Computational methods for data integration

March 21, 2018

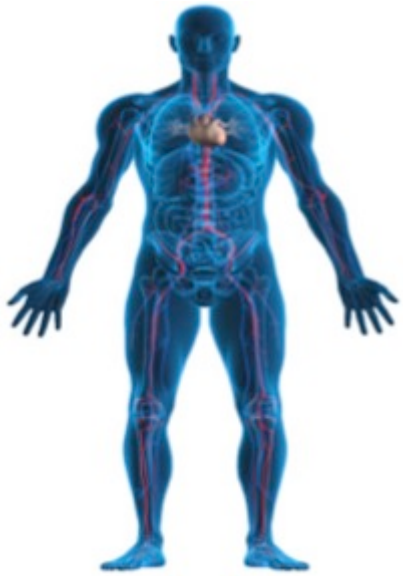
Karan Uppal, PhD
Assistant Professor of Medicine
Director of Computational Metabolomics & Integrative Omics,
Clinical Biomarkers Laboratory
Emory University

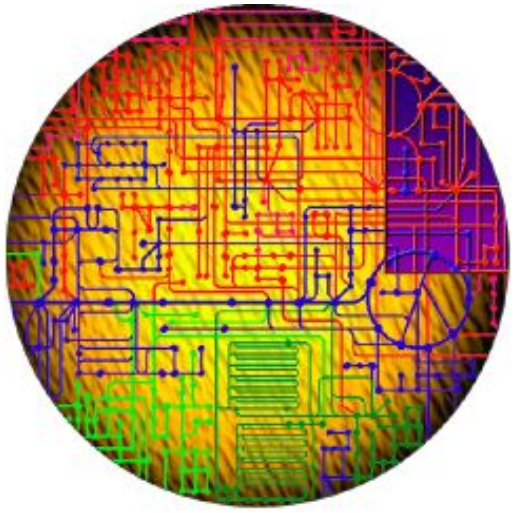
Learning Objectives

- 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

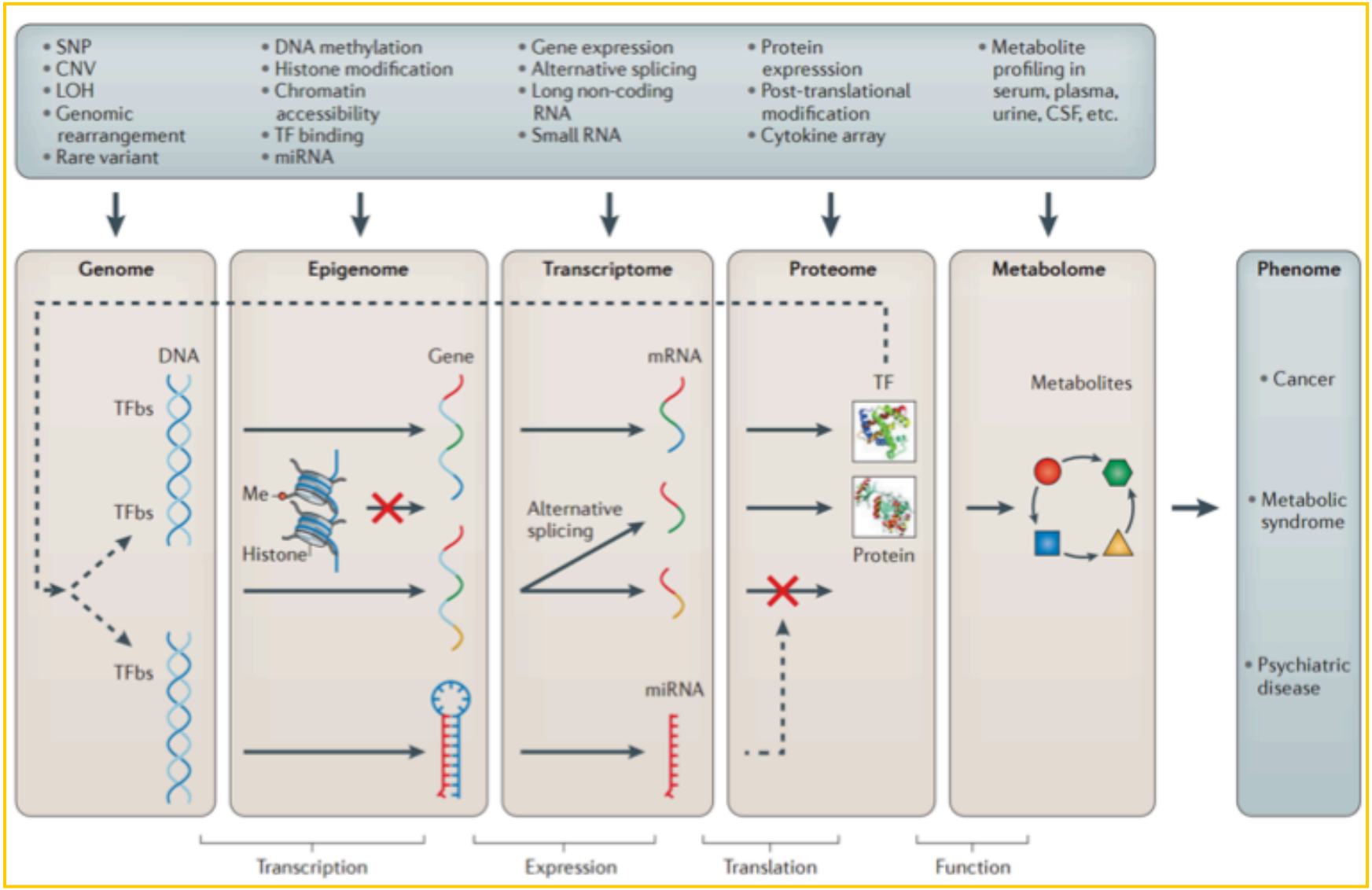




Exposures
Internal measurements
Disease states

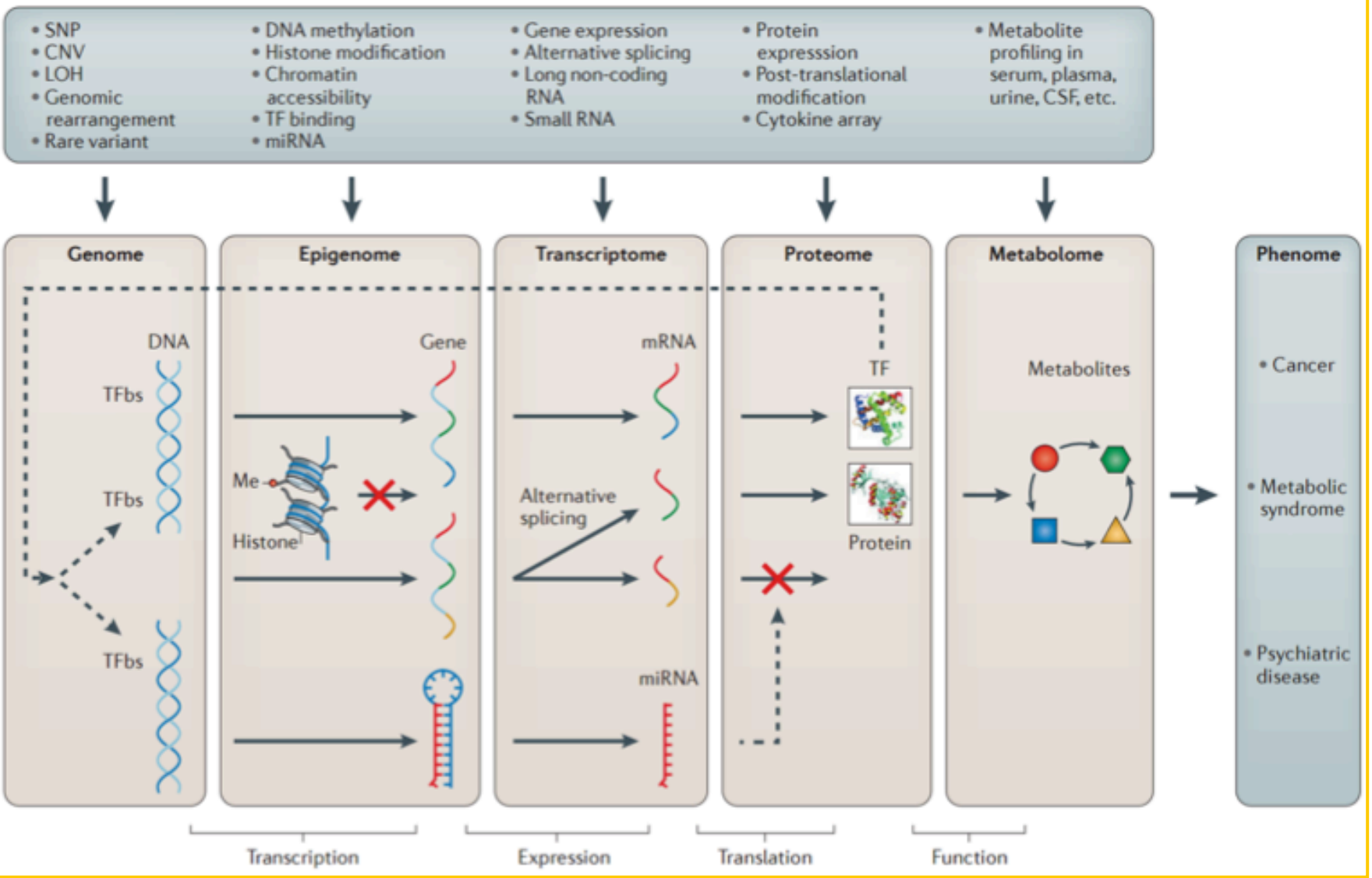
Systems Biology
"*Integrative approach in which scientists study pathways and networks will touch all areas of biology, including drug discovery*"
C. Henry and C. Washington

Dissecting the Biological system via -omics



Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. 2015. Methods of integrating data to uncover genotype-phenotype interactions. Nat Rev Genet *16:85-97.*

Dissecting the Biological system via -omics



"Information Overload": >10,000 variables per –omics experiment

Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. 2015. Methods of integrating data to uncover genotype-phenotype interactions. Nat Rev Genet *16:85-97.*

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

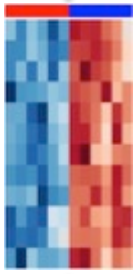
Main approaches for data integration

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

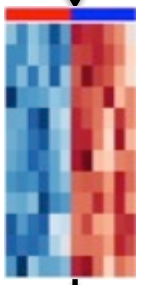


Differentially expressed metabolites

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

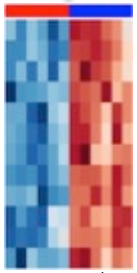
Pathway analysis for genes

Common pathways or pathway rank aggregation

Pathway or knowledge-based integration

Metabolomics data
(n subjects X p metabolites)

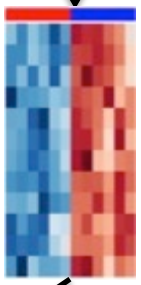
	M1	M2	-	Mp
Subject1	199	19	-	100
Subject2	10	40		90
-	-	-		-
SubjectN	50	30	-	20



