Metabolomic Data Analysis with MetaboAnalyst 5.0

Name: guest5758204379707340938

May 21, 2021

1 Background

The Pathway Analysis module combines results from powerful pathway enrichment analysis with pathway topology analysis to help researchers identify the most relevant pathways involved in the conditions under study.

There are many commercial pathway analysis software tools such as Pathway Studio, MetaCore, or Ingenuity Pathway Analysis (IPA), etc. Compared to these commercial tools, the pathway analysis module was specifically developed for metabolomics studies. It uses high-quality KEGG metabolic pathways as the backend knowledgebase. This module integrates many well-established (i.e. univariate analysis, over-representation analysis) methods, as well as novel algorithms and concepts (i.e. Global Test, GlobalAncova, network topology analysis) into pathway analysis. Another feature is a Google-Map style interactive visualization system to deliver the analysis results in an intuitive manner.

2 Data Input

The Pathway Analysis module accepts either a list of compound labels (common names, HMDB IDs or KEGG IDs) with one compound per row, or a compound concentration table with samples in rows and compounds in columns. The second column must be phenotype labels (binary, multi-group, or continuous). The table is uploaded as comma separated values (.csv).

3 Compound Name Matching

The first step is to standardize the compound labels used in user uploaded data. This is a necessary step since these compounds will be subsequently compared with compounds contained in the pathway library. There are three outcomes from the step - exact match, approximate match (for common names only), and no match. Users should click the textbfView button from the approximate matched results to manually select the correct one. Compounds without match will be excluded from the subsequently pathway analysis.

Table 1 shows the conversion results. Note: *1* indicates exact match, *2* indicates approximate match, and *0* indicates no match. A text file contain the result can be found the downloaded file *name_map.csv*

Table

	Query	Match	HMDB	PubChem	KEGG
1	HMDB0014633	Buspirone	HMDB0014633	2477	C06861
2	HMDB0030844	6,7-Dimethoxy-7-epirosmanol	HMDB0030844	131751086	
3	HMDB0031933	11,12-Dihydroxy-7,14-dimethoxy-8,11,13-abietatrien-20,6-olide	HMDB0031933	76401227	
4	HMDB0035922	Nigakihemiacetal B	HMDB0035922	283906	C08771
5	HMDB0035206	Bakkenolide B	HMDB0035206	101289733	
6	HMDB0240335	NA	NA	NA	NA
7	HMDB0240334	NA	NA	NA	NA
8	HMDB0173904	NA	NA	NA	NA
9	HMDB0173896	NA	NA	NA	NA

