

MicrobiomeAnalyst 2.0

Comprehensive statistical, functional and integrative analysis of microbiome data



Tutorial for Microbiome Metabolomics



MicrobiomeAnalyst -- comprehensive statistical, functional and integrative analysis of microbiome data

[Home](#)

[Formats](#)

[Forum](#)

[Updates](#)

[Resources](#)

[Contact](#)

Marker Data Profiling

Analyze marker gene counts data

Shotgun Data Profiling

Analyze shotgun metagenomics data

Taxon Set Analysis

Discover enriched microbial signatures

Microbiome Metabolomics

Co-analyze microbiome & metabolomics data

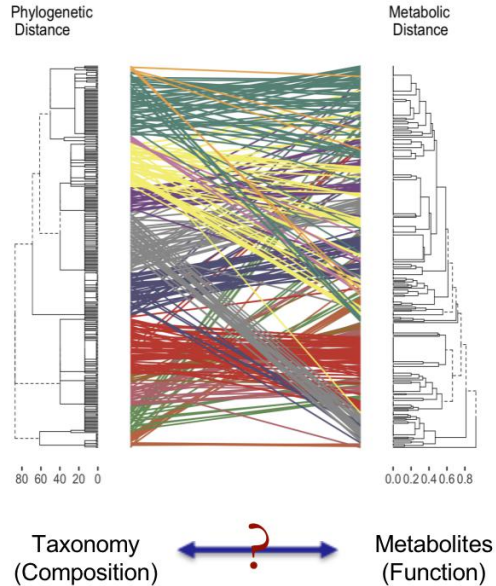
Statistical Meta-analysis

Integrate multiple marker gene data

Raw Data Processing

Convert raw 16S reads to ASV table

Motavition



- Metabolomics is critical to connecting microbial community composition and phenotypes at the level of altered metabolic processes in recent microbiome studies
- General statistical correlation analysis often leads to many false positives, making biological interpretation difficult.
- Integrating high dimensional microbiome and their corresponding metabolomics data remains a significant challenge.

Overview

Goal: to support integrative analysis for microbiome and metabolomics data in both statistically and biologically meaningful perspective

Approches (refer to method selection page for more details):

- **Dimensionality reduction:** Procrustes analysis (PA) and data integration analysis for biomarker discovery using latent components (DIABLO) are implemented to reveal the overall pattern of the paired microbiome and metabolomics datasets.
- **Metabolic network and pathway analysis:** Support pathway enrichment analysis for KO, metabolites and peaks against customized metabolic space based on the taxonomy input and visualization in an interactive network.
- **Microbiome-metabolome correlation analysis:** Provide statistical correlation, model-based correlation analysis as well as integrated correlation analysis

Data type:

- Microbiome data: ASV/OTU/KO count table, taxonomy list
- Metabolomics data: Targeted (metabolite) / Untargeted (Peak) as intensity table or list

I. Start from paired abundance tables:

Data Upload

Please upload your paired microbiome and metabolomics data table separately

The screenshot shows a web interface for data upload. At the top, there are three tabs: 'Abundance table' (selected), 'Feature List', and 'Try our examples'. Below the tabs, there are three sections for data upload, each with a description, radio button options, and an 'Upload' button. The sections are: 'Metadata file' (A text file containing group information), 'Microbiome data' (OTU/ASV counts data and KO abundance data), and 'Metabolomics data' (Targeted metabolomics data and Untargeted metabolomics data). At the bottom, there are 'Previous' and 'Proceed' buttons. Five blue callout boxes with white text provide numbered instructions: 1. Choose the abundance table tab; 2. Click upload the metadata file in .txt or .csv format; 3. Choose OTU/ASV or KO count table (.txt,.csv) to upload; 4. Choose metabolite or peak intensity table to upload; 5. Proceed to integrity check.

1. Choose the abundance table tab

2. Click upload the metadata file in .txt or .csv format.

3. Choose OTU/ASV or KO count table (.txt,.csv) to upload.

4. Choose metabolite or peak intensity table to upload.

5. Proceed to integrity check

- Metadata file should be consistent between the two data types.
- Taxonomy annotation is required for the marker gene data which can be provided as a table or as IDs in OTU/ASV table.
- For metabolites, users need to select the ID type. Names, KEGG IDs and HMDB IDs are accepted. For peaks, general formats (mz, mz__rt and mz@rt) are accepted.

Notes for normalized data upload:

It is highly advised to upload your microbiome abundance table containing raw counts to benefit the best practices for data analysis. However, if your data **has already been normalized**:

- Indicate it is Normalized data using the check box (shown in the red box to the different data types)
- Bypass data filtering and normalization (Optional)
- The function for predicting metabolites from different taxonomy will become inappropriate during data analysis.

