

Metabolomic Data Analysis with MetaboAnalystR

Name: guest1537008872993960577

February 7, 2018

1 Background

The aim of the Network Explorer module is to provide a comprehensive tool to describe user's biological data as a network, as well as to explore interactions between biological entities and their potential impacts to help inform hypothesis generation. This module supports network visualization of both metabolomic and gene list data. Here, we aim to provide support for the integration of transcriptomics and metabolomics data, as well as metagenomics and metabolomics data. MetaboAnalyst 4.0 implements a knowledge-based network approach, whereby user's data (metabolites and genes) can be projected onto five existing biological networks: 1) Pathway-based network discovery, 2) Gene-metabolite interaction network, 3) Metabolite-phenotype interaction network, 4) Metabolite-metabolite interaction network, and 5) Metabolite-gene-phenotype interaction network.

2 Network Explorer Overview

The Network Explorer module consists of three steps: uploading the data, compound name/gene id mapping, selecting the network analysis option, and then viewing the generated network in greater detail.

3 Data Input

For this module, users may upload either a list of metabolites, a list of genes, or both metabolites and genes. MetaboAnalyst 4.0 currently accepts only compound names, HMDB IDs, or KEGG IDs as metabolite identifiers. As well, we only accept Entrez IDs, RefSeqIDs, Genbank accession numbers, ENSEMBL gene accession numbers, Official Gene Symbol IDs, or KEGG orthologs (KO) as gene identifiers. The uploaded list of metabolites and/or genes is then mapped using our internal databases of metabolites and gene annotations. Following this step, users can select which of the five networks to begin exploring their data.

Table 1: Compound Name Mapping

	Query	Match	HMDB	PubChem	KEGG	Comment
1	C00116	Glycerol	HMDB0000131	753	C00116	1
2	C00565	Trimethylamine	HMDB0000906	1146	C00565	1
3	C00033	Acetic acid	HMDB0000042	176	C00033	1
4	C00583	Propylene glycol	HMDB0001881	1030	C00583	1
5	C00022	Pyruvic acid	HMDB0000243	1060	C00022	1
6	C00719	Betaine	HMDB0000043	247	C00719	1
7	C05984	2-Hydroxybutyric acid	HMDB0000008	11266	C05984	1
8	C00207	Acetone	HMDB0001659	180	C00207	1
9	C00065	L-Serine	HMDB0000187	5951	C00065	1
10	C00031	D-Glucose	HMDB0000122	5793	C00031	1
11	C00079	L-Phenylalanine	HMDB0000159	6140	C00079	1
12	C02632	Isobutyric acid	HMDB0001873	6590	C02632	1
13	C00064	L-Glutamine	HMDB0000641	5961	C00064	1
14	C00114	Choline	HMDB0000097	305	C00114	1
15	C00073	L-Methionine	HMDB0000696	6137	C00073	1
16	C00082	L-Tyrosine	HMDB0000158	6057	C00082	1
17	C00186	L-Lactic acid	HMDB0000190	107689	C00186	1
18	C00037	Glycine	HMDB0000123	750	C00037	1
19	C00543	Dimethylamine	HMDB0000087	674	C00543	1
20	C00077	Ornithine	HMDB0000214	6262	C00077	1
21	C00058	Formic acid	HMDB0000142	284	C00058	1
22	C00188	L-Threonine	HMDB0000167	6288	C00188	1
23	C00407	L-Isoleucine	HMDB0000172	6306	C00407	1
24	C00791	Creatinine	HMDB0000562	588	C00791	1
25	C00062	L-Arginine	HMDB0000517	6322	C00062	1
26	C00300	Creatine	HMDB0000064	586	C00300	1

