

NCIBI-RCMI Workshop Translational Bioinformatics

Bioinformatics Tools for the Analysis of Metabolomics Data

Alla Karnovsky, Ph.D.

akarnovs@umich.edu

July 29-30, 2010

University of Michigan

Ann Arbor, Michigan

NCIBI Metabolomics Data Analysis Tools

- **ConceptMetab** - a prototype tool that allows users to examine the relationships among predefined and user - specified compound sets from derived from high-throughput, unbiased metabolomics data or from the literature
- **Metab2MeSH** - a tool that uses statistical approach to reliably and automatically annotate compounds with the concepts defined in MeSH, the National Library of Medicine's controlled vocabulary for biomedical concepts
- **GenePattern Correlation Matrix Module** - a module for calculating pairwise correlations between genes and metabolites
- **MetScape** - Cytoscape plugin for visualization and analysis of metabolomics data

Acknowledgment

NCIBI/CCMB

Terry Weymouth
Glenn Tarcea
Tim Hull
Jing Gao
Alex Ade
Zach Wright
Vasu Mahavisno
Maureen Sartor
Beth Kirschner
Barbara Mirel
Jim Cavalcoli
Youngsheng Bai
Rich Mc Eachin
Paul Trombley
Alex Terzian
H.V. Jagadish
Gilbert Omenn
Brian Athey

Burant Lab:

Chuck Burant
Jinghua Xu
Mary Kay Treutelaar

Kennedy Lab:

Bob Kennedy
Charles Evans
Matt Lorenz

Chris Beecher

Broad Institute

Michael Reich

National Institutes of Health: Grant #U54 DA021519

Michigan Diabetes Research and Training Center

ConceptMetab

ConceptMetab - a prototype tool that allows users to examine the relationships among predefined and user - specified compound sets from derived from high-throughput, unbiased metabolomics data or from the literature

Biological Knowledge Type	# of Concepts	Source
Biological processes	1577	Gene Ontology
Molecular functions	539	Gene Ontology
Cellular components	234	Gene Ontology
Enzyme	319	KEGG
KEGG Metabolic Pathways	74	KEGG

See ConceptMetab poster for details:

ConceptMetab: A New Tool for Testing and Visualizing Relationships Among Sets of Metabolites. *Vasudeva Mahavisno, Glenn Tarcea, Terry Weymouth, Alla Karnovsky, and Maureen A. Sartor*

ConceptMetab

Compound Concept Explorer X

Prostate_cancer

Queried Concept Name: Prostate_cancer
 Concept Type: Experimental Com...
 Compound List Size: 36
 Action:

Kegg ID	Compound Name
C01239	N-Acetyl-beta-D-glucosaminyamine
C01132	N-Acetyl-D-chondrosamine
C00475	Cytidine
C00711	Mallic acid

Select None Select All

Enriched Concepts [Filter Concepts](#)

Selected 0 of 63 Concepts |

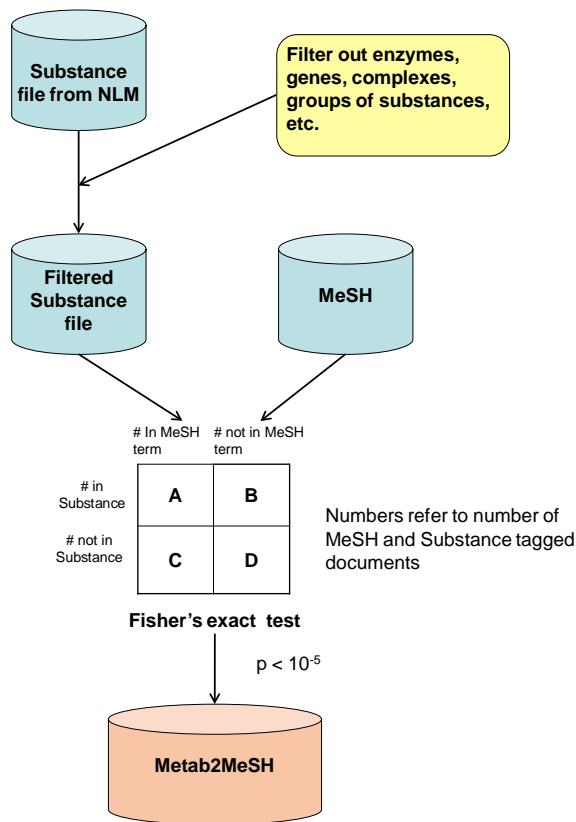
Concept Name	Concept Type Name	Category	Compound List Size	Overlap	P-value	Q-value
<input type="checkbox"/> hydrolase activity, acting on carbon-nitrogen (but not pep	GO Molecular Functi		94	12	1.399493E-6	7.543265E-4
<input type="checkbox"/> amino acid catabolic process	GO Biological Proce		111	10	3.834005E-4	6.046226E-1
<input type="checkbox"/> deaminase activity	GO Molecular Functi		20	5	6.678087E-4	1.799744E-1
<input type="checkbox"/> amine catabolic process	GO Biological Proce		127	10	1.0866E-3	7.641278E-1
<input type="checkbox"/> nitrogen compound catabolic process	GO Biological Proce		132	10	1.453636E-3	7.641278E-1
<input type="checkbox"/> cellular catabolic process	GO Biological Proce		319	15	3.619891E-3	1E0
<input type="checkbox"/> cellular component assembly	GO Biological Proce		96	8	3.943639E-3	1E0
<input type="checkbox"/> hydrolase activity, acting on carbon-nitrogen (but not pep	GO Molecular Functi		50	6	4.032763E-3	7.245532E-1
<input type="checkbox"/> catabolic process	GO Biological Proce		326	15	4.547932E-3	1E0

