



MetaboAnalyst 5.0

A Web-based Tool for Streamlined
Metabolomics Data Analysis

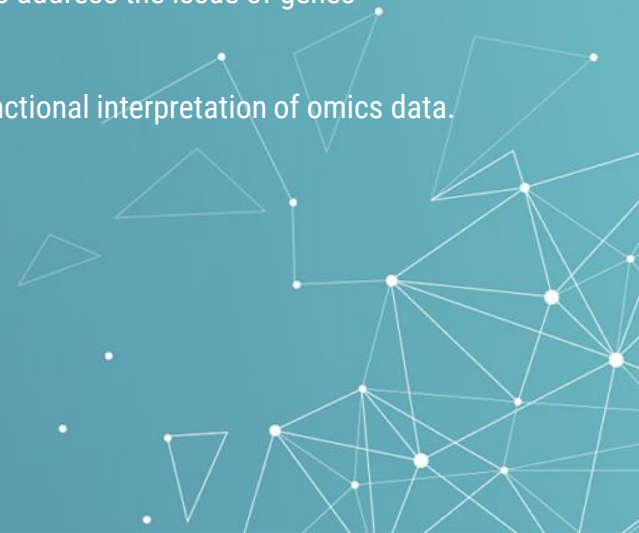
2021.04.15

4. Joint Pathway Analysis

The **Joint Pathway Analysis** module of MetaboAnalyst has added multiple enhancements for Version 5.

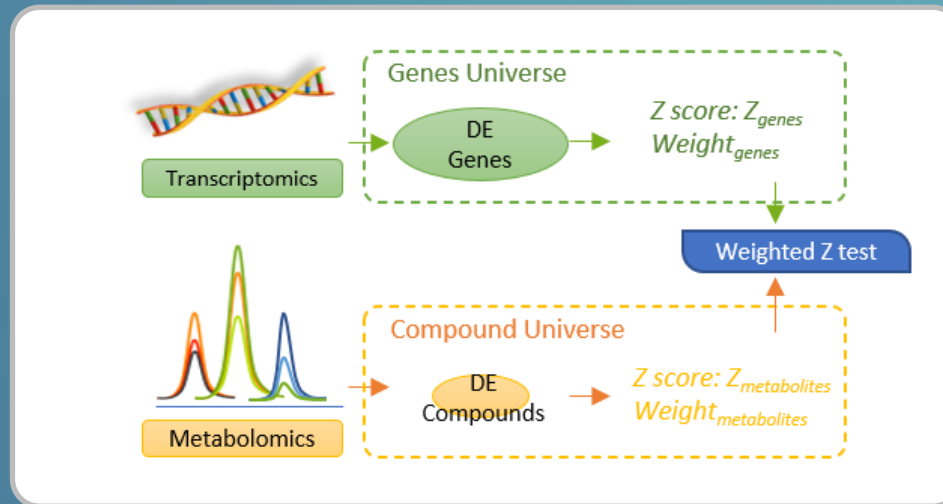
Highlights:

- Added support for an additional 22 model organisms.
- A weighted Z-tests based joint pathway analysis algorithm has been developed to address the issue of genes overwhelming integrated pathway analysis results.
- Multiple integration strategies have been implemented to further enhance the functional interpretation of omics data.



4.0 Knowledge & Background

- Joint Pathway Analysis is based on weighted integration. The weight strategy for different 'universes' ("transcriptomics data universe" and "metabolomics data universe") are performed with a weighted z-test. The weighted z-test is proposed for combination analysis [by Dmitri V. Zaykin](#) and leveraged here for the weighted integration of different datasets with significantly different sizes. Here we specifically prepared a figure to clearly illustrate the mechanism of this weighted integration of different Omics-data in MetaboAnalyst below.



4.0 Knowledge & Background

Specifically, different weights are assigned based on the proportion of genes and metabolites in the specific 'omics 'universes' to balance the influence from the different sizes of the 'omics inputs upon the integrated pathway results. The adjusted P value is estimated with a weighted Z-test as below,

$$P_{merged} = 1 - \Phi \left(\frac{\sum_{i=1}^2 w_i Z_i}{\sqrt{\sum_{i=1}^2 w_i^2}} \right)$$

Where w_i is the weights of the P values of genes or compounds within individual omics "universe" or "pathway space", respectively; Z_i is the Z score of the corresponding P values of single omics data, usually, $Z_i = \Phi^{-1}(1 - P_i)$; P_i is the P values from the enrichment analysis above; Φ denote the standard normal cumulative distribution function.