10	HMDB0173898	NA	NA	NA	NA
11	HMDB0173900	NA	NA	NA	NA
12	HMDB0173906	NA	NA	NA	NA
13	HMDB0173912	NA	NA	NA	NA
14	HMDB0173914	NA	NA	NA	NA
15	HMDB0173902	NA	NA	NA	NA
16	HMDB0161139	NA	NA	NA	NA
17	HMDB0062332	NA	NA	NA	NA
18	HMDB0062325	NA	NA	NA	NA
19	HMDB0059725	NA	NA	NA	NA
20	HMDB0140036	NA	NA	NA	NA
21	HMDB0035091	beta-Citraurin	HMDB0035091	131751663	
22	HMDB0036885	NA	NA	NA	NA
23	HMDB0030898	(22E, 24x)-Ergosta-4,6,8,22-tetraen-3-one	HMDB0030898	12943211	
24	HMDB0000637	Chenodeoxycholic acid glycine conjugate	HMDB0000637	22833540	C05466
25	HMDB0000708	Glycoursodeoxycholic acid	HMDB0000708	12310288	
26	HMDB0184641	NA	NA	NA	NA
27	HMDB0161140	NA	NA	NA	NA
28	HMDB0006898	Chenodeoxyglycocholic acid	HMDB0006898	53477907	C05462
29	HMDB0161141	NA	NA	NA	NA
30	HMDB0000631	Deoxycholic acid glycine conjugate	HMDB0000631	3035026	C05464
31	HMDB0173224	NA	NA	NA	NA
32	HMDB0173223	NA	NA	NA	NA
33	HMDB0173227	NA	NA	NA	NA
34	HMDB0173228	NA	NA	NA	NA
35	HMDB0173226	NA	NA	NA	NA
36	HMDB0173225	NA	NA	NA	NA
37	HMDB0012516	11'-Carboxy-alpha-tocotrienol	HMDB0012516	53481452	
38	HMDB0030140	Adlupulone	HMDB0030140		
39	HMDB0030041	Lupulone	HMDB0030041	51397980	C10706
40	HMDB0173017	NA	NA	NA	NA
41	HMDB0030554	epsilon-Tocopherol	HMDB0030554	9844470	C14154
42	HMDB0037380	NA	NA	NA	NA
43	HMDB0173018	NA	NA	NA	NA
44	HMDB0012958	Gamma-Tocotrienol	HMDB0012958	5282349	C14155
45	HMDB0032106	(3beta,22E,24R)-3-Hydroxyergosta-5,8,22-trien-7-one	HMDB0032106	131751258	
46	HMDB0015073	Salmeterol	HMDB0015073	5152	C07241
47	HMDB0015041	Bimatoprost	HMDB0015041	5311027	
48	HMDB0014792	Latanoprost	HMDB0014792	5311221	
49	HMDB0175691	NA	NA	NA	NA
50	HMDB0034515	Glabrolide	HMDB0034515	14187213	
51	HMDB0035886	Isoglabrolide	HMDB0035886	15559941	
52	HMDB0032837	Ganoderic acid DM	HMDB0032837	131751329	
53	HMDB0040711	NA	NA	NA	NA
54	HMDB0038797	NA	NA	NA	NA
55	HMDB0035434	Isomasticadienonalic acid	HMDB0035434	131751752	
56	HMDB0039615	NA	NA	NA	NA
57	HMDB0172977	NA	NA	NA	NA
58	HMDB0172979	NA	NA	NA	NA
59	HMDB0172978	NA	NA	NA	NA
60	HMDB0172975	NA	NA	NA	NA
61	HMDB0172980	NA	NA	NA	NA
62	HMDB0172973	NA	NA	NA	NA
63	HMDB0172974	NA	NA	NA	NA
64	HMDB0172972	NA	NA	NA	NA
65	HMDB0172971	NA	NA	NA	NA
66	HMDB0172970	NA	NA	NA	NA
67	HMDB0062385	NA	NA	NA	NA
68	HMDB0172976	NA	NA	NA	NA
69	HMDB0031953	3,5-Dihydroxyergosta-7,22-dien-6-one	HMDB0031953	6293947	
70	HMDB0037941	NA	NA	NA	NA
71	HMDB0015694	Nandrolone decanoate	HMDB0015694	9677	C08154
72	HMDB0032108	(3beta,5alpha,6alpha,22E,24R)-Ergosta-7,9(11),22-triene-3,5,6-triol	HMDB0032108	131751260	
73	HMDB0039602	NA	NA	NA	NA
74	HMDB0034541	5,6-Epoxiergosta-8,22-diene-3,7-diol	HMDB0034541	14841775	
75	HMDB0032863	(3beta,5alpha,6alpha,7beta,22E,24R)-5,6-Epoxyergosta-8(14),22-diene-3,7-diol	HMDB0032863	131751336	
76	HMDB0173019	NA	NA	NA	NA
77	HMDB0006225	Ercalcitriol	HMDB0006225	9547243	
78	HMDB0030020	Withanolide B	HMDB0030020	14236712	C00828
79	HMDB0038705	NA	NA	NA	NA
80	HMDB0030037	Euglobal VII	HMDB0030037	131750947	
81	HMDB0030166	Euglobal III	HMDB0030166	131750974	
82	HMDB0030035	Euglobal IVa	HMDB0030035	131750946	
83	HMDB0030038	Euglobal V	HMDB0030038	5317276	
84	HMDB0166712	NA	NA	NA	NA
85	HMDB0170374	NA	NA	NA	NA
86	HMDB0171571	NA	NA	NA	NA
87	HMDB0175475	NA	NA	NA	NA
88	HMDB0175479	NA	NA	NA	NA
89	HMDB0175485	NA	NA	NA	NA
90	HMDB0175481	NA	NA	NA	NA
91	HMDB0175474	NA	NA	NA	NA
92	HMDB0175488	NA	NA	NA	NA
93	HMDB0175477	NA	NA	NA	NA
94	HMDB0175491	NA	NA	NA	NA
95	HMDB0161458	NA	NA	NA	NA
96	HMDB0161457	NA	NA	NA	NA

