

Metabolomics in Practice Day 2

Gary W. Miller, PhD

What do the data look like?

Feature adduct
mass and retention
time

mz	time	mz.min	mz.max
85.02902	91.26769	85.02882	85.02919
85.07661	81.94827	85.07657	85.07677
85.07664	540.0139	85.07649	85.07682
85.08479	115.9833	85.08466	85.08500
85.10138	69.9247	85.09967	85.10174
85.10149	586.7514	85.10022	85.10159
85.57427	115.3154	85.57425	85.57430
85.5743	575.0111	85.57425	85.57448
86.06064	321.544	86.06057	86.06083
86.07595	113.5636	86.07592	86.07597
86.09694	59.24019	86.09659	86.09712
86.097	538.0086	86.09696	86.09717
86.5765	110.2037	86.57639	86.57658
87.04466	283.0571	87.04453	87.04482
87.05521	67.22979	87.05487	87.05595
87.10037	73.77973	87.10031	87.10048
88.00976	28.4435	88.0094	88.01021
88.02217	337.0872	88.022	88.02225
88.02214	23.49906	88.02197	88.02221
88.02217	302.4312	88.02195	88.02237
88.03994	99.20871	88.03976	88.04001
88.05419	495.0292	88.05353	88.05493
88.05839	78.67662	88.05799	88.05900
88.06379	109.4267	88.06365	88.06686
88.06694	457.9176	88.0664	88.06763
88.07625	79.93857	88.07617	88.07643
88.09841	51.20022	88.09816	88.09878
88.1127	277.2989	88.11255	88.11275

Feature adduct quality control
information

NumPres. All. Samples	NumPres. Biological. Samples	median CV	Qscore	Max Intensity
73	13	14.25	7.017544	1.41E+08
77	13	10.84	9.225092	66215199
74	13	31.45	3.17965	19526018
28	5	10.16	3.785584	25719376
77	13	10.53	9.496676	60774656
77	13	14.98	6.675567	11503169
66	11	1.927	43.91042	4.17E+08
66	11	71.34	1.186086	11034116
77	13	37.72	2.651113	66334694
62	10	20.78	3.701784	23527693
72	12	6.357	14.52064	3.41E+08
71	12	38.62	2.390153	17643702
14	2	53.71	0.286439	14551179
55	11	53.51	1.5813	33165270
9	1	15.05	0.511117	38621340
9	1	2.994	2.569241	32508791
26	4	19.21	1.60173	1983668
78	13	11.6	8.62069	63162963
77	13	61.21	1.63372	23337270
73	13	70.78	1.412828	1.08E+08
51	10	38.59	1.993342	67934762
37	9	43.86	1.578449	44013278
24	3	5.501	4.195041	1.17E+08
68	11	5.995	14.11433	3.88E+08
68	12	29.54	3.124837	34701816
78	13	15.45	6.472492	4.05E+08
22	3	7.826	2.948751	15811094
77	13	39.83	2.51067	38562444

Feature adduct abundance

DW.20130823.001	DW.20130823.003	DW.20130823.005
45407773.7	45523695.42	58024001.35
19038655.52	20229930.38	21692560.9
15016532.88	12503391.58	12736614.42
5765998.439	6028270.656	0
24442674.5	22802010.91	22752761.24
5292527.197	5737947.503	4916891.03
298129065.2	303055524	308995269.9
4268419.866	4624121.135	735099.4537
30719179.1	6380770.754	29181996.1
10008059.58	14464156.07	11648212.11
175457466.1	169322621.7	183309889
1110443.943	15275514.83	14271755.26
0	0	0
0	561023.6299	0
0	0	0
0	0	0
0	0	0
497240.4623	0	803482.9587
48980906.54	40636408.33	42969249.27
21265909.72	3945244.131	14626707.3
108223138.3	40500655.28	34468557.81
0	28789120.15	6623435.783
282236695.8	280259689.8	323712389.8
0	22031883.42	15592706.61
1602508967	2638847591	1578323649
0	0	4302269.442
5981178.31	5798194.512	4741684.342

What do the data look like?

Feature adduct
mass and retention
time

m/z	time	m/z	time
85.02902	91.26		
85.07661	81.9482		
85.07664	540.0139		
85.08479	115.9833	85.	
85.10138	69.9247	85.08	
85.10149	586.7514	85.1002	
85.57427	115.3154	85.57425	
85.5743	575.011	85.57425	
86.06064	321.544	86.06057	
86.07595	113.5636	86.07592	
86.09694	59.24019	86.09659	
86.097	538.0066	86.09696	
86.5765	110.2037	86.57639	
87.04466	283.0571	87.04453	
87.05521	67.22979	87.05487	
87.10037	73.77973	87.10031	
88.00976	28.4435	88.0094	
88.02217	337.0872	88.022	
88.02214	23.49906	88.02197	
88.02217	302.4312	88.02195	
88.03994	99.20871	88.03976	
88.05419	495.0292	88.05353	
88.05839	78.67662	88.05799	
88.06379	109.4267	88.06365	
88.06694	457.9176	88.0664	
88.07625	79.93857	88.07617	
88.09841	51.20022	88.09816	88.09878
88.1127	277.2989	88.11255	88.112758

What does this mean??

- This is how we identify molecules in a sample
- Each row describes a unique **feature** (m/z and retention time pair)
- The feature describes the physical properties of an **adduct**
 1. **m/z** – detected mass to charge ratio of feature adduct
 2. **time** – column retention time, i.e. how long after sample is injected does the feature elute?
- The adduct is a ionized version of a **molecule** present in the sample (i.e. $[M+H]^+$)

What do the data look like?

Feature adduct
mass and retention
time

Feature adduct quality control
information

mz	time	mz.min	mz.max	NumPres. All. Samples	NumPres. Biological. Samples	median CV	Qscore	Max Intensity	DW
85.02902	91.26769	85.02882	85.02919	73	13	14.25	7.017544	1.41E+08	
85.07661	81.94827	85.07657	85.07677	77	13	10.84	9.225092	6621514	
85.07664	540.0139	85.07649	85.07682	74	13	31.45	3.17965	19526018	
85.08479	115.9833	85.08466	85.08500	28	5	10.16	3.785584	25719376	
85.10138	69.9247	85.09967	85.10174	77	13	10.53	9.496676	60774656	
85.10149	586.7514	85.10022	85.10159	77	13	14.98	6.675567	11503169	
85.57427	115.3154	85.57425	85.57430	66	11	1.927	43.91042	4.17E+08	
85.5743	575.011	85.57425	85.57448	66	11	71.34	1.186086	11034116	
86.06064	321.544	86.06057	86.06083	77	13	37.72	2.651113	66334694	
86.07595	113.5636	86.07592	86.07597	62	10	20.78	3.701784	23527693	
86.09694	59.24019	86.09659	86.09712	72	12	6.357	14.52064	3.41E+08	
86.097	538.0066	86.09696	86.09717	71	12	38.62	2.390153	17843702	
86.5765	110.2037	86.57639	86.57658	14	2	53.71	0.286439	14551179	
87.04466	283.0571	87.04453	87.04482	55	11	53.51	1.5813	33165270	
87.05521	67.22979	87.05487	87.05595	9	1	15.05	0.511117	38621340	
87.10037	73.77973	87.10031	87.10048	9	1	2.994	2.569241	32508791	
88.00976	28.4435	88.0094	88.01021	26	4	19.21	1.60173	1983668	
88.02217	337.0872	88.022	88.02225	78	13	11.6	8.62069	63162963	
88.02214	23.49906	88.02197	88.02221	77	13	61.21	1.63372	23337270	
88.02217	302.4312	88.02195	88.02237	73	13	70.78	1.412828	1.08E+08	
88.03994	99.20871	88.03976	88.04001	51	10	38.59	1.993342	67934762	
88.05419	495.0292	88.05353	88.05493	37	9	43.86	1.578449	44013278	
88.05839	78.67662	88.05799	88.05900	24	3	5.501	4.195041	1.17E+08	
88.06379	109.4267	88.06365	88.06686	68	11	5.995	14.11433	3.88E+08	
88.06694	457.9176	88.0664	88.06763	68	12	29.54	3.124837	34701816	
88.07625	79.93857	88.07617	88.07643	78	13	15.45	6.472492	4.05E+08	
88.09841	51.20022	88.09816	88.09878	22	3	7.826	2.948751	15811094	0 0 4302269.442
88.1127	277.2989	88.11255	88.11275	77	13	39.83	2.51067	38562444	5981178.31 5798194.512 4741684.342

What does this mean??

- These are quality-control parameters
- We use these to eliminate features that show high variability within technical replicates
- In general, we eliminate features that have >35% variability

What do the data look like?

What does this mean??

This is the **relative abundance** of each feature for each sample

Can determine absolute abundance if each analyte is positively identified and then run against reference standards (difficult to do with any analytes of interest)

88.07023	78.83657	88.07017	88.070458	78	13	15.45	0.472482	4.05E+09
88.09841	51.20022	88.09816	88.098781	22	3	7.826	2.048751	15811094
88.1127	277.2989	88.11255	88.112758	77	13	39.83	2.51067	38562444

Feature adduct abundance

	DW.20130823.001	DW.20130823.003	DW.20130823.005
45407773.7	45523695.42	58024001.35	
19038655.52	20229930.38	21692560.9	
15016532.88	12503391.58	12736614.42	
5765998.439	6028270.656	0	
24442674.5	22802010.91	22752761.24	
5292527.197	5737947.503	4916891.03	
298129065.2	303055524	308995269.9	
4268419.866	4624121.135	735099.4537	
30719179.1	6380770.754	29181996.1	
10008059.58	14464156.07	11648212.11	
175457466.1	169322621.7	183309889	
1110443.943	15275514.83	14271755.26	
51179	0	0	0
65270	0	561023.6299	0
21340	0	0	0
08791	0	0	0
83668	497240.4623	0	803482.9587
62963	48980906.54	40636408.33	42969249.27
37270	21265909.72	3945244.131	14626707.3
3E+08	108223138.3	40500655.28	34468557.81
34762	0	28789120.15	6623435.783
13278	0	0	16762908.27
7E+08	0	0	7083644.261
3E+08	282236695.8	280259689.8	323712389.8
01816	0	22031883.42	15592708.61
	1602508967	2638847591	1578323649
	0	0	4302269.442
	5981178.31	5798194.512	4741684.342

Metabolomics Workflow

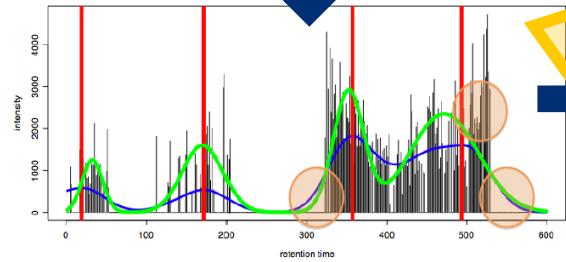
Sample collection
and preparation



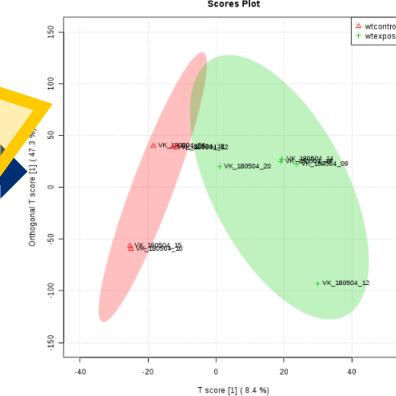
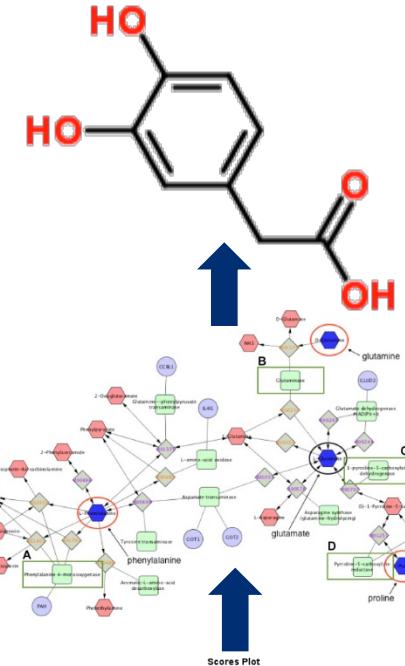
Mass spectrometry



New data processing
(CMS, xMS analyzer,
ComBat)



Data transformation and normalization



Metabolite pred
and verification

Pathway/enrich
analysis (Metsca
MetaCore, cytos

Data mining
(Supervised &
Unsupervised)

Day 2 Practicum

Bottom-up approach



Support or refute hypothesis



Quantify pre-selected analyte
using reference standards



Optimize and verify in real world
samples



Select optimal separation, sample
prep and detection strategy based
on analyte properties



Select analyte(s) based on
hypothesis



**Targeted chemical
profiling**

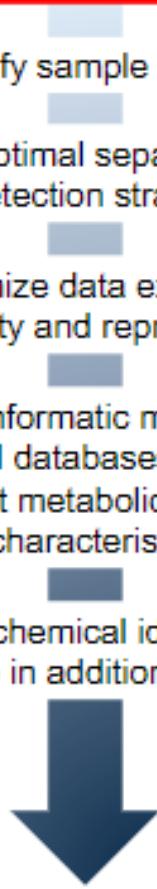
Day 2 Practicum

Bottom-up approach

- 
- Support or refute hypothesis
 - Quantify pre-selected analyte using reference standards
 - Optimize and verify in real world samples
 - Select optimal separation, sample prep and detection strategy based on analyte properties
 - Select analyte(s) based on hypothesis

Targeted chemical profiling

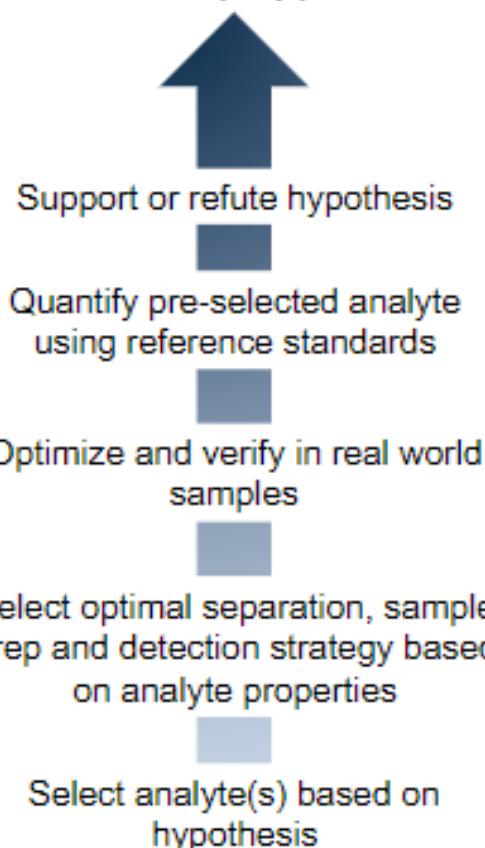
Untargeted chemical profiling

- 
- Identify sample subjects
 - Select optimal separation and detection strategy
 - Maximize data extraction sensitivity and reproducibility
 - Use bioinformatic methods and chemical databases to identify important metabolic phenotype characteristics
 - Confirm chemical identities and validate in additional cohorts

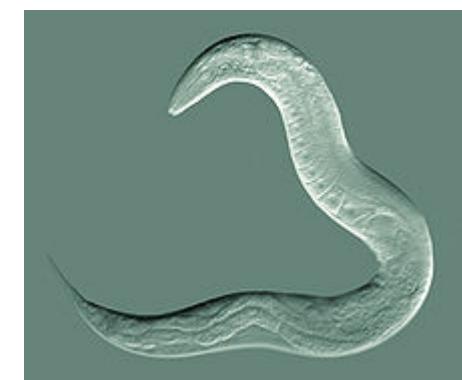
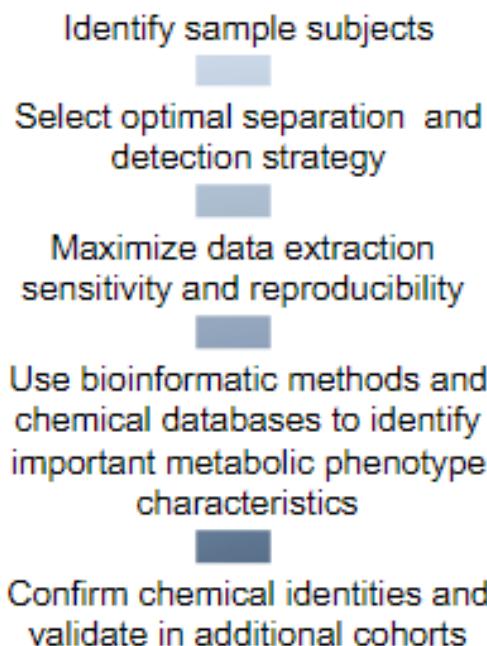
Top-down approach