4.1 Start Joint Pathway Analysis

MetaboAnalyst 5.0 - user-friendly, end-to-end metabolomics data analysis

Module Overview

Input Data Type	Available Modules (click on a module to proceed, or scroll down for more details)					
Raw Spectra (mzML, mzXML or mzData)				LC-MS Spectral Processing		
MS Peaks (peak list or intensity table)			Functional Analysis	Functional Meta-analysis		
Annotated Features (compound list or table)		Enrichment Analysis	Pathway Analysis	Joint-Pathway Analysis	Network Analysis	
Generic Format (.csv or .txt table files)	Statistical Analysis	Biomarker Analysis	Time-series/Two-factor Analysis	Statistical Meta-analysis	Power Analysis	Other Utilities

Click here to start

Show R command history

4.2 Data Upload



MetaboAnalyst 5.0 - user-friendly, end-to-end metabolomics data analysis

TIP: The Fold change is optional. The titles of the 2 columns need to start with '#'.

Upload

Integrative Analysis

Download

Exit

Please upload a gene list and a metabolite list below

Gene list with optional fold changes

```
#Official logFC
AASS -0.139042168
ACAA2 1.401267672
ACADL -2.608712824
ACADM -0.876538515
ACADS 0.150535255
ACADSB -1.637743607
ACHE 2.567118372
ACSM1 -2.348501729
ACTA2 -0.282176735
ACTB 1.559623747
ACTC1 -1.690352151
ADCY1 2.916857724
ADH1A -0.87610472
AGL -0.399133917
AGTR1 -1.078340189
AKR1A1 2.178398898
AKR1B1 -1.077265882
```

Compound list with optional fold changes

```
#KEGG logFC
C00006 0.512160717
C00024 0.351757155
C00026 -2.669056963
C00029 0.379186578
C00031 1.669222153
C00047 -2.492289379
C00049 2.963835134
C00062 -2.558919927
C00064 1.77810046
C00072 0.632536475
C00077 -2.0904508
C00084 0.347392968
C00089 -1.460843412
C00097 3.046798674
C00101 -1.495004303
C00109 0.476718643
C00111 -2.672997377
```

1. Copy and paste your list of genes and metabolites.

2. Specify the ID type.

ID Type: Official Gene Symbol

ID Type: KEGG ID

Specify organism: Homo sapiens (human)

Try our example data

Submit

3. Click "Submit" to upload the data.

4.3 Name Matching Results



MetaboAnalyst 5.0 - from raw spectra to patterns and biological insights



Upload

Integrative Analysis

Diagnose

Set parameter

View result

Download

Exit

The system requires all the IDs (except common compound names) to be matched exactly. The table below shows the matched genes and compounds from the databases. In some cases, users can further perform [approximate match](#) by clicking the **View** link in the Details column. To **remove** a gene or compound from further analysis, click the **Delete** link in the **Details** column.

Compound Name Mapping Gene Name Mapping

Query	Hit	HMDB	KEGG	Details
C00006	NADP	HMDB0000217	C00006	Delete
C00024	Acetyl-CoA	HMDB0001206	C00024	Delete
C00026	Oxoglutaric acid	HMDB0000208	C00026	Delete
C00029	Uridine diphosphate glucose	HMDB0000286	C00029	Delete
C00031	D-Glucose	HMDB0000122	C00031	Delete
C00047	L-Lysine	HMDB0000182	C00047	Delete
C00049	L-Aspartic acid	HMDB0000191	C00049	Delete
C00062	L-Arginine	HMDB0000517	C00062	Delete
C00064	L-Glutamine	HMDB0000541	C00064	Delete
C00072	Ascorbate	-	C00072	Delete
C00077	Ornithine	HMDB0000214	C00077	Delete
C00084	Acetaldehyde	HMDB0000990	C00084	Delete
C00089	Sucrose	HMDB0000258	C00089	Delete
C00097	L-Cysteine	HMDB0000574	C00097	Delete
C00101	Tetrahydrofolic acid	HMDB0001846	C00101	Delete
C00109	2-Ketobutyric acid	HMDB0000005	C00109	Delete
C00111	Dihydroxyacetone phosphate	HMDB0001473	C00111	Delete
C00117	D-Ribose 5-phosphate	HMDB0001548	C00117	Delete
C00122	Fumaric acid	HMDB0000134	C00122	Delete
C00141	Alpha-ketoglutaric acid	HMDB0000019	C00141	Delete

OK
A total of 389 unique genes were uploaded.

