Pathway Analysis with MetaboAnalyst

From a talk by

David Wishart

https://www.youtube.com/watch?v=1EEl9Cze_0I&t=319s

Metabolite Set Enrichment Analysis (MSEA)

http://www.msea.ca

Now part of MetaboAnalyst

- Designed to handle lists of metabolites (with or without concentration data)
- Modeled after Gene Set Enrichment Analysis (GSEA)
- Supports over-representation analysis (ORA), single sample profiling (SSP) and quantitative enrichment analysis (QEA)
- Contains a library of 6300 pre-defined metabolite sets including 85 pathway sets & 850 disease sets

Enrichment Analysis

Purpose: To test if there are biologically meaningful groups of metabolites that are significantly enriched in your data

Biological meaningful in terms of:

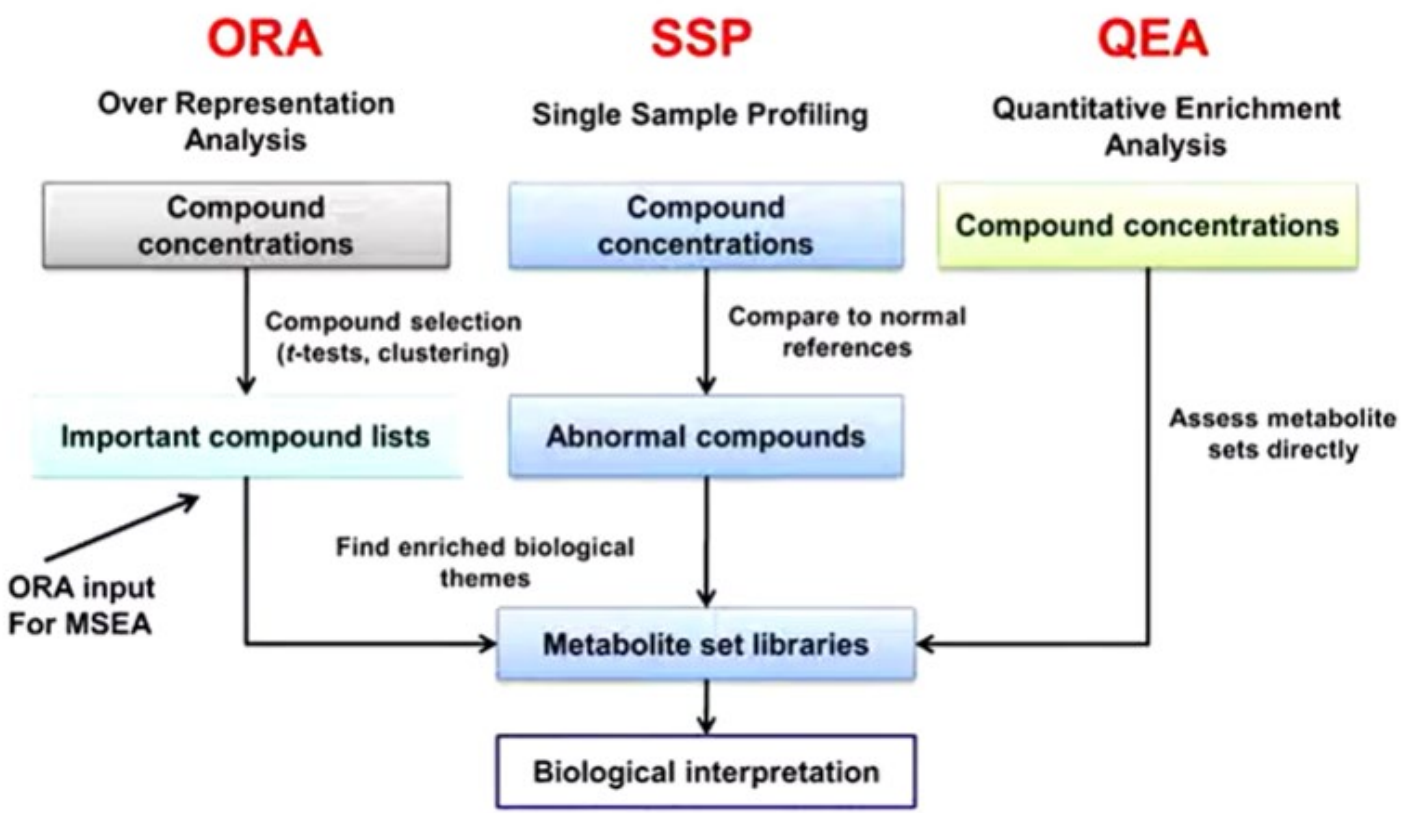
- Pathways
- Disease
- Localization

Currently, MSEA only supports human metabolomic data

MSEA

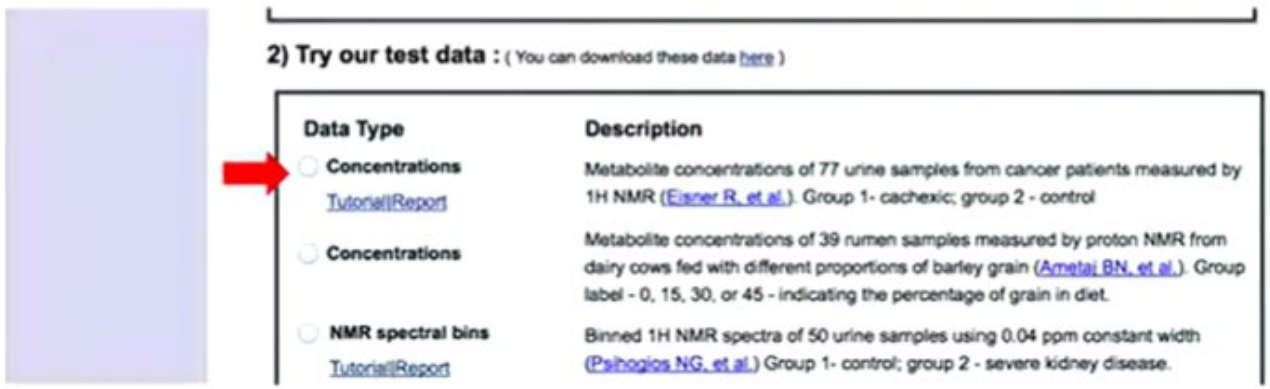
- Accepts 3 kinds of input files
  - list of metabolite names only (ORA – over representation analysis)
  - list of metabolite names + concentration data from a single sample (SSP – single sample profiling)
  - a concentration table with a list of metabolite names + concentrations for multiple samples/patients (QEA – quantitative enrichment analysis)