Table 2: Gene Name Mapping

	Query	Entrez	Symbol	KO	Name
1	1737	1737	DLAT		dihydrolipoamide S-acetyltransferase
2	83440	83440	ADPGK	K08074	ADP-dependent glucokinase
3	3939	3939	LDHA		lactate dehydrogenase A
4	10911	10911	UTS2	K05248	urotensin 2
5	10690	10690	FUT9	K03663	fucosyltransferase 9 (alpha (1,3) fucosyltransferase)
6	10010	10010	TANK	K12650	TRAF family member-associated NFkB activator
7	11224	11224	RPL35		ribosomal protein L35
8	63826	63826	SRR	K12235	serine racemase
9	11031	11031	RAB31		RAB31, member RAS oncogene family
10	4190	4190	MDH1	K00025	malate dehydrogenase 1, NAD (soluble)
11	10782	10782	ZNF274	K12458	zinc finger protein 274
12	10993	10993	SDS		serine dehydratase
13	10455	10455	ECI2		enoyl-CoA delta isomerase 2
14	10963	10963	STIP1	K09553	stress-induced-phosphoprotein 1
15	10282	10282	BET1	K08504	blocked early in transport 1 homolog (S. cerevisiae)
16	11232	11232	POLG2		polymerase (DNA directed), gamma 2, accessory subunit
17	1188	1188	CLCNKB	K05018	chloride channel, voltage-sensitive Kb
18	1629	1629	DBT		dihydrolipoamide branched chain transacylase E2
19	10039	10039	PARP3		poly (ADP-ribose) polymerase family, member 3
20	1140	1140	CHRNA1	K04812	cholinergic receptor, nicotinic, beta 1 (muscle)
21	10019	10019	SH2B3		SH2B adaptor protein 3
22	112755	112755	STX1B		syntaxin 1B
23	10672	10672	GNA13	K04639	guanine nucleotide binding protein (G protein), alpha 13
24	10007	10007	GNPDA1		glucosamine-6-phosphate deaminase 1
25	1213	1213	CLTC	K04646	clathrin, heavy chain (Hc)
26	118788	118788	PIK3AP1		phosphoinositide-3-kinase adaptor protein 1
27	10594	10594	PRPF8		PRPF8 pre-mRNA processing factor 8 homolog (S. cerevisiae)
28	189	189	AGXT	K00830	alanine-glyoxylate aminotransferase
29	11186	11186	RASSF1	K09850	Ras association (RalGDS/AF-6) domain family member 1
30	11264	11264	PXMP4		peroxisomal membrane protein 4, 24kDa
31	108	108	ADCY2	K08042	adenylate cyclase 2 (brain)
32	121130	121130	OR10P1		olfactory receptor, family 10, subfamily P, member 1
33	10390	10390	CEPT1	K13644	choline/ethanolamine phosphotransferase 1
34	10161	10161	LPAR6		lysophosphatidic acid receptor 6
35	10060	10060	ABCC9		ATP-binding cassette, sub-family C (CFTR/MRP), member 9
36	10810	10810	WASF3	K06083	WAS protein family, member 3
37	113612	113612	CYP2U1	K07422	cytochrome P450, family 2, subfamily U, polypeptide 1
38	1119	1119	CHKA		choline kinase alpha
39	10164	10164	CHST4	K04746	carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 4