97	HMDB0161460	NA	NA	NA	NA
98	HMDB0161463	NA	NA	NA	NA
99	HMDB0161464	NA	NA	NA	NA
100	HMDB0161461	NA	NA	NA	NA
101	HMDB0183609	NA	NA	NA	NA
102	HMDB0114751	NA	NA	NA	NA
103	HMDB0062306	NA	NA	NA	NA
104	HMDB0010380	LysoPC(14:1(9Z))	HMDB0010380	24779456	C04230
105	HMDB0032799	(R)-1-O-[b-D-Glucopyranosyl-(1->6)-b-D-glucopyranoside]-1,3-octanediol	HMDB0032799	131751313	
106	HMDB0186557	NA	NA	NA	NA
107	HMDB0186516	NA	NA	NA	NA
108	HMDB0186517	NA	NA	NA	NA
109	HMDB0186519	NA	NA	NA	NA
110	HMDB0177058	NA	NA	NA	NA
111	HMDB0177057	NA	NA	NA	NA
112	HMDB0186520	NA	NA	NA	NA
113	HMDB0186522	NA	NA	NA	NA
114	HMDB0186556	NA	NA	NA	NA
115	HMDB0186553	NA	NA	NA	NA
116	HMDB0186552	NA	NA	NA	NA
117	HMDB0186523	NA	NA	NA	NA
118	HMDB0186526	NA	NA	NA	NA
119	HMDB0186525	NA	NA	NA	NA
120	HMDB0186531	NA	NA	NA	NA
121	HMDB0186528	NA	NA	NA	NA
122	HMDB0186529	NA	NA	NA	NA
123	HMDB0186535	NA	NA	NA	NA
124	HMDB0186534	NA	NA	NA	NA
125	HMDB0186540	NA	NA	NA	NA
126	HMDB0186532	NA	NA	NA	NA
127	HMDB0186537	NA	NA	NA	NA
128	HMDB0186549	NA	NA	NA	NA
129	HMDB0186538	NA	NA	NA	NA
130	HMDB0186543	NA	NA	NA	NA
131	HMDB0186546	NA	NA	NA	NA
132	HMDB0186541	NA	NA	NA	NA
133	HMDB0008680	PC(22:5(4Z,7Z,10Z,13Z,16Z)/22:5(4Z,7Z,10Z,13Z,16Z))	HMDB0008680	53479323	C00157
134	HMDB0008681	PC(22:5(4Z,7Z,10Z,13Z,16Z)/22:5(7Z,10Z,13Z,16Z,19Z))	HMDB0008681	53479325	C00157
135	HMDB0008713	PC(22:5(7Z,10Z,13Z,16Z,19Z)/22:5(4Z,7Z,10Z,13Z,16Z))	HMDB0008713	53479389	C00157
136	HMDB0008714	PC(22:5(7Z,10Z,13Z,16Z,19Z)/22:5(7Z,10Z,13Z,16Z,19Z))	HMDB0008714	53479391	C00157
137	HMDB0008745	PC(22:6(4Z,7Z,10Z,13Z,16Z,19Z)/22:4(7Z,10Z,13Z,16Z))	HMDB0008745	52923727	C00157
138	HMDB0008649	PC(22:4(7Z,10Z,13Z,16Z)/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	HMDB0008649	52923669	C00157
139	HMDB0009607	PE(22:4(7Z,10Z,13Z,16Z)/24:1(15Z))	HMDB0009607	53479904	C00350
140	HMDB0009640	PE(22:5(4Z,7Z,10Z,13Z,16Z)/24:0)	HMDB0009640	53479936	C00350
141	HMDB0009673	PE(22:5(7Z,10Z,13Z,16Z,19Z)/24:0)	HMDB0009673	53479969	C00350
142	HMDB0009736	PE(24:0/22:5(4Z,7Z,10Z,13Z,16Z))	HMDB0009736	53480008	C00350
143	HMDB0009737	PE(24:0/22:5(7Z,10Z,13Z,16Z,19Z))	HMDB0009737	53480009	C00350
144	HMDB0009768	PE(24:1(15Z)/22:4(7Z,10Z,13Z,16Z))	HMDB0009768	53480040	C00350
145	HMDB0114516	NA	NA	NA	NA
146	HMDB0116719	NA	NA	NA	NA
147	HMDB0114568	NA	NA	NA	NA
148	HMDB0114460	NA	NA	NA	NA
149	HMDB0114623	NA	NA	NA	NA
150	HMDB0114648	NA	NA	NA	NA
151	HMDB0114649	NA	NA	NA	NA
152	HMDB0116714	NA	NA	NA	NA
153	HMDB0114434	NA	NA	NA	NA
154	HMDB0114351	NA	NA	NA	NA
155	HMDB0114380	NA	NA	NA	NA
156	HMDB0114408	NA	NA	NA	NA
157	HMDB0008648	PC(22:4(7Z,10Z,13Z,16Z)/22:5(7Z,10Z,13Z,16Z,19Z))	HMDB0008648	53479261	C00157
158	HMDB0008712	PC(22:5(7Z,10Z,13Z,16Z,19Z)/22:4(7Z,10Z,13Z,16Z))	HMDB0008712	53479387	C00157
159	HMDB0008679	PC(22:5(4Z,7Z,10Z,13Z,16Z)/22:4(7Z,10Z,13Z,16Z))	HMDB0008679	53479321	C00157
160	HMDB0008647	PC(22:4(7Z,10Z,13Z,16Z)/22:5(4Z,7Z,10Z,13Z,16Z))	HMDB0008647	53479259	C00157