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

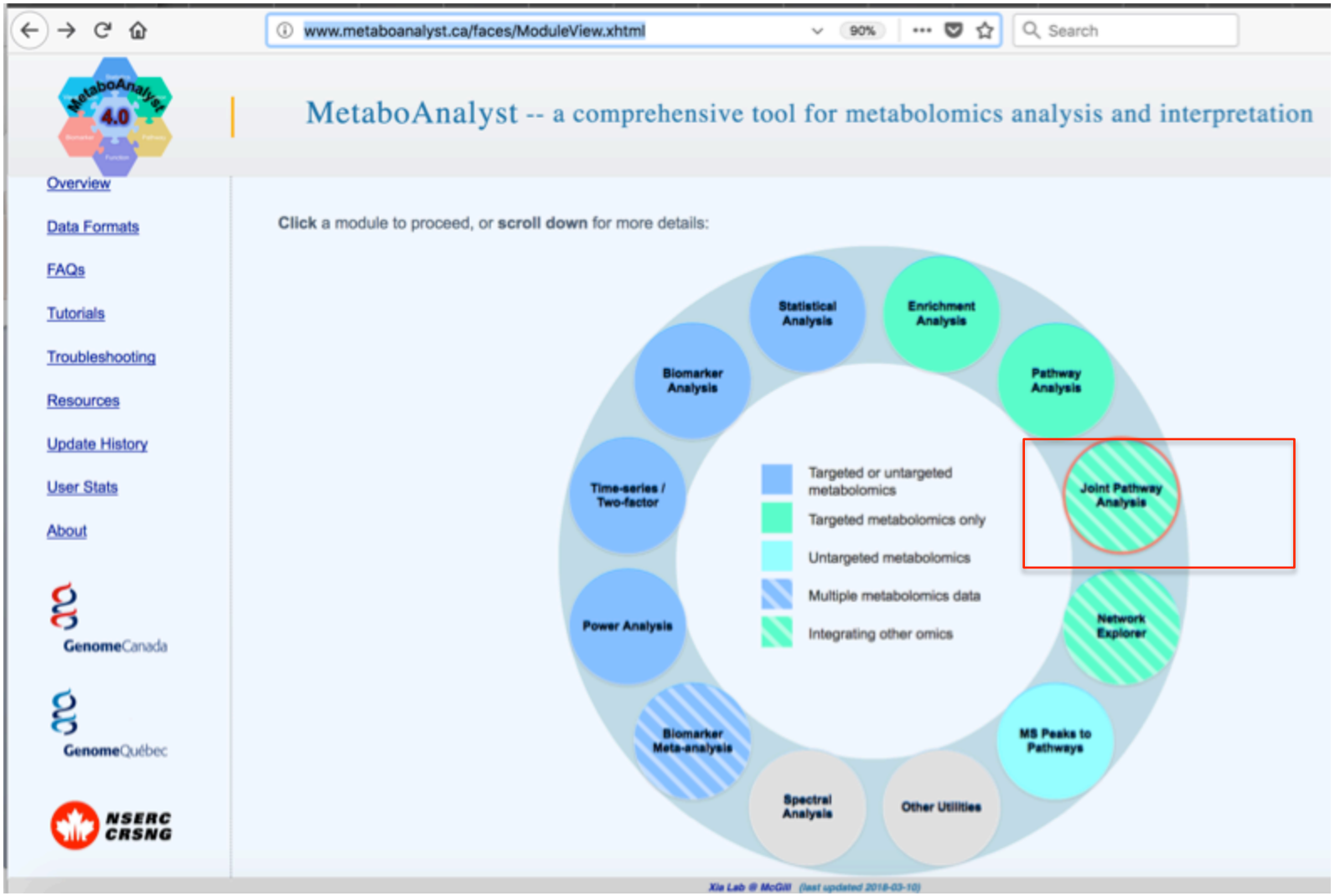


Differentially expressed genes

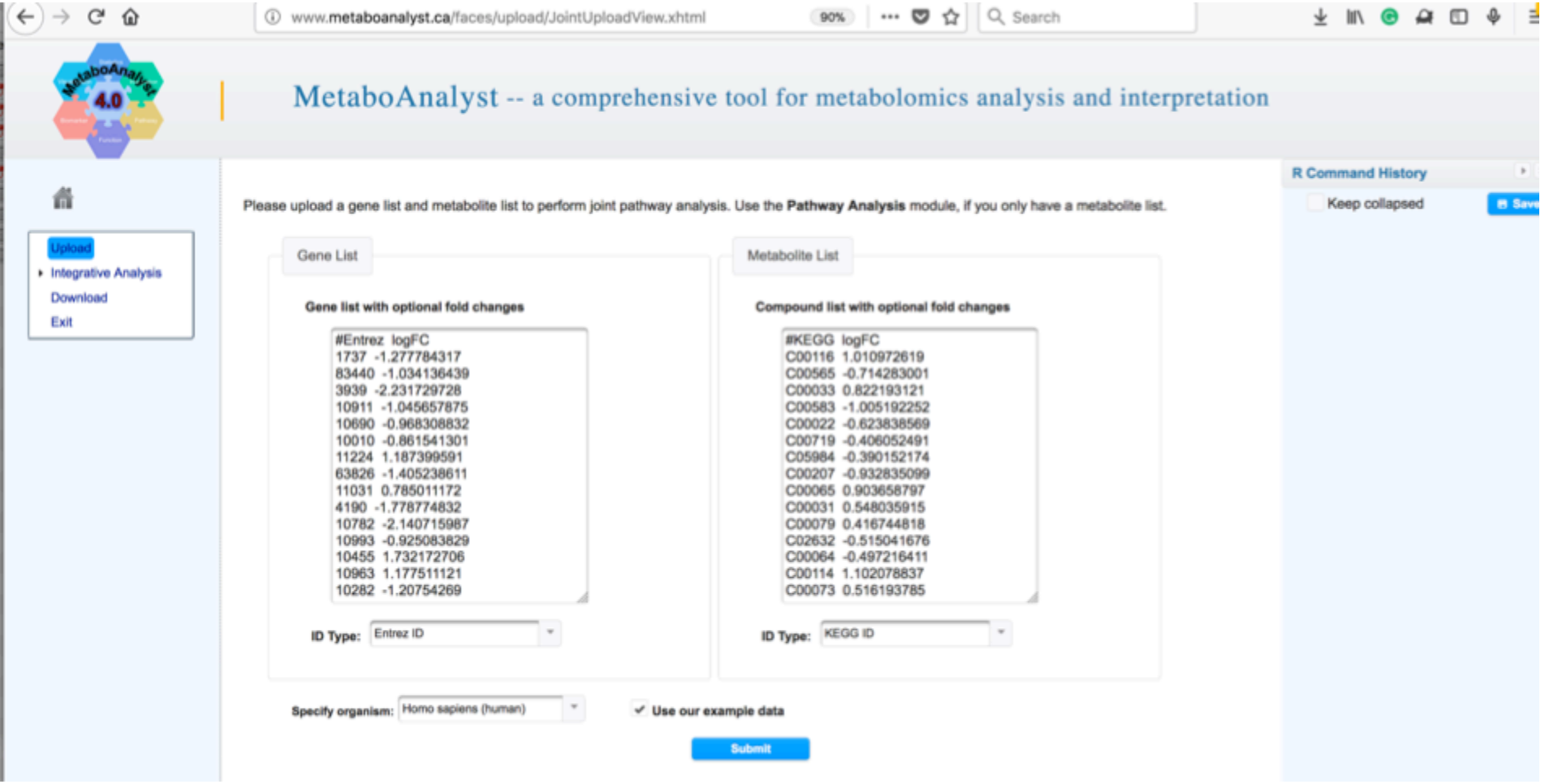
Pathway analysis using genes and metabolites (joint)