OK
Name matching OK, please inspect (and manual correct) the results then proceed.

Results of the name mapping of the uploaded gene / metabolite data to MetaboAnalyst's internal databases. Scroll down and click "Submit" to continue.

4.4 Parameter Selection

1. Select whether to use only metabolic pathways, all pathways, metabolite-only or gene-only pathways.

2. Next, decide on algorithm parameters such as how to combine the gene / metabolite data and how to score important nodes.

3. Scroll down and click "Submit" to perform the Joint Pathway Analysis.

TIP: Please carefully read the instructions of the different options of algorithms before submitting your data.

MetaboAnalyst 5.0 - from raw spectra to patterns and biological insights

Pathway Database

The KEGG pathway was updated in October, 2019 using KEGG API. The integration can occur in two different "universes" defined by **metabolic pathways** or **all pathways**. Metabolic pathways include pathways containing both metabolites and metabolic genes while all pathways include both metabolic pathways as well as gene-only pathways (i.e. regulatory pathways). Users can also perform enrichment analysis for metabolites only using **metabolic pathways (metabolite only)** or for genes only using **all pathways (gene only)**.

Metabolic pathways (integrated)
 All pathways (integrated)
 Metabolic pathways (metabolite only)
 All pathways (gene only)

Algorithm Selection

The topology analysis evaluates the potential importance of a particular molecule (a node) based on its **position** within a pathway. **Degree Centrality** measures the number of links that connect to a node. **Betweenness Centrality** measures the number of shortest paths from all nodes to all the others that pass through a given node. **Closeness Centrality** measures the overall distance from a given node to all other nodes.

For integration methods, there are two general approaches - **tight integration by combining queries** in which genes and metabolites are pooled into a single query and used to perform enrichment analysis within their "pooled universe" or **loose integration by combining p-values** in which enrichment analysis is performed separately for genes and metabolites in their "individual universe", and then individual p-values are combined via **weighted Z-tests**. Moreover, there are three options for computing weights. Let's assume the pathway database contains a total of 100 pathways covering a total of 1000 metabolites and 4000 genes, respectively. Pathway A contains 5 compounds and 45 genes, while pathway B contains 20 compounds and 30 genes.

- **Unweighted** or equal weights (i.e. metabolite: 0.5, gene: 0.5);
- Weights based on the **overall** proportion of each omics within the "universe" (i.e. metabolite: 0.2, gene: 0.8 for all pathways);
- Weights based on the **pathway-level** proportion within individual "pathway space" (i.e. pathway A - metabolite 0.1, gene 0.9; pathway B - metabolite 0.4, gene 0.6)

Note that combining p-values can only be applied to pathways receiving hits from both input types. For pathways with hits from only one input type, p values calculated from their individual universe will be used. In this case, **combining p-values can be viewed as adjusting the confidence level based on new evidence (i.e. input from another omics layer)**. If no new evidence is available, the current confidence level remains.

Enrichment analysis: Hypergeometric Test Fisher's Exact Test

Topology measure: Degree Centrality Betweenness Centrality Closeness Centrality

Integration method: Combine queries
 Combine p values (unweighted)
 Combine p values (overall)
 Combine p values (pathway-level)

Submit

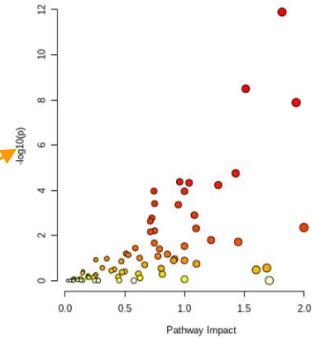
Xia Lab @ McGill (last updated 2020-12-14)