Uploading marker gene counts data:

OTU/ASV table (.txt, .csv, or its zip) Taxonomy included Normalized data

+ Choose ?

Taxonomy table (.txt or .csv) + Choose ?

Taxonomy labels --- Not specified ---

Submit

Uploading KO abundance table:

Data format -- Please Specify --

Data file + Choose ? Normalized data

Submit

Uploading metabolite concentration table:

ID type: --- Please specify ---

Data file: + Choose Normalized data

Submit

Data integrity check:

Home > Data Inspection > Data Processing > Downloads

Uploaded datasets

- test_mic_mmp.csv
Feature: 50
Sample: 16
Norm. Input: No
- test_met_mmp.csv
Feature: 78
Sample: 16
Norm. Input: No

Downloads of the page

- Lib Size View (PDF)
- Lib Size View (SVG)
- Lib Size Data (CSV)

R Command History

Data Integrity Check

Data Check

- Feature abundance table contains raw counts (preferred) or normalized values;
- Features with identical values (i.e. zeros) across all samples will be excluded;
- Features that appear in only one sample will be excluded (considered artifacts);
- For ASV data, which uses actual sequences as IDs, the sequence IDs will be replaced with ASV_1, ASV_2, etc. (refer to the "ASV_ID_mapping.csv" from the Downloads page).

Metadata Check

- For categorical metadata, at least two groups and three replicates per groups are required; a metadata column will be excluded if unique values (i.e. no replicates) are detected.
- For continuous metadata, all values must be numerical.
- Missing values are **not allowed** in metadata.
- Use the [Edit Metadata](#) tab to inspect and manually address the issues

[Text Summary](#) | [Edit Metadata](#)

Data type:	Microbiome data	Metabolomics data
Total feature number:	50	78
Feature with >= 2 counts:	48	78
Percentage of missing value:	0	16
Number of sample names matched metadata:	16	16
Normalized data detected:	No	No

Meta information: 16 samples in metadata; 1 [discrete: 1 continuous: 0] experimental factors with replicates

<< Previous | Proceed >>

Details of uploaded files are displayed here

Available file downloads for each page are displayed here

Check the summary of each dataset here and make sure they are correct.

Proceed to continue

Data processing:

Click here to change the omics type for processing

Data processing status: **Incomplete**

Microbiome Metabolomics

Processing Step	Parameter Selection	Action
Filtering ?	<p>Percentage to remove (%): <input type="range" value="10"/></p> <p>Variance filter</p> <p><input checked="" type="radio"/> Inter-quartile range</p> <p>Based on: <input type="radio"/> Standard deviation</p> <p><input type="radio"/> Coefficient of variation</p> <p>Minimum count: <input type="range" value="4"/></p> <p>Abundance filter</p> <p><input checked="" type="radio"/> Prevalence in samples (%) <input type="range" value="20"/></p> <p><input type="radio"/> Mean abundance value</p> <p><input type="radio"/> Median abundance value</p>	<input type="button" value="Submit"/>
Normalization ?	<p>Data rarefying ? <input type="button" value="Do not rarefy my data"/></p> <p>Data transformation ? <input type="button" value="Do not transform my data"/></p> <p>Data scaling ? <input type="button" value="Total sum scaling (TSS)"/></p>	<input type="button" value="Submit"/>

Data processing status: **Finished**

Microbiome Metabolomics

Processing Step	Parameter Selection	Action
Filtering ?	<p><input type="radio"/> None (less than 5000 features)</p> <p><input checked="" type="radio"/> Interquartile range (IQR)</p> <p><input type="radio"/> Standard deviation (SD)</p> <p><input type="radio"/> Median absolute deviation (MAD)</p> <p><input type="radio"/> Relative standard deviation (RSD = SD/mean)</p> <p><input type="radio"/> Non-parametric relative standard deviation (MAD/median)</p> <p><input type="radio"/> Mean intensity value</p> <p><input type="radio"/> Median intensity value</p>	<input type="button" value="Submit"/> ✓
Normalization ?	<p>Sample normalization ? <input type="button" value="None"/></p> <p>Data transformation ? <input type="button" value="None"/></p> <p>Data scaling ? <input type="button" value="Auto scaling"/></p>	<input type="button" value="Submit"/> ✓

- Different approaches are applied to different omics type
- Background color and ticks suggest the completeness of each processing step , as exemplified by the figure on the right.