Day 2 Practicum

Bottom-up approach



Untargeted chemical profiling



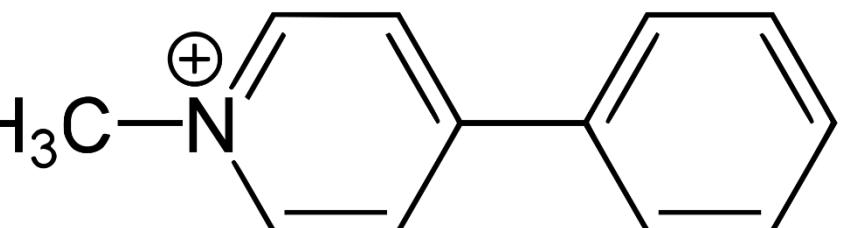
Confirmation or expansion of untargeted results in controlled experiments using laboratory models

Day 2 Practicum



C. elegans

- Short, predictable life span
- Large number of offspring
- Inexpensive
- **Allow for controlled exposures**

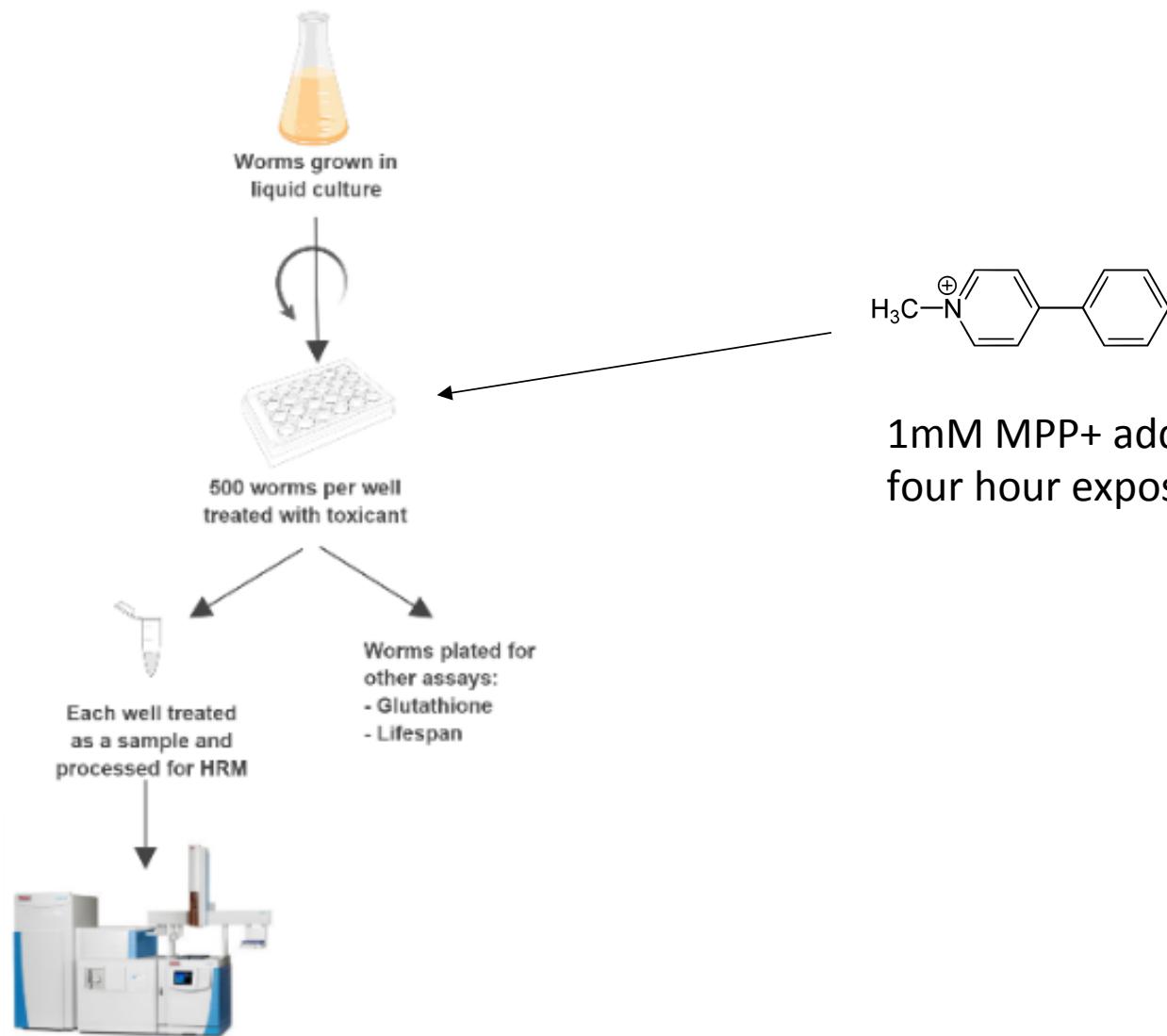


MPP+ Exposure

- MPP+ is a toxic product which kills DA neurons and induces a Parkinsonian state in animal models

Day 2 Practicum

Sample preparation



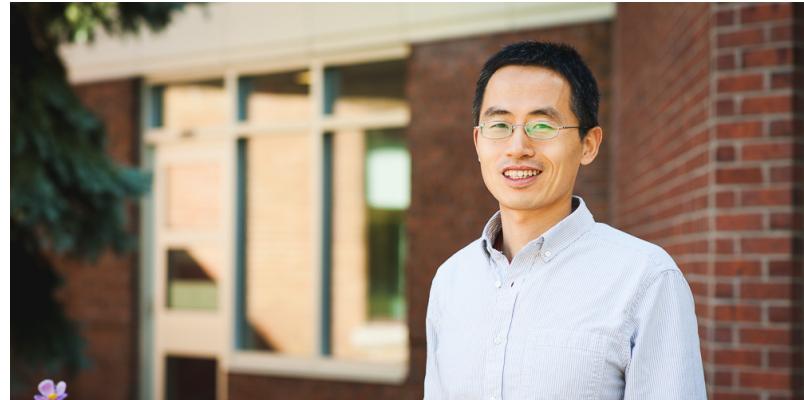
Day 2 Practicum

KAY LET'S DO IT

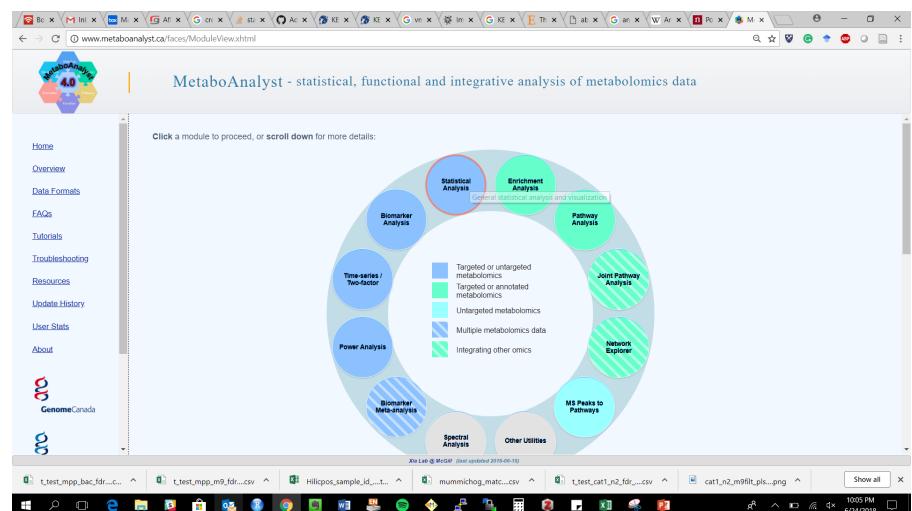
ay 1 we will use a fantastic
nline tool: MetaboAnalyst 4.0

www.metaboanalyst.ca

ee data on github



Jeff Xia at McGill University



The screenshot shows the MetaboAnalyst 4.0 homepage. On the left, a sidebar lists links: Home, Overview, Data Formats, FAQs, Tutorials, Troubleshooting, Resources, Update History, User Stats, and About. Below this is the GenomeCanada logo. The main area features a large circular diagram titled "MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data". The diagram is divided into several colored segments representing different analysis types: Statistical Analysis (red), Enrichment Analysis (green), Pathway Analysis (light green), Joint Pathway Analysis (diagonal stripes), Network Explorer (yellow), MS Peaks to Pathways (blue), Spectral Analysis (light blue), Biomarker Meta-analysis (diagonal stripes), Power Analysis (light blue), Time-series / Two-factor (light blue), Biomarker Analysis (light blue), and Targeted or untargeted metabolomics (blue). A legend at the bottom right identifies the colors: blue for targeted or untargeted metabolomics, teal for targeted or annotated metabolomics, light blue for untargeted metabolomics, light green for multiple metabolomics data, and yellow for integrating other omics.

Day 2 Practicum

. Go to metaboanalyst.ca and select “click here to start”

The screenshot shows the MetaboAnalyst 4.0 website. The top navigation bar includes links for "Home", "Overview", "Data Formats", "FAQs", "Tutorials", "Troubleshooting", "Resources", "Update History", "User Stats", and "About". On the left, there's a sidebar with logos for GenomeCanada and GenomeQuébec, and the NSERC CRSNG logo. The main content area features a large banner with the text "MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data". Below the banner, a "Welcome" section contains a link "[click here to start <<](#)". A "News & Updates" section lists several recent fixes and releases, such as "Fixed issue with pathway visualization (06/18/2018); NEW", "Check out our latest paper on [MetaboAnalyst 4.0](#) NEW", and "Release of MetaboAnalyst 4.0 together with a companion R package [MetaboAnalystR](#). You can still access [version 3.0 here](#) (01/29/2018)". At the bottom of the news section is a "Read more" link. The footer contains a "Please Cite:" section with a list of publications, including references to Nucleic Acids Research and Current Protocols in Bioinformatics.

MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

Welcome [click here to start <<](#)

News & Updates

- Fixed issue with pathway visualization (06/18/2018); **NEW**
- Check out our latest paper on [MetaboAnalyst 4.0](#) **NEW**
- Check out our [OmicsNet](#) for flexible creation & 3D visualization of complex networks integrating metabolites, genes/proteins, miRNAs and transcription factors; **NEW**
- Check out our [MicrobiomeAnalyst](#) for comprehensive analysis of microbiome data;
- Enhanced pathway image generation to deal with concurrency issue (06/13/2018); **NEW**
- Fixed the issues for name mapping and node-click information in pathway visualization (06/12/2018); **NEW**
- Fixed the issue for data editor in biomarker analysis (05/28/2018); **NEW**
- Fixed the issue for sample hold-out analysis in biomarker analysis (04/23/2018); **NEW**
- Fixed the issue with time-series group ordering based on numeric values (04/19/2018); **NEW**
- Enhanced support for SVG export for KEGG global network (04/04/2018); **NEW**
- Release of MetaboAnalyst 4.0 together with a companion R package [MetaboAnalystR](#). You can still access [version 3.0 here](#) (01/29/2018);

[Read more](#)

Please Cite:

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., Wishart, D.S. and Xia, J. (2018) [MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis](#). Nuc. Acids Res. (doi:10.1093/nar/gky310).

Xia, J. and Wishart, D.S. (2016) [Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis](#). Current Protocols in Bioinformatics, 55:14.10.1-14.10.91.

Xia, J., Sinelnikov, I., Han, B., and Wishart, D.S. (2015) [MetaboAnalyst 3.0 - making metabolomics more meaningful](#). Nuc. Acids Res. 43, W251-257.

Xia, J., Mandal, R., Sinelnikov, I., Broadhurst, D., and Wishart, D.S. (2012) [MetaboAnalyst 2.0 - a comprehensive server for metabolomic data analysis](#). Nuc. Acids Res. 40, W127-133.

Xia, J. and Wishart, D.S. (2011) [Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst](#). Nature Protocols 6 (6), 743-760.

Xia, J. and Wishart, D.S. (2011) [Metabolomic data processing, analysis, and interpretation using MetaboAnalyst](#). Current Protocols in Bioinformatics, 14:10.1-14.10.48.

Xia, J., Psychogios, N., Young, N. and Wishart, D.S. (2009) [MetaboAnalyst: a web server for metabolomic data analysis and interpretation](#). Nuc. Acids Res. 37, W652-660.

Xia, J., Broadhurst, D., Wilson, M. and Wishart, D. (2013) [Translational biomarker discovery in clinical metabolomics: an introductory tutorial](#). Metabolomics, 9, 280-299. (biomarker analysis)

Xia, J., Sinelnikov, I., and Wishart, D.S. (2011) [MetTATT - a web-based metabolomic tool for analyzing time-series and two-factor data sets](#). Bioinformatics, 27, 2455-2456 (time-series and two-factor analysis)

Xia, J. and Wishart, D.S. (2010) [MeIPA: a web-based metabolomics tool for pathway analysis and visualization](#). Bioinformatics, 26, 2342-2344 (metabolic pathway analysis)

Xia, J. and Wishart, D.S. (2010) [MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data](#). Nucleic Acids Research, 38, W71-77 (metabolite set enrichment analysis)

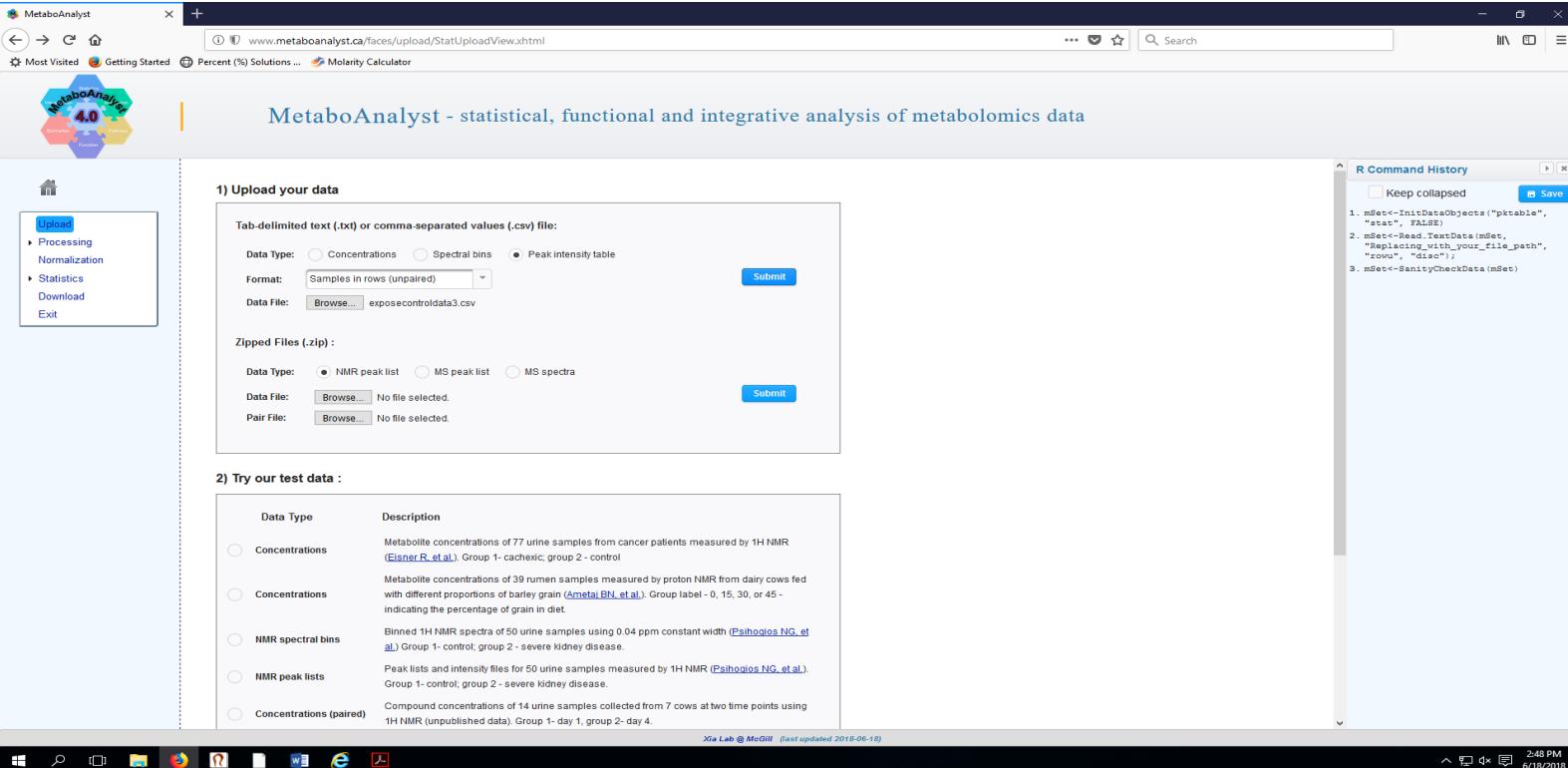
Day 2 Practicum

2. Select “Statistical Analysis”

The screenshot shows the MetaboAnalyst 4.0 website interface. On the left is a vertical sidebar with links to Home, Overview, Data Formats, FAQs, Tutorials, Troubleshooting, Resources, Update History, User Stats, and About. Logos for GenomeCanada, GenomeQuébec, and NSERC CRSNG are also present. The main content area features a large circular diagram with various analysis modules. The "Statistical Analysis" module is highlighted with a red circle and a callout box describing it as "General statistical analysis and visualization". A legend at the bottom of the diagram defines five categories: Targeted or untargeted metabolomics (blue), Targeted or annotated metabolomics (green), Untargeted metabolomics (cyan), Multiple metabolomics data (diagonal stripes), and Integrating other omics (vertical stripes). Other modules shown include Biomarker Analysis, Time-series / Two-factor, Power Analysis, Biomarker Meta-analysis, Spectral Analysis, Other Utilities, MS Peaks to Pathways, Network Explorer, Joint Pathway Analysis, and Enrichment Analysis.

