



EMORY

# Computational methods for data integration

March 21, 2018

Karan Uppal, PhD

Assistant Professor of Medicine

Director of Computational Metabolomics & Integrative Omics,  
Clinical Biomarkers Laboratory

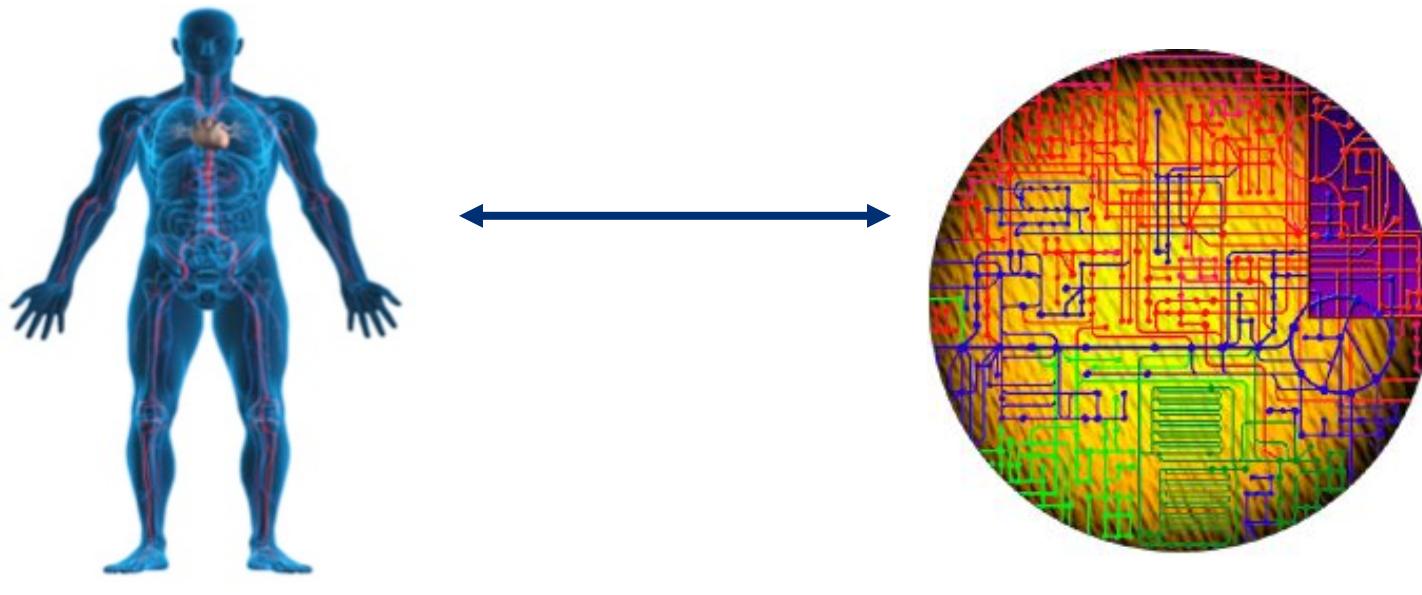
Emory University

# 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

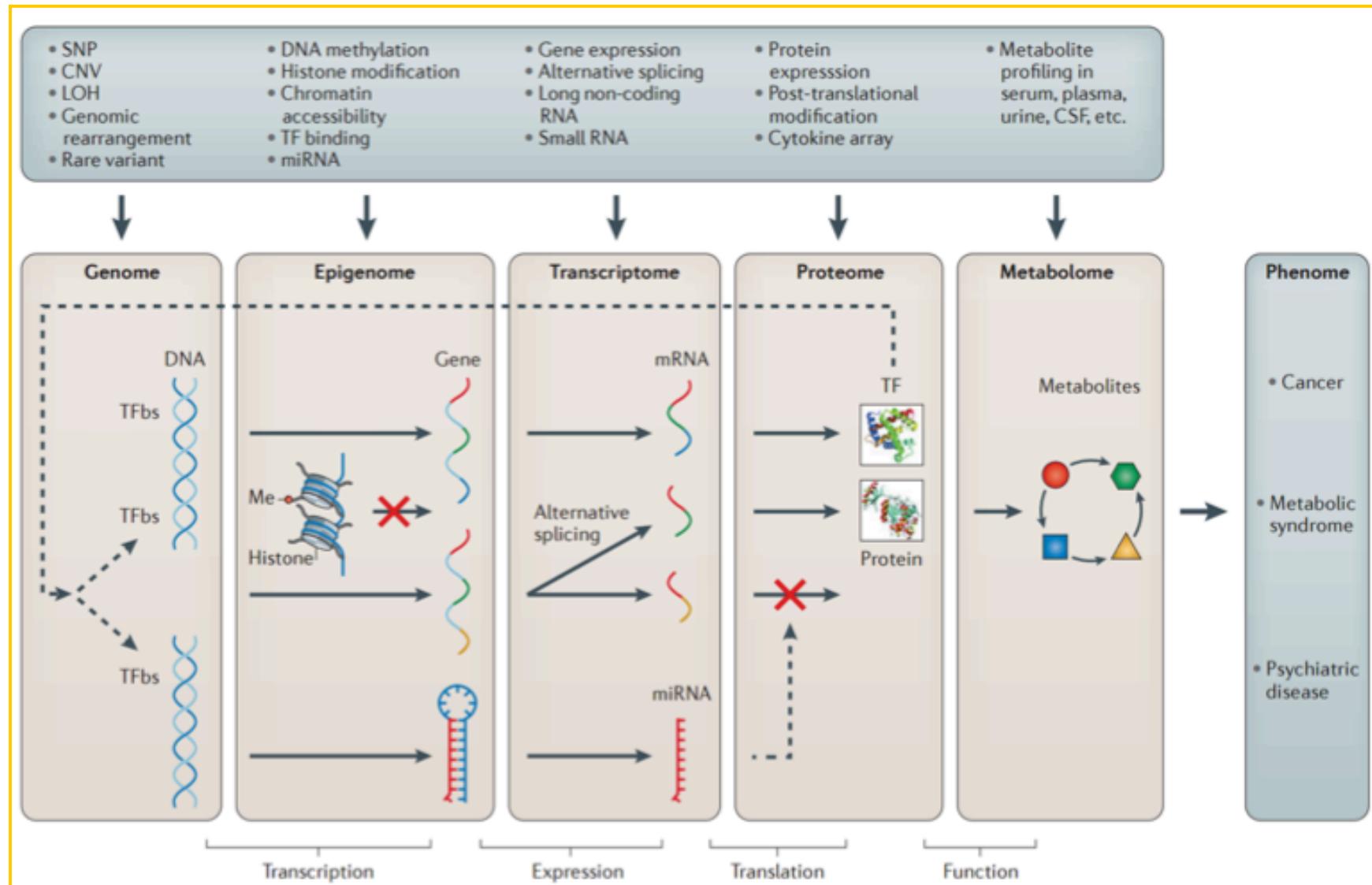


Exposures  
Internal measurements  
Disease states

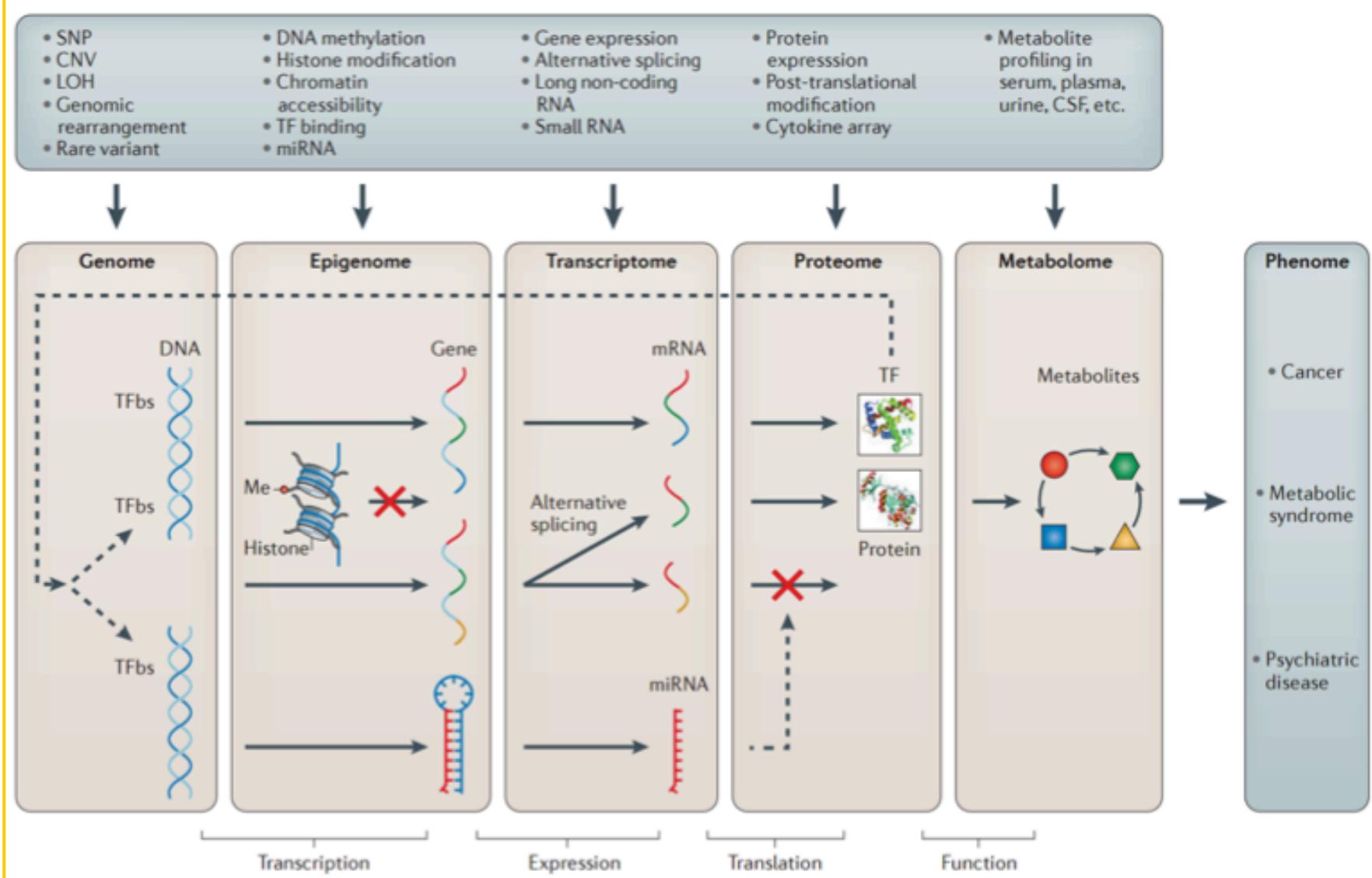
**Systems Biology**  
*“Integrative approach in which scientists study pathways and networks will touch all areas of biology, including drug discovery”*

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, 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

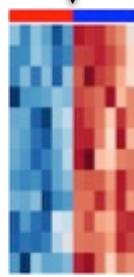
# Main approaches for data integration

- **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	-	M <sub>p</sub>
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	-	G <sub>q</sub>
Subject1	19	19	-	100
Subject2	10	40	-	90
-	-	-	-	-
SubjectN	10	40	-	50



Differentially expressed genes

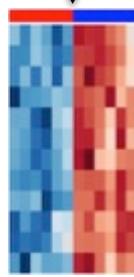
Pathway analysis for genes

Common pathways or pathway rank aggregation

# Pathway or knowledge-based integration

Metabolomics data  
(n subjects X p metabolites)

	M1	M2	-	M <sub>p</sub>
Subject1	199	19	-	100
Subject2	10	40		90
-	-	-		-
SubjectN	50	30	-	20



Differentially expressed metabolites

Transcriptomics data  
(n subjects X q genes)

	G1	G2	-	G <sub>q</sub>
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/faces/ModuleView.xhtml>)

The screenshot shows the MetaboAnalyst 4.0 homepage. On the left, there's a sidebar with links to Overview, Data Formats, FAQs, Tutorials, Troubleshooting, Resources, Update History, User Stats, and About. Logos for GenomeCanada, GenomeQuébec, and NSERC/CRSNG are also present. The main content area features a large circular diagram with various analysis modules. A legend at the bottom identifies the colors: blue for Targeted or untargeted metabolomics, green for Targeted metabolomics only, cyan for Untargeted metabolomics, and striped patterns for Multiple metabolomics data and Integrating other omics. The modules shown are Statistical Analysis, Enrichment Analysis, Pathway Analysis, Network Explorer, MS Peaks to Pathways, Other Utilities, Spectral Analysis, Biomarker Meta-analysis, Power Analysis, Time-series / Two-factor, Biomarker Analysis, and Joint Pathway Analysis. The 'Joint Pathway Analysis' module is highlighted with a red border.

MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation

Click a module to proceed, or scroll down for more details:

- Targeted or untargeted metabolomics
- Targeted metabolomics only
- Untargeted metabolomics
- Multiple metabolomics data
- Integrating other omics

Joint Pathway Analysis

Modules shown in the diagram include:  
Statistical Analysis, Enrichment Analysis, Pathway Analysis, Network Explorer, MS Peaks to Pathways, Other Utilities, Spectral Analysis, Biomarker Meta-analysis, Power Analysis, Time-series / Two-factor, Biomarker Analysis, and Joint Pathway Analysis.

Xia Lab @ McGill (last updated 2018-03-10)

# Upload data and submit

The screenshot shows the MetaboAnalyst 4.0 web interface. At the top, there is a navigation bar with icons for back, forward, search, and other browser functions. The URL is [www.metaboanalyst.ca/faces/upload/JointUploadView.xhtml](http://www.metaboanalyst.ca/faces/upload/JointUploadView.xhtml). On the left, there is a sidebar with a logo for "MetaboAnalyst 4.0" and links for "Upload", "Integrative Analysis", "Download", and "Exit". The main content area has a title "MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation". A message says "Please upload a gene list and metabolite list to perform joint pathway analysis. Use the Pathway Analysis module, if you only have a metabolite list." Below this, there are two sections: "Gene List" and "Metabolite List". Each section contains a text area with sample data and an "ID Type" dropdown menu. The "Gene List" section contains:

```
#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
```

The "Metabolite List" section contains:

```
#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
```

At the bottom, there is a "Specify organism" dropdown set to "Homo sapiens (human)", a checked checkbox for "Use our example data", and a blue "Submit" button.

