

# Computational methods for data integration

July 21, 2017

Karan Uppal, PhD

Assistant Professor of Medicine

Director of Computational Metabolomics & Integrative Omics,  
Clinical Biomarkers Laboratory

Emory University

# Learning Objectives

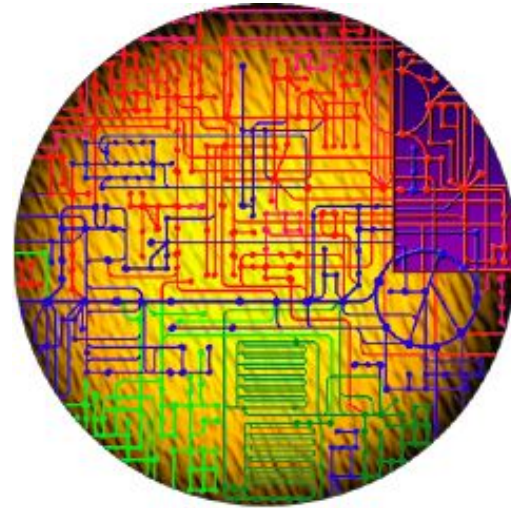
- Understanding of different data integration approaches
- Familiarity with tools for data integration and network visualization

# Introduction: A Systems Biology Framework

- The goal of **Systems Biology**:
  - Systems-level understanding of biological systems
  - Analyze not only individual components, but their interactions as well and emergent behavior



Exposures  
Internal measurements  
Disease states

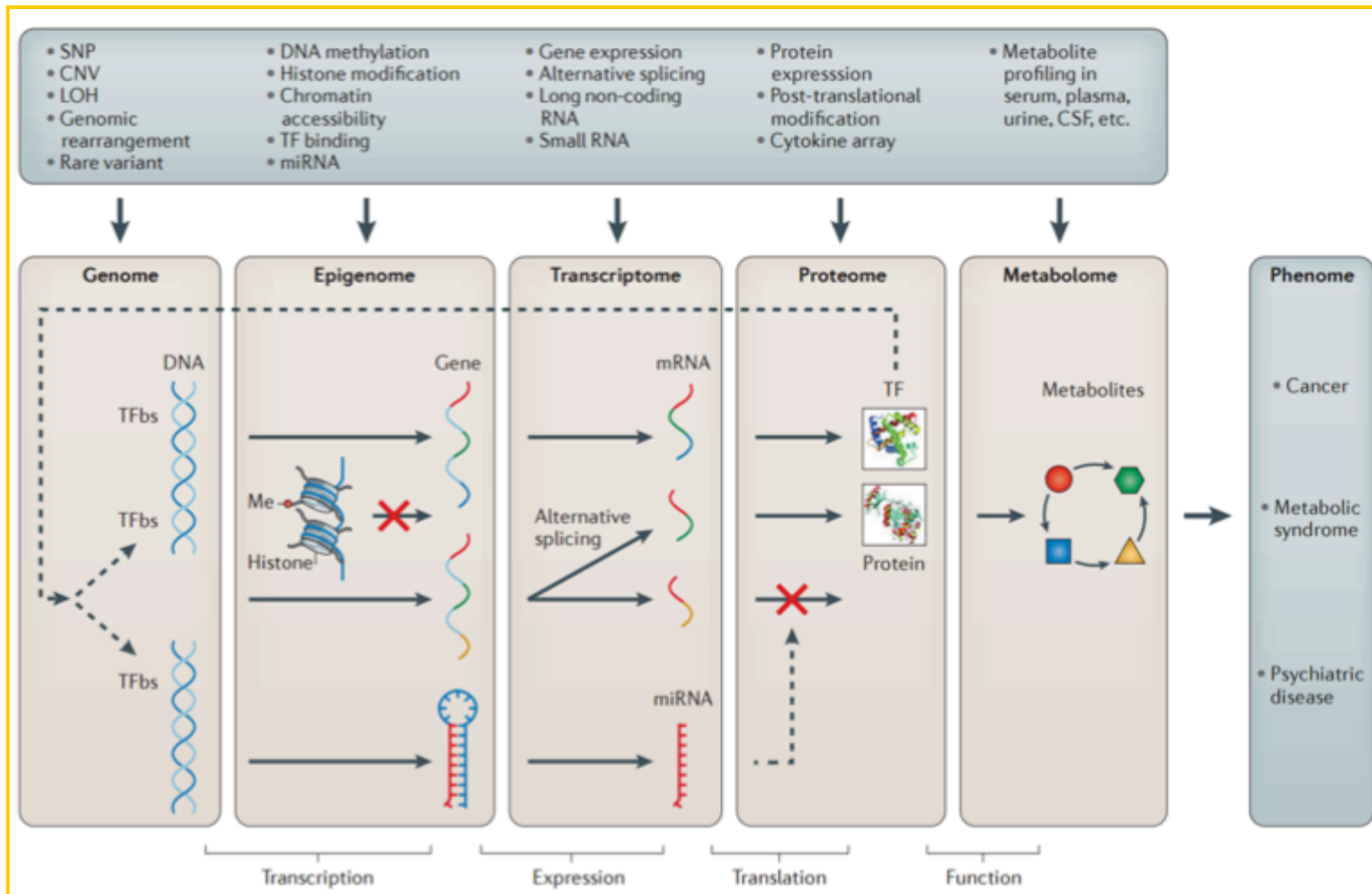


## **Systems Biology**

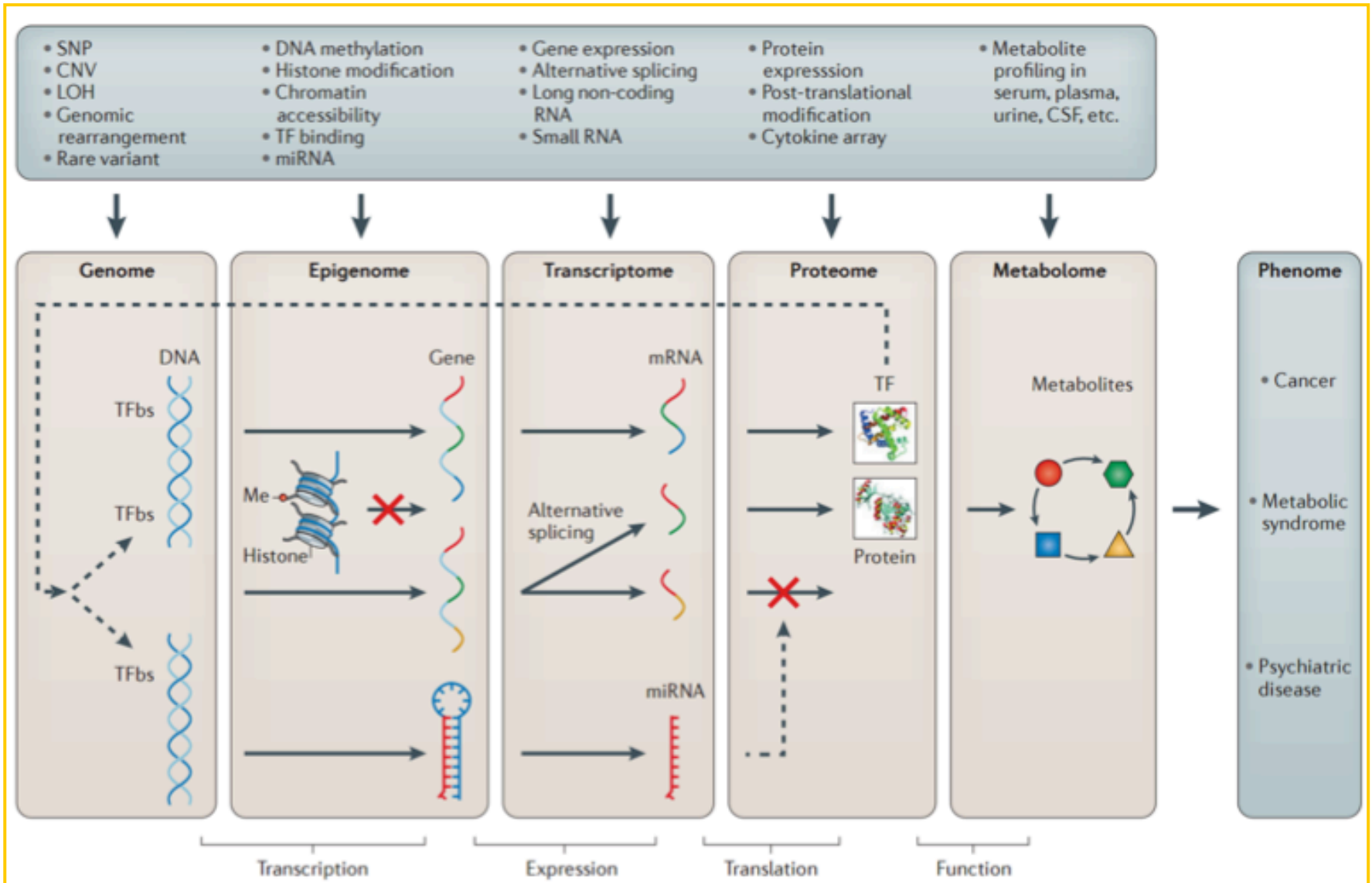
*“Integrative approach in which scientists study pathways and networks will touch all areas of biology, including drug discovery”*

C. Henry and C. Washington

# Dissecting the Biological system via -omics



# Dissecting the Biological system via -omics



**“Information Overload”: >10,000 variables per –omics experiment**

# Why data integration?

- Systems level analysis provides:
  - more detailed overview of underlying mechanisms;
  - exploration of interactions between different biomedical entities (genes, proteins, metabolites, etc.)
- Combining multiple types of data compensates for noise or unreliable information in a single data type
- More confidence in results if multiple sources of evidence pointing to the same gene or pathway

# Paired integrative –omics analysis

- Discover networks of associations or correlated variables (genes, proteins, metabolites, microbiome, epigenetic alterations, clinical variables, etc.) from paired –omics data measured across same samples
  - Univariate or multivariate regression
  - Example: explaining protein abundance with respect to gene expression
- Determine if different –omics data point to same disease mechanism
- Generate novel hypotheses for further investigation

# Main approaches for data integration

- Pathway or knowledge-based integration
  - Datasets are analyzed individually (differentially expressed genes, metabolites, proteins) and integration is performed at the pathway level
  - Examples: MetaboAnalyst, iPEAP, MetScape
- Data-driven integration using meta-dimensional analysis
  - Integration is performed globally such that data from multiple omics layers are combined simultaneously
  - Examples: 3Omics, mixOmics, xMWAS
- Using literature-derived associations for integration
  - Using co-occurrence criteria for establishing relationship
  - Examples: CoPub, ArrowSmith, SEACOIN2.0



# Main approaches for data integration

- **Pathway or knowledge-based integration**
  - Datasets are analyzed individually (differentially expressed genes, metabolites, proteins) and integration is performed at the pathway level
  - Examples: MetaboAnalyst, iPEAP, MetScape, MetaCore
- Data-driven integration using meta-dimensional analysis
  - Integration is performed globally such that data from multiple omics layers are combined simultaneously
  - Examples: 3Omics, mixOmics, xMWAS
- Using literature-derived associations for integration
  - Using co-occurrence criteria for establishing relationship
  - Examples: CoPub, ArrowSmith, SEACOIN2.0