4 Pathway Analysis

In this step, users are asked to select a pathway library, as well as specify the algorithms for pathway enrichment analysis and pathway topology analysis.

4.1 Pathway Library

There are 15 pathway libraries currently supported, with a total of 1173 pathways :

- Homo sapiens (human) [80]
- Mus musculus (mouse) [82]
- Rattus norvegicus (rat) [81]
- Bos taurus (cow) [81]
- Danio rerio (zebrafish) [81]
- Drosophila melanogaster (fruit fly) [79]
- Caenorhabditis elegans (nematode) [78]
- Saccharomyces cerevisiae (yeast) [65]
- Oryza sativa japonica (Japanese rice) [83]
- Arabidopsis thaliana (thale cress) [87]
- Escherichia coli K-12 MG1655 [87]
- Bacillus subtilis [80]
- Pseudomonas putida KT2440 [89]
- Staphylococcus aureus N315 (MRSA/VSSA)[73]
- Thermotoga maritima [57]

Your selected pathway library code is **hsa** (KEGG organisms abbreviation).

4.2 Over Representation Analysis

Over-representation analysis tests if a particular group of compounds is represented more than expected by chance within the user uploaded compound list. In the context of pathway analysis, we are testing if compounds involved in a particular pathway are enriched compared to random hits. MetPA offers two of the most commonly used methods for over-representation analysis:

- Fishers'Exact test
- Hypergeometric Test

Please note, MetPA uses one-tailed Fisher's exact test which will give essentially the same result as the result calculated by the hypergeometric test.

The selected over-representation analysis method is **Hypergeometric test**.

4.3 Pathway Topology Analysis

The structure of biological pathways represent our knowledge about the complex relationships among molecules within a cell or a living organism. However, most pathway analysis algorithms fail to take structural information into consideration when estimating which pathways are significantly changed under conditions of study. It is well-known that changes in more important positions of a network will trigger a more severe impact on the pathway than changes occurred in marginal or relatively isolated positions.

