Metabolomic Data Analysis with MetaboAnalyst 5.0

Name: guest6793469145423477714

October 22, 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 1: Resu

	Query	Match	HMDB	PubChem	KEGG	SMILES
1	1 3-Dimethyluric acid	1,3-Dimethyluric acid	HMDB0001857	70346		CN1C2=C(C(=
2	1-Methylhistidine	1-Methylhistidine	HMDB0000001	92105	C01152	CN1C=C(N=C
3	11a-Hydroxyprogesterone	11a-Hydroxyprogesterone	HMDB0000920	440105	C03747	CC(=O)C1CC[
4	2-Aminoethylphosphonate	Ciliatine	HMDB0011747	339	C03557	C(CP(=O)(O)(
5	3-Phenylbutyric acid	3-Phenylbutyric acid	HMDB0001955	20724		CC(CC(=O)O)
6	4-Guanidinobutanoic acid	4-Guanidinobutanoic acid	HMDB0003464	500	C01035	C(CC(=O)O)C
7	5- Methylthioadenosine	5'-Methylthioadenosine	HMDB0001173	439176	C00170	CSC[C@@H]1[(
8	5-Hydroxytryptophan	5-Hydroxy-L-tryptophan	HMDB0000472	144	C00643	C1=CC2=C(C=
9	5-Methoxytryptamine	5-Methoxytryptamine	HMDB0004095	1833	C05659	COC1=CC2=C