# Pathway or knowledge-based integration

Metabolomics data  
(n subjects X p metabolites)

	M1	M2	-	Mp
Subject1	199	19	-	100
Subject2	10	40		90
-	-	-		-
SubjectN	50	30	-	20

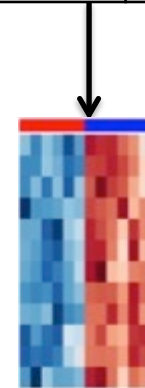


**Differentially  
expressed  
metabolites**

Pathway analysis for metabolites

Transcriptomics data  
(n subjects X q genes)

	G1	G2	-	Gq
Subject1	19	19	-	100
Subject2	10	40	-	90
-	-	-	-	-
SubjectN	10	40	-	50



**Differentially  
expressed  
genes**

Pathway analysis for genes

Common pathways or pathway rank aggregation

# Pathway or knowledge-based integration

Metabolomics data  
(n subjects X p metabolites)

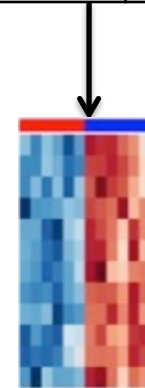
	M1	M2	-	Mp
Subject1	199	19	-	100
Subject2	10	40		90
-	-	-		-
SubjectN	50	30	-	20



**Differentially  
expressed  
metabolites**

Transcriptomics data  
(n subjects X q genes)

	G1	G2	-	Gq
Subject1	19	19	-	100
Subject2	10	40	-	90
-	-	-	-	-
SubjectN	10	40	-	50



**Differentially  
expressed  
genes**

Pathway analysis using genes and metabolites (joint)

# MetaboAnalyst

(<http://www.metaboanalyst.ca>)

← ⓘ | [www.metaboanalyst.ca/faces/ModuleView.xhtml](http://www.metaboanalyst.ca/faces/ModuleView.xhtml) | ↻ 🔍 Search



[Troubleshooting](#)

[Resources](#)

[Update History](#)

[User Stats](#)

[About](#)



also provide clustering and visualization tools to create dendrograms and heatmaps as well as to classify based on random forests and SVM.

### Pathway Analysis

This module supports pathway analysis (integrating enrichment analysis and pathway topology analysis) and visualization for 21 model organisms, including Human, Mouse, Rat, Cow, Chicken, Zebrafish, Arabidopsis thaliana, Rice, Drosophila, Malaria, S. cerevisiae, E.coli. and others, with a total of ~1600 metabolic pathways.

### Time-series/Two-factor Design

This module supports temporal and two-factor data analysis including data overview, two-way ANOVA, and empirical Bayes time-series analysis for detecting distinctive temporal profiles. It also supports ANOVA-simultaneous component analysis (ASCA) to identify major patterns associated with each experimental factor.

### Power Analysis

This module uses pilot data to calculate the minimum number of samples required to detect a statistically significant difference between two populations with a given degree of confidence (called Power Analysis).

### Biomarker Analysis

This module performs various ROC curve based biomarker analyses for a single or multiple biomarkers. It also allows users to manually specify biomarker models as well as new sample prediction.


### Integrated Pathway Analysis

This module performs integrated metabolic pathway analysis on results obtained from combined metabolomics and gene expression studies conducted under the same experimental conditions.

### Other Utilities

This module contains several common utility functions. At this moment, **compound ID conversion**, **batch effect correction** and **lipidomics data analysis** are available.

# Data upload



Upload

Integrative Analysis

Download

Exit

Gene List

Gene list with optional fold changes

#Entrez	logFC
1737	-1.277784317
83440	-1.034136439
3939	-2.231729728
10911	-1.045657875
10690	-0.968308832
10010	-0.861541301
11224	1.187399591
63826	-1.405238611
11031	0.785011172
4190	-1.778774832
10782	-2.140715987
10993	-0.925083829
10455	1.732172706
10963	1.177511121
10282	-1.20754269

ID Type: 

Entrez ID

Metabolite List

Compound list with optional fold changes

#KEGG	logFC
C00116	1.010972619
C00565	-0.714283001
C00033	0.822193121
C00583	-1.005192252
C00022	-0.623838569
C00719	-0.406052491
C05984	-0.390152174
C00207	-0.932835099
C00065	0.903658797
C00031	0.548035915
C00079	0.416744818
C02632	-0.515041676
C00064	-0.497216411
C00114	1.102078837
C00073	0.516193785

ID Type: 

KEGG ID

Specify organism: 

Homo sapiens (human)

☒ Use our example data

Submit

### Enrichment Analysis

Enrichment analysis aims to evaluate whether the observed genes and metabolites in a particular pathway are significantly enriched (appear more than expected by random chance) within the dataset. You can choose over-representation analysis (ORA) based on either hypergeometrics analysis or Fisher's exact method.

- ☒ Hypergeometric Test
- ☐ Fisher's Exact Test

### Topology Analysis

The topology analysis aims to evaluate whether a given gene or metabolite plays an important role in a biological response based on its **position** within a pathway. Degree Centrality measures the number of links that connect to a node (representing either a gene or metabolite) within a pathway; Closeness Centrality measures the overall distance from a given node to all other nodes in a pathway; Betweenness Centrality measures the number of shortest paths from all nodes to all the others that pass through a given node within a pathway.

- ☒ Degree Centrality
- ☐ Betweenness Centrality
- ☐ Closeness Centrality

### Pathway Databases

Users can choose one of three different modes of pathways: - the gene-metabolite mode (default) allows joint-analysis and visualization of both significant genes and metabolites; while the gene-centric or metabolite-centric mode allows users to identify enriched pathways driven by significant genes or metabolites, respectively.

- ☒ Gene-metabolite pathways
- ☐ Gene-centric pathways
- ☐ Metabolite-centric pathways

→ Submit

# Set Parameters



Upload

Integrative Analysis

ID map

Set parameter

Overview

View result

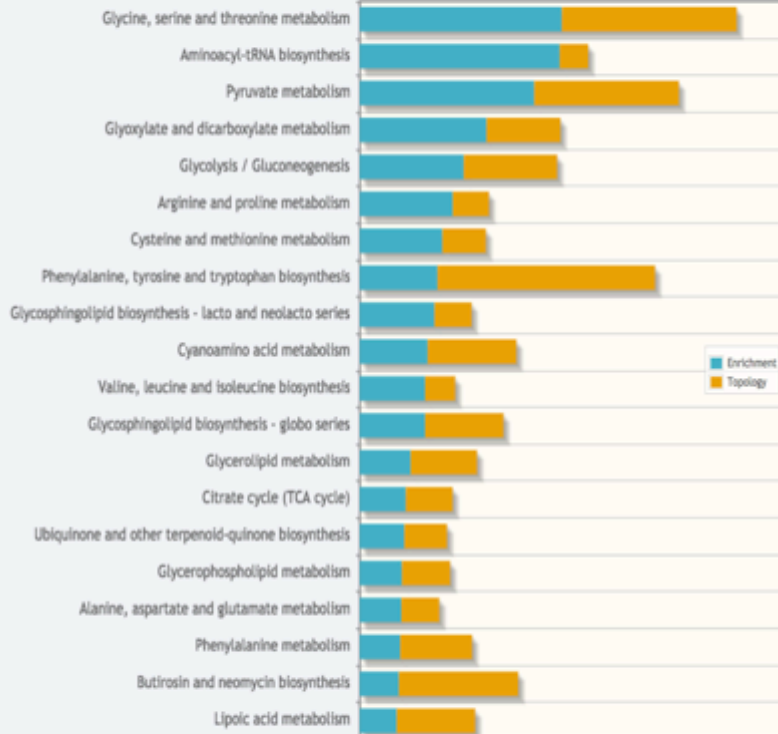
Download

Exit

# Results

The stacked bars below show a summary of the joint evidence from enrichment analysis and topology analysis.

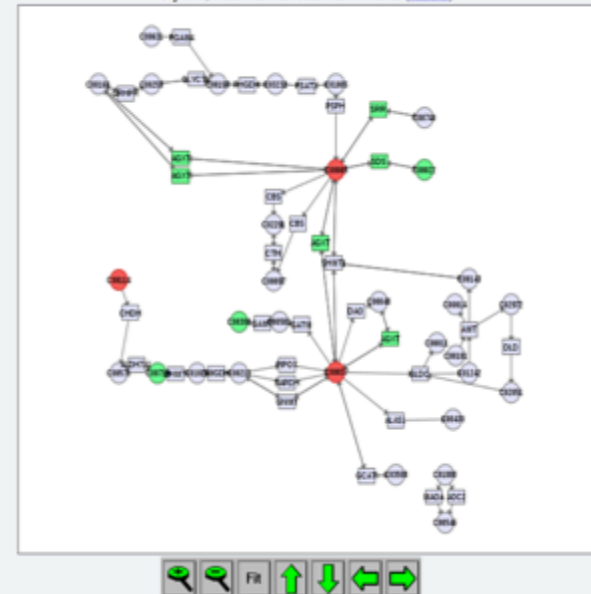
Pathway Analysis Overview



Export Image

View Details

Glycine, serine and threonine metabolism (KEGG)



The matched nodes are highlighted in different colors - red (upregulated), yellow (unknown), green (downregulated) based on fold change (FC) values. Click on a node to show more details.

Pathway	Total	Expected	Hits	P-Value	Topology	View
Glycine, serine and threonine metabolism	68	1.51	9	1.2496E-5	0.96825	<a href="#">View</a>
Aminoacyl-tRNA biosynthesis	87	1.9319	10	1.4183E-5	0.17391	<a href="#">View</a>
Pyruvate metabolism	64	1.4212	8	6.1067E-5	0.80435	<a href="#">View</a>
Glyoxylate and dicarboxylate metabolism	53	1.1769	6	9.4579E-4	0.42	<a href="#">View</a>
Glycolysis / Gluconeogenesis	91	2.0207	7	0.0034979	0.52542	<a href="#">View</a>
Arginine and proline metabolism	102	2.265	7	0.0065963	0.21505	<a href="#">View</a>
Cysteine and methionine metabolism	63	1.3989	5	0.011936	0.25455	<a href="#">View</a>
Phenylalanine, tyrosine and tryptophan biosynthesis	9	0.19985	2	0.015789	1.2	<a href="#">View</a>
Glycosphingolipid biosynthesis - lacto and neolacto series	26	0.57734	3	0.018816	0.22034	<a href="#">View</a>

# Main approaches for data integration

- Pathway or knowledge-based integration
  - Datasets are analyzed individually (differentially expressed genes, metabolites, proteins) and integration is performed at the pathway level
  - Examples: MetaboAnalyst, iPEAP, MetScape
- **Data-driven integration using meta-dimensional analysis**
  - **Integration is performed globally such that data from multiple omics layers are combined simultaneously**
  - **Examples: 3Omics, mixOmics, xMWAS**
- Using literature-derived associations for integration
  - Using co-occurrence criteria for establishing relationship
  - Examples: CoPub, ArrowSmith, SEACOIN2.0



## Metabolomics data (n subjects X p metabolites)

	M1	M2	-	Mn
Subject1	199	19	-	100
Subject2	10	40		90
-	-	-		-
SubjectN	50	30	-	20

## Transcriptomics data (n subjects X q genes)

	G1	G2	-	Gn
Subject1	19	19	-	100
Subject2	10	40	-	90
-	-	-	-	-
SubjectN	10	40	-	50

## Association matrix

	G1	G2	-	Gn
M1	0.4	0.9	-	0.3
M2	0.7	0.1	-	0.5
M3	0.1	0.6		0.8

### Univariate

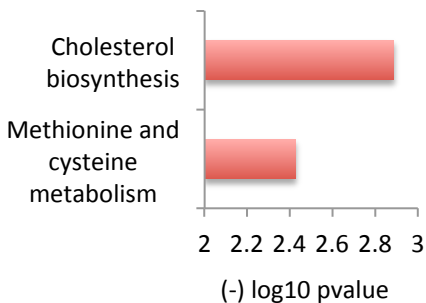
- Pearson, Spearman, Partial Correlation
- Tools: 3Omics, MetabNet, etc.