MetaboAnalyst

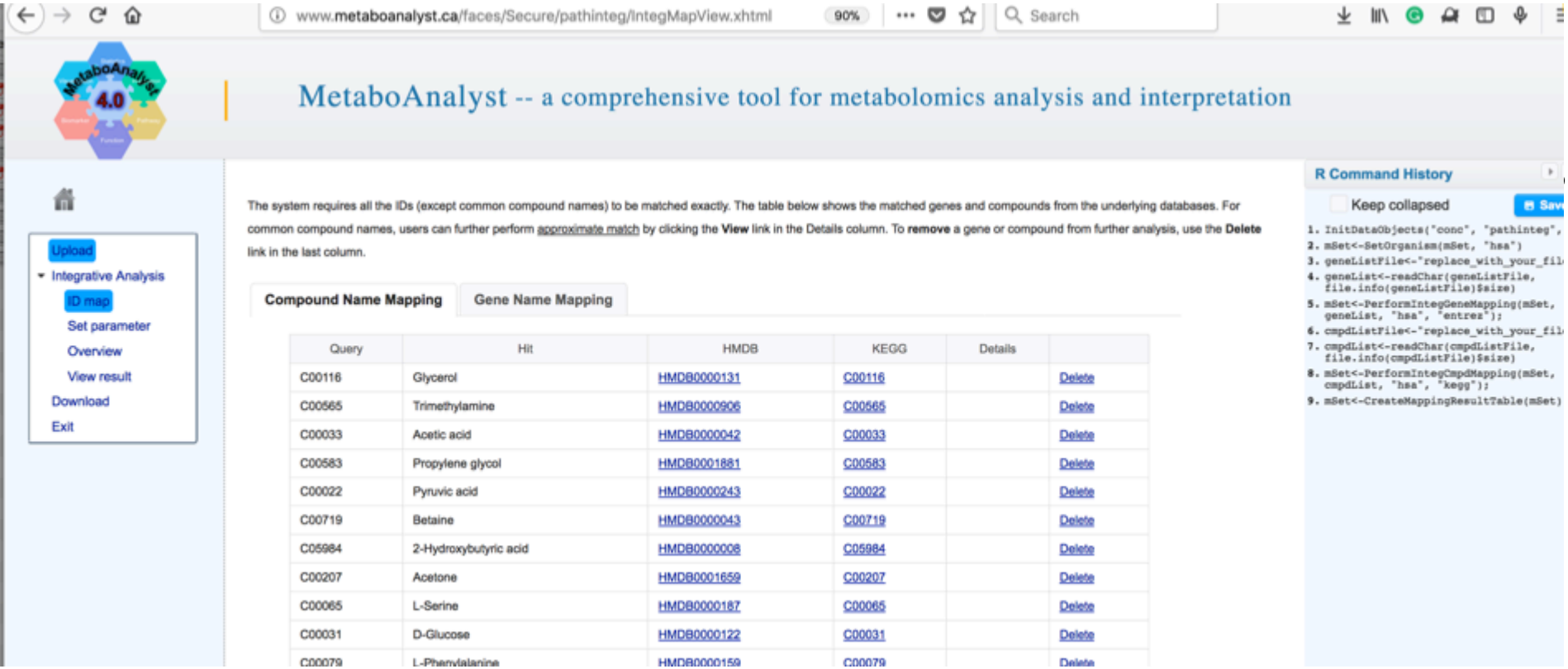
(http://www.metaboanalyst.ca/faces/ModuleView.xhtml)