Draw complete Interactions [Draw Network Graph](#) [Draw Heatmap](#) [Back to Search](#)

Metab2MeSH

Metab2MeSH uses statistical approach to reliably and automatically annotate compounds with the concepts defined in MeSH, the National Library of Medicine's controlled vocabulary for biomedical concepts

<http://metab2mesh.ncibi.org/>



Metab2MeSH – Biological Substance Annotation with MeSH Terms

Search Metab2MeSH About Metab2MeSH

iron

Search by: MeSH term Substance

Metab2MeSH Search

examples: macrophages, phorbol ester

Welcome to Metab2MeSH

Metab2MeSH uses a statistical approach to reliably and automatically annotate biological substances with the concepts defined in MeSH, the National Library of Medicine's controlled vocabulary for biology and medicine. The Metab2MeSH web application searches biological substances or MeSH terms and displays resulting pairs of substances and MeSH terms that match the search term. The substance / MeSH term pairs displayed are those that are significantly associated in PubMed abstracts and are ordered highest to lowest by significance score.

Searching Metab2MeSH

1. Select whether you'd like to search MeSH terms or substances
2. Enter your search term into the search field and click "Metab2MeSH Search"
3. Click on any MeSH term or substance to search again with that specific term
4. Use the history tracker to go back to previous search results

Tips:

- Click the book icon to lookup the MeSH term in NCBI
- Click "download" to download the results into a tab delimited text file

NCBI University of Michigan National Institutes of Health

National Institutes of Health grant #U54 DA021519.
Copyright 2008 by The University of Michigan

Metab2MeSH

Metab2MeSH – Biological Substance Annotation with MeSH Terms



Search Metab2MeSH

About Metab2MeSH

iron Search by: MeSH term Substance

Metab2MeSH Search

examples: macrophages, phorbol ester

history : [iron](#) -> ["Iron"](#) -> ["Anemia, Iron-Deficiency"](#)

202 MeSH headings found matching substance "iron"

[download tab-delimited results](#)

= lookup MeSH heading at NCBI

Substance Name	Registry #	MeSH Heading	MeSH Qualifier	PubMed Articles*	P-Value	Fold Change
Iron	7439-89-6	Anemia, Hypochromic	drug therapy	4074	0.0e-1	65.5
Iron	7439-89-6	Hemochromatosis	genetics	2124	0.0e-1	61.2
Iron	7439-89-6	Anemia, Iron-Deficiency	blood	1677	0.0e-1	56.1
Iron-Dextran Complex	9004-66-4	Anemia, Hypochromic	drug therapy	273	0.0e-1	247.4
Iron	7439-89-6	Anemia	etiology	2690	0.0e-1	19.2
Iron	7439-89-6	Iron Overload	metabolism	816	0.0e-1	57
Iron	7439-89-6	Hemosiderosis	etiology	482	0.0e-1	47.1
Iron	7439-89-6	Siderosis	metabolism	329	0.0e-1	59.7
Iron	7439-89-6	Thalassemia	blood	872	0.0e-1	16.4

