Metabolomic Data Analysis with MetaboAnalyst 5.0

Name: guest325551242525186926

December 7, 2022

1 Background

MSEA or Metabolite Set Enrichment Analysis is a way to identify biologically meaningful patterns that are significantly enriched in quantitative metabolomic data. In conventional approaches, metabolites are evaluated individually for their significance under conditions of study. Those compounds that have passed certain significance level are then combined to see if any meaningful patterns can be discerned. In contrast, MSEA directly investigates if a set of functionally related metabolites without the need to preselect compounds based on some arbitrary cut-off threshold. It has the potential to identify subtle but consistent changes among a group of related compounds, which may go undetected with the conventional approaches.

Essentially, MSEA is a metabolomic version of the popular GSEA (Gene Set Enrichment Analysis) software with its own collection of metabolite set libraries as well as an implementation of user-friendly web-interfaces. GSEA is widely used in genomics data analysis and has proven to be a powerful alternative to conventional approaches. For more information, please refer to the original paper by Subramanian A, and a nice review paper by Nam D, Kim SY. [1]. [2]

2 MSEA Overview

Metabolite set enrichment analysis consists of four steps - data input, data processing, data analysis, and results download. Different analysis procedures are performed based on different input types. In addition, users can also browse and search the metabolite set libraries as well as upload their self-defined metabolite sets for enrichment analysis. Users can also perform metabolite name mapping between a variety of compound names, synonyms, and major database identifiers.

3 Data Input

There are three enrichment analysis algorithms offered by MSEA. Accordingly, three different types of data inputs are required by these three approaches:

- A list of important compound names - entered as a one column data (*Over Representation Analysis (ORA)*);
- A single measured biofluid (urine, blood, CSF) sample- entered as tab separated two-column data with the first column for compound name, and the second for concentration values (*Single Sample Profiling (SSP)*);

[1] Subramanian *Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.*, Proc Natl Acad Sci USA. 2005 102(43): 15545-50

[2] Nam D, Kim SY. *Gene-set approach for expression pattern analysis*, Briefings in Bioinformatics. 2008 9(3): 189-197.

- A compound concentration table - entered as a comma separated (.csv) file with the each sample per row and each metabolite concentration per column. The first column is sample names and the second column for sample phenotype labels (*Quantitative Enrichment Analysis (QEA)*)

You selected Over Representation Analysis (ORA) which requires a list of compound names as input.

4 Data Process

The first step is to standardize the compound labels. It is an essential step since the compound labels will be subsequently compared with compounds contained in the metabolite set library. MSEA has a built-in tool to convert between compound common names, synonyms, identifiers used in HMDB ID, PubChem, ChEBI, BiGG, METLIN, KEGG, or Reactome. **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*