10	5-Methylcytosine	5-Methylcytosine	HMDB0002894	65040	C02376	CC1=C(NC(=O
11	5-Phenylvaleric acid	5-Phenylvaleric acid	HMDB0002043	16757		C1=CC=C(C=
12	a-ketoglutarate	Oxoglutaric acid	HMDB0000208	51	C00026	C(CC(=O)O)C
13	Acetyl glycine	Phenylacetylglycine	HMDB0000821	68144	C05598	C1=CC=C(C=
14	Acetyl proline	NA	NA	NA	NA	NA
15	Acetyl-arginine	N-a-Acetyl-L-arginine	HMDB0004620	67427		CC(=O)N[C@
16	Acetyllysine	N-Alpha-acetyllysine	HMDB0000446	192590	C12989	CC(=O)NC(C
17	Aconitate	cis-Aconitic acid	HMDB0000072	643757	C00417	C(/C(=C/C(=
18	Adenine	Adenine	HMDB0000034	190	C00147	C1=NC2=C(N
19	Adenosine	Adenosine	HMDB0000050	60961	C00212	C1=NC2=C(C
20	ADP	ADP	HMDB0001341	6022	C00008	C1=NC2=C(C
21	ADP-D-Glucose	ADP-glucose	HMDB0006557	16500	C00498	C1=NC2=C(C
22	Alanine	L-Alanine	HMDB0000161	5950	C00041	C[C@@H](C(=
23	Aminoadipic acid	Aminoadipic acid	HMDB0000510	469	C00956	C(CC(C(=O)O
24	AMP	Adenosine monophosphate	HMDB0000045	6083	C00020	C1=NC2=C(C
25	Androstenedione	Androstenedione	HMDB0000053	6128	C00280	C[C@]12CCC(=
26	Arginine	L-Arginine	HMDB0000517	6322	C00062	C(C[C@@H](C
27	Arginine-Alanine	Arginyl-Alanine	HMDB0028702	7020333		CC(NC(=O)C(
28	Arginine-Glutamine	Arginyl-Glutamine	HMDB0028707	9857557		NC(CCCNC(N
29	Arginine-Valine	Arginyl-Valine	HMDB0028722	6992654		CC(C)C(NC(=
30	Ascorbic acid	Ascorbic acid	HMDB0000044	54670067	C01041	C([C@@H]([C@
31	Asparagine	L-Asparagine	HMDB0000168	6267	C00152	C([C@@H](C(=
32	Aspartate	L-Aspartic acid	HMDB0000191	5960	C00049	C([C@@H](C(=
33	Betaine	Betaine	HMDB0000043	247	C00719	C[N+](C)(C)C
34	Carbamoyl phosphate	Carbamoyl phosphate	HMDB0001096	278	C00169	C(=O)(N)OP(=
35	Carnitine	L-Carnitine	HMDB0000062	2724480	C00318	C[N+](C)(C)C
36	Carnosine	Carnosine	HMDB0000033	439224	C00386	C1=C(NC=N1
37	CDP-Choline	Citicoline	HMDB0001413	13804	C00307	C[N+](C)(C)C
38	CDP-Choline (+HAc)	NA	NA	NA	NA	NA
39	Cellobiose	Cellobiose	HMDB0000055	10712	C06422	C([C@@H]1[C@
40	Cholesteryl sulfate	Cholesterol sulfate	HMDB0000653	65076	C18043	C[C@H](CCCC
41	Citramalic acid	Citramalic acid	HMDB0000426	441696	C00815	CC(CC(=O)O)
42	Citrate	Citric acid	HMDB0000094	311	C00158	C(C(=O)O)C(C
43	Citrulline	Citrulline	HMDB0000904	9750	C00327	C(C[C@@H](C
44	CMP	Cytidine monophosphate	HMDB0000095	8117	C00055	C1=CN(C(=O)
45	Creatine	Creatine	HMDB0000064	586	C00300	CN(CC(=O)O)
46	Creatinine	Creatinine	HMDB0000562	588	C00791	CN1CC(=O)N=
47	Cystathionine	L-Cystathionine	HMDB0000099	439258	C02291	C(CSC[C@@H]
48	Cystine	L-Cystine	HMDB0000192	67678	C00491	C([C@@H](C(=
49	Cytidine	Cytidine	HMDB0000089	6253	C00475	C1=CN(C(=O)
50	Cytosine	Cytosine	HMDB0000630	597	C00380	C1=C(NC(=O)
51	D-2-Aminobutyric acid	D-Alpha-aminobutyric acid	HMDB0000650	439691	C02261	CC[C@H](C(=O
52	Deoxycytidine	Deoxycytidine	HMDB0000014	13711	C00881	C1[C@@H]([C@
53	Deoxyguanosine	Deoxyguanosine	HMDB0000085	187790	C00330	C1[C@@H]([C@
54	Dihydroorotate	4,5-Dihydroorotic acid	HMDB0000528	648	C00337	C1C(NC(=O)N
55	Dihydroxyacetone phosphate	Dihydroxyacetone phosphate	HMDB0001473	668	C00111	C(C(=O)COP(
56	Dihydroxybenzeneacetic acid	3,4-Dihydroxybenzeneacetic acid	HMDB0001336	547	C01161	C1=CC(=C(C=
57	Fructose	D-Fructose	HMDB0000660	439709	C02336	C([C@@H]1[C@
58	Fructose-6-phosphate	Fructose 6-phosphate	HMDB0000124	69507	C00085	C([C@H]([C@H
59	gamma-amino butyrate	NA	NA	NA	NA	NA
60	Glucose	D-Glucose	HMDB0000122	5793	C00221	C([C@@H]1[C@
61	Glucose-6-phosphate	Glucose 6-phosphate	HMDB0001401	5958	C00092	C([C@@H]1[C@
62	Glucuronic acid	D-Glucuronic acid	HMDB0000127	94715	C00191	[C@@H]1([C@@
63	Glutamate	L-Glutamic acid	HMDB0000148	33032	C00025	C(CC(=O)O)[C
64	Glutamine	L-Glutamine	HMDB0000641	5961	C00064	C(CC(=O)N)[C
65	Glutaryl-carnitine	NA	NA	NA	NA	NA
66	Glutathione	Glutathione	HMDB0062697	745	C00051	NC(CCC(O)=N
67	Glutathione disulfide	Oxidized glutathione	HMDB0003337	975	C00127	C(CC(=O)NC(
68	Glycerate	Glyceric acid	HMDB0000139	439194	C00258	C([C@H](C(=C
69	Glycerol-3-phosphate	Glycerol 3-phosphate	HMDB0000126	439162	C00093	C([C@H](COP(
70	Glycerophosphocholine	Glycerophosphocholine	HMDB0000086	71920	C00670	C[N+](C)(C)C
71	Glycerophosphocholine (+HAc)	NA	NA	NA	NA	NA
72	Glycine	Glycine	HMDB0000123	750	C00037	C(C(=O)O)N
73	GMP	Guanosine monophosphate	HMDB0001397	6804	C00144	C1=NC2=C(N
74	Guanine	Guanine	HMDB0000132	764	C00242	C1=NC2=C(N
75	Guanosine	Guanosine	HMDB0000133	6802	C00387	C1=NC2=C(N
76	Hippuric acid	Hippuric acid	HMDB0000714	464	C01586	C1=CC=C(C=
77	Histamine	Histamine	HMDB0000870	774	C00388	C1=C(NC=N1
78	Histidine	L-Histidine	HMDB0000177	6274	C00135	C1=C(NC=N1
79	Homoarginine	Homo-L-arginine	HMDB0000670	9085	C01924	C(CCN=C(N)N
80	Homocitrulline	Homocitrulline	HMDB0000679	65072	C02427	C(CCNC(=O)N
81	Homocysteine	Homocysteine	HMDB0000742	778	C00155	C(CS)C(C(=O)
82	Hydroxyoctanoic acid	Hydroxyoctanoic acid	HMDB0000711	94180		CCCCCCC(C(=
83	Hydroxyphenyl pyruvate	4-Hydroxyphenylpyruvic acid	HMDB0000707	979	C01179	C1=CC(=CC=
84	Hydroxyproline	4-Hydroxyproline	HMDB0000725	5810	C01157	C1[C@H](CN[C
85	Hypotaurine	Hypotaurine	HMDB0000965	107812	C00519	C(CS(=O)O)N
86	Hypoxanthine	Hypoxanthine	HMDB0000157	790	C00262	C1=NC2=C(N
87	Indole	Indole	HMDB0000738	798	C00463	C1=CC=C2C(=
88	Inosine	Inosine	HMDB0000195	6021	C00294	C1=NC(=O)C2
89	Inositol	myo-Inositol	HMDB0000211		C00137	O[C@H]1[C@H
90	Isocitrate	Isocitric acid	HMDB0000193	1198	C00311	C(C(C(C(=O)O
91	Isoleucine	L-Isoleucine	HMDB0000172	6306	C00407	CC[C@H](C)[C
92	Itaconic acid	Itaconic acid	HMDB0002092	811	C00490	C=C(CC(=O)O
93	L-Octanoylcarnitine	L-Octanoylcarnitine	HMDB0000791	11953814	C02838	CCCCCCCC(=
94	L-Palmitoylcarnitine	L-Palmitoylcarnitine	HMDB0000222	11953816	C02990	CCCCCCCCCC
95	Lactate	L-Lactic acid	HMDB0000190	61503	C00186	C[C@@H](C(=
96	Leucic acid	D-Leucic acid	HMDB0000624	439960	C03264	CC(C)C[C@H](