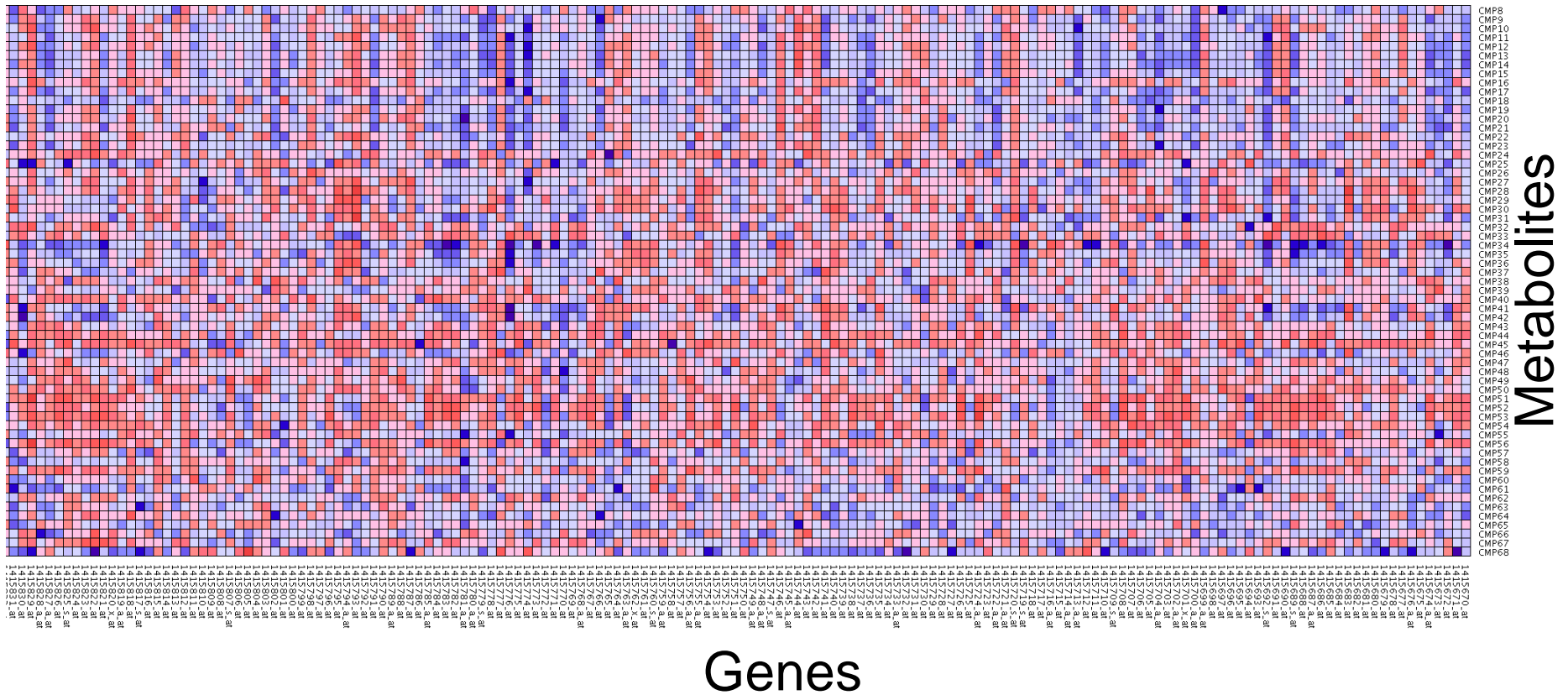
GenePattern Correlation Matrix Module

- GenePattern is a powerful genomic analysis platform developed by Broad Institute (Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP (2006) GenePattern 2.0 Nature Genetics 38 no. 5 (2006): pp500-501 doi:10.1038/ng0506-500.)
- It provides access to variety of tools for gene expression analysis, proteomics, SNP analysis and common data processing tasks.
- **GenePattern Correlation Matrix Module** calculates pairwise correlations between genes and metabolites and creates a matrix that can be imported into other GenePattern tools, e.g. HeatMapView.