	Query	Match	HMDB	PubChem	KEGG	SMILES
1	52931205	SM(d16:1/24:0)		52931205		
2	79437	NA	NA	NA	NA	NA
3	27476	NA	NA	NA	NA	NA
4	96373	NA	NA	NA	NA	NA
5	69726	NA	NA	NA	NA	NA
6	95433	2-Hydroxy-2-methylbutyric acid	HMDB0001987	95433		CCC(C)(O)C(
7	11671	alpha-hydroxy-isobutyric acid	HMDB0000729	11671		CC(C)(O)C(=
8	70679121	2-Octenoylcarnitine	HMDB0013324	70679121		CCCCC\C=C
9	71464477	3-hydroxybutyrylcarnitine		71464477		CC(O)CC(=O
10	87	NA	NA	NA	NA	NA
11	500	NA	NA	NA	NA	NA
12	9378	NA	NA	NA	NA	NA
13	129817528	NA	NA	NA	NA	NA
14	439176	NA	NA	NA	NA	NA
15	11361	NA	NA	NA	NA	NA
16	469	Aminoadipic acid	HMDB0000510	469	C00956	NC(CCCC(=O
17	10467	Arachidic acid	HMDB0002212	10467	C06425	CCCCCCCCC
18	136212424	NA	NA	NA	NA	NA
19	53477833	Arachidyl carnitine	HMDB0006460	53477833		CCCCCCCCC
20	5960	NA	NA	NA	NA	NA
21	123831	NA	NA	NA	NA	NA
22	247	NA	NA	NA	NA	NA
23	2969	Capric acid	HMDB0000511	2969	C01571	CCCCCCCCC
24	71464574	O-(17-carboxyheptadecanoyl)carnitine		71464574		C(C)(C)CC(C
25	5283632	18:1 Cholesterol ester	HMDB00918	5283632	C14641	CCCCCCCC/
26	53477892	CE(20:3(8Z,11Z,14Z))	HMDB0006736	53477892	C02530	CCCCC\C=C
27	10372299	20:5 Cholesterol ester	HMDB06731	10372299		CC/C=C\C/C
28	53477890	CE(22:6(4Z,7Z,10Z,13Z,16Z,19Z))	HMDB0006733	53477890	C02530	CC\C=C/C\C
29	70698937	Cer(d16:1/20:0)		70698937		
30	5283575	Cer(d18:0/22:0)	HMDB0011765	5283575		CCCCCCCCC
31	5283577	Cer(d18:0/24:0)	HMDB0011768	5283577		CCCCCCCCC
32	5283564	Ceramide (d18:1/16:0)	HMDB0004949	5283564	C00195	CCCCCCCCC
33	5283565	Ceramide (d18:1/18:0)	HMDB0004950	5283565	C00195	CCCCCCCCC
34	5283566	Ceramide (d18:1/20:0)	HMDB0004951	5283566	C00195	CCCCCCCCC
35	52931247	FMC-5(d18:1/24:1)		52931247		CCCCCCCC/
36	5283570	Cer(d18:1/26:0)	HMDB0004955	5283570	C00195	CCCCCCCCC
37	5283569	Cer(d18:1/26:1(17Z))	HMDB04954	5283569		CCCCCCCC/
38	132282053	Cer(d18:2/18:0)		132282053		CCC/C=C\C
39	52931120	Cer(d18:2/20:0)		52931120		
40	134725975	Cer(d19:1/22:0)		134725975		
41	134730643	Cer(d20:1/24:0)		134730643		
42	5283572	Cer(d18:0/16:0)	HMDB0011760	5283572		CCCCCCCCC
43	10554446	Cer(d16:0/18:0)		10554446		
44	85366458	NA	NA	NA	NA	NA
45	156960925	NA	NA	NA	NA	NA
46	5997	Cholesterol	HMDB0000067	5997	C00187	[H][C@@]1(CC
47	5754	Cortisol	HMDB0000063	5754	C00735	[H][C@@]12CC
48	225609	Fludrocortisone acetate		225609	C08186	CC(=O)OCC(
49	52924055	NA	NA	NA	NA	NA
50	92094	Delta-Tocopherol	HMDB0002902	92094	C14151	CC(C)CCC[C
51	644078	DG(16:0/16:0/0:0)	HMDB0007098	644078	C00165	[H][C@](CO)(
52	53477956	DG(14:0/20:1(11Z)/0:0)	HMDB0007021	53477956		[H]\C(CCCC
53	3246945	DG(18:0/16:0/0:0)	HMDB0007156	3246945	C00165	[H][C@](CO)(
54	9543714	DG(16:1(9Z)/20:0/0:0)	HMDB0007136	9543714	C00165	[H][C@](CO)(
55	6441562	DG(18:0/18:2(9Z,12Z)/0:0)	HMDB0007161	6441562	C00165	[H]\C(CCCC
56	53477965	DG(14:0/22:2(13Z,16Z)/0:0)	HMDB0007030	53477965		[H][C@](CO)(
57	9543730	NA	NA	NA	NA	NA
58	53478104	DG(18:2(9Z,12Z)/18:1(11Z)/0:0)	HMDB0007246	53478104		[H][C@](CO)(