# Name Mapping – click Submit

The system requires all the IDs (except common compound names) to be matched exactly. The table below shows the matched genes and compounds from the underlying databases. For common compound names, users can further perform [approximate match](#) by clicking the [View](#) link in the Details column. To remove a gene or compound from further analysis, use the [Delete](#) link in the last column.

**Compound Name Mapping**   **Gene Name Mapping**

Query	Hit	HMDB	KEGG	Details	
C00116	Glycerol	<a href="#">HMDB0000131</a>	<a href="#">C00116</a>		<a href="#">Delete</a>
C00565	Trimethylamine	<a href="#">HMDB0000906</a>	<a href="#">C00565</a>		<a href="#">Delete</a>
C00033	Acetic acid	<a href="#">HMDB0000042</a>	<a href="#">C00033</a>		<a href="#">Delete</a>
C00583	Propylene glycol	<a href="#">HMDB0001881</a>	<a href="#">C00583</a>		<a href="#">Delete</a>
C00022	Pyruvic acid	<a href="#">HMDB0000243</a>	<a href="#">C00022</a>		<a href="#">Delete</a>
C00719	Betaine	<a href="#">HMDB0000043</a>	<a href="#">C00719</a>		<a href="#">Delete</a>
C05984	2-Hydroxybutyric acid	<a href="#">HMDB0000008</a>	<a href="#">C05984</a>		<a href="#">Delete</a>
C00207	Acetone	<a href="#">HMDB0001659</a>	<a href="#">C00207</a>		<a href="#">Delete</a>
C00065	L-Serine	<a href="#">HMDB0000187</a>	<a href="#">C00065</a>		<a href="#">Delete</a>
C00031	D-Glucose	<a href="#">HMDB0000122</a>	<a href="#">C00031</a>		<a href="#">Delete</a>
C00079	L-Phenylalanine	<a href="#">HMDB0000159</a>	<a href="#">C00079</a>		<a href="#">Delete</a>

**R Command History**

```
Keep collapsed Save
```

```
1. InitDataObjects("conco", "pathinteg");
2. mSet<-SetOrganism(mSet, "hsa")
3. geneListFile<-"replace_with_your_file"
4. geneList<-readChar(geneListFile,
  file.info(geneListFile)$size)
5. mSet<-PerformIntegGeneMapping(mSet,
  geneList, "hsa", "entrez");
6. cmpdListFile<-"replace_with_your_file"
7. cmpdList<-readChar(cmpdListFile,
  file.info(cmpdListFile)$size)
8. mSet<-PerformIntegCmpdMapping(mSet,
  cmpdList, "hsa", "kegg");
9. mSet<-CreateMappingResultTable(mSet)
```

# Set parameters

MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation

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 hypergeometric 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

R Command History

Keep collapsed  Save

```
1. InitDataObjects("oco", "pathinteg", 1
2. mSet<-SetOrganism(mSet, "hsa")
3. geneListFile<-replace_with_your_file
4. geneList<-readChar(geneListFile,
file.info(geneListFile)$size)
5. mSet<-PerformIntegGeneMapping(mSet,
geneList, "hsa", "entrez")
6. cmpdListFile<-replace_with_your_file
7. cmpdList<-readChar(cmpdListFile,
file.info(cmpdListFile)$size)
8. mSet<-PerformIntegCmpdMapping(mSet,
cmpdList, "hsa", "kegg")
9. mSet<-CreateMappingResultTable(mSet)
10. mSet<-PrepareIntegData(mSet);
```

# Results - Overview

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

**Pathway Analysis Overview**

Pathway	Enrichment (Blue)	Topology (Yellow)	Total
Pyruvate metabolism	~3800	~1200	~5000
Glycosphingolipid biosynthesis - lacto and neolacto series	~3500	~1000	~4500
Glycosphingolipid biosynthesis - globo series	~3000	~1500	~4500
Glycine, serine and threonine metabolism	~2500	~1500	~4000
Lipoic acid metabolism	~2500	~2500	~5000
Glycolysis / Gluconeogenesis	~2000	~2000	~4000
Biosynthesis of unsaturated fatty acids	~1500	~5000	~6500
Glyoxylate and dicarboxylate metabolism	~1500	~1500	~3000
Citrate cycle (TCA cycle)	~1000	~1500	~2500
Glycerophospholipid metabolism	~1000	~1500	~2500
Arginine and proline metabolism	~1000	~500	~1500
Cysteine and methionine metabolism	~500	~500	~1000
Ether lipid metabolism	~500	~3500	~4000
Valine, leucine and isoleucine degradation	~500	~500	~1000
Selenocompound metabolism	~500	~200	~700
Sphingolipid metabolism	~500	~200	~700
Primary bile acid biosynthesis	~500	~100	~600
Steroid biosynthesis	~500	~100	~600
Glycosaminoglycan degradation	~500	~100	~600
Glycerolipid metabolism	~500	~100	~600

**R Command History**

```
Keep collapsed
1. InitiateAnalysis("comc", "pathinteg", R)
2. mNet<-detOrgani(mNet, "hsa")
3. geneListFile<-replace_with_your_file_
4. geneList<-readChar(geneListFile,
file.info(geneListFile)$size)
5. mNet<-PerformIntegGeneMapping(mNet,
geneList, "hsa", "entrez")
6. cmpdListFile<-replace_with_your_file_
7. cmpdList<-readChar(cmpdListFile,
file.info(cmpdListFile)$size)
8. mNet<-PerformIntegCpdMapping(mNet,
cmpdList, "hsa", "kegg")
9. mNet<-CreateMappingResultTable(mNet)
10. mNet<-PrepareIntegBeta(mNet)
11. mNet<-PerformIntegPathwayAnalysis(mNet,
"dc", "hyper", "integ")
12. mNet<-PlotIntegPath(mNet, "heat0920",
513)
```

**Click on “View Details” for detailed results**

# Results - Details

MetaboAnalyst -- a comprehensive tool for metabolomics analysis and interpretation

R Command History

```
Keep collapsed  Save
1. InitiateObjects("conco", "pathinteg");
2. mset<-SetOrganism(mset, "hsa")
3. geneListFile<-`replace_with_your_file
4. geneList<-readChar(geneListFile,
file.info(geneListFile)$size)
5. mset<-PerformIntegGeneMapping(mset,
geneList, "hsa", "entrez");
6. compListFile<-`replace_with_your_file
7. compList<-readChar(compListFile,
file.info(compListFile)$size)
8. mset<-PerformIntegCompMapping(mset,
compList, "hsa", "keep");
9. mset<-CreateMappingResultTable(mset);
10. mset<-PrepareIntegData(mset);
11. mset<-PerformIntegPathwayAnalysis(mset,
"de", "hyper", "integ");
12. mset<-PlotInnessPath(mset, "hsa03620",
513)
```

Pyruvate metabolism (KEGG)

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

Pathway	Total	Expected	Hits	PValue	Topology	View
Pyruvate metabolism	41	1.1233	4	0.023442	0.3913	<a href="#">View</a>
Glycosphingolipid biosynthesis - lacto and neolacto series	26	0.71233	3	0.031941	0.22034	<a href="#">View</a>
Glycosphingolipid biosynthesis - globo series	13	0.35616	2	0.047147	0.44444	<a href="#">View</a>
Other sulfur and chlorine metabolism	99	4.0497	0	0.073306	0.36006	<a href="#">View</a>

Fit

Page 1 of 1000 Last modified 2018-03-10

# 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

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)

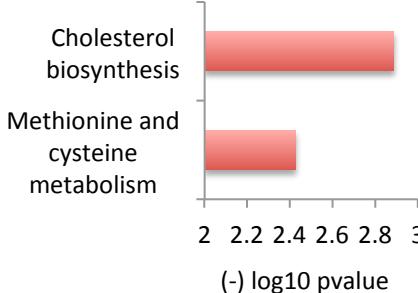
	G1	G2	-	Gn
Subject1	19	19	-	100
Subject2	10	40	-	90
-	-	-	-	-
SubjectN	10	40	-	50

Association matrix

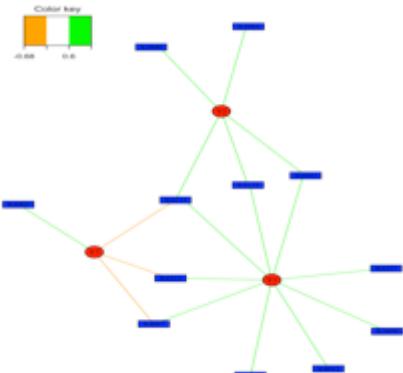
	G1	G2	-	Gn
M1	0.4	0.9	-	0.3
M2	0.7	0.1	-	0.5
M3	0.1	0.6		0.8

# Workflow

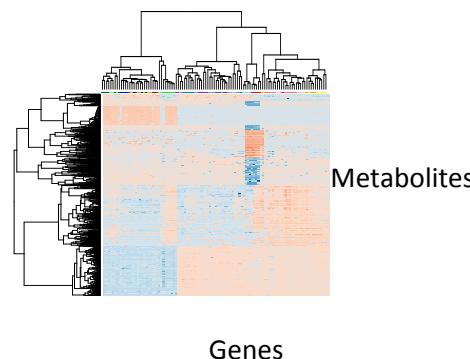
## Pathway enrichment



## Relevance networks



## Clustering



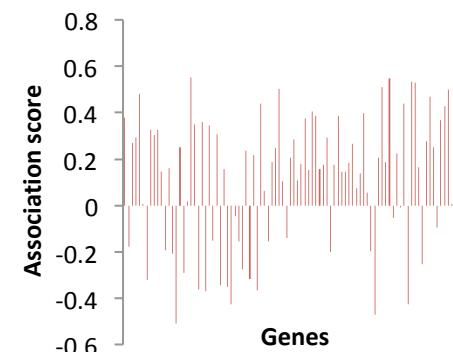
## Univariate

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

## Multivariate

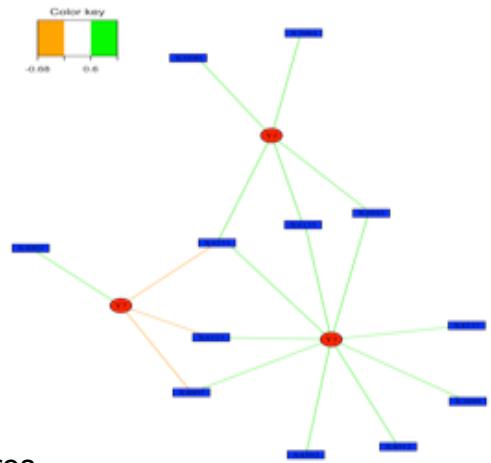
- PLS, CCA, sparse PLS
- Tools: mixOmics (Cao 2009), etc.

## 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 a association threshold (e.g. 0.5) to visualize positive or negative associations



# 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 2018): 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

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.

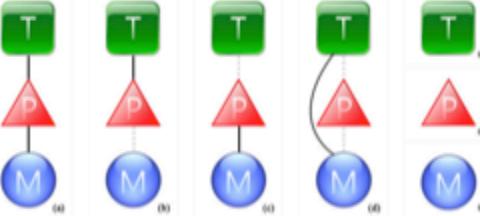
Please select the desired analysis:

a. Transcriptome-Proteome-Metabolome  
b. Transcriptome-Proteome

c. Proteome-Metabolome  
d. Transcriptome-Metabolome

e. Transcriptome only  
f. Proteome only  
g. Metabolome only

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






Copyright (c) 2006-2012 Computational Molecular Design & Metabolomics Laboratory | BEBI | 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>)

Variables	Samples				
	timepoint1	timepoint2	timepoint3	timepoint4	timepoint5
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 Repulsion: 160

Correlation Network Attraction: 80

Refresh

MDDLab  
NATIONAL TAIWAN UNIVERSITY

### Correlation Network of Transcriptomics, Proteomics & Metabolomics

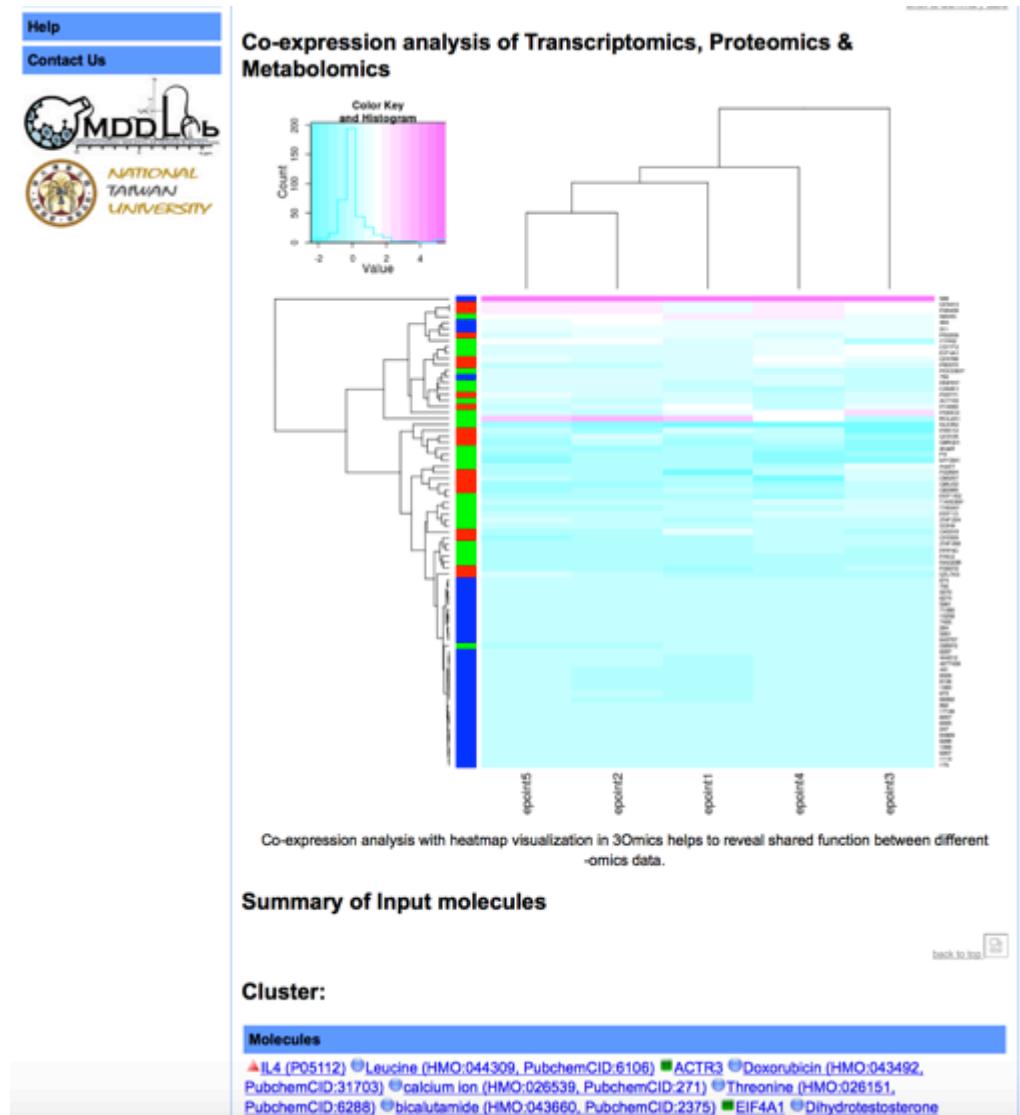
3Omics generates inter-omic correlation network to display the relationship or common patterns in data over time or experimental conditions for all transcripts, proteins and metabolites. Where users may only have two of the three -omics data-sets, 3Omics supplements the missing transcript, protein or 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) ■ PDCD6IP ○ bortezomib (HMO:048610, PubchemCID:387447) ■ PSMC2

