Knowledge that will change your world

Pathway software and databases

Stephen Barnes

University of Alabama at Birmingham

sbarnes@uab.edu

Synopsis

- **Public and proprietary metabolite databases**
  - **Instrument manufacturers provide software and search tools for their proprietary databases**
  - **Usually a single copy of the software is provided**
  - **Sciex and Waters provide additional copies at >$35,000 for each**
  - **Metabolomics service companies have proprietary databases and do not allow investigators submitting samples to have access to them**
    - **This discourages deeper research to be performed on metabolomics data**

Public databases

- North America
  - Metlin (https://metlin.scripps.edu/index.php)
  - Human metabolomics database http://www.hmdb.ca/
  - PubChem http://pubchem.ncbi.nlm.nih.gov/
  - Metabolomics Workbench (http://www.metabolomicsworkbench.org/data/index.php)
- Europe
  - ChemSpider (http://www.chemspider.com/)
- Asia
  - Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/)

Choosing the ions to search

- Go to the Download file from Metaboanalyst and open the file "peak_normalized_rt_mz.csv"
- Use Excel to calculate the p-values for all the metabolites (they're ordered by m/z in the file)
- Now sort the entire file by the p-values
- Choose the ions with p-values <0.001 (there should be 233) and copy them into a Word file and save the file
  - We'll search these in multiple databases

Metlin

METLIN: Metabolite Search

Batch

Simple (Saved Searches) | Advanced | Batch | Fragment | Neutral Loss | MS/MS Spectrum Match | Unknowns

Masses:

417.10292
173.08163
241.10831
499.71989
592.1392

Charge:

Neutral
Positive
Negative

M-H
M-H2O-H
M+Na-2H
M+Cl
M+K-2H
M+FA-H
M-2H
M-3H
M+CH3COO
M+F

*To select multiple Adducts:
- Hit Ctrl + Adducts
- Hit Command + Adducts
Select: all | none

Accuracy (ppm): 5

Display Structure:

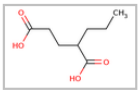
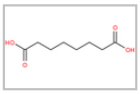
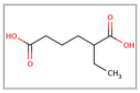
Remove peptides from search:

Find Metabolites | Reset

Metlin output-1

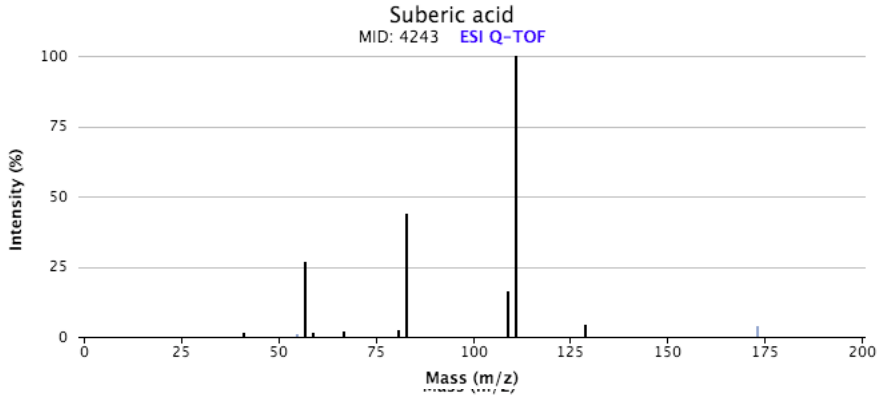
173.08163 m/z

(174.0872 - 174.0906 daltons): 7 Metabolites [M-H]⁻

MetlinID	Mass	Δppm	Name	Formula	CAS	MS/MS	Structure
2998	174.0892	1	2-Propylglutaric acid	C8H14O4	32806-62-5	NO	
4243	174.0892	1	Suberic acid	C8H14O4	505-48-6	View	
45928	174.0892	1	Ethyladipic acid	C8H14O4		NO	

Has MSMS data in this record

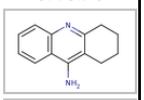
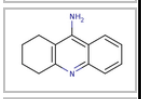
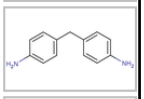
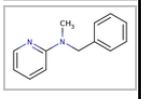
MSMS data of m/z 173.08163



Metlin output-2

197.11032 m/z

(198.1156 - 198.1196 daltons): 4 Metabolites [M-H]⁻

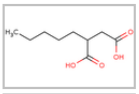
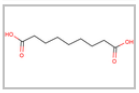
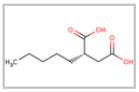
MetlinID	Mass	Δppm	Name	Formula	CAS	MS/MS	Structure
44521	198.1157	9	9-AMINO-1,2,3,4-TETRAHYDROACRIDINE	C13H14N2	NA	View	
2616	198.1157	9	Tacrine	C13H14N2	321-64-2	View	
69942	198.1157	9	4,4'-Methylenedianiline	C13H14N2	101-77-9	NO	
70631	198.1157	9	N-(Phenylmethyl)-N-methyl-2-pyridinamine	C13H14N2		NO	

Unlikely fits – high Δ ppm

Metlin output-3

187.09745 m/z

(188.1029 - 188.1066 daltons): 8 Metabolites [M-H]⁻

MetlinID	Mass	Δppm	Name	Formula	CAS	MS/MS	Structure
44183	188.1049	0	NONIC ACID	C9H16O4		View	
5750	188.1049	0	Nonanedioic acid	C9H16O4		View	
62450	188.1049	0	Nonate	C9H16O4		NO	

Spectra Search Mass Spectrum

MS Search | MS/MS Search | GC/MS Search | 1D NMR Search | 2D NMR Search

HMDB search

Query Masses (Da)

165.0555
425.04502
658.45859
201.11197
501.1806
329.11214
177.05641
269.04555
426.03226
226.03738
199.09784
227.09225
193.05045
490.13636
608.13368

Enter one mass per line (maximum 150 query masses per request)

Ionization

Ion Mode: Negative

Adduct Type: M-H, M+Na-2H, M+Cl, M+K-2H, M+FA-H, M+Hac-H, M+Br

Hold Ctrl ( ) or Command ( ) to select multiple adducts

Molecular Weight Tolerance ±: 10 ppm

Search | Load Example

HMDB search output-1

MS search for 173.08163 m/z

Delta = abs(query mass - adduct mass)