59	91666386	NA	NA	NA	NA	NA
60	131801759	DG(32:1)	HMDB0056040	131801759		[H][C@@](O)(
61	131801868	DG(34:2)	HMDB0056158	131801868		[H][C@](O)(C
62	5497165	DG(36:2)	HMDB0056197	5497165		[H]C(O)(COC
63	53478119	DG(18:3(6Z,9Z,12Z)/18:1(9Z)/0:0)	HMDB0007276	53478119		[H][C@](CO)(
64	94715	NA	NA	NA	NA	NA
65	49661773	NA	NA	NA	NA	NA
66	51090856	NA	NA	NA	NA	NA
67	33032	NA	NA	NA	NA	NA
68	10140	Glycocholic acid	HMDB0000138	10140	C01921	C[C@H](CCC(
69	3035026	glycodeoxycholic acid		3035026	C05464	C(CCC(=O)N
70	12310288	Glycoursodeoxycholic acid	HMDB0000708	12310288		[H][C@@]1(CC
71	22833540	NA	NA	NA	NA	NA
72	439918	NA	NA	NA	NA	NA
73	764	NA	NA	NA	NA	NA
74	9085	NA	NA	NA	NA	NA
75	65072	NA	NA	NA	NA	NA
76	790	NA	NA	NA	NA	NA
77	459122	NA	NA	NA	NA	NA
78	6306	2S-Amino-3S-methylpentanoic acid	HMDB0000172	6306	C00407	CC(C)(N)C(=
79	53627559	NA	NA	NA	NA	NA
80	6426901	2-Methylbutyroylcarnitine	HMDB0000378	6426901		CCC(C)C(=O
81	3845	NA	NA	NA	NA	NA
82	161166	NA	NA	NA	NA	NA
83	53477895	CE(20:2(6Z,9Z))	HMDB0006734	53477895	C01290	CCCCC\C=C
84	53481001	Lactosylceramide (d18:1/24:0)	HMDB0011595	53481001	C01290	[H][C@@](CO[
85	20057309	Lactosylceramide (d18:1/24:1(15Z))	HMDB0004872	20057309	C01290	CCCCCCCCC
86	3893	Dodecanoic acid	HMDB0000638	3893	C02679	CCCCCCCCC
87	76807	NA	NA	NA	NA	NA
88	11197	Tetracosanoic acid	HMDB0002003	11197	C08320	CCCCCCCCC
89	89566	LysoPA(16:0/0:0)	HMDB0007853	89566	C04036	CCCCCCCCC
90	9547179	PA(18:0/0:0)	HMDB07854	9547179		CCCCCCCCC
91	5497152	LysoPA(18:1(9Z)/0:0)	HMDB0007855	5497152	C00416	CCCCCCCC\
92	50990923	PA(18:2(9Z,12Z)/0:0)		50990923		CCCCC/C=C
93	131821850	LysoPA(20:3(5Z,8Z,11Z)/0:0)	HMDB0114747	131821850		CCCCCCCC\
94	131821847	LysoPA(20:4(8Z,11Z,14Z,17Z)/0:0)	HMDB0114742	131821847		CC\C=C/C\C
95	25099673	LysoPA(22:6(4Z,7Z,10Z,13Z,16Z,19Z)/0:0)	HMDB0114755	25099673		CC\C=C/C\C
96	24779491	LysoPC(0:0/18:0)	HMDB0011128	24779491		[H][C@@](CO)
97	85735491	NA	NA	NA	NA	NA
98	312531184	NA	NA	NA	NA	NA
99	497299	LysoPC(18:0)	HMDB0010384	497299	C04230	[H][C@@](O)(
100	53480467	LysoPC(20:3(5Z,8Z,11Z))	HMDB0010393	53480467	C04230	CCCCCCCC\
101	348280585	NA	NA	NA	NA	NA
102	24779476	LysoPC(20:4(5Z,8Z,11Z,14Z))	HMDB0010395	24779476	C04230	CCCCC\C=C
103	24779479	LysoPC(22:0)	HMDB0010398	24779479	C04230	[H][C@@](O)(
104	52924059	LysoPC(22:2(13Z,16Z))	HMDB0010400	52924059	C04230	[H][C@@](O)(
105	52924039	LysoPC(22:4(7Z,10Z,13Z,16Z))	HMDB0010401	52924039	C04230	[H][C@@](O)(
106	53480475	NA	NA	NA	NA	NA
107	85335863	NA	NA	NA	NA	NA
108	53480667	PE(18:0)	HMDB0011129	53480667		[H][C@@](CO)
109	53480924	LysoPE(0:0/18:1(11Z))	HMDB0011475	53480924		[H][C@@](CO)
110	53480926	LysoPE(0:0/18:2(9Z,12Z))	HMDB0011477	53480926		[H][C@@](CO)
111	53480936	LysoPE(0:0/20:4(5Z,8Z,11Z,14Z))	HMDB0011487	53480936		[H][C@@](CO)
112	53480945	LysoPE(0:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	HMDB0011496	53480945		[H][C@@](CO)
113	9547069	LysoPE(16:0/0:0)	HMDB0011503	9547069		[H][C@@](O)(
114	9547068	LysoPE(18:0/0:0)	HMDB0011130	9547068	C21484	[H][C@@](O)(
115	42607465	LysoPE(20:4(5Z,8Z,11Z,14Z)/0:0)	HMDB0011517	42607465		[H][C@@](O)(
116	52925133	LysoPE(22:4(7Z,10Z,13Z,16Z)/0:0)	HMDB0011523	52925133		[H][C@@](O)(
117	42607484	PG(18:0/0:0)		42607484		CCCCCCCCC
118	9547135	PG(18:1(9Z)/0:0)		9547135		CCCCCCCC/
119	52927437	PG(18:2(9Z,12Z)/0:0)		52927437		CCCCC/C=C
120	42607497	PI(20:4(5Z,8Z,11Z,14Z)/0:0)		42607497		CCCCC/C=C
121	5962	NA	NA	NA	NA	NA
122	525	NA	NA	NA	NA	NA
123	5283467	MG(18:1(11E)/0:0/0:0)		5283467		CCCCCC/C=
124	5283468	Monooleoylglycerol	HMDB0094684	5283468		[H]\C(CCCC
125	5319879	MG(0:0/18:1(9Z)/0:0)	HMDB0011537	5319879		[H]C(CO)(CO
126	673	NA	NA	NA	NA	NA
127	496	NA	NA	NA	NA	NA
128	440810	NA	NA	NA	NA	NA
129	65095	NA	NA	NA	NA	NA
130	92919	NA	NA	NA	NA	NA
131	107461	NA	NA	NA	NA	NA
132	92832	NA	NA	NA	NA	NA
133	102175	NA	NA	NA	NA	NA
134	67427	NA	NA	NA	NA	NA
135	65065	NA	NA	NA	NA	NA
136	185	NA	NA	NA	NA	NA
137	25561	NA	NA	NA	NA	NA
138	131802901	NA	NA	NA	NA	NA
139	129397	NA	NA	NA	NA	NA
140	10221026	NA	NA	NA	NA	NA
141	445063	NA	NA	NA	NA	NA
142	74839	NA	NA	NA	NA	NA
143	700653	NA	NA	NA	NA	NA
144	936	NA	NA	NA	NA	NA
145	6450015	Linoelaidyl carnitine	HMDB0006461	6450015		CCCCC\C=C