### Multivariate

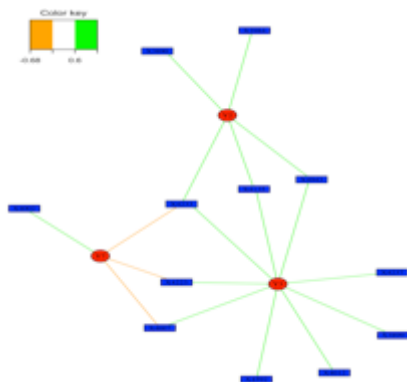
- PLS, CCA, sparse PLS
- Tools: mixOmics (Cao 2009), etc.

# Workflow

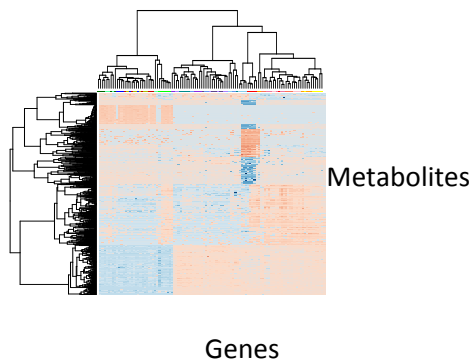
## Pathway enrichment



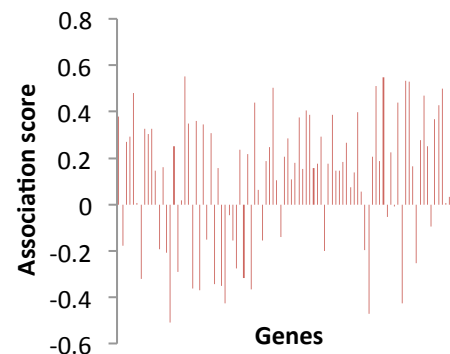
## Relevance networks



## Clustering

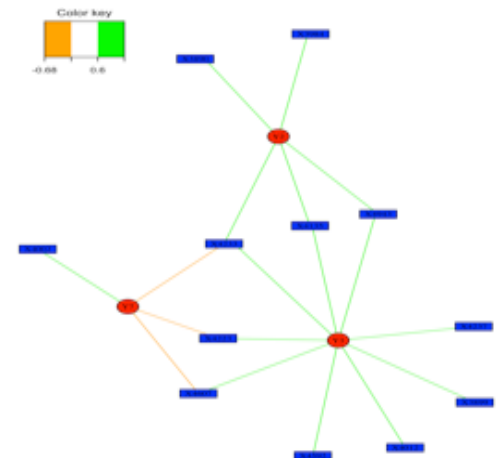


## Targeted investigation (e.g.: Arginine x Transcriptome)



# Relevance networks

- What is a network (or graph)?
  - A set of nodes (vertices) and edges (links)
  - Edges describe a relationship (e.g. correlation) between the nodes
- What is a relevance network?
  - Networks of highly-correlated biomedical/clinical entities (Butte 2000; PNAS)
  - Metabolomics x Proteomics, Transcriptomics x Proteomics, Metabolomics x Microbiome, Metabolomics x Clinical variables/phenotypes, etc.
  - Generate a bipartite graph network using an association threshold (e.g. 0.5) to visualize positive or negative associations



Circles: microbial species  
Rectangles: metabolome features

# Methods for generating relevance networks

- Univariate
  - Pairwise Pearson or Spearman correlation between data from different biomedical/clinical technologies (Butte et al. 2000, Uppal et al. 2015)
  - 3Omics (Kuo 2013; a web-based tool for analysis, integration and visualization of human transcriptome, proteome and metabolome data)
  - MetabNet (Uppal 2015; R package for performing pairwise correlation analysis and generating relevance networks)
- Multivariate
  - Multivariate regression techniques such as partial least squares (PLS), sparse partial least squares regression (sPLS), multilevel sparse partial least squares (msPLS) regression, etc.
  - mixOmics (Cao et al. 2009, Liquet et al. 2012; R package for integration and variable selection using multivariate regression)
  - xMWAS (Uppal et al. Submitted): R package for data-driven integration and differential network analysis

# Univariate methods

# 3Omics (Kuo et al. BMC Systems Biology 2013)

- A web-based tool for analyzing, integrating and visualizing transcriptomic, proteomic and metabolomic data
- <http://3omics.cmdm.tw/>

# 3Omics - homepage



The screenshot shows the 3Omics homepage in a web browser. The browser's address bar displays "3omics.cmdm.tw". The page features a navigation menu on the left with links for "Project Features", "Overview", "Name-ID Converter", "Help", and "Contact Us". Below the menu are logos for MDDL and National Taiwan University. The main content area is titled "Overview" and describes 3Omics as a web-based systems biology visualization tool for integrating human transcriptomic, proteomic, and metabolomic data. It lists five analysis types: 1. Transcriptomics-Proteomics-Metabolomics, 2. Transcriptomics-Proteomics, 3. Proteomics-Metabolomics, 4. Transcriptomics-Metabolomics, and 5. Proteomics only. A diagram illustrates the relationships between Transcripts (T), Proteins (P), and Metabolites (M) using colored shapes and lines. A legend at the bottom identifies the shapes: green square for Transcripts, red triangle for Proteins, and blue circle for Metabolites. It also defines the line types: a solid line for Correlation and a dashed line for Literature-derived relationship. The footer contains copyright information (© 2006-2012 Computational Molecular Design & Metabolomics Laboratory | BEB | NTU) and contact details, along with a recommendation to use Google Chrome or Mozilla Firefox.

3omics.cmdm.tw

3Omics

**Project Features**  
Overview  
Name-ID Converter  
Help  
Contact Us

**Overview**

**3Omics: A web based systems biology visualization tool for integrating human transcriptomic, proteomic and metabolomic data**

3Omics is a one-click web tool for visualizing and rapidly integrating multiple inter- or intra-transcriptomic, proteomic, and metabolomic human data. It covers and connects cascades from transcripts, proteins, and metabolites and provides five commonly used analyses including correlation network, co-expression, phenotype generation, KEGG/HumanCyc pathway enrichment, and GO enrichment.

Please select the desired analysis:

- 1. Transcriptomics-Proteomics-Metabolomics
- 2. Transcriptomics-Proteomics
- 3. Proteomics-Metabolomics
- 4. Transcriptomics-Metabolomics
- 5. Proteomics only

Please refer to the help page for more details about each integrating method.

Legend:  
T: Transcripts  
P: Proteins  
M: Metabolites  
—: Correlation  
---: Literature-derived relationship

Copyright (c) 2006-2012 Computational Molecular Design & Metabolomics Laboratory | BEB | NTU | Contact us  
We recommend using the latest version of Google Chrome or Mozilla Firefox to get the best experience using 3Omics services.

# Features

- Correlation analysis and network visualization
  - Pairwise Pearson correlation analysis
- Database-derived relationships in correlation analysis
  - Uses an internal database based on NCBI Entrez gene, Uniprot proteins, and KEGG metabolites to determine gene-protein-metabolite relationship
- Coexpression analysis
  - Two-way hierarchical clustering analysis
  - Rows: variables (Genes + proteins + metabolites, genes+metabolites, etc.)
  - Columns: samples
- Phenotype analysis
  - Uses OMIM databases to link genes with phenotypes
- Pathway and Gene Ontology Enrichment analysis
  - Using KEGG, HumanCyc, and DAVID

# Data upload

Please select the desired analysis.

a. Transcriptomics-Proteomics-Metabolomics  
b. Transcriptomics-Proteomics

c. Proteomics-Metabolomics  
d. Transcriptomics-Metabolomics

e. Transcriptomics only  
f. Proteomics only  
g. Metabolomics only

Please refer to the help page for more details about each integrating method.



[← Back](#)

User may upload three kinds of -omic expression data. All analyses will be performed.

☐ [Use example data](#) [?](#)

## Transcriptomics

No file selected. [?](#)

GenBank ID: e.g. [NAT1](#), [ABL1](#)

## Proteomics

No file selected. [?](#)

Uniprot Accession: e.g. [P31946](#), [P62258](#)

## Metabolomics

No file selected. [?](#)



# Data format

(<http://3omics.cmdm.tw/help.php#examples>)

	Samples					
	timepoint1	timepoint2	timepoint3	timepoint4	timepoint5	
Variables	akap9	-0.24	-0.6	-0.47	-0.38	-0.31
	macf1	-0.3	-0.3	0.48	0.07	-0.36
	RNPEP	0.24	0.85	0.15	0.79	0.69
	SDHA	0.1	0.37	0.18	0.23	0.33
	EEF1B2	-0.04	-0.31	0.06	-0.39	-0.46
	EEF1D	0.07	0.29	0.22	0.75	0.47
	EIF4A1	0.42	0.65	0.66	0.97	0.78
	WARS	1.47	1.72	0.58	1.79	1.69
	G3BP2	0.15	0.09	0.1	0.2	-0.22
	PAK2	-0.21	-0.14	-0.15	-0.31	-0.4
	PPP4C	-0.13	0.05	-0.09	0.21	-0.12
	ZNF224	-0.06	0.31	0.17	0.27	0.61
	ZNF268	-0.23	0.08	0.01	0.1	-0.1
	TRRAP	0.07	-0.12	0.41	0.45	-0.09
	RAD23B	-0.07	-0.32	-0.02	-0.02	-0.44
	TARDBP	0.23	0.18	0.39	0.63	0.23
	CSTF2	0.51	0.65	0.71	1.18	0.89
	PSMC2	0.82	0.57	1.15	1.75	0.58
	F8	-0.19	-0.02	-0.35	-0.82	-0.81
	MYOM1	-0.28	-0.29	-0.54	-1.06	-1.03
	ACTR3	0.57	0.48	0.39	0.32	0.72
	ITPR2	0.62574	1.771	-0.057392	1.2612	1.7769
	NUCB2	-1.1943	-0.96016	-0.71549	-1.1877	-0.70604
	CAMK1	0.33342	0.87499	0.059355	0.062122	0.53605
	BCL2A1	2.2913	3.8479	-0.12343	1.6604	3.3933
	PDCD6IP	0.46362	0.88049	0.20539	0.36177	0.62012

# Correlation analysis

Help

Contact Us

Parameters Section



How to set up parameters?