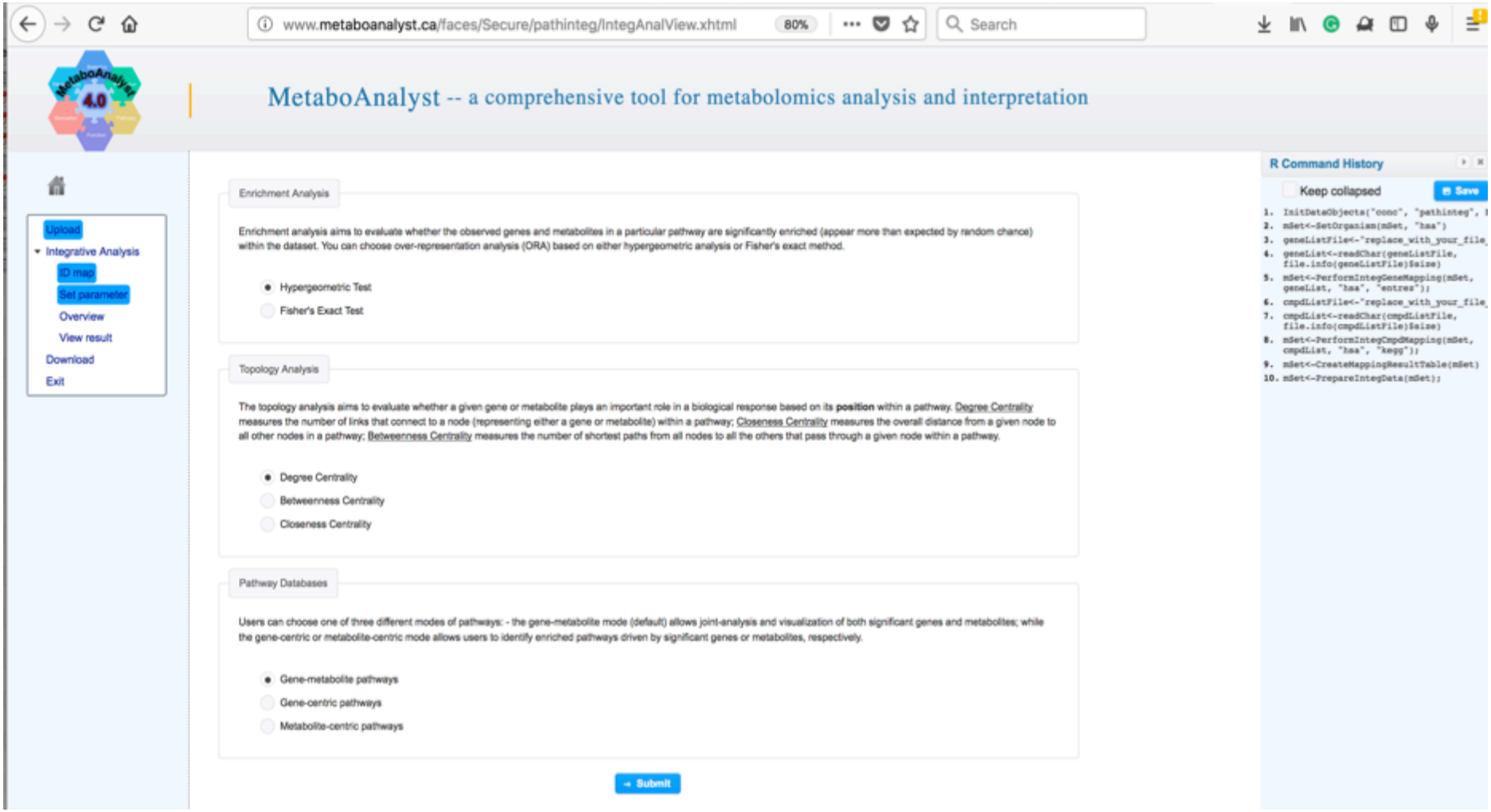
Upload data and submit



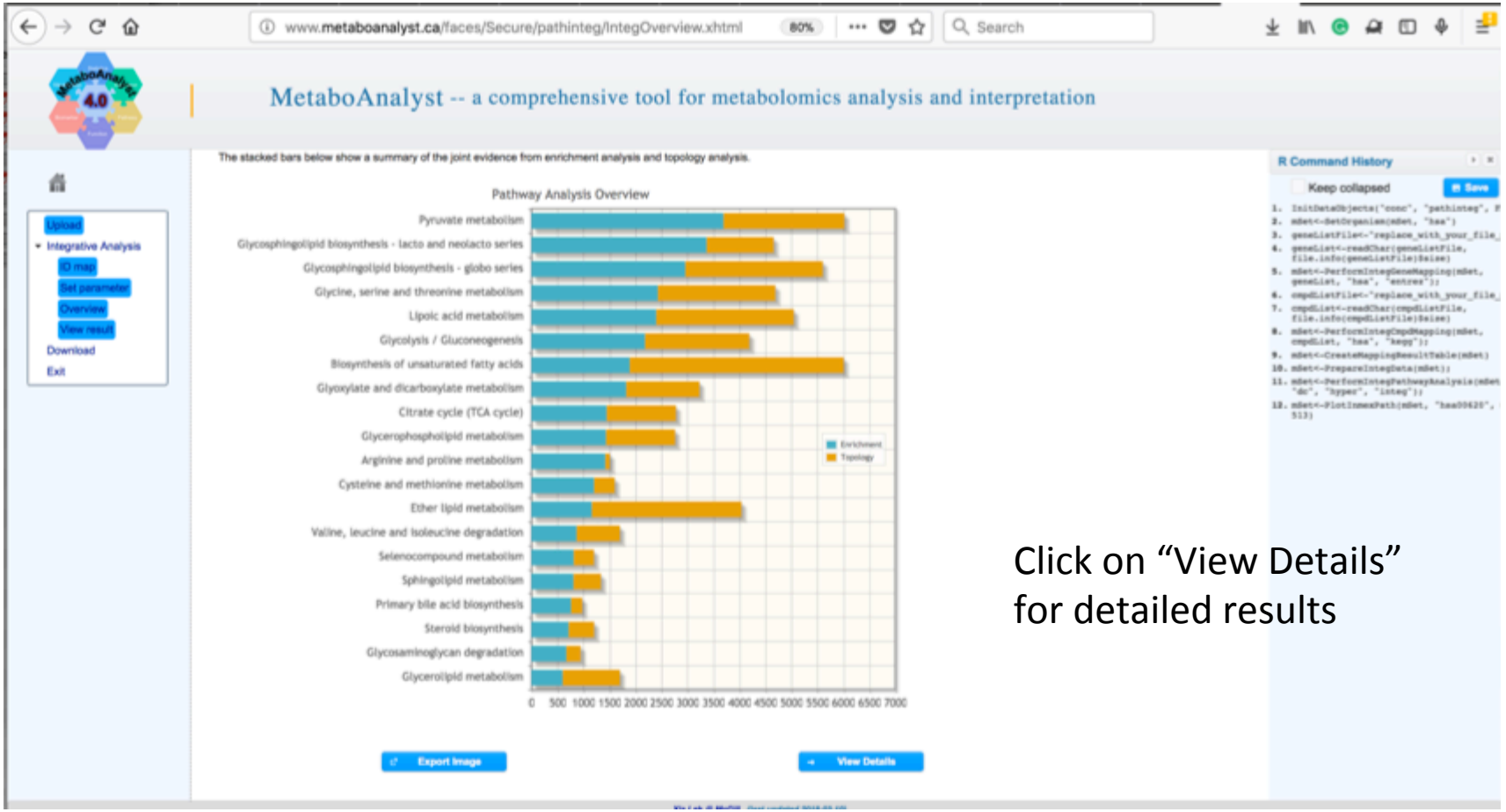
Name Mapping – click Submit



Set parameters

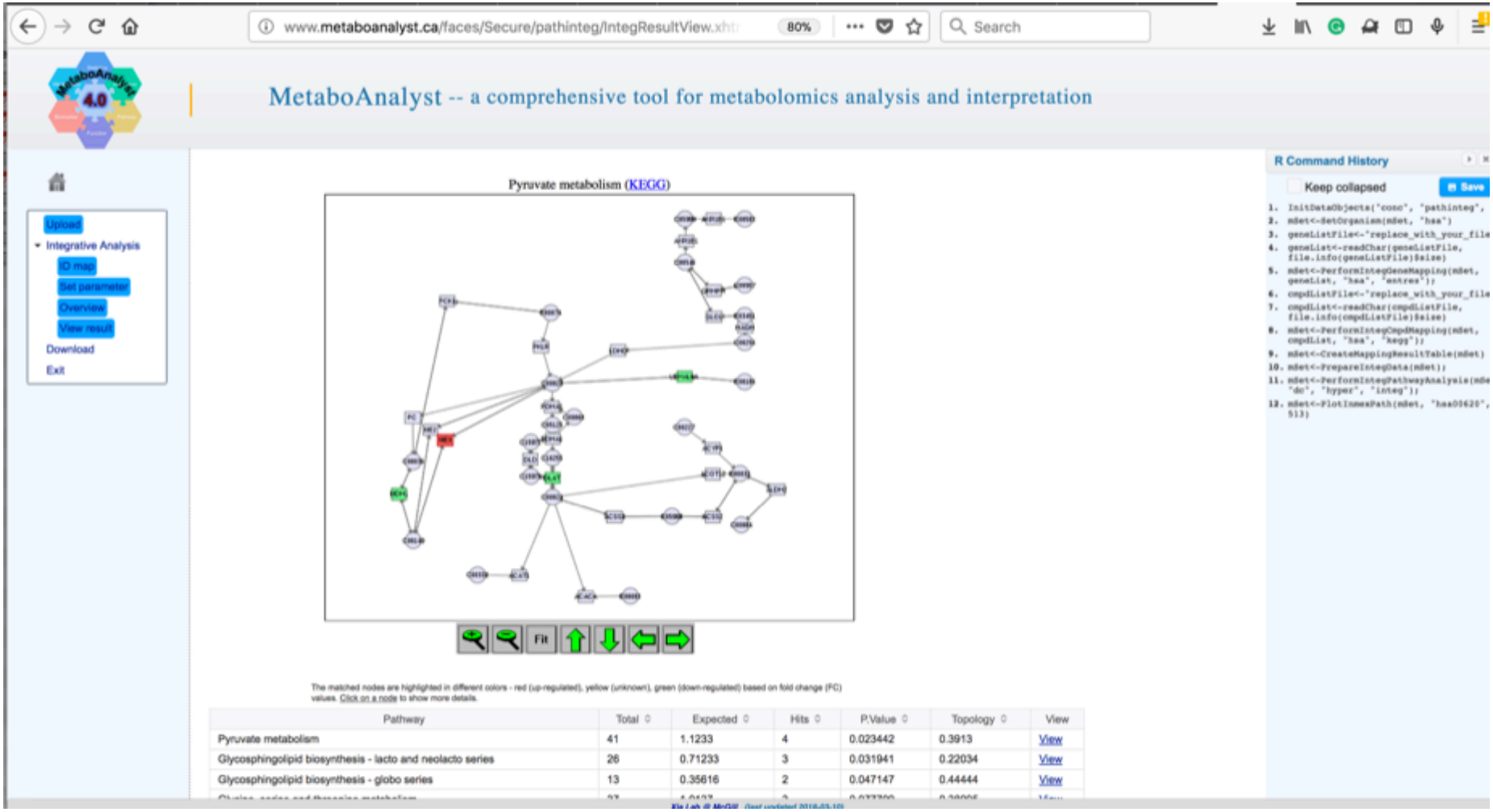


Results - Overview



Click on “View Details” for detailed results

Results - Details



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

Workflow

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

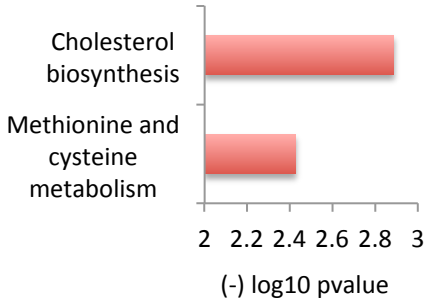
Univariate

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

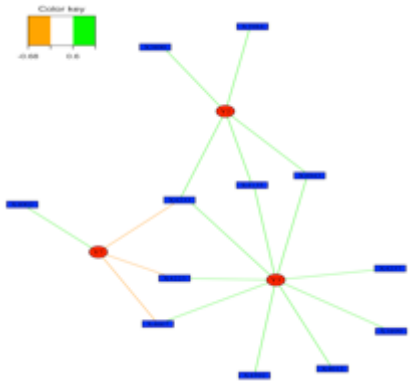
Multivariate

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