The MSEA Approach

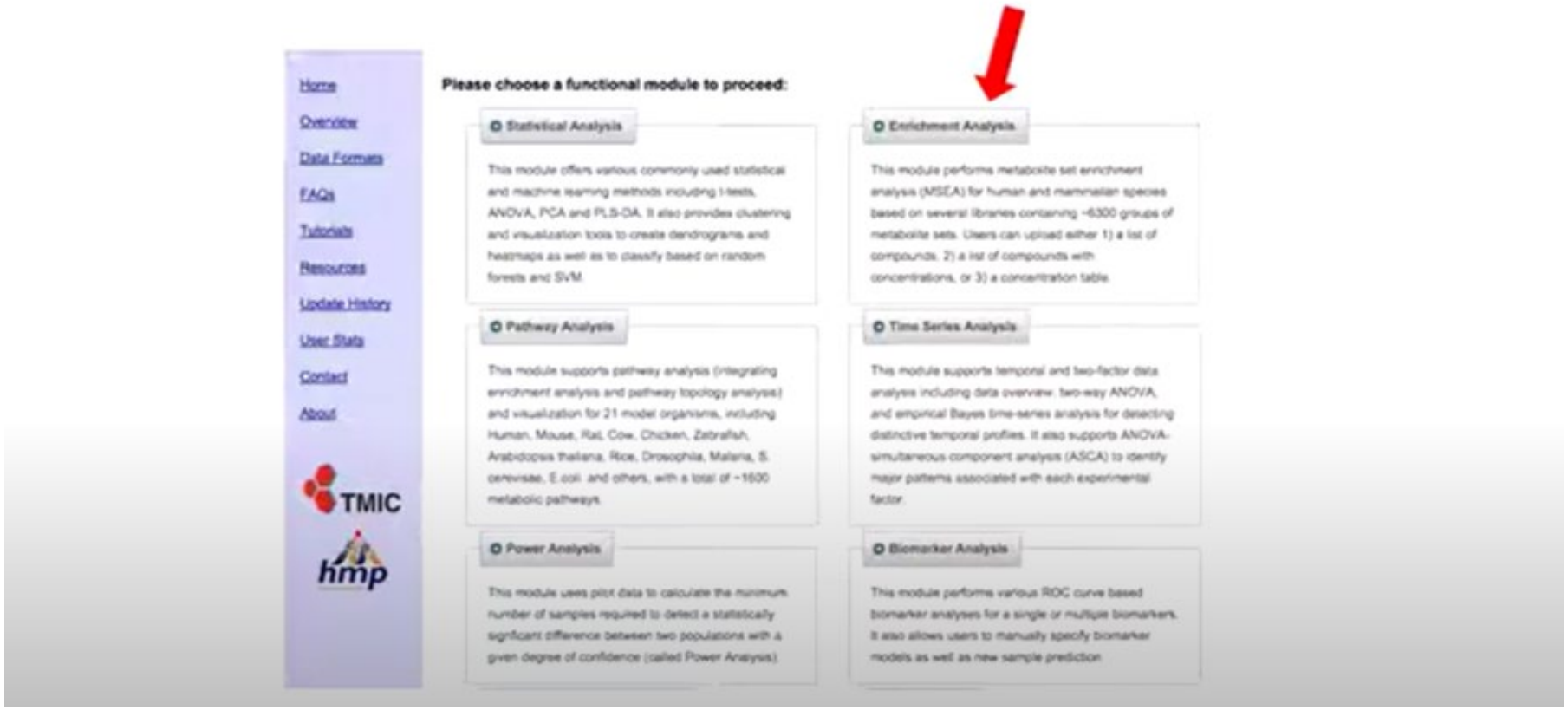


Data Set Selected

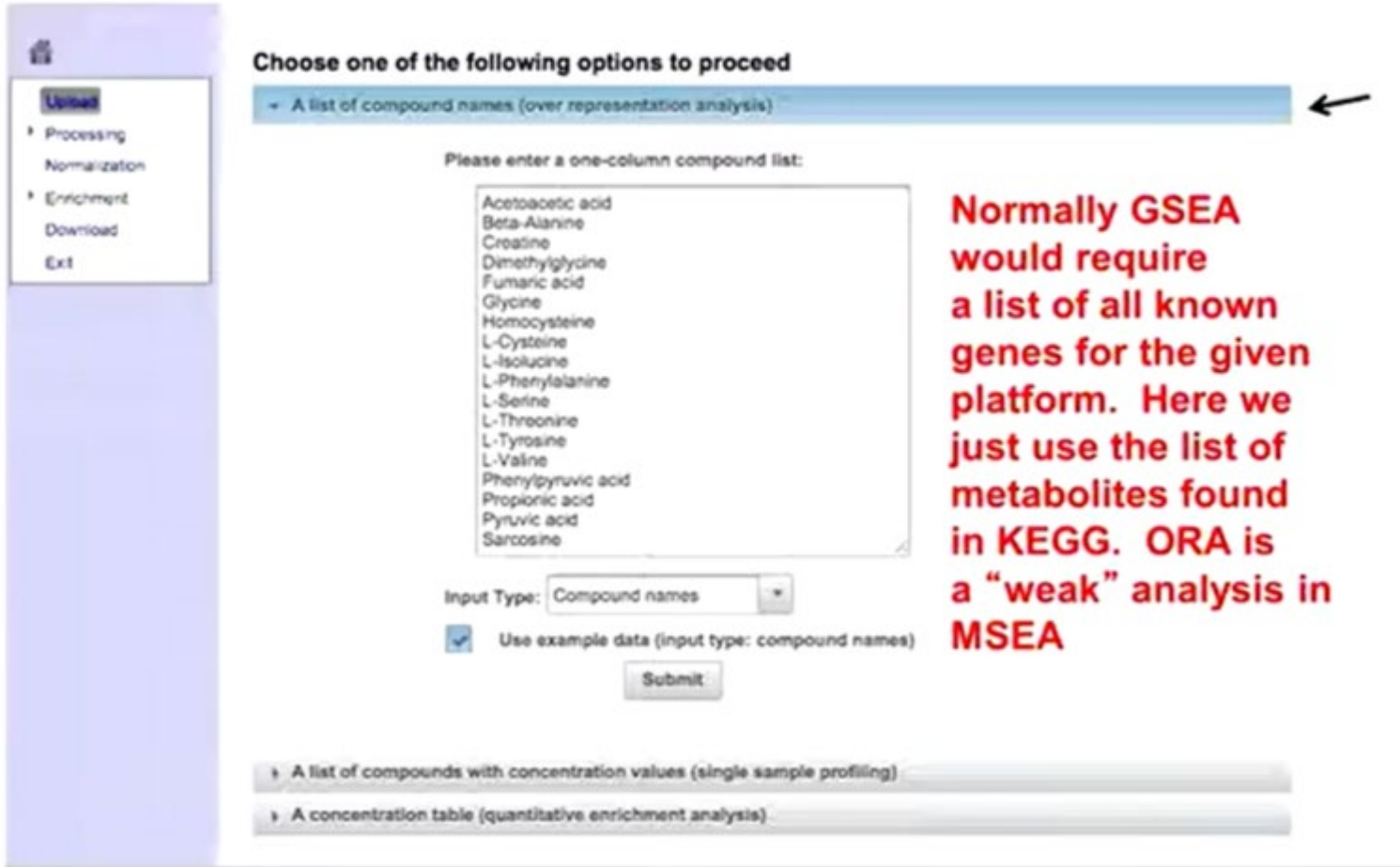
- Here we are using a collection of metabolites identified by NMR (compound list + concentrations) from the urine from 77 lung and colon cancer patients, some of whom were suffering from cachexia (muscle wasting)



Start with a Compound List for ORA

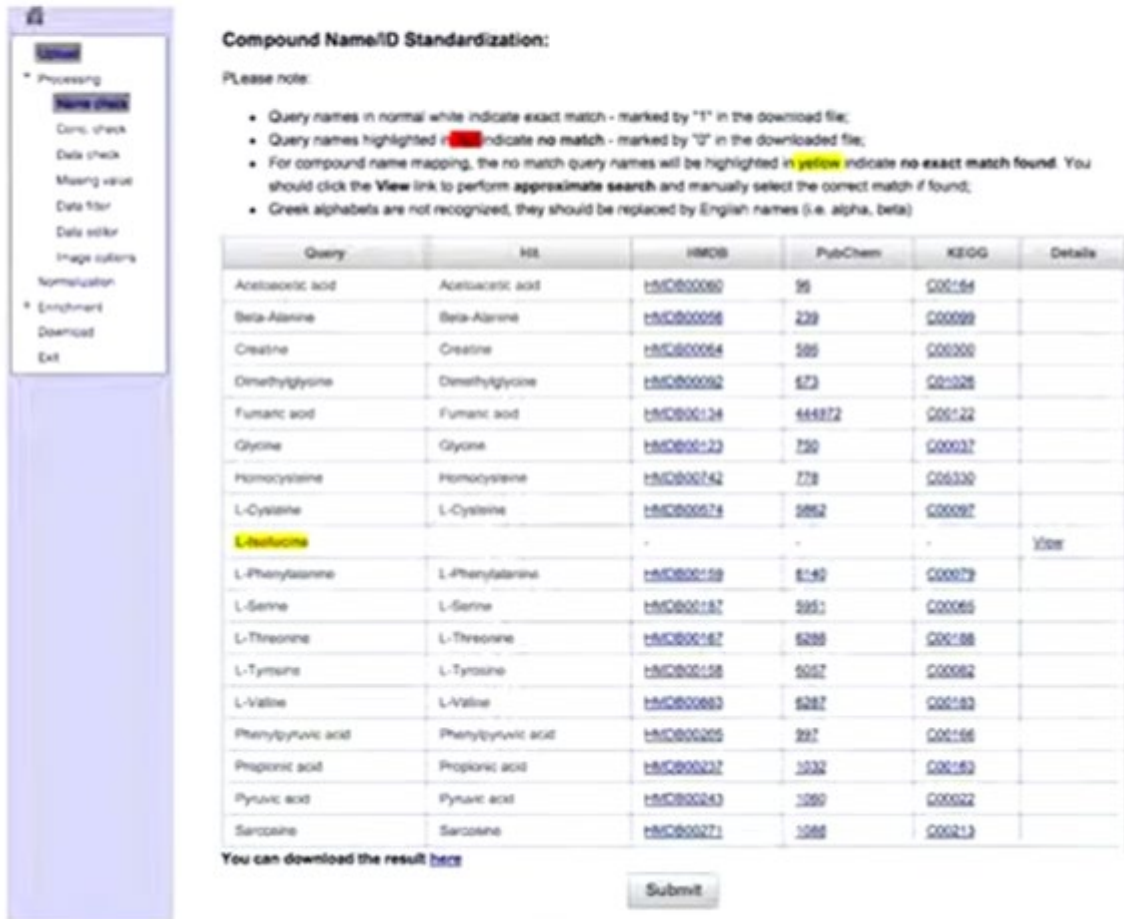


Upload Compound List

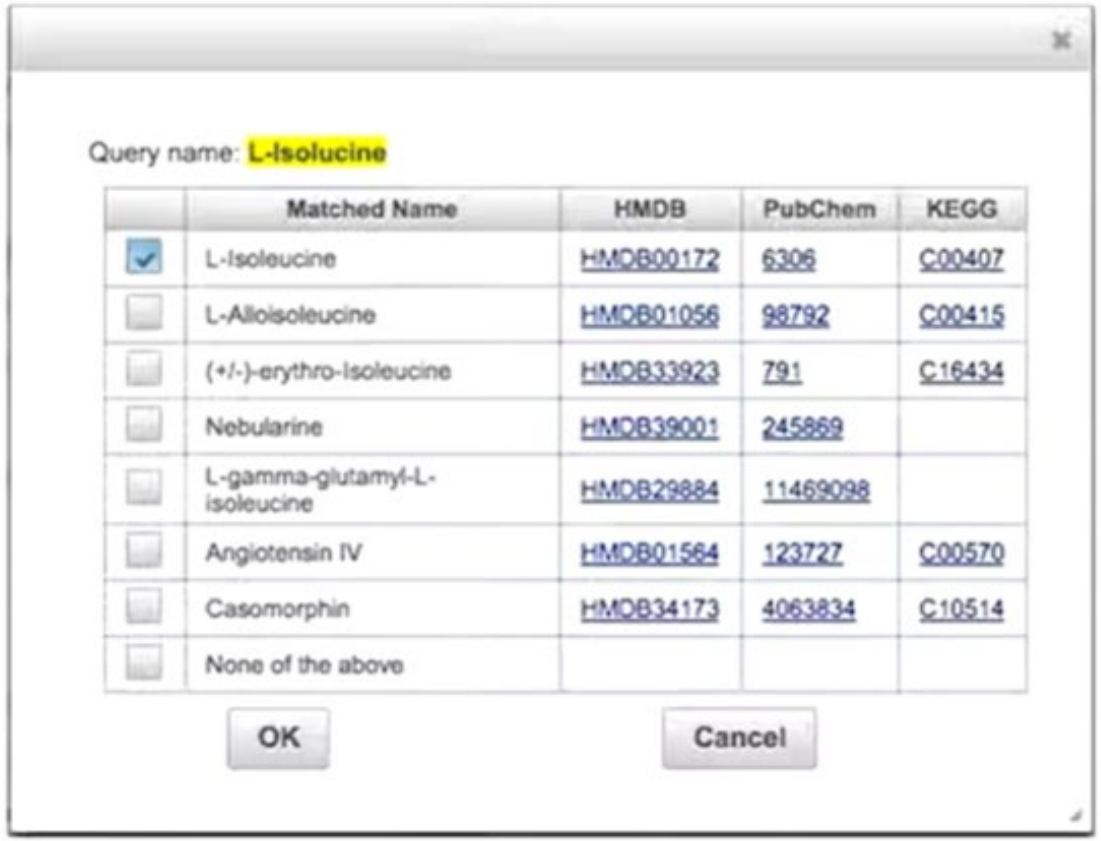


Normally GSEA would require a list of all known genes for the given platform. Here we just use the list of metabolites found in KEGG. ORA is a "weak" analysis in MSEA