Correlation Coefficient Threshold  
0.9

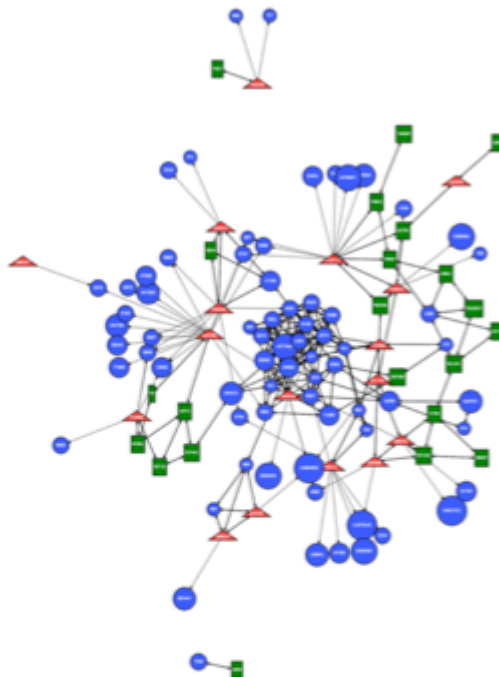
Correlation Network Regulation  
160

Correlation Network Attraction  
80

Refresh

## Correlation Network of Transcriptomics, Proteomics & Metabolomics



3Omics generates inter-omic correlation network to display the relationship or common patterns in data over time or experimental conditions for all transcripts, proteins and metabolites. Where users may only have two of the three -omics data-sets, 3Omics supplements the missing transcript, protein or metabolite information by searching [iHOP database](#).

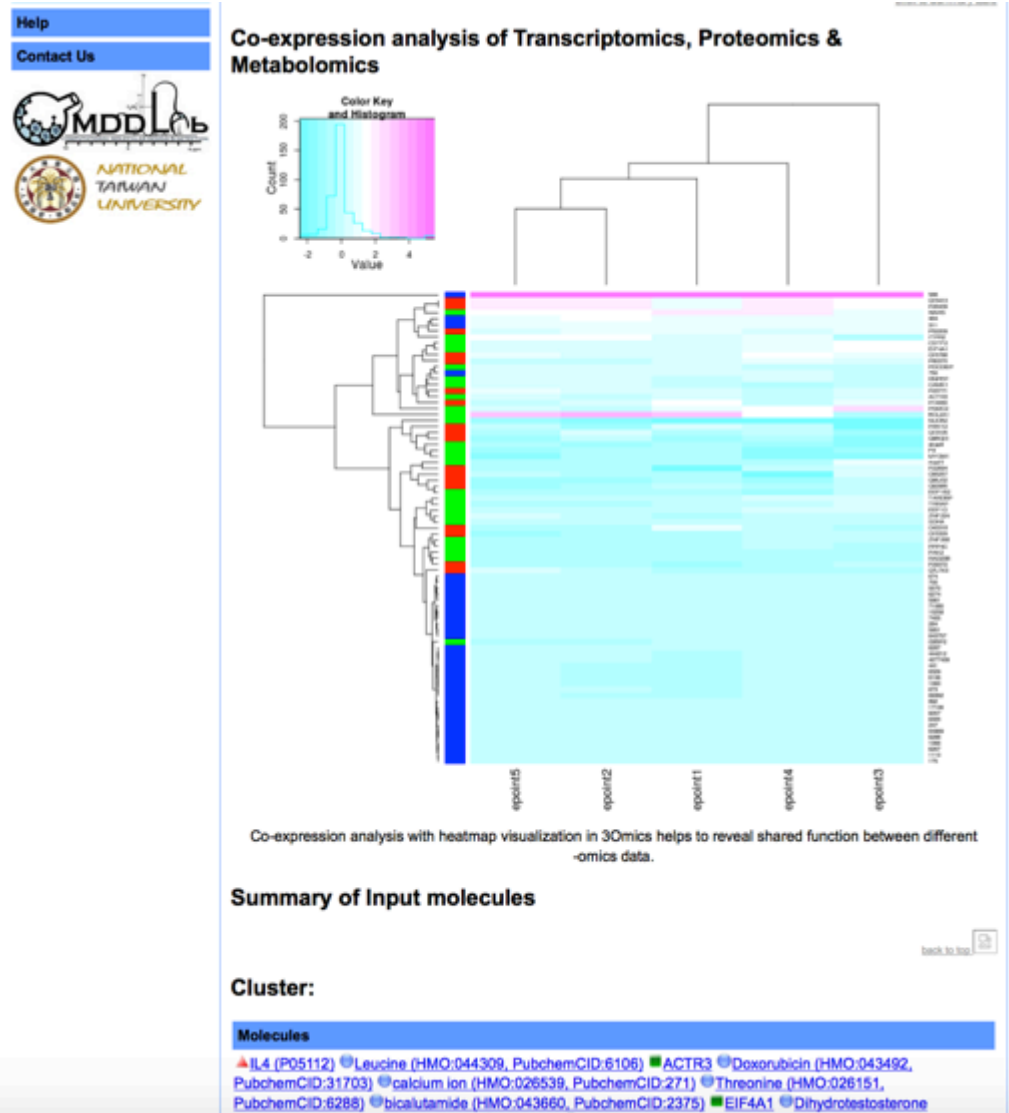
### Summary of Input molecules

Cluster:

Molecules

▲ IL4 (P05112) ● Leucine (HMO:044309, PubchemCID:6106) ■ ACTR3 ● Doxorubicin (HMO:043492, PubchemCID:31703) ● calcium ion (HMO:026539, PubchemCID:271) ● Threonine (HMO:026151, PubchemCID:6288) ● bicalutamide (HMO:043660, PubchemCID:2375) ■ EIF4A1 ● Dihydrotestosterone (HMO:025783, PubchemCID:10635) ■ PDCC6IP ● bortezomib (HMO:048610, PubchemCID:387447) ■ PSMC2

# Co-expression analysis



# Phenotype analysis



**Project Features**  
Overview  
Name-ID Converter  
**Help**  
Contact Us



**Correlation Network**  
**Coexpression Profile**  
**Phenotype Analysis**  
**Pathway Analysis**  
**GO Enrichment Analysis**

[Click to Summary table](#)

## Phenotype Analysis

A phenotype is defined as any observable characteristic or trait of an organism arising from gene expression, the influence of environmental factors, and the interactions between them. With phenotype-gene association from OMIM, genes and genetic disorders containing information to relate genes in the human genome with specific phenotypes can be identified.

The Transcriptomics data you've input have been used to search through the OMIM database, and the related phenotype and genes can be listed as below:  
Please click the link for description and molecular genetic information on OMIM website.

Human-related Phenotype	Related-Gene
<a href="#">[OMIM: 611820] LONG QT SYNDROME 11</a>	 <a href="#">akap9</a>
<a href="#">[OMIM: 256008] LEIGH SYNDROME</a>	 <a href="#">SDHA</a>
<a href="#">[OMIM: 612086] AMYOTROPHIC LATERAL SCLEROSIS 10, WITH OR WITHOUT FRONTOTEMPORAL DEMENTIA WITH TDP43 INCLUSIONS</a>	 <a href="#">TARDBP</a>
<a href="#">[OMIM: 306700] HEMOPHILIA A COAGULATION FACTOR VIII, INCLUDED</a>	 <a href="#">F8</a>

## Summary of Input molecules

[back to top](#)

### Cluster:

**Molecules**  
[▲ IL4 \(P05112\)](#) [● Leucine \(HMO:044309, PubchemCID:6106\)](#) [■ ACTR3](#) [● Doxorubicin \(HMO:043492, PubchemCID:31703\)](#) [● calcium ion \(HMO:026539, PubchemCID:271\)](#) [● Threonine \(HMO:026151, PubchemCID:6288\)](#) [● bicalutamide \(HMO:043660, PubchemCID:2375\)](#) [■ EIF4A1](#) [● Dihydrotestosterone \(HMO:025783, PubchemCID:10635\)](#) [■ PDCD6IP](#) [● bortezomib \(HMO:048610, PubchemCID:387447\)](#) [■ PSMC2](#) [● trigonelline \(HMO:033252, PubchemCID:5570\)](#) [■ TARDBP](#) [▲ RYR3 HBRR \(Q15413\)](#) [● dimethylamine \(HMO: PubchemCID:674\)](#) [▲ HSD3B1 3BH HSDB3A \(P14060\)](#) [● Hydrocortisone \(HMO:043177, PubchemCID:5754\)](#) [● Tyrosine \(HMO:026152, PubchemCID:6057\)](#) [● Methotrexate \(HMO:042925, PubchemCID:126941\)](#) [● formic acid \(HMO:044577, PubchemCID:284\)](#) [● hippuric acid \(HMO:033093, PubchemCID:464\)](#) [● Testosterone Propionate \(HMO:043961, PubchemCID:5995\)](#) [● Androsterone \(HMO:027989, PubchemCID:5879\)](#) [■ MYOM1](#) [● Leucine \(HMO:042148, PubchemCID:857\)](#) [● zinc fluoride \(HMO:040479, PubchemCID:24551\)](#) [● 3d0b \(HMO:049721, PubchemCID:24812721\)](#) [● Mifepristone \(HMO:043298, PubchemCID:55245\)](#) [▲ MAP3K7 TAK1 \(Q43318\)](#) [● Aconitic Acid \(HMO:033434, PubchemCID:444212\)](#) [● Indican \(HMO:049137, PubchemCID:10258\)](#) [● Estradiol \(HMO:026665, PubchemCID:5757\)](#) [● NTH \(HMO:049464, PubchemCID:5289054\)](#) [● Inositol \(HMO:036496,](#)

# Pathway analysis





Project Features

Overview

Name-ID Converter

Help

Contact Us

Correlation Network

Coexpression Profile

Phenotype Analysis

Pathway Analysis

GO Enrichment Analysis

Click to Summary table


### Pathway analysis - Normal, non-enrichment

[KEGG section](#) | [HumanCyc section](#)