40	10335	10335	MRV1		murine retrovirus integration site 1 homolog
41	10682	10682	EBP	K01824	emopamil binding protein (sterol isomerase)
42	10154	10154	PLXNC1		plexin C1
43	10105	10105	PIF	K09565	peptidylprolyl isomerase F
44	10623	10623	POLR3C		polymerase (RNA) III (DNA directed) polypeptide C (62kD)
45	10059	10059	DNM1L	K17065	dynamitin 1-like
46	10333	10333	TLR6	K10169	toll-like receptor 6
47	10500	10500	SEMA6C		sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6C
48	10855	10855	HPSE	K07964	heparanase
49	1176	1176	AP3S1		adaptor-related protein complex 3, sigma 1 subunit
50	10134	10134	BACAP31		B-cell receptor-associated protein 31
51	11164	11164	NUDT5	K13987	nudix (nucleoside diphosphate linked moiety X)-type motif 5
52	10093	10093	ARPC4	K05755	actin related protein 2/3 complex, subunit 4, 20kDa
53	10505	10505	SEMA4F		sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4F
54	1147	1147	CHUK		conserved helix-loop-helix ubiquitous kinase
55	118672	118672	PSTK	K10837	phosphoserine-tRNA kinase
56	10850	10850	CLC27	K16598	chemokine (C-C motif) ligand 27
57	10005	10005	ACOT8		acyl-CoA thioesterase 8
58	10251	10251	SPRY3		sprouty homolog 3 (Drosophila)
59	10652	10652	YKT6	K08516	YKT6 v-SNARE homolog (S. cerevisiae)
60	1002	1002	CDH4	K06797	cadherin 4, type 1, R-cadherin (retinal)
61	114049	114049	WBSR22		Williams Beuren syndrome chromosome region 22
62	10585	10585	POMT1		protein-O-mannosyltransferase 1
63	10235	10235	RASGRP2	K12361	RAS guanyl releasing protein 2 (calcium and DAG-regulated)
64	11184	11184	MAP4K1		mitogen-activated protein kinase kinase kinase 1
65	10913	10913	EDAR	K05162	ectodysplasin A receptor
66	1017	1017	CDK2	K02206	cyclin-dependent kinase 2
67	1138	1138	CHRNA5	K04807	cholinergic receptor, nicotinic, alpha 5 (neuronal)
68	10458	10458	BAIAP2		BAI1-associated protein 2
69	10808	10808	HSPH1		heat shock 105kDa/110kDa protein 1
70	11338	11338	U2AF2	K12837	U2 small nuclear RNA auxiliary factor 2
71	10402	10402	ST3GAL6		ST3 beta-galactoside alpha-2,3-sialyltransferase 6
72	10929	10929	SRSF8		serine/arginine-rich splicing factor 8
73	10798	10798	OR511		olfactory receptor, family 5, subfamily I, member 1
74	100137049	100137049	PLA2G4B		phospholipase A2, group IVB (cytosolic)
75	11019	11019	LIAS		lipoic acid synthetase
76	11255	11255	HRH3	K04151	histamine receptor H3
77	115653	115653	KIR3DL3		killer cell immunoglobulin-like receptor, three domains, long cytoplasmic tail, 3
