RCMI Workshop, July 29-30, 2010

Bioinformatics Tools for the Analysis of Metabolomic Data

Introduction

Metabolomics is a rapidly evolving field that holds promise to provide insights into genotype – phenotype relationships in diabetes, cancers, and other complex diseases. Due to progress in analytical instrumentation, it is now possible to obtain quantitative data on large number of metabolites under a variety of experimental conditions and track time and/or condition-dependent changes in metabolite-related processes. One of the major informatics tasks is mapping metabolite data onto relevant metabolic networks and linking them with other types of high throughput molecular data (e.g. proteomics, expression profiling).

Approximately 150 million people worldwide have type 2 diabetes (T2D) and this number is predicted to double to 300 million by 2025. The pathogenesis of T2D is complex and involves the interaction of genetic and environmental factors, although the precise manner in which the genetic, environmental, and pathophysiologic factors interact to lead to the clinical onset of T2D is not known.

Db/db diabetic mouse is us one if the most extensively studied animal models of T2D. These mice lack a functioning leptin receptor resulting in defective leptin-mediated signal transduction, which causes chronic overeating, obesity, insulin resistance, hyperglycemia, dyslipidemia, and other manifestations including kidney disease and fibrosis. The paper published in 2008 by Connor and coauthors (1) describes the results of the NMR-based unbiased metabolomics study that characterized the differences in metabolites between urine samples from the diabetic db/db and control db/+ mice.

In this tutorial we will use two publically available bioinformatics tools Metscape 1.0 (Karnovsky et al., 2009, Gao et al., 2010) and Pathway Tools/Omics Viewer (Krieger et al., 2004) to analyze the data and compare the results.

Metabolomics Data

Comparison of db/db *versus* db/+ NMR data showed that approximately two thirds of the spectrum showed discernable changes in the urinary metabolite profile related to disease effects. Following statistical analysis (ANOVA, PCA, PLS-DA) authors were able to identify 66 metabolites, including several that were previously confirmed as arising from gut microflora (Table 1).

Table 1. Summary of polar metabolites differentiating between db/db and db/+ mice (from Connor et al., 2010)

Metabolite	CAS number	Direction of change	MFC D11	*p* value	Assignment confidence	Pathway information
N-Acetylaspartate	997-55-7	DOWN	1.07	0.0116611	2	Alanine and aspartate metabolism
Mucicacid	526-99-8	UP	2.34	0.0000000	3	Ascorbate and aldarate metabolism
Phosphocreatine	67-07-2	UP	2.34	0.000000	2	Creatine biosynthesis and amino acid metabolism, glycine,