#### KEGG Pathway analysis

back to top

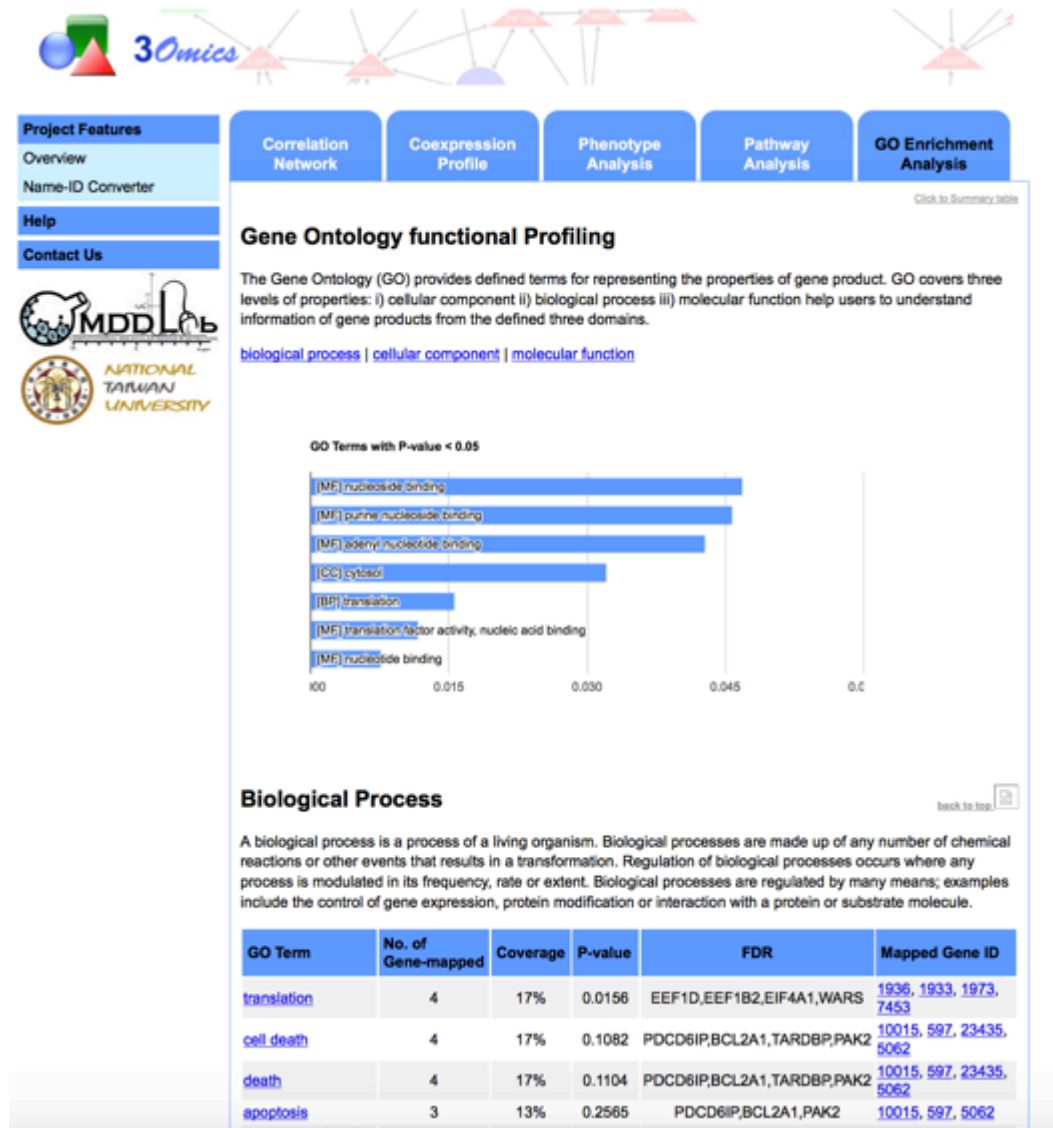
KEGG pathway enrichment analysis operates upon metabolomic data to reveal enriched pathways in a KEGG Pathway database by ranking the biological pathways commonly shared by metabolites.

The enriched KEGG metabolic pathways are listed on the bottom of the page.  
Please click  to see mapped pathway images on KEGG Pathway.

(Normal Mode) Show Records: 20

Metabolic Pathways	Hits
( hsa01100 ) Metabolic pathways - Homo sapiens (human)  <a href="#">Acetate</a> * <a href="#">D-Alanine</a> * <a href="#">L-Asparagine</a> * <a href="#">Betaine</a> * <a href="#">Citrate</a> * <a href="#">Ethanolamine</a> * <a href="#">Formate</a> * <a href="#">6-Deoxy-L-galactose</a> * <a href="#">L-Glutamine</a> * <a href="#">Glycine</a> * <a href="#">L-Histidine</a> * <a href="#">N,N-Dimethylglycine</a> * <a href="#">Pyruvate</a> * <a href="#">Pyridine-2,3-dicarboxylate</a> * <a href="#">L-Serine</a> * <a href="#">Succinate</a> * <a href="#">L-Tryptophan</a> * <a href="#">L-Tyrosine</a> * <a href="#">N(p)-Methyl-L-histidine</a>	19
( hsa00970 ) Aminoacyl-tRNA biosynthesis - Homo sapiens (human)  <a href="#">L-Asparagine</a> * <a href="#">L-Glutamine</a> * <a href="#">Glycine</a> * <a href="#">L-Histidine</a> * <a href="#">L-Serine</a> * <a href="#">L-Threonine</a> * <a href="#">L-Tryptophan</a> * <a href="#">L-Tyrosine</a>	8
( hsa00250 ) Alanine, aspartate and glutamate metabolism - Homo sapiens (human)  <a href="#">Acetate</a> * <a href="#">L-Asparagine</a> * <a href="#">L-Glutamine</a> * <a href="#">Glycine</a> * <a href="#">Pyruvate</a> * <a href="#">Succinate</a> * <a href="#">L-Tyrosine</a>	7
( hsa00280 ) Valine, leucine and isoleucine degradation - Homo sapiens (human)  <a href="#">Acetate</a> * <a href="#">Glycine</a> * <a href="#">Pyruvate</a> * <a href="#">Succinate</a> * <a href="#">L-Tryptophan</a> * <a href="#">L-Tyrosine</a>	6
( hsa00270 ) Cysteine and methionine metabolism - Homo sapiens (human)  <a href="#">Betaine</a> * <a href="#">N,N-Dimethylglycine</a> * <a href="#">Pyruvate</a> * <a href="#">L-Serine</a> * <a href="#">L-Tryptophan</a> * <a href="#">L-Tyrosine</a>	6
( hsa00330 ) Arginine and proline metabolism - Homo sapiens (human)  <a href="#">Acetate</a> * <a href="#">L-Glutamine</a> * <a href="#">Glycine</a> * <a href="#">Pyruvate</a> * <a href="#">Succinate</a> * <a href="#">L-Tyrosine</a>	6
( hsa00360 ) Phenylalanine metabolism - Homo sapiens (human)  <a href="#">Acetate</a> * <a href="#">Glycine</a> * <a href="#">L-Histidine</a> * <a href="#">L-Tryptophan</a> * <a href="#">L-Tyrosine</a>	5
( hsa00340 ) Histidine metabolism - Homo sapiens (human)  <a href="#">Acetate</a> * <a href="#">L-Histidine</a> * <a href="#">L-Tryptophan</a> * <a href="#">L-Tyrosine</a> * <a href="#">N(p)-Methyl-L-histidine</a>	5
( hsa00260 ) Glycine, serine and threonine metabolism - Homo sapiens (human)  <a href="#">Betaine</a> * <a href="#">Glycine</a> * <a href="#">N,N-Dimethylglycine</a> * <a href="#">Pyruvate</a> * <a href="#">L-Serine</a>	5
( hsa00520 ) Amino sugar and nucleotide sugar metabolism - Homo sapiens (human) 	4

# GO Enrichment Analysis



# Multivariate methods

# Generating relevance network using sPLS or msPLS techniques (Cao 2009, Liquet 2012)

- sparse partial least squares (sPLS) regression or multilevel partial least squares (msPLS) method
- One-step procedure for variable selection as well as integration
- Comparison of different multivariate integration techniques showed that sPLS generates (Cao 2009)
- msPLS – for repeated measures
- Implemented in the R package mixOmics
- Generates association matrix and allows visualization of associations using bipartite relevance networks (Liquet 2012)



# sPLS method

- sPLS is a variable selection and dimensionality reduction method that allows integration of heterogeneous omics data from same set of samples
- Robust approximation of Pearson correlation using regression and latent (principal) variates
- Eg: transcriptome (matrix X) and metabolome (matrix Y) data  
where,  
matrix X is an  $n \times p$  matrix that includes  $n$  samples and  $p$  metabolites  
matrix Y is an  $n \times q$  matrix that includes  $n$  samples and  $q$  genes

Objective function

$\max \text{cov}(X_u, Y_v)$

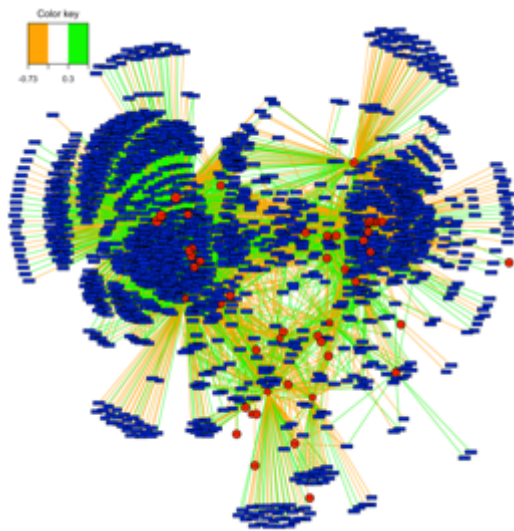
where

$u_1, u_2 \dots u_H$  and  $v_1, v_2 \dots v_H$  are the loading vectors

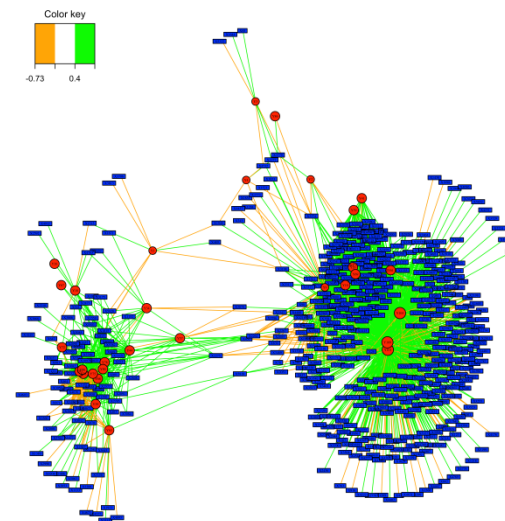
H is the number of PLS-DA dimensions

A Lasso based optimization is used to select most relevant variables

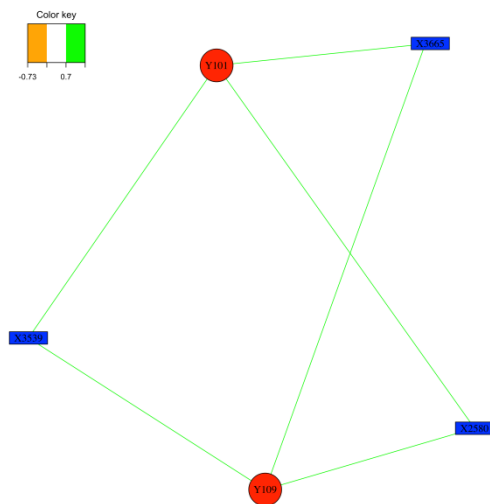
**Case Study 2: Application of sPLS technique for integrative –omics.** Microbiome-Metabolome Wide Association Study of Lung BAL: Global integration of 5930 m/z features with 153 microbial species using sparse Partial Least Squares regression (Cribbs et al. Microbiome 2016)