Data processing:

Home > Data Processing > Data Inspection > Downloads

Data scaling

Uploaded datasets

- test_mic_mmp.csv
Feature: 50
Sample: 16
Norm. Input: No
Finished
- test_met_mmp.csv
Feature: 78
Sample: 16
Norm. Input: No
Finished

Downloads of the page

- Density Plot (PDF)
- Density Plot (SVG)
- PCA Plot (PDF)
- PCA Plot (SVG)

Command History

Apply auto scale as default

PCA Overview Density Plot

Auto scale can be set as default to make the datasets in the same scale

When all the processing is finished, users can click proceed on the bottom to continue

Overall pattern of the two datasets according to their processing result presenting by PCA and density plot

Microbiome

Metabolomics

Conditions

- MI
- NOMI

Principal Component Analysis (PCA) is a popular technique to project high-dimensional data into lower dimensions to visually identify patterns. It highlights similarities and differences between the different samples using linear transformation.

<< Previous

>> Proceed

Comparison analysis:

- Aim to identify significant features from individual omics.
- MaAsLin2 is used for microbiome and Limma is used for metabolomics data. Both methods are based on general linear models.

The screenshot shows the MaAsLin2 web interface with the following fields:

- Primary metadata:** group (dropdown)
- Comparison:** T1D vs. NC (dropdowns)
- Covariates (control for):** location (dropdown with an 'x' icon)
- Blocking factor:** -- Unspecified -- (dropdown)
- Adjusted p-value cutoff:** For microbiome data: 0.2; For metabolomics data: 0.1
- Submit** button

The main metadata of interest included as a 'fixed effect' in the linear model whose coefficient are displayed in the results table.

If primary metadata is categorical, specify the comparison of interest.

Specify all variables that you'd like to account for as 'fixed effects' here.

Specify the variables that you'd like to account for as 'random effects' here.

Both covariates and blocking factor will impact statistics extracted for the primary metadata.

Set the significant level for each omics data. Note the number indicate the adjusted p-value

Microbiome Result Metabolomics Result

Feature	Log2FC ↑↓	P_value ↑↓

Comparison analysis:











- Comparison results are displayed as tables including fold change, raw p-value and adjusted p-value. Feature details can be viewed as box plot by clicking the picture icon.

Check the comparison results for each omics type by selecting the corresponding tab

Microbiome Result Metabolomics Result

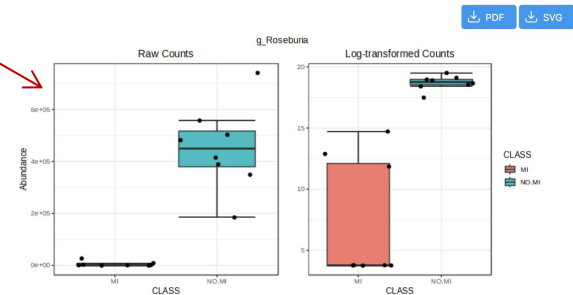
Taxonomy level: Genus

The table below shows at most 500 features ranked by their p-value with significant features highlighted in orange.

Name ↑↓	Log2FC ↑↓	P_value ↑↓	FDR ↑↓	View	
g_Roseburia	6.99	3.98E-9	1.47E-7		
g_Mucispirillum	6.17				
g_Enterococcus	-4.84				
g_Acetatifactor	2.53				
g_Oscillibacter	2.98	0.464	1.58E-5	1.16E-4	
g_unclassified_Lachnospiraceae	2.94	0.465	1.89E-5	1.16E-4	
g_Desulfovibrio	-1.52	0.264	4.93E-5	2.61E-4	
g_unclassified_Tyzzarella	3.67	0.657	6.79E-5	3.14E-4	
g_Akkermansia	-8.03	1.46	7.87E-5	3.23E-4	
g_Brevetella	-5.98	1.04	1.69E-4	6.25E-4	

The result for different taxonomy can be updated and checked here.

Feature Details View



Methods Selection:

Explore Overall Patterns via Dimensionality Reduction

Procrustes analysis (PA) is a simple visualization technique that superimposes the principal components of two datasets at a low-dimensional space. Procrustes essentially computes reduced dimensions for each data set using a method similar to PCA. Then, one of the reduced dimension matrices is rotated until it has maximum similarity with the other. Scores from both matrices are plotted at the same time, with pairs belonging to the same sample connected by a line. Procrustes is asymmetric, therefore the order that the 'omics datasets are uploaded will impact the results. ([more details...](#))

Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO) is a supervised method for multi-omics biomarker exploration. It is based on a generalized version of PLS (multi block PLS-DA) that seeks to find related multi-dimensional components that maximally separate sample labels. DIABLO is symmetric with respect to the 'omics data, therefore the order that 'omics datasets are uploaded will not impact the results. ([more details...](#))

Method selection:

Procrustes ▾
Procrustes
DIABLO

Submit

Obtain Functional Insights via Pathways & Network

Microbiome data and metabolomics data are projected into KEGG metabolic network for visual exploration as well as enrichment analysis. The integration strategies are based on microbiome data types.