146	47205608	NA	NA	NA	NA	NA
147	88490793	NA	NA	NA	NA	NA
148	6441392	NA	NA	NA	NA	NA
149	6441392	O-oleoylcarnitine		6441392		CCCCCCCC/
150	514186	NA	NA	NA	NA	NA
151	9547158	PA(16:0/18:1(11Z))	HMDB0007858	9547158	C00416	[H][C@@](CO
152	9547167	PA(16:0/18:2(9Z,12Z))	HMDB0007860	9547167	C00416	CCCCCCCC
153	24779559	PA(18:0/18:2(9Z,12Z))	HMDB0007861	24779559	C00416	CCCCCCCC
154	53478602	PA(18:1(11Z)/18:1(11Z))	HMDB0007862	53478602	C00416	[H][C@@](CO
155	6613	NA	NA	NA	NA	NA
156	24778634	PC(14:0/20:4(5Z,8Z,11Z,14Z))	HMDB0007883	24778634	C00157	CCCCCCCC
157	24778686	PC(16:0/18:0)	HMDB0007970	24778686	C00157	[H][C@@](CO
158	18631368	PC(18:0/16:0)	HMDB0008034	18631368	C00157	[H][C@@](CO
159	5288075	PC(18:2(9Z,12Z)/18:2(9Z,12Z))	HMDB0008138	5288075	C00157	[H][C@@](CO
160	24778979	PC(18:2(9Z,12Z)/20:4(5Z,8Z,11Z,14Z))	HMDB0008147	24778979	C00157	CCCCC\C=C
161	24778633	PC(14:0/20:0)	HMDB0007878	24778633	C00157	[H][C@@](CO
162	53478747	PC(18:1(11Z)/22:2(13Z,16Z))	HMDB0008086	53478747	C00157	CCCCCC\C=
163	6443157	PC(O-34:2)	HMDB0011151	6443157		CCCCCCCCC
164	53481709	PC(o-16:1(9Z)/18:2(9Z,12Z))	HMDB0013413	53481709		CCCCCC\C=
165	6443065	PC(o-18:0/20:4(8Z,11Z,14Z,17Z))	HMDB0013420	6443065		[H][C@@](CO
166	24779386	PC(P-16:0/18:2(9Z,12Z))	HMDB0011211	24779386		CCCCCCCCC
167	9546726	PE(16:0/18:1(11Z))	HMDB0008926	9546726	C00350	[H][C@@](CO
168	9546747	PE(16:0/18:2(9Z,12Z))	HMDB0008928	9546747	C00350	[H][C@@](CO
169	53479587	PE(16:0/20:3(5Z,8Z,11Z))	HMDB0008935	53479587	C00350	[H][C@@](CO
170	9546800	PE(16:0/20:4(5Z,8Z,11Z,14Z))	HMDB0008937	9546800	C00350	[H][C@@](CO
171	445757	PE(18:0/16:0)	HMDB0008989	445757	C00350	[H][C@@](CO
172	53479609	PE(18:0/18:1(11Z))	HMDB0008992	53479609	C00350	[H][C@@](CO
173	9546749	PE(18:0/18:2(9Z,12Z))	HMDB0008994	9546749	C00350	[H][C@@](CO
174	52924329	PE(18:0/18:3(6Z,9Z,12Z))	HMDB0008995	52924329	C00350	[H][C@@](CO
175	42627598	PE(18:0/20:3(5Z,8Z,11Z))	HMDB0009001	42627598	C00350	[H][C@@](CO
176	9547058	PE(O-18:1(1Z)/20:4(5Z,8Z,11Z,14Z))	HMDB0005779	9547058	C00350	[H][C@@](CO
177	52924901	PE(18:0/22:4(7Z,10Z,13Z,16Z))	HMDB0009009	52924901	C21481	[H][C@@](CO
178	53479611	PE(18:0/22:5(4Z,7Z,10Z,13Z,16Z))	HMDB0009010	53479611	C00350	[H][C@@](CO
179	9546798	PE(18:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	HMDB0009012	9546798	C00350	[H][C@@](CO