A. Association threshold: 0.3



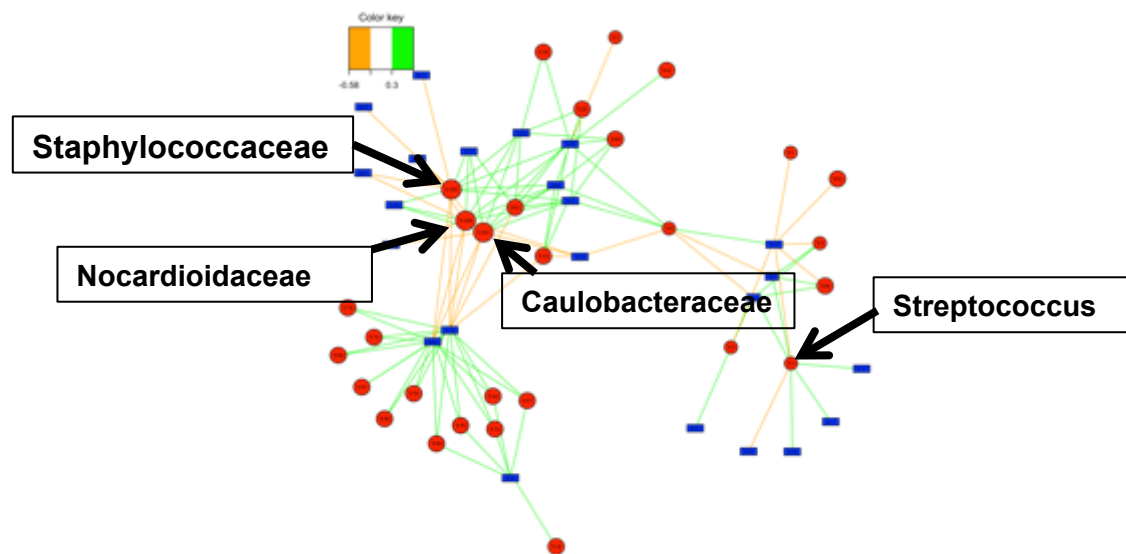
B. Association threshold: 0.4



C. Association threshold: 0.7

## Legend

Circles: microbial species  
Rectangles: metabolome features



D. Using only subset of metabolic features also associated with HIV status (+ve or -ve)

# Integrating more than two datasets

Proteomics data  
(n subjects X s proteins)

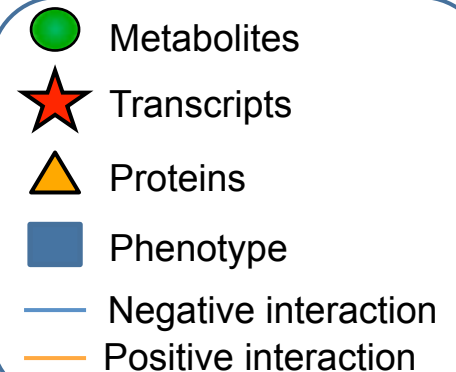
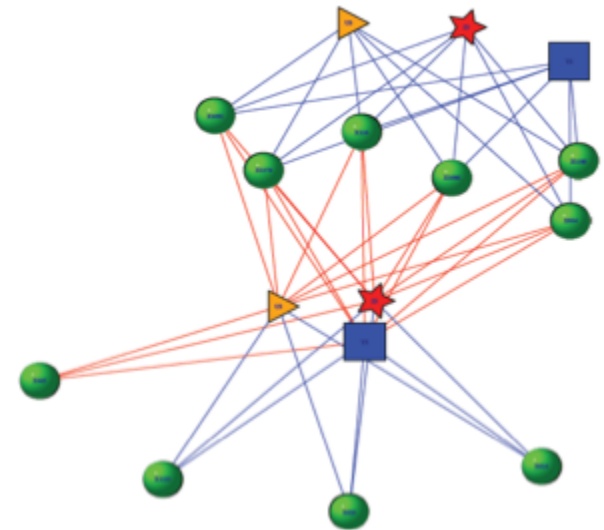
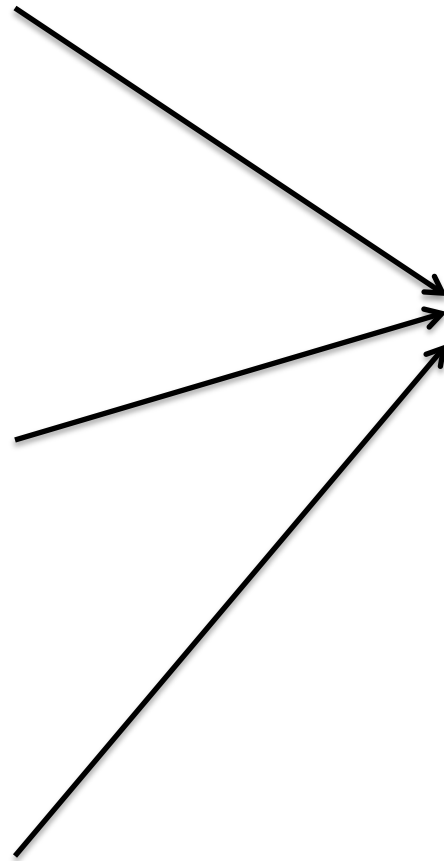
	E1	E2	-	Es
Subject1	199	19	-	100
Subject2	10	40		90
-	-	-		-
SubjectN	50	30	-	20

Metabolomics data  
(n subjects X p metabolites)

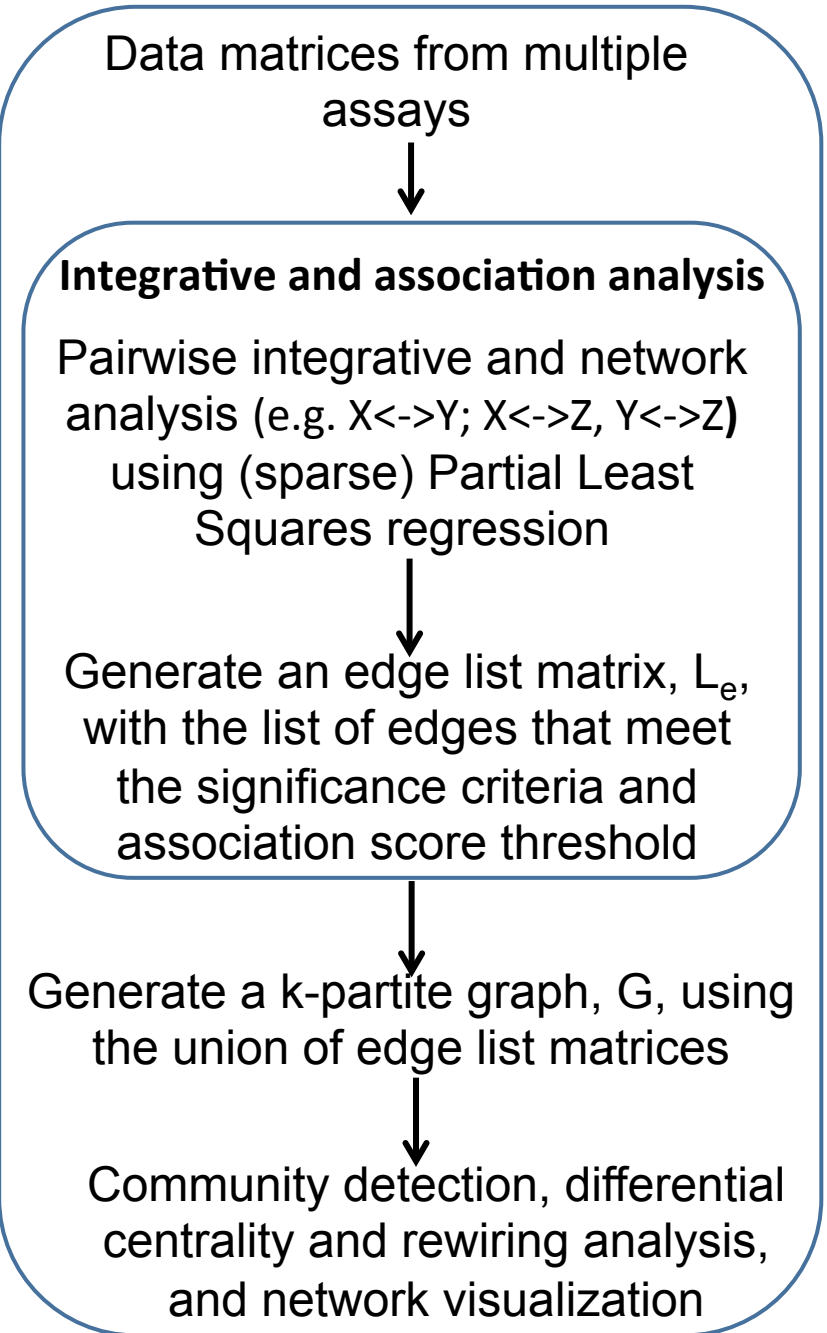
	M1	M2	-	Mp
Subject1	199	19	-	100
Subject2	10	40		90
-	-	-		-
SubjectN	50	30	-	20

Transcriptomics data  
(n subjects X q genes)

	G1	G2	-	Gq
Subject1	19	19	-	100
Subject2	10	40	-	90
-	-	-	-	-
SubjectN	10	40	-	50



A.

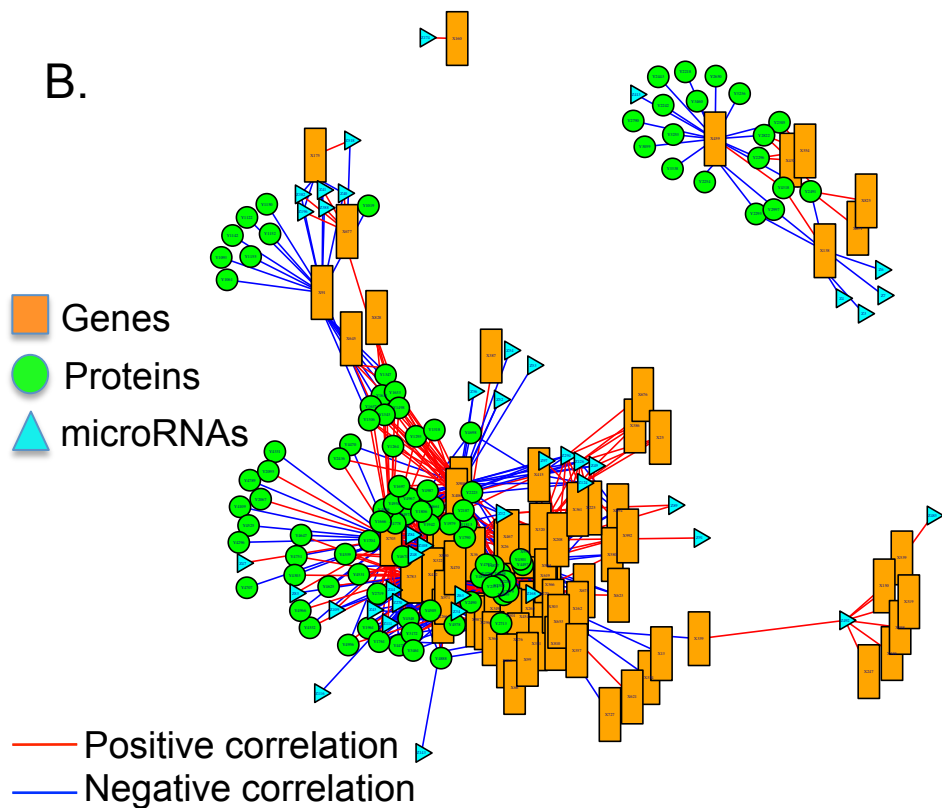


# xMWAS: R package for data integration and differential network analysis

(Uppal et al. Submitted to Bioinformatics)

URL: <https://sourceforge.net/projects/xmwas/>

B.