The pathway topology analysis uses two well-established node centrality measures to estimate node importance - **degree centrality** and **betweenness centrality**. Degree centrality is defined as the number of links occurred upon a node. For a directed graph there are two types of degree: in-degree for links come from other nodes, and out-degree for links initiated from the current node. Metabolic networks are directed graph. Here we only consider the out-degree for node importance measure. It is assumed that nodes upstream will have regulatory roles for the downstream nodes, not vice versa. The betweenness centrality measures the number of shortest paths going through the node. Since the metabolic network is directed, we use the relative betweenness centrality for a metabolite as the importance measure. The degree centrality measure focuses more on local connectivities, while the betweenness centrality measure focuses more on global network topology. For more detailed discussions on various graph-based methods for analyzing biological networks, please refer to the article by Tero Aittokallio, T. et al. [1]

Please note, for comparison among different pathways, the node importance values calculated from centrality measures are further normalized by the sum of the importance of the pathway. Therefore, the total/maximum importance of each pathway is 1; the importance measure of each metabolite node is actually the percentage w.r.t the total pathway importance, and the pathway impact value is the cumulative percentage from the matched metabolite nodes.

Your selected node importance measure for topological analysis is **relative betweenness centrality**.

5 Pathway Analysis Result

The results from pathway analysis are presented graphically as well as in a detailed table.

A Google-map style interactive visualization system was implemented to facilitate data exploration. The graphical output contains three levels of view: **metabolome view**, **pathway view**, and **compound view**. Only the metabolome view is shown below. Pathway views and compound views are generated dynamically based on your interactions with the visualization system. They are available in your downloaded files.

[1] Tero Aittokallio and Benno Schwikowski. *Graph-based methods for analyzing networks in cell biology*, Briefings in Bioinformatics 2006 7(3):243-255

