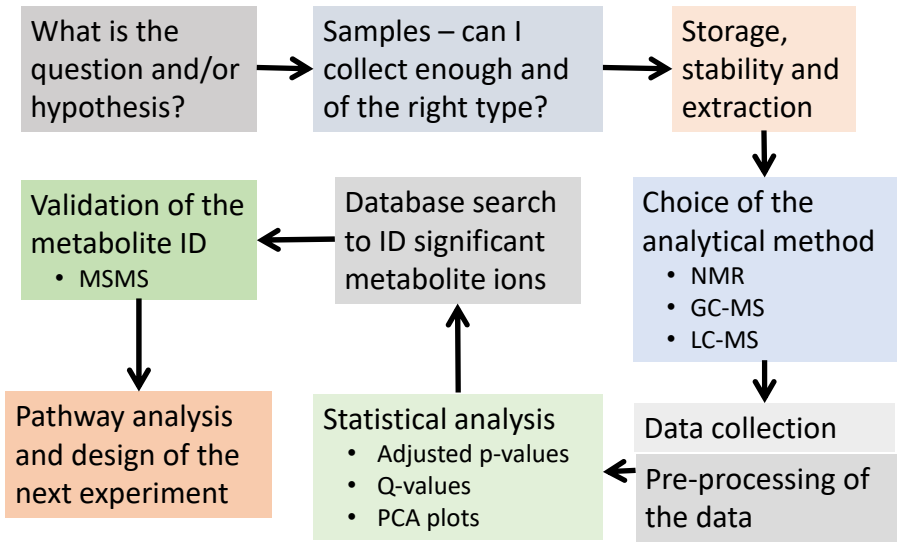
Stephen Barnes, PhD
University of Alabama at Birmingham
sbarnes@uab.edu

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



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*."

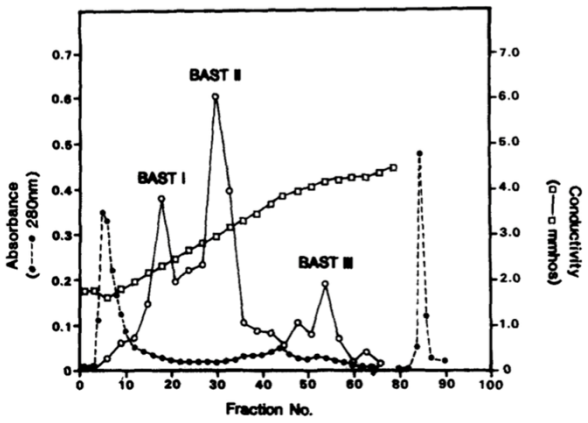
Prosser et al., EMBO Reports 2014

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 2017 website**
- https://www.uab.edu/proteomics/metabolomics/workshop/2017/day4/32-SLi%20-%20Pathway%20and%20Network%20Analysis%20for%20Metabolomics-Mummichog.pdf
- https://www.uab.edu/proteomics/metabolomics/workshop/2017/videos/li_day4.html