# Co-expression analysis



Rows: Variables  
Columns: Samples

# Phenotype analysis



The 3Omics platform interface is shown, featuring a navigation bar with links to Project Features, Correlation Network, Coexpression Profile, Phenotype Analysis, Pathway Analysis, and GO Enrichment Analysis. Below the navigation bar, there is a section titled "Phenotype Analysis" which includes a brief description of what a phenotype is, a list of related genes and phenotypes from OMIM, and a summary of input molecules.

**Project Features**

- Overview
- Name-ID Converter

**Help**

**Contact Us**

**MDD Lab**

**NATIONAL TAIWAN UNIVERSITY**

### 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
[OMIM: 611820] LONG QT SYNDROME 11	akap9
[OMIM: 256000] LEIGH SYNDROME	SDHA
[OMIM: 612068] AMYOTROPHIC LATERAL SCLEROSIS 10, WITH OR WITHOUT FRONTOTEMPORAL DEMENTIA WITH TDP43 INCLUSIONS	TARDBP
[OMIM: 306700] HEMOPHILIA A COAGULATION FACTOR VIII, INCLUDED	F8

### 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:043860, 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 (O43318) ● 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

The 3Omics logo features three colored shapes (blue sphere, green triangle, red square) above the word "3Omics". Below the logo is a network diagram showing nodes connected by various colored arrows (red, blue, green) forming a complex web.

**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**

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) • Acetate • D-Alanine • L-Asparagine • Betaine • Citrate • Ethanolamine • Formate • 6-Deoxy-L-galactose • L-Glutamine • Glycine • L-Histidine • N,N-Dimethylglycine • Pyruvate • Pyridine-2,3-dicarboxylate; • L-Serine • Succinate • L-Tryptophan • L-Tyrosine • N(pi)-Methyl-L-histidine	19
( hsa00970 ) Aminocacyl-tRNA biosynthesis - Homo sapiens (human) • L-Asparagine • L-Glutamine • Glycine • L-Histidine • L-Serine • L-Threonine • L-Tryptophan • L-Tyrosine	8
( hsa00250 ) Alanine, aspartate and glutamate metabolism - Homo sapiens (human) • Acetate • L-Asparagine • L-Glutamine • Glycine • Pyruvate • Succinate • L-Tyrosine	7
( hsa00280 ) Valine, leucine and isoleucine degradation - Homo sapiens (human) • Acetate • Glycine • Pyruvate • Succinate • L-Tryptophan • L-Tyrosine	6
( hsa00270 ) Cysteine and methionine metabolism - Homo sapiens (human) • Betaine • N,N-Dimethylglycine • Pyruvate • L-Serine • L-Tryptophan • L-Tyrosine	6
( hsa00330 ) Arginine and proline metabolism - Homo sapiens (human) • Acetate • L-Glutamine • Glycine • Pyruvate • Succinate • L-Tyrosine	6
( hsa00360 ) Phenylalanine metabolism - Homo sapiens (human) • Acetate • Glycine • L-Histidine • L-Tryptophan • L-Tyrosine	5
( hsa00340 ) Histidine metabolism - Homo sapiens (human) • Acetate • L-Histidine • L-Tryptophan • L-Tyrosine • N(pi)-Methyl-L-histidine	5
( hsa00260 ) Glycine, serine and threonine metabolism - Homo sapiens (human) • Betaine • Glycine • N,N-Dimethylglycine • Pyruvate • L-Serine	5
( hsa00520 ) Amino sugar and nucleotide sugar metabolism - Homo sapiens (human)	4

# GO Enrichment Analysis



3Omics

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



Correlation Network   Coexpression Profile   Phenotype Analysis   Pathway Analysis   GO Enrichment Analysis

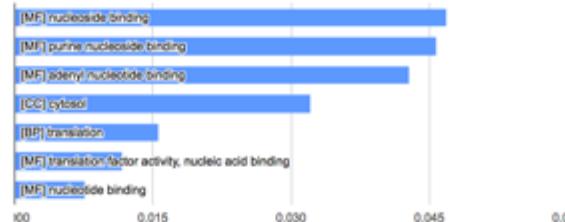
Click to Summary table

## Gene Ontology functional Profiling

The Gene Ontology (GO) provides defined terms for representing the properties of gene product. GO covers three levels of properties: i) cellular component ii) biological process iii) molecular function help users to understand information of gene products from the defined three domains.

[biological process](#) | [cellular component](#) | [molecular function](#)

### GO Terms with P-value < 0.05



## Biological Process

back to top

A biological process is a process of a living organism. Biological processes are made up of any number of chemical reactions or other events that results in a transformation. Regulation of biological processes occurs where any process is modulated in its frequency, rate or extent. Biological processes are regulated by many means; examples include the control of gene expression, protein modification or interaction with a protein or substrate molecule.

GO Term	No. of Gene-mapped	Coverage	P-value	FDR	Mapped Gene ID
<a href="#">translation</a>	4	17%	0.0156	EEF1D,EEF1B2,EIF4A1,WARS	<a href="#">1936</a> , <a href="#">1933</a> , <a href="#">1973</a> , <a href="#">7453</a>
<a href="#">cell death</a>	4	17%	0.1082	PDCD6IP,BCL2A1,TARDBP,PAK2	<a href="#">10015</a> , <a href="#">597</a> , <a href="#">23435</a> , <a href="#">5062</a>
<a href="#">death</a>	4	17%	0.1104	PDCD6IP,BCL2A1,TARDBP,PAK2	<a href="#">10015</a> , <a href="#">597</a> , <a href="#">23435</a> , <a href="#">5062</a>
<a href="#">apoptosis</a>	3	13%	0.2565	PDCD6IP,BCL2A1,PAK2	<a href="#">10015</a> , <a href="#">597</a> , <a href="#">5062</a>

# Multivariate methods

# Generating relevance network using sPLS or msPLS techniques (Cao 2009, Lique 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 (Lique 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

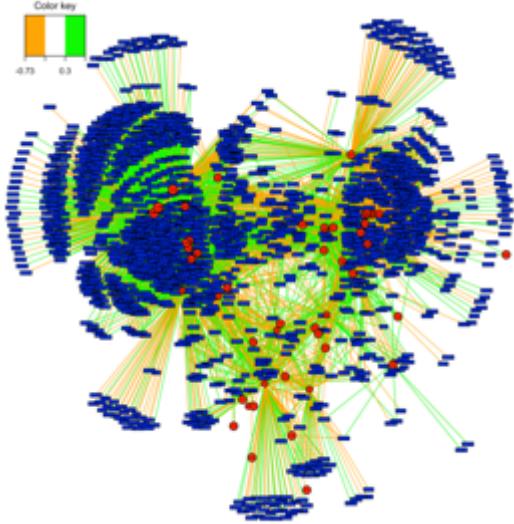
$u_1, u_2 \dots u_H$  and  $v_1, v_2 \dots v_H$  are the loading vectors

H is the number of PLS-DA dimensions

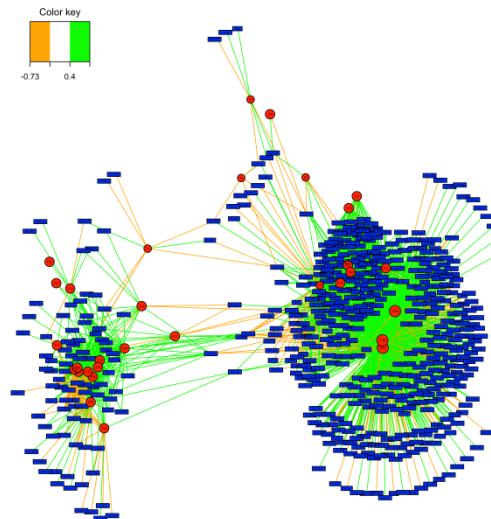
A Lasso based optimization is used to select most relevant variables

**Case Study: 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)

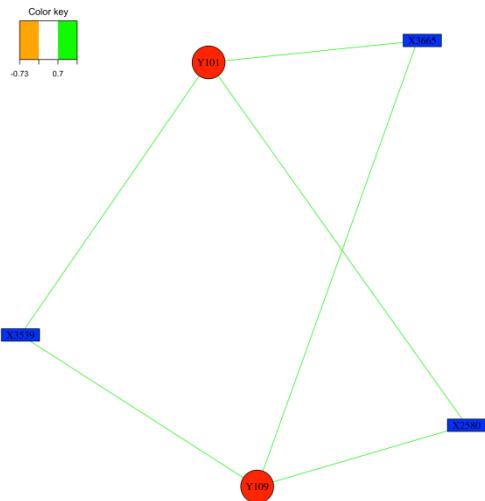
**Legend**  
Circles: microbial species  
Rectangles: metabolome features



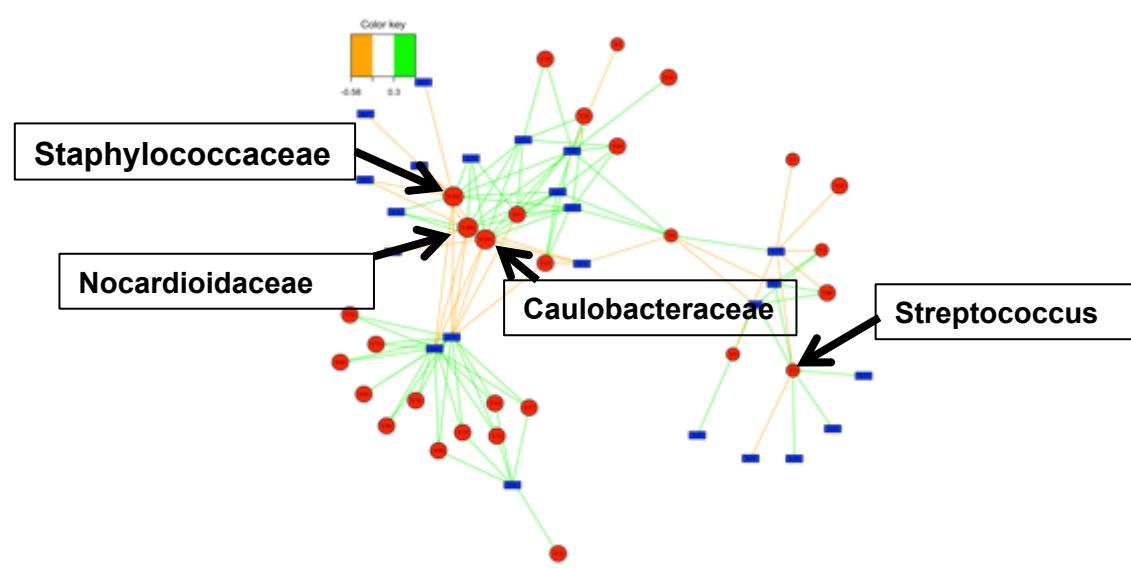
A. Association threshold: 0.3



B. Association threshold: 0.4



C. Association threshold: 0.7



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

Proteomics data  
(n subjects X s proteins)

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

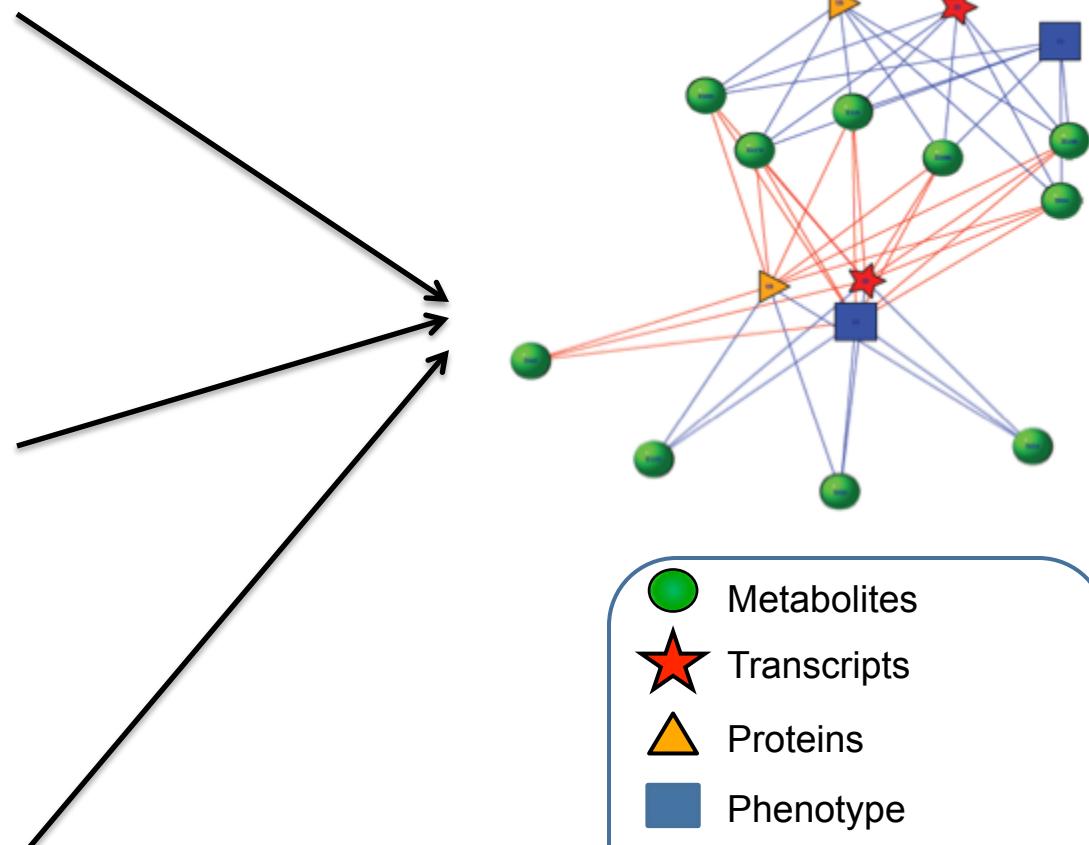
Metabolomics data  
(n subjects X p metabolites)

	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

# Integrating more than two datasets



- Metabolites
- ★ Transcripts
- ▲ Proteins
- Phenotype
- Negative interaction
- Positive interaction

A.