4.5 Parameter Selection

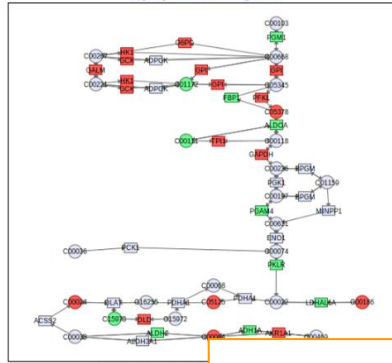


MetaboAnalyst 5.0 - from raw spectra to patterns and biological insights

Overview of Pathway Analysis



Glycolysis or Gluconeogenesis



Click the corresponding Pathway Name to view its graphical presentation; click Match Status to view the pathway members (with matched genes)

Pathway Name	Match Status	p	-log(p)	Holm p	
Glycolysis or Gluconeogenesis	28/81	1.2805E-12	11.893	1.0766E-10	
Fructose and mannose metabolism	19/49	3.1764E-9	8.4981	2.6364E-7	1.3341E-7 1.5128 KEGG
Fructose and sucrose metabolism	20/47	1.2988E-8	7.8855	1.065E-6	3.6367E-7 1.9348 KEGG
Starch and sucrose metabolism	15/43	1.7821E-5	4.754	0.0014273	3.7004E-4 1.4286 KEGG
Galactose metabolism	16/51	4.1966E-5	4.3771	0.0033573	6.5122E-4 0.96 KEGG
Arginine biosynthesis	11/27	4.6516E-5	4.3324	0.0036748	6.5122E-4 1.0385 KEGG
Retinol metabolism	15/47	5.8685E-5	4.2315	0.0045774	7.0422E-4 1.2826 KEGG
beta-Alanine metabolism	14/44	1.082E-4	3.9658	0.0083311	0.0010314 0.74419 KEGG
Pantothenic acid and CoA biosynthesis	12/34	1.1051E-4	3.9566	0.0083989	0.0010314 1.0 KEGG
Lysine degradation	14/49	3.8705E-4	3.4122	0.029029	0.0032513 0.75 KEGG
Amino sugar and nucleotide sugar metabolism	19/79	4.3262E-4	3.3639	0.032014	0.0033036 0.94872 KEGG

Scatterplot summary of pathway analysis results. Double-click a node to view the corresponding KEGG pathway.

View the pathway matching details by clicking the Match Status.

Matched Features

Pathway	Members
Fructose and mannose metabolism	D-Sorbitol; D-Fructose; D-Mannose; beta-D-Fructose 2,6-bisphosphate; D-Mannose 6-phosphate; L-Fucose 1-phosphate; GDP-4-dehydro-6-deoxy-D-mannose; GDP-mannose; D-Mannose 1-phosphate; beta-D-Fructose 6-phosphate; D-Glyceraldehyde 3-phosphate; D-Fructose 1-phosphate; D-Glyceraldehyde beta-D-Fructose 1,6-bisphosphate; 6-Deoxy-L-galactose; L-Fucosate; alpha-D-Glucose; GDP-L-fucose; Glycerone phosphate; 2-Dehydro-3-deoxy-L-fucosate; AKR1B1, ADR, ALDR1, ALR2, AR..., SORD, HEL, S-95n, RDH, SDH, SORD1, XDH, KHK, HK1, HK, HK1-ta, HK1-tb, HK1-tc, HKD, HKI, HMSNR, HXK1, RP79, hexokinase..., PFKFB1, F6PK, HL2K, PFK..., MPI, CDG1B, PMI, PM1; PAM1; PM1, 1, PM1h-22, Sec53..., FPGT, GFFP, 1S1A3, FX, P35B, SDR4E1, GMD5, GMD, SDR3E1, GMPPB, LGMDR19, MDDG14, MDDGB14, MDDGC14..., PFKL, ATP-PFK, PFK-B, PFK-L..., TPI1, HEL, S-49, TIM, TPI, TPI0, ALDOA, ALDA, GSD12, HEL, S-87n..., TKFC, DAK, NET45, FBP..., FCSK, 1110046B12RIK, COGF2, FUK, ENOSF1, FUCC, RTS, TYMSAS

Link to the KEGG database of the specific pathway.

Thanks

*If you have any questions please read through the FAQs or contact us at
[Zhiqiang.pang\[at\]xialab.ca](mailto:Zhiqiang.pang@xialab.ca) or [Jasmine.chong\[at\]xialab.ca](mailto:Jasmine.chong@xialab.ca).*