78	11040	11040	PIM2	K08806	pim-2 oncogene
79	10267	10267	RAMP1	K08447	receptor (G protein-coupled) activity modifying protein 1
80	1142	1142	CHRNA3	K04814	cholinergic receptor, nicotinic, beta 3 (neuronal)
81	1146	1146	CHRNA3	K04818	cholinergic receptor, nicotinic, gamma (muscle)
82	11065	11065	UBE2C		ubiquitin-conjugating enzyme E2C
83	10501	10501	SEMA6B		sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6B
84	1019	1019	CDK4	K02089	cyclin-dependent kinase 4
85	1132	1132	CHRM4	K04132	cholinergic receptor, muscarinic 4
86	10825	10825	NEU3		sialidase 3 (membrane sialidase)
87	10842	10842	PPP1R17		protein phosphatase 1, regulatory subunit 17
88	11025	11025	LILRB3		leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 3
89	10947	10947	AP3M2		adaptor-related protein complex 3, mu 2 subunit
90	120066	120066	OR5P3		olfactory receptor, family 5, subfamily P, member 3
91	1203	1203	CLN5	K12390	ceroid-lipofuscinosis, neuronal 5
92	116444	116444	GRIN3B	K05214	glutamate receptor, ionotropic, N-methyl-D-aspartate 3B
93	10477	10477	UBE2E3		ubiquitin-conjugating enzyme E2E 3
94	10558	10558	SPTLC1		serine palmitoyltransferase, long chain base subunit 1
95	118460	118460	EXOSC6		exosome component 6
96	11332	11332	ACOT7	K17360	acyl-CoA thioesterase 7
97	120796	120796	OR56A1		olfactory receptor, family 56, subfamily A, member 1
98	10130	10130	PDIA6		protein disulfide isomerase family A, member 6
99	11154	11154	AP4S1	K12403	adaptor-related protein complex 4, sigma 1 subunit
100	10490	10490	VT11B		vesicle transport through interaction with t-SNAREs homolog 1B (yeast)
101	1158	1158	CKM		creatine kinase, muscle
102	11009	11009	IL24		interleukin 24
103	10342	10342	TFG	K09292	TRK-fused gene
104	113189	113189	CHST14		carbohydrate (N-acetylglucosamine 4-O) sulfotransferase 14
105	10725	10725	NFAT5		nuclear factor of activated T-cells 5, tonicity-responsive
106	1050	1050	CEBPA	K09055	CCAAT/enhancer binding protein (C/EBP), alpha
107	103	103	ADAR		adenosine deaminase, RNA-specific
108	10201	10201	NME6		non-metastatic cells 6, protein expressed in (nucleoside-diphosphate kinase)
109	10802	10802	SEC24A		SEC24 family, member A (S. cerevisiae)
110	10555	10555	AGPAT2		1-acylglycerol-3-phosphate O-acyltransferase 2 (lysophosphatidic acid acyltransferase, beta)
111	100133941	100133941	CD24	K06469	CD24 molecule
112	1179	1179	CLCA1	K05027	chloride channel accessory 1
113	10535	10535	RNASEH2A	K10743	ribonuclease H2, subunit A