Perform Compound Name Standardization



Name Standardization (cont.)

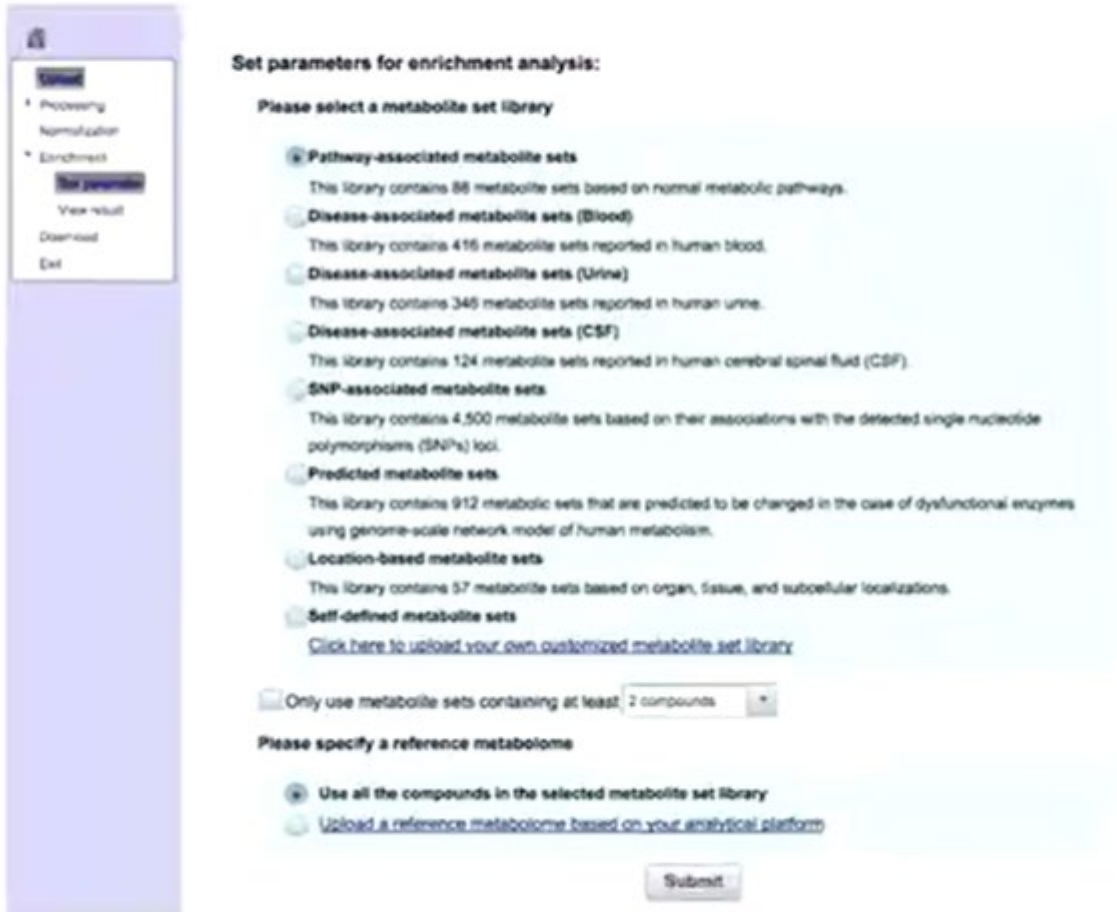


Query name: L-Isolucine

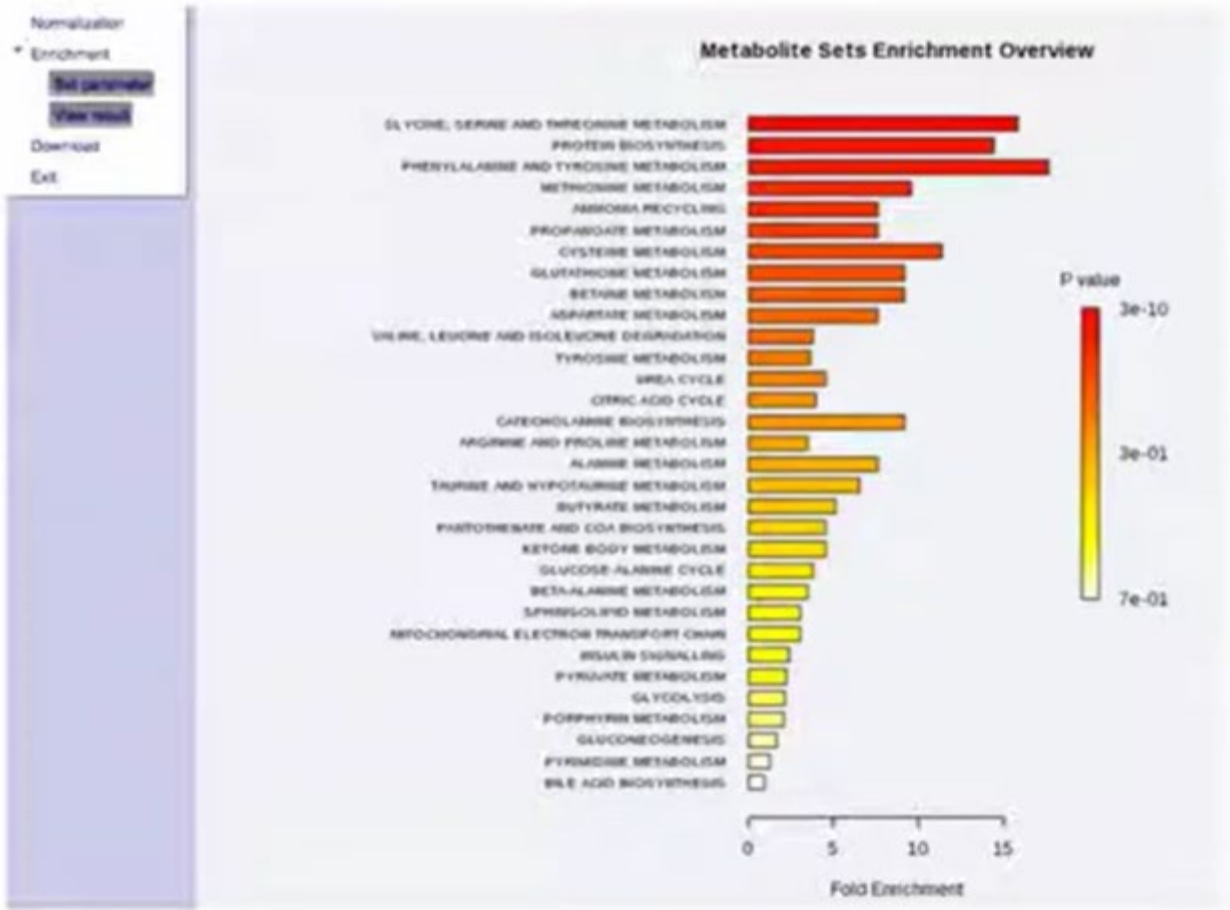
	Matched Name	HMDB	PubChem	KEGG
☑	L-Isoleucine	HMDB00172	6306	C00407
☐	L-Alloisoleucine	HMDB01056	98792	C00415
☐	(+/-)-erythro-Isoleucine	HMDB33923	791	C16434
☐	Nebularine	HMDB39001	245869	
☐	L-gamma-glutamyl-L-isoleucine	HMDB29884	11469098	
☐	Angiotensin IV	HMDB01564	123727	C00570
☐	Casomorphin	HMDB34173	4063834	C10514
☐	None of the above			