Can be accessed from NCIBI portal (<http://portal.ncibi.org/gateway/>)

Please contact me for specific instructions on how to access this tool
(akarnovs@umich.edu)

Visualizing Gene-Metabolite Correlation Matrix as Heatmap



Metscape

Objectives:

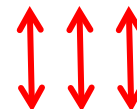
- create the framework for the analysis of metabolomic data in the context of human metabolic pathways
- develop a tool that would link metabolite data with other types of high throughput molecular data by incorporating prior knowledge of metabolites, genes, proteins, pathways and molecular interactions

Framework for Data Analysis

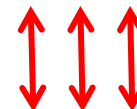
- A number of public databases contain information about, metabolites, reactions, human metabolic networks:

- KEGG (Kanehisa et al., 2008)
- BIGG (Recon1 - Duarte et al., 2007)
- EHMN (Ma et al., 2007)
- HMDB (Wishart et al., 2009)

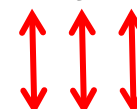
Compounds



Reactions



Enzymes (EC)



Genes

Exploring Data in MetScape

The image displays the MetScape 1.0 software interface, which is a plug-in for Cytoscape. It shows the search process for a specific compound and the resulting network visualization of the TCA cycle.

Welcome to MetScape 1.0

Enter Compound(s) | Load Experiment File | Show Pathway

Query Metabolic Database

Enter Compound(s); e.g. c00024 c00020

Search

Cytoscape Desktop (New Session)

Control Panel | Edit | Files | MetScape Layout

Initial seed compound

Compound

Expansion seed compound

Expansion compound

Reaction

Expansion reaction

Reaction with single enzyme

Reaction with multiple enzymes

Substrate compound

Product compound

Color range [2015.264264 to 19162.64]

Size range [50 to 200]

TCA cycle

9-oxaloacetate

8-(s)-malate

7-fumarate

6-succinate

5-succinyl-coa

4-α-ketoglutarate

3-isocitrate

2-oxoglutarate

1-acetyl-Coa

Control Panel

EC	Description	Enzyme	Pathway	Type
PE1706	Coa+ succ 10' 5' hydroxyl-oxalate 8' 8' E' 12'	2.3.1.10 (pepco-CoA-synthetase)	3-oxo-10'-oxalate-α-ketoglutarate beta-oxidation	Reaction
PE1460	Coa+ succ 10' 3' hydroxyl-oxalate 8' 2' succinate 4'	2.3.1.10 (pepco-CoA-synthetase)	3-oxo-10'-oxalate-α-ketoglutarate beta-oxidation	Reaction
PE1569	Acetyl-Coa+ Dihydroxyacetate=Coa+ Acetylshufc	2.3.1.11 (ethionine/phenylethylamine acetyltransferase)	Glycolysis and gluconeogenesis	Reaction
PE1577	Acetyl-Coa+ Glyoxylate=Coa+ Glyoxylate	2.3.1.21 (pepco-CoA synthetase)	Glycolysis and gluconeogenesis	Reaction
PE1695	3-Oxoglutarate+Coa+H2O=Acetyl-Coa+CO2+H2O	2.3.1.10 (pepco-CoA-synthetase)	Propionate metabolism	Reaction
PE1696	Acetyl-Coa+ Malate+Coa=Coa+ Succinyl-Coa	2.3.1.10 (pepco-CoA-synthetase)	Succinyl-Coa synthetase	Reaction
PE1678	2-Acetyl-Coa+ Acetylshufc=Coa	2.3.1.10 (pepco-CoA-synthetase)	Calculated fatty acid beta-oxidation	Reaction
PE1703	Coa+ succ 10' 10' hydroxyl-oxalate 8' 8' E' 12'	2.3.1.10 (pepco-CoA-synthetase)	3-oxo-10'-oxalate-α-ketoglutarate beta-oxidation	Reaction
PE1674	Acetyl-Coa+ Malate=Coa+ Succinyl-Coa	2.3.1.10 (pepco-CoA-synthetase)	3-oxo-10'-oxalate-α-ketoglutarate beta-oxidation	Reaction