Show 10 entries

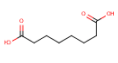
Search

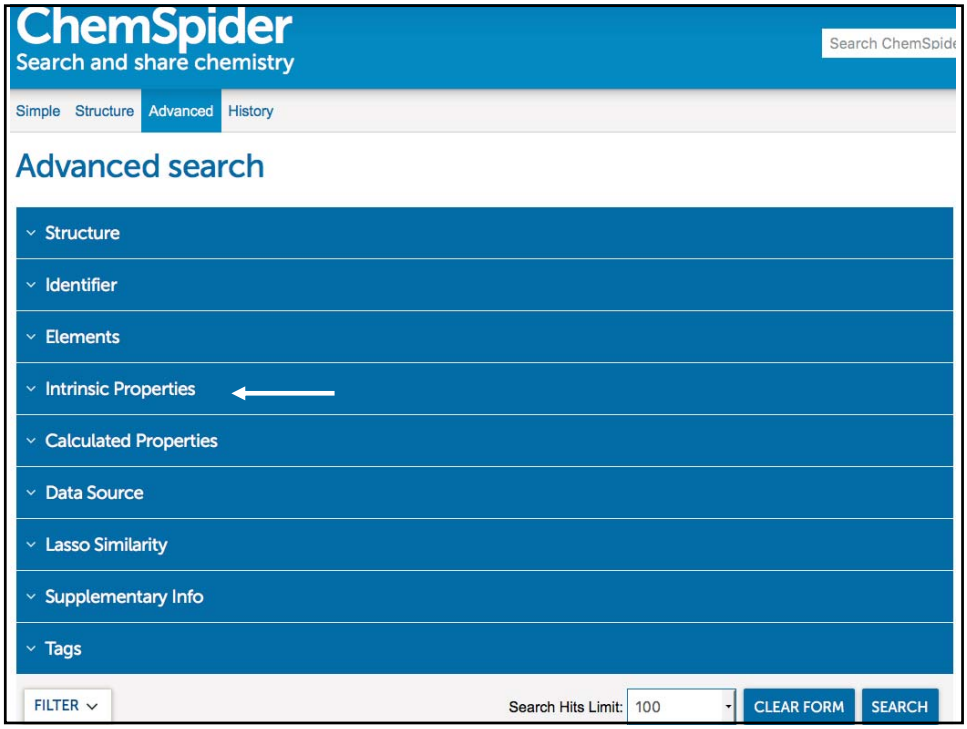
Compound	Name	Adduct	Adduct MW (Da)	Compound MW (Da)	Delta
HMDB00893	Suberic acid	M-H	173.081933	174.089208936	0.000303
HMDB02023	Ethyladipic acid	M-H	173.081933	174.089208936	0.000303
HMDB33838	Diethyl succinate	M-H	173.081933	174.089208936	0.000303
HMDB59727	2,4-Dimethyladipic acid	M-H	173.081933	174.089208936	0.000303
HMDB59757	3-Methylpimelic acid	M-H	173.081933	174.089208936	0.000303
HMDB60684	2-Propylglutaric acid	M-H	173.081933	174.089208936	0.000303

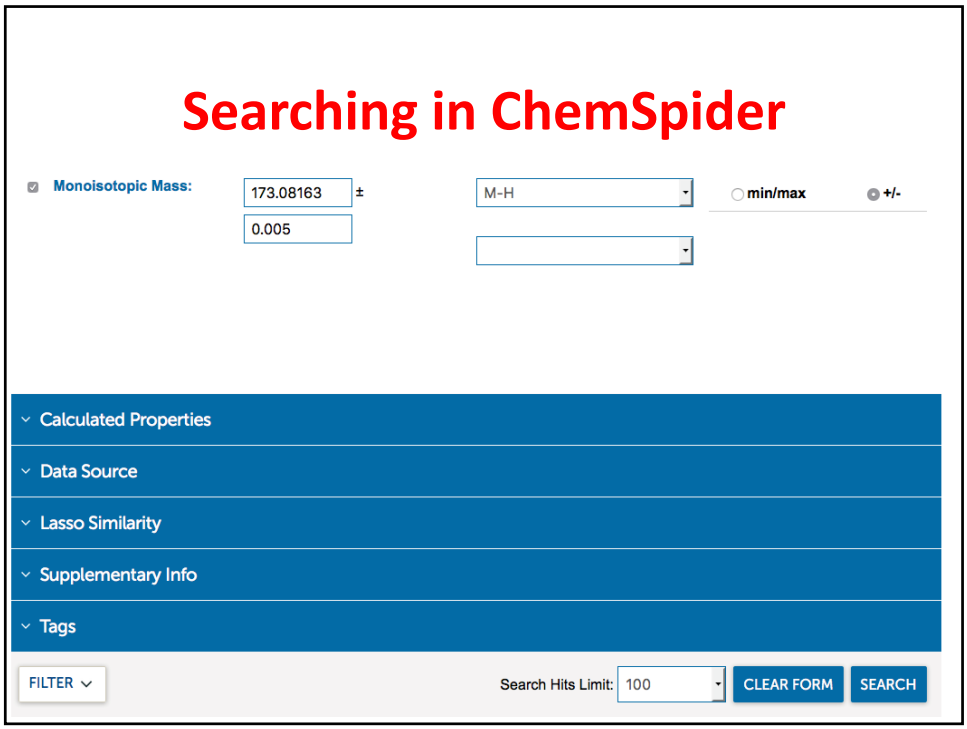
Showing 1 to 6 of 6 entries

Previous 1 Next

Click on HMDB record

Record Information	
Version	3.6
Creation Date	2005-11-16 15:48:42 UTC
Update Date	2015-12-02 05:10:07 UTC
HMDB ID	HMDB00893
Secondary Accession Numbers	• HMDB00837
Metabolite Identification	
Common Name	**Suberic acid**
Description	Suberic acid, also octanedioic acid, is a dicarboxylic acid, with formula C6H12(COOH)2. It is present in the urine of patients with Fatty Acid Oxidation Disorders (PMID 10404733). A metabolic breakdown product derived from oleic acid. Elevated levels of this unsaturated dicarboxylic acid are found in individuals with dicarboxylic acid and medium-chain acyl-CoA dehydrogenase deficiency (MCAD).
Structure	 MOL SDF PDB SMILES InChI
Synonyms	1. 1,6-Dicarboxyhexane 2. 1,6-Hexanedicarboxylate 3. 1,6-Hexanedicarboxylic acid 4. 1,8-Octanedioate 5. 1,8-Octanedioic acid 6. Cork acid