OK Cancel

Select a Metabolite Set Library



Result


Normalization
Enrichment
Set parameter
View result
Download
Exit
Metabolite Sets Enrichment Overview
GLYCINE, SERINE AND THREONINE METABOLISM
PROTEIN BIOSYNTHESIS
PHENYLALANINE AND TYROSINE METABOLISM
METHIONINE METABOLISM
AMMONIA RECYCLING
PROPANOATE METABOLISM
CYSTEINE METABOLISM
GLUTATHIONE METABOLISM
BETAINE METABOLISM
ASPARTATE METABOLISM
VALINE, LEUCINE AND ISOLEUCINE DEGRADATION
TYROSINE METABOLISM
UREA CYCLE
CITRIC ACID CYCLE
CATECHOLAMINE BIOSYNTHESIS
ARGININE AND PROLINE METABOLISM
ALANINE METABOLISM
TAURINE AND HYPOTAURINE METABOLISM
BUTYRATE METABOLISM
PANTOTHENATE AND COA BIOSYNTHESIS
KETONE BODY METABOLISM
GLUCOSE-ALANINE CYCLE
BETA-ALANINE METABOLISM
SPHINGOLIPID METABOLISM
MITOCHONDRIAL ELECTRON TRANSPORT CHAIN
INSULIN SIGNALLING
PYRUVATE METABOLISM
GLYCOLYSIS
PORPHYRIN METABOLISM
GLUCONEOGENESIS
PYRIMIDINE METABOLISM
BILE ACID BIOSYNTHESIS
P value
3e-10
3e-01
7e-01
0
5
10
15
Fold Enrichment

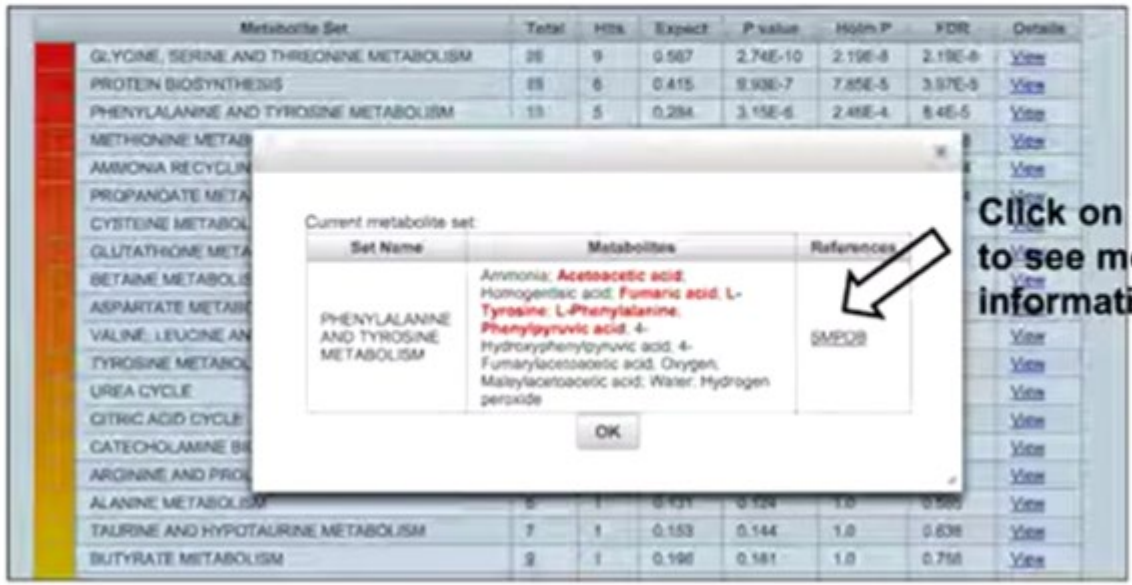
Result (cont.)

Click on details to see more

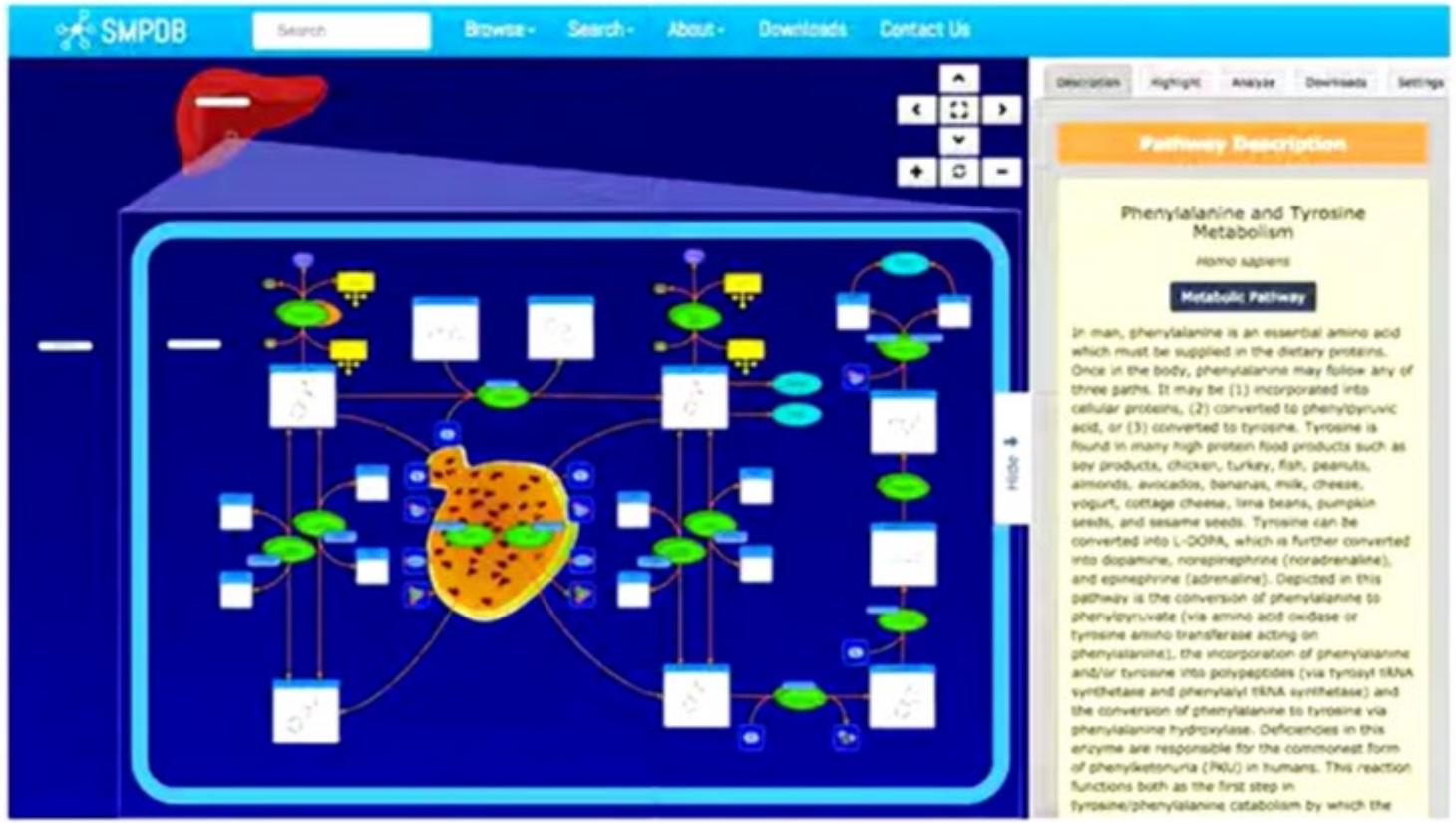
Metabolite Set	Total	Hits	Expect	P value	Holm P	FDR	Details
GLYCINE, SERINE AND THREONINE METABOLISM	20	9	0.567	2.74E-10	2.19E-8	2.19E-8	View
PROTEIN BIOSYNTHESIS	19	8	0.415	9.93E-7	7.85E-5	3.97E-5	View
PHENYLALANINE AND TYROSINE METABOLISM	13	5	0.284	3.15E-6	2.46E-4	8.4E-5	View
METHIONINE METABOLISM	24	5	0.524	8.98E-5	0.00691	0.0018	View
AMMONIA RECYCLING	18	3	0.393	0.00581	0.441	0.0774	View
PROPANOATE METABOLISM	18	3	0.393	0.00581	0.441	0.0774	View
CYSTEINE METABOLISM	8	2	0.175	0.0117	0.883	0.133	View
GLUTATHIONE METABOLISM	10	2	0.218	0.0183	1.0	0.162	View
BETAINE METABOLISM	10	2	0.218	0.0183	1.0	0.162	View
ASPARTATE METABOLISM	13	2	0.362	0.0261	1.0	0.209	View
VALINE, LEUCINE AND ISOLEUCINE DEGRADATION	36	3	0.786	0.0397	1.0	0.288	View
TYROSINE METABOLISM	30	3	0.829	0.0456	1.0	0.304	View
UREA CYCLE	20	2	0.436	0.0677	1.0	0.417	View
CITRIC ACID CYCLE	23	2	0.502	0.0868	1.0	0.496	View
CATECHOLAMINE BIOSYNTHESIS	5	1	0.109	0.105	1.0	0.536	View
ARGININE AND PROLINE METABOLISM	26	2	0.567	0.107	1.0	0.536	View
ALANINE METABOLISM	6	1	0.131	0.124	1.0	0.585	View
TAURINE AND HYPOTAURINE METABOLISM	7	1	0.153	0.144	1.0	0.638	View
BUTYRATE METABOLISM	9	1	0.196	0.181	1.0	0.756	View
PANTOTHENATE AND COA BIOSYNTHESIS	10	1	0.218	0.199	1.0	0.758	View
KETONE BODY METABOLISM	10	1	0.218	0.199	1.0	0.758	View
GLUCOSE-ALANINE CYCLE	12	1	0.262	0.234	1.0	0.851	View
BETA-ALANINE METABOLISM	13	1	0.284	0.251	1.0	0.873	View
SPHINGOLIPID METABOLISM	15	1	0.327	0.284	1.0	0.908	View
MITOCHONDRIAL ELECTRON TRANSPORT CHAIN	15	1	0.327	0.284	1.0	0.908	View
INSULIN SIGNALLING	16	1	0.415	0.345	1.0	1.0	View
PYRUVATE METABOLISM	20	1	0.436	0.36	1.0	1.0	View
GLYCOLYSIS	21	1	0.458	0.374	1.0	1.0	View
PORPHYRIN METABOLISM	22	1	0.49	0.398	1.0	1.0	View
GLUCONEOGENESIS	27	1	0.589	0.454	1.0	1.0	View
PYRIMIDINE METABOLISM	36	1	0.785	0.556	1.0	1.0	View
BILE ACID BIOSYNTHESIS	49	1	1.07	0.672	1.0	1.0	View