- Marker genes data will be used to customize the metabolic network for enrichment analysis of metabolomics data. Users can click a node to view the most correlated microbes of metabolites
- For the metagenomics data, both KOs and metabolomic features will be projected to the selected network for integration analysis.

Compound origin:

All uploaded taxa/ko ▾
All uploaded taxa/ko
Significant Taxa
Bacteria
Bacteria and Human
Whole metabolic map

Submit

Tune the KEGG metabolic network for enrichment analysis. Significant taxa is only applicable for marker gene input.

Discover Metabolite-microbe Correlation

The relationships between the paired microbe and metabolite are intuitively presented using an interactive heatmap. Two types of heatmaps are provided including:

- Correlation heatmap is based on the statistical correlation.
- Prediction heatmap is based on GEM-based predictive models. The predictive models are logistic regression models trained based on high-quality genome-scale metabolic models (GEMs) that can predict the potential of each metabolite production across different taxonomy levels. Prediction heatmap is only suitable for integrating different taxonomy and metabolites.

Heatmap type:

Correlation heatmap ▾
Correlation heatmap
Prediction heatmap

Submit

Choose the heatmap mode here. Prediction heatmap is **not** suitable for KO and peaks input.

Procrustes analysis summary:

Procrustes analysis summary

The resulting plot of procrustes analysis along with key statistics is shown below.

Summary statistics of Procrustes analysis are displayed below. Procrustes sum of squares represents

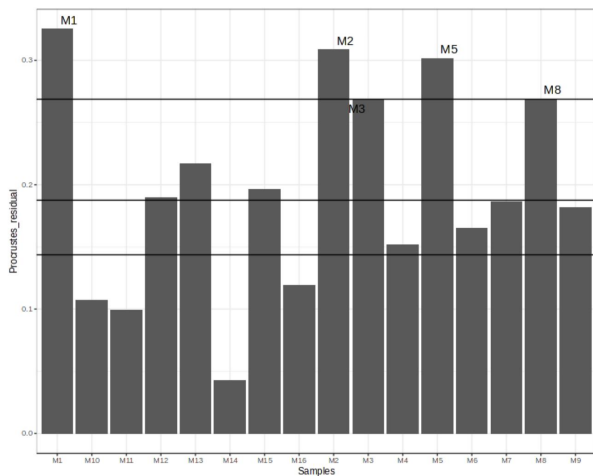
The shorter the distance, the better

Procrustes Sum of Squares: 0.4793
Correlation in a symmetric Procrustes rotation: 0.7216
Significance: 0.001

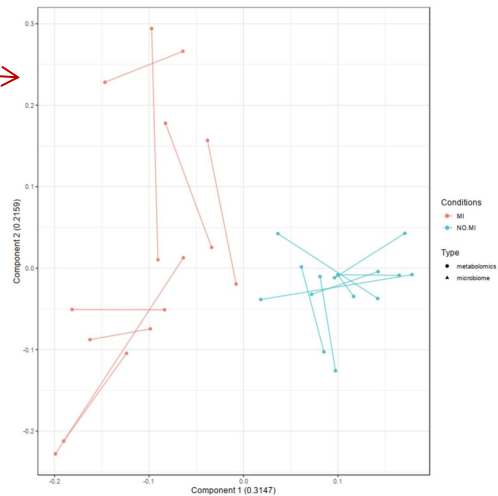
Check statistical summary here

Diagnostic

Sample space



View plot summary



Plot of procrustes residual differences which is based on the sum of squared deviations. The higher the residual difference, the more different is the paired sample between both omics. The three horizontal lines correspond to 75, median and 25 percentile of all residuals. Please refer to [vegan](#) for more details.

<< Previous

>> Proceed

Proceed to visualization

Procruste analysis visualization:

Use these options to customize the taxonomy level, appearance of the nodes and edges, or download the images.

Omics Space: Multi-omics View Type: Score plot (samples) Taxonomy: Genus Node Style: -- Specify -- Edge Style: -- Specify -- Download: -- Specify -- Advanced Options