ChemSpider
Search and share chemistry
Search ChemSpide
Simple
Structure
Advanced
History
Advanced search
Structure
Identifier
Elements
Intrinsic Properties
Calculated Properties
Data Source
Lasso Similarity
Supplementary Info
Tags
FILTER
Search Hits Limit:
100
CLEAR FORM
SEARCH

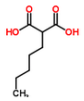

Searching in ChemSpider
Monoisotopic Mass:
173.08163
±
0.005
M-H
min/max
+/-
Calculated Properties
Data Source
Lasso Similarity
Supplementary Info
Tags
FILTER
Search Hits Limit:
100
CLEAR FORM
SEARCH

ChemSpider output

Found 1374 results

Search term: **MM >= 174.083906 AND MM <= 174.093906 AND abs(Monoisotopic_Mass - 174.088906451991) as mass_defect**

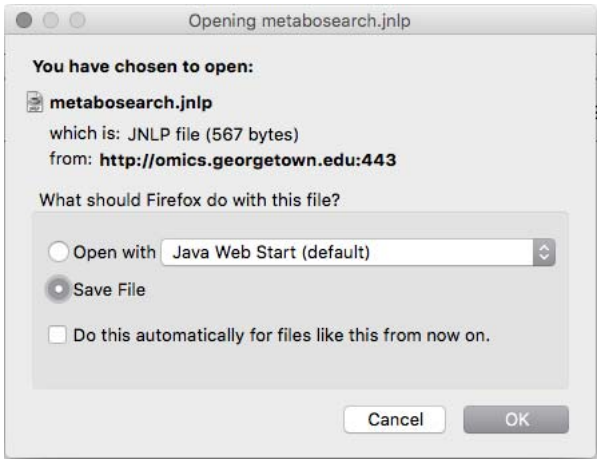
1 2 3 4 5

ID	Structure	Molecular Formula	Molecular Weight	# of Data Sources	# of References	# of PubMed	# of RSC	Mass defect ▲
10025		$C_8H_{14}O_4$	174.1944	109	264	73	275	0.0003
11364		$C_8H_{14}O_4$	174.1944	76	121	0	173	0.0003
11539		$C_8H_{14}O_4$	174.1944	17	18	0	4	0.0003

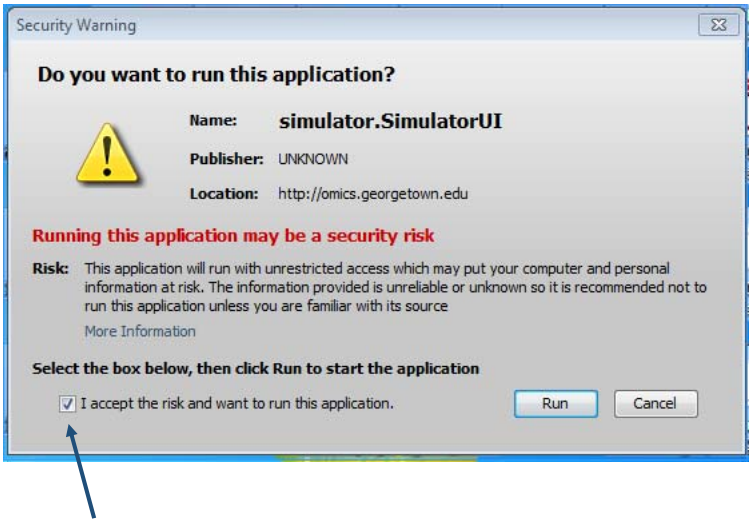
MetaboSearch

- **Find MetaboSearch using your browser**
  - **it's at http://omics.georgetown.edu/metabosearch.html**
  - **Note that this software may not work (for me, it won't allow me to output and save the data)**