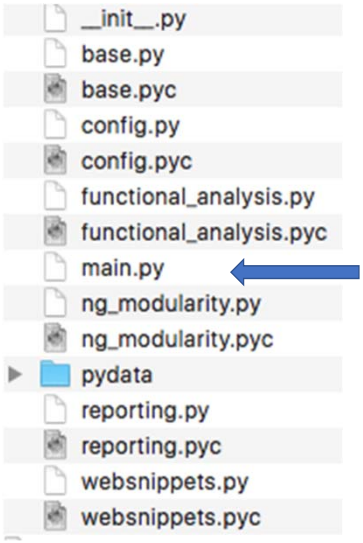
Using mummichog

- **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

- The URL for mummichog is http://mummichog.org/
  - Download the .zip file for mummichog-1.0.9
  - Unzipping it will create a mummichog-1.0.9 folder
  - Inside the mummichog-1.0.9 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	J	K	L	M	N	O
1		mz	rt	neg_c1	neg_c2	neg_c3	neg_c4	neg_c5	neg_c6	neg_g1	neg_g2	neg_g3	neg_g4	neg_g5	neg_g6
2	74.02843/14	74.02843	14.33	0.12664495	-0.0303119	0.03509817	0.06467715	-0.0575027	-0.0636304	-0.0006229	-0.0462304	0.14253023	-0.1355194	0.02930268	-0.0644355
3	74.02952/17	74.02952	17.5	0.01547948	-0.0270041	-0.018878	0.06646159	-0.0321898	0.34679445	-0.0438417	0.01097769	-0.0051323	-0.1393008	-0.0954005	-0.0779661
4	76.02753/16	76.02753	16.21	0.04629944	-0.0099944	-0.0269583	0.02469518	-0.0630185	0.03956184	-0.0297107	0.06570166	0.00349874	-0.1118317	0.03818951	0.02356723
5	79.96091/18	79.96091	18.46	-0.0432693	-0.078762	0.23218973	0.01667164	0.15731869	0.001864	-0.039098	-0.0685515	-0.0179537	-0.0964785	-0.0523582	-0.0115729
6	79.96084/17	79.96084	17.38	-0.1361527	-0.1148119	-0.0651033	-0.1374084	-0.1126426	-0.1508236	0.27096547	-0.1045456	0.12626328	-0.1567398	0.39634765	0.18465152
7	79.96117/15	79.96117	15.16	0.22992768	-0.0969451	-0.0291716	0.07885047	-0.0226638	-0.0312488	-0.0959724	0.07760922	0.0785857	-0.1497007	0.04781246	-0.0870831
8	92.05243/13	92.05243	13.21	-0.0102274	0.03889331	0.1556366	-0.0472297	0.14594445	-0.1250232	-0.0760256	-0.0035282	-0.058765	-0.0281401	-0.0492668	0.05773169
9	92.05353/15	92.05353	15.86	0.0525857	0.0082471	0.12125539	-0.0353331	0.0415483	-0.0997245	-0.0372895	0.01300243	0.01779581	-0.0660077	-0.0558565	0.03977657
10	105.02165/5	105.02165	5.19	-0.1162267	-0.1137777	0.28308714	-0.1465759	0.22467928	-0.0808206	-0.2041561	0.24300149	0.04164454	-0.1093876	-0.1368178	0.11534994
11	106.0684/17	106.0684	17.26	0.04214713	0.0591476	-0.0011793	0.04890675	-0.0100977	0.02902541	-0.1005581	-0.0246235	0.03841387	-0.0580234	-0.0427857	0.01962693
12	107.05242/1	107.05242	15.11	0.48632249	-0.3621466	-0.1106521	0.21518693	-0.0852372	-0.0254121	-0.1577704	0.01434695	0.15316498	-0.2382998	0.14654682	-0.03605
13	108.05603/1	108.05603	15.12	0.11817402	-0.1075493	-0.0093153	0.09802415	-0.0523433	-0.0262978	-0.0413214	-0.0028402	0.07604416	-0.044924	0.00687064	-0.0145217
14	111.08192/1	111.08192	14.93	-0.07111	0.03946878	0.14456981	-0.0531731	0.0034443	-0.0521122	-0.0564861	0.16605359	-0.0605724	-0.037144	-0.0333031	0.01036446
15	114.05963/1	114.05963	18.77	0.08778912	-0.0200643	0.00957514	-0.0411797	0.05286783	-0.1563856	-0.0295392	0.06590721	-0.0019699	0.04754588	-0.0437751	0.02922858
16	114.06137/1	114.06137	16.39	0.06881548	-0.0269546	0.04287468	-0.0153606	0.07861427	-0.0863962	-0.0055338	-0.0670997	0.00326145	-0.0235086	0.01865675	0.01263087
17	115.04191/9	115.04191	9.98	0.09041689	0.08253192	0.05081909	0.02218947	-0.0099546	-0.0321693	-0.0020963	0.05689716	-0.0534827	-0.1352419	-0.058544	-0.0113657
18	116.0382/20	116.0382	20.88	-0.0226216	-0.0281222	-0.0341195	0.03951147	-0.0037243	0.0862031	0.00772989	0.09456858	0.05044538	-0.0943176	-0.0598532	-0.0357
19	116.04406/1	116.04406	16.45	-0.0994239	-0.0430451	-0.0268517	0.10466009	-0.0007153	0.33732245	-0.0819278	0.2416521	-0.0158087	-0.2436366	-0.0837632	-0.0884624
20	116.0533/18	116.0533	18.02	0.10582384	0.17466492	0.10994354	0.13958668	0.0333693	-0.132664	-0.0228965	-0.0645426	0.02069976	-0.2795106	-0.0081542	-0.0763201
21	118.05356/1	118.05356	11.16	0.0230979	-0.0104565	0.03052541	0.00560681	0.04506756	0.08220221	-0.0066199	0.00813188	-0.066882	-0.0376179	-0.0733911	0.00033564
22	119.05252/1	119.05252	12.83	0.15193627	-0.0179565	0.09974712	-0.0061033	0.05409237	-0.2017757	-0.1556025	-0.0585596	0.11971313	-0.1021928	0.14515855	-0.0284571
23	120.01586/1	120.01586	17.66	0.06824196	0.06919243	0.01502251	0.04292943	0.03741188	0.09189158	-0.0482252	-0.0145089	-0.0747095	-0.1173351	-0.0280702	-0.0418409
24	120.04371/1	120.04371	15.87	0.08711151	0.00683223	0.2863333	-0.0844461	0.14831342	-0.2327929	-0.1175346	0.12278738	-0.0171705	-0.201191	-0.1030023	0.1047596
25	121.03147/1	121.03147	14.47	-0.0063402	0.09192284	0.37937981	0.09735868	0.00845184	-0.2792755	-0.1582626	0.34349423	0.00839008	-0.3433544	-0.2148383	0.07307349