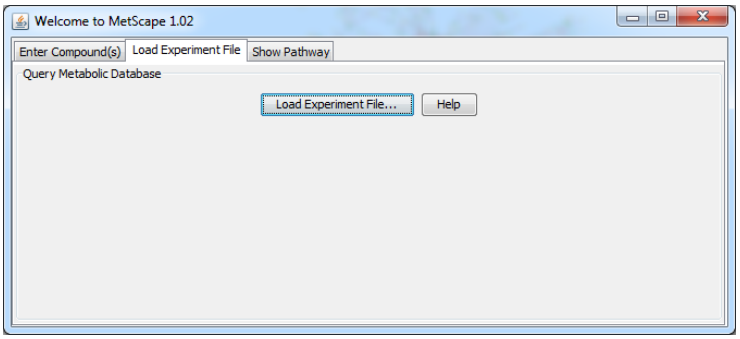
Metabolite	CAS number	Direction of change	MFC D11	*p* value	Assignment confidence	Pathway information
Creatinine	60-27-5	UP	1.24	0.0011194	1	serine and threonine metabolism
Glycine	56-40-6	UP	1.88	0.000000	1	
Guanidinoacetate	352-97-6	UP	1.50	0.0480000	1	
Butyrylglycine	107-92-6	DOWN	1.55	0.000000	1	Fatty acid metabolism
Malonate	141-82-2	UP	1.39	0.0000000	3	
Carnitine	541-15-1	UP	1.24	0.0003512	3	
2-Hydroxy-*N*-valerate	109-52-4	DOWN	1.17	0.0000002	3	
Caproylglycine	142-62-1	DOWN	1.70	0.0000000	1	
N-Acetylglutamate	1188-37-0	UP	1.27	0.000000	2	Glutamate pathway (link with urea cycle)
Choline	62-49-7	UP	1.28	0.0001503	2	Glycine, serine and threonine metabolism
Threonine	72-19-5	UP	1.70	0.0000000	4	
Valerylglycine	109-52-4	DOWN	1.29	0.0000010	1	
Glucose	50-99-7	UP	102.97	0.000000	1	Glycolysis
Lactate	50-21-5	UP	1.89	0.000000	1	
Alanine	56-41-7	UP	1.29	0.0000179	1	Glycolysis, ala and asp metabolism
Acetate	71-50-1	UP	1.24	0.0125364	1	Glycolysis, fatty acid β-oxidation
Formate	64-18-6	UP	1.86	0.0007927	1	Glyoxylate and dicarboxylate
2-Oxoadipate	3184-35-8	UP	1.29	0.0000004	2	Lysine degradation
Lysine	56-87-1	UP	1.10	0.0490000	4	Lysine degradation, aminoacyl tRNA biosynthesis
Glutaric acid	110-94-1	DOWN	1.29	0.0000003	2	Lysine degradation, fatty acid metabolism
Methionine	63-68-3	UP	1.24	0.0009814	1	Methionine metabolism
N1-Methylnicotinamide	3106-60-3	UP	1.86	0.000000	1	Nicotinate, nicotinamide metabolism
N1-Methylnicotinic acid	535-83-1	UP	1.60	0.0000967	1	
Nicotinamide-*n*-oxide	1986-81-8	UP	1.54	0.0031113	2	
N-Methyl-2-pyridone-5-carboxamide	701-44-0	UP	1.88	0.000000	1	
N-Methyl-4-pyridone-3-carboxamide	3128-29-8	UP	2.99	0.0000016	1	
Pantothenate	79-83-4	UP	2.30	0.0000340	3	Pantothenate and coA biosynthesis, β-alanine metabolism
3-Ureidopropanoate	462-88-4	UP	1.23	0.0000495	1	Purine metabolism
Allantoin	97-59-6	UP	1.14	0.0490000	1	
Orotate	65-86-1	UP	1.24	0.0440571	1	Pyrimidine
Sucrose	57-50-1	UP	2.55	0.000000	1	Starch and glucose
Isocaproylglycine	646-07-1	DOWN	1.23	0.0024995	1	Steroid hormone production
Taurine	107-35-7	DOWN	1.50	0.0490000	1	Taurine and hypotaurine
(s)-Malate	97.67.6	UP	2.15	0.0000000	1	tca
2-Oxoglutarate	328-50-7	UP	2.45	0.0000024	1	
cis-Aconitate	585-84-2	DOWN	1.01	0.0490000	1	
Citrate	77-92-9	UP	3.84	0.0000017	1	
Fumarate	110-17-8	UP	3.20	0.000000	1	
Succinate	110-15-6	UP	1.31	0.0076474	1	
Indoxyl sulfate	487-94-5	DOWN	1.06	0.0490000	2	Tryptophan metabolism
Tyrosine	60-18-4	UP	1.36	0.0019804	2	Tyrosine metabolism
Citrulline	627-77-0	UP	1.46	0.0000000	3	Urea cycle
L-Argininosuccinic acid	C034-06	UP	1.70	0.0000000	3	
N-Acetyl citrulline	627-77-0	UP	1.22	0.0000357	4	
Ornithine	70-26-8	UP	1.30	0.0000000	3	
2-Hydroxy isobutyrate	594-61-6	UP	1.21	0.0000031	2	val, leu and ileu degradation
2-Hydroxyisovalerate	498-36-2	UP	2.30	0.0000340	3	
2-Oxo-3-methyl-*N*-valerate	1460-34-0	DOWN	1.23	0.0000216	2	
2-Oxoisocaproate	328-50-7	DOWN	1.16	0.0208412	2	
2-Oxoisovalerate	759-05-7	DOWN	1.22	0.0002984	2	
Isobutyrlglycine	79-31-2	DOWN	1.29	0.0000010	1	