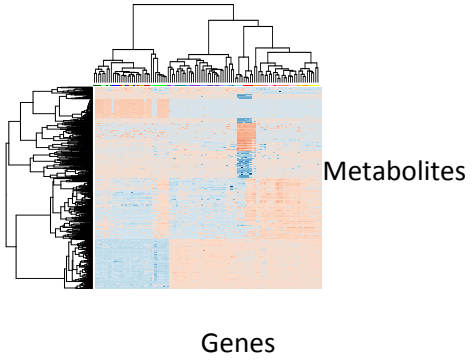
Pathway enrichment



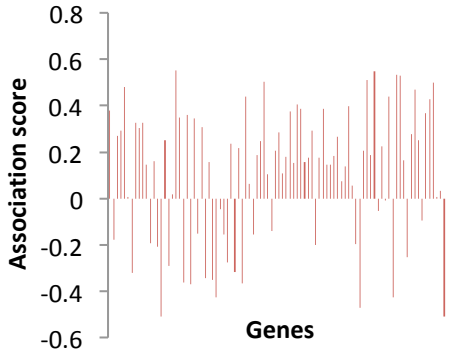
Relevance networks



Clustering

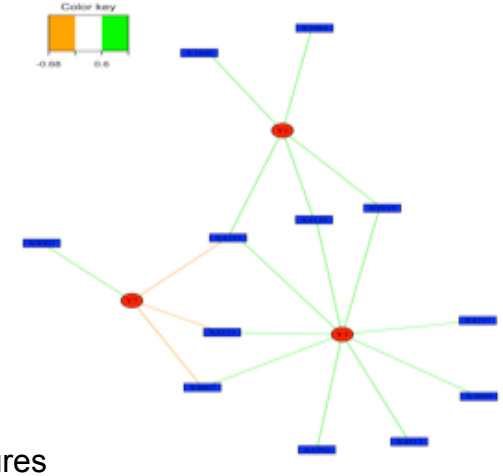


Targeted investigation (e.g.: Arginine x Transcriptome)



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 a 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 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

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

Browse... No file selected.

GenBank ID: e.g. NAT1, ABL1

Proteomics

Browse... No file selected.

Uniprot Accession: e.g. P31946, P62258

Metabolomics

Browse... No file selected.