Open a new Excel file and transfer data as follows

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	mz	rt							neg_c1	neg_c2	neg_c3	neg_c4	neg_c5	neg_c6
2	74.02843	14.33							0.12664495	-0.0303119	0.03509817	0.06467715	-0.0575027	-0.0636304
3	74.02952	17.5							0.01547948	-0.0270041	-0.018878	0.06646159	-0.0321898	0.34679445
4	76.02753	16.21							0.04629944	-0.0099944	-0.0269583	0.02469518	-0.0630185	0.03956184
5	79.96091	18.46							-0.0432693	-0.078762	0.23218973	0.01667164	0.15731869	0.001864
6	79.96084	17.38							-0.1361527	-0.1148119	-0.0651033	-0.1374084	-0.1126426	-0.1508236
7	79.96117	15.16							0.22992768	-0.0969451	-0.0291716	0.07885047	-0.0226638	-0.0312488
8	92.05243	13.21							-0.0102274	0.03889331	0.1556366	-0.0472297	0.14594445	-0.1250232
9	92.05353	15.86							0.0525857	0.0082471	0.12125539	-0.0353331	0.0415483	-0.0997245
10	105.02165	5.19							-0.1162267	-0.1137777	0.28308714	-0.1465759	0.22467928	-0.0808206
11	106.0684	17.26							0.04214713	0.0591476	-0.0011793	0.04890675	-0.0100977	0.02902541
12	107.05242	15.11							0.48632249	-0.3621466	-0.1106521	0.21518693	-0.0852372	-0.0254121
13	108.05603	15.12							0.11817402	-0.1075493	-0.0093153	0.09802415	-0.0523433	-0.0262978
14	111.08192	14.93							-0.07111	0.03946878	0.14456981	-0.0531731	0.0034443	-0.0521122
15	114.05963	18.77							0.08778912	-0.0200643	0.00957514	-0.0411797	0.05286783	-0.1563856
16	114.06137	16.39							0.06881548	-0.0269546	0.04287468	-0.0153606	0.07861427	-0.0863962
17	115.04191	9.98							0.09041689	0.08253192	0.05081909	0.02218947	-0.0099546	-0.0321693
18	116.0382	20.88							-0.0226216	-0.0281222	-0.0341195	0.03951147	-0.0037243	0.0862031
19	116.04406	16.45							-0.0994239	-0.0430451	-0.0268517	0.10466009	-0.0007153	0.33732245
20	116.0533	18.02							0.10582384	0.17466492	0.10994354	0.13958668	0.0333693	-0.132664
21	118.05356	11.16							0.0230979	-0.0104565	0.03052541	0.00560681	0.04506756	0.08220221
22	119.05252	12.83							0.15193627	-0.0179565	0.09974712	-0.0061033	0.05409237	-0.2017757
23	120.01586	17.66							0.06824196	0.06919243	0.01502251	0.04292943	0.03741188	0.09189158
24	120.04371	15.87							0.08711151	0.00683223	0.2863333	-0.0844461	0.14831342	-0.2327929
25	121.03147	14.47							-0.0063402	0.09192284	0.37937981	0.09735868	0.00845184	-0.2792755

Columns b and c in the Excel file

Working space

Data from Excel file

Creating the p-values

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	mz	rt							neg_c1	neg_c2	neg_c3	neg_c4	neg_c5	neg_c6
2	74.02843	14.33		=ttest(i2:n2,o2:t2,2,2)					0.12664495	-0.0303119	0.03509817	0.06467715	-0.0575027	-0.0636304
3	74.02952	17.5							0.01547948	-0.0270041	-0.018878	0.06646159	-0.0321898	0.34679445
4	76.02753	16.21							0.04629944	-0.0099944	-0.0269583	0.02469518	-0.0630185	0.03956184
5	79.96091	18.46							-0.0432693	-0.078762	0.23218973	0.01667164	0.15731869	0.001864
6	79.96084	17.38							-0.1361527	-0.1148119	-0.0651033	-0.1374084	-0.1126426	-0.1508236
7	79.96117	15.16							0.22992768	-0.0969451	-0.0291716	0.07885047	-0.0226638	-0.0312488
8	92.05243	13.21							-0.0102274	0.03889331	0.1556366	-0.0472297	0.14594445	-0.1250232
9	92.05353	15.86							0.0525857	0.0082471	0.12125539	-0.0353331	0.0415483	-0.0997245
10	105.02165	5.19							-0.1162267	-0.1137777	0.28308714	-0.1465759	0.22467928	-0.0808206
11	106.0684	17.26							0.04214713	0.0591476	-0.0011793	0.04890675	-0.0100977	0.02902541
12	107.05242	15.11							0.48632249	-0.3621466	-0.1106521	0.21518693	-0.0852372	-0.0254121
13	108.05603	15.12							0.11817402	-0.1075493	-0.0093153	0.09802415	-0.0523433	-0.0262978
14	111.08192	14.93							-0.07111	0.03946878	0.14456981	-0.0531731	0.0034443	-0.0521122
15	114.05963	18.77							0.08778912	-0.0200643	0.00957514	-0.0411797	0.05286783	-0.1563856
16	114.06137	16.39							0.06881548	-0.0269546	0.04287468	-0.0153606	0.07861427	-0.0863962
17	115.04191	9.98							0.09041689	0.08253192	0.05081909	0.02218947	-0.0099546	-0.0321693
18	116.0382	20.88							-0.0226216	-0.0281222	-0.0341195	0.03951147	-0.0037243	0.0862031
19	116.04406	16.45							-0.0994239	-0.0430451	-0.0268517	0.10466009	-0.0007153	0.33732245
20	116.0533	18.02							0.10582384	0.17466492	0.10994354	0.13958668	0.0333693	-0.132664
21	118.05356	11.16							0.0230979	-0.0104565	0.03052541	0.00560681	0.04506756	0.08220221
22	119.05252	12.83							0.15193627	-0.0179565	0.09974712	-0.0061033	0.05409237	-0.2017757
23	120.01586	17.66							0.06824196	0.06919243	0.01502251	0.04292943	0.03741188	0.09189158
24	120.04371	15.87							0.08711151	0.00683223	0.2863333	-0.0844461	0.14831342	-0.2327929
25	121.03147	14.47							-0.0063402	0.09192284	0.37937981	0.09735868	0.00845184	-0.2792755

In column D, enter the ttest function, where i2:n2 are the control samples and o2:t2 the genistein samples. The "2,2" represents two-tailed and unpaired options. Copy this formula down the entire column. Then copy (control or command c) the column and paste into column C (as a value). Then delete column D.