Settings

Background

Floor

Wall

Shadow

Axis

Metadata highlight Advanced

Node display Update

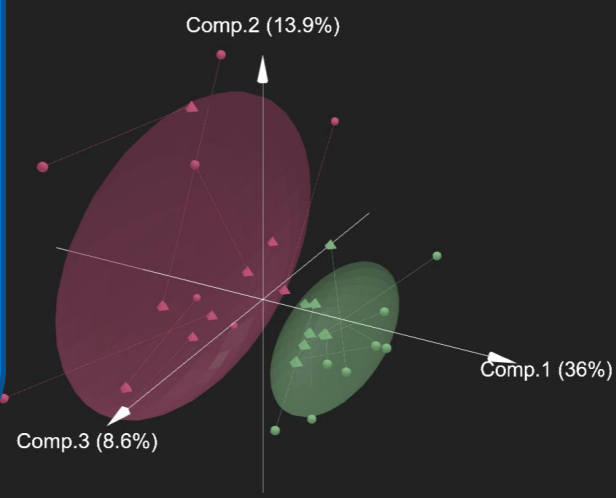
Metadata selection

Metadata of interest Submit

<input type="checkbox"/>	Name	Size	Color	Edit
<input checked="" type="checkbox"/>	M1	16		
<input type="checkbox"/>	NO...	16		

Use these options for overall bvisualization settings

- Scatter plot tools:
- Change highlight color
 - Reset
 - Rotate view on click
 - Drag view on click
 - Zoom in
 - Zoom out
 - Switch inset and main view
 - Remove inset view
 - Set ellipsoids



Click a sample or feature to see more info here

Current Selection

- MI
 - M1
 - M2
 - M3
 - M4
 - M5
 - M6

DIABLO analysis summary:

DIABLO analysis summary

The summary plots displayed below can be used to visualize the effectiveness of the dimension reduction before proceeding to the visual analytics page.

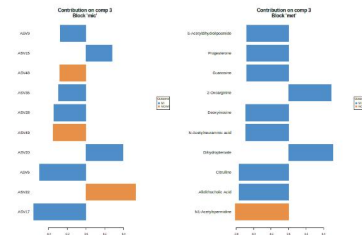
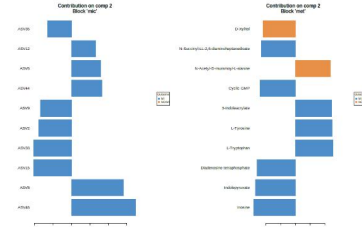
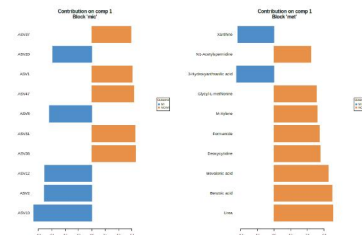
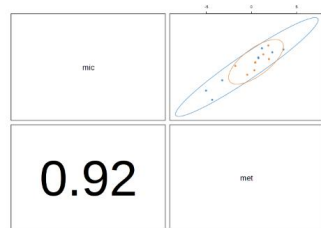
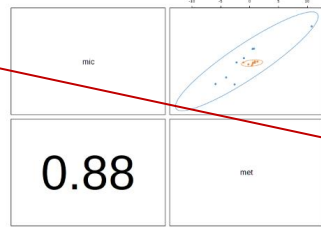
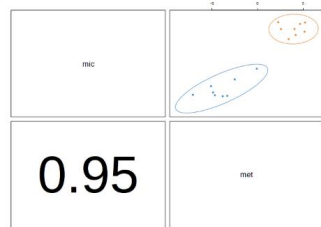
- **Diagnostic** tab shows how well the top components explain the variability in the integrated data set.
- **Sample space** tab visualizes overall distribution of samples in 2D space.
- **Loading** tab focuses on the contributions of features to the top components.



View plot summary



Proceed to visualization



<< Previous

>> Proceed

DIABLO analysis visualization:

Omics Space: **Multi-omics (Microbiome)**
View Type: **Score plot (samples)**
Taxonomy: **Genus**
Node Style: **-- Specify --**
Edge Style: **-- Specify --**
Download: **-- Specify --**
Advanced Options

Settings

Background:

Floor:

Wall:

Shadow:

Axis:

Metadata:

Node display: **Default** Update

Metadata selection

Metadata of interest: **CLASS** Submit

<input type="checkbox"/>	Name	Size	Color	Edit
<input checked="" type="checkbox"/>	MI	8	Red	✎
<input type="checkbox"/>	NO MI	8	Green	✎

Current Selection

- MI
- M1
- M2
- M3
- M4
- M5

Benzaldehyde

g_Roseburia

Comp.1 (35.7919%)

Comp.3 (8.26391%)

Metabolome Result Table

Feature	Log2FC	P-val	Color
2-Oxo-4-methylthioib	-1.710	0.000178	
2-Oxoarginine	-1.680	0.000232	
3-Hydroxyanthranilic	1.570	0.000588	
3-Indoleacrylate	1.560	0.000645	
3,4-Dihydroxyphenyl	-1.500	0.000994	
4-Aminobutyraldehye	-1.490	0.00109	
5-Methylthioadenosin	-1.400	0.00221	
Allolihocholic Acid	-1.380	0.00252	
Benzaldehyde	1.370	0.00268	
Benzoic acid	-1.360	0.00297	
Citrulline	-1.350	0.00303	
Cyclic GMP	-1.330	0.00346	

