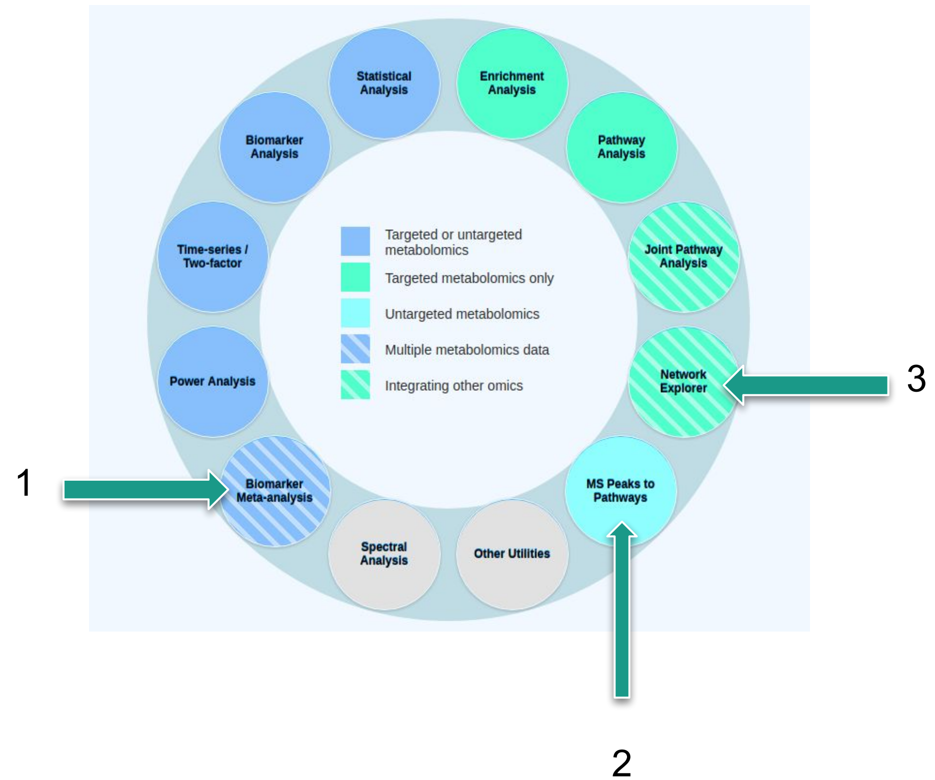
MetaboAnalyst 4.0 Tutorial

Overview of new modules: Biomarker meta-analysis, MS Peaks to Pathways, and Network Explorer

Goal for this tutorial

To introduce users to the 3 new modules in MetaboAnalyst Version 4.0:

1) Biomarker Meta-Analysis
2) MS Peaks to Pathway
3) Network Explorer



1) What is Biomarker Meta-Analysis?

- The combination of multiple independent studies investigating the same condition in similar populations is termed “horizontal integration” or “meta-analysis”.
- Leverages the collective power of multiple studies to overcome noise, bias, and small effect sizes to improve the precision in identifying true patterns within data.
- In metabolomics, biomarker validation is challenging due to inconsistencies in identified biomarkers amongst similar experiments.
- Solution: Performing meta-analysis across similar studies will increase the sample size and the power to identify robust and precise biomarkers of disease.
- Therefore the aim of the **Biomarker Meta-Analysis** module is the integration of individual metabolomic studies to identify consistent and robust biomarkers of disease.

Steps for Biomarker Meta-Analysis

1. Users must upload individual datasets in tabular form.
2. Differential enrichment analysis is performed to compute summary level-statistics for each feature (e.g. p-value) for each individual study.
3. The summary level-statistical results from all studies are combined, and meta-analysis is performed using one of several statistical options: **combining p-values**, **vote counting**, or **direct merging** of data into a mega-dataset.
4. The results can be visualized as a Venn diagram to view all possible combinations of shared features between the datasets.

MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation

Home

Overview

Data Formats

FAQs

Tutorials

Troubleshooting

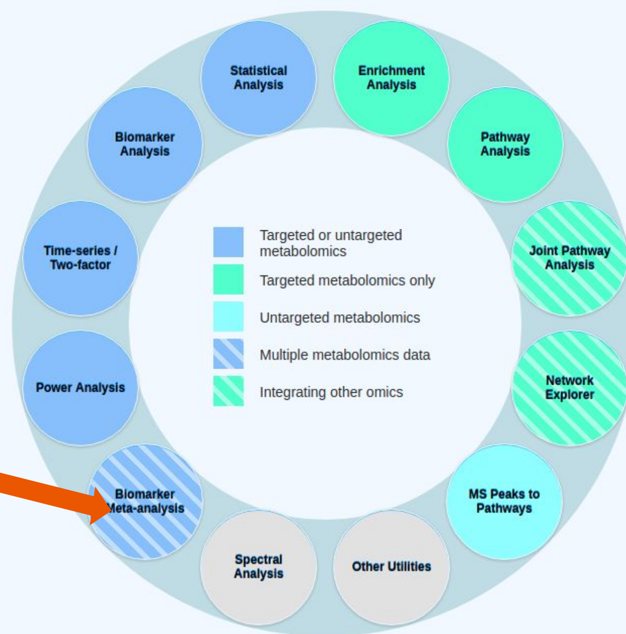
Resources

Update History

User Stats

About

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



Xia Lab @ McGill (last updated 2018-02-02)

Biomarker Meta-Analysis: Data Preparation

Prior to uploading the data, the user must clean the datasets in a spreadsheet program like Excel:

- At least 25% of features must be consistent between all datasets (named compounds, spectral bins, or peaks).
- Metadata must be consistent across all studies (e.g. Cancer vs Control labels for all datasets).
- Sample identifiers must be unique across all studies.

	A	B	C	D	E	F	G	H	I	J
1	Samples	140225dlvsa44_1	140226dlvsa30_1	140226dlvsa36_1	140227dlvsa36_1	140225dlvsa25_1	140225dlvsa33_1	140225dlvsa49_1	140226dlvsa12_1	140226dlvsa14_1
2	Class	Adenocarcinoma	Adenocarcinoma	Adenocarcinoma	Adenocarcinoma	Control	Control	Control	Control	Control
3	1_5-anhydroglucitol	6799	17473	38267	12027	10561	19520	7778	23835	7063
4	1-monoolein	165	411	525	726	459	350	1221	673	158
5	1-monopalmitin	107	100	195	122	86	210	189	190	96
6	1-monostearin	67	125	209	200	182	223	139	184	70
7	2_3_5-trihydroxypyrazine NIST	34	54	45	107	49	73	84	58	20
8	2_3-dihydroxybutanoic acid NIST	74	146	183	152	58	115	115	184	17
9	2-deoxyerythritol	334	765	474	495	240	534	535	532	277
10	2-deoxytetronic acid NIST	762	1830	1356	1128	345	1670	1895	1248	597
11	2-hydroxybutanoic acid	7786	15277	7794	9810	1816	15962	17303	10140	8135
12	2-hydroxyglutaric acid	233	1274	1021	1533	979	1785	2182	1274	232
13	2-hydroxyhippuric acid	84	82	107	109	784	153	80	93	118
14	2-hydroxyvaleric acid	1392	1047	876	1170	926	1020	1430	934	1257
15	2-ketoisocaproic acid	2094	1229	1021	2840	708	1609	1699	1275	2247
16	3-aminoisobutyric acid	891	473	362	387	742	507	1105	553	735
17	3-hydroxybutanoic acid	5015	2336	1509	4757	905	4786	4152	3013	5299
18	3-phosphoglycerate	131	61	52	980	114	86	110	84	112
19	4-hydroxyproline	1389	5112	4323	4082	5184	2426	4082	1942	1236
20	5-hydroxynorvaline NIST	197	142	264	178	1065	272	409	223	227
21	5-methoxytryptamine	314	304	246	83	178	269	389	118	246
22	acetophenone NIST	299	971	801	1732	1206	1037	1105	779	442