Data matrices from multiple assays



### Integrative and association analysis

Pairwise integrative and network analysis (e.g.  $X \leftrightarrow Y$ ;  $X \leftrightarrow Z$ ,  $Y \leftrightarrow Z$ ) using (sparse) Partial Least Squares regression



Generate an edge list matrix,  $L_e$ , with the list of edges that meet the significance criteria and association score threshold



Generate a k-partite graph,  $G$ , using the union of edge list matrices

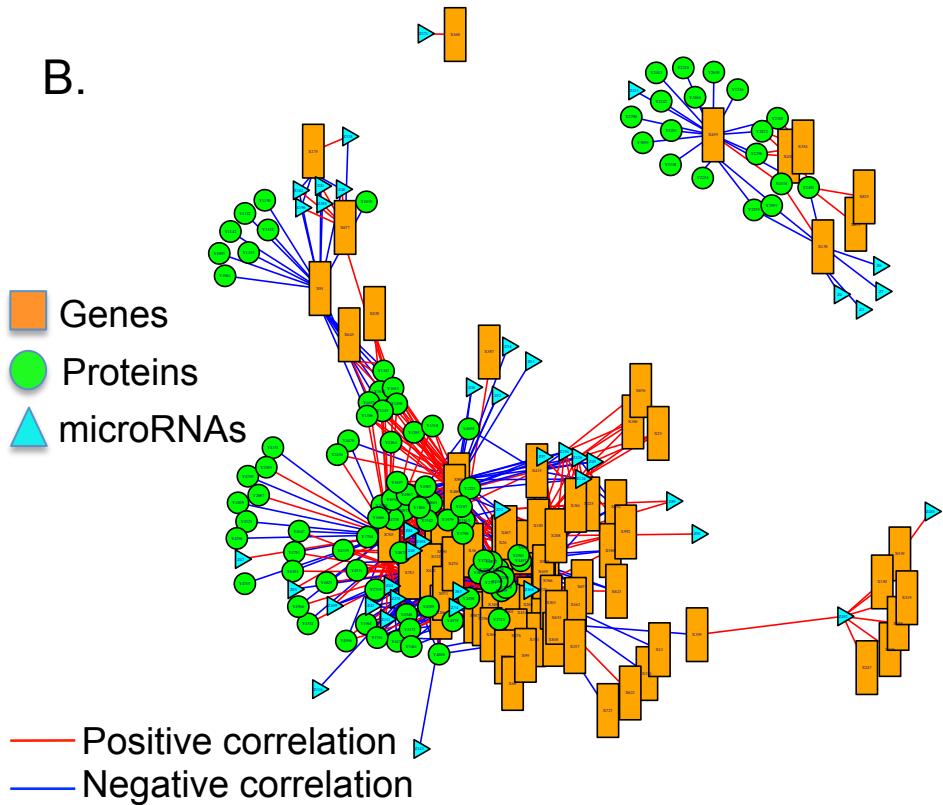


Community detection, differential centrality and rewiring analysis, and network visualization

# xMWAS: R package for data integration and differential network analysis (Uppal 2018, Bioinformatics)

URL: <https://kuppal.shinyapps.io/xmwas/>

B.



# xMWAS: <https://kuppal.shinyapps.io/xmwas/>

https://kuppal.shinyapps.io/xmwas/

## xMWAS - a data-driven integration and network analysis tool (v0.54)

Introduction   Analysis   Help and Support

**xMWAS provides an automated workflow for data integration, network visualization, clustering, and differential network analysis of up to four datasets from biochemical and phenotypic assays, and omics platforms.**

For installing xMWAS locally in R run:

```
library(devtools);install_github("kuppal2/xMWAS")
```

Data matrices from multiple assays

↓

**Integrative and association analysis**  
Pairwise integrative and network analysis  
(e.g. X<->Y; X<->Z, Y<->Z) using (sparse)  
Partial Least Squares regression

↓

Generate an edge list matrix,  $L_e$ , with the list of edges that meet the significance criteria and association score threshold

↓

Generate a graph, G, using the union of edge list matrices

↓

Community detection, differential centrality and rewiring analysis, and network visualization

Legend:

- Genes (Orange square)
- Proteins (Green circle)
- microRNAs (Blue triangle)

— Positive correlation  
— Negative correlation

Citation: Uppal K, Ma C, Go YM, Jones DP. xMWAS: a data-driven integration and differential network analysis tool. *Bioinformatics*. 2017 Oct 23. PMID: 29069296

Maintained by Chunyu Ma ([chunyu.ma@emory.edu](mailto:chunyu.ma@emory.edu)) and Karan Uppal ([kuppal2@emory.edu](mailto:kuppal2@emory.edu)) at [Clinical Biomarkers Laboratory](#), Emory University, Atlanta, GA, USA

# Step 1. Upload data files

xMWAS - a data-driven integration and network analysis tool (v0.54)

Introduction Analysis Help and Support

Input Files

Choose Files (see help and support)

Parameter Settings

1. Data preparation and filtering
2. Integration and association analysis
3. Centrality analysis
4. Graphical options

Select input file for dataset A ('.csv' or '.txt', 100MB limit)  
Browse... No file selected

Name for dataset A: datasetA

Select input file for dataset B ('.csv' or '.txt', 100MB limit)  
Browse... No file selected

Name for dataset B: datasetB

Add more datasets: + -

Choose a class labels file ('.csv' or '.txt'):  
Browse... No file selected

Output folder name: Default: xwasresults

Are there repeated measurements?  
 True - Paired (repeated measures)  
 False - Unpaired (case-control & multiclass)

Compare classes?  
 True  
 False

Use example data?  
 True  
 False

<https://kuppal.shinyapps.io/xmwas/>  
(See: Help & Support)

# Input data format

Metabolomics data  
(p metabolites x n subjects)

	Subject1	Subject2	-	Subject N
Metabolite 1	199	19	-	100
Metabolite 2	10	40		90
-	-	-		-
Metabolite p	50	30	-	20

Class labels file

	Class
Subject1	Control
Subject2	Control
-	-
SubjectN	Tumor

Transcriptomics data  
(q genes x n subjects)

	Subject1	Subject2	-	Subject N
Gene 1	19	19	-	100
Gene 2	10	40	-	90
-	-	-	-	-
Gene q	10	40	-	50

The screenshot shows a web browser window with the URL <https://kuppal.shinyapps.io/xmwas/>. The page title is "xMWAS - a data-driven integration and network analysis tool (v0.54)". The top navigation bar includes links for "Introduction", "Analysis" (which is selected), and "Help and Support". The main content area contains sections for "User Manual" (with a link to the user manual), "Input File Format (no missing values allowed)" (with a table for Dataset File Format and Class Label File Format), and "Class Label File Format (repeated measure with one factor)" (with a table). The browser's address bar shows the URL, and the top right corner has standard browser controls like search and refresh.

## xMWAS - a data-driven integration and network analysis tool (v0.54)

Introduction   Analysis   Help and Support

### User Manual:

Click [here](#) to see the user manual.

### Input File Format (no missing values allowed):

#### Dataset File Format:

mrna_id	CNS.SF_268	CNS.SF_295	CNS.SF_539	CNS.SNB_19	CNS.SNB_75	CNS.U251
5-HT3C2_1_mrna	0.53	-0.42	0	0.5	-0.27	0.43
A1BG-AS1_2_mrna	0.35	0.54	0.8	-0.24	-0.88	-0.1
A2LD1_3_mrna	-0.05	-1.04	0.85	0.12	-0.36	-0.3
A2MP1_4_mrna	-1.09	-1.13	0	-0.43	-0.6	0.42
A4GALT_5_mrna	-0.86	-0.46	-0.57	0.43	1.38	0.4

Each row is a feature and each column is filename

#### Class Label File Format (multiclass):

FileName	Class
CNS.SF_268	CNS
CNS.SF_295	CNS
CNS.SF_539	CNS
CNS.SNB_19	CNS
CNS.SNB_75	CNS
CNS.U251	CNS
LE.CCRF_CEM	LE
LE.HL_60	LE
LE.K_562	LE
LE.MOLT_4	LE
LE.RPMI_8226	LE
LE.SR	LE

Two columns: the first column is filename; the second column is class. Each row is the information of a file.

#### Class Label File Format (repeated measure with one factor):

FileName	Subject	Factor1
S1	Subject001	TP1
S2	Subject002	TP1
S3	Subject003	TP1
S4	Subject004	TP1
S5	Subject005	TP1
SC	Subject006	TP1

# Step 2. Data preprocessing and filtering

The screenshot shows a web browser window for the *xMWAS* tool at <https://kuppal.shinyapps.io/xmwas/>. The page title is *xMWAS - a data-driven integration and network analysis tool (v0.54)*. The navigation bar includes links for **Introduction**, **Analysis** (which is active), and **Help and Support**. On the left, a sidebar lists steps: **1. Data preparation and filtering** (selected), **2. Integration and association analysis**, **3. Centrality analysis**, and **4. Graphical options**. The main content area contains several input fields:

- Relative Standard Deviation (RSD) Threshold (rows):** A dropdown menu set to **1**.
- Maximum number of datasetA variables to select based on RSD:** A dropdown menu set to **1000**.
- Maximum number of datasetB variables to select based on RSD:** A dropdown menu set to **1000**.
- Minimum ratio of number of samples with a non-missing value to the total number of samples for a variable (rows):** A dropdown menu set to **0.7**.
- How are the missing values represented in the data?:** A dropdown menu currently empty.

# Step 3. Set parameters for integration and association analysis

The screenshot shows the xMWAS web application interface. At the top, there is a browser header with navigation icons, a URL bar showing <https://kuppal.shinyapps.io/xmwas/>, and a search bar. Below the header, the title "xMWAS - a data-driven integration and network analysis tool (v0.54)" is displayed.

The main interface has a navigation menu with tabs: "Introduction", "Analysis" (which is selected), and "Help and Support". On the left, a sidebar lists steps: "Input Files", "Parameter Settings", "1. Data preparation and filtering", "2. Integration and association analysis" (which is highlighted with a blue background), "3. Centrality analysis", and "4. Graphical options".

The "Parameter Settings" section is titled "Pairwise integrative analysis". It includes the following controls:

- "Choose a data integration method": A dropdown menu set to "sPLS".
- "Choose PLS mode": A dropdown menu set to "canonical".
- "Number of components to use in PLS model": An input field containing the value "5".
- "Find optimal number of PLS components?": A radio button group where the "True" option is selected.
- "Maximum number of datasetA variables to select in sPLS": An input field containing the value "1000".
- "Maximum number of datasetB variables to select in sPLS": An input field containing the value "1000".

The "Association analysis" section includes the following controls:

- "Correlation Threshold": An input field containing the value "0.4".
- "P-value Threshold For Student's T-test": An input field containing the value "0.05".

# Step 4. Select method for centrality analysis

The screenshot shows the xMWAS shiny app interface. At the top, there is a browser header with icons for back, forward, refresh, and a search bar labeled "Search". Below the header, the title "xMWAS - a data-driven integration and network analysis tool (v0.54)" is displayed. A navigation bar with tabs for "Introduction", "Analysis" (which is selected), and "Help and Support" is visible. On the left, a sidebar titled "Input Files" contains a "Choose Files (see help and support)" button. The main content area is titled "Method for centrality analysis:" and contains a dropdown menu with the following options: eigenvector, betweenness, degree.count, degree.weight, and closeness. The "eigenvector" option is currently selected. To the right of the dropdown, a list of descriptions for each method is provided:

- Eigenvector: based on the number and quality of connections
- Betweenness: based on the extent to which a node lies on the path between other nodes
- Degree.count: based on the number of connections
- Degree.weight: based on the magnitude of edges (association scores)
- Closeness: based on the closeness of a node to all other nodes

- Eigenvector: based on the number and quality of connections
- Betweenness: based on the extent to which a node lies on the path between other nodes
- Degree.count: based on the number of connections
- Degree.weight: based on the magnitude of edges (association scores)
- Closeness: based on the closeness of a node to all other nodes

# Step 5. Set graphical options

The screenshot shows a web browser window for the *xMWAS* tool at <https://kuppal.shinyapps.io/xmwas/>. The page title is *xMWAS - a data-driven integration and network analysis tool (v0.54)*. The navigation menu includes [Introduction](#), [Analysis](#) (which is active), and [Help and Support](#). On the left, a sidebar lists steps: [Input Files](#), [Choose Files \(see help and support\)](#), [Parameter Settings](#), [1. Data preparation and filtering](#), [2. Integration and association analysis](#), [3. Centrality analysis](#), and [4. Graphical options](#) (highlighted in blue). The main content area contains several input fields:

- Size of the Labels:** A dropdown menu set to 0.25.
- Size of the Nodes:** A dropdown menu set to 7.
- Seed for Random Number Generator:** A dropdown menu set to 100.
- Maximum number of associations to include in the network (any numeric value >0 or -1 to use all):** A dropdown menu set to -1.
- Use dataset A as reference?** A radio button group where  True is unselected and  False is selected.