Day 2 Practicum

3. Attach data and select proper data type and format (in this case, it is a *peak intensity table* organized in *rows*)
load the mpp_bac-fdr0.95 file from github



The screenshot shows the MetaboAnalyst 4.0 web interface. On the left, a sidebar menu includes 'Upload' (which is currently selected), 'Processing', 'Normalization', 'Statistics', 'Download', and 'Exit'. The main content area is titled 'MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data'. Step 1, 'Upload your data', is active. It contains fields for 'Data Type' (radio buttons for 'Concentrations', 'Spectral bins', and 'Peak intensity table', with 'Peak intensity table' selected), 'Format' (dropdown menu showing 'Samples in rows (unpaired)'), 'Data File' (button to 'Browse...' a file named 'exposcontroldata3.csv'), and 'Submit' (blue button). Below this, there are sections for 'Zipped Files (.zip)' and 'Pair File', each with similar input fields. Step 2, 'Try our test data', lists several pre-defined datasets with their descriptions. On the right, the 'R Command History' panel displays the R code used for the analysis:

```
1. mSet<-InitDataObjects("ptable", "stat", FALSE)
2. mSet<-Read.TextData(mSet, "Replacing_with_your_file_path",
"read", "data");
3. mSet<-SanityCheckData(mSet)
```

Day 2 Practicum

4. Press “Skip”

The screenshot shows the MetaboAnalyst 4.0 web application. The left sidebar has a navigation menu with 'Upload' selected, followed by 'Processing', 'Pre-process', 'Data check' (which is highlighted in blue), 'Missing value', 'Data filter', 'Data editor', 'Image options', 'Normalization', 'Statistics', 'Download', and 'Exit'. The main content area is titled 'MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data'. A section titled 'Data Integrity Check:' lists four steps: 1. Checking the class labels - at least three replicates are required in each class. 2. If the samples are paired, the pair labels must conform to the specified format. 3. The data (except class labels) must not contain non-numeric values. 4. The presence of missing values or features with constant values (i.e. all zeros). Below this is a dashed box containing 'Data processing information:' with the following details: 'Checking data content ...passed', 'Samples are in rows and features in columns', 'The uploaded file is in comma separated values (.csv) format.', 'The uploaded data file contains 10 (samples) by 6151 (peaks(mz/rt)) data matrix.', 'Samples are not paired.', '2 groups were detected in samples.', 'Only English letters, numbers, underscore, hyphen and forward slash (/) are allowed.', 'Other special characters or punctuations (if any) will be stripped off.', 'All data values are numeric.', 'A total of 0 (0%) missing values were detected.', 'By default, these values will be replaced by a small value.', 'Click Skip button if you accept the default practice', and 'Or click Missing value imputation to use other methods'. At the bottom of this box are two buttons: 'Missing value estimation' and 'Skip'. On the right side, there is an 'R Command History' panel with the following code:

```
1. mSet<-InitDataObjects("pktable", "stat", FALSE)
2. mSet<-Read.TextData(mSet, "Replacing_with_your_file_path", "roww", "disc");
3. mSet<-SanityCheckData(mSet)
```

Keep collapsed Save

Day 2 Practicum

Select “None” for Data Filtering if data has already been filtered in R or Excel (in this case, we have pre-filtered the data). If data has not already been filtered, submit the default IQR filter.

The screenshot shows the MetaboAnalyst 4.0 web interface. The left sidebar contains navigation links: Home, Upload, Processing (Pre-process, Data check, Data filter, Missing value, Data editor, Image options), Normalization, Statistics, Download, and Exit. The main content area is titled "MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data". A sub-section titled "Data Filtering:" explains the purpose of filtering and provides empirical rules for different variable counts. It notes that the "None" option is only for less than 4000 features. A dashed box highlights the filtering options, which include a checkbox for "Filtering features if their RSDs are > [25%] in QC samples" and a radio button for "None (less than 5000 features)". Other options listed are Interquartile range (IQR), Standard deviation (SD), Median absolute deviation (MAD), Relative standard deviation (RSD = SD/mean), Non-parametric relative standard deviation (MAD/median), Mean intensity value, and Median intensity value. On the right, the "R Command History" panel shows the R code used for data processing:

```
1. mSet<-InitDataObjects("pktable", "stat", FALSE)
2. mSet<-Read.TextData(mSet, "Replacing_with_your_file_path", "row", "disc");
3. mSet<-SanityCheckData(mSet)
4. mSet<-ReplaceMin(mSet);
```

Day 2 Practicum

5. Select “None” for Data Filtering if data has already been filtered in R not

The screenshot shows a Windows desktop environment with several open windows. In the center is the MetaboAnalyst 4.0 web application, which displays a table of predicted pathway activity profiles based on the dataset "mummichog". The table includes columns for Pathway Name, Total, Hits (all), Hits (sig.), Fisher's P, EASE Score, Gamma P, and Details. The pathways listed include Terpenoid backbone biosynthesis, Tyrosine metabolism, Vitamin B6 metabolism, Glycerophospholipid metabolism, Glycerolipid metabolism, Aminoacyl-tRNA biosynthesis, Alanine, aspartate and glutamate metabolism, Purine metabolism, Amino sugar and nucleotide sugar metabolism, Pyruvate metabolism, Sulfur metabolism, Glutathione metabolism, Citrate cycle (TCA cycle), Cysteine and methionine metabolism, Glycolysis or Gluconeogenesis, Glycine, serine and threonine metabolism, Arginine and proline metabolism, Pyrimidine metabolism, Fructose and mannose metabolism, and Galactose metabolism. The R Command History window on the right shows the R code used to process the data, including reading the dataset and performing Mummichog analysis. The taskbar at the bottom shows various open files and applications.

Predicted pathway activity profiles based on mummichog

Pathway Name	Total	Hits (all)	Hits (sig.)	Fisher's P	EASE Score	Gamma P	Details
Terpenoid backbone biosynthesis	10	6	4	0.018296	0.10327	2.8806E-4	View
Tyrosine metabolism	25	13	5	0.10715	0.26453	7.5935E-4	View
Vitamin B6 metabolism	8	6	3	0.10505	0.35506	0.0013256	View
Glycerophospholipid metabolism	24	7	3	0.1569	0.43532	0.002195	View
Glycerolipid metabolism	9	3	2	0.10906	0.49662	0.0032528	View
Aminoacyl-tRNA biosynthesis	66	19	5	0.34901	0.56528	0.0051075	View
Alanine, aspartate and glutamate metabolism	20	14	4	0.3205	0.5694	0.0052499	View
Purine metabolism	54	26	6	0.45171	0.64065	0.0085239	View
Amino sugar and nucleotide sugar metabolism	27	16	4	0.42464	0.66628	0.010198	View
Pyruvate metabolism	18	5	2	0.27404	0.68204	0.011403	View
Sulfur metabolism	7	5	2	0.27404	0.68204	0.011403	View
Glutathione metabolism	25	12	3	0.46219	0.73934	0.017327	View
Citrate cycle (TCA cycle)	20	6	2	0.35853	0.74744	0.018415	View
Cysteine and methionine metabolism	26	13	3	0.52016	0.78013	0.023678	View
Glycolysis or Gluconeogenesis	25	14	3	0.57462	0.81525	0.031396	View
Glycine, serine and threonine metabolism	22	11	2	0.69636	0.92057	0.083968	View
Arginine and proline metabolism	28	15	2	0.84836	0.96873	0.15938	View
Pyrimidine metabolism	40	26	2	0.98209	0.99767	0.36954	View
Fructose and mannose metabolism	20	13	1	0.95127	1.0	1.0	View
Galactose metabolism	13	10	1	0.9016	1.0	1.0	View

R Command History

```
Keep collapsed Save
1. mSet<-initMtaObjects("mass_all", "mummichog", FALSE)
2. mSet<-ReadPeakListData(mSet, "Replacing_with_your_file_path");
3. mSet<-UpdateMummichogParameters(mSet, "0.1", "positive", 0.1);
4. mSet<-SanityCheckMummichogData(mSet)
5. mSet<-PerformMummichog(mSet, "cel_kegg", "fisher", "gamma")
6. mSet<-PerformMummichog(mSet, "cel_kegg", "fisher", "gamma")
```

ory Save

```
objects("pktable",
Data(mSet,
"your_file_path",
;
ckData(mSet)
n(mSet);
```

Xia Lab @ McGill (last updated 2018-06-18)

t_test_mpp_bac_fdr....csv t_test_mpp_m9_fdr....csv Hilicpos_sample_id....t... mummichog_matc....csv t_test_cat1_n2_fdr....csv cat1_n2_m9filt_pls....png Show all 10:11 PM 6/24/2018

Day 2 Practicum

data has already been normalized in R using a log base 10 transformation, select “None” for all. If data has already been normalized, use Metaboanalyst’s generalized log transformation, or play around with which normalization best normalizes the data.

The screenshot shows the MetaboAnalyst 4.0 web interface. The left sidebar has a navigation menu with 'Normalization' selected. The main content area contains three main sections: 'Sample normalization', 'Data transformation', and 'Data scaling'. Each section lists several normalization methods with 'Specify' links. The 'R Command History' panel on the right shows the R code used for the analysis:

```
1. mSet<-InitDataObjects("pktable", "stat", FALSE)
2. mSet<-Read.TextData(mSet, "Replacing_with_your_file_path", "row1", "disc");
3. mSet<-SanityCheckData(mSet)
4. mSet<-ReplaceMin(mSet);
```

MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

Data Integrity Check:

1. Checking the class labels - at least three replicates are required in each class.
2. If the samples are paired, the pair labels must conform to the specified format.
3. The data (except class labels) must not contain non-numeric values.
4. The presence of missing values or features with constant values (i.e. all zeros)

Data processing information:

Checking data content ...passed
Samples are in columns and features in rows.
The uploaded file is in comma separated values (.csv) format.
The uploaded data file contains 10 (samples) by 6254 (peaks(mz/rt)) data matrix.
Samples are not paired.
2 groups were detected in samples.
Only English letters, numbers, underscore, hyphen and forward slash (/) are allowed.
Other special characters or punctuations (if any) will be stripped off.
All data values are numeric.
A total of 0 (0%) missing values were detected.
By default, these values will be replaced by a small value.
Click **Skip** button if you accept the default practice
Or click **Missing value imputation** to use other methods

R Command History

Keep collapsed Save

```
1. mSet<-InitDataObjects("pktable", "stat", FALSE)
2. mSet<-Read.TextData(mSet, "Replacing_with_your_file_path", "colu", "disc");
3. mSet<-SanityCheckData(mSet)
```

Xia Lab @ McGill (last updated 2018-06-18)

10:05 PM 6/24/2018

Bo x M Int x box M x Afi x G cr x sta x Ac x KE x KE x Im x G vn x KE x Th x ab x G an x W Ar x n Po x M x

www.metaboanalyst.ca/faces/Secure/process/FilterView.xhtml

MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

variables, many of them are from baseline noises. Filtering can usually improve the results. For details, please refer to the paper by [Hackstadt, et al.](#)

Non-informative variables can be characterized in three groups: 1) variables of **very small values** (close to baseline or detection limit) - these variables can be detected using mean or median; 2) variables that are **near-constant values** throughout the experiment conditions (housekeeping or homeostasis) - these variables can be detected using standard deviation (SD); or the robust estimate such as interquartile range (IQR); and 3) variables that show **low repeatability** - this can be measured using QC samples using the relative standard deviation(RSD = SD/mean). Features with high percent RSD should be removed from the subsequent analysis (the suggested threshold is 20% for LC-MS and 30% for GC-MS). For data filtering based on the first two categories, the following empirical rules are applied during data filtering:

- **Less than 250 variables:** 5% will be filtered;
- **Between 250 - 500 variables:** 10% will be filtered;
- **Between 500 - 1000 variables:** 25% will be filtered;
- **Over 1000 variables:** 40% will be filtered;

Please note, in order to reduce the computational burden to the server, the **None** option is only for less than 4000 features. Over that, if you choose None, the IQR filter will still be applied. In addition, the maximum allowed number of variables is 8000. If over 8000 variables were left after filtering, only the top 8000 will be used in the subsequent analysis.

Filtering features if their RSDs are > % in QC samples

None (less than 5000 features)
 Interquartile range (IQR)
 Standard deviation (SD)
 Median absolute deviation (MAD)
 Relative standard deviation (RSD = SD/mean)
 Non-parametric relative standard deviation (MAD/median)
 Mean intensity value
 Median intensity value

Submit **Proceed**

R Command History

Keep collapsed **Save**

```
1. mSet<-InitDataObjects("pktable", "stat", F  
ALSE)  
2. mSet<-Read.TextData(mSet, "Replacing_with_  
your_file_path", "colu", "disc");  
3. mSet<-SanityCheckData(mSet)  
4. mSet<-ReplaceMin(mSet);
```



www.metaboanalyst.ca/faces/Secure/process/NormalizationView.xhtml



MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

Sample normalization

- None
- Sample-specific normalization (i.e. weight, volume) [Specify](#)
- Normalization by sum
- Normalization by median
- Normalization by reference sample (PQN) [Specify](#)
- Normalization by a pooled sample from group [Specify](#)
- Normalization by reference feature [Specify](#)
- Quantile normalization

Data transformation

- None
- Log transformation (generalized logarithm transformation or glog)
- Cube root transformation (take cube root of data values)

Data scaling

- None
- Mean centering (mean-centered only)
- Auto scaling (mean-centered and divided by the standard deviation of each variable)
- Pareto scaling (mean-centered and divided by the square root of standard deviation of each variable)
- Range scaling (mean-centered and divided by the range of each variable)

R Command History

Keep collapsed Save

```
1. mSet<-InitDataObjects("pktable", "stat", F ALSE)
2. mSet<-Read.TextData(mSet, "Replacing_with_ your_file_path", "colu", "disc");
3. mSet<-SanityCheckData(mSet)
4. mSet<-ReplaceMin(mSet);
```

Xia Lab @ McGill (last updated 2018-06-18)

t_test_mpp_bac_fdr....csv ^ t_test_mpp_m9_fdr....csv ^ Hilicpos_sample_id....t... ^ mummichog_matc....csv ^ t_test_cat1_n2_fdr....csv ^ cat1_n2_m9filt_pls....png ^ Show all X

10:06 PM 6/24/2018

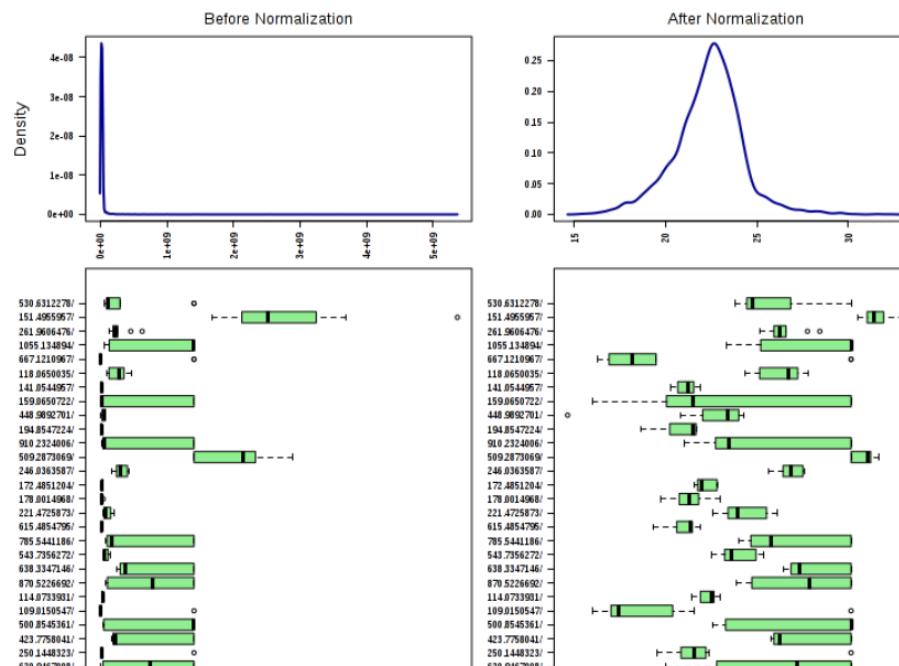
MetaboA



Normalization Result:

Please note: the boxplots show at most 50 features/samples due to space limitation; the density plots are based on all data

Feature View Sample View



- None
 - Sample-specific
 - Normalization by sample
 - Normalization by column
 - Normalization by row
 - Quantile normalization
- Data transformation
- None
 - Log transformation
 - Cube root transformation
- Data scaling
- None
 - Mean centering (n)
 - Auto scaling (n)
 - Pareto scaling (n)
 - Range scaling (n)

Normalize

R Command History

```
Keep collapsed  Save   
1. mSet<-InitDataObjects("pktable", "stat", FALSE)  
2. mSet<-Read.TextData(mSet, "Replacing_with_your_file_path", "colu", "disc")  
3. mSet<-SanityCheckData(mSet)  
4. mSet<-ReplaceMin(mSet);  
5. mSet<-Normalization(mSet, "NULL", "LogNorm", "NULL", ratio=FALSE, ratioNum=20)  
6. mSet<-PlotNormSummary(mSet, "norm_0_", "png", 72, width=NA)  
7. mSet<-PlotSampleNormSummary(mSet, "snorm_0_", "png", 72, width=NA)
```

MetaboAnalyst 4.0 - statistical, functional and integrative analysis of metabolomics data

Select an analysis path to explore :

Univariate Analysis

- [Fold Change Analysis](#)
- [T-tests](#)
- [Volcano plot](#)

One-way Analysis of Variance (ANOVA)

- [Correlation Analysis](#)
- [Pattern Searching](#)

Chemometrics Analysis

- [Principal Component Analysis \(PCA\)](#)
- [Partial Least Squares - Discriminant Analysis \(PLS-DA\)](#)
- [Sparse Partial Least Squares - Discriminant Analysis \(sPLS-DA\)](#)
- [Orthogonal Partial Least Squares - Discriminant Analysis \(orthoPLS-DA\)](#)

Feature Identification

- [Significance Analysis of Microarray \(and Metabolites\) \(SAM\)](#)
- [Empirical Bayesian Analysis of Microarray \(and Metabolites\) \(EBAM\)](#)

Cluster Analysis

- Hierarchical Clustering: [Dendrogram](#) [Heatmaps](#)
- Partitional Clustering: [K-means](#) [Self Organizing Map \(SOM\)](#)

Classification & Feature Selection

- [Random Forest](#)
- [Support Vector Machine \(SVM\)](#)

R Command History

Keep collapsed Save

```
1. mSet<-InitDataObjects("pktable", "stat", F  
ALSE)  
2. mSet<-Read.TextData(mSet, "Replacing_with_  
your_file_path", "col", "disc");  
3. mSet<-SanityCheckData(mSet);  
4. mSet<-ReplaceNaN(mSet);  
5. mSet<-Normalization(mSet, "NULL", "LogNor  
m", "NULL", ratio=FALSE, ratioNum=20);  
6. mSet<-PlotNormSummary(mSet, "norm_0_", "pn  
g", 72, width=NA);  
7. mSet<-PlotSampleNormSummary(mSet, "snorm_0_  
_", "png", 72, width=NA);
```

Xia Lab @ McGill (last updated 2018-06-18)

t_test_mpp_bac_fdr....c... t_test_mpp_m9_fdr....csv Hilicpos_sample_id....t... mummichog_matc....csv t_test_cat1_n2_fdr....csv cat1_n2_m9filt_pls....png Show all



MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

R Command History

Keep collapsed Save

```

1. mSet<-InitDataObjects("pktable", "stat", FALSE)
2. mSet<-Read.TextData(mSet, "Replacing_with_your_file_path", "colv", "disc")
3. mSet<-SanityCheckData(mSet)
4. mSet<-ReplaceMin(mSet);
5. mSet<-Normalization(mSet, "NULL", "LogNorm", "NULL", ratio=FALSE, ratioNum=20)
6. mSet<-PlotNormSummary(mSet, "norm_0", "png", 72, width=NA)
7. mSet<-PlotSampleNormSummary(mSet, "snorm_0", "png", 72, width=NA)
8. mSet<-PCA.Anal(mSet)
9. mSet<-PlotPCAPairSummary(mSet, "pca_pair_0", "png", 72, width=NA, 5)
10. mSet<-PlotPCAScree(mSet, "pca_scree_0", "png", 72, width=NA, 5)
11. mSet<-PlotPCA2DScore(mSet, "pca_score2d_0", "png", 72, width=NA, 1,2,0.95,1,0)
12. mSet<-PlotPCALoading(mSet, "pca_loading_0", "png", 72, width=NA, 1,2,"scatter", 1);
13. mSet<-PlotPCABiplot(mSet, "pca_biplot_0", "png", 72, width=NA, 1,2)
14. mSet<-PlotPCA3DScoreImg(mSet, "pca_score3d_0", "png", 72, width=NA, 1,2,3, 40)

```

Display pairwise score plot for top 5 PCs

PC 1 37.9 %

PC 2 14.2 %

PC 3 8.7 %

PC 4 8 %

Xia Lab @ McGill (last updated 2018-06-18)

Upload Processing Normalization Statistics Fold change T-test Volcano plot ANOVA Correlations PatternHunter PCA PLSDA sPLSDA OrthoPLSDA SAM EBAM Dendrogram Heatmap SOM K-means RandomForest SVM Download

Overview Scree Plot 2D Scores Plot 3D Scores Plot Loadings Plot Biplot

t_test_mpp_bac_fdr...csv t_test_mpp_m9_fdr...csv Hilicpos_sample_id....t... mummichog_matc....csv t_test_cat1_n2_fdr...csv cat1_n2_m9filt_pls....png Show all

10:06 PM 6/24/2018

MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

R Command History

```

 Keep collapsed  Save
1. mSet<-InitDataObjects("pktable", "stat", FALSE)
2. mSet<-Read.TextData(mSet, "Replacing_with _your_file_path", "colu", "disc");
3. mSet<-SanityCheckData(mSet)
4. mSet<-ReplaceMin(mSet);
5. mSet<-Normalization(mSet, "NULL", "LogNor m", "NULL", ratio=FALSE, ratioNum=20)
6. mSet<-PlotNormSummary(mSet, "norm_0_", "png", 72, width=NA)
7. mSet<-PlotSampleNormSummary(mSet, "snorm_0_", "png", 72, width=NA)
8. mSet<-PCA.Anal(mSet)
9. mSet<-PlotPCAPairSummary(mSet, "pca_pair_0_", "png", 72, width=NA, 5)
10. mSet<-PlotPCAScreet(mSet, "pca_scree_0_", "png", 72, width=NA, 5)
11. mSet<-PlotPCA2DScore(mSet, "pca_score2d_0 _", "png", 72, width=NA, 1,2,0.95,1,0)
12. mSet<-PlotPCALoading(mSet, "pca_loading_0 _", "png", 72, width=NA, 1,2,"scatter", 1);
13. mSet<-PlotPCABiplot(mSet, "pca_biplot_0 _", "png", 72, width=NA, 1,2)
14. mSet<-PlotPCA3DScoreImg(mSet, "pca_score3 d_0_", "png", 72, width=NA, 1,2,3, 40)

```

Overview **Scree Plot** **2D Scores Plot** **3D Scores Plot** **Loadings Plot** **Biplot**

Specify PC on X-axis: 1

Specify PC on Y-axis: 2

Display 95% confidence regions:

Display sample names:

Use grey-scale colors:

X axis Y axis All

Scores Plot

2 (14.2 %)

Xia Lab @ McGill (last updated 2018-06-18)

Upload Processing Normalization Statistics Fold change T-test Volcano plot ANOVA Correlations PatternHunter PCA PLSDA sPLSDA OrthoPLSDA SAM EBAM Dendrogram Heatmap SOM K-means RandomForest SVM Download

t_test_mpp_bac_fdr....c... t_test_mpp_m9_fdr....csv Hilicpos_sample_id....t... mummichog_matc....csv t_test_cat1_n2_fdr....csv cat1_n2_m9filt_pls....png Show all

10:06 PM 6/24/2018

MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

[Overview](#)[Scree Plot](#)[2D Scores Plot](#)[3D Scores Plot](#)[Loadings Plot](#)[Biplot](#)

Specify PC on X-axis:

1

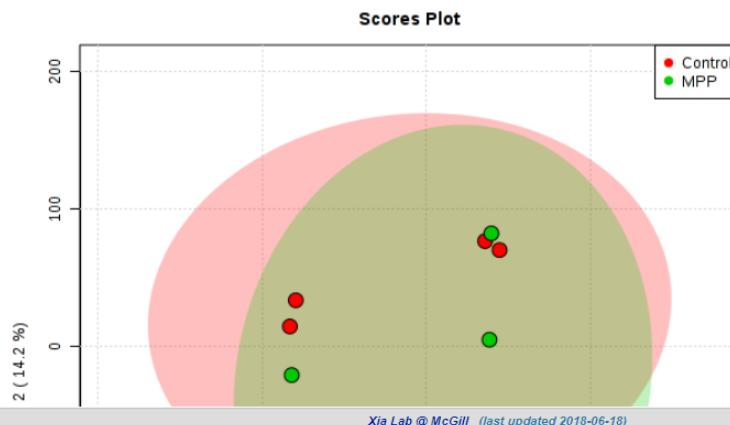
Specify PC on Y-axis:

2

Display 95% confidence regions:

Display sample names:

Use grey-scale colors:

[Flip Image](#) X axis Y axis All[Update](#)

R Command History

 Keep collapsed[Save](#)

```
1. mSet<-InitDataObjects("pktable", "stat", FALSE)
2. mSet<-Read.TextData(mSet, "Replacing_with_<your_file_path>", "col", "disc")
3. mSet<-SanityCheckData(mSet)
4. mSet<-ReplaceNan(mSet);
5. mSet<-Normalizem(mSet, "NULL", "LogNorm", "NULL", ratio=FALSE, ratioNum=20)
6. mSet<-PlotNormSummary(mSet, "norm_0_", "png", 72, width=NA)
7. mSet<-PlotSampleNormSummary(mSet, "snorm_0_", "png", 72, width=NA)
8. mSet<-PCA.Anal(mSet)
9. mSet<-PlotPCAPairSummary(mSet, "pca_pair_0_", "png", 72, width=NA, 5)
10. mSet<-PlotPCAScree(mSet, "pca_scree_0_", "png", 72, width=NA, 5)
11. mSet<-PlotPCA2DScore(mSet, "pca_score2d_0_", "png", 72, width=NA, 1,2,0.95,1,0)
12. mSet<-PlotPCALoading(mSet, "pca_loading_0_", "png", 72, width=NA, 1,2,"scatter", 1);
13. mSet<-PlotPCABiplot(mSet, "pca_biplot_0_", "png", 72, width=NA, 1,2)
14. mSet<-PlotPCA3DScoreImg(mSet, "pca_score3d_0_", "png", 72, width=NA, 1,2,3, 40)
15. mSet<-PlotPCA2DScore(mSet, "pca_score2d_1_", "png", 72, width=NA, 1,2,0.95,0,0)
```

- [Upload](#)
- [Processing](#)
- [Normalization](#)
- [Statistics](#)
 - [Fold change](#)
 - [T-test](#)
 - [Volcano plot](#)
 - [ANOVA](#)
 - [Correlations](#)
 - [PatternHunter](#)
- [PCA](#)
 - [PLSDA](#)
 - [sPLSDA](#)
 - [OrthoPLSDA](#)
 - [SAM](#)
 - [EBAM](#)
 - [Dendrogram](#)
 - [Heatmap](#)
 - [SOM](#)
 - [K-means](#)
 - [RandomForest](#)
 - [SVM](#)
 - [Download](#)



MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

R Command History

Keep collapsed Save

```

1. mSet<-InitDataObjects("pktable", "stat", FALSE)
2. mSet<-Read.TextData(mSet, "Replacing_with_your_file_path", "col1", "disc");
3. mSet<-SanityCheckData(mSet)
4. mSet<-ReplaceMin(mSet);
5. mSet<-Normalization(mSet, "NULL", "Log Norm", "NULL", ratio=FALSE, ratioNum=20)
6. mSet<-PlotNormSummary(mSet, "norm_0_", "png", 72, width=NA)
7. mSet<-PlotSampleNormSummary(mSet, "sno_norm_0_", "png", 72, width=NA)
8. mSet<-PCA.Anal(mSet)
9. mSet<-PlotPCA2DPairSummary(mSet, "pca_pair_0_", "png", 72, width=NA, 5)
10. mSet<-PlotPCAScree(mSet, "pca_scree_0_0", "png", 72, width=NA, 5)
11. mSet<-PlotPCA2DScore(mSet, "pca_score2_d_0_0", "png", 72, width=NA, 1, 2, 0.95, 1, 0)
12. mSet<-PlotPCALoading(mSet, "pca_loading_0_0", "png", 72, width=NA, 1, 2, "scatter", 1);
13. mSet<-PlotPCABiplot(mSet, "pca_biplot_0_0", "png", 72, width=NA, 1, 2)
14. mSet<-PlotPCA3DScoreImg(mSet, "pca_score3d_0_0", "png", 72, width=NA, 1, 2, 3, 40)
15. mSet<-PlotPCA2DScore(mSet, "pca_score2_d_1_0", "png", 72, width=NA, 1, 2, 0.95, 0, 0)
16. mSet<-PLSR.Anal(mSet, reg=TRUE)
17. mSet<-PlotPLSPairSummary(mSet, "pls_pair_0_0", "png", 72, width=NA, 5)
18. mSet<-PlotPLS2DScore(mSet, "pls_score2_d_0_0", "png", 72, width=NA, 1, 2, 0.95, 1, 0)
19. mSet<-PlotPLS3DScoreImg(mSet, "pls_score3d_0_0", "png", 72, width=NA, 1, 2, 3, 40)

```

Display pairwise score plot for top 5 components

Component 1 22 %

Component 2 24.7 %

Component 3 11.6 %

Component 4 8.1 %

Xia Lab @ McGill (last updated 2018-06-18)

Upload Processing Normalization Statistics Fold change T-test Volcano plot ANOVA Correlations PatternHunter PCA PLSDA sPLSDA OrthoPLSDA SAM EBAM Dendrogram Heatmap SOM K-means RandomForest SVM Download

Overview 2D Scores Plot 3D Scores Plot Loadings Plot Cross Validation Imp. Features Permutation

t_test_mpp_bac_fdr....csv t_test_mpp_m9_fdr....csv Hilicpos_sample_id....t... mummichog_matc....csv t_test_cat1_n2_fdr....csv cat1_n2_m9filt_pls....png Show all

MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

R Command History

```

1. mSet<-InitDataObjects("pktable", "stat", FALSE)
2. mSet<-Read.TextData(mSet, "Replacing_with_your_file_path", "colu", "disc")
3. mSet<-SanityCheckData(mSet)
4. mSet<-ReplaceMin(mSet);
5. mSet<-Normalization(mSet, "NULL", "Log Norm", "NULL", ratio=FALSE, ratioNum=20)
6. mSet<-PlotNormSummary(mSet, "norm_0_", "png", 72, width=NA)
7. mSet<-PlotSampleNormSummary(mSet, "snorm_0_", "png", 72, width=NA)
8. mSet<-PCA.Anal(mSet)
9. mSet<-PlotPCAPairSummary(mSet, "pca_pai_r_0_", "png", 72, width=NA, 5)
10. mSet<-PlotPCAScreee(mSet, "pca_scree_0_0", "png", 72, width=NA, 5)
11. mSet<-PlotPCA2DScore(mSet, "pca_score2_d_0_0", "png", 72, width=NA, 1,2,0.95, 1,0)
12. mSet<-PlotPCALoading(mSet, "pca_loadin_g_0_0", "png", 72, width=NA, 1,2,'scatter', 1);
13. mSet<-PlotPCAbiplot(mSet, "pca_biplot_0_0", "png", 72, width=NA, 1,2)
14. mSet<-PlotPCA3DScoreImg(mSet, "pca_sco re3d_0_0", "png", 72, width=NA, 1,2,3, 40)
15. mSet<-PlotPCA2DScore(mSet, "pca_score2_d_1_0", "png", 72, width=NA, 1,2,0.95, 0,0)
16. mSet<-PLSR.Anal(mSet, reg=TRUE)
17. mSet<-PlotPLSPairSummary(mSet, "pls_pai_r_0_0", "png", 72, width=NA, 5)
18. mSet<-PlotPLS2DScore(mSet, "pls_score2_d_0_0", "png", 72, width=NA, 1,2,0.95, 1,0)
19. mSet<-PlotPLS3DScoreImg(mSet, "pls_sco re3d_0_0", "png", 72, width=NA, 1,2,3, 40)
20. mSet<-PlotPLS3DScoreImg(mSet, "pls_sco re3d_0_0", "png", 72, width=NA, 1,2,3, 40)

```

Scores Plot

Xia Lab @ McGill (last updated 2018-06-18)

Recent Files

- t_test_mpp_bac_fdr....csv
- t_test_mpp_m9_fdr....csv
- Hilicpos_sample_id....tsv
- mumimichog_matc....csv
- t_test_cat1_n2_fdr....csv
- cat1_n2_m9filt_pls....png

Taskbar

10:07 PM 6/24/2018

Not secure | www.metaboanalyst.ca/faces/Secure/analysis/TtestView.xhtml

MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

T Tests

Note, for large data set (> 1000 variables), both the paired information and the group variance will be ignored, and the default parameters will be used for t-tests to save computational time. If you choose non-parametric tests (Wilcoxon rank-sum test), the group variance will be ignored.