97	Leucine	L-Leucine	HMDB0000687	6106	C00123	CC(C)C[C@@H
98	Levulinic Acid	Levulinic acid	HMDB0000720	11579		CC(=O)CCC(=
99	Lysine	L-Lysine	HMDB0000182	5962	C00047	C(CCN)C[C@@
100	Lysine-Glutamine	Lysyl-Glutamine	HMDB0028949	196305		NCCCCC(N)C
101	Malate	L-Malic acid	HMDB0000156	222656	C00149	C([C@@H](C(=
102	Mandelic acid	Mandelic acid	HMDB0000703	439616	C01984	C1=CC=C(C=
103	Mannose-6-phosphate	Mannose 6-phosphate	HMDB0001078	439198	C00275	C([C@@H]1[C@
104	Methionine	L-Methionine	HMDB0000696	6137	C00073	CSCC[C@@H](
105	Methionine sulfoxide	Methionine sulfoxide	HMDB0002005	847	C02989	CS(=O)CCC(C
106	Methyloxovaleric acid (Ketoleucine)	NA	NA	NA	NA	NA
107	Methylphenyllactate	3-Methylphenylacetic acid	HMDB0002222	12121		CC1=CC(=CC
108	N N N-Trimethyllysine	N6,N6,N6-Trimethyl-L-lysine	HMDB0001325	440120	C03793	C[N+](C)(C)C
109	N-acetyl-glutamate	N-Acetylglutamic acid	HMDB0001138	185	C00624	CC(=O)NC(CC
110	N-acetyl-glutamine	N-Acetylglutamine	HMDB0006029	25561		CC(=O)NC(CC
111	N-Acetyl-L-alanine	N-Acetyl-L-alanine	HMDB0000766	88064		C[C@@H](C(=
112	N-acetyl-L-ornithine	N-Acetylornithine	HMDB0003357	439232	C00437	CC(=O)N[C@@
113	NAD+	NAD	HMDB0000902	5893	C00003	C1=CC(=C[N-
114	NAD+ (SIM)	NA	NA	NA	NA	NA
115	NADH (SIM)	NA	NA	NA	NA	NA
116	NADP+ (SIM)	NA	NA	NA	NA	NA
117	NADPH (SIM)	NA	NA	NA	NA	NA
118	NegIS_Alanine-2H4-15N	NA	NA	NA	NA	NA
119	NegIS_Glucose-2H7-13C6	NA	NA	NA	NA	NA
120	NegIS_Glutamine-2H5-15N2	NA	NA	NA	NA	NA
121	NegIS_Glycine-13C2_15N1	NA	NA	NA	NA	NA
122	NegIS_Inosine-15N4	NA	NA	NA	NA	NA
123	NegIS_Lysine-2H8	NA	NA	NA	NA	NA
124	NegIS_Malate-2H3	NA	NA	NA	NA	NA
125	NegIS_Serine-13C3-15N1	NA	NA	NA	NA	NA
126	NegIS_Thymine-2H4	NA	NA	NA	NA	NA
127	NG-dimethyl-L-arginine	Asymmetric dimethylarginine	HMDB0001539	123831	C03626	CN(C)C(=NCC
128	Nicotinamide riboside	Nicotinamide riboside	HMDB0000855	439924	C03150	C1=CC(=C[N-
129	Nicotinate	Nicotinic acid	HMDB0001488	938	C00253	C1=CC(=CN=
130	o-acetyl-L-serine	O-Acetylserine	HMDB0003011	99478	C00979	CC(=O)OC[C@
131	O-Acylcarnitine	O-Acylcarnitine		5355	C02301	
132	O-Butanoylcarnitine	Butyrylcarnitine	HMDB0002013	439829	C02862	CCCC(=O)OC
133	O-Decanoyl-L-carnitine	O-decanoyl-L-carnitine	HMDB0062631	11953821	C03299	[H][C@@](CC([
134	O-Phosphorylethanolamine	O-Phosphoethanolamine	HMDB0000224	1015	C00346	C(COP(=O)(O
135	O-Propanoylcarnitine	Propionylcarnitine	HMDB0000824	107738	C03017	CCC(=O)OC(C
136	Ornithine	Ornithine	HMDB0000214	6262	C00077	C(C[C@@H](C(
137	Pantothenate	Pantothenic acid	HMDB0000210	6613	C00864	CC(C)(CO)C(C
138	Phenylalanine	L-Phenylalanine	HMDB0000159	6140	C00079	C1=CC=C(C=
139	Phenylpropanolamine	Phenylpropanolamine	HMDB0001942	26934	C02343	C[C@H]([C@H]
140	Phosphocholine	Phosphorylcholine	HMDB0001565	8691	C00588	C[N+](C)(C)C
141	Phosphocholine (+HAc)	NA	NA	NA	NA	NA
142	PosIS_Alanine-2H4-15N	NA	NA	NA	NA	NA
143	PosIS_Glutamine-2H5-15N2	NA	NA	NA	NA	NA
144	PosIS_Inosine-15N4	NA	NA	NA	NA	NA
145	PosIS_Lysine-2H8	NA	NA	NA	NA	NA
146	PosIS_Serine-13C3-15N1	NA	NA	NA	NA	NA
147	Proline	L-Proline	HMDB0000162	145742	C00148	C1C[C@H](NC
148	Purine	Purine	HMDB0001366	1044	C15587	C1=C2C(=NC=
149	Pyroglutamate	Pyroglutamic acid	HMDB0000267	7405	C01879	C1CC(=O)N[C
150	Pyrroline-5-carboxylic acid	NA	NA	NA	NA	NA
151	Pyruvate	Pyruvic acid	HMDB0000243	1060	C00022	CC(=O)C(=O)
152	Raffinose	Raffinose	HMDB0003213	10542	C00492	C([C@@H]1[C@
153	Ribitol	Ribitol	HMDB0000508		C00474	OC[C@H](O)[C
154	Ribose phosphate	Ribose 1-phosphate	HMDB0001489	439236	C00620	C([C@@H]1[C@
155	S-adenosyl-L-methionine	S-Adenosylmethionine	HMDB0001185	16757548	C00019	C[S+](CC[C@H
156	Serine	Serine	HMDB0062263	5951	C00716	N[C@@H](CO)
157	Sorbitol	Sorbitol	HMDB0000247	5780	C00794	C([C@H]([C@H
158	Spermine	Spermine	HMDB0001256	1103	C00750	C(CCNCCCN)
159	Stearic acid	Stearic acid	HMDB0000827	5281	C01530	CCCCCCCCC
160	Succinate	Succinic acid	HMDB0000254	1110	C00042	C(CC(=O)O)C
161	Sucrose	Sucrose	HMDB0000258	5988	C00089	C([C@@H]1[C@
162	Taurine	Taurine	HMDB0000251	1123	C00245	C(CS(=O)(=O
163	Thiamine	Thiamine	HMDB0000235	1130	C00378	CC1=C(SC=[N
164	Threonic acid	Threonic acid	HMDB0000943	151152	C01620	C([C@H]([C@@
165	Threonine	L-Threonine	HMDB0000167	6288	C00188	C[C@H]([C@@H
166	Thymidine	Thymidine	HMDB0000273	5789	C00214	CC1=CN(C(=O
167	Tridecanoic acid	Tridecanoic acid	HMDB0000910	656741	C17076	CCCCCCCCCC
168	Trigonelline	Trigonelline	HMDB0000875	5570	C01004	C[N+]1=CC=C
169	Tryptamine	Tryptamine	HMDB0000303	1150	C00398	C1=CC=C2C(=
170	Tryptophan	L-Tryptophan	HMDB0000929	6305	C00078	C1=CC=C2C(=
171	Tyramine	Tyramine	HMDB0000306	5610	C00483	C1=CC(=CC=
172	Tyrosine	L-Tyrosine	HMDB0000158	6057	C00082	C1=CC(=CC=
173	UDP-D-Glucose	Uridine diphosphate glucose	HMDB0000286	53477679	C00029	C1=CN(C(=O)
174	UDP-N-acetyl-glucosamine	Uridine diphosphate-N-acetylglucosamine	HMDB0000290	9547196	C00043	CC(=O)N[C@@
175	UMP	Uridine 5'-monophosphate	HMDB0000288	6030	C00105	C1=CN(C(=O)
176	Uracil	Uracil	HMDB0000300	1174	C00106	C1=CNC(=O)
177	Uric acid	Uric acid	HMDB0000289	1175	C00366	C12=C(NC(=O
178	Uridine	Uridine	HMDB0000296	6029	C00299	C1=CN(C(=O)
179	Valine	L-Valine	HMDB0000883	6287	C00183	CC(C)[C@@H](
180	Xanthine	Xanthine	HMDB0000292	1188	C00385	C1=NC2=C(N