# Step 6. Click on “Start processing”

Screenshot of the xMWAS web application interface.

The URL in the browser is <https://kuppal.shinyapps.io/xmwas/>.

The page title is *xMWAS - a data-driven integration and network analysis tool (v0.54)*.

Navigation tabs at the top include **Introduction**, **Analysis** (selected), and **Help and Support**.

**Input Files** section:

- Choose Files (see help and support)

**Parameter Settings** section:

1. Data preparation and filtering
2. Integration and association analysis
3. Centrality analysis
- 4. Graphical options** (selected)

Configuration fields for Graphical Options:

- Size of the Labels: 0.25
- Size of the Nodes: 7
- Seed for Random Number Generator: 100
- Maximum number of associations to include in the network (any numeric value >0 or -1 to use all): -1
- Use dataset A as reference?  
 True  False

Action buttons at the bottom left:

- Start processing** (highlighted with a blue border)
- Download results

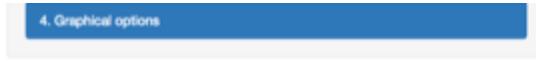
**Output** section:

Slide to go to next figure: 0

Citation: Uppal K, Ma C, Go YM, Jones DR. xMWAS: a data-driven integration and differential network analysis tool. Bioinformatics. 2017 Oct 23. PMID: 29069296  
Maintained by Chunyu Ma ([chunyu.ma@emory.edu](mailto:chunyu.ma@emory.edu)) and Karan Uppal ([kuppal2@emory.edu](mailto:kuppal2@emory.edu)) at Clinical Biomarkers Laboratory, Emory University, Atlanta, GA, USA

A message box in the bottom right corner says: You're using example data and it is processing now. Your results will be available for download shortly.

# Results

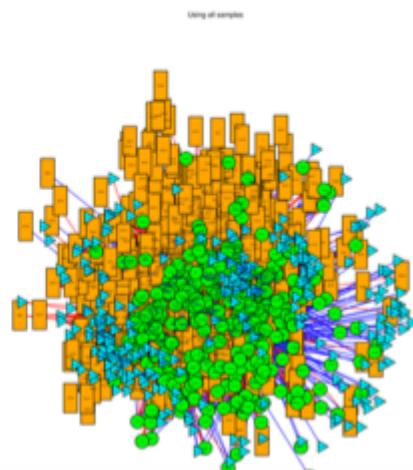


Processing complete. Please click on Download to save the results.

Start processing   

## Output

Slide to go to next figure:

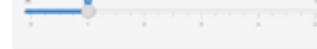


Processing complete. Please click on Download to save the results.

Start processing   

## Output

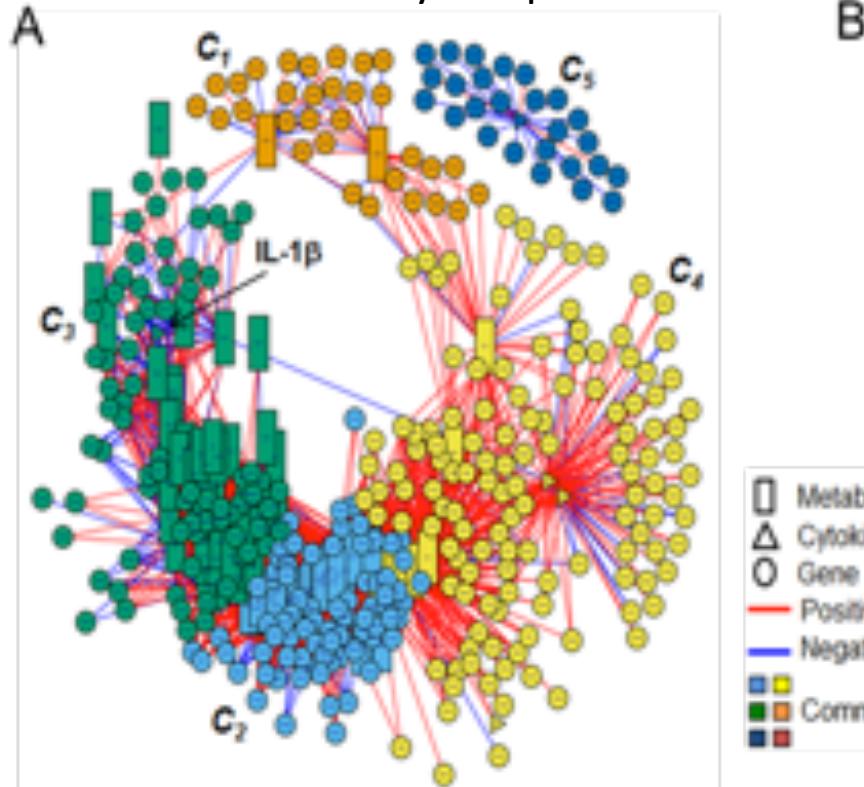
Slide to go to next figure:



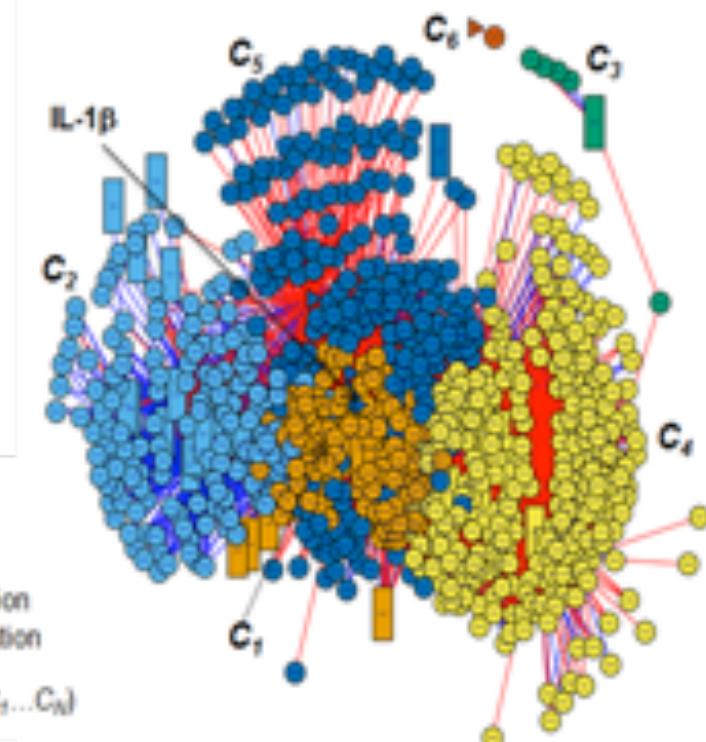
## **Case Study: Application of xMWAS for integrative –and differential network analysis of more than 2 dataests.**

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

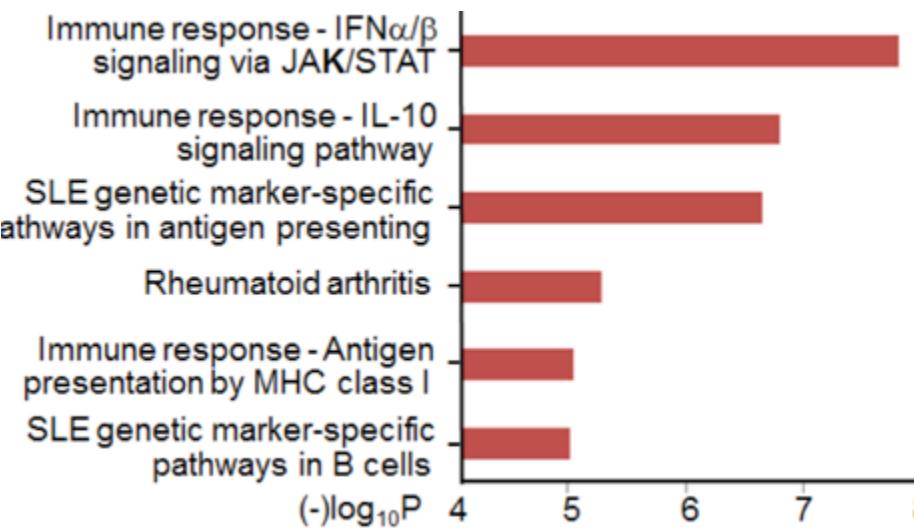
Control only samples



H1N1 only samples



C



# 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: HiPub, CoPub, ArrowSmith**

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

NCBI Resources How To

PubMed "breast cancer" Search

Create RSS Create alert Advanced

Article types Summary 20 per page Sort by Most Recent

Clinical Trial

Review

Customize ...

Results: 1 to 20 of 191685

Send to: <<First <Prev Page 1 of 9685 Next> Last>

Text availability Abstract

Free full text

Full text

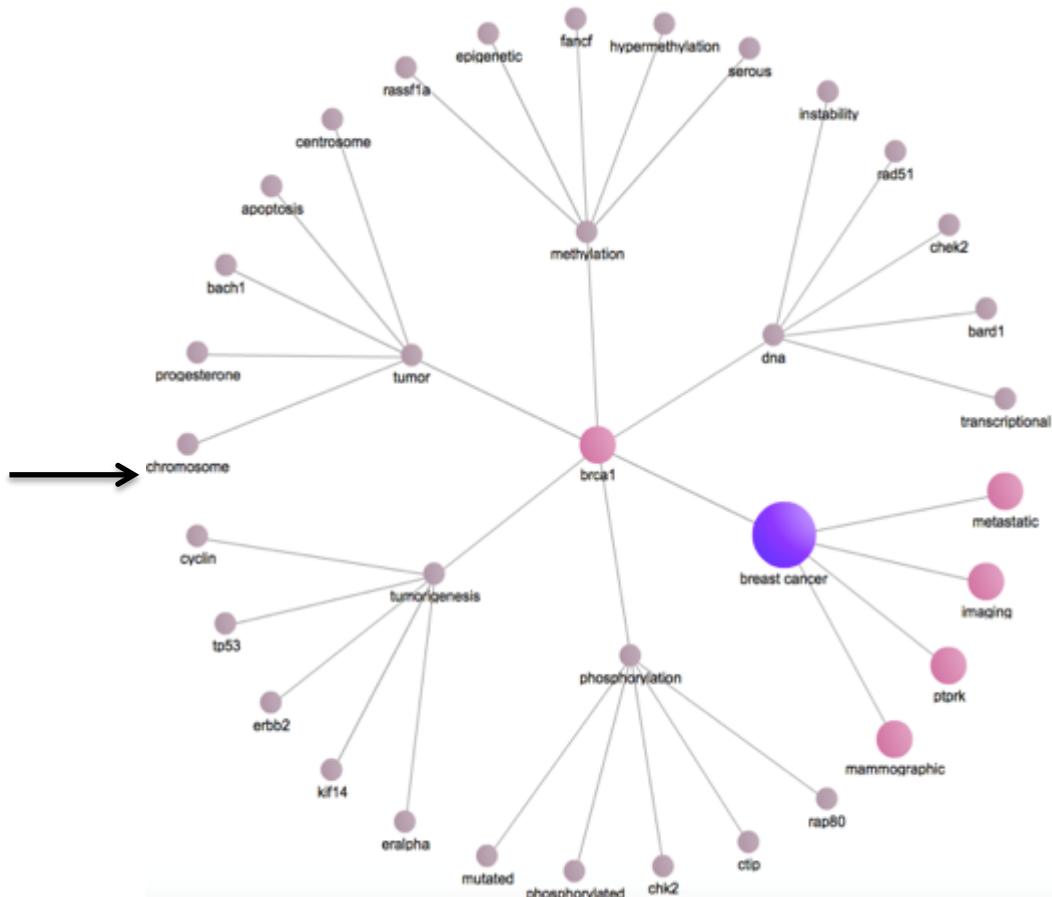
PubMed Commons

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](#)



# 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 Advanced Search Help

Format: Abstract

PLoS One, 2013 Aug 29;8(8):e73059. doi: 10.1371/journal.pone.0073059. eCollection 2013.

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 hypereosinophils 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

Verleporfin induces apoptosis and eliminates cancer stem-like cells in [Am J Cancer Res. 2016]

Review The Role of New Tyrosine Kinase Inhibitors in Chronic Myeloid Leu [Cancer Res. 2016]

See reviews... See all...