Analysis type: Unpaired
 Group variance: Equal
 Non-parametric tests
 Adjusted P-value (FDR) cutoff: 0.95

Submit

Click on a point to view, drag to zoom

-log10(raw P-value)

Reset

Xia Lab @ McGill (last updated 2018-06-18)

R Command History

OK
 A total of 4433 significant features were found.

```

1. mSet<-InitDataObjects("pktable", "stat", FALSE)
2. mSet<-Read.TextData(mSet, "Replacing_With_your_file_path", "colu", "disc");
3. mSet<-SanityCheckData(mSet)
4. mSet<-ReplaceMin(mSet);
5. mSet<-Normalization(mSet, "NULL", "Log Norm", "NULL", ratio=FALSE, ratioNum=20)
6. mSet<-PlotNormSummary(mSet, "norm_0", "png", 72, widthNA)
7. mSet<-PlotSampleNormSummary(mSet, "sno_rm_0", "png", 72, widthNA)
8. mSet<-PCA.Anal(mSet)
9. mSet<-PlotPCAPairSummary(mSet, "pca_pai_r_0", "png", 72, widthNA, 5)
10. mSet<-PlotPCAScreen(mSet, "pca_screen_0", "png", 72, widthNA, 5)
11. mSet<-PlotCA2DScore(mSet, "pca_score2_d_0", "png", 72, widthNA, 1,2,0.95, 1,0)
12. mSet<-PlotPCALoading(mSet, "pca_loading_0", "png", 72, widthNA, 1,2,"scatter", 1)
13. mSet<-PlotPCABiplot(mSet, "pca_biplot_0", "png", 72, widthNA, 1,2)
14. mSet<-PlotPCA3DScoreImg(mSet, "pca_score3d_0", "png", 72, widthNA, 1,2,3, 40)
15. mSet<-PlotPCA2DScore(mSet, "pca_score2_d_1", "png", 72, widthNA, 1,2,0.95, 0,0)
16. mSet<-PLSR.Anal(mSet, reg=TRUE)
17. mSet<-PlotPLSPairSummary(mSet, "pls_pai_r_0", "png", 72, widthNA, 5)
18. mSet<-PlotPLS2DScore(mSet, "pls_score2_d_0", "png", 72, widthNA, 1,2,0.95, 1,0)
19. mSet<-PlotPLS3DScoreImg(mSet, "pls_score3d_0", "png", 72, widthNA, 1,2,3, 40)
  
```

t_test_mpp_bac_fdr....csv t_test_mpp_m9_fdr....csv Hilicpos_sample_id....t... mummichog_matc....csv t_test_cat1_n2_fdr....csv cat1_n2_m9filt_pls....png Show all 10:07 PM 6/24/2018

MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

A heatmap provides intuitive visualization of a data table. Each colored cell on the map corresponds to a concentration value in your data table, with samples in rows and features/compounds in columns. You can use a heatmap to identify samples/features that are unusually high/low. [Tip 1: choose Do not re-organize samples/rows to show the natural contrast among groups \(with each group in a block\)](#) [Tip 2: choose Display top # of features ranked by t-tests to retain the most contrasting patterns](#)

Distance Measure: Euclidean

Clustering Algorithm: Ward

Color Contrast: Default

Data Source: Normalized data

Standardization: Autoscale features

View Mode : Overview

Do not reorganize: Samples

Use top: 25 T-test / ANOVA

Show cell borders

Show only group averages

Submit

Xia Lab @ McGill (last updated 2018-06-18)

R Command History

```

Keep collapsed Save
1. mSet<-InitDataObjects("pktable", "sta
t", FALSE)
2. mSet<-readTextData(mSet, "Replacing_w
ith_your_file_path", "colu", "disc");
3. mSet<-sanityCheckData(mSet)
4. mSet<-ReplaceIn(mSet)
5. mSet<-Normalization(mSet, "NULL", "Log
Norm", "NULL", ratio=FALSE, ratioNum=2
, 0)
6. mSet<-PlotNormSummary(mSet, "norm_0",
"png", 72, width=NA)
7. mSet<-PlotSampleNormSummary(mSet, "sno
rm_0", "png", 72, width=NA)
8. mSet<-PCA.Anal(mSet)
9. mSet<-PlotPCAPairSummary(mSet, "pca_pa
ir_0", "png", 72, width=NA, 5)
10. mSet<-PlotPCAScreene(mSet, "pca_screee_0
", "png", 72, width=NA, 5)
11. mSet<-PlotPCA2DScore(mSet, "pca_score2
_d_0", "png", 72, width=NA, 1,2,0.95,
1,0)
12. mSet<-PlotPCALoading(mSet, "pca_loadin
g_0", "png", 72, width=NA, 1,2,"scatt
er", 1)
13. mSet<-PlotPCABiplot(mSet, "pca_biplot_
0", "png", 72, width=NA, 1,2)
14. mSet<-PlotPCA2DScore(mSet, "pca_score2
_d_1", "png", 72, width=NA, 1,2,3,
40)
15. mSet<-PlotPCA2DScore(mSet, "pca_score2
_d_1", "png", 72, width=NA, 1,2,0.95,
0,0)
16. mSet<-PLSR.Anal(mSet, reg=TRUE)
17. mSet<-PlotPLSPairSummary(mSet, "pls_pa
ir_0", "png", 72, width=NA, 5)
18. mSet<-PlotPLS2DScore(mSet, "pls_score2
_d_0", "png", 72, width=NA, 1,2,0.95,
1,0)
19. mSet<-PlotPLS3DScoreImg(mSet, "pls_sco
re3d_0", "png", 72, width=NA, 1,2,3,
40)
20. mSet<-PlotPLSLoading(mSet, "pls_loadin
g_0", "png", 72, width=NA, 1, 2,"scat
ter", 1)
21. mSet<-PLSDA.CV(mSet, "L", 5, "Q2")
22. mSet<-PlotPLS.Classification(mSet, "pls
_cv_0", "png", 72, width=NA)
23. mSet<-PlotPLS.Imp(mSet, "pls_imp_0",
"png", 72, width=NA, "vip", "Comp. 1",
15, FALSE)
24. mSet<-PlotPLS2DScore(mSet, "pls_score2
_d_1", "png", 72, width=NA, 1,2,0.95,
0,0)
25. mSet<-ttests.Anal(mSet, F, 0.05, FALS
E, TRUE)
26. mSet<-PlotTT(mSet, "tt_0", "png", 72,
width=NA)

```

t_test_mpp_bac_fdr....csv t_test_mpp_m9_fdr....csv Hilicpos_sample_id....t... mummichog_matc....csv t_test_cat1_n2_fdr....csv cat1_n2_m9filt_pls....png Show all 10:08 PM 6/24/2018

Not secure | www.metaboanalyst.ca/faces/Secure/analysis/HeatmapView.xhtml

MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

MetaboAnalyst 4.0

- [Upload](#)
- [Processing](#)
- [Normalization](#)
- [Statistics](#)
 - [Fold change](#)
 - [T-test](#)
 - [Volcano plot](#)
 - [ANOVA](#)
 - [Correlations](#)
 - [PatternHunter](#)
 - [PCA](#)
 - [PLSDA](#)
 - [sPLSDA](#)
 - [OrthoPLSDA](#)
 - [SAM](#)
 - [EBAM](#)
 - [Dendrogram](#)
 - [Heatmap](#)
 - [SOM](#)
 - [K-means](#)
 - [RandomForest](#)
 - [SVM](#)
- [Download](#)
- [Exit](#)

Distance Measure: Euclidean

Clustering Algorithm: Ward

Color Contrast: Default

Data Source: Normalized data

Standardization: Autoscale features

Overview Detail View (< 2000 features)

Do not reorganize: Samples

Use top: 50 T-test / ANOVA

Show cell borders

Show only group averages

Submit

R Command History

```

1. mSet<-InitDataObjects("pktable", "sta
t", FALSE)
2. mSet<-Read.TextData(mSet, "Replacing_w
ith_your_file_path", "colu", "disc");
3. mSet<-SanityCheckData(mSet)
4. mSet<-ReplaceMin(mSet);
5. mSet<-Normalization(mSet, "NULL", "Log
Norm", "NULL", ratio=FALSE, ratioNum=2
0)
6. mSet<-PlotNormSummary(mSet, "norm_0",
"png", 72, widthNA)
7. mSet<-PlotSampleNormSummary(mSet, "sno
rm_0", "png", 72, widthNA)
8. mSet<-PCA.Anal(mSet)
9. mSet<-PlotPCAPaiSummary(mSet, "pca_pa
ir_0", "png", 72, widthNA, 5)
10. mSet<-PlotPCAScreen(mSet, "pca_scree
_0", "png", 72, widthNA, 5)
11. mSet<-PlotPCA2DScore(mSet, "pca_score2
d_0", "png", 72, widthNA, 1, 2, 0.05,
1, 0)
12. mSet<-PlotPCALoading(mSet, "pca_loadin
g_0", "png", 72, widthNA, 1, 2, "scatt
er", 1)
13. mSet<-PlotPCABiplot(mSet, "pca_biplot_
0", "png", 72, widthNA, 1, 2)
14. mSet<-PlotPCA2DScore(mSet, "pca_score2
d_1", "png", 72, widthNA, 1, 2, 0.05,
0, 0)
15. mSet<-PLSR.Anal(mSet, reg=TRUE)
16. mSet<-PlotPLSPaiSummary(mSet, "pls_pa
ir_0", "png", 72, widthNA, 5)
17. mSet<-PlotPLS2DScore(mSet, "pls_score2
d_0", "png", 72, widthNA, 1, 2, 0.05,
1, 0)
18. mSet<-PlotPLS3DScoreImg(mSet, "pls_sc
oreid_0", "png", 72, widthNA, 1, 2, 3,
40)
19. mSet<-PlotPLSLoading(mSet, "pls_loadin
g_0", "png", 72, widthNA, 1, 2, "scat
ter", 1)
20. mSet<-PlotPLSClassification(mSet, "pl
s_clas_0", "png", 72, widthNA)
21. mSet<-PLSDA.CV(mSet, "L", 5, "Q2")
22. mSet<-PlotPLS.Classification(mSet, "pl
s_clas_0", "png", 72, widthNA)
23. mSet<-PlotPLS.Imp(mSet, "pls_imp_0",
"png", 72, widthNA, "vip", "Comp. 1",
15, FALSE)
24. mSet<-PlotPLS2DScore(mSet, "pls_score2
d_1", "png", 72, widthNA, 1, 2, 0.05,
0, 0)
25. mSet<-Ttests.Anal(mSet, F, 0.05, FA
LSE, TRUE)
26. mSet<-PlotTT(mSet, "tt_0", "png", 72,
widthNA)

```

t_test_mpp_bac_fdr....csv t_test_mpp_m9_fdr....csv Hilicpos_sample_id....t... mummichog_matc....csv t_test_cat1_n2_fdr....csv cat1_n2_m9filt_pls....png Show all 10:08 PM 6/24/2018

www.metaboanalyst.ca/faces/ModuleView.xhtml

MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

MetaboAnalyst 4.0

Click a module to proceed, or scroll down for more details:

Targeted or untargeted metabolomics
Targeted or annotated metabolomics
Untargeted metabolomics
Multiple metabolomics data
Integrating other omics

MS Peaks to Pathways

LC-MS peaks to pathway activities

Statistical Analysis

Biomarker Analysis

Pathway Analysis

Joint Pathway Analysis

Network Explorer

Other Utilities

Spectral Analysis

Biomarker Meta-analysis

Power Analysis

Time-series / Two-factor

Statistical Analysis

Enrichment Analysis

Pathway Analysis

Joint Pathway Analysis

Network Explorer

Other Utilities

Spectral Analysis

Biomarker Meta-analysis

Power Analysis

Time-series / Two-factor

Xia Lab @ McGill (last updated 2018-06-18)

t_test_mpp_bac_fdr....csv

t_test_mpp_m9_fdr....csv

Hilicpos_sample_id....t

mummichog_matc....csv

t_test_cat1_n2_fdr....csv

cat1_n2_m9filt_pls....png

Show all

NSERC CRSNG

Genome Canada

Genome Québec

Home

Overview

Data Formats

FAQs

Tutorials

Troubleshooting

Resources

Update History

User Stats

About

Upload the t_test_mpp_bac... file

(this was downloaded from the first analysis that was performed)

www.metaboanalyst.ca/faces/upload/PeakUploadView.xhtml

MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

Upload a peak list profile

Mass Accuracy (ppm): 0.1 (editable)

Analytical Mode: Positive Mode

P-value Cutoff: 0.1 (editable)

Choose Data File: Choose File t_test_mpp_bac..._5_metab.txt

Use the example data

An example peak list data obtained from untargeted metabolomics of human monocyte-derived dendritic cells (moDC) under stimulation by a strain of yellow fever virus (YF17D, vaccine strain) collected using Orbitrap LC-MS (positive mode, human samples, p.value cutoff: 0.0001).

Submit

R Command History

Keep collapsed Save

```
1. mSet<-InitDataObjects("mass_all", "mummichog", FALSE)
2. mSet<-Read.PeakListData(mSet, "Replacing_with_your_file_path");
3. mSet<-UpdateMummichogParameters(mSet, "0.1", "positive", 0.1);
4. mSet<-SanityCheckMummichogData(mSet)
```

Xia Lab @ McGill (last updated 2018-06-18)

t_test_mpp_bac_fdr....csv t_test_mpp_m9_fdr....csv Hilicpos_sample_id.....t... mummichog_matc....csv t_test_cat1_n2_fdr....csv cat1_n2_m9filt_pls....png Show all

MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

Data Integrity Check:

1. Checking the class labels - at least three replicates are required in each class.
2. If the samples are paired, the pair labels must conform to the specified format.
3. The data (except class labels) must not contain non-numeric values.
4. The presence of missing values or features with constant values (i.e. all zeros)

Data processing information:

Checking data content ...passed

A total of 4433 input mz features were retained for further analysis

The optimal number of significant features ~300.