**Case Study 3: Application of xMWAS for integrative –and differential network analysis of more than 2 dataests.**

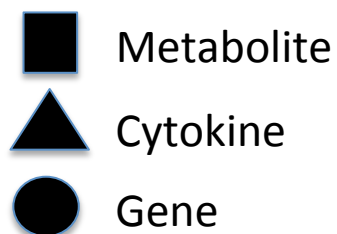
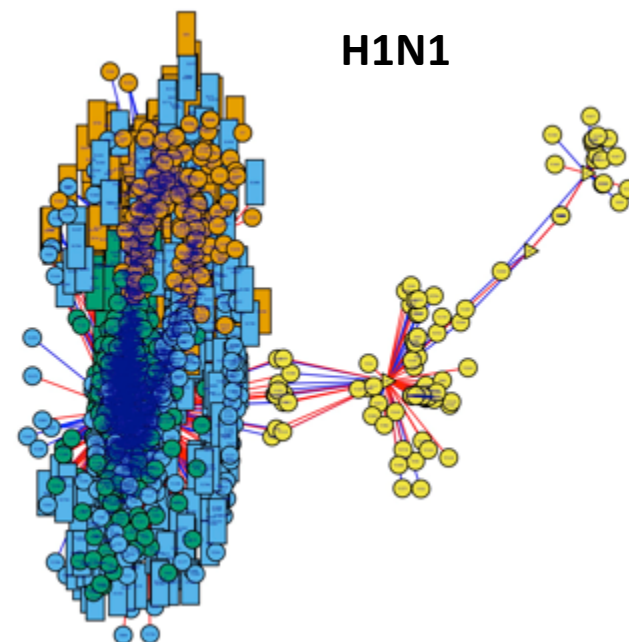
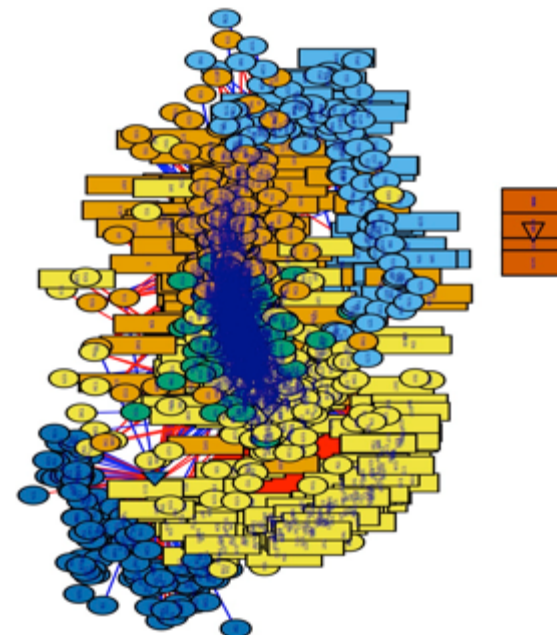
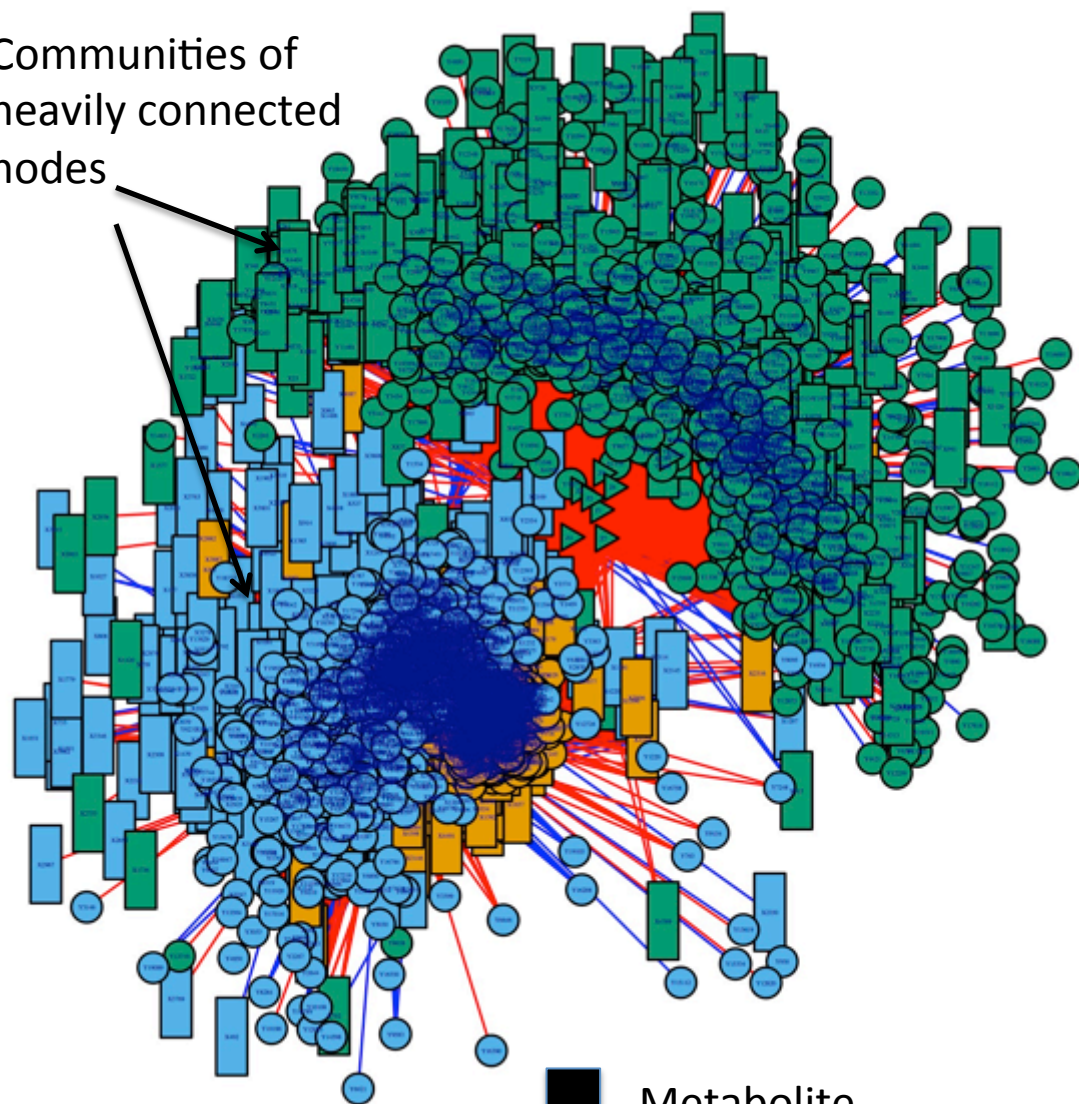
Integrative network analysis of cytokine, metabolome and transcriptome datasets from a study of H1N1 virus infection of mice (Chandler et al. 2016)

**All samples (Control + H1N1)**

**Control**

**H1N1**

Communities of heavily connected nodes



Pathway Name	IL-1beta	IL-6	IL-10	TNFAalpha	IP-1	IFNgamma
Immune response_Alternative complement pathway	X	X	X	X	X	X
GTP metabolism	X	X	X	X	X	
Cytoskeleton remodeling_TGF, WNT and cytoskeletal remodeling	X	X		X	X	X
Cytoskeleton remodeling_Cytoskeleton remodeling	X	X		X	X	X
Alternative complement cascade disruption in age-related macular degeneration	X	X		X	X	X
ATP metabolism	X	X		X	X	
Immune response_Lectin induced complement pathway		X	X		X	X
Regulation of lipid metabolism_Regulation of lipid metabolism by niacin and isoprenaline	X	X		X		
Development_Oligodendrocyte differentiation from adult stem cells		X			X	X
Cell adhesion_Chemokines and adhesion	X			X		
Expression targets of Tissue factor signaling in cancer	X			X		
Cell adhesion_ECM remodeling		X				X
Immune response_Classical complement pathway			X			X

**Pathway analysis of genes** that were found to be significantly associated with the six cytokines at *p*<0.05 using MetaCore



Pathway Name	IL-1beta	IL-6	IL-10	TNFalpha	IP-1	IFNgamma
Carnitine shuttle	X	X	X	X	X	
Glycosphingolipid metabolism	X	X	X	X	X	
Vitamin D <sub>3</sub> (cholecalciferol) metabolism	X	X	X	X	X	
Vitamin B <sub>6</sub> (pyridoxine) metabolism		X		X	X	X
Biopterin metabolism	X	X		X	X	
Saturated fatty acids beta-oxidation		X		X	X	
Vitamin E metabolism	X		X	X		
Nitrogen metabolism		X			X	
Urea cycle/amino group metabolism					X	X
Drug metabolism - cytochrome P450					X	X
Tryptophan metabolism					X	X
Arginine and proline metabolism						X
Bile acid biosynthesis				X		
Fatty acid metabolism			X			
Limonene and pinene degradation			X			
Linoleate metabolism			X			
Tyrosine metabolism		X				

**Pathway analysis of metabolic features** that were found to be significantly associated with the six cytokines at  $p<0.05$  using Mummichog

# Main approaches for data integration

- Pathway or knowledge-based integration
  - Datasets are analyzed individually (differentially expressed genes, metabolites, proteins) and integration is performed at the pathway level
  - Examples: MetaboAnalyst, iPEAP, MetScape
- Data-driven integration using meta-dimensional analysis
  - Integration is performed globally such that data from multiple omics layers are combined simultaneously
  - Examples: 3Omics, mixOmics, xMWAS
- **Using literature-derived associations for integration**
  - **Using co-occurrence criteria for establishing relationship**
  - **Examples: HiPub, CoPub, ArrowSmith**

# Text mining tools for literature-based relation discovery biomedical text

NCBI Resources How To

PubMed.gov  
US National Library of Medicine  
National Institutes of Health

PubMed "breast cancer" Search

Create RSS Create alert Advanced

Article types  
Clinical Trial  
Review  
Customize ...