Full text links

PLOS PMC Full text

Save items Add to Favorites

Similar articles

Antitumor activity of S116836, a novel tyrosine kinase inhibitor, against Imatin [Oncotarget. 2014]

Cyclin-dependent kinase 7/9 inhibitor SNS-032 abrogates FIP1-like-1 p16 [Clin Cancer Res. 2012]

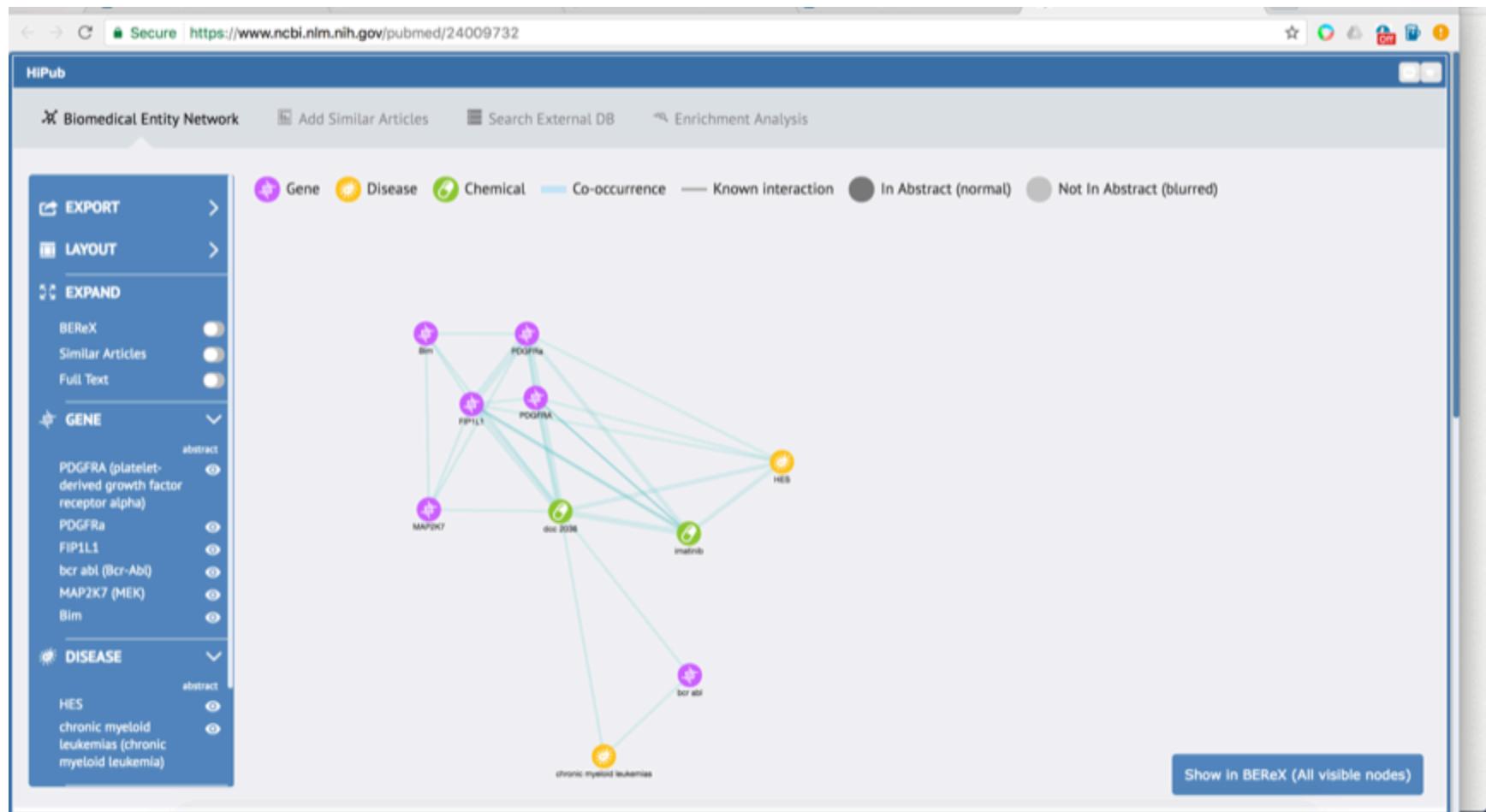
Pontatinib efficiently kills imatinib-resistant chronic eosinophilic leukemia cells ha [Mol Cancer. 2014]

Review [FIP1L1-PDGFR $\alpha$  positive chronic eosinophilic [Zhonghua Xue Ye Xue Za Zhi. 2013]

Review The FIP1L1-PDGFR $\alpha$  fusion tyrosine kinase in hypereosinophilic [Blood. 2004]

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

# Hands-on exercises

# 1. HiPub

Go to: <http://hipub.korea.ac.kr/>

The screenshot shows a web browser window with the URL 'hipub.korea.ac.kr' in the address bar. The page content is as follows:

- i. HiPub is available in [Chrome Browser](#)
- ii. HiPub is designed for [PubMed](#) and [PubMed Central](#).
- iii. HiPub also works in,
  - All journals available in the AACR ([www.aacrjournals.org](http://www.aacrjournals.org)) sites
  - All journals available in the ASCO ([www.ascopubs.org](http://www.ascopubs.org)) sites
  - All journals available in BioMedCentral (BMC) ([www.biomedcentral.com](http://www.biomedcentral.com)) sites
  - All journals available in the Oxford University Press ([www.oxfordjournals.org](http://www.oxfordjournals.org)) sites

**2. Installation and Usage**

- i. Install the newest version of [Chrome Browser](#) (skip this step if you have Chrome Browser)
- ii. Install HiPub through [Google Chrome Web Store](#)
- iii. Go to [PubMed](#) or [PubMed Central](#) and HiPub will automatically work.
- iv. Usage Example.
  - i. PubMed Article (Abstract) : [PMID24009732](#)
  - ii. PubMed Article (Abstract) : [PMID21262914](#)
  - iii. PubMed Article (Abstract) : [PMID21262914](#)
  - iv. PubMed Central Article (Full Text) : [PMC3733554](#)
  - v. PubMed Central Article (Full Text) : [PMC3641357](#)

**3. Documentation**

User Manual : [HiPub User Manual](#)

**4. Citation**

Kyubum Lee, Wonho Shin, Byounggun Kim, Sunwon Lee, Yonghwa Choi, Sunkyu Kim, Minji Jeon, Aik Choon Tan\*, and Jaewoo Kang\*  
HiPub: translating PubMed and PMC texts to networks for knowledge discovery

# Steps for installing HiPub

- Install Chrome browser:

<https://www.google.com/chrome/browser/desktop/>

- Install the HiPub plugin:

<https://chrome.google.com/webstore/detail/hipub/jlbtmiklekmigmbmcodhjgdpooldjcjam>

- Test installation:

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

- You should see an annotated title and abstract when you go to the PubMed page above as shown on the next slide

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 +

PLoS One, 2013 Aug 29;8(8):e73059. doi: 10.1371/journal.pone.0073059. eCollection 2013.

## 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 hypereosinophilic 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](https://doi.org/10.1371/journal.pone.0073059)

[Indexed for MEDLINE] Free PMC Article

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



Full text links

OPEN ACCESS TO FULL TEXT PLOS ONE PMC Full text FREE

Save items Add to Favorites

Similar articles

- Antitumor activity of S116836, a novel tyrosine kinase inhibitor, against imatin [Oncotarget. 2014]
- Cyclin-dependent kinase 7/9 inhibitor SNS-032 abrogates FIP1-like-1 pfa [Clin Cancer Res. 2012]
- Ponatinib efficiently kills imatinib-resistant chronic eosinophilic leukemia cells ha [Mol Cancer. 2014]
- Review [FIP1L1-PDGFR $\alpha$  positive chronic eosinophilic [Zhonghua Xue Ye Xue Za Zhi. 2013]
- Review The FIP1L1-PDGFR $\alpha$  fusion tyrosine kinase in hypereosinophilic [Blood. 2004]

See reviews... See all...

Cited by 5 PubMed Central articles

- Verteporfin induces apoptosis and eliminates cancer stem-like cells ir [Am J Cancer Res. 2016]
- Review The Role of New Tyrosine Kinase Inhibitors in Chronic Myeloid Leu [Cancer Lett. 2016]

Previous Diabetic dermopathy for HiPub

## 2. xMWAS: Web version

- URL: <https://kuppal.shinyapps.io/xmwas/>
- Input files:  
[https://github.com/kuppal2/xMWAS/upload/  
master/example manual tutorial](https://github.com/kuppal2/xMWAS/upload/master/example_manual_tutorial)
  - exh1n1\_transcriptome.txt
  - exh1n1\_metabolome.txt
  - exh1n1\_cytokine.txt
  - exh1n1\_classlabels.txt

https://kuppal.shinyapps.io/xmwas/ 80% Search

xMWAS - a data-driven integration and network analysis tool (v0.54)

Introduction Analysis Help and Support

**Input Files**

Choose Files (see help and support)

**Parameter Settings**

1. Data preparation and filtering
2. Integration and association analysis
3. Centrality analysis
4. Graphical options

Select input file for dataset A ('.csv' or '.txt', 100MB limit)  
Browse... exh1n1\_metabolome.txt

Name for dataset A:  
metabolome

Select input file for dataset B ('.csv' or '.txt', 100MB limit)  
Browse... exh1n1\_transcriptome.txt

Name for dataset B:  
transcriptome

Select input file for dataset C ('.csv' or '.txt', 100MB limit)  
Browse... exh1n1\_cytokine.txt

Name for dataset C:  
cytokine

Add more datasets: + -

Choose a class labels file ('.csv' or '.txt'):  
Browse... exh1n1\_classlabels.txt

Output folder name:  
Default: xwasresults

Are there repeated measurements?  
 True - Paired (repeated measures)  
 False - Unpaired (case-control & multiclass)

Compare classes?  
 True  
 False

Use example data?  
 True  
 False

Start processing

**Output**

Slide to go to next figure:

Cloud Drive > Documents > Lectures > UAB march21 2018 > Software int karanuppal

## xMWAS - a data-driven integration and network analysis tool (v0.54)

[Introduction](#)[Analysis](#)[Help and Support](#)

### Input Files

Choose Files (see help and support)

### Parameter Settings

1. Data preparation and filtering

2. Integration and association analysis

3. Centrality analysis

4. Graphical options

Relative Standard Deviation (RSD) Threshold  
(rows):

Maximum number of datasetA variables to select based

on RSD:

Maximum number of datasetB variables to select based  
on RSD:

Maximum number of datasetC variables to select based  
on RSD:

Minimum ratio of number of samples with a non-missing value to the total number of  
samples for a variable (rows):

How are the missing values represented in  
the data?:

xMWAS - a data-driven integration and network analysis tool (v0.54)

Introduction Analysis Help and Support

**Input Files**

Choose Files (see help and support)

**Parameter Settings**

1. Data preparation and filtering

2. Integration and association analysis

3. Centrality analysis

4. Graphical options

**Pairwise integrative analysis**

Choose a data integration method: sPLS

Choose PLS mode: canonical

Number of components to use in PLS model: 5

Find optimal number of PLS components? True

Maximum number of datasetA variables to select in sPLS: 100

Maximum number of datasetB variables to select in sPLS: 100

Maximum number of datasetC variables to select in sPLS: 100

**Association analysis**

Correlation Threshold: 0.7

P-value Threshold For Student's T-test: 0.05

Start processing

**Output**

Slide to go to next figure: 5

Citation: Uppal K, Ma C, Go YM, Jones DP. xMWAS: a data-driven integration and differential network analysis tool. *Bioinformatics*. 2017 Oct 23. PMID: 29066296  
Maintained by Chunyu Ma ([chunyu.ma@emory.edu](mailto:chunyu.ma@emory.edu)) and Karan Uppal ([kuppal2@emory.edu](mailto:kuppal2@emory.edu)) at Clinical Biomarkers Laboratory, Emory University, Atlanta, GA, USA

shinyapps Download ... Inbox (1,822)... DLFM - add... PLOS Comp... translating P... Highlighter f... Chrome Web... HiPub - Chro... HiPub - Chro... xMWAS - v0... +

## xMWAS - a data-driven integration and network analysis tool (v0.54)

Introduction Analysis Help and Support

**Input Files**

Choose Files (see help and support)

**Parameter Settings**

1. Data preparation and filtering
2. Integration and association analysis
3. Centrality analysis
4. Graphical options

Size of the Labels: 0.25

Size of the Nodes: 7

Seed for Random Number Generator: 100

Maximum number of associations to include in the network (any numeric value >0 or -1 to use all): -1

Use dataset A as reference?  
 True  False

Start processing  Download results

**Output**

Slide to go to next figure:

Starting processing now. Your results will be available for download shortly. The processing time depends on the number of variables. Please use the data filtering options to reduce the run time.

Citation: Uppal K, Ma C, Go YM, Jones DP. xMWAS: a data-driven integration and differential network analysis tool. Bioinformatics. 2017 Oct 23. PMID: 29069296

# Results



# Download results

The screenshot shows a file browser window with the following details:

- Title Bar:** xmwasresults2018032111005
- Toolbar:** View, Arrange, Action, Share, Edit Tags.
- Search Bar:** Search (with a magnifying glass icon).
- Table Headers:** Name, Date Modified, Size, Kind.
- File List:** A list of files and a folder, all modified today at 7:11 AM or 7:10 AM, except for LogWed\_Mar\_21\_07\_10\_05\_2018.txt which was modified at 7:10 AM.
- Folder Structure:** The folder contains a 'pairwise\_results' folder which contains three files: datasetA\_x\_datasetB\_association\_matrix\_threshold0.7.txt, datasetA\_x\_datasetB\_association\_matrix\_threshold0.7.png, and datasetA\_x\_datasetB\_association\_matrix\_threshold0.7.txt.
- File Types:** Plain Text, PNG image, Document, comma...values, Folder.
- File Sizes:** Various sizes including 12 KB, 18 KB, 4 KB, 789 bytes, 1.6 MB, 160 KB, 1.5 MB, 524 KB, 126 bytes, 22 KB, --, 389 KB, 2.4 MB, 363 KB, 2 KB, and 5.5 MB.