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 **cel** (KEGG organisms abbreviation).

4.2 Pathway Enrichment Analysis

Pathway enrichment analysis usually refers to quantitative enrichment analysis directly using the compound concentration values, as compared to compound lists used by over-representation analysis. As a result, it is more sensitive and has the potential to identify **subtle but consistent** changes amongst compounds involved in the same biological pathway.

Many procedures have been developed in the last decade for quantitative enrichment analysis, the most famous being the Gene Set Enrichment Analysis. Many new and improved methods have been implemented since. The enrichment analysis is based on GlobalTest and GlobalAncova. Both methods support enrichment analysis with binary, multi-group, as well as continuous phenotypes. The p-values can be approximated based on the asymptotic distribution without using permutations which is computationally very intensive and is not suitable for web applications. Please note, when sample sizes are small, the approximated p values may be slightly less accurate compared to p values obtained by using a permutation-based method (for details, please refer to the paper by Goeman, J.J. et al. [1] and by

[1] Jelle J. Goeman and Peter Buhlmann. *Analyzing gene expression data in terms of gene sets: methodological issues*, Bioinformatics 2007 23(8):980-987

Hummel, M. et al. [2]) However, since our focus is to identify the most relevant pathways within the pathways in the library, we are more interested in the rank of the pathway, not its absolute p-value. Therefore, this disadvantage may be tolerated.

The selected pathway enrichment analysis method is **Globaltest**.

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. [3]

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.