A total of 373 significant mz features were found based on the selected p-value cutoff: 0.1

[Missing value estimation](#)

[Skip](#)

R Command History

```
1. mSet<-InitDataObjects("mass_all", "mummichog", FALSE)
2. mSet<-Read.PeakListData(mSet, "Replacing_with_your_file_path");
3. mSet<-UpdateMummichogParameters(mSet, "0_1", "positive", 0.1);
4. mSet<-SanityCheckMummichogData(mSet)
```

Keep collapsed [Save](#)

Xia Lab @ McGill (last updated 2018-06-18)

t_test_mpp_bac_fdr....csv t_test_mpp_m9_fdr....csv Hilicpos_sample_id....t... mummichog_matc....csv t_test_cat1_n2_fdr....csv cat1_n2_m9filt_pls....png Show all

10:19 PM 6/24/2018

MetaboAnalyst 4.0

Please select a pathway library:

Mammals

- Homo sapiens (human) [MFN]
- Homo sapiens (human) [BioCyc]
- Homo sapiens (human) [KEGG]
- Mus musculus (mouse) [BioCyc]
- Mus musculus (mouse) [KEGG]
- Rattus norvegicus (rat) [KEGG]
- Bos taurus (cow) [KEGG]

Birds

- Gallus gallus (chicken) [KEGG]

Fish

- Danio rerio (zebrafish) [KEGG]
- Danio rerio (zebrafish) [MTF]

Insects

- Drosophila melanogaster (fruit fly) [KEGG]
- Drosophila melanogaster (fruit fly) [BioCyc]

Nematodes

- Caenorhabditis elegans (nematode) [KEGG]

Fungi

- Saccharomyces cerevisiae (yeast) [KEGG]
- Saccharomyces cerevisiae (yeast) [BioCyc]

Plants

- Oryza sativa japonica (Japanese rice) [KEGG]
- Arabidopsis thaliana (thale cress) [KEGG]

Parasites

- Schistosoma mansoni [KEGG]
- Plasmodium falciparum 3D7 (Malaria) [KEGG]
- Trypanosoma brucei [KEGG]

Prokaryotes

- Escherichia coli K-12 MG1655 [KEGG]
- Bacillus subtilis [KEGG]
- Pseudomonas putida KT2440 [KEGG]
- Staphylococcus aureus N315 (MRSA/VSSA) [KEGG]
- Thermotoga maritima [KEGG]
- Synechococcus elongatus PCC7942 [KEGG]
- Mesorhizobium loti [KEGG]

Submit

R Command History

```
Keep collapsed Save
1. mSet<-InitDataObjects("mass_all", "mummichog", FALSE)
2. mSet<-Read.PeakListData(mSet, "Replacing_with_your_file_path");
3. mSet<-UpdateMummichogParameters(mSet, "0.1", "positive", 0.1);
4. mSet<-SanityCheckMummichogData(mSet)
5. mSet<-PerformMummichog(mSet, "cel_kegg", "fisher", "gamma")
```

Xia Lab @ McGill (last updated 2018-06-18)

t_test_mpp_bac_fdr....csv ^ t_test_mpp_m9_fdr....csv ^ Hilicpos_sample_id....csv ^ mummichog_matc....csv ^ t_test_cat1_n2_fdr....csv ^ cat1_n2_m9filt_pls....png ^ Show all X

10:10 PM 6/24/2018

MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

R Command History

```
Keep collapsed
1. mSet<-InitDataObjects("mass_og", FALSE)
2. mSet<-Read.PeakListData(mSet, ith_your_file_path");
3. mSet<-UpdateMummichogParameter 1, "positive", 0.1);
4. mSet<-SanityCheckMummichog(mSet, "fisher", "gamma")
5. mSet<-PerformMummichog(mSet, "fisher", "gamma")
6. mSet<-PerformMummichog(mSet, "fisher", "gamma")
```

Predicted pathway activity profiles based on mummichog

Pathway Hits Compound Hits Explore Results in Network

Pathway Name	Total	Hits (all)	Hits (sig.)	Fisher's P	EASE Score	Gamma P	Details
Terpenoid backbone biosynthesis	10	6	4	0.018296	0.10327	2.8806E-4	View
Tyrosine metabolism	25	13	5	0.10715	0.26453	7.5935E-4	View
Vitamin B6 metabolism	8	6	3	0.10505	0.35506	0.0013256	View
Glycerophospholipid metabolism	24	7	3	0.1569	0.43532	0.002195	View
Glycerolipid metabolism	9	3	2	0.10906	0.49662	0.0032528	View
Aminoacyl-tRNA biosynthesis	66	19	5	0.34901	0.56528	0.0051075	View
Alanine, aspartate and glutamate metabolism	20	14	4	0.3205	0.5694	0.0052499	View
Purine metabolism	54	26	6	0.45171	0.64065	0.0085239	View
Amino sugar and nucleotide sugar metabolism	27	16	4	0.42464	0.66628	0.010198	View
Pyruvate metabolism	18	5	2	0.27404	0.68204	0.011403	View
Sulfur metabolism	7	5	2	0.27404	0.68204	0.011403	View
Glutathione metabolism	25	12	3	0.46219	0.73934	0.017327	View
Citrate cycle (TCA cycle)	20	6	2	0.35853	0.74744	0.018415	View
Cysteine and methionine metabolism	26	13	3	0.52016	0.78013	0.023678	View
Glycolysis or Gluconeogenesis	25	14	3	0.57462	0.81525	0.031396	View
Glycine, serine and threonine metabolism	22	11	2	0.69636	0.92057	0.083968	View
Arginine and proline metabolism	28	15	2	0.84836	0.96873	0.15938	View
Pyrimidine metabolism	40	26	2	0.98209	0.99767	0.36954	View
Fructose and mannose metabolism	20	13	1	0.95127	1.0	1.0	View
Galactose metabolism	13	10	1	0.9016	1.0	1.0	View

Xia Lab @ McGill (last updated 2018-06-18)

t_test_mpp_bac_fdr....csv t_test_mpp_m9_fdr....csv Hilicpos_sample_id....t... mummichog_matc....csv t_test_cat1_n2_fdr....csv cat1_n2_m9filt_pls....png

10:11 PM 6/24/2018

Bo x M Int x box Ma x G Afl x G cre x sta x Ac x KE x KE x G vn x Im x G KE x E Th x ab x G an x W Ar x n Po x M i x

www.metaboanalyst.ca/faces/Secure/mummichog/LibraryView.xhtml

MetaboAnalyst 4.0 - Statistical, Functional and Integrative Analysis of Metabolomics Data

Please select a pathway library:

Mammals

- Homo sapiens (human) [MFN]
- Homo sapiens (human) [BioCyc]
- Homo sapiens (human) [KEGG]
- Mus musculus (mouse) [BioCyc]
- Mus musculus (mouse) [KEGG]
- Rattus norvegicus (rat) [KEGG]
- Bos taurus (cow) [KEGG]

Birds

- Gallus gallus (chicken) [KEGG]

Fish

- Danio rerio (zebrafish) [KEGG]
- Danio rerio (zebrafish) [MTF]

Insects

- Drosophila melanogaster (fruit fly) [KEGG]
- Drosophila melanogaster (fruit fly) [BioCyc]

Nematodes

- Caenorhabditis elegans (nematode) [KEGG]

Fungi

- Saccharomyces cerevisiae (yeast) [KEGG]
- Saccharomyces cerevisiae (yeast) [BioCyc]

Plants

- Oryza sativa japonica (Japanese rice) [KEGG]
- Arabidopsis thaliana (thale cress) [KEGG]

Parasites

- Schistosoma mansoni [KEGG]
- Plasmodium falciparum 3D7 (Malaria) [KEGG]
- Trypanosoma brucei [KEGG]
- Escherichia coli K-12 MG1655 [KEGG]
- Bacillus subtilis [KEGG]
- Pseudomonas putida KT2440 [KEGG]
- Staphylococcus aureus N315 (MRSA/VSSA) [KEGG]
- Thermotoga maritima [KEGG]
- Synechococcus elongatus PCC7942 [KEGG]
- Mesorhizobium loti [KEGG]

Prokaryotes

Submit

R Command History

Keep collapsed Save

```
1. mSet<-InitDataObjects("mass_all", "mummichog", FALSE)
2. mSet<-Read.PeakListData(mSet, "Replacing_with_your_file_path");
3. mSet<-UpdateMummichogParameters(mSet, "0.1", "positive", 0.1);
4. mSet<-SanityCheckMummichogData(mSet)
5. mSet<-PerformMummichog(mSet, "cel_kegg", "fisher", "gamma")
6. mSet<-PerformMummichog(mSet, "cel_kegg", "fisher", "gamma")
7. mSet<-PerformMummichog(mSet, "hsa_mfn", "fisher", "gamma")
```

Xia Lab @ McGill (last updated 2018-06-18)

t_test_mpp_bac_fdr....c... t_test_mpp_m9_fdr....csv Hilicpos_sample_id....t... mummichog_matc....csv t_test_cat1_n2_fdr....csv cat1_n2_m9filt_pls....png Show all 10:11 PM 6/24/2018

MetaboAnalyst - statistical, functional and integrative analysis of metabolomics data

Predicted pathway activity profiles based on mummichog

R Command History

Keep collapsed Save

```

1. mSet<-InitDataObjects("mass_all", "mummichog", FALSE)
2. mSet<-Read.PeakListData(mSet, "Replacing_with_your_file_path");
3. mSet<-UpdateMummichogParameters(mSet, "0.1", "positive", 0.1);
4. mSet<-SanityCheckMummichogData(mSet)
5. mSet<-PerformMummichog(mSet, "cel_kegg", "fisher", "gamma")
6. mSet<-PerformMummichog(mSet, "cel_kegg", "fisher", "gamma")
7. mSet<-PerformMummichog(mSet, "hsa_mfn", "fisher", "gamma")
8. mSet<-PerformMummichog(mSet, "hsa_mfn", "fisher", "gamma")

```

Pathway Hits

Pathway Name	Total	Hits (all)	Hits (sig.)	Fisher's P	EASE Score	Gamma P	Details
Tyrosine metabolism	160	89	19	0.10041	0.1545	0.020486	View
Glycine, serine, alanine and threonine metabolism	88	41	10	0.10289	0.18915	0.022851	View
TCA cycle	31	16	5	0.096045	0.23983	0.026848	View
Keratan sulfate degradation	68	6	3	0.054968	0.24355	0.02717	View
Urea cycle/amino group metabolism	85	44	10	0.14843	0.25279	0.027986	View
Vitamin B3 (nicotinate and nicotinamide) metabolism	28	17	5	0.11921	0.27749	0.0303	View
Squalene and cholesterol biosynthesis	55	23	6	0.1456	0.29491	0.032055	View
Lysine metabolism	52	23	6	0.1456	0.29491	0.032055	View
Vitamin B5 - CoA biosynthesis from pantothenate	12	7	3	0.085129	0.30729	0.03337	View
Vitamin B6 (pyridoxine) metabolism	11	7	3	0.085129	0.30729	0.03337	View
Aminosugars metabolism	69	24	6	0.16978	0.32852	0.035761	View
Methionine and cysteine metabolism	94	44	9	0.2572	0.39646	0.044749	View
Glycolysis and Gluconeogenesis	49	38	8	0.24861	0.39724	0.044865	View
CoA Catabolism	7	3	2	0.067811	0.40443	0.045956	View
Valine, leucine and isoleucine degradation	65	28	6	0.27999	0.46399	0.056193	View
Dynorphin metabolism	8	4	2	0.12154	0.49906	0.063391	View
Butanoate metabolism	34	23	5	0.29835	0.50729	0.065227	View
Caffeine metabolism	11	11	3	0.24929	0.54065	0.073306	View
Tryptophan metabolism	94	63	11	0.42073	0.55398	0.076845	View
N-Glycan Degradation	16	5	2	0.18183	0.5787	0.08394	View

Xia Lab @ McGill (last updated 2018-06-18)

t_test_mpp_bac_fdr....csv t_test_mpp_m9_fdr....csv Hilicpos_sample_id....t... mummichog_matc....csv t_test_cat1_n2_fdr....csv cat1_n2_m9filt_pls....png Show all

10:12 PM 6/24/2018



Science
of the

ation Science Book
consider it to be a
vement not only
out of a prodigious
work of the highest
its subject matter is
to the interests of
community, but also
k's visual qualities
n value, which will
widely read beyond
field."

eatured in more
articles:

ure
journal of science
science
ne full of tools for discovery.

Katy Börner

Katy Börner is the Victor H. Yngve Professor of Information Science in the Department of Information and Library Science, School of Informatics and Computing, Adjunct Professor at the Department of Statistics in the College of Arts and Sciences, Core Faculty of Cognitive Science, Research Affiliate of the Center for Complex Networks and Systems Research and Biocomplexity Institute, Member of the Advanced Visualization Laboratory, Leader of the Information Visualization Lab, and Founding Director of the Cyberinfrastructure for Network Science Center at Indiana University in Bloomington, IN and Visiting Professor at the Royal Netherlands Academy of Arts and Sciences (KNAW) in The Netherlands. She is a curator of the international *Places & Spaces: Mapping Science* exhibit. She holds a MS in Electrical Engineering from the University of Technology in Leipzig, 1991 and a Ph.D. in Computer Science from the University of Kaiserslautern, 1997. She became an American Association for the Advancement of Science (AAAS) Fellow in 2012.

[Home](#) | [CV](#) | [Biography](#) | [Talks & Presentations](#) | [Photos](#) | [CNS Facebook](#) | [Exhibit Facebook](#)

[RESEARCH](#) | [TEACHING](#) | [SERVICE](#) | [MEDIA](#)

RESEARCH

Katy's research focuses on the development of data analysis and visualization techniques for information access, understanding, and management. She is particularly interested in the study of the structure and evolution of scientific disciplines; the analysis and visualization of online activity; and the development of cyberinfrastructures for large scale scientific collaboration and computation.

Selected Books

- 2014: *Atlas of Knowledge*, MIT Press.
- 2014: *Visual Insights: A Practical Guide to Making Sense of Data*, MIT Press. (Co-authors with David F. Pollay) (Book preview and [IVMOOC Course](#))

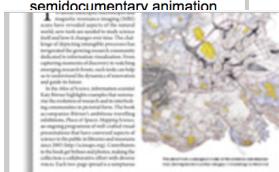
- [Google Scholar](#)
- [PubMed](#)
- [Mendeley-ReaderMeter](#)
- [US Amazon](#)
- [VIVO@IU](#)



email: katy@indiana.edu
Phone: 812-855-3256
Fax: 812-855-6166
[Contact Info](#) and [Directions](#)

UPCOMING TALKS & APPEARANCES ([see all](#))

- Apr 25, 2013 DASER talk at NAS, Washington, DC.
 Mar 22, 2013 [TEDx](#) at Buskirk-Chumley Theater in Bloomington, IN.
 Jan 24, 2013 "Visualizing What We Know" Talk using [Betazone](#) at WEF, Davos.
 Jan 11, 2013 [Humanexus](#) semidocumentary animation



Additional resources for visualizing data

Shneiderman, Ben. *Mapping science: Ben Shneiderman enjoys a tome full of tools for discovery*. *Nature*. Dec 23, 2010.

Science INFORMATION SCIENCE

Bounds and Vision

Mason A. Porter



Porter, Mason A. *Bounds and Vision*. *Science*. Feb 11, 2011

AMERICAN

- 2012: *VIVO: A Semantic Approach to Scholarly Networking and Discovery*, Morgan & Claypool Publishers. (Co-Edited with Mike Conlon, Jon Corson-Rikert, and Ying Ding)
- 2011: *Models of Science Dynamics*, Springer.
- 2010: *Atlas of Science*, MIT Press. (Images & References)
- 2009: *Data on Federal Research and Development Investments: A Pathway to Modernization*, NAS.
- 2003: *Visual Interfaces to Digital Libraries*, Springer.

Selected Papers ([see all 170+](#))

1. Skupin, André, Joseph R. Biberstine, and Katy Börner. (2013) *Visualizing the Topical Structure of the Medical Sciences: A Self-Organizing Map Approach*. *PLoS ONE* 8 (3): e58779.
2. Börner, Katy, Klavans, Richard, Patek, Michael, Zoss, Angela, Biberstine, Joseph R., Light, Robert, Larivière, Vincent, and Boyack, Kevin W.. (2012). *Design and Update of a Classification System: The UCSD Map of Science*. *PLoS One* 7(7): e39464.
3. Börner, Katy. (2011). *Plug-and-Play Macrosopes*. *Communications of the ACM*. Vol. 54(3), 60-69, ACM Press.
4. Boyack, Kevin W., Newman, David, Duhon, Russell Jackson, Klavans, Richard, Patek, Michael, Biberstine, Joseph R., Schijvenaars, Bob, Skupin, Andre, Ma, Nianli & Börner, Katy. (2011). *Clustering More Than Two Million Biomedical Publications: Comparing the Accuracies of Nine Text-Based Similarity Approaches*. *PLoS ONE*. Vol. 6(3), 1-11.
5. Börner, Katy, Contractor, Noshir S., Falk-Krzesinski, Holly J., Fiore, Stephen M., Hall, Kara L., Keyton, Joann, Spring, Bonnie, Stokols, Daniel, Trochim, William & Uzzi, Brian. (2010). *A Multi-Level Systems Perspective for the Science of Team Science*. In *Science Translational Medicine*. Vol. 2(49), 49(cm24).
6. Börner, Katy, Sanyal, Soma & Vespignani, Alessandro. (2007). *Network Science*. In Cronin, Blaise (Eds.), *Annual Review of Information Science & Technology* (Vol. 41, pp. 537-607), chapter 12, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology.
7. Börner, Katy, Chen, Chaomei & Boyack, Kevin W. (2003). *Visualizing Knowledge Domains*. In Cronin, Blaise (Eds.), *Annual Review of Information Science & Technology* (Vol. 37, pp. 179-255), chapter 5, American Society for Information Science and Technology, Medford, NJ.
8. Börner, Katy, Maru, Jeegaa & Goldstone, Robert. (2004). *The Simultaneous Evolution of Author and Paper Networks*. *Proceedings of the National Academy of Sciences of the United States of America*. Vol. 101, 5266-5273.
9. Mane, Ketan K. & Börner, Katy. (2004). *Mapping Topics and Topic Bursts in PNAS*.