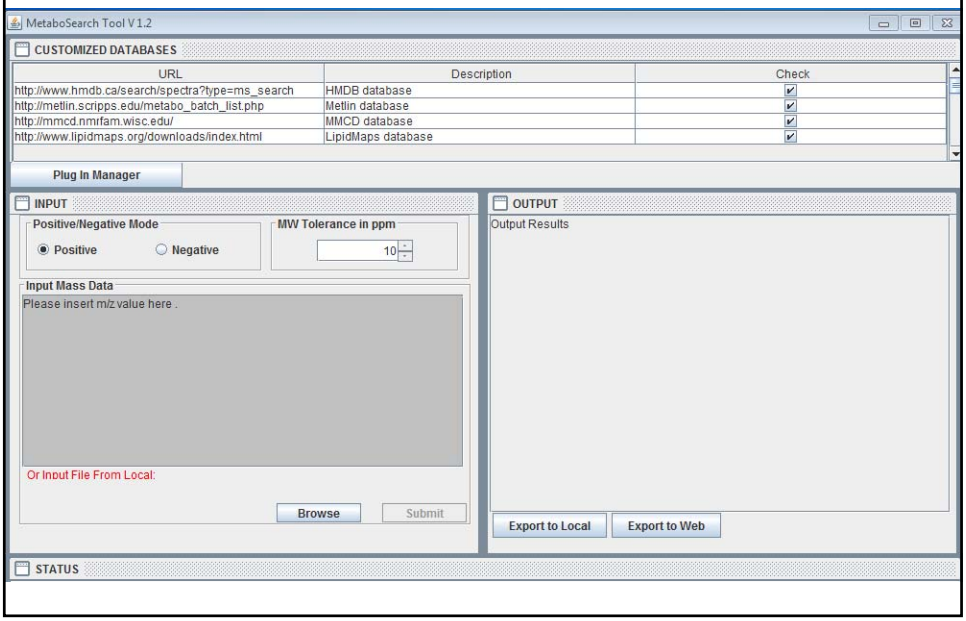
Download at Ressom lab



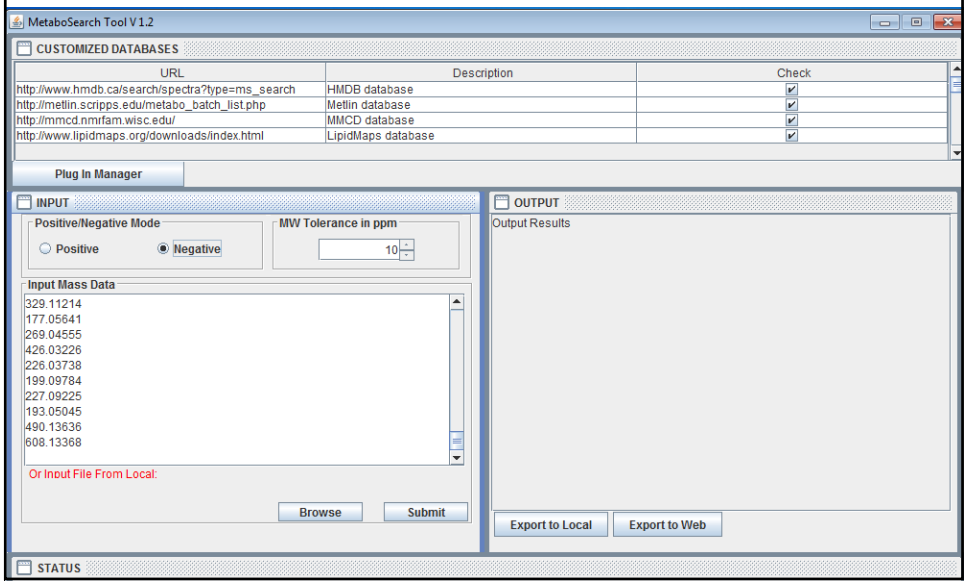
Opening MetaboSearch gives you this



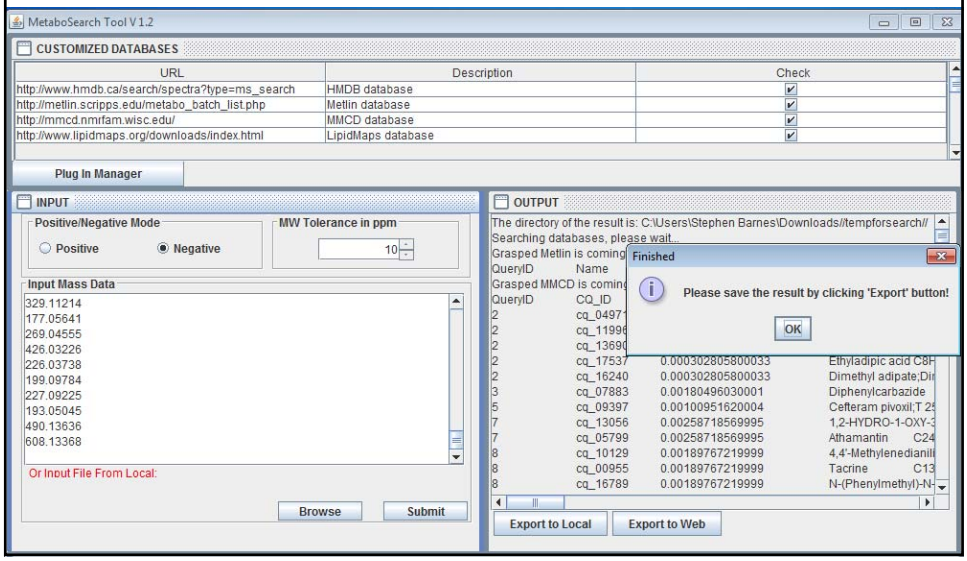
MetaboSearch set up page



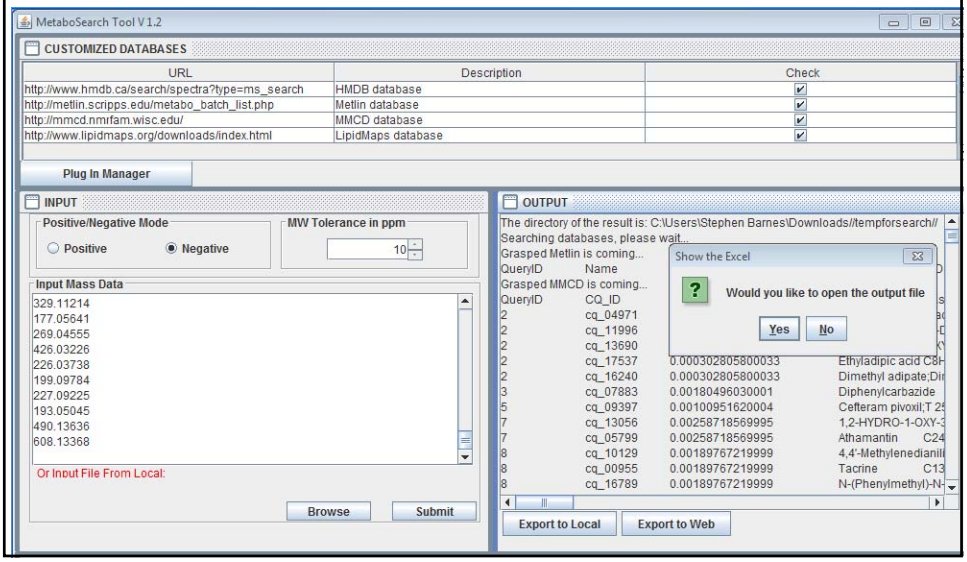
Data loaded – note negative ions



MetaboSearch completed



Opening Metabosearch file



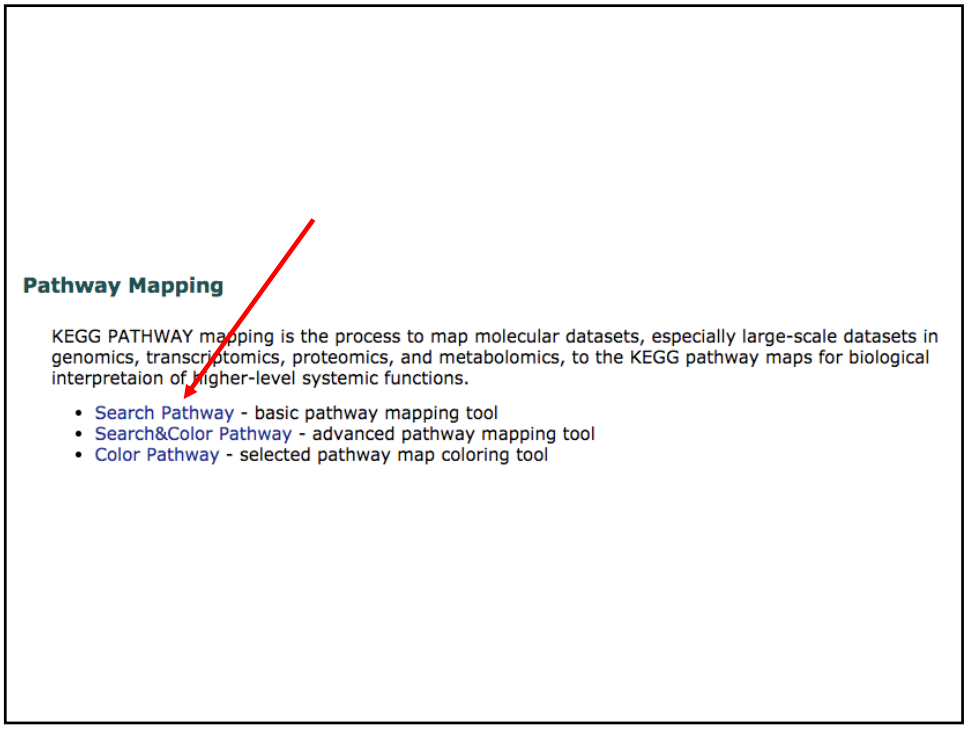
MetaboSearch output

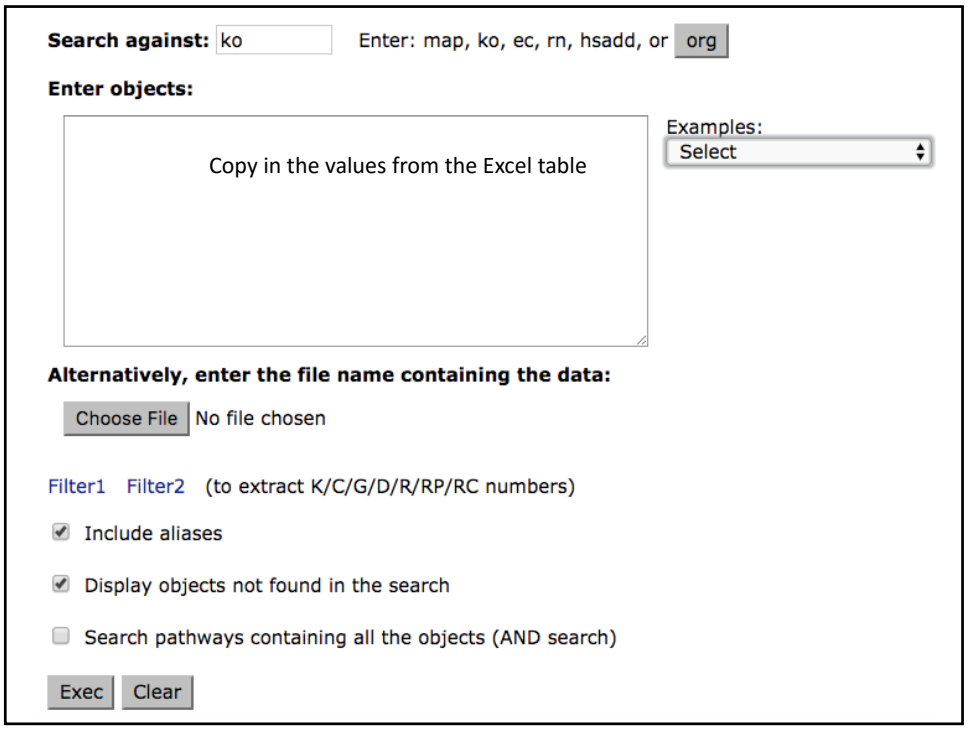
A	B	C	D	E	F	G	H	I	J	K	L	M
Query_ID	Query_m/	Input_RT	Name	Formula	Exact_Ma	KEGG ID	PubChem	PubChem	HMDB ID	Databases	dppm	Delta
1	417.1029	-	-	-	-	-	-	-	-	-	-	-
2	173.0816	-	[O4]-ACETOXY-2,3-DIDEOXYFI	C8H14O4	174.0892	-	-	-	-	MMCD	1.665336	2.90E-04
2	173.0816	-	2-(ALPHA-HYDROXYISOVALER	C8H14O4	174.0892	-	-	-	-	MMCD	1.665336	2.90E-04
2	173.0816	-	Ethyladipic acid	C8H14O4	174.0892	-	-	152459	HMDB020	MMCD	1.665336	2.90E-04
2	173.0816	-	Dimethyl adipate;Dimethyl h	C8H14O4	174.0892	C14570	-	-	-	MMCD	1.665336	2.90E-04
2	173.0816	-	Suberic acid	C8H14O4	174.0892	C08278	10457;-	10457	HMDB008	LIPIDMaps	1.665336	2.90E-04