Example dataset highlighting class labels and unique sample identifiers

Biomarker Meta-Analysis: Data Format

Datasets must be in tabular form and uploaded individually:

- Concentration table, spectral binned data, or a peak intensity table.
- Tables may either be in .csv or .txt format
- Class labels must be present, and only 2 classes are accepted (i.e. Cancer vs. Healthy)

	A	B	C	D	E	F	G	H	I	J
1	Samples	140225dlvsa44_1	140226dlvsa30_1	140226dlvsa36_1	140227dlvsa36_1	140225dlvsa25_1	140225dlvsa33_1	140225dlvsa49_1	140226dlvsa12_1	140226dlvsa14_1
2	Class	Adenocarcinoma	Adenocarcinoma	Adenocarcinoma	Adenocarcinoma	Control	Control	Control	Control	Control
3	1_5-anhydroglucitol	6799	17473	38267	12027	10561	19520	7778	23835	7063
4	1-monoolein	165	411	525	726	459	350	1221	673	158
5	1-monopalmitin	107	100	195	122	86	210	189	190	96
6	1-monostearin	67	125	209	200	182	223	139	184	70
7	2_3_5-trihydroxypyrazine NIST	34	54	45	107	49	73	84	58	20
8	2_3-dihydroxybutanoic acid NIST	74	146	183	152	58	115	115	184	17
9	2-deoxyerythritol	334	765	474	495	240	534	535	532	277
10	2-deoxytetronic acid NIST	762	1830	1356	1128	345	1670	1895	1248	597
11	2-hydroxybutanoic acid	7786	15277	7794	9810	1816	15962	17303	10140	8135
12	2-hydroxyglutaric acid	233	1274	1021	1533	979	1785	2182	1274	232
13	2-hydroxyhippuric acid	84	82	107	109	784	153	80	93	118
14	2-hydroxyvaleric acid	1392	1047	876	1170	926	1020	1430	934	1257
15	2-ketoisocaproic acid	2094	1229	1021	2840	708	1609	1699	1275	2247
16	3-aminoisobutyric acid	891	473	362	387	742	507	1105	553	735
17	3-hydroxybutanoic acid	5015	2336	1509	4757	905	4786	4152	3013	5299
18	3-phosphoglycerate	131	61	52	980	114	86	110	84	112
19	4-hydroxyproline	1389	5112	4323	4082	5184	2426	4082	1942	1236
20	5-hydroxynorvaline NIST	197	142	264	178	1065	272	409	223	227
21	5-methoxytryptamine	314	304	246	83	178	269	389	118	246
22	acetophenone NIST	299	971	801	1732	1206	1037	1105	779	442

Example dataset highlighting class labels and unique sample identifiers

MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation

- Upload
- Meta analysis
- Result table
- Venn diagram
- Download
- Exit

Use the panel below to upload and prepare each individual data. Click the individual cell to activate each process. Click **Add New** to add a new data set. The maximum total number of samples allowed is **1000**. When all data sets have been processed, Click **Proceed** to proceed. Click the **Try Examples** button if you want to use example datasets to explore the functions available.

Data Upload	Sanity Check	Visualization	Normalization	DE Analysis	Data Summary	Include
Upload	Process	View	Normalize	Analyze	Detail	✔

Add New

Adjust study batch effect

1. Upload each dataset individually
2. Perform sanity check on data
3. Visualize data as box-plots
4. Normalize data
5. Perform DE analysis
6. View summary of DE analysis
7. Click to include in meta-analysis or not