114	10273	10273	STUB1		STIP1 homology and U-box containing protein 1, E3 ubiquitin protein ligase
115	10580	10580	SORBS1		sorbin and SH3 domain containing 1
116	10629	10629	TAF6L		TAF6-like RNA polymerase II, p300/CBP-associated factor (PCAF)-associated factor, 65kDa
117	10523	10523	CHERP	K12841	calcium homeostasis endoplasmic reticulum protein
118	10992	10992	SF3B2		splicing factor 3b, subunit 2, 145kDa
119	1032	1032	CDKN2D		cyclin-dependent kinase inhibitor 2D (p19, inhibits CDK4)
120	116983	116983	ACAP3		ArfGAP with coiled-coil, ankyrin repeat and PH domains 3
121	10229	10229	COQ7	K06134	coenzyme Q7 homolog, ubiquinone (yeast)
122	10938	10938	EHD1	K12483	EH-domain containing 1
123	10239	10239	AP3S2		adaptor-related protein complex 3, sigma 2 subunit
124	11112	11112	HIBADH		3-hydroxyisobutyrate dehydrogenase
125	11231	11231	SEC63	K09540	SEC63 homolog (S. cerevisiae)
126	1139	1139	CHRNA7	K04809	cholinergic receptor, nicotinic, alpha 7 (neuronal)
127	10478	10478	SLC25A17		solute carrier family 25 (mitochondrial carrier; peroxisomal membrane protein, 34kDa), member 17
128	1022	1022	CDK7	K02202	cyclin-dependent kinase 7
129	11001	11001	SLC27A2		solute carrier family 27 (fatty acid transporter), member 2
130	1136	1136	CHRNA3	K04805	cholinergic receptor, nicotinic, alpha 3 (neuronal)
131	119679	119679	OR52J3		olfactory receptor, family 52, subfamily J, member 3
132	10483	10483	SEC23B		Sec23 homolog B (S. cerevisiae)
133	1021	1021	CDK6	K02091	cyclin-dependent kinase 6
134	10277	10277	UBE4B		ubiquitination factor E4B
135	10327	10327	AKR1A1		aldo-keto reductase family 1, member A1 (aldehyde reductase)
136	10075	10075	HUWE1		HECT, UBA and WWE domain containing 1, E3 ubiquitin protein ligase
137	119749	119749	OR4C46		olfactory receptor, family 4, subfamily C, member 46
138	10718	10718	NRG3	K05457	neuregulin 3
139	1080	1080	CFTR	K05031	cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7)
140	10723	10723	SLC12A7		solute carrier family 12 (potassium/chloride transporters), member 7
141	1160	1160	CKMT2		creatine kinase, mitochondrial 2 (sarcomeric)
142	10438	10438	C1D		C1D nuclear receptor corepressor
143	1031	1031	CDKN2C		cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)
144	10317	10317	B3GALT5	K03877	UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 5
145	119692	119692	OR51S1		olfactory receptor, family 51, subfamily S, member 1
146	1081	1081	CGA	K08522	glycoprotein hormones, alpha polypeptide
147	10563	10563	CXCL13	K10032	chemokine (C-X-C motif) ligand 13
148	10054	10054	UBA2		ubiquitin-like modifier activating enzyme 2
149	10915	10915	TCERG1		transcription elongation regulator 1
150	11017	11017	SNRNP27	K12846	small nuclear ribonucleoprotein 27kDa (U4/U6.U5)
151	10008	10008	KCNE3	K04897	potassium voltage-gated channel, Isk-related family, member 3
152	10381	10381	TUBB3		tubulin, beta 3 class III
153	10919	10919	EHMT2		euchromatic histone-lysine N-methyltransferase 2
154	10683	10683	DLL3		delta-like 3 (Drosophila)
155	117157	117157	SH2D1B		SH2 domain containing 1B
156	119682	119682	OR51L1		olfactory receptor, family 51, subfamily L, member 1
157	1152	1152	CKB		creatine kinase, brain
158	10991	10991	SLC38A3		solute carrier family 38, member 3
159	1020	1020	CDK5	K02090	cyclin-dependent kinase 5
160	115727	115727	RASGRP4	K12363	RAS guanyl releasing protein 4
161	10873	10873	ME3		malic enzyme 3, NADP(+)-dependent, mitochondrial
162	10717	10717	AP4B1	K12401	adaptor-related protein complex 4, beta 1 subunit
163	10053	10053	AP1M2		adaptor-related protein complex 1, mu 2 subunit