2	173.0816	-	Ethyladipic acid	C8H14O4	174.0892	-	152459;-	152459	HMDB020	LIPIDMaps	1.665336	2.90E-04
2	173.0816	-	Diethyl succinate	C8H14O4	174.0892	-	-	31249	HMDB338	HMDB	1.665336	2.90E-04
2	173.0816	-	Suberic acid;Cork acid;1,8-Oct	C8H14O4	174.0892	C08278	153742	10457	HMDB008	MMCD	1.665336	2.90E-04
3	241.1077	-	Diphenylcarbazide	C13H14N4	242.1168	C11232	151941	-	-	MMCD	7.436648	0.001801
4	499.7199	-	-	-	-	-	-	-	-	-	-	-
5	592.1392	-	Cefteram pivoxil;T 2588	C22H27N9	593.1475	C13147	192387	54885	-	MMCD	1.646408	9.77E-04
6	155.0716	-	-	C8H12O3	156.0786	-	5312943	-	-	LIPIDMaps	1.857506	2.90E-04
6	155.0716	-	8-Hydroxy-5,6-octadienoic ac	C8H12O3	156.0786	-	-	-	HMDB311	HMDB	1.564215	2.44E-04
6	155.0716	-	2,5-Dimethyl-4-ethoxy-3(2H)-	C8H12O3	156.0786	-	-	3017596	HMDB322	HMDB	1.564215	2.44E-04
7	429.1893	-	Mammea E/BA	C24H30O7	430.1992	-	-	11517592	HMDB308	HMDB	6.029752	0.002594
7	429.1893	-	Armillarilin	C24H30O7	430.1992	-	-	21126389	HMDB316	HMDB	6.029752	0.002594
7	429.1893	-	Athamantin	C24H30O7	430.1992	C09123	11315	442051	-	MMCD	6.029752	0.002594
7	429.1893	-	1,2-HYDRO-1-OXY-3,4-HYDRO	C24H30O7	430.1992	-	-	-	-	MMCD	6.029752	0.002594