Metabolite	CAS number	Direction of change	MFC D11	*p* value	Assignment confidence	Pathway information
Isovalerate	503-74-2	DOWN	1.55	0.000000	3	
Isovalerylglycine	503-74-2	DOWN	1.16	0.0000146	1	
Methylmalonate	516-05-2	DOWN	1.15	0.0000251	4	
Valine	72-18-4	UP	1.20	0.0490000	4	

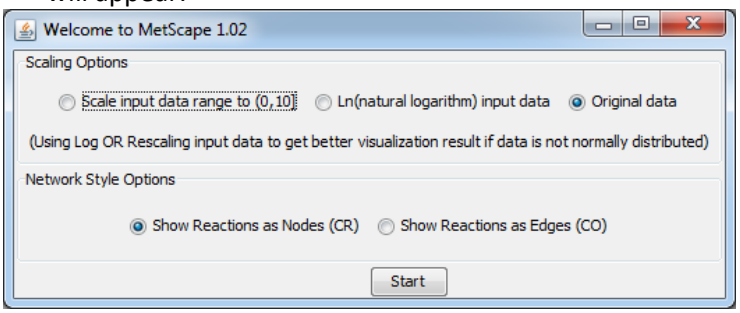
We will be viewing the differentiating metabolites in the context metabolic pathways.

Loading data into Metscape 1.0

1. Launch Cytoscape by double clicking the Icon on the desktop.
2. On the top menu, go to Plugins → Metscape plugin → Build Compound Reaction Network. A query dialogue box appears with three tabs on top. Each tab lets you conduct a different type of query.
3. Select the second tab - "Load Experiment File".



4. Click on the Load Experiment File button. A dialog box will appear. Select the data file class/Desktop/Metabolomics/Diff_metabolites_Metscape. Click open. The following dialog box will appear:



Select "scale input data range to (0, 10)". Leave "Show Reactions as Nodes" button selected. We will explore the other option later.

5. We will now explore the network that consists of the experimental compounds. You will notice that the compounds form several separate subnetworks.
6. At this point, you are also able to view experimentally measured compound in the context of known biochemical reactions and pathways.

Please refer to Metscape Plugin User Guide for specific plugin features. The manual also contains detailed instructions for downloading and installing Cytoscape and the Metscape plugin on your computer.

Loading data into the OMICS viewer

1. Open the web browser and go to http://biocyc.org/expression.html
2. Select "Homo sapiens" from the organism drop-down list on the upper right
3. Click "Browse" button and select the following file: class/Desktop/Metabolomics/Diff_metabolites_Omics_viewer
4. Select "single data column" and "0-centered scale"
5. Select "Compound names and/or identifiers from the drop-down list
6. Enter "1" into the "Data column (numerator in ratios)" field
7. We will leave the default values for the rest of the parameters.

For detailed description of the Omics Viewer features please go to

http://biocyc.org/ov-expr.shtml or watch a webinar at http://biocyc.org/webinar.shtml

References

Connor SC, Hansen MK, Corner A, Smith RF, Ryan TE. Integration of metabolomics and transcriptomics data to aid biomarker discovery in type 2 diabetes. Mol Biosyst. 2010 May 10;6(5):909-21.

Gao J, Tarcea VG, Karnovsky A, Mirel BR, Weymouth TE, Beecher CW, Cavalcoli JD, Athey BD, Omenn GS, Burant CF, Jagadish HV.(2010). Metscape: a Cytoscape plug-in for visualizing and interpreting metabolomic data in the context of human metabolic networks. Bioinformatics. Apr 1;26(7):971-3.

Karnovsky A**,** J. Gao, V.G. Tarcea, C. Beecher, C.F. Burant, B. Mirel, H.V. Jagadish, G.S. Omenn (2009). Bioinformatics Framework for the Analysis and Interpretation of Metabolomic Data. AMIA Translational Bioinformatics Summit.

Krieger CJ, Zhang P, Mueller LA, Wang A, Paley S, Arnaud M, Pick J, Rhee SY, Karp PD. (2004). MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res. Jan 1;32(Database issue):D438-42.