180	53479623	PE(18:1(11Z)/18:0)	HMDB0009024	53479623	C00350	[H][C@@](CO
181	53479626	PE(18:1(11Z)/18:2(9Z,12Z))	HMDB0009027	53479626	C00350	[H][C@@](CO
182	52924564	PE(20:1(11Z)/20:4(5Z,8Z,11Z,14Z))	HMDB0009267	52924564	C00350	[H][C@@](CO
183	53479554	PE(14:0/22:1(13Z))	HMDB0008842	53479554	C00350	[H][C@@](CO
184	9546755	NA	NA	NA	NA	NA
185	140348918	NA	NA	NA	NA	NA
186	6140	NA	NA	NA	NA	NA
187	14410593	NA	NA	NA	NA	NA
188	11902892	NA	NA	NA	NA	NA
189	15047	NA	NA	NA	NA	NA
190	3756	NA	NA	NA	NA	NA
191	493570	NA	NA	NA	NA	NA
192	52931165	SM(d16:1/20:0)		52931165		
193	52931183	SM(d16:1/22:0)		52931183		
194	52931189	SM(d16:1/23:0)		52931189		
195	52931201	SM(d16:1/24:1)		52931201		
196	134765538	SM(d17:1/16:0)		134765538		
197	46891684	SM(d18:1/23:0)	HMDB0012105	46891684	C00550	CCCCCCCCC
198	52931235	NA	NA	NA	NA	NA
199	52931237	SM(d18:2/22:0)		52931237		
200	52931209	SM(d18:2/23:0)		52931209		
201	52931217	SM(d18:2/24:0)		52931217		
202	52931173	SM(d19:1/18:0)		52931173		
203	52931133	SM(d16:1/16:0)		52931133		
204	11433862	SM(d18:1/14:0)	HMDB0012097	11433862		CCCCCCCCC
205	5283560	Sphingosine 1-phosphate	HMDB0000277	5283560	C06124	CCCCCCCCC
206	1110	Succinic acid	HMDB0000254	1110	C00042	O=C(O)CCC(
207	20849086	NA	NA	NA	NA	NA
208	169148	NA	NA	NA	NA	NA
209	169148	NA	NA	NA	NA	NA
210	6675	Taurocholic acid	HMDB0000036	6675	C05122	[H][C@@]1(CC
211	6440260	Tetracosenoic acid	HMDB0029799	6440260		CCCCCCCCC
212	71464539	(5Z,8Z)-tetradecadienoylcarnitine		71464539		CCCCC/C=C
213	53477791	Tetradecanoylcarnitine	HMDB0005066	53477791		CCCCCCCCC
214	192669	NA	NA	NA	NA	NA
215	11147	TG(16:0/16:0/16:0)	HMDB0005356	11147		[H]C(COC(=C
216	56937945	NA	NA	NA	NA	NA
217	56938423	NA	NA	NA	NA	NA
218	9544045	NA	NA	NA	NA	NA
219	56937262	NA	NA	NA	NA	NA
220	53481034	TG(15:0/16:0/20:4(5Z,8Z,11Z,14Z))	HMDB0011702	53481034		[H][C@](COC(
221	131750341	NA	NA	NA	NA	NA
222	99647498	NA	NA	NA	NA	NA
223	131750338	NA	NA	NA	NA	NA
224	131750395	NA	NA	NA	NA	NA
225	131761089	TG(53:1)	HMDB0050337	131761089		[H][C@@](CO
226	131764316	TG(53:3)	HMDB0053676	131764316		[H][C@@](CO
227	131766065	TG(53:4)	HMDB0055545	131766065		[H][C@@](CO
228	131766212	TG(53:5)	HMDB0055692	131766212		[H][C@@](CO
229	131750340	NA	NA	NA	NA	NA
230	131750353	NA	NA	NA	NA	NA
231	131750358	NA	NA	NA	NA	NA
232	131750342	NA	NA	NA	NA	NA