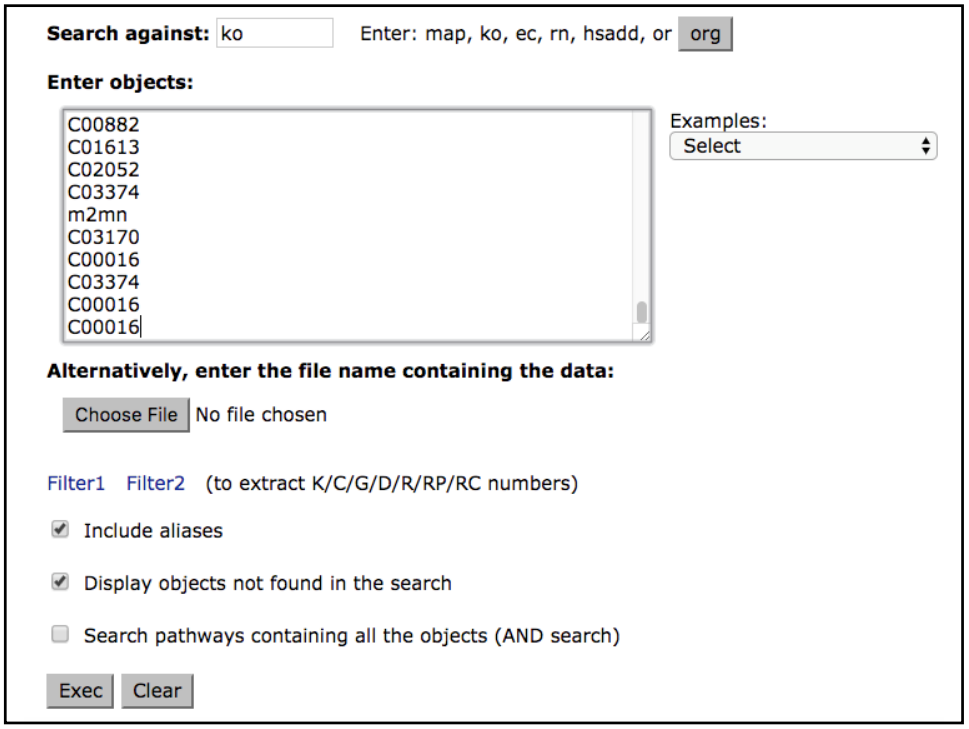
Going to the KEGG Pathway website

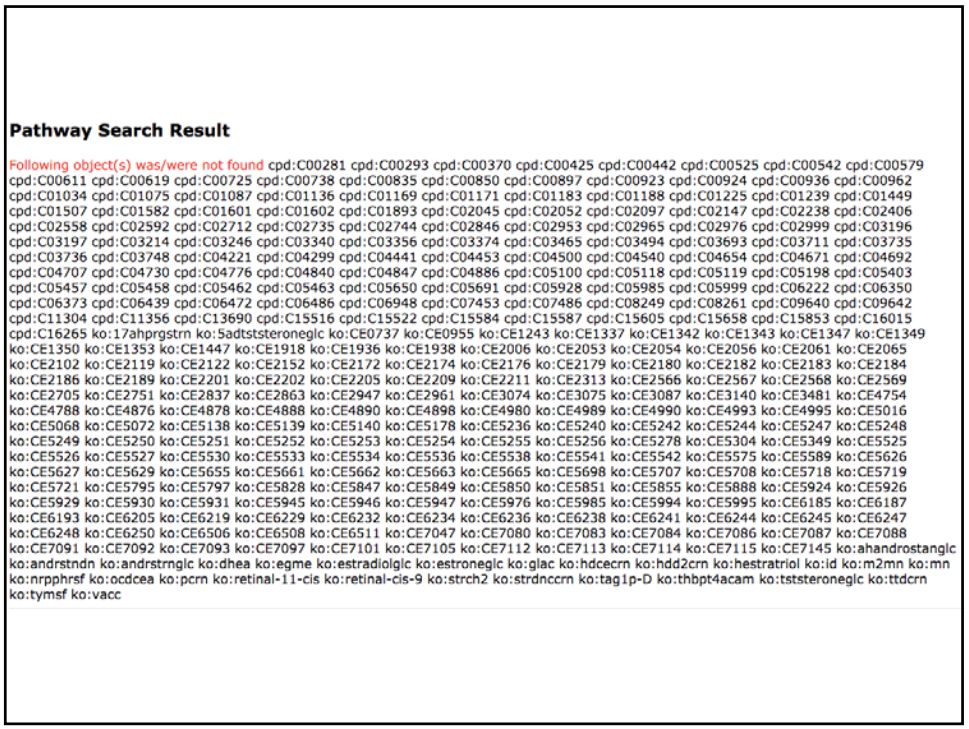
- http://www.genome.jp/kegg/pathway.html
- We'll take the compounds detected in MetaboSearch that have KEGG identifiers
- These can be mapped to pathways

- If Metabosearch does not work, then we will use the Excel file on the class website for 2-10-17 class