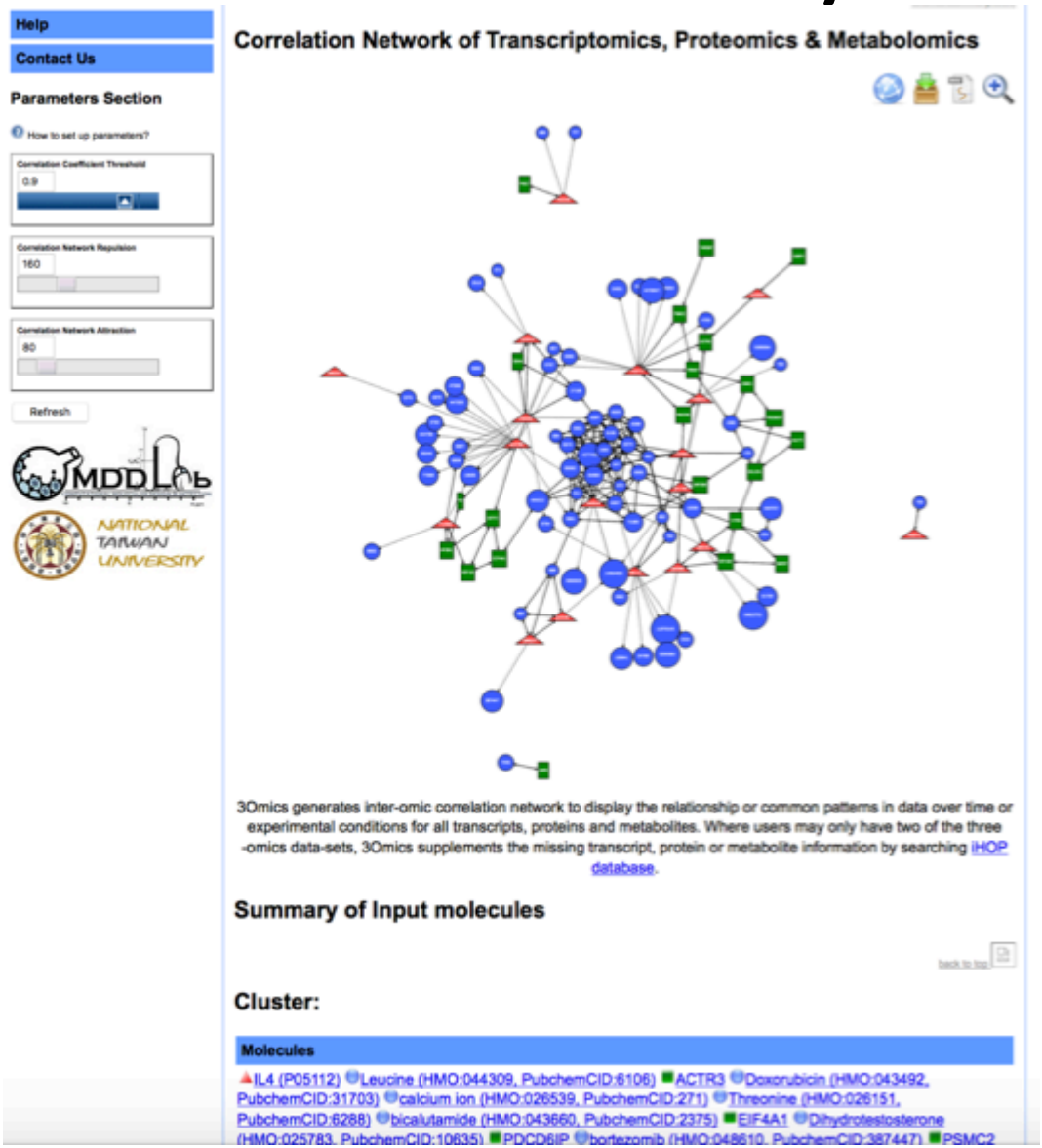
Data format

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

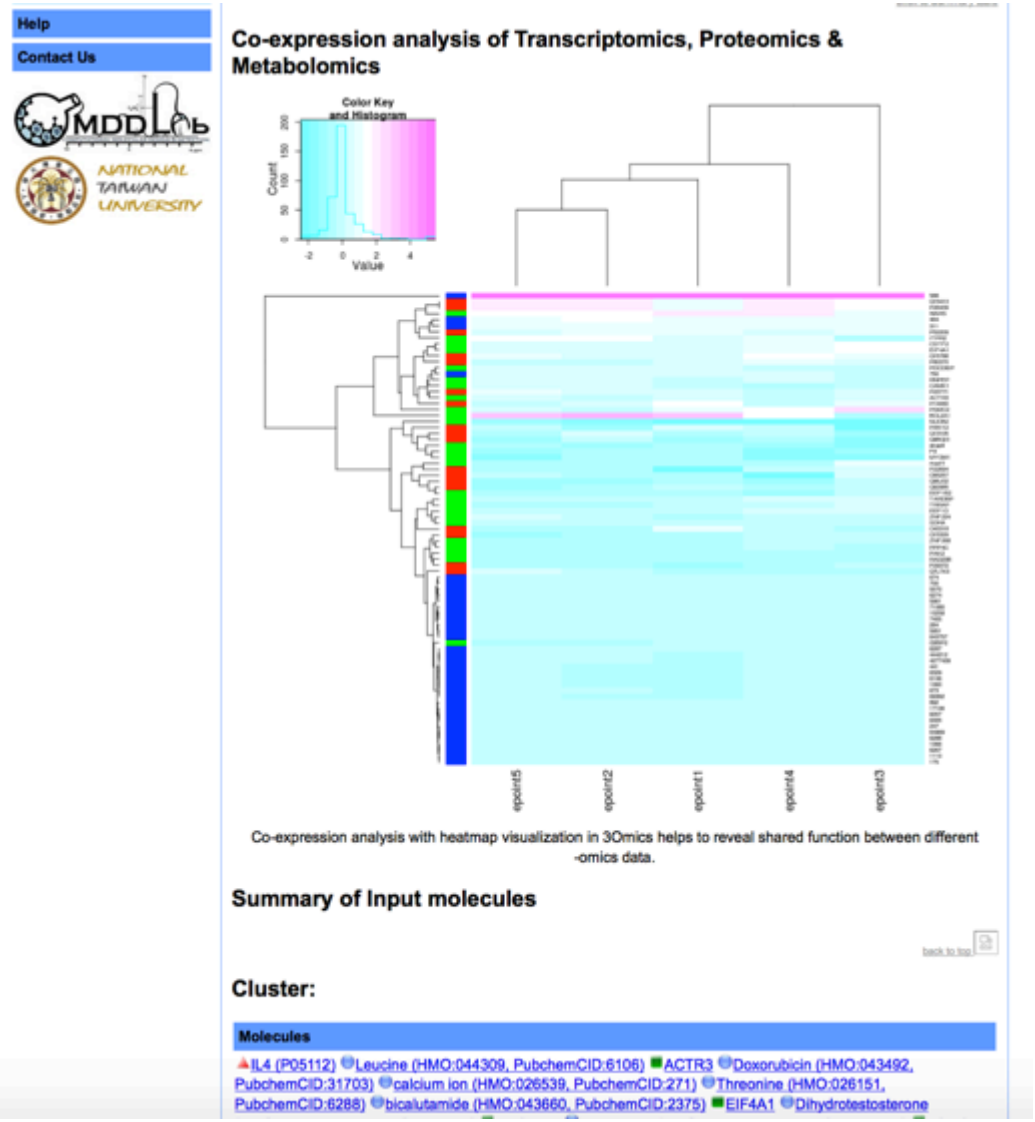
Samples

Variables	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



Co-expression analysis



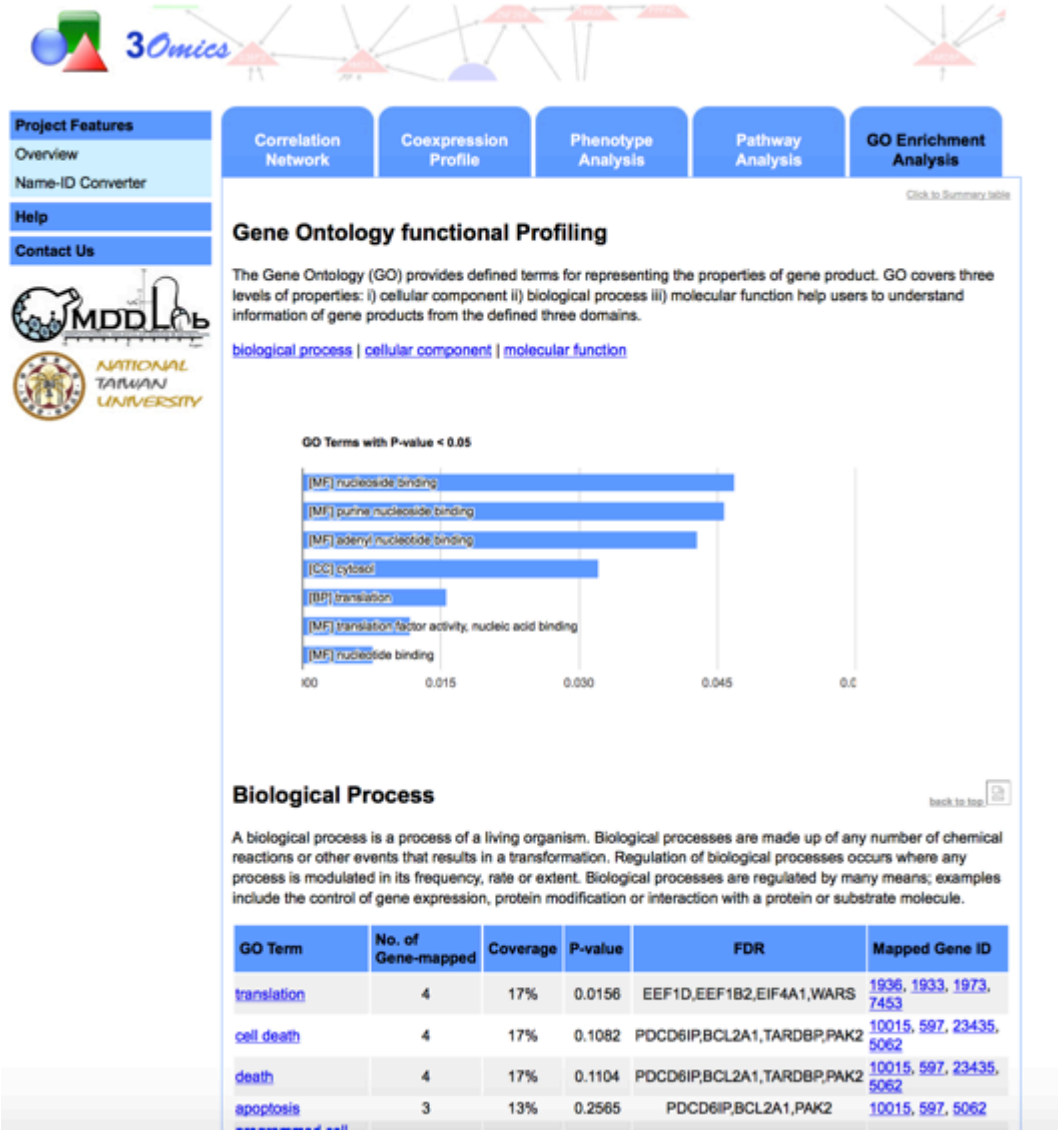
Rows: Variables

Columns: Samples

Phenotype analysis

Pathway analysis

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 $cov(X_u, Y_v)$