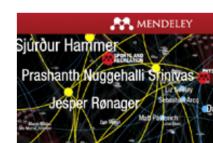
Spring Data Visualization Liter
Workshop @ CNS, IUE
Nov Plug-and-Play Macros
3-4 Workshop @ CNS, IUE
Nov CNS Open House.

Talk Series on Network Complex Systems and Brown Bag Talk Se

VISUALIZATIONS

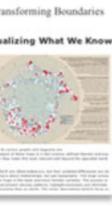


NARCIS: Network of Experts & Organizations in the Netherlands, for Data Archiving and Networked Services (DANS) Netherlands Academy of Sciences (KNAW), by Linda R. Michael J. Stamper, Katy Börner, Baars, and Andrea Scharr



Chin Hua Kong, Katy Börner

Proceedings of the National Academy of Sciences of the United States of America. Vol. 101, 5287-5290.



J. Visualizing
transforming
American Scientist.
2011.

Katy Börner: Atlas of
what we know. June
11.

Classroom Use in
ge Visualization by
in, SDSU
Schedule)

see all)

ork Times

Seeing Farther:
Alter Science

a darkened theater a

Edited Journals

- Börner, Katy, Glänzel, Wolfgang, Scharnhorst, Andrea & van den Besselaar, Peter (Eds.), *Modeling science: studying the structure and dynamics of science*. (2011). *Scientometrics*, Springer.
- Börner, Katy & Scharnhorst, Andrea (Eds.), *Visual Conceptualizations and Models of Science*. (2009). *Special Issue on the Science of Science: Conceptualizations and Models of Science, Journal of Informetrics*, Vol. 3(3), Elsevier.
- Skupin, Andre & Börner, Katy (Eds.), Special Issue on Mapping Humanity's Knowledge and Expertise in the Digital Domain. (2007). *Environment and Planning B: Planning and Design*, Vol. 34(5), Pion.
- Börner, Katy & Navarro-Prieto, Raquel (Eds.), *Special Issue on Collaborative Information Visualization Environments*. (2005). *PRESENCE*, Vol. 14(1), MIT Press.
- Börner, Katy & Mostafa, Javed (Eds.), Special Issue on Information Visualization Interfaces for Retrieval and Analysis (2005), *International Journal on Digital Libraries*, Vol. 5(1), Springer-Verlag.
- Shiffrin, Richard & Börner, Katy (Eds.), *Mapping Knowledge Domains*. (2004). *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 101(Suppl. 1).

Selected Presentations (see all)



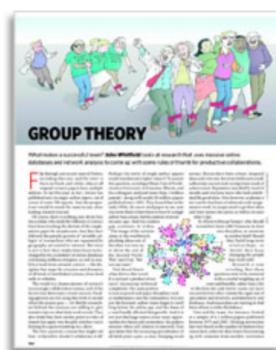
Katy Börner's TEDx Bloomington talk on *Maps & Macroscopes -- Gaining Insights from BIG Data*. March 2013.



Katy Börner is featured in *FuturICT - New Science and Techno Manage Our Complex, Connected World*. Oct 2011.



Katy Börner presents *Plug-and-Play Macroscopes*. Communications ACM, March 2011.



Whitfield, John. (2008). *Group Theory*. In *Nature*. Vol. 455, 720-723.

WIRED SCIENCE
NEWS FOR YOUR NEURONS

Michael J. Stamper (2011) Mendeley's Evolving Network of Expertise and Knowledge, 2011 Mendeley Binary Battle.



Joseph R. Biberstine, Katy Börner, and Michael J. Stamper (2011) Sources and Sinks of Life Time in U.S. Air Travel, Sunbelt Conference.



From: www.katayb.com



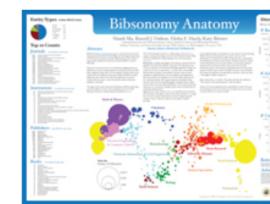
From: www.katayb.com



Katy Börner presents *Scholarly Data, Network Science and (Google) Maps* at Google, Inc., Mountain View California. Jan 2007.



Joseph Biberstine, Russell J. Duhon, Katy Börner, Elisha Hardy, and Skupin (2010) *A Semantic Lands Self-Organizing Map*, Sunbelt Conference.



Nianli Ma, Russell J. Duhon, Elisha Hardy, Katy Börner (2009) *Bibsonomy Anatomy*, Sunbelt Conference.



Weixia (Bonnie) Huang, Russell J. Duhon, Elisha F. Hardy, Katy Börner (2008) *Research Collaborations Chinese Academy of Sciences, CAS, Display at NSLC*, CAS, China

Exhibit

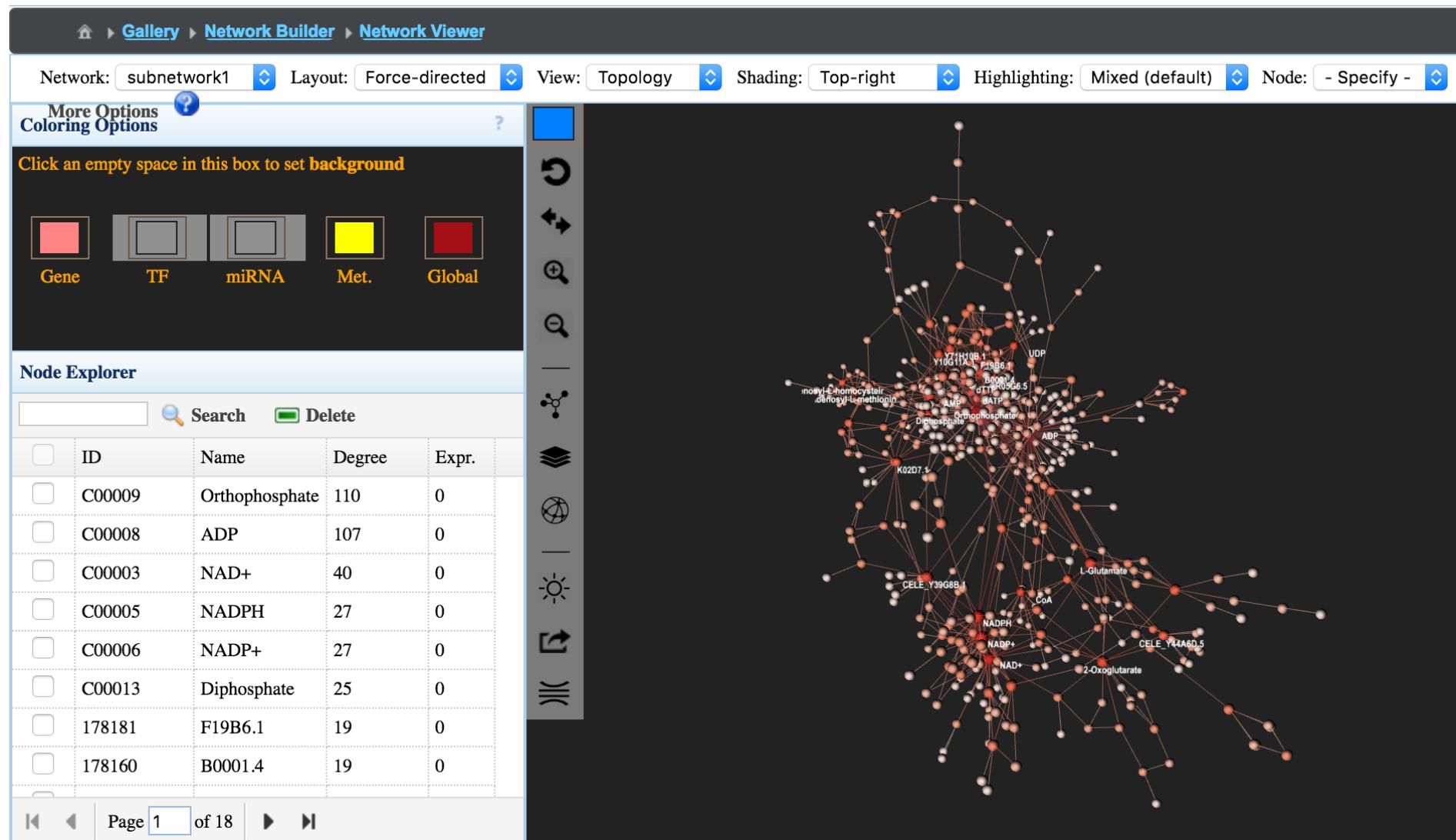
The international *Places & Spaces: Mapping Science* exhibit was created in 2005 to introduce maps of science in support of the navigation, management, understanding, and communication of data, knowledge, and expertise. As of 2011, the exhibit features 70 maps by 189 authors from 11 countries. Each year, a new themed set of 10 maps is added via an open call for maps and peer review by the exhibit advisory board and invited experts. Over the last 7 years, the exhibit has been on display at more than 200 venues in 19 countries on 6 continents.

Data

VIVO National Researcher Network: CNS implemented all social network visualizations in support of researcher networking and analysis

- Visualizations in VIVO
- Hands-on Tutorial
- Explore the Growing Network





Network: subnetwork1 Layout: Force-directed View: Topology Shading: Top-right Highlighting: Mixed (default) Node: - Specify - E

More Options
Coloring Options ?

Click an empty space in this box to set background

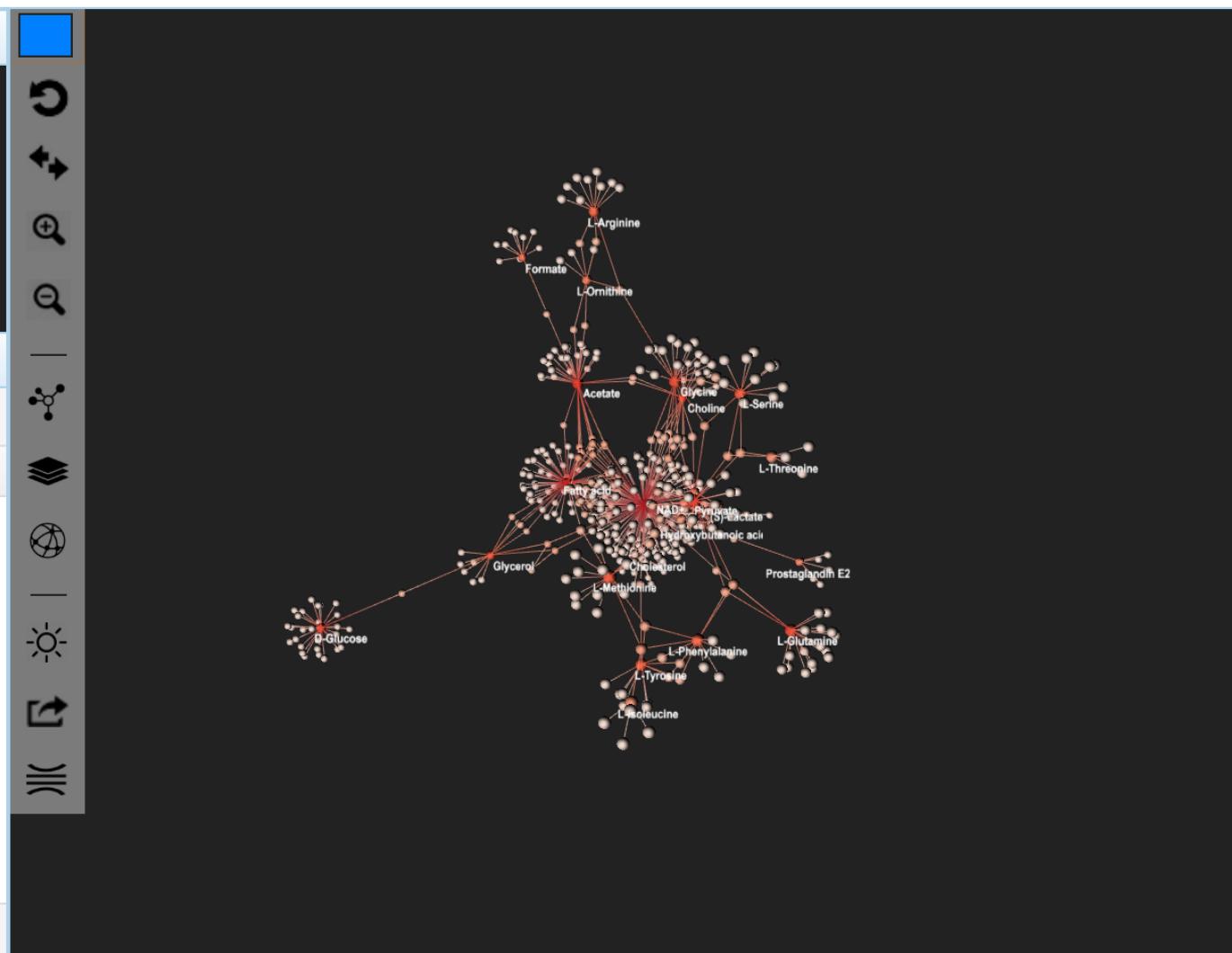


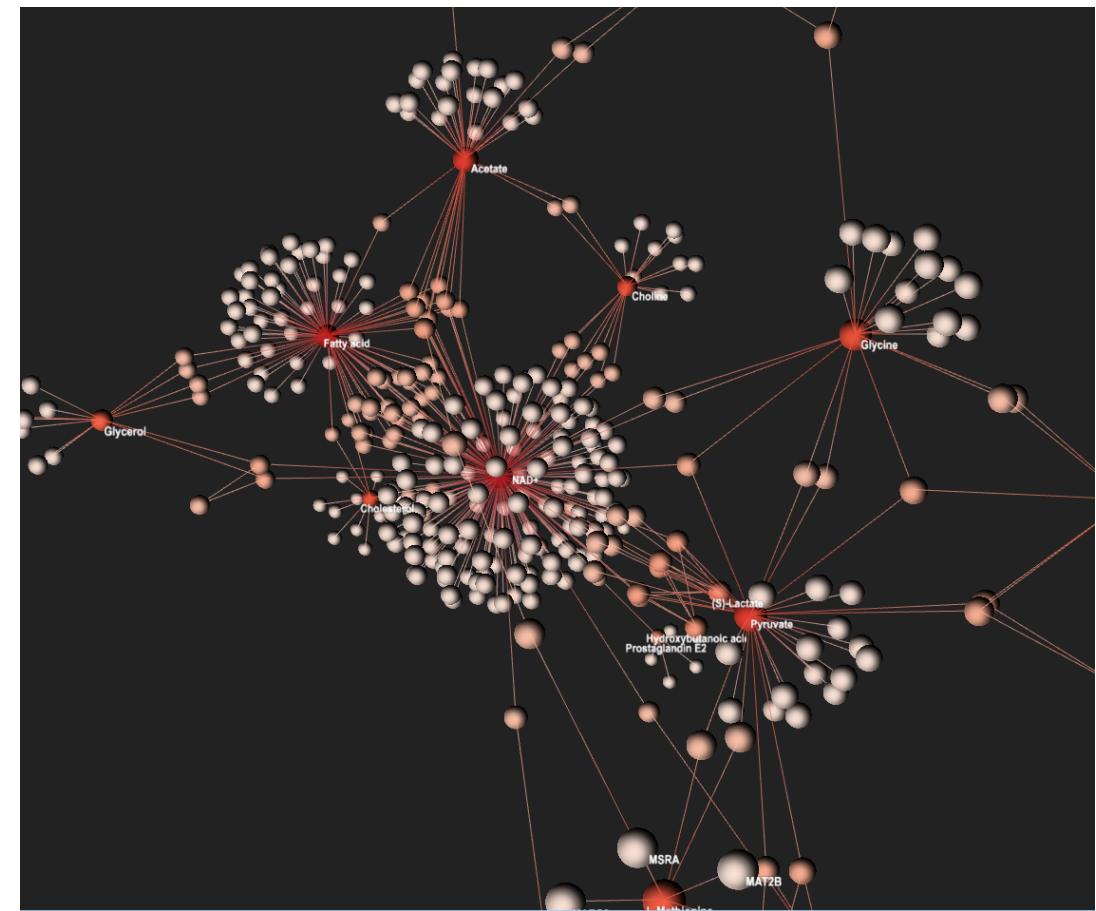
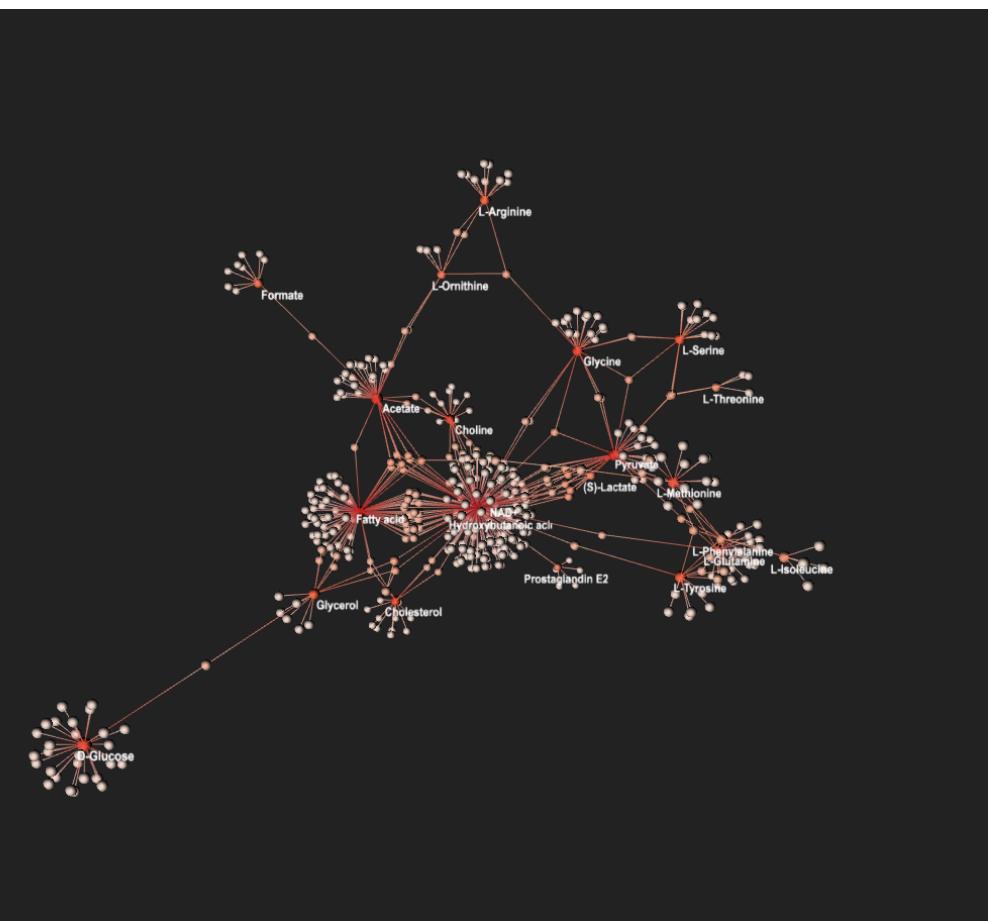
Node Explorer