[2]Manuela Hummel, Reinhard Meister and Ulrich Mansmann. *GlobalANCOVA: exploration and assessment of gene group effects*, Bioinformatics 2008 24(1):78-85

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

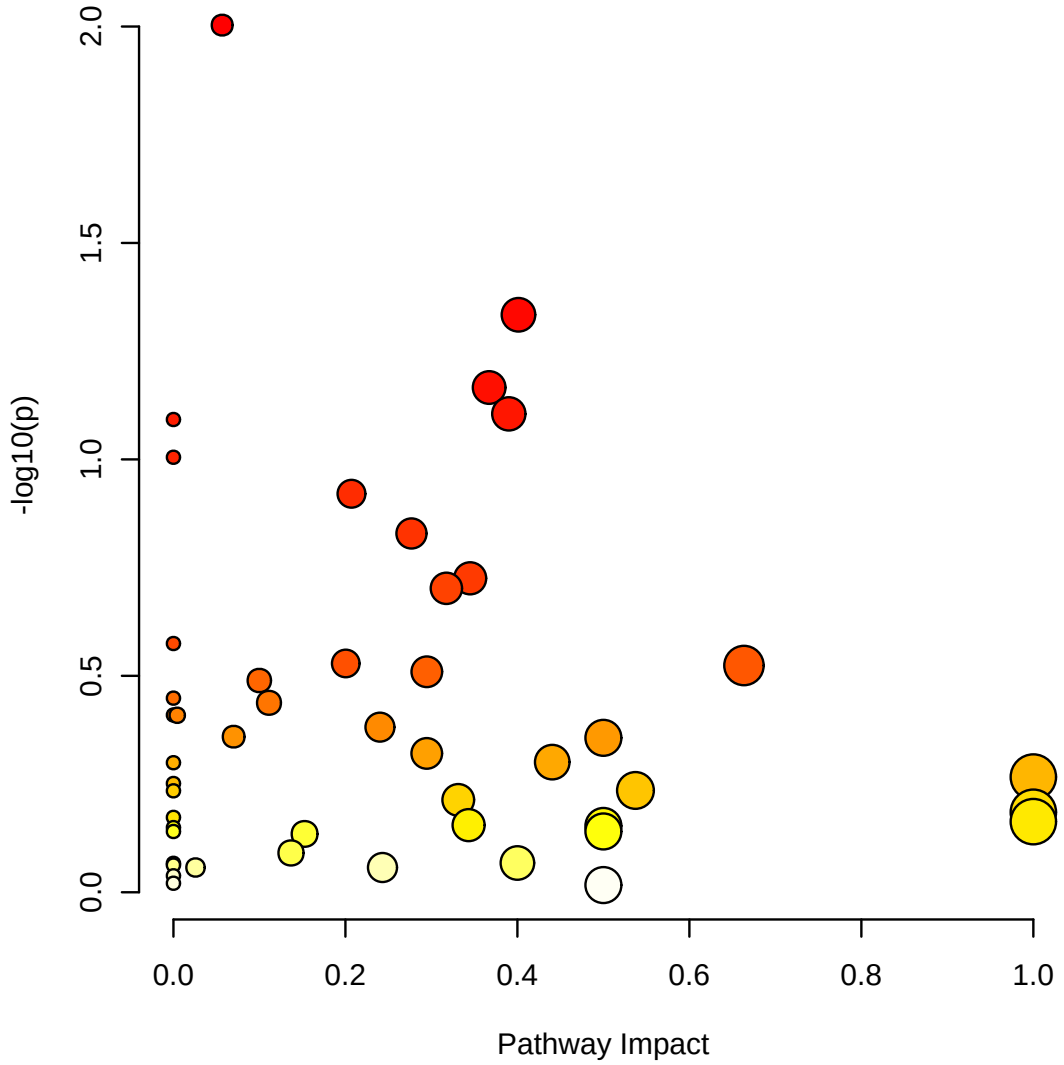

Figure 1: Summary of Pathway Analysis

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 Cmpd	Hits	Raw p	-log10(p)	Holm adjust	FDR	Impact
Galactose metabolism	15	2	9.93E-03	2.00E+00	4.76E-01	4.76E-01	0.06
Glyoxylate and dicarboxylate metabolism	31	9	4.63E-02	1.33E+00	1.00E+00	7.76E-01	0.40
Citrate cycle (TCA cycle)	20	7	6.82E-02	1.17E+00	1.00E+00	7.76E-01	0.37
Starch and sucrose metabolism	14	4	7.85E-02	1.11E+00	1.00E+00	7.76E-01	0.39
Fatty acid degradation	38	1	8.09E-02	1.09E+00	1.00E+00	7.76E-01	0.00
Pentose phosphate pathway	21	2	9.89E-02	1.00E+00	1.00E+00	7.91E-01	0.00
Fructose and mannose metabolism	16	3	1.20E-01	9.21E-01	1.00E+00	8.23E-01	0.21
Glycine, serine and threonine metabolism	28	5	1.48E-01	8.29E-01	1.00E+00	8.90E-01	0.28
Pyruvate metabolism	19	3	1.88E-01	7.25E-01	1.00E+00	9.02E-01	0.35
Cysteine and methionine metabolism	34	7	1.99E-01	7.02E-01	1.00E+00	9.02E-01	0.32
Valine, leucine and isoleucine biosynthesis	8	4	2.66E-01	5.75E-01	1.00E+00	9.02E-01	0.00
Purine metabolism	60	14	2.96E-01	5.29E-01	1.00E+00	9.02E-01	0.20
Alanine, aspartate and glutamate metabolism	20	9	2.99E-01	5.24E-01	1.00E+00	9.02E-01	0.66
Glycerolipid metabolism	13	3	3.09E-01	5.09E-01	1.00E+00	9.02E-01	0.29
Glycolysis / Gluconeogenesis	25	4	3.24E-01	4.89E-01	1.00E+00	9.02E-01	0.10
Aminoacyl-tRNA biosynthesis	45	18	3.56E-01	4.48E-01	1.00E+00	9.02E-01	0.00
Pentose and glucuronate interconversions	15	2	3.65E-01	4.38E-01	1.00E+00	9.02E-01	0.11
Porphyrin and chlorophyll metabolism	20	1	3.89E-01	4.10E-01	1.00E+00	9.02E-01	0.00
Sulfur metabolism	11	1	3.90E-01	4.09E-01	1.00E+00	9.02E-01	0.00
Amino sugar and nucleotide sugar metabolism	29	5	4.15E-01	3.81E-01	1.00E+00	9.02E-01	0.24
Phosphatidylinositol signaling system	26	1	4.37E-01	3.59E-01	1.00E+00	9.02E-01	0.07