Dynamic Expression

Time Series

Expression_U[7 to 6632] Control Node Color

1

2

3

4

5

Play

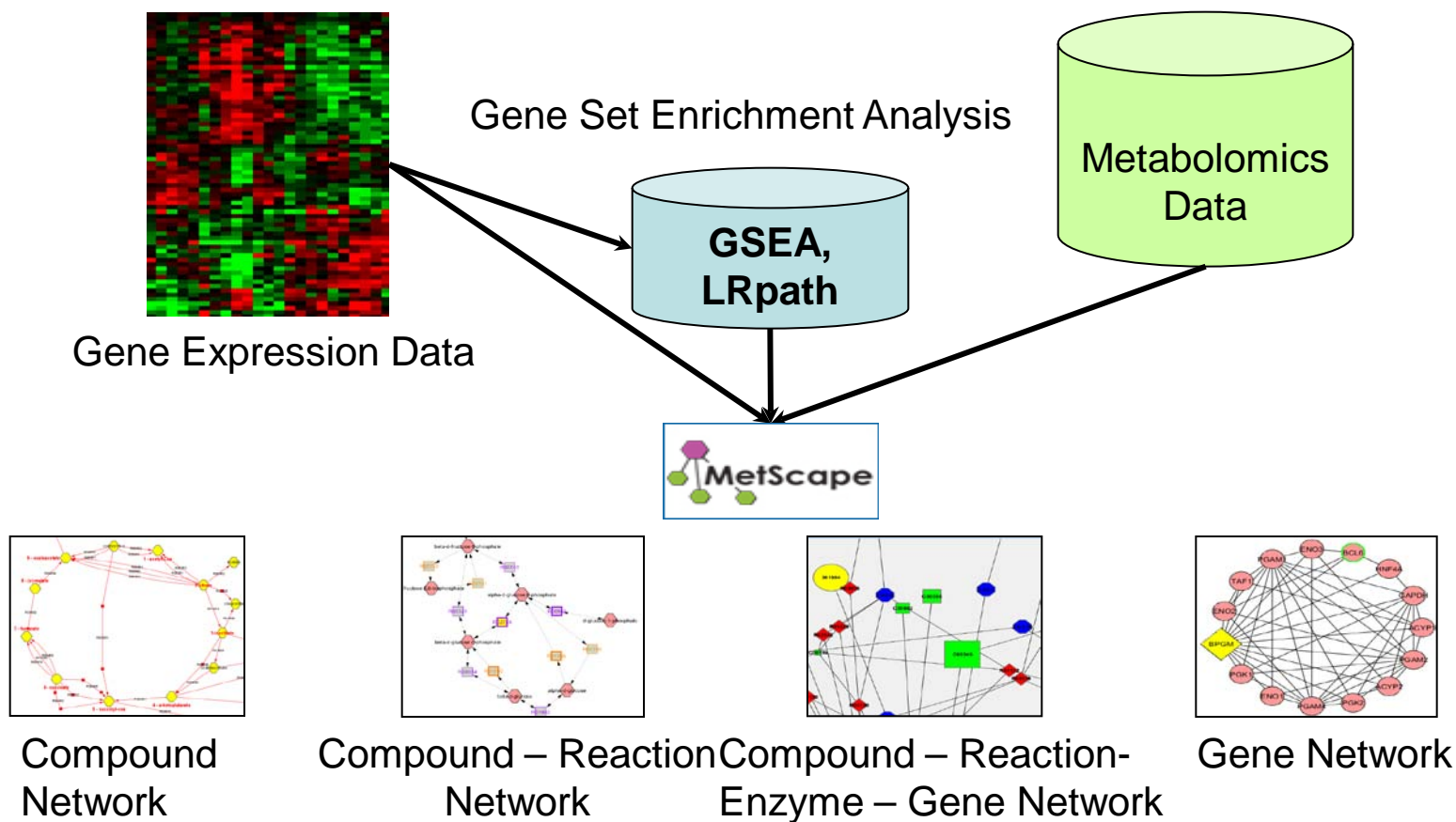
Pause

Stop

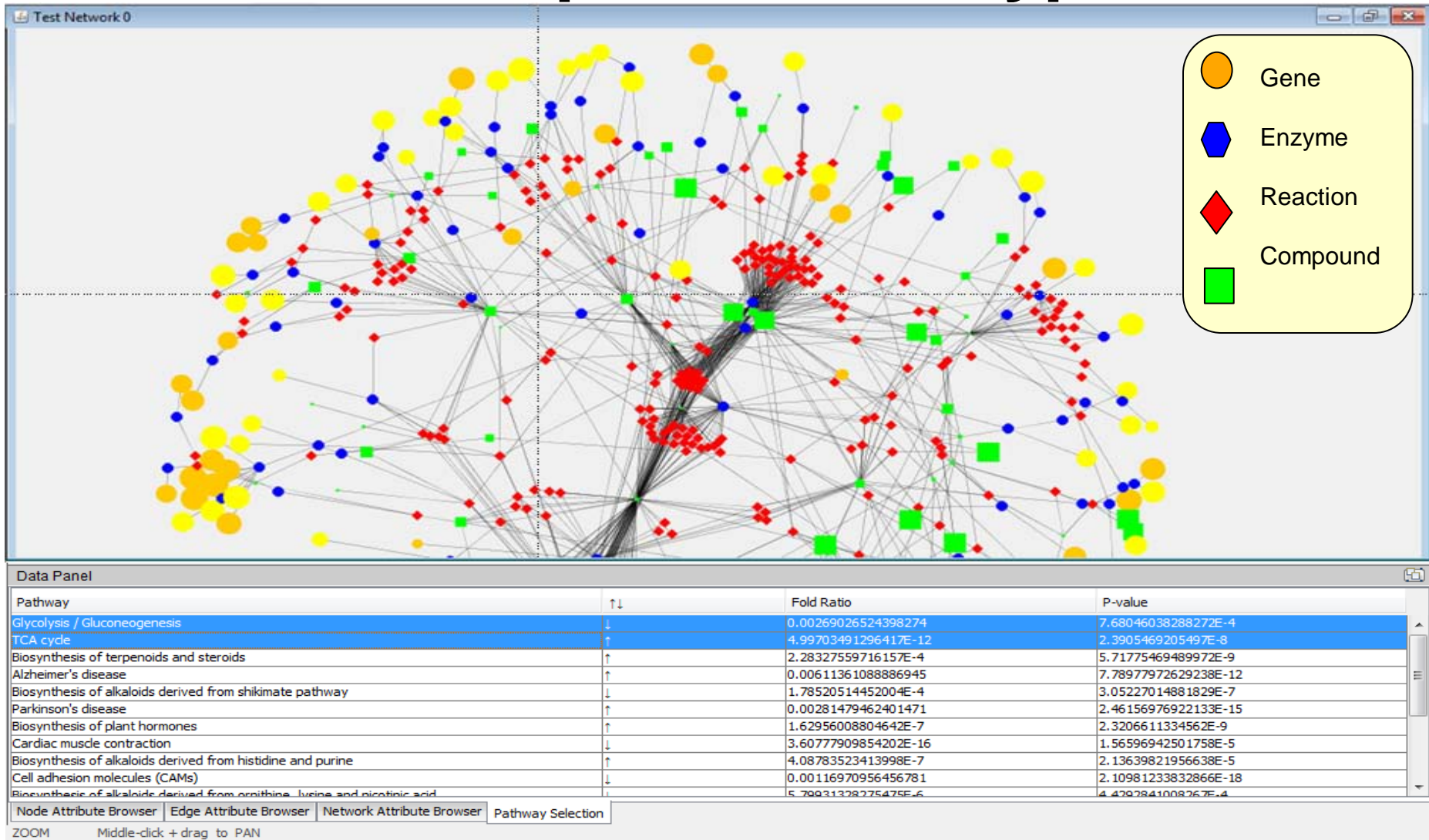
Duration

MetScape is a plug-in for Cytoscape (Shannon et al., 2003)
<http://www.cytoscape.org>

Next Generation Metscape Workflow



MetScape 2.0 Prototype



We Need Your Feedback!

You can help us improve our tools by providing your feedback

Please take a minute to complete the survey
at the end of this session

Metscape 1.0 and Pathway tools/Omics viewer evaluation survey		
Please indicate on a scale from 1 to 5 (5 – strongly agree, 1 – strongly disagree) how satisfied you were with various aspects of 2 programs.		
	Metscape	Or
1. Loading experimental data		
- Data upload process is easy and efficient		
- It is easy to un-do, re-do and make corrections in the process of data loading if needed		
- It is easy to verify that the data was loaded correctly		
- The input values accepted by the program adequately describe the data (e.g. compound concentration, fold change)		
- I would like to be able to view additional derived values (e.g. enrichment scores) .		
2. Building and visualizing network		
- The parameters for a desired network view are easy to set.		
- The parameters clearly describe the resulting view		
3. Network view		

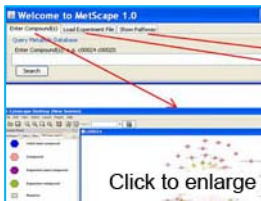
Background Information for Today's Demo

- Approximately 150 million people worldwide have type 2 diabetes (T2D) and this number is predicted to double to 300 million by 2025
- Db/db diabetic mouse is one of the most extensively studied animal models of T2D.
 - Db/db 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.
- We will use the NMR-based unbiased metabolomics data set from Connor et al.*
 - characterized the differences in metabolites between urine samples from the diabetic db/db and control db/+ mice.
 - identified 66 metabolites, including several that were previously confirmed as arising from gut microflora

*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.