R Command History

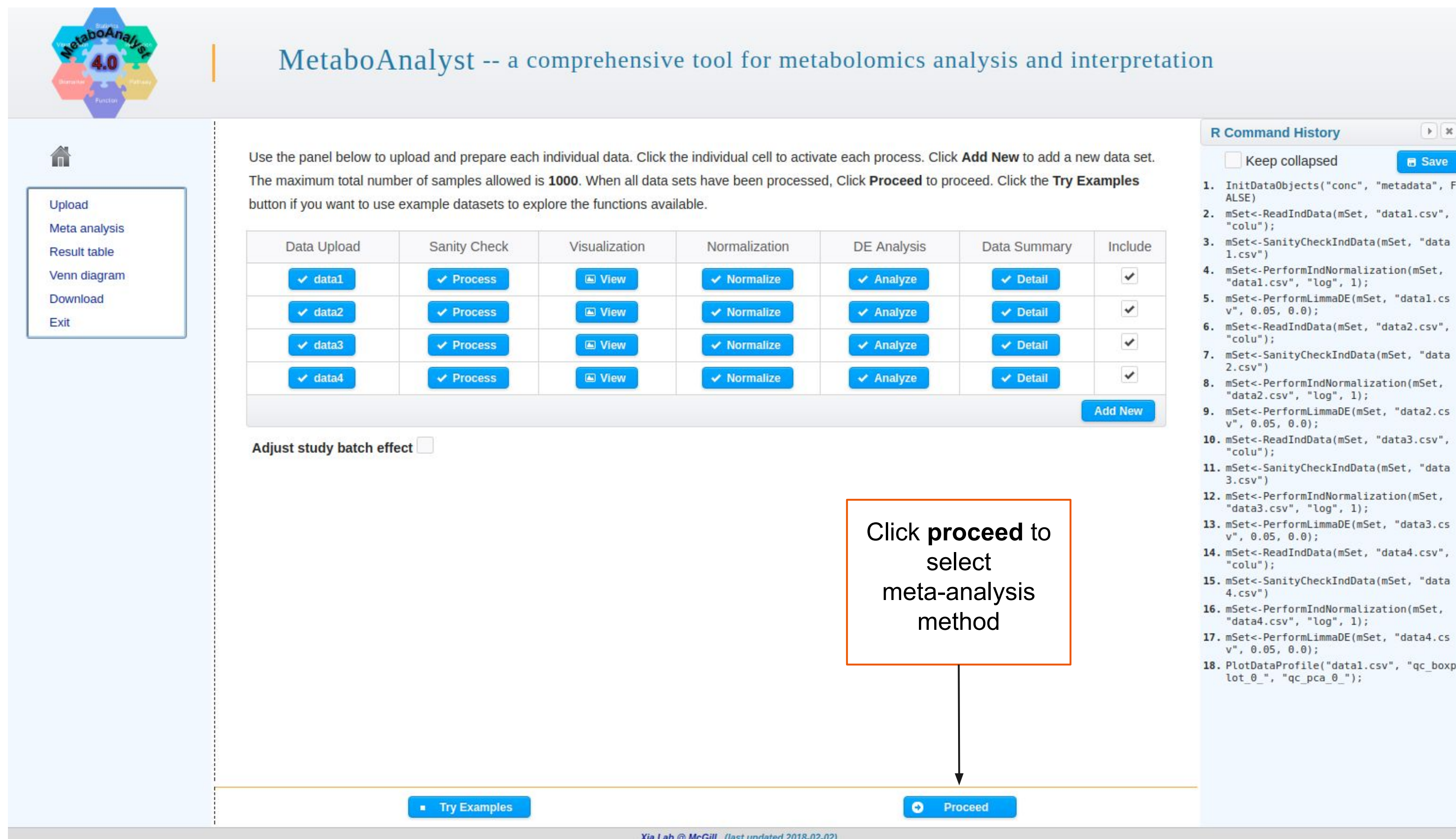
Keep collapsed   Save

Try our example data here

Try Examples   Proceed

Xia Lab @ McGill (last updated 2018-02-02)

Screenshot using the example data



R Command History appears real-time and is ordered sequentially

Screenshot of meta-analysis methods

MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation

- Upload
- Meta analysis
- Result table
- Venn diagram
- Download
- Exit

1. Select **one** of three methods to perform meta-analysis

Combining P Values

There are two widely used methods to combine p values from multiple studies for information integration - the Fisher's method (-2*∑Log(p)) and the Stouffer's method (based on inverse normal transformation). Stouffer's method incorporates weight (i.e. based on sample sizes) into the calculation; while Fisher's method is known as a 'weight-free' method. They usually have very similar performance. However, in metabolomic meta-analysis, larger sample sizes do not warrant larger weights as the quality of each study can vary. Users should choose to apply Stouffer's method only when all studies are of similar qualities (i.e. same analytical platforms with similar levels of missing values).

Select a method: Fisher's method

Set a significance level: 0.05

Submit

Vote Counting

This is the simplest method to perform meta-analysis. Differentially expressed metabolites are first selected based on a threshold to obtain a list of significant features from each study. The vote for each feature can then be calculated as the total number of times it appears as significant in all features lists. The final significant features can be selected based on the **minimal number of votes** set by the user.

Set a significance level: 0.05

Set the minimal number of votes: 2

Submit

Direct Merging

This approach directly merges all datasets into a mega-dataset and then analyzes it as a single dataset. It should only be used when all datasets are very similar (i.e. collected by the same lab using the same analytical platforms)

Set a significance level: 0.05

Submit

Previous | Proceed

2. Click **proceed** to view results

R Command History

Keep collapsed | Save

```
1.  InitDataObjects("conc", "metadata", FALSE)
2.  mSet<-ReadIndData(mSet, "data1.csv", "colu");
3.  mSet<-SanityCheckIndData(mSet, "data1.csv")
4.  mSet<-PerformIndNormalization(mSet, "data1.csv", "log", 1);
5.  mSet<-PerformLimmaDE(mSet, "data1.csv", 0.05, 0.0);
6.  mSet<-ReadIndData(mSet, "data2.csv", "colu");
7.  mSet<-SanityCheckIndData(mSet, "data2.csv")
8.  mSet<-PerformIndNormalization(mSet, "data2.csv", "log", 1);
9.  mSet<-PerformLimmaDE(mSet, "data2.csv", 0.05, 0.0);
10. mSet<-ReadIndData(mSet, "data3.csv", "colu");
11. mSet<-SanityCheckIndData(mSet, "data3.csv")
12. mSet<-PerformIndNormalization(mSet, "data3.csv", "log", 1);
13. mSet<-PerformLimmaDE(mSet, "data3.csv", 0.05, 0.0);
14. mSet<-ReadIndData(mSet, "data4.csv", "colu");
15. mSet<-SanityCheckIndData(mSet, "data4.csv")
16. mSet<-PerformIndNormalization(mSet, "data4.csv", "log", 1);
17. mSet<-PerformLimmaDE(mSet, "data4.csv", 0.05, 0.0);
18. mSet<-CheckMetaDataConsistency(mSet, F);
```

Screenshot of meta-analysis example results

MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation

- Upload
- Meta analysis
- Result table
- Venn diagram
- Download
- Exit

Top metabolite features identified in the meta-analysis

The statistics from individual data analysis are given in columns with the corresponding dataset names. You can either adjust its content or sort the table.

Data summary: Log fold change | Sort by: CombinedPval | Order: Ascending | Update | Search | Download

ID	data1	data2	data3	data4	CombinedTstat	CombinedPval	View
adenosine-5-phosphate	-1.7535	-0.81954	-0.89017	-1.6839	-187.97	0.0	View
pyrophosphate	-1.6539	-0.68801	-1.0106	-1.6181	-174.04	0.0	View
pyruvic acid	-1.7275	-0.02462	-1.154	-0.18216	-123.22	0.0	View
maltotriose	-0.57042	-0.78002	-0.62125	-0.34062	-45.974	8.2354E-6	View
glutamine	0.25055	0.58333	0.9243	0.23142	42.535	2.9447E-5	View
lactamide	-0.16134	-0.33998	-0.99086	-0.14049	-37.845	1.8363E-4	View
citrulline	0.17019	0.70856	0.64683	0.23439	34.702	5.1871E-4	View
lactic acid	-0.048584	-0.13516	-1.0411	0.010808	-34.79	5.1871E-4	View
alpha ketoglutaric acid	-0.52005	-0.28152	-0.58456	-0.40263	-32.543	0.0011327	View
cystine	0.21575	0.80808	0.38177	0.20646	31.319	0.0015355	View
taurine	0.0050866	-0.21137	-0.88704	-0.27037	-31.389	0.0015355	View
maltose	-0.35858	-0.69713	-0.38259	-0.22187	-30.371	0.0020746	View
fructose	0.54515	0.28723	0.60315	0.1174	29.697	0.0025194	View
asparagine	0.26279	0.37151	0.66667	0.28026	29.053	0.0030375	View
oxalic acid	-0.24872	-0.55199	-0.44125	-0.32291	-27.222	0.0059119	View