Submit

The Matched Metabolite Set



Phenylalanine and Tyrosine Metabolism in SMPDB



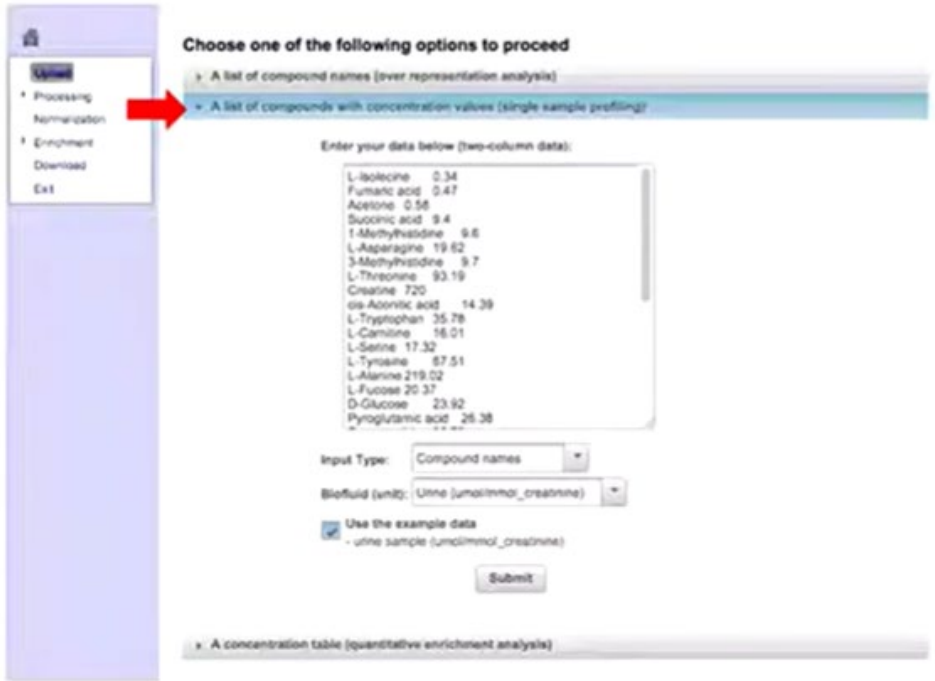
so over-representation analysis is actually a veryweak form of doing pathway analysis

it's just giving you as an exampleit's arguably the easiest one to do because

all you need to do isprovide a simple list of differentially changed metabolites

Single Sample Profiling (SSP)

(Basically used by a physician to analyze a patient)



Single sample profiling is basically what you would do if you were analyzing a single individual and **you're trying to determine whether they're sick or healthy**

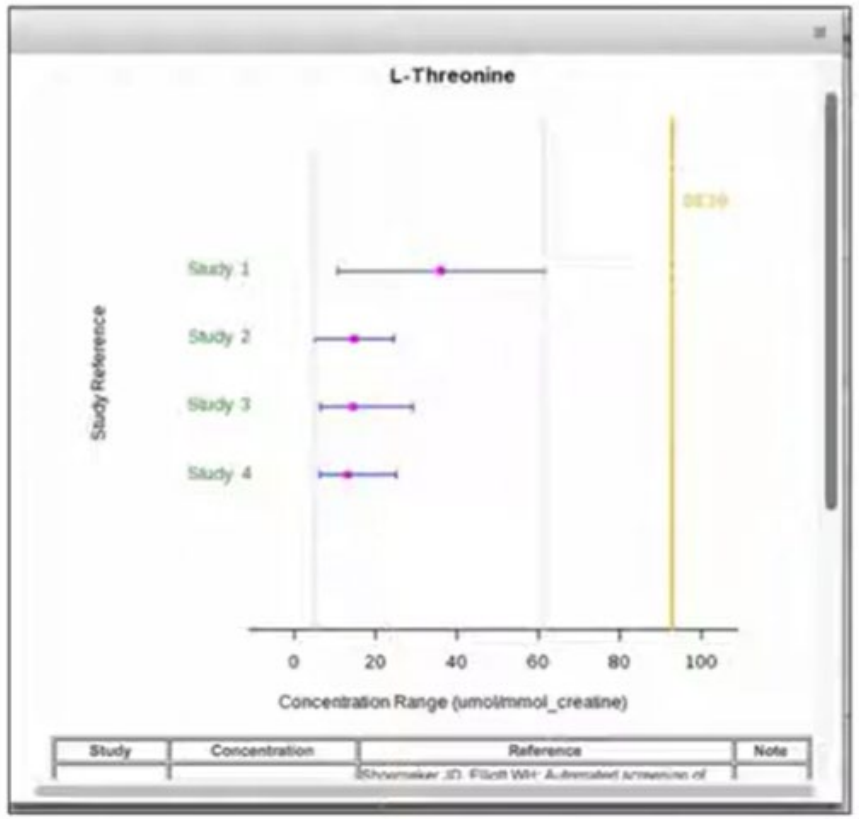
You might be taking out a profile of an individual and here instead of just providing a list of a town length to actuallyproviding a list of metabolites

You have to indicate whether it's blood or serum or urine or wherever but

As long as it knows what type of bio fluid is working with and as long as it knows that it's human because this is what it's goingto be referencing

Concentration Comparison

Concentration Comparison (cont.)



it will do essentially a concentration comparison so for each of the compounds that are in there it will provide a list of from the known or relative concentration ranges that are expected to be normal and how far whether things are significantly off and abnormal or or normal you can then further view these things and

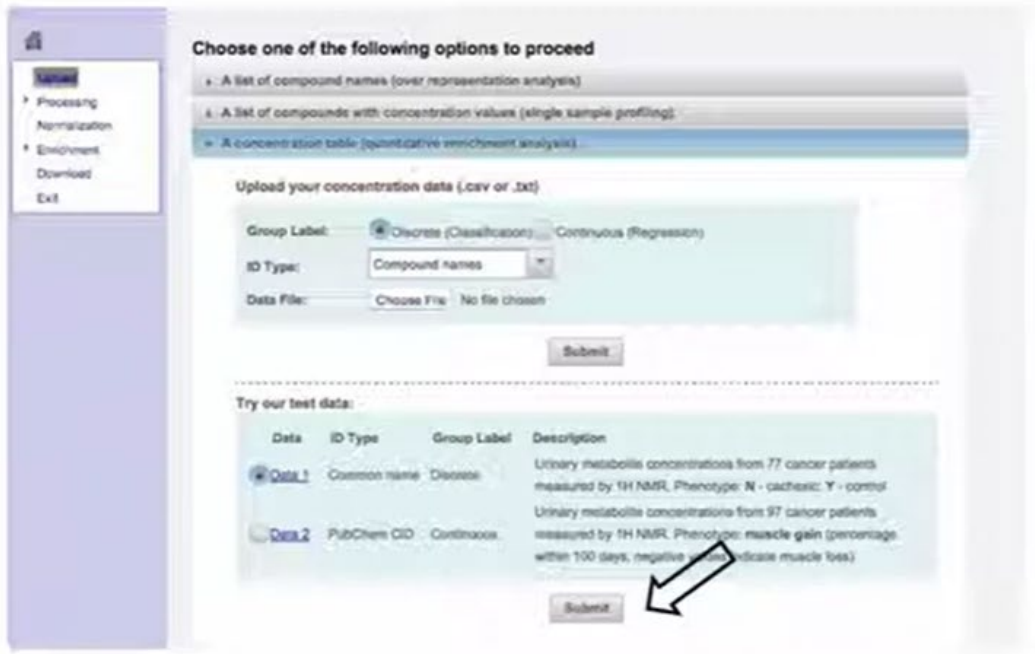
where we're looking at Concentration Comparison (cont.) levels for threonine

in this particular individual they're read out inurine with something like 90 micro molar and

the information that's beencollected on humans in urine over last number of years there's four studies that were published all the values are typically reported orsomewhere between 20 and 40 micro molar