233	131750354	NA	NA	NA	NA	NA
234	131750359	NA	NA	NA	NA	NA
235	131750405	NA	NA	NA	NA	NA
236	14390011	TG(18:3(9Z,12Z,15Z)/18:2(9Z,12Z)/18:3(9Z,12Z,15Z))	HMDB0010507	14390011		[H]C(COC(=C
237	131764957	TG(55:2)	HMDB0054360	131764957		[H]\C(CCCCC
238	56936760	NA	NA	NA	NA	NA
239	131765229	TG(55:4)	HMDB0054657	131765229		[H][C@@](CO
240	131750355	NA	NA	NA	NA	NA
241	131750373	NA	NA	NA	NA	NA
242	25240380	NA	NA	NA	NA	NA
243	9544625	NA	NA	NA	NA	NA
244	56939382	NA	NA	NA	NA	NA
245	56938699	NA	NA	NA	NA	NA
246	9544695	NA	NA	NA	NA	NA
247	131750360	NA	NA	NA	NA	NA
248	131750406	NA	NA	NA	NA	NA
249	131750428	NA	NA	NA	NA	NA
250	53480557	TG(18:3(9Z,12Z,15Z)/18:2(9Z,12Z)/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	HMDB0010510	53480557		[H][C@](COC(
251	131750370	NA	NA	NA	NA	NA
252	131750412	NA	NA	NA	NA	NA
253	9545265	NA	NA	NA	NA	NA
254	131750424	NA	NA	NA	NA	NA
255	131750375	NA	NA	NA	NA	NA
256	131766428	TG(58:9)	HMDB0055940	131766428		[H][C@@](CO
257	22833596	CAR(4:1(2Me))		22833596		
258	6305	NA	NA	NA	NA	NA
259	5610	NA	NA	NA	NA	NA
260	736715	NA	NA	NA	NA	NA
261	64959	NA	NA	NA	NA	NA
262	4369188	Alpha-Carotene	HMDB0003993	4369188	C05433	C\C(\C=C\C
263	5280489	B-Carotene	HMDB0000561	5280489	C02094	C\C(\C=C\C

