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.
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.
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)

Example dataset highlighting class labels and unique sample identifiers
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

Try our example data here
Click proceed to select meta-analysis method.
1. Select one of three methods to perform meta-analysis

2. Click proceed to view results

Biomarker Meta-Analysis

There are two widely used methods to combine p-values from multiple studies for information integration – the Fisher’s method (2*ΣLog(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 quality (i.e., same analytical platforms with similar levels of missing values).

- Select a method: Fisher’s method
- Set a significance level: 0.05

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

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

Screenshot of meta-analysis methods
Biomarker Meta-Analysis

Top metabolite features identified in the meta-analysis

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

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

2. Click submit to view resulting Venn Diagram
1. Click anywhere inside the Venn Diagram to explore common features between the different datasets.

2. The common features will be listed above.

3. Download results here.
Metabolomic Data Analysis with MetaboAnalystR

Name: guest6017560821136707149

February 7, 2018

1 Background

The combination of multiple independent metabolomics studies investigating the same condition in similar populations, which is often termed horizontal integration, or horizontal metabolomic meta-analysis. The aim of metabolomic meta-analysis is to leverage the collective power of multiple studies to overcome potential biases, 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-Analysisto module for the integration of individual metabolomic studies to identify consistent and robust biomarkers of disease. This module supports these 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 an 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
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

---

**Try the example data here**
Metabo Analyst -- a comprehensive tool for metabolomics analysis and interpretation

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 281 significant mz features were found based on the selected p-value cutoff: 1e-04.

Click Skip to continue

R Command History

appears
real-time
and is
ordered
sequentially
1. Select the organism library that best matches your data
2. Scroll down and click **Submit** to continue
Biomarker Meta-Analysis

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

Click view to see detailed hits for each pathway

Click Explore Results in Network to visualize your results on a global KEGG metabolic network
Biomarker Meta-Analysis

Screenshot of MS Peaks to Pathways Network View

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

Mapped features are highlighted in user-selected color

View compounds within the selected pathway

MS Peaks to Pathways analysis results
Biomarker Meta-Analysis

Screenshot of MS Peaks to Pathways Network View

Nodes that are significantly enriched in the pathway are fully colored and larger than non-enriched nodes.

Double click a node to view match details for the corresponding compound.
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 user 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.
Upload a list of genes and/or a list of metabolites. You can first try our example data.

1. Copy and paste either a list of genes, a list of metabolites, or both
2. Specify the ID type
3. Click Submit to upload your data and continue
Results of the name mapping of the uploaded data to MetaboAnalyst's internal database. Scroll down and click **Submit** to continue.

**R Command History** appears real-time and is ordered sequentially.
Select one network to explore your data.
Biomarker Meta-Analysis

View details of subnetwork construction from user’s data

Click Proceed to visualize your network
Biomarker Meta-Analysis

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
Biomarker Meta-Analysis

Screenshot of Network Explorer View

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

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.