4 Network Explorer

The Network Explorer analysis module complements MetaboAnalyst's joint-Pathway Analysis module by allowing the identification of connections that cross pathway boundaries (e.g. metabolite-disease interactions) as well as enabling a more global view of the pathways which may not be obvious when examined individually. The Network Explorer module currently supports five types of biological networks including the KEGG global metabolic network, a gene-metabolite interaction network, a metabolite-disease interaction network, a metabolite-metabolite interaction network, and a metabolite-gene-disease interaction network. The last four networks are created based on information gathered from HMDB and STITCH databases, and are applicable to human studies only. The integration of network topological analysis, interactive network exploration, and functional enrichment analysis provides users with different views on their data. Interpreting metabolomic data and/or gene expression data in such a context is far more insightful, and will lead to the generation of testable experimental hypotheses.

The selected interaction network is: Gene-Metabolite Interaction Network. Figure 1. shows the plot of the created interaction network. The plot (Figure 1) shows feature names, which may be hard for users to decipher in the generated plot within this report. It is therefore recommended that users with large datasets to use the Network Explorer to visually explore their data online, where high-quality SVG/PNG images can be saved. As well, users will be able to directly edit their created network online.

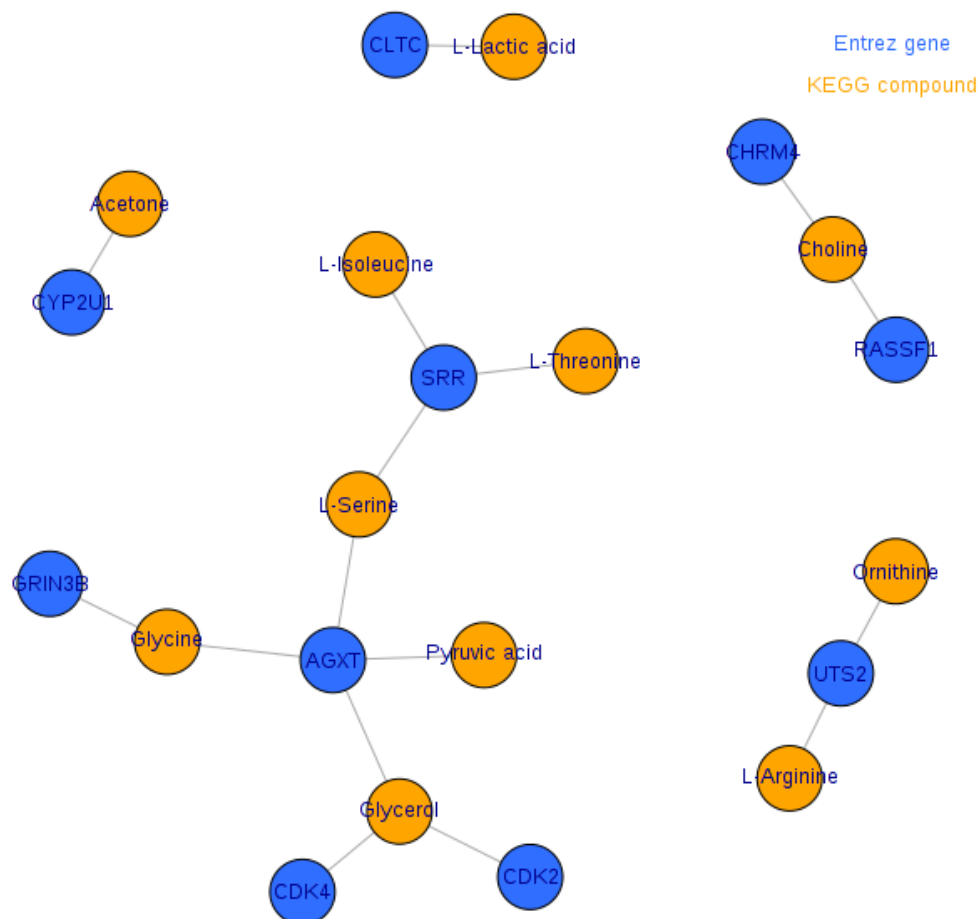


Figure 1: Plot of the selected interaction network. Grey nodes represent diseases, blue nodes represent genes, and yellow nodes represent metabolites.

5 Appendix: R Command History

```
[1] "InitDataObjects(\"conc\", \"network\", FALSE)"
[2] "mSet<-SetOrganism(mSet, \"hsa\")"
[3] "geneListFile<-\"replace_with_your_file_name\""
[4] "geneList<-readChar(geneListFile, file.info(geneListFile)$size)"
[5] "mSet<-PerformIntegGeneMapping(mSet, geneList, \"hsa\", \"entrez\");"
[6] "cmpdListFile<-\"replace_with_your_file_name\""
[7] "cmpdList<-readChar(cmpdListFile, file.info(cmpdListFile)$size)"
[8] "mSet<-PerformIntegCmpdMapping(mSet, cmpdList, \"hsa\", \"kegg\");"
[9] "mSet<-CreateMappingResultTable(mSet)"
[10] "mSet<-PrepareNetworkData(mSet);"
[11] "mSet<-SearchNetDB(mSet, \"pheno\", \"gene_metabolites\", FALSE, 0.5)"
[12] "mSet<-CreateGraph(mSet)"
[13] "mSet<-SaveTransformedData(mSet)"
```

The report was generated on Wed Feb 7 20:56:26 2018 with R version 3.4.1 (2017-06-30).