Microbiome Result Table

Feature	Log2FC	P-val	Color
g_Roseburia	6.990	3.98e-9	
g_Moryella	6.170	1.36e-8	
g_Moryella	-4.840	9.64e-7	
g_Acetatifactor	2.530	0.0000847	
g_Oscillibacter	2.980	0.000158	
g_unclassified_Lachn	2.940	0.000189	
g_Desulfovibrio	-1.520	0.000493	
g_unclassified_Tyzer	3.670	0.0000679	
g_Akkermansia	-8.030	0.0000787	
g_Prevotella	-5.280	0.000169	
g_Ruminiclostridium	2.130	0.000536	
g_unclassified_Clostr	3.410	0.000940	
g_Intestinimonas	2.400	0.00175	
g_Moribaculaceae_ur	-6.360	0.00186	
g_unclassified_Enter	-5.890	0.00375	
g_Parabacteroides	2.680	0.00654	
g_Mollicutes_RF39_1	3.120	0.00752	
g_Tyzerella	2.740	0.0101	
g_Bifidobla	-1.350	0.0113	

Switch omics type here

Switch view mode among score plot, loading plot and biplot

Insert view

Select the features of interest to highlight in the scatter plot

Pathway & network analysis:

Customize the appearance of the network here

Style: **KEGG style** Background: **Black** Pathway name: **Hide** Gene name: **Show** Compound name: **Show** Download: **--Please Select--** Highlight:

Metabolomics Enrichment Analysis

Submit Save

<input type="checkbox"/>	Name	Hits	P-value	P-adjust	Color
<input type="checkbox"/>	Arginine biosynthesis	3/19	0.0000050	0.0001588	
<input type="checkbox"/>	Purine metabolism	7/83	0.0000075	0.0001588	
<input checked="" type="checkbox"/>	Tryptophan metabolism	4/32	0.0000217	0.0003047	
<input type="checkbox"/>	Aminobenzoate degradati	2/19	0.0000885	0.0007423	
<input type="checkbox"/>	Toluene degradation	2/12	0.0000885	0.0007423	
<input type="checkbox"/>	Nitrogen metabolism	2/15	0.0001102	0.0007714	
<input type="checkbox"/>	Benzoate degradation	1/36	0.0001414	0.0008487	
<input type="checkbox"/>	Atrazine degradation	1/7	0.0001725	0.0009095	
<input type="checkbox"/>	Alanine, aspartate and glut	1/23	0.0004848	0.0015664	
<input type="checkbox"/>	Carbapenem biosynthesis	1/3	0.0004848	0.0015664	
<input type="checkbox"/>	Glyoxylate and dicarboxyl	1/44	0.0004848	0.0015664	
<input type="checkbox"/>	Butanoate metabolism	1/36	0.0004848	0.0015664	
<input type="checkbox"/>	Taurine and hypotaurine re	1/17	0.0004848	0.0015664	
<input type="checkbox"/>	Histidine metabolism	3/22	0.0006528	0.0019584	
<input type="checkbox"/>	Porphyrin metabolism	3/68	0.0011490	0.0032175	
<input type="checkbox"/>	D-Amino acid metabolism	3/43	0.0013611	0.0035725	
<input type="checkbox"/>	Arginine and proline meta	2/46	0.0015468	0.0038216	
<input type="checkbox"/>	Glutathione metabolism	2/22	0.0018631	0.0043477	
<input type="checkbox"/>	Terpenoid backbone biosy	1/23	0.0021811	0.0048215	
<input type="checkbox"/>	Pyrimidine metabolism	4/54	0.0046312	0.0097250	
<input type="checkbox"/>	Cysteine and methionine r	3/54	0.0099852	0.0199704	
<input type="checkbox"/>	Phenylalanine, tyrosine an	6/27	0.0116875	0.0223125	
<input type="checkbox"/>	Cyanoamino acid metabol	1/23	0.0132930	0.0242745	
<input type="checkbox"/>	Tyrosine metabolism	2/39	0.0195066	0.0341366	
<input type="checkbox"/>	Lysine biosynthesis	1/22	0.0350567	0.0588945	

Pathway enrichment result. Select the pathway of interest to highlight in the network

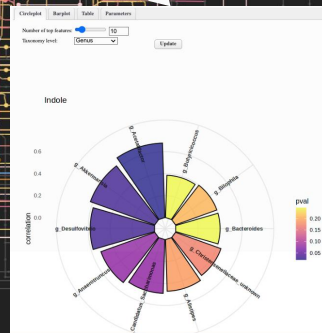
Click the node to see the most correlated taxa for the selected metabolite