(1 of 3) 1 2 3 15

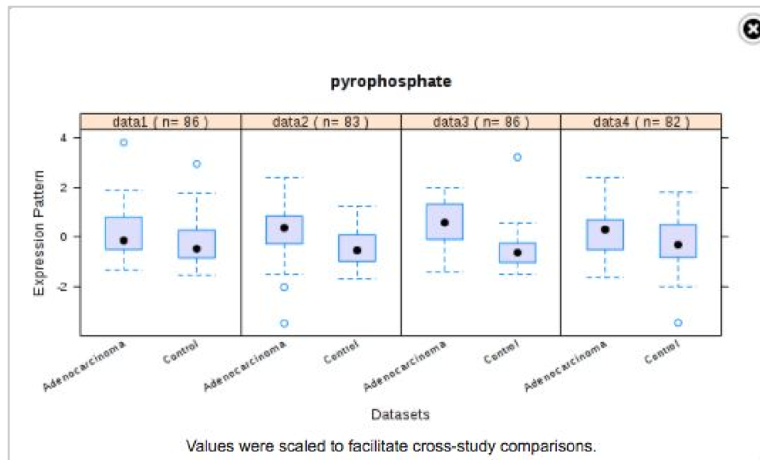
Previous | Venn Diagram

Click **View** for box-plot of your selected feature across all the data sets

Click **Venn Diagram** to view data in Venn

R Command History

```
7.  mSet<-SanityCheckIndData(mSet, "data2.csv")
8.  mSet<-PerformIndNormalization(mSet, "data2.csv", "log", 1);
9.  mSet<-PerformLimmaDE(mSet, "data2.csv", 0.05, 0.0);
10. mSet<-ReadIndData(mSet, "data3.csv", "colu");
11. mSet<-SanityCheckIndData(mSet, "data3.csv")
12. mSet<-PerformIndNormalization(mSet, "data3.csv", "log", 1);
13. mSet<-PerformLimmaDE(mSet, "data3.csv", 0.05, 0.0);
14. mSet<-ReadIndData(mSet, "data4.csv", "colu");
15. mSet<-SanityCheckIndData(mSet, "data4.csv")
16. mSet<-PerformIndNormalization(mSet, "data4.csv", "log", 1);
17. mSet<-PerformLimmaDE(mSet, "data4.csv", 0.05, 0.0);
18. mSet<-CheckMetaDataConsistency(mSet, F);
19. mSet<-PerformPvalCombination(mSet, "f
```



Example of a box-plot of pyrophosphate across the 4 different datasets.

From the image, pyrophosphate is consistently more expressed in patients with Adenocarcinoma than in healthy patients.

MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation

- Upload
- Meta analysis
- Result table
- Venn diagram
- Download
- Exit

The statistics from individual data analysis are given in columns with the corresponding dataset names. You can either adjust its content or sort the table.

Data summary: Log fold change | Sort by: CombinedPval | Order: Ascending | Update | Search | Download

ID	data1	data2	data3	data4	CombinedTstat		
adenosine-5-phosphate	-1.7535					0.0	
pyrophosphate	-1.6539					0.0	
pyruvic acid	-1.7275					0.0	
maltotriose	-0.57042					8.2	
glutamine	0.25055					2.9	
lactamide	-0.16134					1.8	
citrulline	0.17019					5.1	
lactic acid	-0.048584					5.1871E-4	View
alpha ketoglutaric acid	-0.52005					0.0	
cystine	0.21575					0.0	
taurine	0.0050866					0.0	
maltose	-0.35858					0.0	
fructose	0.54515	0.28723	0.60315	0.1174	29.697	0.0	
asparagine	0.26279	0.37151	0.66667	0.28026	29.053	0.0030375	View
oxalic acid	-0.24872	-0.55199	-0.44125	-0.32291	-27.222	0.0059119	View

(1 of 3) 1 2 3 15

Previous | Venn Diagram

The max number of datasets that can be compared is **four**. Datasets without significant hits will be excluded.

Name	DE #	Include
data1	3	✓
data2	12	✓
data3	21	✓
data4	2	✓
meta_dat	34	✓

Cancel | Submit

1. Click here to select datasets to be included in the meta-analysis (max 4)

2. Click **submit** to view resulting Venn Diagram

R Command History

Keep collapsed | Save

```
1. InitDataObjects("conc", "metadata", FALSE)
2. mSet<-ReadIndData(mSet, "data1.csv", "colu");
3. mSet<-SanityCheckIndData(mSet, "data1.csv")
4. mSet<-PerformIndNormalization(mSet, "data1.csv", "log", 1);
5. mSet<-PerformLimmaDE(mSet, "data1.csv", 0.05, 0.0);
6. mSet<-ReadIndData(mSet, "data2.csv", "colu");
7. mSet<-SanityCheckIndData(mSet, "data2.csv")
8. mSet<-PerformIndNormalization(mSet, "data2.csv", "log", 1);
9. mSet<-PerformLimmaDE(mSet, "data2.csv", 0.05, 0.0);
10. mSet<-ReadIndData(mSet, "data3.csv", "colu");
11. mSet<-SanityCheckIndData(mSet, "data3.csv")
12. mSet<-PerformIndNormalization(mSet, "data3.csv", "log", 1);
13. mSet<-PerformLimmaDE(mSet, "data3.csv", 0.05, 0.0);
14. mSet<-ReadIndData(mSet, "data4.csv", "colu");
15. mSet<-SanityCheckIndData(mSet, "data4.csv")
16. mSet<-PerformIndNormalization(mSet, "data4.csv", "log", 1);
17. mSet<-PerformLimmaDE(mSet, "data4.csv", 0.05, 0.0);
18. mSet<-CheckMetaDataConsistency(mSet, F);
19. mSet<-PerformPvalCombination(mSet, "fisher", 0.05)
20. mSet<-GetMetaResultMatrix(mSet, "fc")
```

Screenshot of meta-analysis venn diagram view

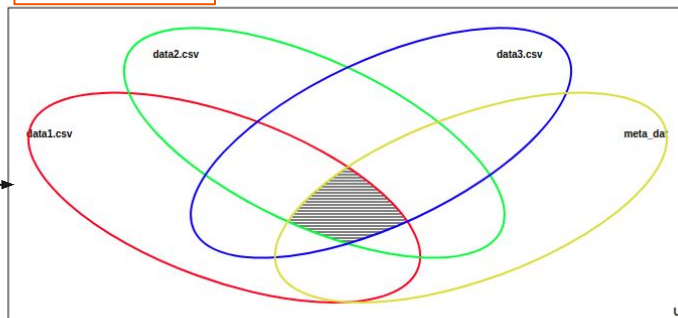
MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation

- Upload
- Meta analysis
- Result table
- Venn diagram
- Download
- Exit

Different regions in the Venn diagram represent all possible comparisons among the data sets.

- At most **four** datasets can be compared at the same time. Datasets without hits will **NOT** be shown here;
- The area of the region does **NOT** relate to the number of features;
- ...sponding features on the left panel.