The second step is to check concentration values. For SSP analysis, the concentration must be measured in *umol* for blood and CSF samples. The urinary concentrations must be first converted to *umol/mmol_creatinine* in order to compare with reported concentrations in literature. No missing or negative values are allowed in SSP analysis. The concentration data for QEA analysis is more flexible. Users can upload either the original concentration data or normalized data. Missing or negative values are allowed (coded as *NA*) for QEA.

5 Selection of Metabolite Set Library

Before proceeding to enrichment analysis, a metabolite set library has to be chosen. There are seven built-in libraries offered by MSEA:

- Metabolic pathway associated metabolite sets (*currently contains 99 entries*);
- Disease associated metabolite sets (reported in blood) (*currently contains 344 entries*);
- Disease associated metabolite sets (reported in urine) (*currently contains 384 entries*)
- Disease associated metabolite sets (reported in CSF) (*currently contains 166 entries*)
- Metabolite sets associated with SNPs (*currently contains 4598 entries*)
- Predicted metabolite sets based on computational enzyme knockout model (*currently contains 912 entries*)
- Metabolite sets based on locations (*currently contains 73 entries*)
- Drug pathway associated metabolite sets (*currently contains 461 entries*)

In addition, MSEA also allows user-defined metabolite sets to be uploaded to perform enrichment analysis on arbitrary groups of compounds which researchers want to test. The metabolite set library is simply a two-column comma separated text file with the first column for metabolite set names and the second column for its compound names (**must use HMDB compound name**) separated by "; ". Please note, the built-in libraries are mainly from human studies. The functional grouping of metabolites may not be valid. Therefore, for data from subjects other than human being, users are suggested to upload their self-defined metabolite set libraries for enrichment analysis.

6 Enrichment Analysis

Over Representation Analysis (ORA) is performed when a list of compound names is provided. The list of compound list can be obtained through conventional feature selection methods, or from a clustering algorithm, or from the compounds with abnormal concentrations detected in SSP, to investigate if some biologically meaningful patterns can be identified.

ORA was implemented using the *hypergeometric test* to evaluate whether a particular metabolite set is represented more than expected by chance within the given compound list. One-tailed p values are provided after adjusting for multiple testing. **Figure 2** below summarizes the result.