Name	Date Modified	Size	Kind
class-wise_centrality_matrix.txt	Today at 7:11 AM	12 KB	Plain Text
cluster_membership_centrality_table.txt	Today at 7:11 AM	18 KB	Plain Text
InputParameters.txt	Today at 7:10 AM	4 KB	Plain Text
LogWed_Mar_21_07_10_05_2018.txt	Today at 7:10 AM	789 bytes	Plain Text
Multidata_Network_threshold0.7_communities.png	Today at 7:11 AM	1.6 MB	PNG image
Multidata_Network_threshold0.7_linkmatrix.txt	Today at 7:10 AM	160 KB	Plain Text
Multidata_Network_threshold0.7.png	Today at 7:10 AM	1.5 MB	PNG image
Multidata_Network_threshold0.7cytoscapeall.gml	Today at 7:10 AM	524 KB	Document
Network_stats.csv	Today at 7:10 AM	126 bytes	comma...values
NodeID_Name_mapping.txt	Today at 7:10 AM	22 KB	Plain Text
pairwise_results	Today at 7:11 AM	--	Folder
datasetA_x_datasetB_association_matrix_threshold0.7.txt	Today at 7:10 AM	389 KB	Plain Text
datasetA_x_datasetB_association_matrix_threshold0.7.png	Today at 7:10 AM	2.4 MB	PNG image
datasetA_x_datasetB_association_matrix_threshold0.7.txt	Today at 7:10 AM	363 KB	Plain Text
README.txt	Today at 7:11 AM	2 KB	Plain Text
xmwasresults2018032111005.zip	Today at 7:11 AM	5.5 MB	ZIP archive

### 3. xMWAS R package installation instructions for Windows

- Install R: <https://cran.cnr.berkeley.edu/>

- Install R dependencies

- R command for installation:

```
source("https://bioconductor.org/biocLite.R");
biocLite(c("GO.db","graph","RBGL","impute","preprocessCore"),dependencies=TRUE);
install.packages(c("devtools","WGCNA","mixOmics","snow","igraph","plyr","plsgenomics")
,dependencies=TRUE,type="binary", repos="http://cran.r-project.org")
```

- Install R package xMWAS

- R command for installation:

```
library(devtools); install_github("kuppal2/xMWAS")
```

- Test installation:

- R command for loading the package:

```
library(xMWAS)
```

# xMWAS R package installation instructions for Mac OS X

- Install Xquartz: <https://www.xquartz.org/>
- Install R: <https://cran.cnr.berkeley.edu/>
- Install R dependencies

- R command for installation:

```
source("https://bioconductor.org/biocLite.R");
biocLite(c("GO.db","graph","RBGL","impute","preprocessCore"),dependencies=TRUE);
install.packages(c("devtools","WGCNA","mixOmics","snow","igraph","plyr","plsgenomics")
,dependencies=TRUE,type="source", repos="http://cran.r-project.org")
```

- Install R package xMWAS
  - R command for installation:  
**library(devtools); install\_github("kuppal2/xMWAS")**
- Test installation:
  - R command for loading the package: **library(xMWAS)**

# R script for xMWAS using the example dataset

(URL: [https://github.com/kuppal2/xMWAS/blob/master/example\\_manual\\_tutorial/example\\_xmwas\\_runscript\\_v0.5.R](https://github.com/kuppal2/xMWAS/blob/master/example_manual_tutorial/example_xmwas_runscript_v0.5.R))

```
#load package
library(xMWAS)

#example dataset that includes metabolome, transcriptome, and cytokine data from the H1N1 mice study (Chandler 2016)
data(exh1n1)
data(classlabels_casecontrol) #example classlabels file for case vs control design
data(classlabels_repeatmeasures) #example classlabels file for repeat measures design
xMat<-exh1n1$metabolome
yMat<-exh1n1$transcrptome
zMat<-exh1n1$cytokine
classlabels<-exh1n1$classlabels

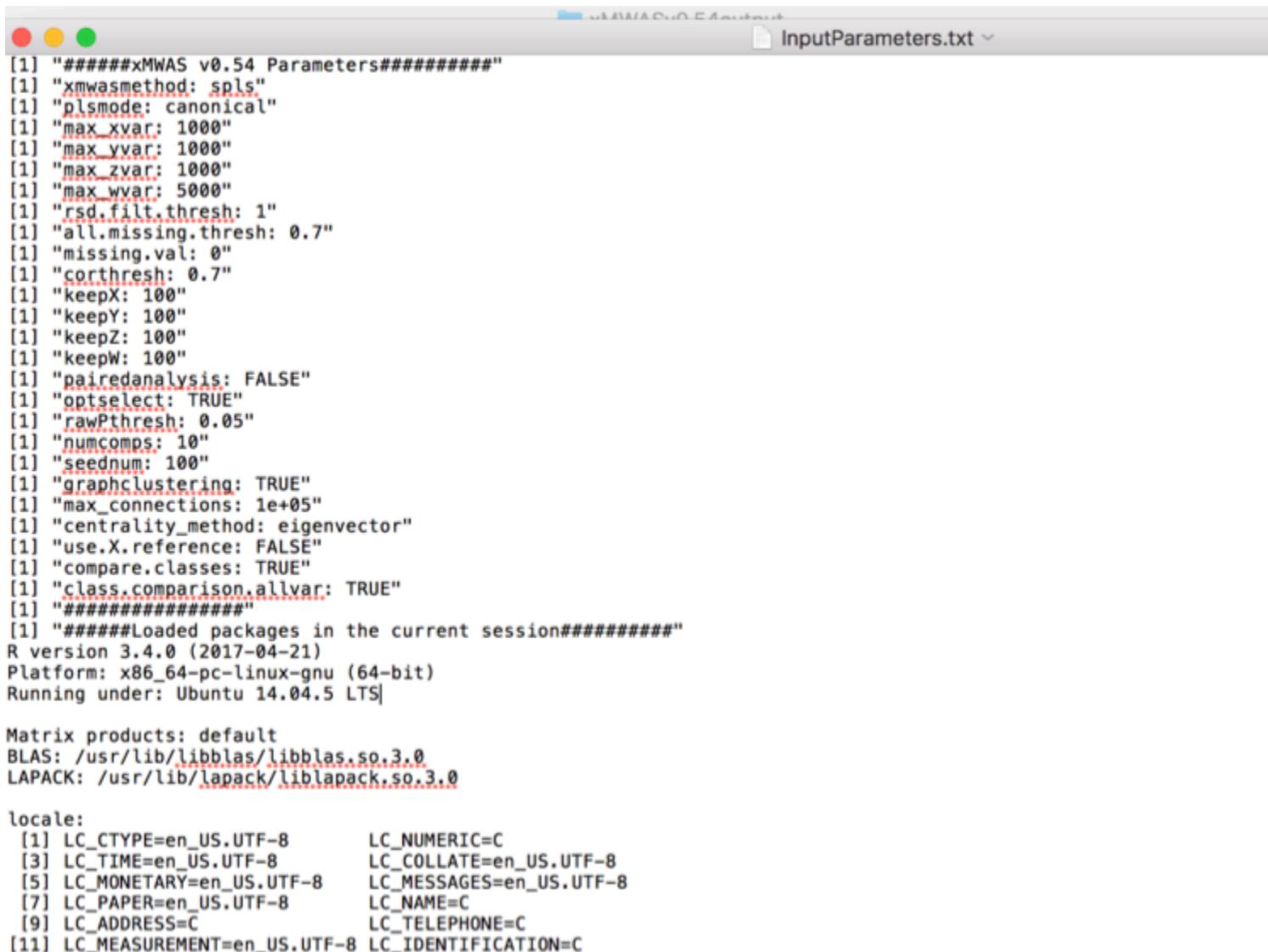
output<-/home/kuppal2/xMWASv0.54output/"

#call the run_xmwas() function:
xmwas_res<-run_xmwas(Xome_data=xMat,Yome_data=yMat,Zome_data=zMat,Wome_data=NA,outloc=output,
classlabels=classlabels,class_fname=NA,xmwasmethod="spls",plsmode="canonical",max_xvar=1000,max_yvar=1000,
max_zvar=1000,max_wvar=1000,rsd.filt.thresh=1,corthresh=0.7,keepX=100,keepY=100,keepZ=100,keepW=100,
pairedanalysis=FALSE,optselect=TRUE,rawPthresh=0.05,numcomps=10,net_edge_colors=c("blue","red"),
net_node_colors=c("orange", "green", "cyan", "pink"),Xname="X",Yname="Y",Zname="Z",Wname="W",
net_node_shape=c("rectangle","circle","triangle","star"),all.missing.thresh=0.7,missing.val=0,
seednum=100,label.cex=0.2,vertex.size=6,graphclustering=TRUE,interactive=FALSE,max_connections=100000,
centrality_method="eigenvector",use.X.reference=FALSE,removeRda=TRUE,compare.classes=TRUE,class.comparison.allvar=TRUE
) suppressWarnings(try(sink(file=NULL),silent=TRUE))
```

# xMWAS output

xMWASv0.54output			
View	Arrange	Action	Share
Name	Date Modified	Size	Kind
class-wise_centrality_matrix.txt	Today at 11:29 PM	68 KB	Plain Text
cluster_membership_centrality_table.txt	Today at 11:30 PM	32 KB	Plain Text
Control	Today at 11:32 PM	--	Folder
H1N1	Today at 11:30 PM	--	Folder
InputParameters.txt	Today at 11:30 PM	4 KB	Plain Text
LogTue_Mar_20_23_00_40_2018.txt	Today at 11:29 PM	5 KB	Plain Text
Multidata_Network_threshold0.7_communities.png	Today at 11:29 PM	1.9 MB	PNG image
Multidata_Network_threshold0.7_linkmatrix.txt	Today at 11:32 PM	541 KB	Plain Text
Multidata_Network_threshold0.7.png	Today at 11:32 PM	1.6 MB	PNG image
Multidata_Network_threshold0.7cytoscapeall.gml	Today at 11:32 PM	1.7 MB	Document
Network_stats.csv	Today at 11:31 PM	143 bytes	comma...values
NodeID_Name_mapping.txt	Today at 11:32 PM	44 KB	Plain Text
pairwise_results	Today at 11:31 PM	--	Folder
Rplots.pdf	Today at 11:31 PM	4 KB	Adobe...ument
Rplots1.pdf	Today at 11:30 PM	4 KB	Adobe...ument
Rplots2.pdf	Today at 11:31 PM	4 KB	Adobe...ument
X_x_Y_association_matrix_threshold0.7.txt	Today at 11:31 PM	1.1 MB	Plain Text
X_x_Y_association_network_threshold0.7.png	Today at 11:31 PM	3.4 MB	PNG image
X_x_YBoolean_association_matrix_threshold0.7.txt	Today at 11:30 PM	980 KB	Plain Text
X_x_Z_association_matrix_threshold0.7.txt	Today at 11:31 PM	5 KB	Plain Text
X_x_Z_association_network_threshold0.7.png	Today at 11:31 PM	1.8 MB	PNG image
X_x_ZBoolean_association_matrix_threshold0.7.txt	Today at 11:30 PM	5 KB	Plain Text
Y_x_Z_association_matrix_threshold0.7.txt	Today at 11:30 PM	6 KB	Plain Text
Y_x_Z_association_network_threshold0.7.png	Today at 11:31 PM	2.1 MB	PNG image
Y_x_ZBoolean_association_matrix_threshold0.7.txt	Today at 11:30 PM	5 KB	Plain Text
README.txt	Today at 11:29 PM	2 KB	Plain Text

# Input Parameters

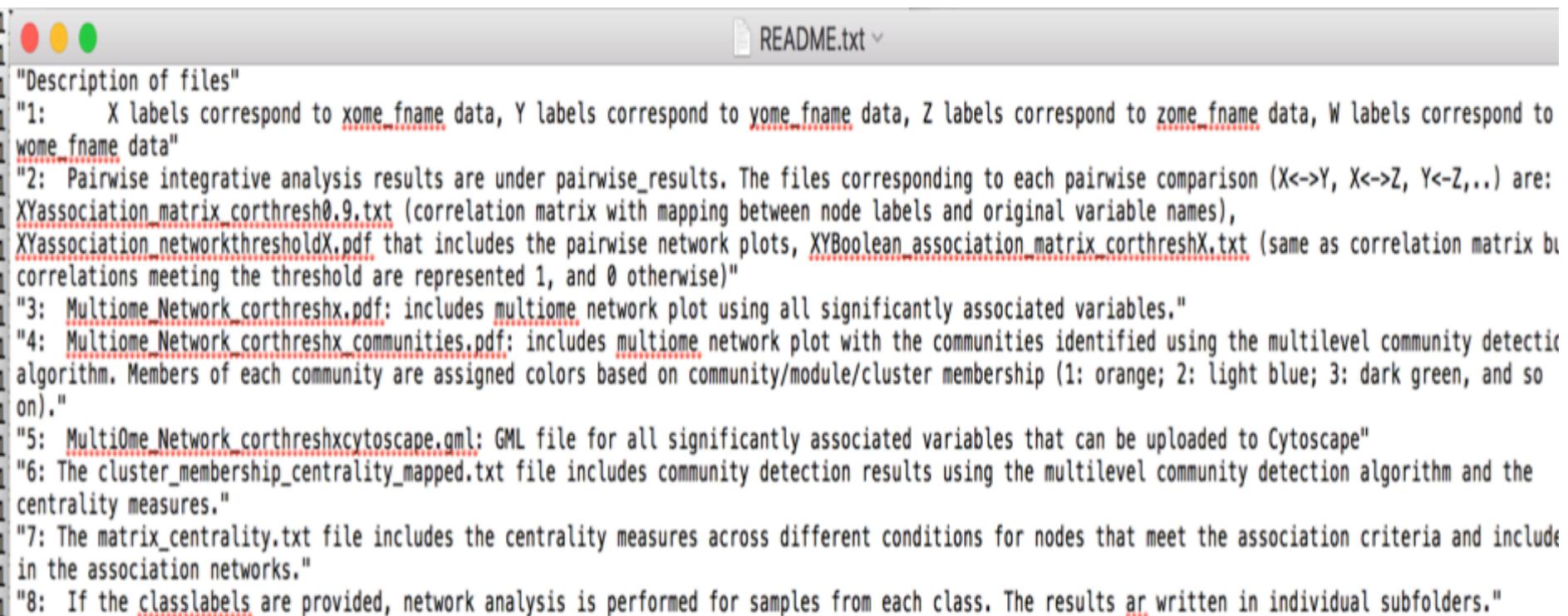


```
[1] #####xMWAS v0.54 Parameters#####
[1] "xmwasmethod: spls"
[1] "plsmode: canonical"
[1] "max_xvar: 1000"
[1] "max_yvar: 1000"
[1] "max_zvar: 1000"
[1] "max_wvar: 5000"
[1] "rsd_filt.thresh: 1"
[1] "all.missing.thresh: 0.7"
[1] "missing.val: 0"
[1] "corthresh: 0.7"
[1] "keepX: 100"
[1] "keepY: 100"
[1] "keepZ: 100"
[1] "keepW: 100"
[1] "pairedanalysis: FALSE"
[1] "optselect: TRUE"
[1] "rawPthresh: 0.05"
[1] "numcomps: 10"
[1] "seednum: 100"
[1] "graphclustering: TRUE"
[1] "max_connections: 1e+05"
[1] "centrality_method: eigenvector"
[1] "use.X.reference: FALSE"
[1] "compare.classes: TRUE"
[1] "class.comparison.allvar: TRUE"
[1] #####
[1] #####Loaded packages in the current session#####
R version 3.4.0 (2017-04-21)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 14.04.5 LTS

Matrix products: default
BLAS: /usr/lib/libblas/libblas.so.3.0
LAPACK: /usr/lib/lapack/liblapack.so.3.0

locale:
[1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
[3] LC_TIME=en_US.UTF-8       LC_COLLATE=en_US.UTF-8
[5] LC_MONETARY=en_US.UTF-8   LC_MESSAGES=en_US.UTF-8
[7] LC_PAPER=en_US.UTF-8      LC_NAME=C
[9] LC_ADDRESS=C              LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
```