3. Download results here



1. Click anywhere inside the **Venn Diagram** to explore common features between the different datasets

Total:2

- adenosine-5-phosphate
- pyrophosphate

2. The common features will be listed above

R Command History

Keep collapsed   Save

```
1.  InitDataObjects("conc", "metadata", FALSE)
2.  mSet<-ReadIndData(mSet, "data1.csv", "colu");
3.  mSet<-SanityCheckIndData(mSet, "data1.csv")
4.  mSet<-PerformIndNormalization(mSet, "data1.csv", "log", 1);
5.  mSet<-PerformLimmaDE(mSet, "data1.csv", 0.05, 0.0);
6.  mSet<-ReadIndData(mSet, "data2.csv", "colu");
7.  mSet<-SanityCheckIndData(mSet, "data2.csv")
8.  mSet<-PerformIndNormalization(mSet, "data2.csv", "log", 1);
9.  mSet<-PerformLimmaDE(mSet, "data2.csv", 0.05, 0.0);
10. mSet<-ReadIndData(mSet, "data3.csv", "colu");
11. mSet<-SanityCheckIndData(mSet, "data3.csv")
12. mSet<-PerformIndNormalization(mSet, "data3.csv", "log", 1);
13. mSet<-PerformLimmaDE(mSet, "data3.csv", 0.05, 0.0);
14. mSet<-ReadIndData(mSet, "data4.csv", "colu");
15. mSet<-SanityCheckIndData(mSet, "data4.csv")
16. mSet<-PerformIndNormalization(mSet, "data4.csv", "log", 1);
17. mSet<-PerformLimmaDE(mSet, "data4.csv", 0.05, 0.0);
18. mSet<-CheckMetaDataConsistency(mSet, F);
19. mSet<-PerformPvalCombination(mSet, "fisher", 0.05)
20. mSet<-GetMetaResultMatrix(mSet, "fc")
21. mSet<-PrepareVennData(mSet);
```

Xia Lab @ McGill (last updated 2018-02-02)

Example of analysis-report

Analysis_Report.pdf 1 / 7

Metabolomic Data Analysis with MetaboAnalystR

Name: guest6017956821136707109

February 7, 2018

1 Background

The combination of multiple independent metabolomics studies investigating the same condition in similar populations, which is often termed âĂIJhorizontal integrationâĂİ, or âĂIJmetabolomic meta-analysisâĂİ. The aim of metabolomic meta-analysis is to leverage the collective power of multiple studies to overcome potential noise, bias, and small effect sizes to improve the precision in identifying true patterns within data. Specifically, biomarker identification remains a large area of research in metabolomics, and their validation is challenging due to inconsistencies in identified biomarkers amongst similar experiments. Performing meta-analysis across similar studies will thereby increase the sample size and the power to identify robust and precise biomarkers of disease. The aim of the Meta-Analysis module for the integration of individual metabolomic studies to identify consistent and robust biomarkers of disease. This module supports three methods for performing meta-analysis: 1) Combining p-values, 2) Vote counting, and 3) Direct merging of data into a mega-dataset.

2 Meta-Analysis Overview

The Meta-Analysis module consists of six steps: 1) uploading the individual datasets; 2) data processing of each individual dataset, however it is suggested that the data-processing steps are consistent amongst the studies; 3) differential expression analysis of individual datasets; 4) data integrity check prior to meta-analysis ; 5) selection of the statistical method for meta-analysis, and 6) visualization of results as a Venn diagram to view all possible combinations of shared features between the datasets.

3 Data Input

The Meta-Analysis module accepts individual datasets which must be prepared by users prior to being uploaded. In general, the datasets must have been collected under comparable experimental conditions/share the same hypothesis or have the same mechanistic underpinnings. At the moment, the module only supports two-group comparisons (ex: control vs disease). Further, the module accepts either a compound concentration table, spectral binned data, or a peak intensity table. The format of the data must be specified, identifying whether the samples are in rows or columns, or may either be .csv or .txt files.

2) What is MS Peaks to Pathways?

- High-throughput analysis and functional interpretation of untargeted MS-based metabolomics data is a major bottleneck
- A promising approach is to shift the unit of analysis from individual compounds to pathways - similar to GSEA/MSEA
- Mummichog algorithm (Li et al. 2013) bypasses the bottleneck of identification prior to pathway analysis, leveraging a priori pathway/network knowledge to directly infer biological activity
- The **MS Peaks to Pathways** module implements this algorithm in a user-friendly interface, including an expanded library of 21 organisms derived from KEGG metabolic pathways

Steps for MS Peaks to Pathways

1) Users must upload a table containing three-columns, m/z features, p-values, and statistical scores (t-scores/fold-change values) --- see example below
2) Users must specify the mass accuracy and ion mode of their MS instrument, and the p-value cutoff
3) Users must select an organism's library from which to perform pathway analysis
4) View pathway analysis results
5) Visualize results in a global KEGG metabolic network

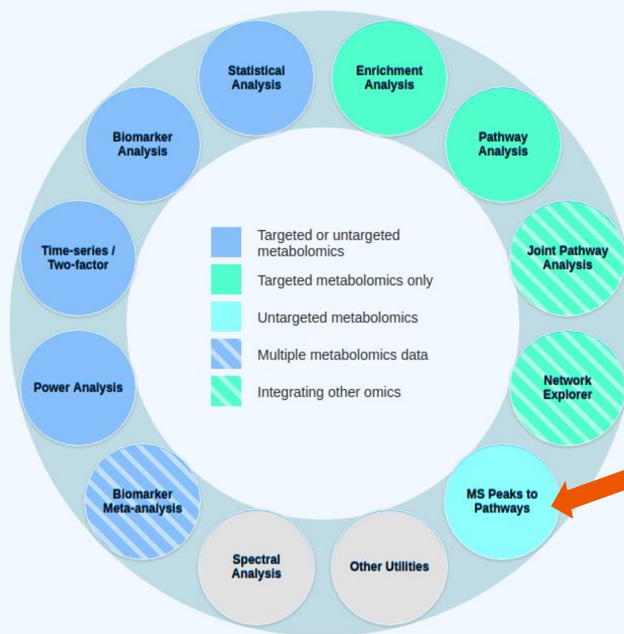
Example of a dataset to upload: user's data must have identical column titles, **m.z**, **p.value**, and **t.score**

	A	B	C
1	m.z	p.value	t.score
2	304.2979	1.02E-10	14.7179316
3	177.1024	1.62E-10	14.2666
4	345.0277	1.72E-10	-14.209195
5	491.0325	1.83E-10	-14.146348
6	258.0048	2.17E-10	-13.987636
7	483.1205	2.22E-10	-13.967634
8	694.9937	2.81E-10	-13.745172
9	270.9767	3.27E-10	13.6060705
10	371.604	3.53E-10	-13.534483
11	316.5773	3.71E-10	13.4893333
12	451.0505	4.04E-10	-13.412347

MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation

- [Home]
- [Overview]
- [Data Formats]
- [FAQs]
- [Tutorials]
- [Troubleshooting]
- [Resources]
- [Update History]
- [User Stats]
- [About]