Move the p-values into place

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	mz	rt							neg_c1	neg_c2	neg_c3	neg_c4	neg_c5	neg_c6
2	74.02843	14.33	0.62498766						0.12664495	-0.0303119	0.03509817	0.06467715	-0.0575027	-0.0636304
3	74.02952	17.5	0.09755203						0.01547948	-0.0270041	-0.018878	0.06646159	-0.0321898	0.34679445
4	76.02753	16.21	0.91160426						0.04629944	-0.0099944	-0.0269583	0.02469518	-0.0630185	0.03956184
5	79.96091	18.46	0.09191752						-0.0432693	-0.078762	0.23218973	0.01667164	0.15731869	0.001864
6	79.96084	17.38	0.02230474						-0.1361527	-0.1148119	-0.0651033	-0.1374084	-0.1126426	-0.1508236
7	79.96117	15.16	0.51106098						0.22992768	-0.0969451	-0.0291716	0.07885047	-0.0226638	-0.0312488
8	92.05243	13.21	0.30913979						-0.0102274	0.03889331	0.1556366	-0.0472297	0.14594445	-0.1250232
9	92.05353	15.86	0.43063321						0.0525857	0.0082471	0.12125539	-0.0353331	0.0415483	-0.0997245
10	105.02165	5.19	0.87614126						-0.1162267	-0.1137777	0.28308714	-0.1465759	0.22467928	-0.0808206
11	106.0684	17.26	0.04045387						0.04214713	0.0591476	-0.0011793	0.04890675	-0.0100977	0.02902541
12	107.05242	15.11	0.77873557						0.48632249	-0.3621466	-0.1106521	0.21518693	-0.0852372	-0.0254121
13	108.05603	15.12	0.8669316						0.11817402	-0.1075493	-0.0093153	0.09802415	-0.0523433	-0.0262978
14	111.08192	14.93	0.9405459						-0.07111	0.03946878	0.14456981	-0.0531731	0.0034443	-0.0521122
15	114.05963	18.77	0.57836225						0.08778912	-0.0200643	0.00957514	-0.0411797	0.05286783	-0.1563856
16	114.06137	16.39	0.49754071						0.06881548	-0.0269546	0.04287468	-0.0153606	0.07861427	-0.0863962
17	115.04191	9.98	0.06870494						0.09041689	0.08253192	0.05081909	0.02218947	-0.0099546	-0.0321693
18	116.0382	20.88	0.72944876						-0.0226216	-0.0281222	-0.0341195	0.03951147	-0.0037243	0.0862031
19	116.04406	16.45	0.34605835						-0.0994239	-0.0430451	-0.0268517	0.10466009	-0.0007153	0.33732245
20	116.0533	18.02	0.04602802						0.10582384	0.17466492	0.10994354	0.13958668	0.0333693	-0.132664
21	118.05356	11.16	0.01325537						0.0230979	-0.0104565	0.03052541	0.00560681	0.04506756	0.08220221
22	119.05252	12.83	0.71343749						0.15193627	-0.0179565	0.09974712	-0.0061033	0.05409237	-0.2017757
23	120.01586	17.66	0.00018591						0.06824196	0.06919243	0.01502251	0.04292943	0.03741188	0.09189158
24	120.04371	15.87	0.45750166						0.08711151	0.00683223	0.2863333	-0.0844461	0.14831342	-0.2327929
25	121.03147	14.47	0.47915684						-0.0063402	0.09192284	0.37937981	0.09735868	0.00845184	-0.2792755

Excel function to calculate t-score

- (AVERAGE(F2:H2)-AVERAGE(I2:K2))/SQRT((STDEV(F2:H2)^2)/6+(STDEV(I2:K2)^2)/6)
- The squared Standard Deviation for each group should be divided by the number of samples in the group. Adjust this as needed for other experiments.
- It's easy to make mistakes, so I calculate the M1-M2 first, then do the pooled variance

Calculating difference in means

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	mz	rt	p-value		Delta				neg_c1	neg_c2	neg_c3	neg_c4	neg_c5	neg_c6
2	74.02843	14.33	0.62498766		=average(I2:n2)-average(o2:t2)				0.12664495	-0.0303119	0.03509817	0.06467715	-0.0575027	-0.0636304
3	74.02952	17.5	0.09755203						0.01547948	-0.0270041	-0.018878	0.06646159	-0.0321898	0.34679445
4	76.02753	16.21	0.91160426						0.04629944	-0.0099944	-0.0269583	0.02469518	-0.0630185	0.03956184
5	79.96091	18.46	0.09191752						-0.0432693	-0.078762	0.23218973	0.01667164	0.15731869	0.001864
6	79.96084	17.38	0.02230474						-0.1361527	-0.1148119	-0.0651033	-0.1374084	-0.1126426	-0.1508236
7	79.96117	15.16	0.51106098						0.22992768	-0.0969451	-0.0291716	0.07885047	-0.0226638	-0.0312488
8	92.05243	13.21	0.30913979						-0.0102274	0.03889331	0.1556366	-0.0472297	0.14594445	-0.1250232
9	92.05353	15.86	0.43063321						0.0525857	0.0082471	0.12125539	-0.0353331	0.0415483	-0.0997245
10	105.02165	5.19	0.87614126						-0.1162267	-0.1137777	0.28308714	-0.1465759	0.22467928	-0.0808206
11	106.0684	17.26	0.04045387						0.04214713	0.0591476	-0.0011793	0.04890675	-0.0100977	0.02902541
12	107.05242	15.11	0.77873557						0.48632249	-0.3621466	-0.1106521	0.21518693	-0.0852372	-0.0254121
13	108.05603	15.12	0.8669316						0.11817402	-0.1075493	-0.0093153	0.09802415	-0.0523433	-0.0262978
14	111.08192	14.93	0.9405459						-0.07111	0.03946878	0.14456981	-0.0531731	0.0034443	-0.0521122
15	114.05963	18.77	0.57836225						0.08778912	-0.0200643	0.00957514	-0.0411797	0.05286783	-0.1563856
16	114.06137	16.39	0.49754071						0.06881548	-0.0269546	0.04287468	-0.0153606	0.07861427	-0.0863962
17	115.04191	9.98	0.06870494						0.09041689	0.08253192	0.05081909	0.02218947	-0.0099546	-0.0321693
18	116.0382	20.88	0.72944876						-0.0226216	-0.0281222	-0.0341195	0.03951147	-0.0037243	0.0862031
19	116.04406	16.45	0.34605835						-0.0994239	-0.0430451	-0.0268517	0.10466009	-0.0007153	0.33732245
20	116.0533	18.02	0.04602802						0.10582384	0.17466492	0.10994354	0.13958668	0.0333693	-0.132664
21	118.05356	11.16	0.01325537						0.0230979	-0.0104565	0.03052541	0.00560681	0.04506756	0.08220221
22	119.05252	12.83	0.71343749						0.15193627	-0.0179565	0.09974712	-0.0061033	0.05409237	-0.2017757
23	120.01586	17.66	0.00018591						0.06824196	0.06919243	0.01502251	0.04292943	0.03741188	0.09189158
24	120.04371	15.87	0.45750166						0.08711151	0.00683223	0.2863333	-0.0844461	0.14831342	-0.2327929
25	121.03147	14.47	0.47915684						-0.0063402	0.09192284	0.37937981	0.09735868	0.00845184	-0.2792755