# Description of output: Readme.txt

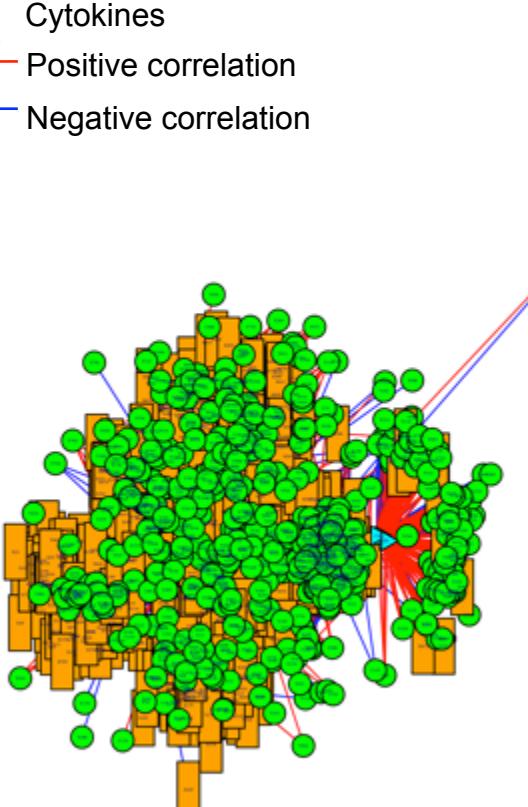
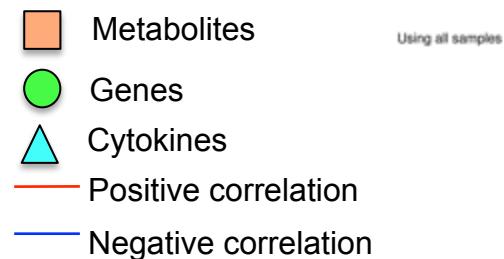


The screenshot shows a text editor window titled "README.txt". The content of the file is a series of numbered instructions describing the output files generated by the analysis. The text is color-coded with red highlights on certain file names and paths.

```
"Description of files"
"1: X labels correspond to xome_fname data, Y labels correspond to yome_fname data, Z labels correspond to zome_fname data, W labels correspond to wome_fname data"
"2: Pairwise integrative analysis results are under pairwise_results. The files corresponding to each pairwise comparison (X<->Y, X<->Z, Y<-Z,...) are: XYassociation_matrix_corthresh0.9.txt (correlation matrix with mapping between node labels and original variable names), XYassociation_networkthresholdX.pdf that includes the pairwise network plots, XYBoolean_association_matrix_corthreshX.txt (same as correlation matrix but correlations meeting the threshold are represented 1, and 0 otherwise)"
"3: Multiome Network_corthreshx.pdf: includes multiome network plot using all significantly associated variables."
"4: Multiome Network_corthreshx_communities.pdf: includes multiome network plot with the communities identified using the multilevel community detection algorithm. Members of each community are assigned colors based on community/module/cluster membership (1: orange; 2: light blue; 3: dark green, and so on)."
"5: MultiOme_Network_corthreshxcytoscape.gml: GML file for all significantly associated variables that can be uploaded to Cytoscape"
"6: The cluster_membership_centrality_mapped.txt file includes community detection results using the multilevel community detection algorithm and the centrality measures."
"7: The matrix_centrality.txt file includes the centrality measures across different conditions for nodes that meet the association criteria and include in the association networks."
"8: If the classlabels are provided, network analysis is performed for samples from each class. The results are written in individual subfolders."
```

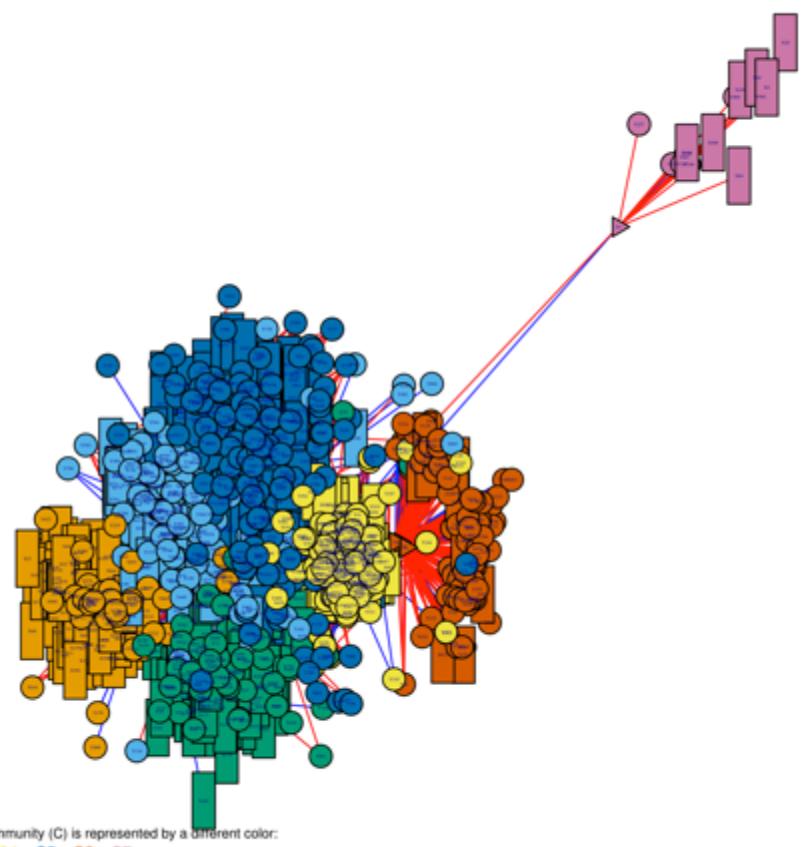
# Network graphs

A. Colored by data type



B. Colored by community membership

(Edges) Red: +ve correlation; Blue: -ve correlation  
(Nodes) Rectangle: X; Circle: Y; Triangle: Z  
Using all samples



# Community detection and centrality analysis

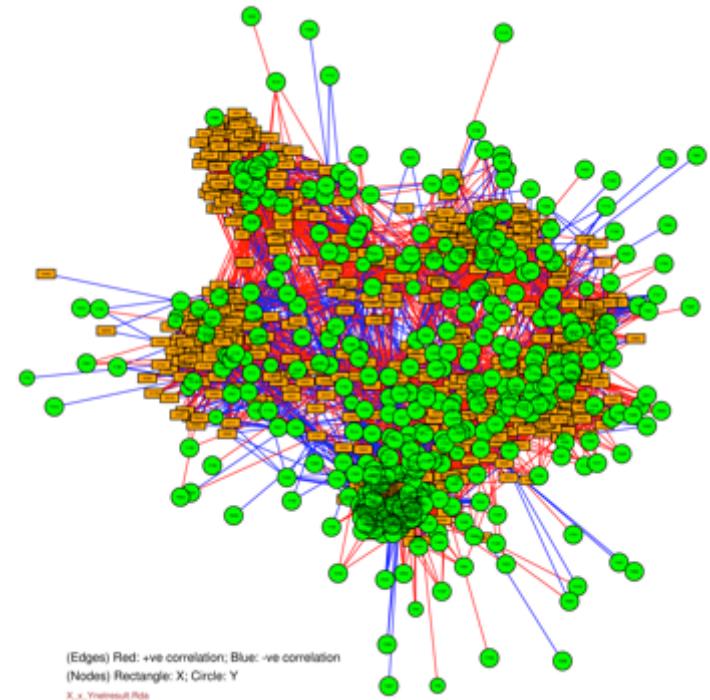
The screenshot shows a Microsoft Excel spreadsheet titled "cluster\_membership\_centrality\_table.txt". The ribbon menu includes Home, Layout, Tables, Charts, SmartArt, Formulas, Data, and Review. The Font tab is selected, showing Calibri (Body) at size 12, bold, italic, and underline options. The Alignment tab shows wrap text and center alignment. The Number tab shows general format with decimal and percentage options. The table has columns A through L and rows 1 through 23. Column A contains row numbers 1 to 23. Column B contains "Node" and "Cluster" labels, followed by numerical values. Column C contains "Name" labels, followed by numerical values. Column D contains "centrality\_vec" labels, followed by numerical values.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Node	Cluster	Name	centrality_vec								
2	X531		4 210.0948032		1							
3	X842		4 461.7220187		0.991							
4	X955		4 461.2204565		0.99							
5	X950		4 522.2812747		0.985							
6	X248		4 504.296908		0.978							
7	X316		4 509.3782577		0.976							
8	X952		4 438.7301667		0.971							
9	X961		4 437.2206166		0.968							
10	X765		4 723.5387798		0.959							
11	X450		4 367.3348744		0.956							
12	X340		4 508.374864		0.955							
13	X714		4 486.2942826		0.945							
14	X441		4 772.6193686		0.929							
15	X880		4 237.0863286		0.918							
16	X440		4 889.4960109		0.913							
17	X733		4 601.5178843		0.912							
18	X215		4 770.6033418		0.906							
19	X891		4 209.0914685		0.901							
20	X452		4 417.3350253		0.899							
21	X769		4 289.1559214		0.898							
22	X234		4 771.6068006		0.89							
23	X697		4 205.1199506		0.889							

# Pairwise results – X<->Y, X<->Z, Y<->Z

xMWASv0.54output

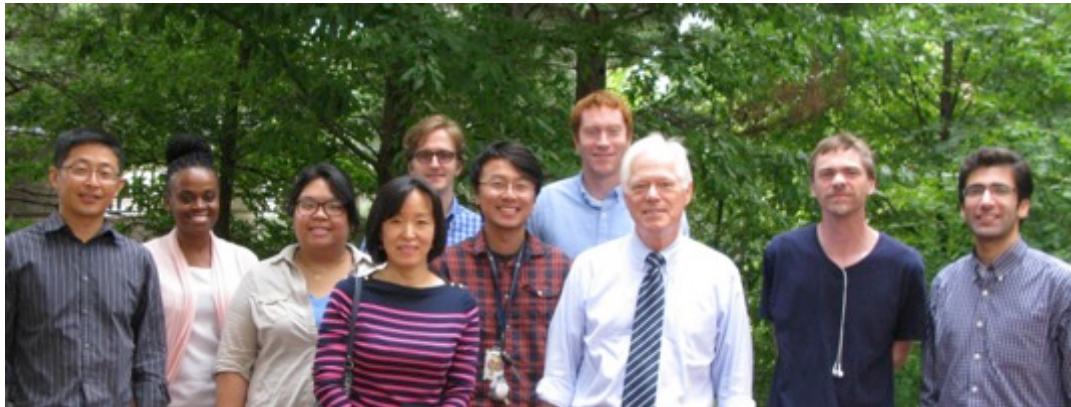
Name	Date Modified	Size	Kind
class-wise_centrality_matrix.txt	Today at 11:29 PM	68 KB	Plain Text
cluster_membership_centrality_table.txt	Today at 11:30 PM	32 KB	Plain Text
Control	Today at 11:32 PM	--	Folder
H1N1	Today at 11:30 PM	--	Folder
InputParameters.txt	Today at 11:30 PM	4 KB	Plain Text
LogTue_Mar_20_00_40_2018.txt	Today at 11:29 PM	5 KB	Plain Text
Multidata_Network_threshold0.7.communities.png	Today at 11:29 PM	1.9 MB	PNG image
Multidata_Network_threshold0.7.linkmatrix.txt	Today at 11:32 PM	541 KB	Plain Text
Multidata_Network_threshold0.7.png	Today at 11:32 PM	1.6 MB	PNG image
Multidata_Network_threshold0.7.cytoscapeall.gml	Today at 11:32 PM	1.7 MB	Document
Network_stats.csv	Today at 11:31 PM	143 bytes	comma...values
NodeID_Name_mapping.txt	Today at 11:32 PM	44 KB	Plain Text
pairwise_results	Today at 11:31 PM	--	Folder
Rplots.pdf	Today at 11:31 PM	4 KB	Adobe...cument
Rplots1.pdf	Today at 11:30 PM	4 KB	Adobe...cument
Rplots2.pdf	Today at 11:31 PM	4 KB	Adobe...cument
X_X_Y_association_matrix_threshold0.7.txt	Today at 11:31 PM	1.1 MB	Plain Text
X_X_Y_association_network_threshold0.7.png	Today at 11:31 PM	3.4 MB	PNG image
X_x_YBoolean_association_matrix_threshold0.7.txt	Today at 11:30 PM	980 KB	Plain Text
X_x_Z_association_matrix_threshold0.7.txt	Today at 11:31 PM	5 KB	Plain Text
X_x_Z_association_network_threshold0.7.png	Today at 11:31 PM	1.8 MB	PNG image
X_x_ZBoolean_association_matrix_threshold0.7.txt	Today at 11:30 PM	5 KB	Plain Text
Y_x_Z_association_matrix_threshold0.7.txt	Today at 11:30 PM	6 KB	Plain Text
Y_x_Z_association_network_threshold0.7.png	Today at 11:31 PM	2.1 MB	PNG image
Y_x_ZBoolean_association_matrix_threshold0.7.txt	Today at 11:30 PM	5 KB	Plain Text
README.txt	Today at 11:29 PM	2 KB	Plain Text





EMORY

## 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?