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



Click here to start

Xia Lab @ McGill (last updated 2018-02-02)

MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation

- Upload
- Data check
- Set parameter
- View result
- Metabolic network
- Download
- Exit

Upload a peak list profile

Mass accuracy: 10 ppm

Analytical Mode: Positive Mode Negative Mode

P-value Cutoff: 1.0E-4 (editable)

Choose Data File: Choose file No file chosen

Use the example data

Dataset An example peak list data obtained from untargeted metabolomics using Orbitrap LC-MS (positive mode, human samples, p.value cutoff: 0.0001)

Submit

Try the example data here

1. Specify the mass-accuracy of your MS instrument
2. Specify the mode of your MS instrument
3. Specify the p-value cutoff between DE features
4. Upload your file
5. Click **Submit** to continue

R Command History

Keep collapsed Save

Xia Lab @ McGill (last updated 2018-02-02)

MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation

- Upload
- Data check
- Set parameter
- View result
- Metabolic network
- Download
- Exit

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 3934 input mz features were retained for further analysis

The optimal number of significant features ~300.

A total of 261 signficant mz features were found based on the selected p-value cutoff: 1e-04

Missing value estimation   Skip

Click **Skip** to continue

Data Integrity Check performs a check on your uploaded data to ensure it is suitable for further analysis

R Command History

Keep collapsed   Save

```
1. InitDataObjects("mass_all", "mummichog", FALSE)
2. mSet<-Read.PeakListData(mSet, "Replacing_with_your_file_path");
3. mSet<-UpdateMummichogParameters(mSet, "three", "positive", 1.0E-4);
4. mSet<-SanityCheckMummichogData(mSet)
```

R Command History appears real-time and is ordered sequentially

MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation

- Upload
- Data check
- Set parameter
- View result
- Metabolic network
- Download
- Exit

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]

1. Select the organism library that best matches your data

2. Scroll down and click **Submit** to continue

R Command History

Keep collapsed Save

```
1. InitDataObjects("mass_all", "mummichog", FALSE)
2. mSet<-Read.PeakListData(mSet, "Replacing_with_your_file_path");
3. mSet<-UpdateMummichogParameters(mSet, "three", "positive", 1.0E-4);
4. mSet<-SanityCheckMummichogData(mSet)
```

Xia Lab @ McGill (last updated 2018-02-02)

Screenshot of MS Peaks to Pathways example results

MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation

- Upload
- Data check
- Set parameter
- View result
- Metabolic network
- Download
- Exit

Predicted pathway activity profiles based on Mummichog:

Click View under Match Details to view the compounds within each pathway (with matched compounds highlighted). The detailed pathways and matched compounds tables can be downloaded using the links at the bottom.

Explore Results in Network

Pathway Name	Total	Hits (all)	Hits (sig.)	Fisher's Pvalue	EASE Score	Gamma Pvalue	Match Details
Vitamin B9 (folate) metabolism	33	8	3	0.061284	0.24646	0.0011975	View
Drug metabolism - cytochrome P450	53	11	3	0.13848	0.38654	0.0026042	View
Glycosphingolipid biosynthesis - globoseries	16	5	2	0.11403	0.47249	0.0042576	View
Glycosphingolipid biosynthesis - ganglioseries	62	6	2	0.15795	0.53617	0.006189	View
Sialic acid metabolism	107	15	3	0.27033	0.55362	0.0068691	View
Tryptophan metabolism	94	42	6	0.40228	0.58196	0.0081517	View
N-Glycan biosynthesis	48	7	2	0.20437	0.59226	0.0086803	View
Vitamin E metabolism	54	7	2	0.20437	0.59226	0.0086803	View
Phosphatidylinositol phosphate metabolism	59	8	2	0.25207	0.64166	0.011793	View
Glycerophospholipid metabolism	156	18	3	0.37673	0.65677	0.012976	View
Glycolysis and Gluconeogenesis	49	19	3	0.41186	0.68674	0.015735	View
Methionine and cysteine metabolism	94	21	3	0.48018	0.74034	0.022487	View
Glycosphingolipid metabolism	67	16	2	0.59885	0.87363	0.06139	View
Aminosugars metabolism	69	16	2	0.59885	0.87363	0.06139	View
Pyrimidine metabolism	70	21	2	0.74822	0.93467	0.1107	View
Tyrosine metabolism	160	47	3	0.94362	0.98475	0.23503	View
C5-Branched dibasic acid metabolism	10	3	1	0.32235	1.0	1.0	View
Chondroitin sulfate degradation	37	3	1	0.32235	1.0	1.0	View
Hexose phosphorylation	20	7	1	0.59805	1.0	1.0	View
Galactose metabolism	41	12	1	0.79196	1.0	1.0	View

1 2 3

Download Result Tables: Pathway Hits Compound Hits

R Command History

Keep collapsed Save

```
1. InitDataObjects("mass_all", "mummichog", FALSE)
2. mSet<-Read.PeakListData(mSet, "Replacing_with_your_file_path");
3. mSet<-UpdateMummichogParameters(mSet, "three", "positive", 1.0E-4);
4. mSet<-SanityCheckMummichogData(mSet)
```

A table of results containing ranked pathways enriched in user-uploaded data

Click **Explore Results in Network** to visualize your results on a global KEGG metabolic network

Click **view** to see detailed hits for each pathway

Xia Lab @ McGill (last updated 2018-02-02)

Screenshot of MS Peaks to Pathways Network View

Module ▸ Upload ▸ Data Check ▸ Library ▸ Result Table ▸ Metabolic network ▸ Download

Style: KEGG style Background: White Pathway name: Hide Download: --Please Select-- Highlight:

Top toolbar containing different menus to customize the network (change background colour, download image)

MS Peaks to Pathways analysis results

	Name	Hits	Sigs	P-value	Color
☐	Vitamin B9 (folate) metab	8	3	0.061284	
☐	Drug metabolism - cytoch	11	4	0.13848	
☐	Glycosphingolipid biosynt	5	2	0.11403	
☐	Glycosphingolipid biosynt	6	3	0.15795	
☑	Sialic acid metabolism	15	7	0.27033	
☑	Tryptophan metabolism	42	13	0.40228	
☐	N-Glycan biosynthesis	7	3	0.20437	
☐	Vitamin E metabolism	7	2	0.20437	
☐	Phosphatidylinositol phos	8	3	0.25207	
☐	Glycerophospholipid meta	18	4	0.37673	
☐	Glycolysis and Gluconeog	19	6	0.41186	
☐	Methionine and cysteine n	21	3	0.48018	
☐	Glycosphingolipid metabc	16	5	0.59885	
☐	Aminosugars metabolism	16	2	0.59885	
☐	Pyrimidine metabolism	21	2	0.74822	
☐	Tyrosine metabolism	47	3	0.94362	
☐	C5-Branched dibasic acid	3	2	0.32235	
☐	Chondroitin sulfate degrac	3	1	0.32235	
☐	Hexose phosphorylation	7	5	0.59805	
☐	Galactose metabolism	12	8	0.79196	
☐	Keratan sulfate degradatio	7	1	0.59805	
☐	Alkaloid biosynthesis II	3	1	0.32235	
☐	Vitamin B1 (thiamin) meta	5	1	0.47784	
☐	Glycosphingolipid biosynt	3	1	0.32235	
☐	Propanoate metabolism	9	2	0.6909	
☐	Blood Group Biosynthesis	4	1		