Copy this formula down the entire column.

Calculating pooled variance

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	mz	rt	p-value		Delta	SD1	SD2	SQRT	neg_c1	neg_c2	neg_c3	neg_c4	neg_c5	neg_c6
2	74.02843	14.33	0.62498766		0.02499175	=((STDEV(i2:n2))^2/6)			0.12664495	-0.0303119	0.03509817	0.06467715	-0.0575027	-0.0636304
3	74.02952	17.5	0.09755203		0.1168879				0.01547948	-0.0270041	-0.018878	0.06646159	-0.0321898	0.34679445
4	76.02753	16.21	0.91160426		0.00352843				0.04629944	-0.0099944	-0.0269583	0.02469518	-0.0630185	0.03956184
5	79.96091	18.46	0.09191752		0.09533758				-0.0432693	-0.078762	0.23218973	0.01667164	0.15731869	0.001864
6	79.96084	17.38	0.02230474		-0.2389808				-0.1361527	-0.1148119	-0.0651033	-0.1374084	-0.1126426	-0.1508236
7	79.96117	15.16	0.51106098		0.0429163				0.22992768	-0.0969451	-0.0291716	0.07885047	-0.0226638	-0.0312488
8	92.05243	13.21	0.30913979		0.05266469				-0.0102274	0.03889331	0.1556366	-0.0472297	0.14594445	-0.1250232
9	92.05353	15.86	0.43063321		0.02952629				0.0525857	0.0082471	0.12125539	-0.0353331	0.0415483	-0.0997245
10	105.02165	5.19	0.87614126		0.01678849				-0.1162267	-0.1137777	0.28308714	-0.1465759	0.22467928	-0.0808206
11	106.0684	17.26	0.04045387		0.05598331				0.04214713	0.0591476	-0.0011793	0.04890675	-0.0100977	0.02902541
12	107.05242	15.11	0.77873557		0.03935382				0.48632249	-0.3621466	-0.1106521	0.21518693	-0.0852372	-0.0254121
13	108.05603	15.12	0.8669316		0.00689749				0.11817402	-0.1075493	-0.0093153	0.09802415	-0.0523433	-0.0262978
14	111.08192	14.93	0.9405459		0.00369587				-0.07111	0.03946878	0.14456981	-0.0531731	0.0034443	-0.0521122
15	114.05963	18.77	0.57836225		-0.0224658				0.08778912	-0.0200643	0.00957514	-0.0411797	0.05286783	-0.1563856
16	114.06137	16.39	0.49754071		0.02053099				0.06881548	-0.0269546	0.04287468	-0.0153606	0.07861427	-0.0863962
17	115.04191	9.98	0.06870494		0.0679445				0.09041689	0.08253192	0.05081909	0.02218947	-0.0099546	-0.0321693
18	116.0382	20.88	0.72944876		0.01237564				-0.0226216	-0.0281222	-0.0341195	0.03951147	-0.0037243	0.0862031
19	116.04406	16.45	0.34605835		0.09064885				-0.0994239	-0.0430451	-0.0268517	0.10466009	-0.0007153	0.33732245
20	116.0533	18.02	0.04602802		0.14357477				0.10582384	0.17466492	0.10994354	0.13958668	0.0333693	-0.132664
21	118.05356	11.16	0.01325537		0.05868112				0.0230979	-0.0104565	0.03052541	0.00560681	0.04506756	0.08220221
22	119.05252	12.83	0.71343749		0.02664677				0.15193627	-0.0179565	0.09974712	-0.0061033	0.05409237	-0.2017757
23	120.01586	17.66	0.00018591		0.10822993				0.06824196	0.06919243	0.01502251	0.04292943	0.03741188	0.09189158
24	120.04371	15.87	0.45750166		0.07045047				0.08711151	0.00683223	0.2863333	-0.0844461	0.14831342	-0.2327929
25	121.03147	14.47	0.47915684		0.09716582				-0.0063402	0.09192284	0.37937981	0.09735868	0.00845184	-0.2792755

Repeat this formula in column G (change i2:n2 to o2:t2). Then in column H enter the function = SQRT(f2+g2). Then highlight and copy line 2 in columns F, G and H – paste all the way to the bottom of the file.