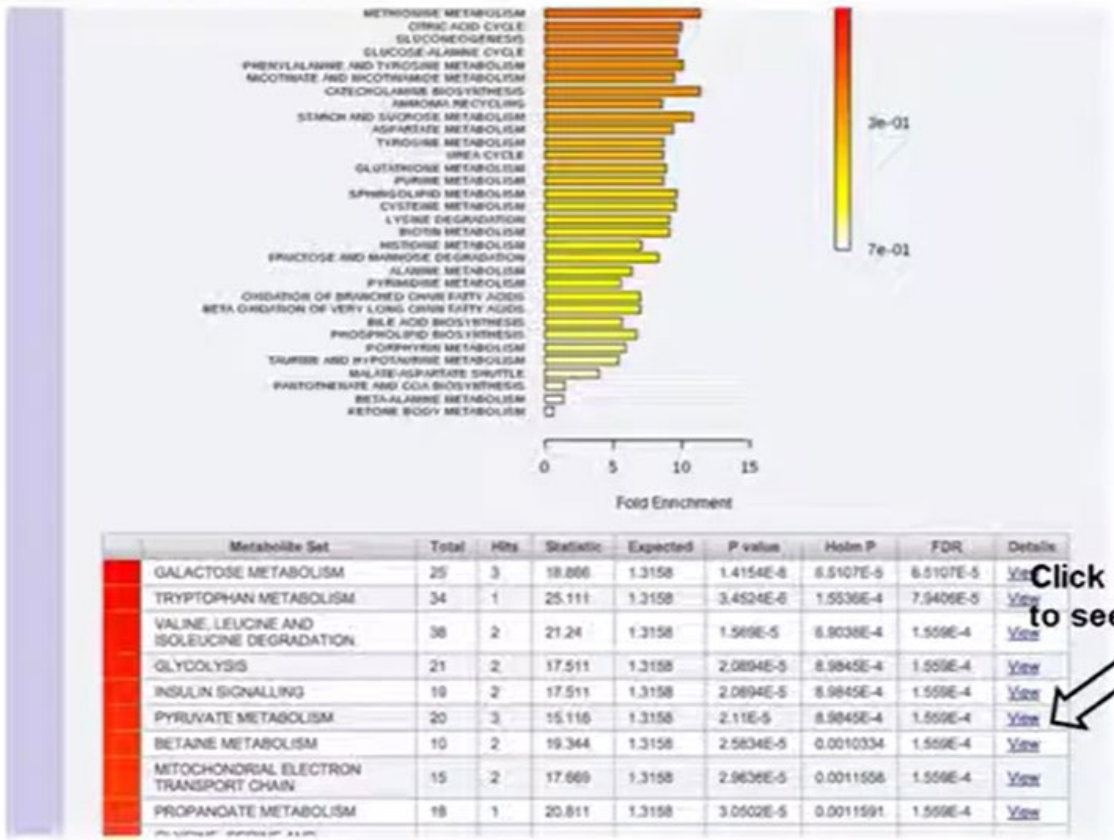
so this individual clearly has abnormally high values

Quantitative Enrichment Analysis (QEA)



now you're looking at a population with a whole bunch of metabolite concentrations and in this case we'll be indicating names andconcentrations this case got a sample data step you can upload

Result

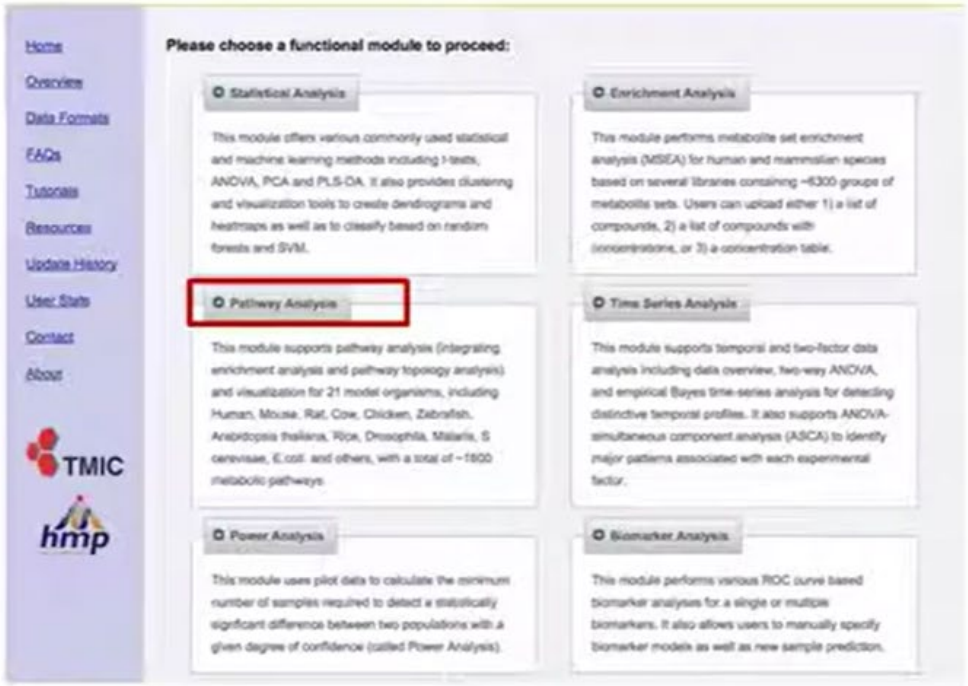


Metabolite Set	Total	Hits	Statistic	Expected	P value	Holm P	FDR	Details
GALACTOSE METABOLISM	25	3	18.886	1.3158	1.4154E-6	6.5107E-5	6.5107E-5	View
TRYPTOPHAN METABOLISM	34	1	25.111	1.3158	3.4524E-6	1.5536E-4	7.9406E-5	View
VALINE, LEUCINE AND ISOLEUCINE DEGRADATION	38	2	21.24	1.3158	1.569E-5	6.9038E-4	1.559E-4	View
GLYCOLYSIS	21	2	17.511	1.3158	2.0894E-5	8.9845E-4	1.559E-4	View
INSULIN SIGNALLING	19	2	17.511	1.3158	2.0894E-5	8.9845E-4	1.559E-4	View
PYRUVATE METABOLISM	20	3	15.116	1.3158	2.11E-5	8.9845E-4	1.559E-4	View
BETAINE METABOLISM	10	2	19.344	1.3158	2.5834E-5	0.0010334	1.559E-4	View
MITOCHONDRIAL ELECTRON TRANSPORT CHAIN	15	2	17.669	1.3158	2.9636E-5	0.0011556	1.559E-4	View
PROPANOATE METABOLISM	18	1	20.811	1.3158	3.0502E-5	0.0011591	1.559E-4	View

Click on details to see more

The Matched Metabolite Set

Select a Module (Pathway Analysis)



While enrichment analysis tells you which diseases pathways other things are are modified

this one is largely more focused on pathways but

it looks not just at the Association ofmetabolites in pathways but also

It considers the pathway structureswhether metabolites that are being changed or representative of hubs orspokes in the pathway whether they play a central role and

rather thanjust being restricted only to humans as the over-representation or MSC ais this allows you to look at things from 21 model organisms

Pathway Analysis

- **Purpose: to extend and enhance metabolite set enrichment analysis for pathways by**
  - **Considering pathway structures**
  - **Supporting pathway visualization**
- **Currently supports analysis for 21 diverse (model) organisms such as humans, mouse, Drosophila, Arabidopsis, E. coli, yeast, etc. (KEGG pathways only)**

While enrichment analysis tells you which diseases pathways other things are are modified

this one is largely more focused on pathways but

it looks not just at the Association ofmetabolites in pathways but also

It considers the pathway structureswhether metabolites that are being changed or representative of hubs orspokes in the pathway whether they play a central role and

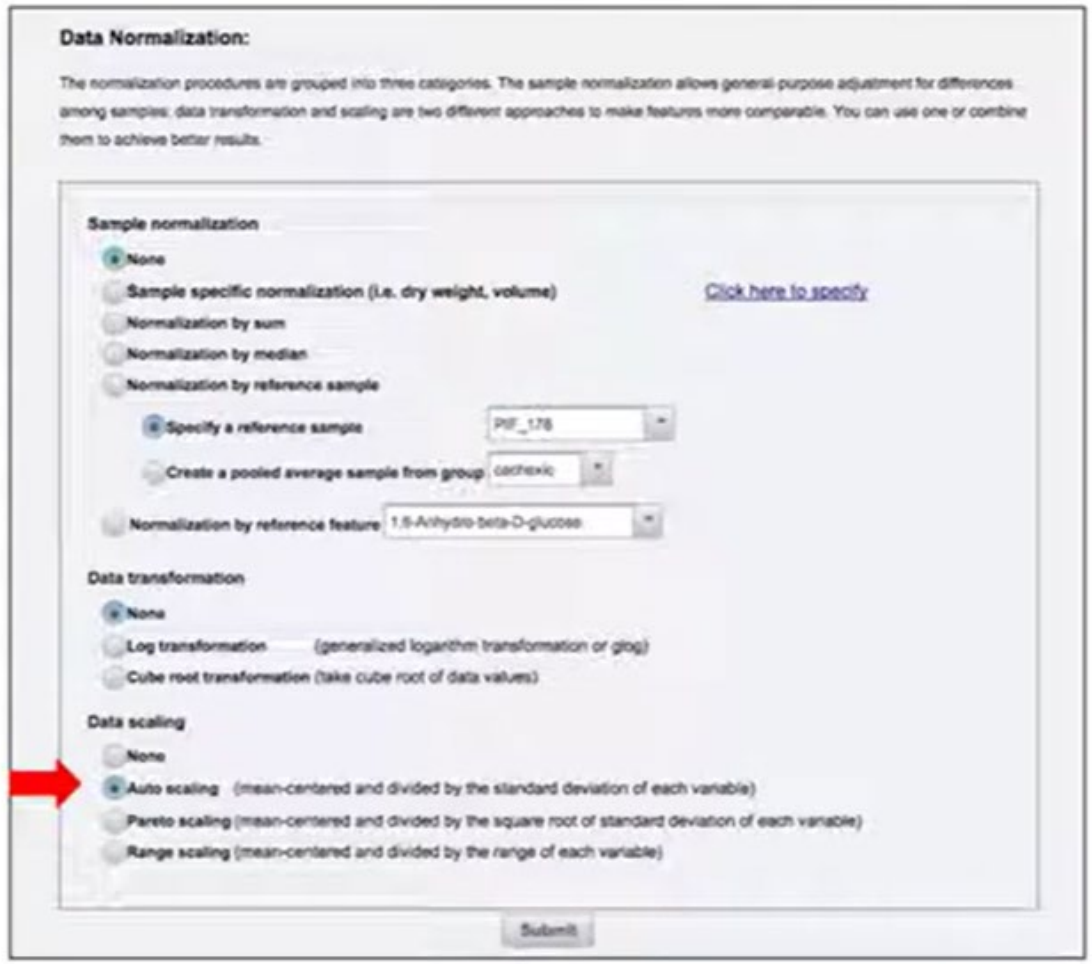
rather thanjust being restricted only to humans as the over-representation or MSC ais this allows you to look at things from 21 model organisms