Hits

Tryptophan metabolism
(signficant hits in red)

- C00025
- C00026
- C00020
- CE2119
- C03722

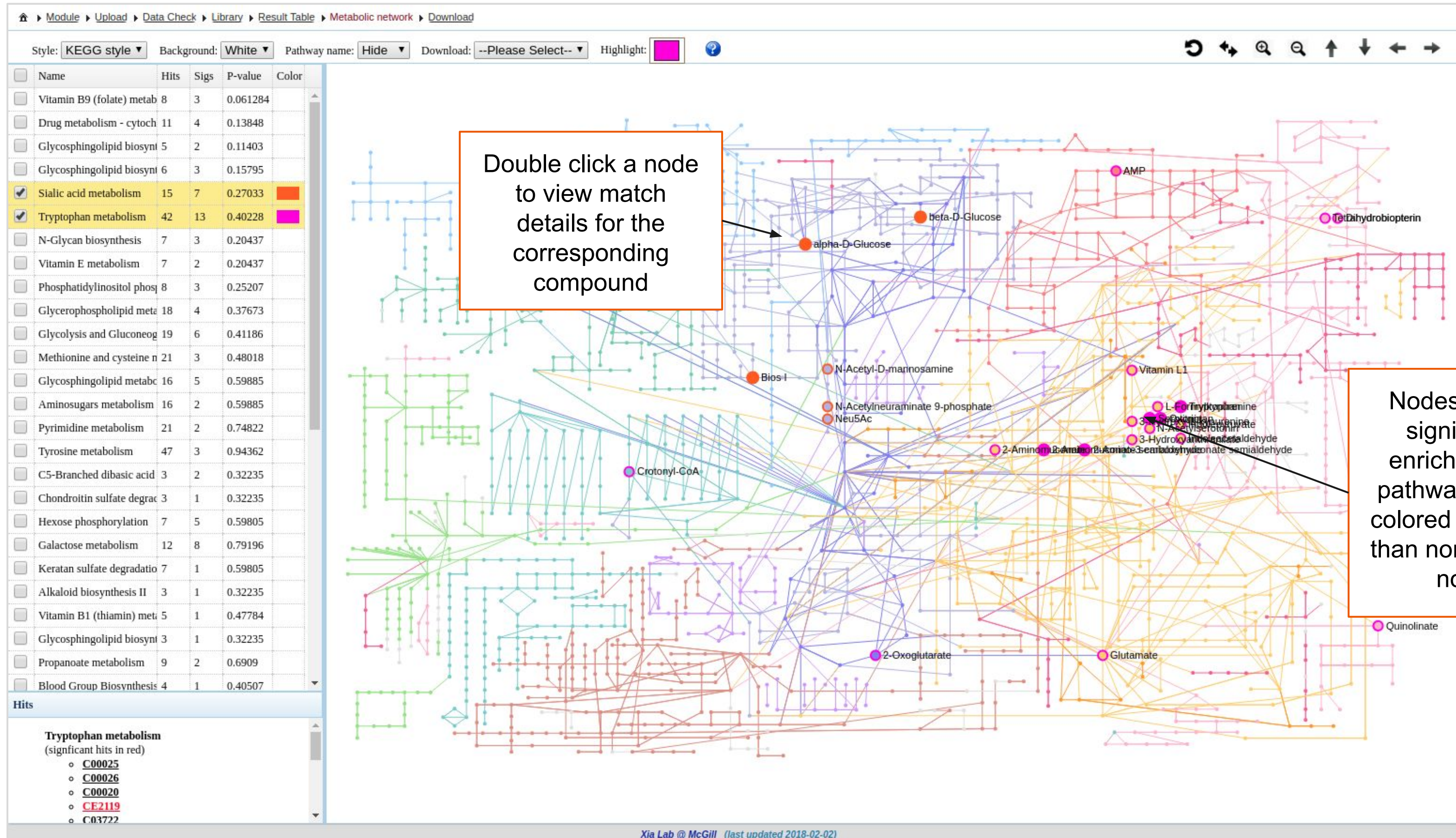
View compounds within the selected pathway

beta-D-Glucose
alpha-D-Glucose
Tetrahydrobiopterin
Bios I
N-Acetyl-D-mannosamine
Vitamin L1
N-Acetylneuraminate 9-phosphate
Neu5Ac
Crotonyl-CoA
Quinolinate
2-Oxoglutarate
Glutamate

Mapped features are highlighted in user-selected color

Xia Lab @ McGill (last updated 2018-02-02)

Screenshot of MS Peaks to Pathways Network View



3) Network Explorer

- Integrating multiple omics data and interpreting these results at a systems level is a significant challenge
- Biological networks are an intuitive and flexible vehicle to convey a priori knowledge with users data at a systems level
- The **Network Explorer** module provides users an easy-to-use tool that permits mapping of metabolites and/or genes onto any of the 5 molecular interactions networks:
  - KEGG global metabolic network, gene-metabolite interaction network, metabolite-disease interaction network, metabolite-metabolite interaction network, and a metabolite-gene-disease interaction network

MetaboAnalyst
4.0

MetaboAnalyst -- a comprehensive tool for m analysis and interpretation
Try our example data here
Upload
ID mapping
Set parameter
Network viewer
Download
Exit
Upload a list of genes and/or a list of metabolites. You can first try our example data.
R Command History
Keep collapsed
Save
Gene List
Gene list with optional fold changes
Metabolite List
Compound list with optional fold changes
1. Copy and paste either a list of genes, a list of metabolites, or both
ID Type:
--- Not Specified ---
ID Type:
-- Please specify
2. Specify the ID type
Submit
3. Click **Submit** to upload your data and continue
Xia Lab @ McGill (last updated 2018-02-02)

MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation

- Upload
- ID mapping
- Set parameter
- Network viewer
- Download
- Exit

Results of the name mapping of the uploaded data to MetaboAnalyst's internal database. Scroll down and click **Submit** to continue.