Ready to calculate the t-score

	A	B	C	D	E	F	G	H
1	mz	rt	p-value		Delta	SD1	SD2	SQRT
2	74.02843	14.33	0.62498766		0.02499175	0.00095971	0.00149639	0.04955905
3	74.02952	17.5	0.09755203		0.1168879	0.00355169	0.00053865	0.06395577
4	76.02753	16.21	0.91160426		0.00352843	0.00030326	0.00065711	0.03098984
5	79.96091	18.46	0.09191752		0.09533758	0.00244615	0.00016999	0.05114824
6	79.96084	17.38	0.02230474		-0.2389808	0.00015357	0.00767939	0.08850404
7	79.96117	15.16	0.51106098		0.0429163	0.0022687	0.00169756	0.0629782
8	92.05243	13.21	0.30913979		0.05266469	0.00202803	0.00038791	0.04915223
9	92.05353	15.86	0.43063321		0.02952629	0.0009716	0.00032092	0.03595173
10	105.02165	5.19	0.87614126		0.01678849	0.00615556	0.00486771	0.10499176
11	106.0684	17.26	0.04045387		0.05598331	0.00013038	0.00043584	0.02379534
12	107.05242	15.11	0.77873557		0.03935382	0.01439324	0.0041932	0.1363321
13	108.05603	15.12	0.8669316		0.00689749	0.00128625	0.00032344	0.0401209
14	111.08192	14.93	0.9405459		0.00369587	0.00110164	0.00123364	0.04832476
15	114.05963	18.77	0.57836225		-0.0224658	0.00121305	0.00031648	0.0391091
16	114.06137	16.39	0.49754071		0.02053099	0.00068495	0.00016572	0.02916628
17	115.04191	9.98	0.06870494		0.0679445	0.00040903	0.00070081	0.03331432
18	116.0382	20.88	0.72944876		0.01237564	0.00037477	0.00083569	0.0347916
19	116.04406	16.45	0.34605835		0.09064885	0.00416246	0.00424109	0.09167088
20	116.0533	18.02	0.04602802		0.14357477	0.00203571	0.00194051	0.06305726
21	118.05356	11.16	0.01325537		0.05868112	0.00017431	0.00020731	0.01953508
22	119.05252	12.83	0.71343749		0.02664677	0.00253227	0.00244106	0.07052183
23	120.01586	17.66	0.00018591		0.10822993	0.00012621	0.00022844	0.01883218
24	120.04371	15.87	0.45750166		0.07045047	0.00551658	0.00279343	0.09115928
25	121.03147	14.47	0.47915684		0.09716582	0.00752674	0.00994438	0.13217836

Divide the value in line 2/column E by line 2/column H – copy that line down the rest of column D. Then insert a column to the left of it (column D becomes column E). Copy column E and paste special (as a value) into column D. You can now delete all the columns to the right of column D.

Complete the file and save it as a .txt file

	A	B	C	D
1	mz	rt	p-value	T-score
2	74.02843	14.33	0.62498766	0.50428222
3	74.02952	17.5	0.09755203	1.82763652
4	76.02753	16.21	0.91160426	0.11385747
5	79.96091	18.46	0.09191752	1.86394636
6	79.96084	17.38	0.02230474	-2.7002253
7	79.96117	15.16	0.51106098	0.68144702
8	92.05243	13.21	0.30913979	1.07146075
9	92.05353	15.86	0.43063321	0.82127602
10	105.02165	5.19	0.87614126	0.15990297
11	106.0684	17.26	0.04045387	2.35270064
12	107.05242	15.11	0.77873557	0.28866146
13	108.05603	15.12	0.8669316	0.17191763
14	111.08192	14.93	0.9405459	0.07647979
15	114.05963	18.77	0.57836225	-0.5744404
16	114.06137	16.39	0.49754071	0.70392897
17	115.04191	9.98	0.06870494	2.03949861
18	116.0382	20.88	0.72944876	0.35570756
19	116.04406	16.45	0.34605835	0.98885111
20	116.0533	18.02	0.04602802	2.27689522
21	118.05356	11.16	0.01325537	3.00388489
22	119.05252	12.83	0.71343749	0.37785138
23	120.01586	17.66	0.00018591	5.74707256
24	120.04371	15.87	0.45750166	0.77282826
25	121.03147	14.47	0.47915684	0.73511135

You may need to add the .txt extension after saving the file if the saved file name doesn't have it.

Then place the file in the mummichog folder.

Now go to the terminal mode (or command line mode)

Sequence of steps invoking mummichog on a Mac

Stephens-MacBook-Air-3:~ stephenbarnes$ cd /Applications

Stephens-MacBook-Air-3:Applications stephenbarnes$ ls

Pymol	ls.app
Python 2.7	mummichog-1.0.9
Python 3.2	mummichog-2.0.4
Python 3.3	networkx-1.10-py3.4.egg

Stephens-MacBook-Air-3:Applications stephenbarnes$ cd mummichog-1.0.9

Stephens-MacBook-Air-3:mummichog-1.0.9 stephenbarnes$ ls

mummichog
mummichog_manual.html
test

Code to run mummichog

Stephens-MacBook-Air-3:mummichog-1.0.9 stephenbarnes$ mummichog/main.py -c 0.05 -f test/class_neg.txt -m negative -p 100 -o class_neg_out

Where mummichog/main.py is the program
-c 0.05 is the cutoff for significance
-f is the path for the .txt file
-p is the number of permutations sampling the non-significant ions
-m is the charge (default is positive)
-o is the path of the output file