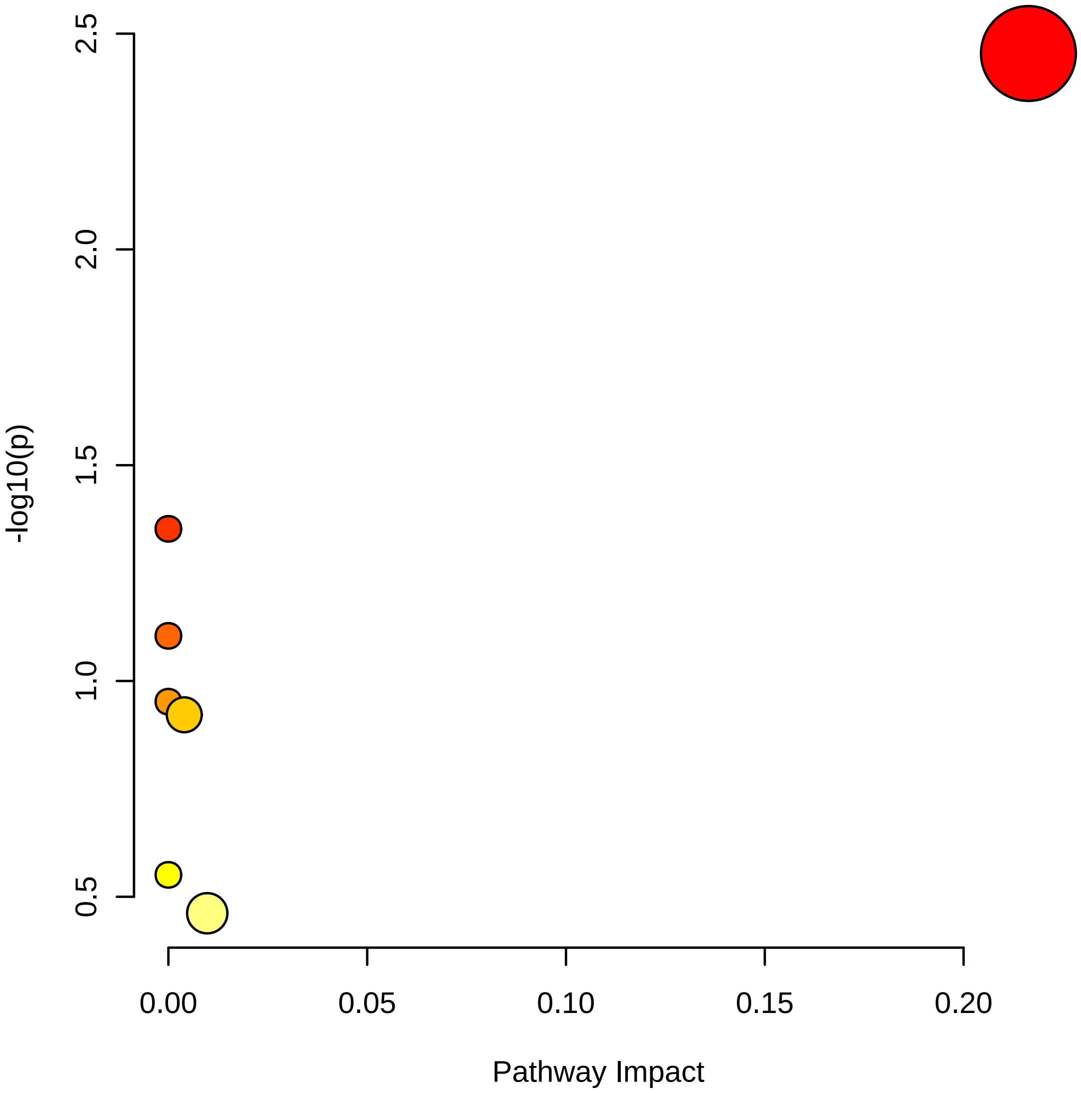

Figure 1: Summary of Pathway Analysis

The table below shows the detailed results from the pathway analysis. Since we are testing many pathways at the same time, the statistical p values from enrichment analysis are further adjusted for multiple testings. In particular, the **Total** is the total number of compounds in the pathway; the **Hits** is the actually matched number from the user uploaded data; the **Raw p** is the original p value calculated from the enrichment analysis; the **Holm p** is the p value adjusted by Holm-Bonferroni method; the **FDR p** is the p value adjusted using False Discovery Rate; the **Impact** is the pathway impact value calculated from pathway topology analysis.

Table 2: Result from Pathway Analysis

	Total	Expected	Hits	Raw p	-log10(p)	Holm adjust	FDR	Impact
Glycerophospholipid metabolism	36	0.33	3	3.52E-03	2.45E+00	2.95E-01	2.95E-01	0.22
Linoleic acid metabolism	5	0.05	1	4.44E-02	1.35E+00	1.00E+00	1.00E+00	0.00
Ubiquinone and other terpenoid-quinone biosynthesis	9	0.08	1	7.86E-02	1.10E+00	1.00E+00	1.00E+00	0.00
alpha-Linolenic acid metabolism	13	0.12	1	1.12E-01	9.52E-01	1.00E+00	1.00E+00	0.00
Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	14	0.13	1	1.20E-01	9.22E-01	1.00E+00	1.00E+00	0.00
Arachidonic acid metabolism	36	0.33	1	2.81E-01	5.51E-01	1.00E+00	1.00E+00	0.00
Primary bile acid biosynthesis	46	0.42	1	3.45E-01	4.62E-01	1.00E+00	1.00E+00	0.01

6 Appendix: R Command History

```
 [1] "mSet<-InitDataObjects(\"conc\", \"pathora\", FALSE)"
 [2] "cmpd.vec<-c(\"HMDB0014633\",\"HMDB0030844\",\"HMDB0031933\",\"HMDB0035922\",\"HMDB0035206\",\"
 [3] "mSet<-Setup.MapData(mSet, cmpd.vec);"
 [4] "mSet<-CrossReferencing(mSet, \"hmdb\");"
 [5] "mSet<-CreateMappingResultTable(mSet)"
 [6] "mSet<-SetKEGG.PathLib(mSet, \"hsa\", \"current\")"
 [7] "mSet<-SetMetabolomeFilter(mSet, F);"
 [8] "mSet<-CalculateOraScore(mSet, \"rbc\", \"hyperg\")"
 [9] "mSet<-PlotPathSummary(mSet, F, \"path_view_0_\", \"png\", 72, width=NA)"
[10] "mSet<-SaveTransformedData(mSet)"
[11] "mSet<-PreparePDFReport(mSet, \"guest5758204379707340938\")\n"
```

The report was generated on Fri May 21 16:13:52 2021 with R version 4.0.5 (2021-03-31).