Data Set Selected

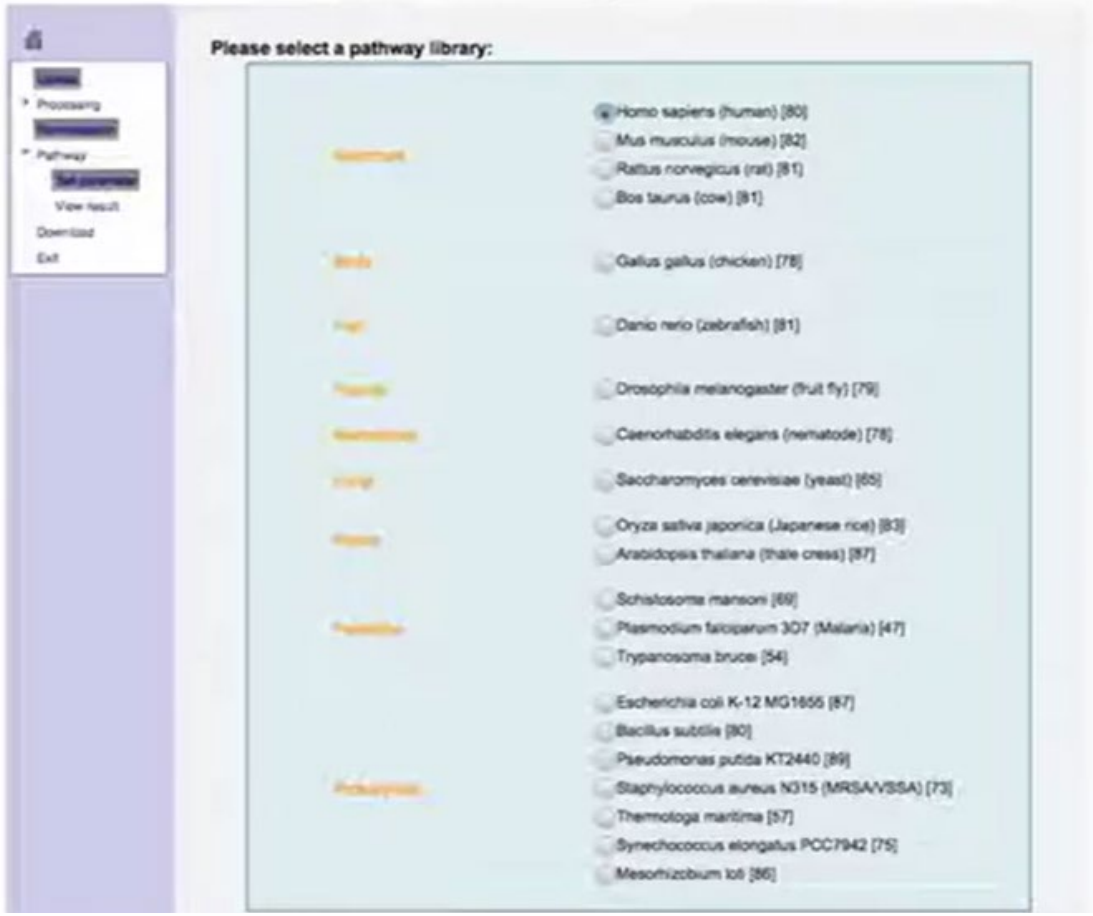
- Here we are using a collection of metabolites identified by NMR (compound list + concentrations) from the urine from 77 lung and colon cancer patients, some of whom were suffering from cachexia (muscle wasting)


Please enter a one-column compound list:
Input Type:
-- Please specify
Use our example data
Submit
Or upload a concentration table (.csv or .txt):
Group Label:
Discrete (Classification)
Continuous (Regression)
ID Type:
-- Please specify
Data File:
No file chosen
Use the example data
Data
Description
Dataset
Urinary metabolite concentrations from 77 cancer patients measured by 1H NMR. Phenotype:

Perform Data Normalization



Select Pathway Libraries



Perform Network Topology Analysis

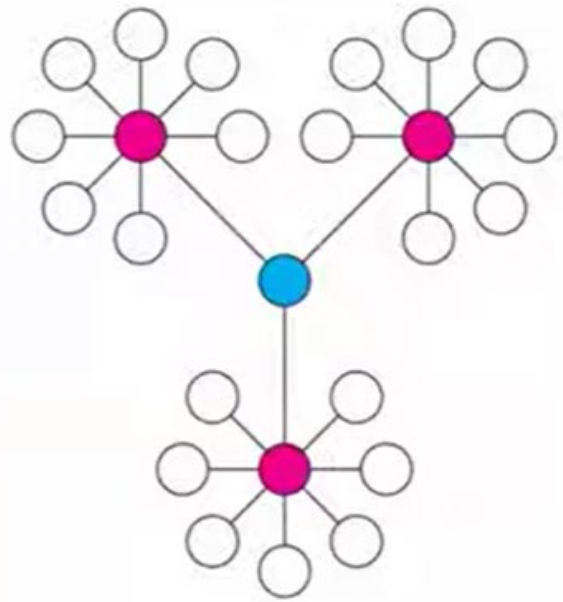
Pathway Position Matters

Which positions are important?

- Hubs
  - Nodes that are highly connected (red ones)
- Bottlenecks
  - Nodes on many shortest paths between other nodes (blue ones)