Note that the *up arrow* key brings up the previous command line

First stage of mummichog

```
Stephens-MacBook-Air-3:mummichog-1.0.9 stephenbarnes$ mummichog/main.py -c 0.05 -f test/class_neg.txt
-m negative -p 100 -o class_neg_out

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

             o0                        ooooooooo
           oO0   00000  ooooo       ooo oooo
     oO0   0       ooooo  oooooo ooooo
    oooO           oooooo          oooo ooooo
        Oooo   o      000000   oooo   ooooooooo
            ooooo  oooo
                 o

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

mummichog version 1.0.9

Pygraphviz is not found. Skipping...
Started @ Sun Feb 25 19:37:12 2018

Loading metabolic network MFN_1.10.2...
cpds with MW: 2016
Got 653 significant features from 3162 references

Pathway Analysis...
query_set_size = 361 compounds
total_feature_num = 1031 compounds
```

Mummichog searching

```
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 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 30 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
Null distribution is estimated on 2590 random modules
User data yield 31 network modules

Got ActivityNetwork of 58 metabolites.
```

Problem encountered in Windows 7

From Shuzhao Li

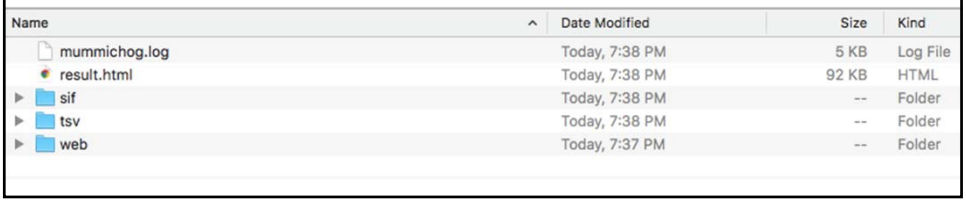
This problem is described at mummichog.org

> 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".

Also, MetaboAnalyst 4 has a module using mummichog 1 now:
http://www.metaboanalyst.ca/faces/upload/PeakUploadView.xhtml

Mummichog export

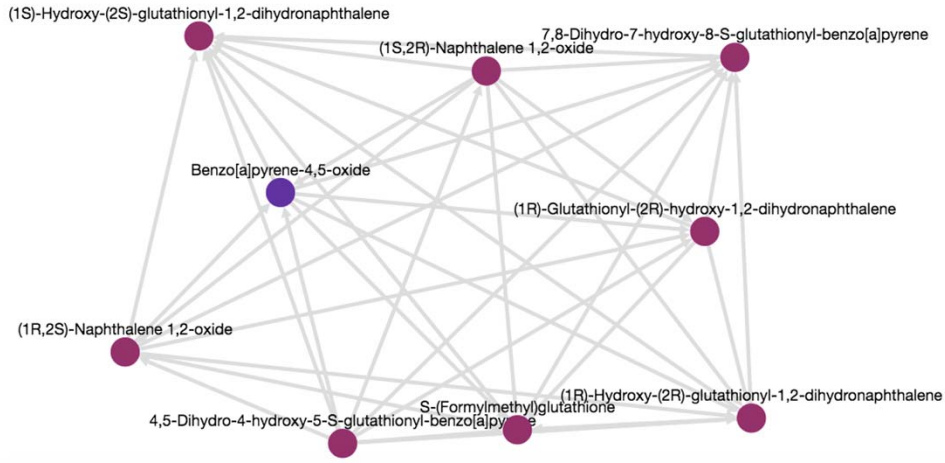
```
Annotation was written to
1519609025.95.class_neg_out/tsv/_tentative_featurematch_class_neg_out (.tsv and .xlsx)
Pathway analysis report was written to
1519609025.95.class_neg_out/tsv/mcg_pathwayanalysis_class_neg_out (.tsv and .xlsx)

Modular analysis report was written to
1519609025.95.class_neg_out/tsv/mcg_modularanalysis_class_neg_out (.tsv and .xlsx)

Inspected network report was written to
1519609025.95.class_neg_out/tsv/InspectedNodes_ActivityNetwork.tsv

Worksheet of top metabolites was written to
1519609025.95.class_neg_out/tsv/mcg_metabolite_worksheet_class_neg_out (.tsv and .xlsx)

Exporting top modules to 1519609025.95.class_neg_out/sif/...

HTML report was written to
1519609025.95.class_neg_out/result.html
```



Pathways significantly different

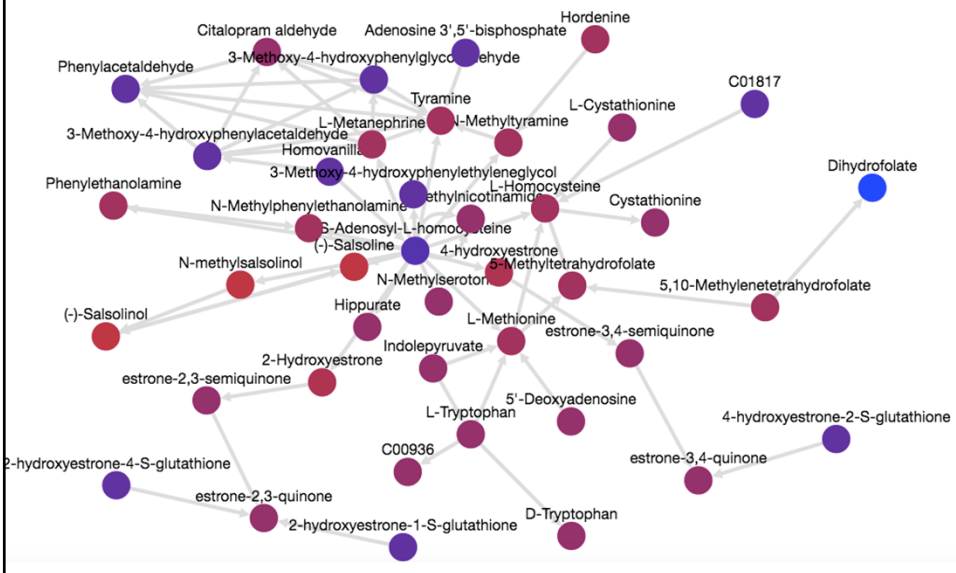
Pathways	overlap_size	pathway_size	p-value
Porphyrin metabolism	13	24	0.00066
Prostaglandin formation from dihomo gama-linoleic acid	3	3	0.00184
Glycine, serine, alanine and threonine metabolism	7	13	0.00185
Urea cycle/amino group metabolism	11	24	0.00215
Vitamin B9 (folate) metabolism	6	12	0.0039
Dynorphin metabolism	3	4	0.00441
Fatty acid activation	4	7	0.00525
Saturated fatty acids beta-oxidation	4	7	0.00525
Geraniol degradation	2	2	0.01011
Fatty acid oxidation	2	2	0.01011
De novo fatty acid biosynthesis	4	9	0.01713
N-Glycan Degradation	3	6	0.01927
Keratan sulfate degradation	3	6	0.01927