Goals of This Tutorial

- Use two publically available bioinformatics tools - Metscape 1.0 (NCIBI) and Pathway Tools/Omics Viewer (SRI) to analyze the data and compare the results.



Metscape

<http://portal.ncibi.org/gateway/metscape.html>



Omics Viewer

<http://biocyc.org/expression.html>

Input Files

Metscape

	1 Fold change
C01585	0.182683524
C00246	0.212247974
C08262	0.212247974
C00245	0.22313016
C00489	0.275270783
C02632	0.275270783
C00671	0.292292578
C00141	0.295230167
C00803	0.310366941
C02170	0.316636769
C01042	0.343008517
C00417	0.36421898
C00047	3.004166024

Omics Viewer

CID	Fold Change
C01585	-1.7
C00246	-1.55
C08262	-1.55
C00245	-1.5
C00489	-1.29
C02632	-1.29
C00671	-1.23
C00141	-1.22
C00803	-1.17
C02170	-1.15
C01042	-1.07
C00417	-1.01
C00047	1.1

All materials from today's demo will be available at the NCIBI portal

<http://portal.ncibi.org/gateway/>

Pathways Identified by Metscape

Amino Acid Metabolism	Glycine, serine, alanine and threonine metabolism
	Histidine metabolism
	Lysine metabolism
	Methionine and cysteine metabolism
	Tyrosine metabolism
	Urea cycle and metabolism of arginine, proline, glutamate, aspartate and asparagine
Carbohydrate metabolism	Butanoate metabolism
	Fructose and mannose metabolism
	Galactose metabolism
	Glycolysis and Gluconeogenesis
	Phytanic acid peroxisomal oxidation
	TCA cycle
Lipid metabolism	Bile acid biosynthesis
	Glycerophospholipid metabolism
	Leukotriene metabolism
	Lipoate metabolism
Metabolism of Cofactors and Vitamins	Biopterin metabolism
	Porphyrin metabolism
	Valine, leucine and isoleucine degradation
	Vitamin B3 (nicotinate and nicotinamide) metabolism
	Vitamin B5 - CoA biosynthesis from pantothenate
	Vitamin B9 (folate) metabolism
	Vitamin H (biotin) metabolism
Nucleotide Metabolism	Purine metabolism
	Pyrimidine metabolism

Found 25 unique pathways

Finding Significant Pathways

Ingenuity Pathway Analysis
(commercial tool)

ConceptMetab
(NCIBI)

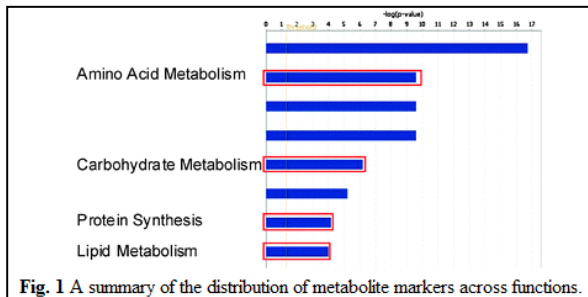


Fig. 1 A summary of the distribution of metabolite markers across functions.

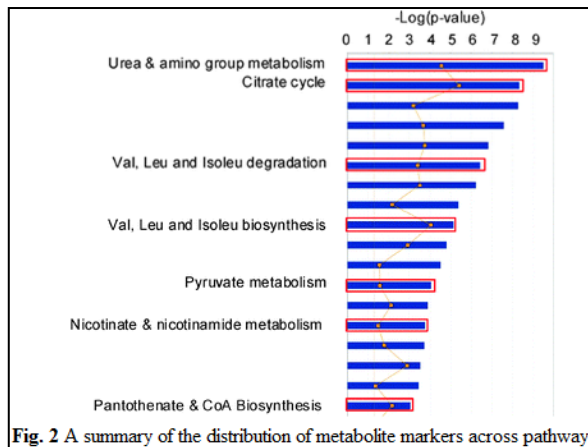


Fig. 2 A summary of the distribution of metabolite markers across pathway.