KO / Metabolite Hits

- Tryptophan metabolism**
- o L-Tryptophan
 - o Indole
 - o 3-Hydroxyanthranilic acid
 - o Indolepyruvate

Note: IDs in red are not presented in the network

Detail information of the selected pathway



Correlation analysis:

Heatmap Visualization & Correlation:

Taxonomy level: Genus

Color contrast: Default

Heatmap font size: Column: 10.0 Row: 10.0

Similarity Method: Distance Correlation test

Correlation threshold: 0.2

Correlation significance: 0.1

Correlation sign: Positive only

Overlay with prediction heatmap

Prediction database: AGORA

Prediction potential score: 0.5

Prediction differential significance: 0.05

Parameters for correlation heatmap including correlation method, threshold, significance and direction

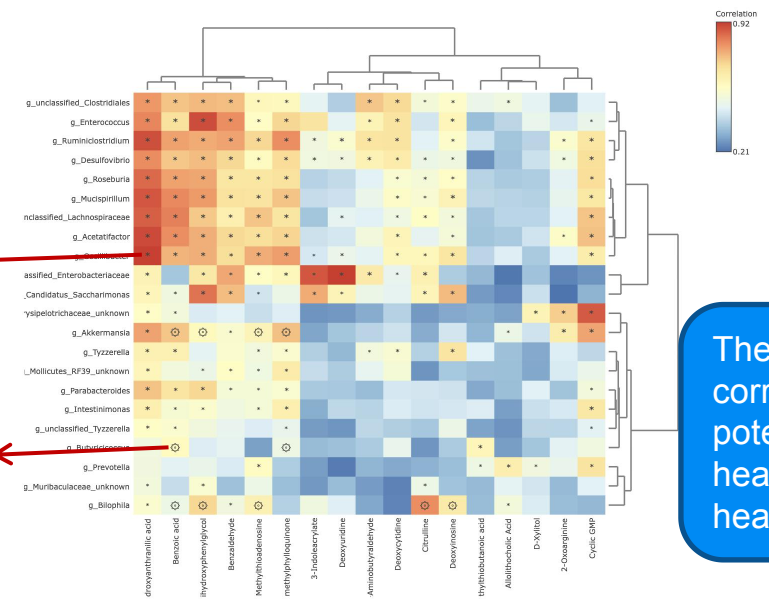
Set taxonomy level, color and font size for heatmap

Asterisk means the given correlation passes the setting significance

Diamond indicate the given correlation is identified as significant by both statistical correlation result and prediction result

Parameters for prediction heatmap including GEM database used for prediction model, potential score and significance

The color indicates the level of correlation and average potential score in the correlation heatmap and prediction heatmap, respectively.



II. Start from paired lists:

Data Upload
Please upload your paired microbiome and metabolomics data table separately

Abundance table **Feature List** Try our examples

1. Choose the Feature List tab

2. Specify the taxonomy level and ID type. Click upload to paste input your list in the text box.

3. Specify the metabolite or peak list here and upload accordingly.

4. Proceed to continue

Microbiome data: Taxonomy level: Phylum ID type: --- Please specify ---

Metabolomics data: Data Type: Targeted (compound list) ID type: --- Please specify ---

Metabolite only

--- Please specify ---
Taxon names
NCBI IDs

Targeted (compound list)
Targeted (compound list)
Untargeted (peak list)

--- Please specify ---
--- Please specify ---
Compound Name
HMDB ID
KEGG ID

Uploading taxon list:

Uploading compound list:

Uploading peak list:

Ion Mode: Positive Mode
Mass Tolerance (ppm): 5.0 (editable)
Retention Time: Not present
Ranked by (1 column only): P values Fold change T scores
Data File: + Choose

<< Previous >> Proceed

The image shows a web interface for data upload. The main section is titled 'Data Upload' and has a sub-header 'Please upload your paired microbiome and metabolomics data table separately'. There are three tabs: 'Abundance table', 'Feature List' (which is selected), and 'Try our examples'. Below the tabs, there are two main sections: 'Microbiome data' and 'Metabolomics data'. The 'Microbiome data' section has a dropdown menu for 'Taxonomy level' (set to 'Phylum') and another dropdown for 'ID type' (set to '--- Please specify ---'). The 'Metabolomics data' section has a dropdown for 'Data Type' (set to 'Targeted (compound list)') and another dropdown for 'ID type' (set to '--- Please specify ---'). There are two 'Upload' buttons, one for each section. To the right, there are three inset windows: 'Uploading taxon list:', 'Uploading compound list:', and 'Uploading peak list:'. The 'Uploading taxon list' window has a text box and a 'Submit' button. The 'Uploading compound list' window has a text box and a 'Submit' button. The 'Uploading peak list' window has several dropdowns and radio buttons for 'Ion Mode', 'Mass Tolerance', 'Retention Time', and 'Ranked by', and a 'Data File' section with a '+ Choose' button and a 'Submit' button. At the bottom, there are navigation buttons: '<< Previous' and '>> Proceed'. There are also four blue callout boxes with white text providing numbered instructions: 1. Choose the Feature List tab; 2. Specify the taxonomy level and ID type. Click upload to paste input your list in the text box.; 3. Specify the metabolite or peak list here and upload accordingly.; 4. Proceed to continue.