Module 1

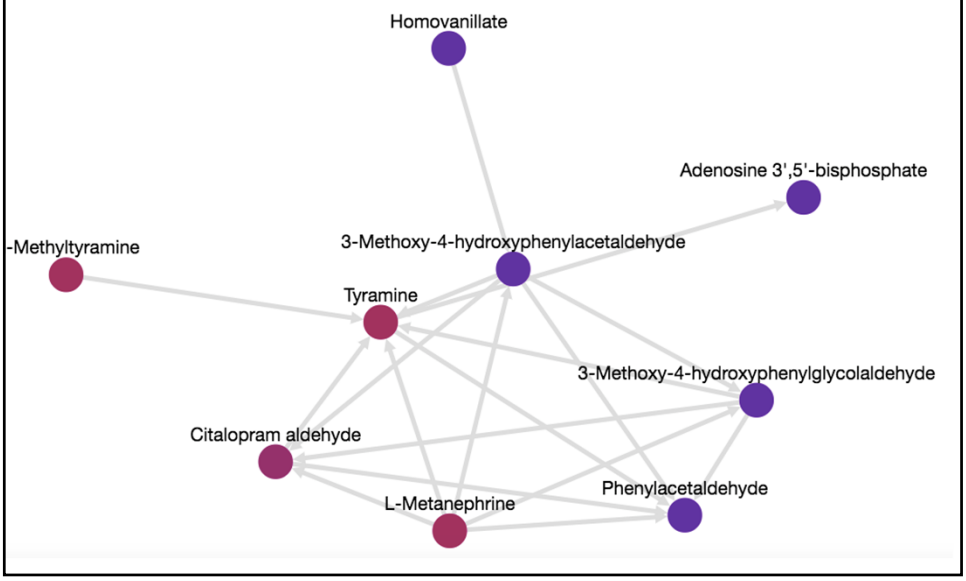


Aromatic oxidized hydrocarbons, or GSH and ROS?

Module 3

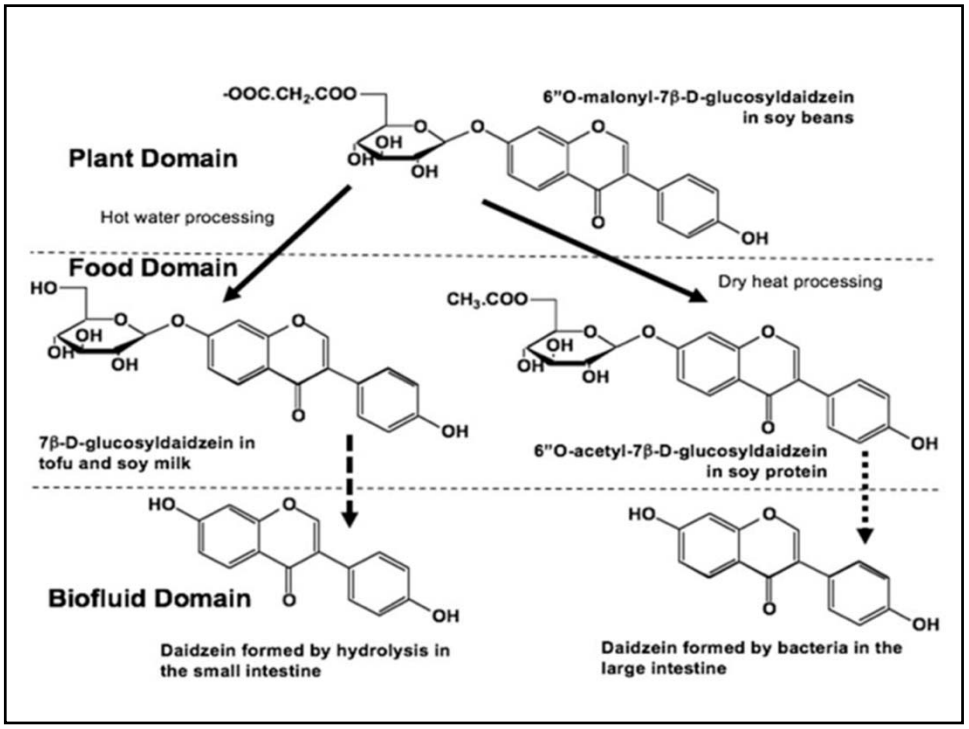


Module 5



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)



Mummichog in XCMSonline
It's called Connections