Ascorbate and aldarate metabolism	6	3	4.40E-01	3.57E-01	1.00E+00	9.02E-01	0.50
Nicotinate and nicotinamide metabolism	14	4	4.78E-01	3.21E-01	1.00E+00	9.02E-01	0.29
Arginine and proline metabolism	27	7	5.00E-01	3.01E-01	1.00E+00	9.02E-01	0.44
Butanoate metabolism	12	3	5.02E-01	2.99E-01	1.00E+00	9.02E-01	0.00
Ubiquinone and other terpenoid-quinone biosynthesis	5	2	5.42E-01	2.66E-01	1.00E+00	9.02E-01	1.00
Phosphonate and phosphinate metabolism	5	1	5.61E-01	2.51E-01	1.00E+00	9.02E-01	0.00
Glutathione metabolism	25	6	5.82E-01	2.35E-01	1.00E+00	9.02E-01	0.54
beta-Alanine metabolism	13	2	5.83E-01	2.34E-01	1.00E+00	9.02E-01	0.00
Tyrosine metabolism	21	4	6.12E-01	2.13E-01	1.00E+00	9.02E-01	0.33
D-Glutamine and D-glutamate metabolism	5	3	6.53E-01	1.85E-01	1.00E+00	9.02E-01	1.00
Lysine degradation	21	2	6.71E-01	1.73E-01	1.00E+00	9.02E-01	0.00
Phenylalanine, tyrosine and tryptophan biosynthesis	4	3	6.87E-01	1.63E-01	1.00E+00	9.02E-01	1.00
Pyrimidine metabolism	40	10	7.00E-01	1.55E-01	1.00E+00	9.02E-01	0.34
Arginine biosynthesis	6	5	7.03E-01	1.53E-01	1.00E+00	9.02E-01	0.50
Valine, leucine and isoleucine degradation	40	3	7.09E-01	1.49E-01	1.00E+00	9.02E-01	0.00
Phenylalanine metabolism	6	2	7.24E-01	1.41E-01	1.00E+00	9.02E-01	0.50
Nitrogen metabolism	6	2	7.24E-01	1.40E-01	1.00E+00	9.02E-01	0.00
Tryptophan metabolism	30	3	7.33E-01	1.35E-01	1.00E+00	9.02E-01	0.15
Inositol phosphate metabolism	26	4	8.11E-01	9.09E-02	1.00E+00	9.35E-01	0.14
Thiamine metabolism	7	1	8.56E-01	6.76E-02	1.00E+00	9.35E-01	0.40
Pantothenate and CoA biosynthesis	18	3	8.57E-01	6.69E-02	1.00E+00	9.35E-01	0.00
Selenocompound metabolism	18	1	8.65E-01	6.28E-02	1.00E+00	9.35E-01	0.00
Sphingolipid metabolism	17	1	8.77E-01	5.72E-02	1.00E+00	9.35E-01	0.03
Glycerophospholipid metabolism	32	6	8.77E-01	5.72E-02	1.00E+00	9.35E-01	0.24
Histidine metabolism	8	1	9.16E-01	3.81E-02	1.00E+00	9.56E-01	0.00
Propanoate metabolism	23	1	9.53E-01	2.09E-02	1.00E+00	9.62E-01	0.00
Taurine and hypotaurine metabolism	6	2	9.62E-01	1.66E-02	1.00E+00	9.62E-01	0.50