Search Delete

	ID	Name	Degree	Expr.
<input type="checkbox"/>	C00003	NAD+	189	0.25225
<input type="checkbox"/>	C00162	Fatty acid	82	0.3633
<input type="checkbox"/>	C00033	Acetate	35	0.82219
<input type="checkbox"/>	C00022	Pyruvate	35	-0.6238
<input type="checkbox"/>	C00031	D-Glucose	28	0.54804
<input type="checkbox"/>	C00037	Glycine	22	0.69285
<input type="checkbox"/>	C00064	L-Glutamine	19	-0.4972
<input type="checkbox"/>	C00114	Choline	19	1.1021

Page 1 of 16





Additional tools

Human Metabolome Database (HMDB)

metlin

superGrouper

ummichog

Cytoscape

msPANDA

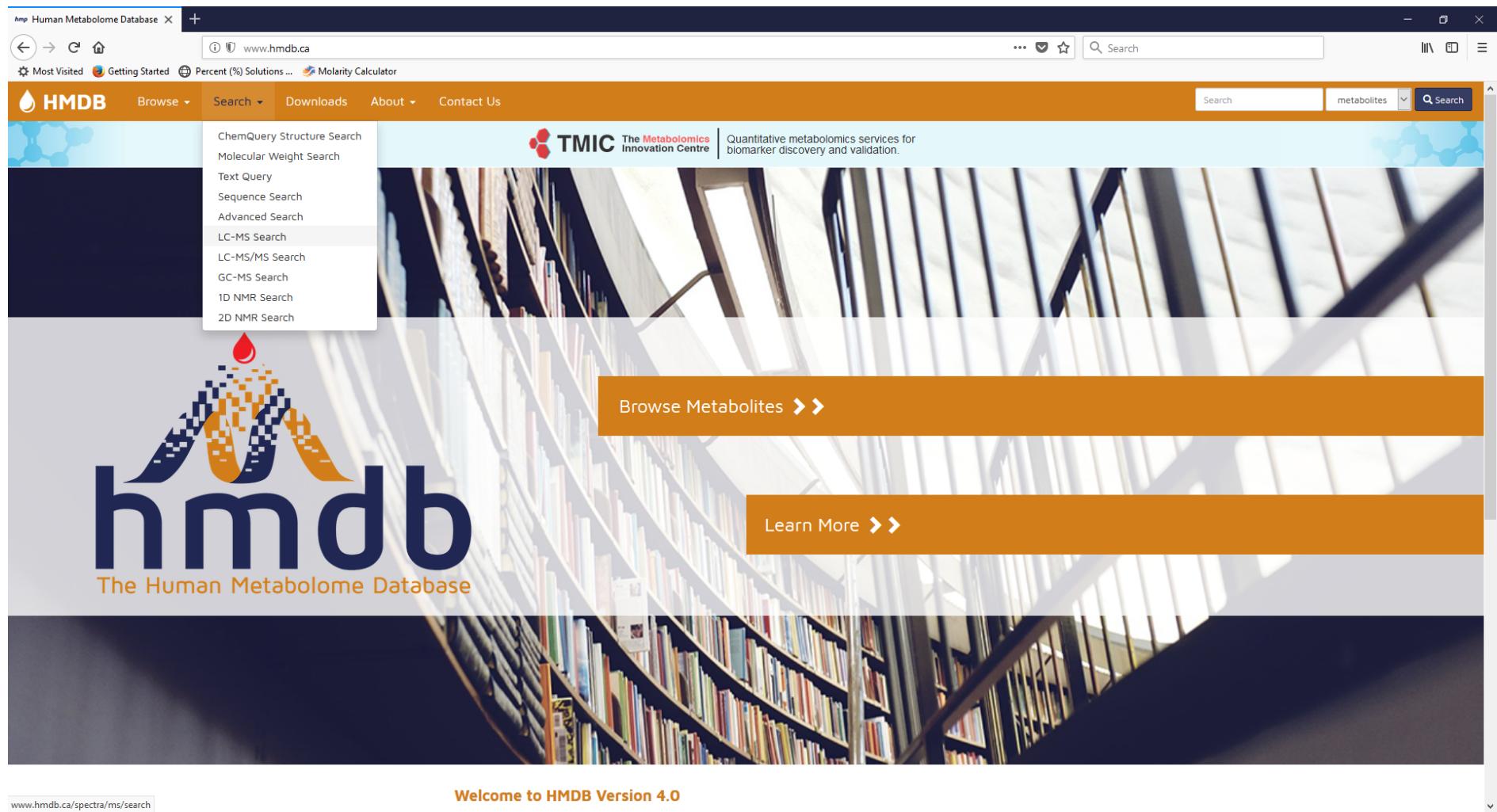
omicsNet

<http://clinicalmetabolomics.org/init/default/software>

links to various Emory-developed programs

Metabolite Identification

1. Go to hmdb.ca and click Search -> LC-MS Search



Metabolite Identification

2. Copy in m/z value. For MPP+ data, select positive ion mode, M+H and M+Na adducts, and \pm 10 ppm

The screenshot shows the HMDB spectra search interface. At the top, there is a navigation bar with links for 'Human Metabolome Database', 'Most Visited', 'Getting Started', 'Percent (%) Solutions ...', and 'Molarity Calculator'. Below the navigation bar is a header with the 'HMDB' logo, a search bar containing 'metabolites', and a 'Search' button. To the right of the search bar is a 'TMIC' logo with the text 'The Metabolomics Innovation Centre' and a subtitle 'Your source for quantitative metabolomics technologies and bioinformatics.'

The main search area is titled 'Spectra Search Mass Spectrum'. It features several search tabs: 'LC-MS Search' (selected), 'LC-MS/MS Search', 'GC-MS Search', '1D NMR Search', and '2D NMR Search'. Below these tabs is a 'Query Masses (Da)' input field containing the value '296.168720645931'. A note next to the input field says 'Enter one mass per line (maximum 700 query masses per request)'.
Under the 'Ionization' section, the 'Ion Mode' is set to 'Positive'. The 'Adduct Type' dropdown menu is open, showing options: Unknown, M+H, M-2H2O+H, M-H2O+H, M-H2O+NH4, M+Li, M+NH4, and M+Na. The 'M+H' and 'M+Na' options are highlighted with blue selection bars.
At the bottom left, there is a 'Molecular Weight Tolerance \pm ' input field with '10' and 'ppm' selected.

Metabolite Identification

3. Click on the link under “Compound” for more information on each metabolite.

Spectra Search Mass Spectrum

LC-MS Search LC-MS/MS Search GC-MS Search 1D NMR Search 2D NMR Search

▼ Search options

Search Results

MS search for 296.168720645931 m/z

Download Results As CSV

Delta = (abs(query mass - adduct mass)/adduct mass)*1000000

Compound	Name	Formula	Monoisotopic Mass	Adduct	Adduct M/Z	Delta (ppm)
HMDB0028722	Arginyl-Valine	C11H23N5O3	273.1801	M+Na	296.1693	2 m/z calculator
HMDB0029121	Valyl-Arginine	C11H23N5O3	273.1801	M+Na	296.1693	2 m/z calculator
HMDB0042026	Tertatolol	C16H25NO2S	295.1606	M+H	296.1679	3 m/z calculator

Showing 1 to 3 of 3 entries

Previous 1 Next

This project is supported by the Canadian Institutes of Health Research (award #111062), Alberta Innovates - Health Solutions, and by The Metabolomics Innovation Centre (TMIC), a nationally-funded research and core facility that supports a wide range of cutting-edge metabolomic studies. TMIC is funded by Genome Alberta, Genome British Columbia, and Genome Canada, a not-for-profit organization that is leading Canada's national genomics strategy with \$900 million in funding from the federal government.



Metabolite Identification

1. Go to metlin.scripps.edu and click Metlin-> Simple

The screenshot shows the homepage of the METLIN database at metlin.scripps.edu/landing_page.php?pgcontent=mainPage. The top navigation bar includes links for Home, Metlin (with dropdown), isoMETLIN, Sign Up, Forgot Password?, Enter email address, Enter password, and Login. A search bar is also present. A dropdown menu is open under the 'Metlin' link, showing options: Simple, Advanced, Batch, Fragment, Neutral Loss, and MS/MS Spectrum Match. The main background features a complex network graph and several chemical structures and spectra. Text overlays include 'The original and most comprehensive MS/MS metabolite database', 'Latest News and Articles', and 'Analytical Chemistry 2018 - METLIN: A Technology Platform for Identifying Knowns and Unknowns'. Below the main content area are three sections: 'Metabolite Searching', 'Tandem Mass Spectrometry', and 'Metabolites'.

Simple

Advanced

Batch

Fragment

Neutral Loss

MS/MS Spectrum Match

METLIN

The original and most comprehensive MS/MS metabolite database

Latest News and Articles

Analytical Chemistry 2018 - METLIN: A Technology Platform for Identifying Knowns and Unknowns

Metabolite Searching

METLIN has multiple searching capabilities including single, batch, precursor ion, neutral loss, accurate mass, and fragment searches. The popular [similarity search algorithm](#) for unknowns

Tandem Mass Spectrometry

METLIN represents the largest MS/MS collection of data with the database generated at multiple collision energies and in positive and negative ionization modes. The data is generated on

Metabolites

Created in 2003, METLIN now includes over a million molecules ranging from lipids, steroids, plant & bacteria metabolites, small peptides, carbohydrates, exogenous drugs/metabolites,

10550 North Torrey Pines Road BCC-007, La Jolla, CA 92037 USA • (858) 784-5415, Fax (858) 784-9496

[Home](#) | [Privacy Policy](#) | [Terms of Use](#) | [Contact Us](#)

Metabolite Identification

2. You will be prompted to type in your username and password. Set up an account if you have not done so already.

The screenshot shows the homepage of the METLIN website (https://metlin.scripps.edu/landing_page.php). The page has a dark background with a network graph overlay. A central modal dialog box is displayed, prompting the user to log in with their XCMSOnline username and password. The dialog contains the text "Please log in using your XCMSOnline's username/password to run Simple Search!" and an "OK" button. Below the dialog, the METLIN logo is visible, along with a list of metabolite names such as PYRINE, TRYPTOPHAN, PHOSPHATE, CHOLESTEROL, TESTOSTERONE, GLUCOSE, and SERIN. At the bottom of the page, there are three main sections: "Metabolite Searching", "Tandem Mass Spectrometry", and "Metabolites".

Please log in using your XCMSOnline's username/password to run Simple Search!

OK

PYRINE TRYPTOPHAN PHOSPHATE CHOLESTEROL TESTOSTERONE GLUCOSE
TESTO GLUCOSE CHOLESTEROL ACYLAMINO ACYLAMINO
PYRU' GLUCOSE CHOLESTEROL CHOLINE CHOLINE CHOLINE
GLUC NICO SERIN PYRU'
TESTOSTERONE GLUCOSE PHOSPHATE CHOLESTEROL OXALOSUCCINIC ACID GALACTOSE
GLUCOSE CHOLESTEROL OXALOSUCCINIC ACID GALACTOSE FUMARIC ACID
NICO SERIN TRYPTOPHAN PHOSPHOCHELINE ADENINE DINUCLEOTIDE OXALOSUCCINIC ACID GALACTOSE
SERINE TRYPTOPHAN PHOSPHOCHELINE ACYLAMINO ACYLAMINO THIOPHENE GLYCOSIDE

The original and most comprehensive MS/MS metabolite database

Latest News and Articles

Analytical Chemistry 2018 - METLIN: A Technology Platform for Identifying Knowns and Unknowns

Metabolite Searching

METLIN has multiple searching capabilities including single, batch, precursor ion, neutral loss, accurate mass, and fragment searches. The popular [similarity search algorithm](#) for unknown

Tandem Mass Spectrometry

METLIN represents the largest MS/MS collection of data with the database generated at multiple collision energies and in positive and negative ionization modes. The data is generated on

Metabolites

Created in 2003, METLIN now includes over a million molecules ranging from lipids, steroids, plant & bacteria metabolites, small peptides, carbohydrates, exogenous drugs/metabolites,

10550 North Torrey Pines Road BCC-007, La Jolla, CA 92037 USA, (858) 784-3415, Fax (858) 784-9496

Metabolite Identification

3. Copy in m/z value. For MPP+ data, select \pm 10 ppm, positive ion mode, and M+H and M+Na adducts. Search.

The screenshot shows the METLIN website interface. On the left, there is a "Simple Search" form with the following parameters:

- Mass: 296.168720645931
- Tolerance: 10 PPM
- Charge: Positive
- Adducts: M+H, M+NH4, M+Na
- Peptides: Add Peptides to Search
- Toxicants: Add Toxicants to Search

Below the search form are two buttons: "Search" (green) and "Clear".

The main area features a large, dark background with a network of interconnected nodes. Overlaid on this network is a large, semi-transparent watermark containing the word "METLIN" in large blue letters, surrounded by various metabolite names such as CHOLINE, SERINE, PYRUVIC ACID, TESTOSTERONE, GLUCOSE, NICOTINAMIDE, and TRYPTOPHAN. Below the watermark, the text "The original and most comprehensive MS/MS metabolite database" is visible. In the bottom left corner of the main area, there is a small inset showing a mass spectrum plot with several peaks.

At the bottom of the page, there are three sections with descriptive text and links:

- Metabolite Searching**: METLIN has multiple searching capabilities including single, batch, precursor ion, neutral loss, accurate mass, and fragment searches. The popular [similarity search algorithm](#) for unknown
- Tandem Mass Spectrometry**: METLIN represents the largest MS/MS collection of data with the database generated at multiple collision energies and in positive and negative ionization modes. The data is generated on
- Metabolites**: Created in 2003, METLIN now includes over a million molecules ranging from lipids, steroids, plant & bacteria metabolites, small peptides, carbohydrates, exogenous drugs/metabolites,

At the very bottom of the page, the address 10550 North Torrey Pines Road BCC-007, La Jolla, CA 92037 USA - (858) 784-9415, Fax (858) 784-9496 is listed.

Acknowledgements

- Rachel Cliburn Branco, Ph.D.
 - Vrinda Kalia, MPH
 - William Elsworth (high school student)
 - Josh Bradner, M.S.
-
- Dean Jones, Ph.D., Emory University
 - Doug Walker, Ph.D., Mt. Sinai (former Emory postdoc)
 - Karan Uppal, Ph.D. Emory University
 - Megan Niedzwiecki, Ph.D., Mt. Sinai (former Emory postdoc)
 - Shuzhao Li, Ph.D. Emory University