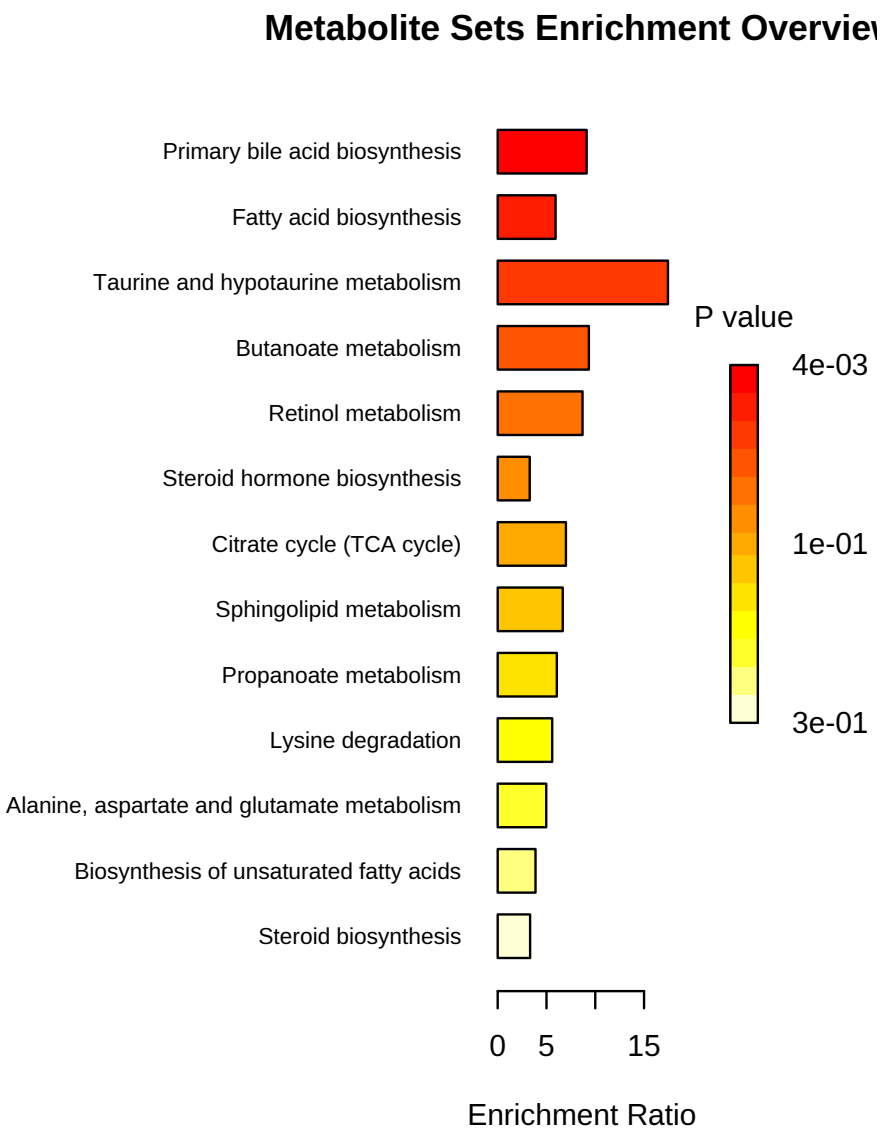

Figure 1: Summary Plot for Over Representation Analysis (ORA)

Table 2: Result from Over Representation Analysis

	total	expected	hits	Raw p	Holm p	FDR
Primary bile acid biosynthesis	46	0.33	3	3.51E-03	2.95E-01	2.95E-01
Fatty acid biosynthesis	47	0.34	2	4.23E-02	1.00E+00	1.00E+00
Taurine and hypotaurine metabolism	8	0.06	1	5.60E-02	1.00E+00	1.00E+00
Butanoate metabolism	15	0.11	1	1.03E-01	1.00E+00	1.00E+00
Retinol metabolism	16	0.12	1	1.09E-01	1.00E+00	1.00E+00
Steroid hormone biosynthesis	85	0.61	2	1.20E-01	1.00E+00	1.00E+00
Citrate cycle (TCA cycle)	20	0.14	1	1.35E-01	1.00E+00	1.00E+00
Sphingolipid metabolism	21	0.15	1	1.41E-01	1.00E+00	1.00E+00
Propanoate metabolism	23	0.17	1	1.53E-01	1.00E+00	1.00E+00
Lysine degradation	25	0.18	1	1.66E-01	1.00E+00	1.00E+00
Alanine, aspartate and glutamate metabolism	28	0.20	1	1.84E-01	1.00E+00	1.00E+00
Biosynthesis of unsaturated fatty acids	36	0.26	1	2.30E-01	1.00E+00	1.00E+00
Steroid biosynthesis	42	0.30	1	2.64E-01	1.00E+00	1.00E+00

7 Appendix: R Command History

```
 [1] "mSet<-InitDataObjects(\"conc\", \"msetora\", FALSE)"
 [2] "cmpd.vec<-c(\"52931205\",\"79437\",\"27476\",\"96373\",\"69726\",\"95433\",\"11671\",\"7067912
 [3] "mSet<-Setup.MapData(mSet, cmpd.vec);"
 [4] "mSet<-CrossReferencing(mSet, \"pubchem\", lipid = T);"
 [5] "mSet<-CreateMappingResultTable(mSet)"
 [6] "mSet<-SetMetabolomeFilter(mSet, F);"
 [7] "mSet<-SetCurrentMsetLib(mSet, \"kegg_pathway\", 2);"
 [8] "mSet<-CalculateHyperScore(mSet)"
 [9] "mSet<-PlotORA(mSet, \"ora_0_\", \"net\", \"png\", 72, width=NA)"
[10] "mSet<-PlotEnrichDotPlot(mSet, \"ora\", \"ora_dot_0_\", \"png\", 72, width=NA)"
[11] "mSet<-CalculateHyperScore(mSet)"
[12] "mSet<-PlotORA(mSet, \"ora_1_\", \"net\", \"png\", 72, width=NA)"
[13] "mSet<-PlotEnrichDotPlot(mSet, \"ora\", \"ora_dot_1_\", \"png\", 72, width=NA)"
[14] "mSet<-SaveTransformedData(mSet)"
[15] "mSet<-PreparePDFReport(mSet, \"guest325551242525186926\")\n"
```

The report was generated on Wed Dec 7 07:20:05 2022 with R version 4.2.2 (2022-10-31), OS system: Linux, version: -Ubuntu SMP Thu Oct 13 08:03:55 UTC 2022 .