6 Appendix: R Command History

```
 [1] "mSet<-InitDataObjects(\"conc\", \"pathqea\", FALSE)"
 [2] "mSet<-Read.TextData(mSet, \"Replacing_with_your_file_path\", \"rowu\", \"disc\");"
 [3] "mSet<-SanityCheckData(mSet)"
 [4] "mSet<-ReplaceMin(mSet);"
 [5] "mSet<-CrossReferencing(mSet, \"name\");"
 [6] "mSet<-CreateMappingResultTable(mSet)"
 [7] "mSet<-PerformDetailMatch(mSet, \"1 3-Dimethyluric acid\");"
 [8] "mSet<-GetCandidateList(mSet);"
 [9] "mSet<-SetCandidate(mSet, \"1 3-Dimethyluric acid\", \"1,3-Dimethyluric acid\");"
[10] "mSet<-PerformDetailMatch(mSet, \"5- Methylthioadenosine\");"
[11] "mSet<-GetCandidateList(mSet);"
[12] "mSet<-SetCandidate(mSet, \"5- Methylthioadenosine\", \"5'-Methylthioadenosine\");"
[13] "mSet<-PerformDetailMatch(mSet, \"Acetyl glycine\");"
[14] "mSet<-GetCandidateList(mSet);"
[15] "mSet<-SetCandidate(mSet, \"Acetyl glycine\", \"Phenylacetylglycine\");"
[16] "mSet<-PerformDetailMatch(mSet, \"Acetyl proline\");"
[17] "mSet<-GetCandidateList(mSet);"
[18] "mSet<-PerformDetailMatch(mSet, \"Acetyl-arginine\");"
[19] "mSet<-GetCandidateList(mSet);"
[20] "mSet<-SetCandidate(mSet, \"Acetyl-arginine\", \"N-a-Acetyl-L-arginine\");"
[21] "mSet<-PerformDetailMatch(mSet, \"Arginine-Alanine\");"
[22] "mSet<-GetCandidateList(mSet);"
[23] "mSet<-PerformDetailMatch(mSet, \"Arginine-Glutamine\");"
[24] "mSet<-GetCandidateList(mSet);"
[25] "mSet<-SetCandidate(mSet, \"Arginine-Glutamine\", \"Arginyl-Glutamine\");"
[26] "mSet<-PerformDetailMatch(mSet, \"Arginine-Alanine\");"
[27] "mSet<-GetCandidateList(mSet);"
[28] "mSet<-SetCandidate(mSet, \"Arginine-Alanine\", \"Arginyl-Alanine\");"
[29] "mSet<-PerformDetailMatch(mSet, \"Arginine-Valine\");"
[30] "mSet<-GetCandidateList(mSet);"
[31] "mSet<-SetCandidate(mSet, \"Arginine-Valine\", \"Arginyl-Valine\");"
[32] "mSet<-PerformDetailMatch(mSet, \"CDP-Choline (+HAc)\");"
[33] "mSet<-GetCandidateList(mSet);"
[34] "mSet<-PerformDetailMatch(mSet, \"Dihydroxybenzeneacetic acid\");"
[35] "mSet<-GetCandidateList(mSet);"
[36] "mSet<-SetCandidate(mSet, \"Dihydroxybenzeneacetic acid\", \"3,4-Dihydroxybenzeneacetic acid\"
[37] "mSet<-PerformDetailMatch(mSet, \"gamma-amino butyrate\");"
[38] "mSet<-GetCandidateList(mSet);"
[39] "mSet<-SetCandidate(mSet, \"gamma-amino butyrate\", \"Gamma-Aminobutyric acid\");"
[40] "mSet<-SetCandidate(mSet, \"gamma-amino butyrate\", \"Gamma-Aminobutyric acid\");"
[41] "mSet<-PerformDetailMatch(mSet, \"Glutaryl-carnitine\");"
[42] "mSet<-GetCandidateList(mSet);"
[43] "mSet<-PerformDetailMatch(mSet, \"Glycerophosphocholine (+HAc)\");"
[44] "mSet<-GetCandidateList(mSet);"
[45] "mSet<-PerformDetailMatch(mSet, \"Hydroxyphenyl pyruvate\");"
[46] "mSet<-GetCandidateList(mSet);"
[47] "mSet<-SetCandidate(mSet, \"Hydroxyphenyl pyruvate\", \"4-Hydroxyphenylpyruvic acid\");"
[48] "mSet<-PerformDetailMatch(mSet, \"Lysine-Glutamine\");"
[49] "mSet<-GetCandidateList(mSet);"
[50] "mSet<-SetCandidate(mSet, \"Lysine-Glutamine\", \"Lysyl-Glutamine\");"
[51] "mSet<-PerformDetailMatch(mSet, \"Methyloxovaleric acid (Ketoleucine)\");"
[52] "mSet<-GetCandidateList(mSet);"
[53] "mSet<-PerformDetailMatch(mSet, \"Methylphenyllactate\");"
[54] "mSet<-GetCandidateList(mSet);"
[55] "mSet<-SetCandidate(mSet, \"Methylphenyllactate\", \"3-Methylphenylacetic acid\");"
[56] "mSet<-PerformDetailMatch(mSet, \"N N N-Trimethyllysine\");"
```

```
 [57] "mSet<-GetCandidateList(mSet);"
 [58] "mSet<-SetCandidate(mSet, \"N N N-Trimethyllysine\", \"N6,N6,N6-Trimethyl-L-lysine\");"
 [59] "mSet<-PerformDetailMatch(mSet, \"N-acetyl-glutamine\");"
 [60] "mSet<-GetCandidateList(mSet);"
 [61] "mSet<-SetCandidate(mSet, \"N-acetyl-glutamine\", \"N-Acetylglutamine\");"