Pathway Mapping
KEGG PATHWAY mapping is the process to map molecular datasets, especially large-scale datasets in genomics, transcriptomics, proteomics, and metabolomics, to the KEGG pathway maps for biological interpretaion of higher-level systemic functions.
Search Pathway - basic pathway mapping tool
Search&Color Pathway - advanced pathway mapping tool
Color Pathway - selected pathway map coloring tool


Search against: ko
Enter: map, ko, ec, rn, hsadd, or org
Enter objects:
Copy in the values from the Excel table
Examples:
Select
Alternatively, enter the file name containing the data:
Choose File
No file chosen
Filter1 Filter2 (to extract K/C/G/D/R/RP/RC numbers)
Include aliases
Display objects not found in the search
Search pathways containing all the objects (AND search)
Exec
Clear



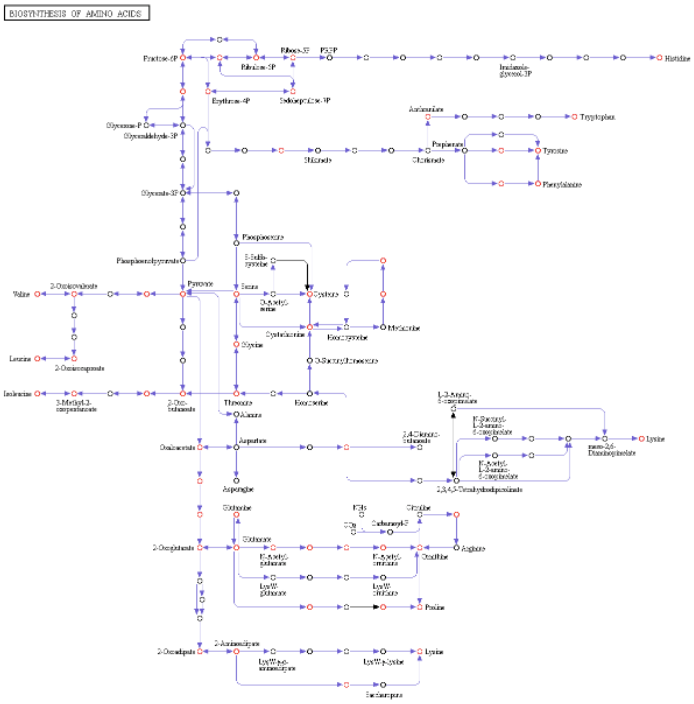


Pathway Search Result

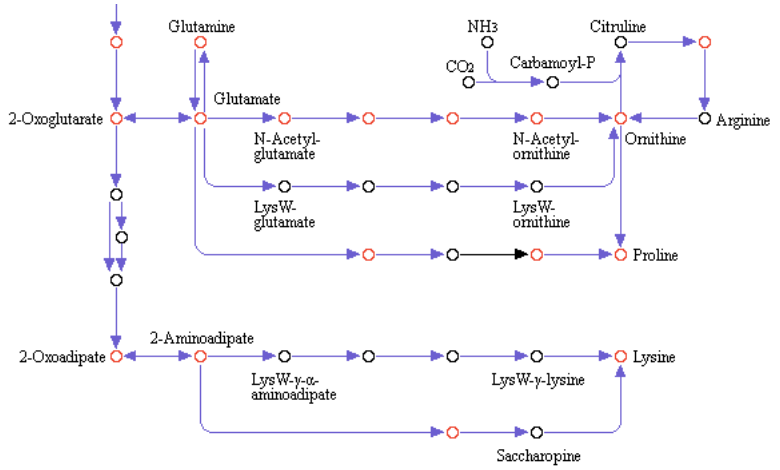
Following object(s) was/were not found cpd:C00281 cpd:C00293 cpd:C00370 cpd:C00425 cpd:C00442 cpd:C00525 cpd:C00542 cpd:C00579 cpd:C00611 cpd:C00619 cpd:C00725 cpd:C00738 cpd:C00835 cpd:C00850 cpd:C00897 cpd:C00923 cpd:C00924 cpd:C00936 cpd:C00962 cpd:C01034 cpd:C01075 cpd:C01087 cpd:C01136 cpd:C01169 cpd:C01171 cpd:C01183 cpd:C01188 cpd:C01225 cpd:C01239 cpd:C01449 cpd:C01507 cpd:C01582 cpd:C01601 cpd:C01602 cpd:C01893 cpd:C02045 cpd:C02052 cpd:C02097 cpd:C02147 cpd:C02238 cpd:C02406 cpd:C02558 cpd:C02592 cpd:C02712 cpd:C02735 cpd:C02744 cpd:C02846 cpd:C02953 cpd:C02965 cpd:C02976 cpd:C02999 cpd:C03196 cpd:C03197 cpd:C03214 cpd:C03246 cpd:C03340 cpd:C03356 cpd:C03374 cpd:C03465 cpd:C03494 cpd:C03693 cpd:C03711 cpd:C03735 cpd:C03736 cpd:C03748 cpd:C04221 cpd:C04299 cpd:C04441 cpd:C04453 cpd:C04500 cpd:C04540 cpd:C04654 cpd:C04671 cpd:C04692 cpd:C04707 cpd:C04730 cpd:C04776 cpd:C04840 cpd:C04847 cpd:C04886 cpd:C05100 cpd:C05118 cpd:C05119 cpd:C05198 cpd:C05403 cpd:C05457 cpd:C05458 cpd:C05462 cpd:C05463 cpd:C05650 cpd:C05691 cpd:C05928 cpd:C05985 cpd:C05999 cpd:C06222 cpd:C06350 cpd:C06373 cpd:C06439 cpd:C06472 cpd:C06486 cpd:C06948 cpd:C07453 cpd:C07486 cpd:C08249 cpd:C08261 cpd:C09640 cpd:C09642 cpd:C11304 cpd:C11356 cpd:C13690 cpd:C15516 cpd:C15522 cpd:C15584 cpd:C15587 cpd:C15605 cpd:C15658 cpd:C15853 cpd:C16015 cpd:C16265 ko:17ahprgstrn ko:5adtststeronegic ko:CE0737 ko:CE0955 ko:CE1243 ko:CE1337 ko:CE1342 ko:CE1343 ko:CE1347 ko:CE1349 ko:CE1350 ko:CE1353 ko:CE1447 ko:CE1918 ko:CE1936 ko:CE1938 ko:CE2006 ko:CE2053 ko:CE2054 ko:CE2056 ko:CE2061 ko:CE2065 ko:CE2102 ko:CE2119 ko:CE2122 ko:CE2152 ko:CE2172 ko:CE2174 ko:CE2176 ko:CE2179 ko:CE2180 ko:CE2182 ko:CE2183 ko:CE2184 ko:CE2186 ko:CE2189 ko:CE2201 ko:CE2202 ko:CE2205 ko:CE2209 ko:CE2211 ko:CE2313 ko:CE2566 ko:CE2567 ko:CE2568 ko:CE2569 ko:CE2705 ko:CE2751 ko:CE2837 ko:CE2863 ko:CE2947 ko:CE2961 ko:CE3074 ko:CE3075 ko:CE3087 ko:CE3140 ko:CE3481 ko:CE4754 ko:CE4788 ko:CE4876 ko:CE4878 ko:CE4888 ko:CE4890 ko:CE4898 ko:CE4980 ko:CE4989 ko:CE4990 ko:CE4993 ko:CE4995 ko:CE5016 ko:CE5068 ko:CE5072 ko:CE5138 ko:CE5139 ko:CE5140 ko:CE5178 ko:CE5236 ko:CE5240 ko:CE5242 ko:CE5244 ko:CE5247 ko:CE5248 ko:CE5249 ko:CE5250 ko:CE5251 ko:CE5252 ko:CE5253 ko:CE5254 ko:CE5255 ko:CE5256 ko:CE5278 ko:CE5304 ko:CE5349 ko:CE5525 ko:CE5526 ko:CE5527 ko:CE5530 ko:CE5533 ko:CE5534 ko:CE5536 ko:CE5538 ko:CE5541 ko:CE5542 ko:CE5575 ko:CE5589 ko:CE5626 ko:CE5627 ko:CE5629 ko:CE5655 ko:CE5661 ko:CE5662 ko:CE5663 ko:CE5665 ko:CE5698 ko:CE5707 ko:CE5708 ko:CE5718 ko:CE5719 ko:CE5721 ko:CE5795 ko:CE5797 ko:CE5828 ko:CE5847 ko:CE5849 ko:CE5850 ko:CE5851 ko:CE5855 ko:CE5888 ko:CE5924 ko:CE5926 ko:CE5929 ko:CE5930 ko:CE5931 ko:CE5945 ko:CE5946 ko:CE5947 ko:CE5976 ko:CE5985 ko:CE5994 ko:CE5995 ko:CE6185 ko:CE6187 ko:CE6193 ko:CE6205 ko:CE6219 ko:CE6229 ko:CE6232 ko:CE6234 ko:CE6236 ko:CE6238 ko:CE6241 ko:CE6244 ko:CE6245 ko:CE6247 ko:CE6248 ko:CE6250 ko:CE6506 ko:CE6508 ko:CE6511 ko:CE7047 ko:CE7080 ko:CE7083 ko:CE7084 ko:CE7086 ko:CE7087 ko:CE7088 ko:CE7091 ko:CE7092 ko:CE7093 ko:CE7097 ko:CE7101 ko:CE7105 ko:CE7112 ko:CE7113 ko:CE7114 ko:CE7115 ko:CE7145 ko:ahandrostanglc ko:andrstndn ko:andrstrnglc ko:dhea ko:egme ko:estradiolglc ko:estronegic ko:glac ko:hdcecrn ko:hdd2crn ko:hestratriol ko:id ko:m2mn ko:mn ko:nrpphrsf ko:ocdcea ko:pcrn ko:retinal-11-cis ko:retinal-cis-9 ko:strch2 ko:strdnccrn ko:tag1p-D ko:thbpt4acam ko:tststeronegic ko:ttdcrn ko:tymsf ko:vacc