Compound Name Mapping | Gene Name Mapping

Query	Hit	HMDB	KEGG	Details	
C00116	Glycerol	HMDB0000131	C00116		Delete
C00565	Trimethylamine	HMDB0000906	C00565		Delete
C00033	Acetic acid	HMDB0000042	C00033		Delete
C00583	Propylene glycol	HMDB0001881	C00583		Delete
C00022	Pyruvic acid	HMDB0000243	C00022		Delete
C00719	Betaine	HMDB0000043	C00719		Delete
C05984	2-Hydroxybutyric acid	HMDB0000008	C05984		Delete
C00207	Acetone	HMDB0001659	C00207		Delete
C00065	L-Serine	HMDB0000187	C00065		Delete
C00031	D-Glucose	HMDB0000122	C00031		Delete
C00079	L-Phenylalanine	HMDB0000159	C00079		Delete
C02632	Isobutyric acid	HMDB0001873	C02632		Delete
C00064	L-Glutamine	HMDB0000641	C00064		Delete
C00114	Choline	HMDB0000097	C00114		Delete
C00073	L-Methionine	HMDB0000696	C00073		Delete
C00082	L-Tyrosine	HMDB0000158	C00082		Delete
C00186	L-Lactic acid	HMDB0000190	C00186		Delete
C00037	Glycine	HMDB0000123	C00037		Delete
C00543	Dimethylamine	HMDB0000087	C00543		Delete
C00077	Ornithine	HMDB0000214	C00077		Delete
C00058	Formic acid	HMDB0000142	C00058		Delete
C00188	L-Threonine	HMDB0000167	C00188		Delete
C00407	L-Isoleucine	HMDB0000172	C00407		Delete

R Command History

Keep collapsed | Save

```
1. InitDataObjects("conc", "network", FALSE)
2. mSet<-SetOrganism(mSet, "hsa")
3. geneListFile<-"replace_with_your_file_name"
4. geneList<-readChar(geneListFile, file.info(geneListFile)$size)
5. mSet<-PerformIntegGeneMapping(mSet, geneList, "hsa", "entrez");
6. cmpdListFile<-"replace_with_your_file_name"
7. cmpdList<-readChar(cmpdListFile, file.info(cmpdListFile)$size)
8. mSet<-PerformIntegCmpdMapping(mSet, cmpdList, "hsa", "kegg");
9. mSet<-CreateMappingResultTable(mSet)
```

R Command History appears real-time and is ordered sequentially

Xia Lab @ McGill (last updated 2018-02-02)

MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation

- Upload
- ID mapping
- Set parameter
- Network viewer
- Download
- Exit

Select one network to explore your data

Users can choose one of five different modes of networks analysis:

KEGG Global Metabolic Network

Users can map metabolites and enzymes/KOs (KEGG Orthologs), and then visually explore the results in the KEGG global KEGG metabolic network (ko01100). This feature is specially suitable to integrate results from joint **metabolomics and metagenomics** studies.

Gene-Metabolite Interaction Network

The gene-metabolite interaction network enables exploration and visualization of interactions between functionally related metabolites and genes. The chemical and human gene associations were extracted from STITCH, such that only highly confident interactions are used. Most of associations in STITCH are based on co-mentions highlighted in PubMed Abstracts including reactions from similar chemical structures and similar molecular activities.

Metabolite-Disease Interaction Network

The metabolite-disease interaction network enables exploration of disease-related metabolites. The associations were obtained from HMDB. Disease association have been added to HMDB via the Human Metabolome Project's literature curation team.

Metabolite-Metabolite Interaction Network

The metabolite-metabolite interaction network helps to highlight potential functional relationships between a wide set of annotated metabolites. The chemical-chemical associations for the metabolites network were extracted from STITCH, such that only highly confident interactions are used. Most of associations in STITCH are based on co-mentions highlighted in PubMed Abstracts including reactions from similar chemical structures and similar molecular activities.

Metabolite-Gene-Disease Interaction Network

The metabolite-gene-disease interaction network provides a global view of potential functional relationships between metabolites, connected genes, and target diseases. The network is an integration of gene-metabolite, metabolite-disease and gene-disease interaction networks.

R Command History

Keep collapsed | Save

```
1. InitDataObjects("conc", "network", FALSE)
2. mSet<-SetOrganism(mSet, "hsa")
3. geneListFile<-"replace_with_your_file_name"
4. geneList<-readChar(geneListFile, file.info(geneListFile)$size)
5. mSet<-PerformIntegGeneMapping(mSet, geneList, "hsa", "entrez");
6. cmpdListFile<-"replace_with_your_file_name"
7. cmpdList<-readChar(cmpdListFile, file.info(cmpdListFile)$size)
8. mSet<-PerformIntegCmpdMapping(mSet, cmpdList, "hsa", "kegg");
9. mSet<-CreateMappingResultTable(mSet)
10. mSet<-PrepareNetworkData(mSet);
```

MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation

- Upload
- ID mapping
- Set parameter
- Network viewer
- Download
- Exit

Mapping Overview

The input metabolites and genes (seeds) are mapped to the selected molecular interaction network to create subnetworks containing these seeds and their direct neighbours (i.e. first-order subnetworks). The procedure often produces one big subnetwork ("continent") with several smaller ones ("islands"). Subnetworks with at least 3 nodes are listed below. You can visually explore them in the next step. These subnetworks can be downloaded as SIF (simple interaction format) files to be explored in other tools (i.e. Cytoscape).

Networks	Nodes	Edges	Seeds	Interactions (.SIF)
subnetwork1	60	77	9	Download
subnetwork2	5	4	1	Download
subnetwork3	5	4	1	Download
subnetwork4	3	2	0	Download
subnetwork5	3	2	1	Download
subnetwork6	3	2	1	Download
subnetwork7	3	2	1	Download
subnetwork8	3	2	1	Download

View details of subnetwork construction from user's data

Previous

Proceed

Click Proceed to visualize your network

R Command History

Keep collapsed

Save

```
1. InitDataObjects("conc", "network", FALSE)
2. mSet<-SetOrganism(mSet, "hsa")
3. geneListFile<-"replace_with_your_file_name"
4. geneList<-readChar(geneListFile, file.info(geneListFile)$size)
5. mSet<-PerformIntegGeneMapping(mSet, geneList, "hsa", "entrez");
6. cmpdListFile<-"replace_with_your_file_name"
7. cmpdList<-readChar(cmpdListFile, file.info(cmpdListFile)$size)
8. mSet<-PerformIntegCmpdMapping(mSet, cmpdList, "hsa", "kegg");
9. mSet<-CreateMappingResultTable(mSet)
10. mSet<-PrepareNetworkData(mSet);
11. mSet<-SearchNetDB(mSet, "pheno", "global", FALSE, 0.5)
12. mSet<-CreateGraph(mSet)
```

Screenshot of Network Explorer View

Top toolbar containing different menus to customize the network (change background colour, download image)

Click on a feature to zoom-in on it in the network

Table of features showing its name, the node degree, the betweenness centrality, and the expression level

Use your mouse to zoom-in and out of your network, as well as highlight, drag and drop nodes

Screenshot of Network Explorer View

Details of current selection, showing their name and databank IDs

Users can also perform Path Explorer to view all possible paths from one feature to another

Results of functional enrichment analysis on selected nodes. Users can highlight all nodes of the selected pathway in the network.