Name Mapping:

Input feature Mapping

All the input microbiome and metabolomics features need to be mapped to a specific taxon or compound for potential prediction. Two databases, Agora and EMBL GEMs are used for this purpose. To remove a taxon or compound from further analysis, use the Delete button.

GEMs are genome-scale metabolic models which can be used for metabolic prediction. The tables below show the matched taxon and compounds. To remove a

Check the name mapping results for each omics type by selecting the corresponding tab.

Taxon Name Mapping

Compound Name Mapping

For microbial taxonomy, KEGG, GEM (Agora and EMBL) and NCBI ID are provided.

For metabolites, KEGG and GEMs ID are provided.

Unwanted features can be deleted here.

Query	KEGG Hits	Agora Hits	EMBL Hits	NCBI Taxonomy	
Haemophilus_D	Haemophilus_D	Haemophilus_D	Haemophilus_D	724	Delete
Bacteroides	Bacteroides	Bacteroides	Bacteroides	816	Delete
Gemella	Gemella	Gemella	Gemella	1378	Delete
Faecalibacterium	Faecalibacterium	Faecalibacterium	Faecalibacterium	216851	Delete
Barnesiella	Barnesiella	Barnesiella	Barnesiella	397864	Delete
Pseudoruminococcus					Delete
Slackia_A	Slackia_A	Slackia_A	Slackia_A	84108	Delete
Granulicatella	Granulicatella	Granulicatella	Granulicatella	117563	Delete
Fusobacterium_C	Fusobacterium_C	Fusobacterium_C	Fusobacterium_C	848	Delete
Streptococcus	Streptococcus	Streptococcus	Streptococcus	1301	Delete
Acidaminococcus	Acidaminococcus	Acidaminococcus	Acidaminococcus	904	Delete

Methods Selection:

Please choose an integration method to proceed

Explore Overall Patterns via Dimensionality Reduction

Procrustes analysis (PA) is a simple visualization technique that superimposes the principal components of two datasets at a low-dimensional space. Procrustes essentially computes reduced dimensions for each data set using a method similar to PCA. Then, one of the reduced dimension matrices is rotated until it has maximum similarity with the other. Scores from both matrices are plotted at the same time, with pairs belonging to the same sample connected by a line. Procrustes is asymmetric, therefore the order that the 'omics datasets are uploaded will impact the results. ([more details ...](#))

Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO) is a supervised method for multi-omics biomarker exploration. It is based on a generalized version of PLS (multi block PLS-DA) that seeks to find related multi-dimensional components that maximally separate sample labels. DIABLO is symmetric with respect to the 'omics data, therefore the order that 'omics datasets are uploaded will not impact the results. ([more details ...](#))

Method selection:

Procrustes ▾

Submit

Not applicable for list input

Obtain Functional Insights via Pathways & Network

Microbiome data and metabolomics data are projected into KEGG metabolic network for visual exploration as well as enrichment analysis. The integration strategies are based on microbiome data types.

- Marker genes data will be used to customize the metabolic network for enrichment analysis of metabolomics data. Users can click a node to view the most correlated microbes of metabolites
- For the metagenomics data, both KOs and metabolomic features will be projected to the selected network for integration analysis.

Compound origin:

All uploaded taxa/ko ▾

Submit

Metabolic network and Heatmap visualization is same as the table input

Discover Metabolite-microbe Correlations via Statistics & GEMs

The relationships between the paired microbiome and metabolomics are intuitively presented using an interactive heatmap. Two

- Correlation heatmap is based on the statistical correlation.
- Prediction heatmap is based on GEM-based predictive models. The predictive models are logistic regression models to predict the potential of each metabolite production across different taxonomy levels. Prediction heatmap is only suitable for integration

Heatmap type:

Prediction heatmap ▾ ?

Submit

Only prediction heatmap is applicable for list input. It shows the potential of each taxon to produce each metabolite

The End



For more information, visit Tutorials, Resources
and Contact pages on www.microbiomeanalyst.ca
Also visit our forum for FAQs on www.omicsforum.ca