 [62] "mSet<-PerformDetailMatch(mSet, \"N-acetyl-L-ornithine\");"
 [63] "mSet<-GetCandidateList(mSet);"
 [64] "mSet<-SetCandidate(mSet, \"N-acetyl-L-ornithine\", \"N-Acetylornithine\");"
 [65] "mSet<-PerformDetailMatch(mSet, \"NG-dimethyl-L-arginine\");"
 [66] "mSet<-GetCandidateList(mSet);"
 [67] "mSet<-SetCandidate(mSet, \"NG-dimethyl-L-arginine\", \"Asymmetric dimethylarginine\");"
 [68] "mSet<-PerformDetailMatch(mSet, \"Phosphocholine (+HAc)\");"
 [69] "mSet<-GetCandidateList(mSet);"
 [70] "mSet<-PerformDetailMatch(mSet, \"Pyrroline-5-carboxylic acid\");"
 [71] "mSet<-GetCandidateList(mSet);"
 [72] "mSet<-SetCandidate(mSet, \"Pyrroline-5-carboxylic acid\", \"1-Pyrroline-5-carboxylic acid\");
 [73] "mSet<-PerformDetailMatch(mSet, \"Ribose phosphate\");"
 [74] "mSet<-GetCandidateList(mSet);"
 [75] "mSet<-SetCandidate(mSet, \"Ribose phosphate\", \"Ribose 1-phosphate\");"
 [76] "mSet<-PreparePrenormData(mSet)"
 [77] "mSet<-Normalization(mSet, \"NULL\", \"NULL\", \"NULL\", ratio=FALSE, ratioNum=20)"
 [78] "mSet<-PlotNormSummary(mSet, \"norm_1_\", \"png\", 72, width=NA)"
 [79] "mSet<-PlotSampleNormSummary(mSet, \"snorm_1_\", \"png\", 72, width=NA)"
 [80] "mSet<-SetKEGG.PathLib(mSet, \"cel\", \"current\")"
 [81] "mSet<-SetMetabolomeFilter(mSet, F);"
 [82] "mSet<-CalculateQeaScore(mSet, \"rbc\", \"gt\")"
 [83] "mSet<-PlotPathSummary(mSet, F, \"path_view_1_\", \"png\", 72, width=NA)"
 [84] "mSet<-PlotKEGGPath(mSet, \"Starch and sucrose metabolism\",576, 480, \"png\", NULL)"
 [85] "mSet<-RerenderMetPAGraph(mSet, \"zoom1634927200216.png\",576.0, 480.0, 100.0)"
 [86] "mSet<-PlotKEGGPath(mSet, \"Starch and sucrose metabolism\",576, 480, \"png\", NULL)"
 [87] "mSet<-PlotKEGGPath(mSet, \"Alanine, aspartate and glutamate metabolism\",576, 480, \"png\", NU
 [88] "mSet<-PlotKEGGPath(mSet, \"Glyoxylate and dicarboxylate metabolism\",576, 480, \"png\", NULL)
 [89] "mSet<-PlotKEGGPath(mSet, \"Pyruvate metabolism\",576, 480, \"png\", NULL)"
 [90] "mSet<-PlotKEGGPath(mSet, \"Citrate cycle (TCA cycle)\",576, 480, \"png\", NULL)"
 [91] "mSet<-PlotKEGGPath(mSet, \"Glycine, serine and threonine metabolism\",576, 480, \"png\", NULL
 [92] "mSet<-PlotKEGGPath(mSet, \"Galactose metabolism\",576, 480, \"png\", NULL)"
 [93] "mSet<-PlotKEGGPath(mSet, \"Fructose and mannose metabolism\",576, 480, \"png\", NULL)"
 [94] "mSet<-PlotKEGGPath(mSet, \"Glyoxylate and dicarboxylate metabolism\",576, 480, \"png\", NULL)
 [95] "mSet<-PlotKEGGPath(mSet, \"Citrate cycle (TCA cycle)\",576, 480, \"png\", NULL)"
 [96] "mSet<-PlotKEGGPath(mSet, \"Starch and sucrose metabolism\",576, 480, \"png\", NULL)"
 [97] "mSet<-PlotKEGGPath(mSet, \"Galactose metabolism\",576, 480, \"png\", NULL)"
 [98] "mSet<-PlotKEGGPath(mSet, \"Glyoxylate and dicarboxylate metabolism\",576, 480, \"png\", NULL)
 [99] "mSet<-PlotKEGGPath(mSet, \"Glyoxylate and dicarboxylate metabolism\",576, 480, \"png\", NULL)
[100] "mSet<-PlotKEGGPath(mSet, \"Glyoxylate and dicarboxylate metabolism\",576, 480, \"png\", NULL)
[101] "mSet<-PlotKEGGPath(mSet, \"Glyoxylate and dicarboxylate metabolism\",576, 480, \"png\", NULL)
[102] "mSet<-PlotKEGGPath(mSet, \"Galactose metabolism\",576, 480, \"png\", NULL)"
[103] "mSet<-PlotKEGGPath(mSet, \"Glyoxylate and dicarboxylate metabolism\",576, 480, \"png\", NULL)
[104] "mSet<-PlotKEGGPath(mSet, \"Galactose metabolism\",576, 480, \"png\", NULL)"
[105] "mSet<-PlotKEGGPath(mSet, \"Glyoxylate and dicarboxylate metabolism\",576, 480, \"png\", NULL)
[106] "mSet<-SaveTransformedData(mSet)"
[107] "mSet<-PreparePDFReport(mSet, \"guest6793469145423477714\")\n"
```

The report was generated on Fri Oct 22 14:47:51 2021 with R version 4.1.1 (2021-08-10).