Summary • 20 per page • Sort by Most Recent • Send to: ▼

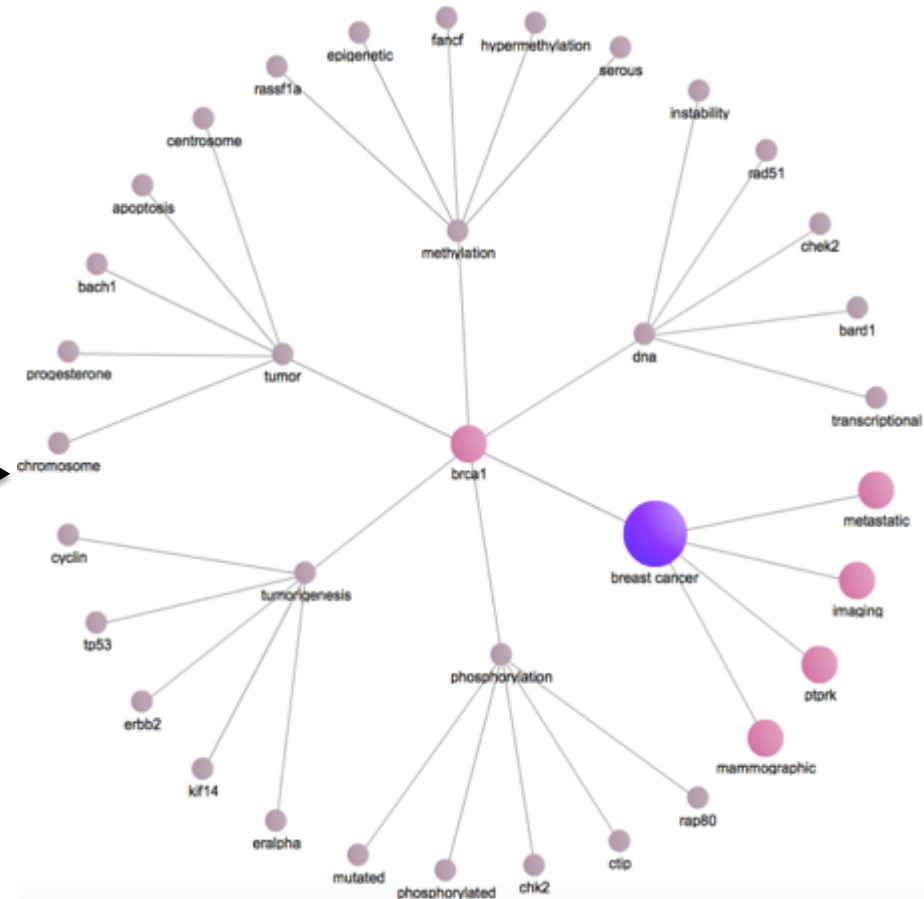
Results: 1 to 20 of 191685

<< First < Prev Page 1 of 9585 Next > Last >>

☐ Targeting ceramide metabolic pathway induces apoptosis in human breast cancer cell lines.

1. Vethakanraj HS, Babu TA, Sudarsanan GB, Duraisamy PK, Kumar SA.  
Biochem Biophys Res Commun. 2015 Jul 15; pii: S0006-291X(15)30278-3. doi: 10.1016/j.bbrc.2015.07.047. [Epub ahead of print]  
PMID: 26188095  
[Similar articles](#)

PubMed  
Crossref



# HiPub (Lee 2016):

## <http://hipub.korea.ac.kr/>

Secure <https://www.ncbi.nlm.nih.gov/pubmed/24009732>

NCBI Resources How To Sign in to NCBI

PubMed US National Library of Medicine National Institutes of Health Advanced Search Help

Format: Abstract

Send to

Full text links

Save items

Similar articles

Cited by 5 PubMed Central articles

**The conformational control inhibitor of tyrosine kinases DCC-2036 is effective for imatinib-resistant cells expressing T674I FIP1L1-PDGFR $\alpha$ .**

Shen Y<sup>1</sup>, Shi X, Pan J.

Author information

**Abstract**

The cells expressing the T674I point mutant of FIP1-like-1-platelet-derived growth factor receptor alpha (FIP1L1-PDGFR $\alpha$ ) in hyper-eosinophilia syndrome (HES) are resistant to imatinib and some second-generation tyrosine kinase inhibitors (TKIs). There is a desperate need to develop therapy to combat this acquired drug resistance. DCC-2036 has been synthesized as a third-generation TKI to combat especially the Bcr-Abl T315I mutant in chronic myeloid leukemia. This study evaluated the effect of DCC-2036 on FIP1L1-PDGFR $\alpha$ -positive cells, including the wild type (WT) and the T674I mutant. The in vitro effects of DCC-2036 on the PDGFR $\alpha$  signal pathways, proliferation, cell cycling and apoptosis of FIP1L1-PDGFR $\alpha$ -positive cells were investigated, and a nude mouse xenograft model was employed to assess the in vivo antitumor activity. We found that DCC-2036 decreased the phosphorylated levels of PDGFR $\alpha$  and its downstream targets without apparent effects on total protein levels. DCC-2036 inhibited proliferation, and induced apoptosis with MEK-dependent up-regulation of the pro-apoptotic protein Bim in FIP1L1-PDGFR $\alpha$ -positive cells. DCC-2036 also exhibited in vivo antineoplastic activity against cells with T674I FIP1L1-PDGFR $\alpha$ . In summary, FIP1L1-PDGFR $\alpha$ -positive cells are sensitive to DCC-2036 regardless of their sensitivity to imatinib. DCC-2036 may be a potential compound to treat imatinib-resistant HES.

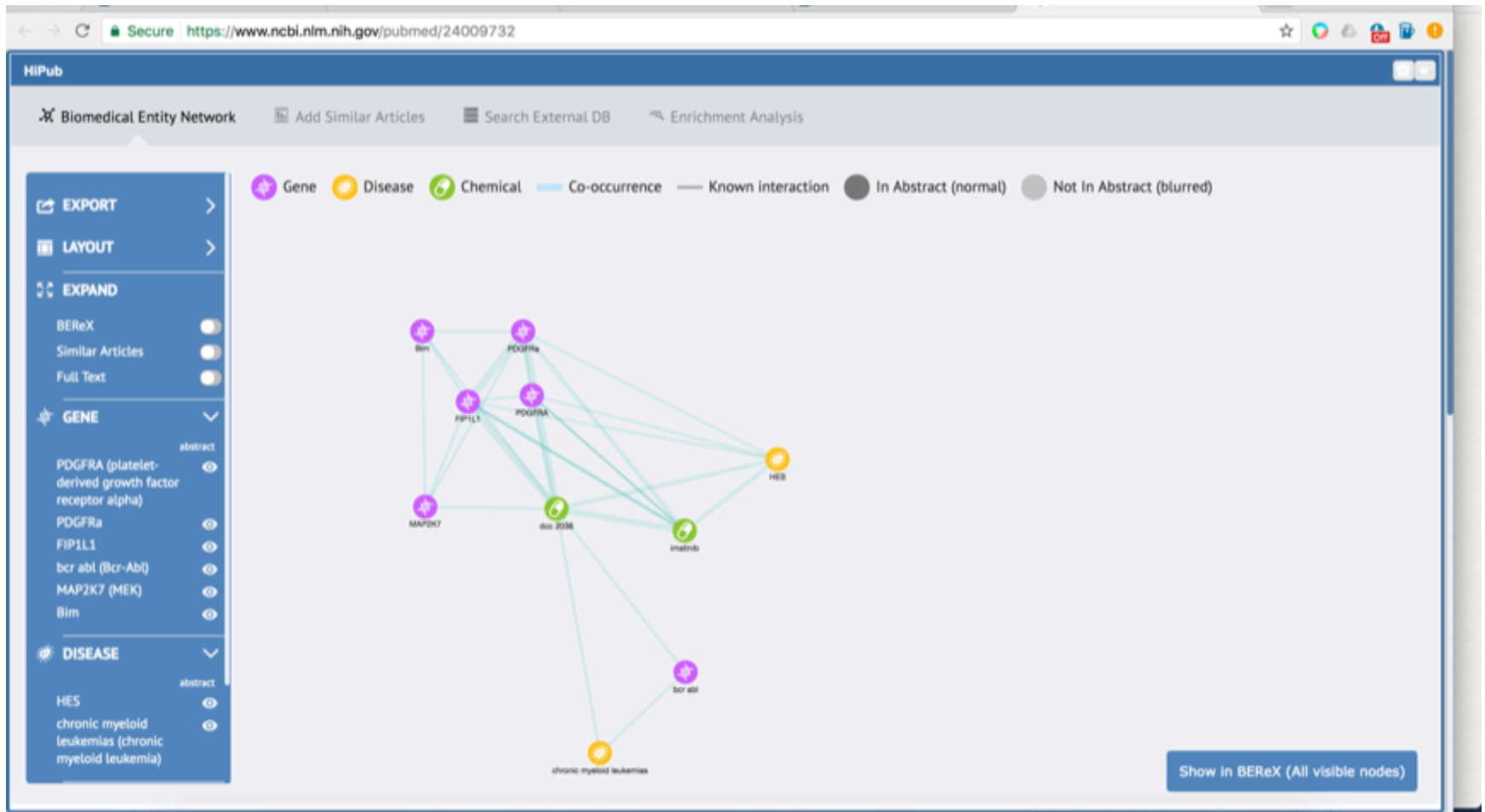
PMID: 24009732 PMCID: PMC3756952 DOI: 10.1371/journal.pone.0073059

[Indexed for MEDLINE] Free PMC Article

Images from this publication. See all images (7) Free text

HiPub

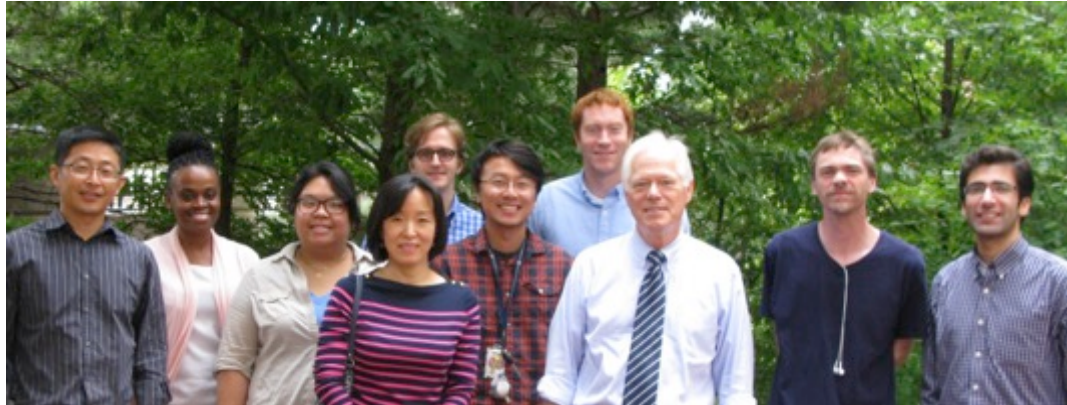
# HiPub (Lee 2016):



# Summary

- Various tools and techniques are available for integrating and visualization multi –omics data
- Integrative –omics drives systems biology and could play a critical role in personalized medicine

## Clinical Biomarkers Laboratory



Dean Jones, Young-Mi Go, Shuzaho Li, Karan Uppal, Douglas Walker, Josh Chandler, Sophia Banton, Ken Liu, Vilinh Tran, Michael Orr, Bill Liang (not shown)

Lab website: <http://clinicalmetabolomics.org/>

Funding: ES025632, ES023485, ES019776,  
OD018006, HL095479, EY022618,  
HHSN272201200031C

# Questions?