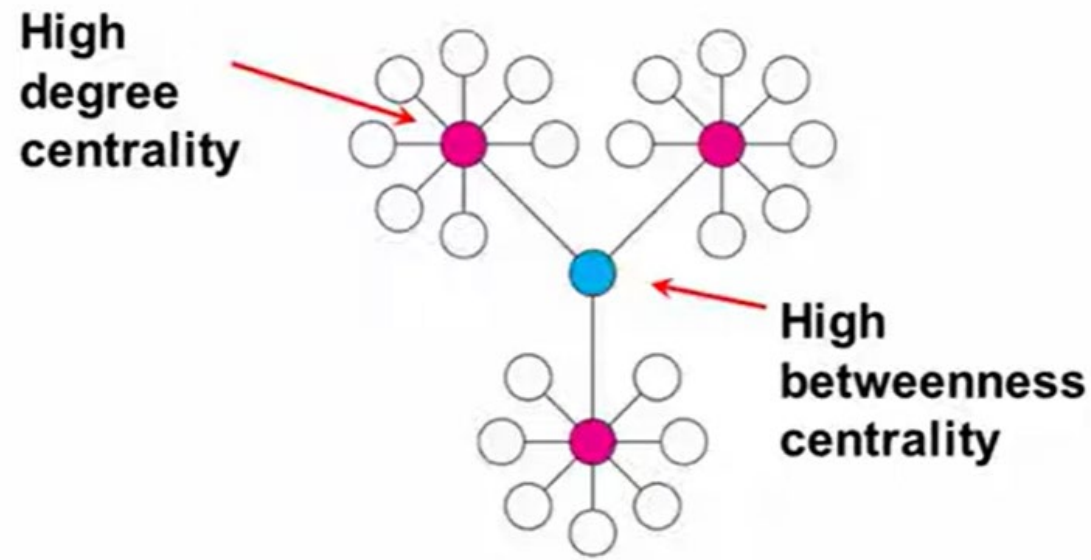
Graph theory

- Degree centrality
- Betweenness centrality

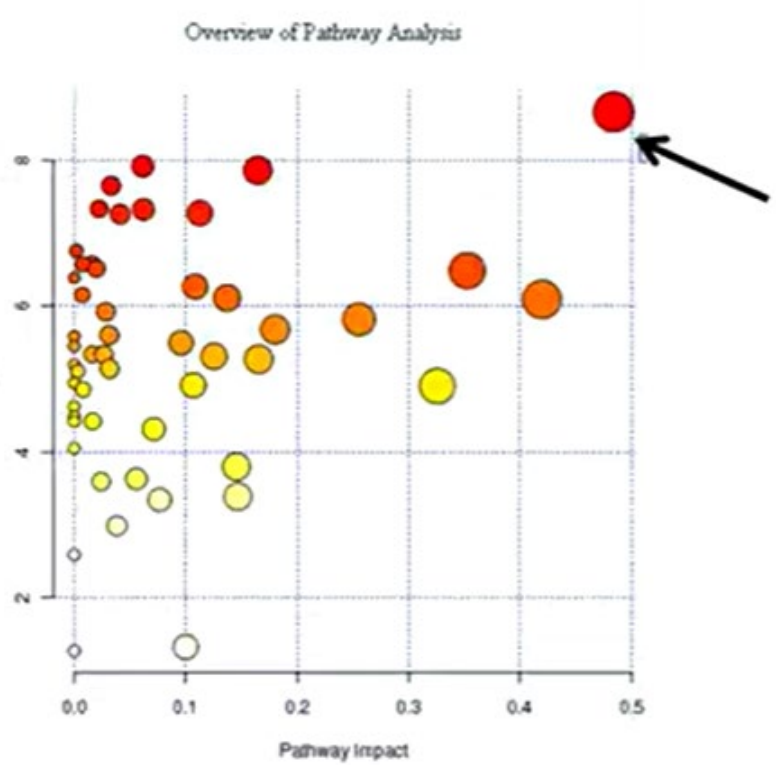


Junker *et al. BMC Bioinformatics* 2006

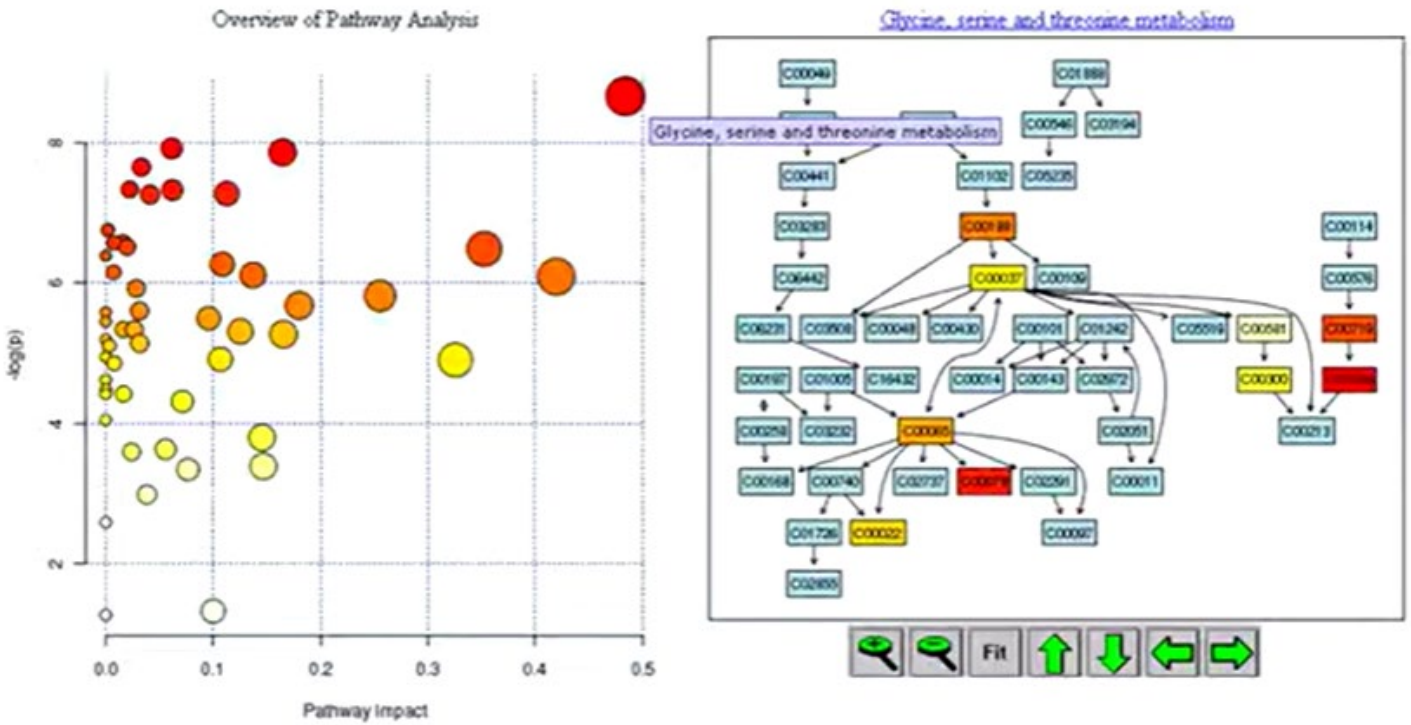
Which Node is More Important?



Pathway Visualization



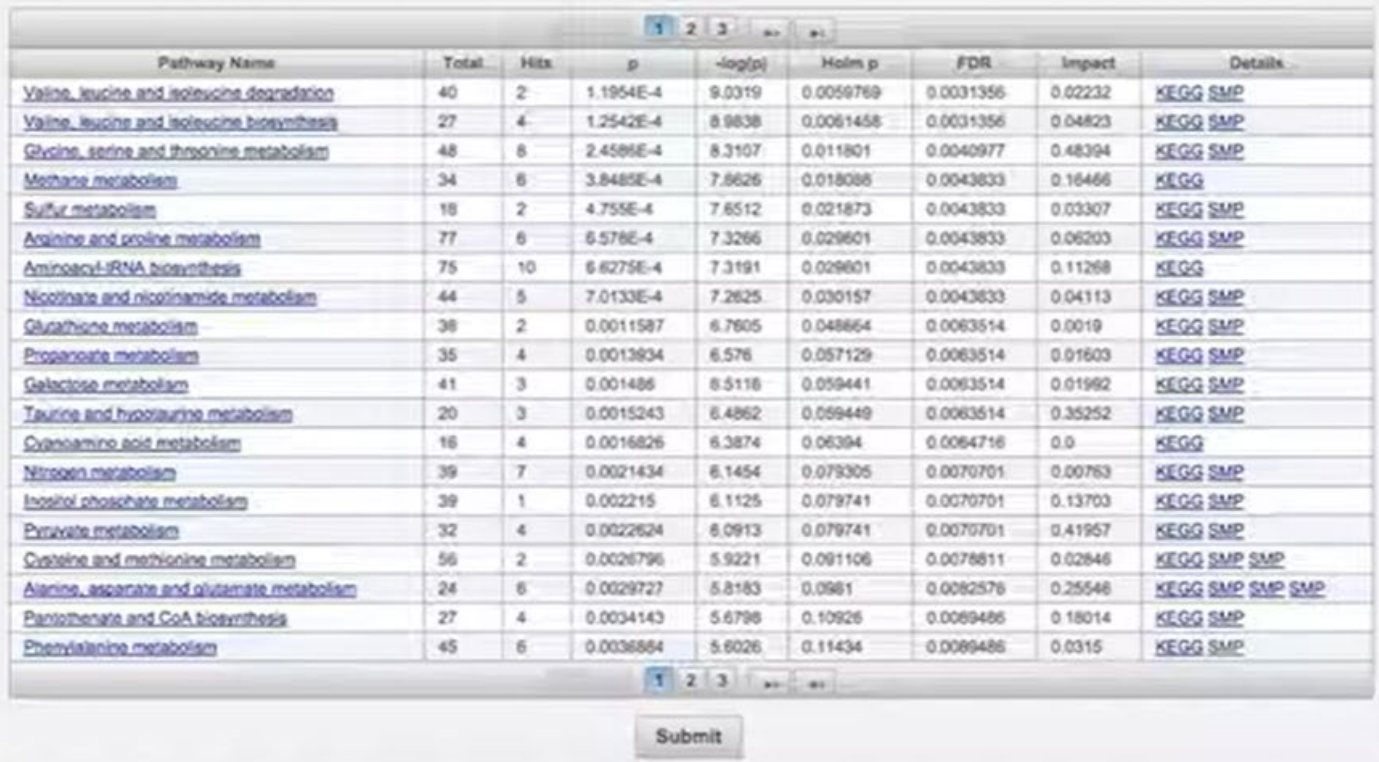
Pathway Visualization



Pathway Impact

- Incorporates parameters such as the log fold-change of the DE metabolites, the statistical significance of the set of pathway genes and the topology of the signaling pathway
- Combines the pathway topology with the over-representation evidence

Result



1 2 3 ▸▸ ▸|

Pathway Name	Total	Hits	p	-log(p)	Holm p	FDR	Impact	Details
Valine, leucine and isoleucine degradation	40	2	1.1954E-4	9.0319	0.0059769	0.0031356	0.02232	KEGG SMP
Valine, leucine and isoleucine biosynthesis	27	4	1.2542E-4	8.9838	0.0061458	0.0031356	0.04823	KEGG SMP
Glycine, serine and threonine metabolism	48	6	2.4586E-4	8.3107	0.011801	0.0040977	0.48394	KEGG SMP
Methane metabolism	34	6	3.8485E-4	7.8626	0.018086	0.0043833	0.16466	KEGG
Sulfur metabolism	18	2	4.755E-4	7.6512	0.021873	0.0043833	0.03307	KEGG SMP
Arginine and proline metabolism	77	6	6.578E-4	7.3266	0.029601	0.0043833	0.06203	KEGG SMP
Aminoacyl-tRNA biosynthesis	75	10	6.6275E-4	7.3191	0.029601	0.0043833	0.11268	KEGG
Nicotinate and nicotinamide metabolism	44	5	7.0133E-4	7.2625	0.030157	0.0043833	0.04113	KEGG SMP
Glutathione metabolism	38	2	0.0011587	6.7605	0.048664	0.0063514	0.0019	KEGG SMP
Propanoate metabolism	35	4	0.0013934	6.576	0.057129	0.0063514	0.01603	KEGG SMP
Galactose metabolism	41	3	0.001486	6.5118	0.059441	0.0063514	0.01992	KEGG SMP
Taurine and hypotaurine metabolism	20	3	0.0015243	6.4862	0.059449	0.0063514	0.35252	KEGG SMP
Cyanoamino acid metabolism	16	4	0.0016826	6.3874	0.06394	0.0064716	0.0	KEGG
Nitrogen metabolism	39	7	0.0021434	6.1454	0.079305	0.0070701	0.00763	KEGG SMP
Inositol phosphate metabolism	39	1	0.002215	6.1125	0.079741	0.0070701	0.13703	KEGG SMP
Pyruvate metabolism	32	4	0.0022624	6.0913	0.079741	0.0070701	0.41957	KEGG SMP
Cysteine and methionine metabolism	56	2	0.0026796	5.9221	0.091106	0.0078811	0.02846	KEGG SMP SMP
Alanine, aspartate and glutamate metabolism	24	6	0.0029727	5.8183	0.0981	0.0082576	0.25546	KEGG SMP SMP SMP
Pantothenate and CoA biosynthesis	27	4	0.0034143	5.6796	0.10926	0.0089486	0.18014	KEGG SMP
Phenylalanine metabolism	45	6	0.0036884	5.6026	0.11434	0.0089486	0.0315	KEGG SMP

1 2 3 ▸▸ ▸|

Submit