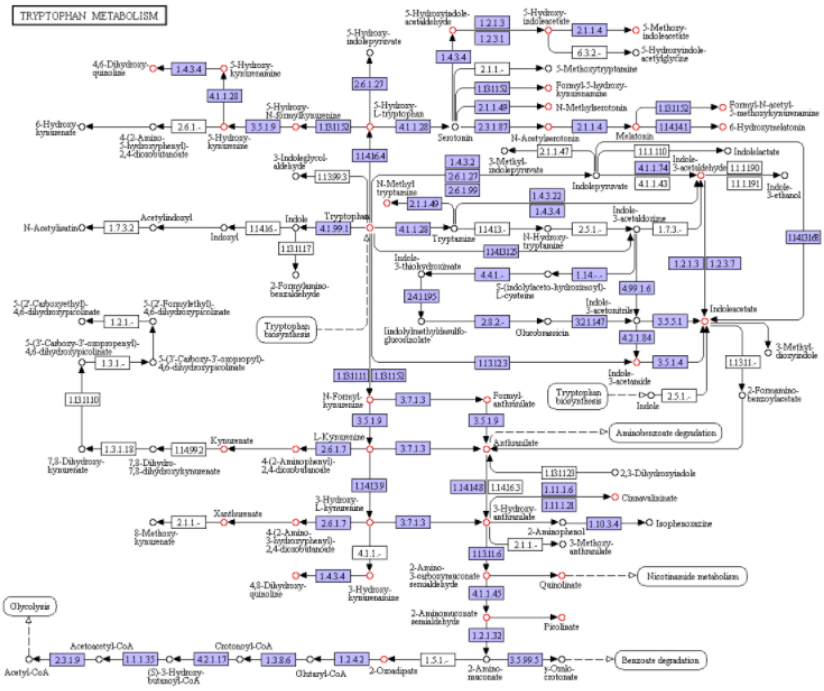
Show all objects

- ko01100 Metabolic pathways (399)
- ko01110 Biosynthesis of secondary metabolites (142)
- ko01120 Microbial metabolism in diverse environments (115)
- ko01130 Biosynthesis of antibiotics (94)
- ko00140 Steroid hormone biosynthesis (63)
- ko01230 Biosynthesis of amino acids (52)
- ko00350 Tyrosine metabolism (43)
- ko02010 ABC transporters (43)
- ko00980 Metabolism of xenobiotics by cytochrome P450 (38)
- ko00380 Tryptophan metabolism (37)



Expanded amino acid metabolism





Expanded tryptophan metabolism