  where

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

  H is the number of PLS-DA dimensions

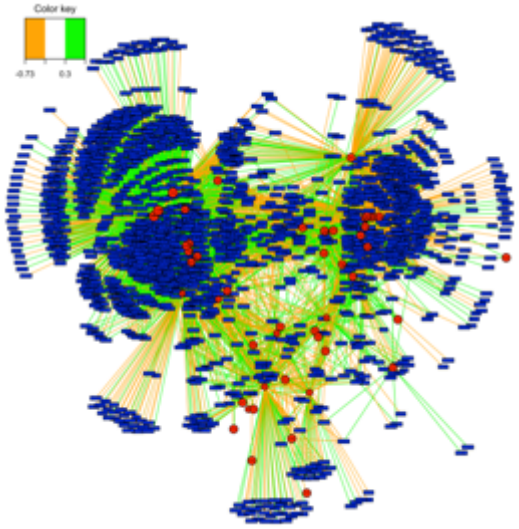
  A Lasso based optimization is used to select most relevant variables

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)

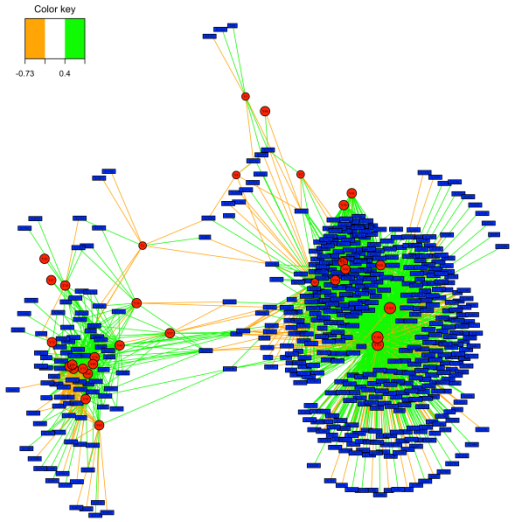
Legend

Circles: microbial species

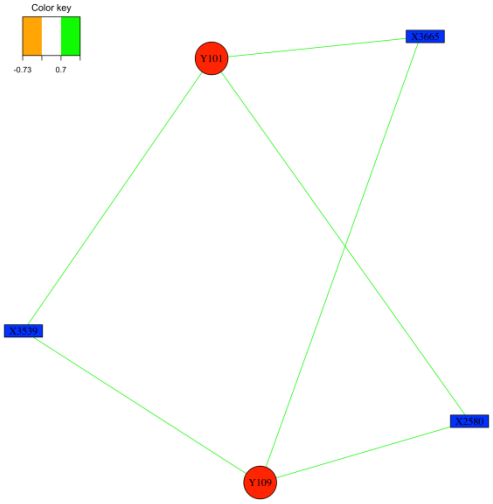
Rectangles: metabolome features



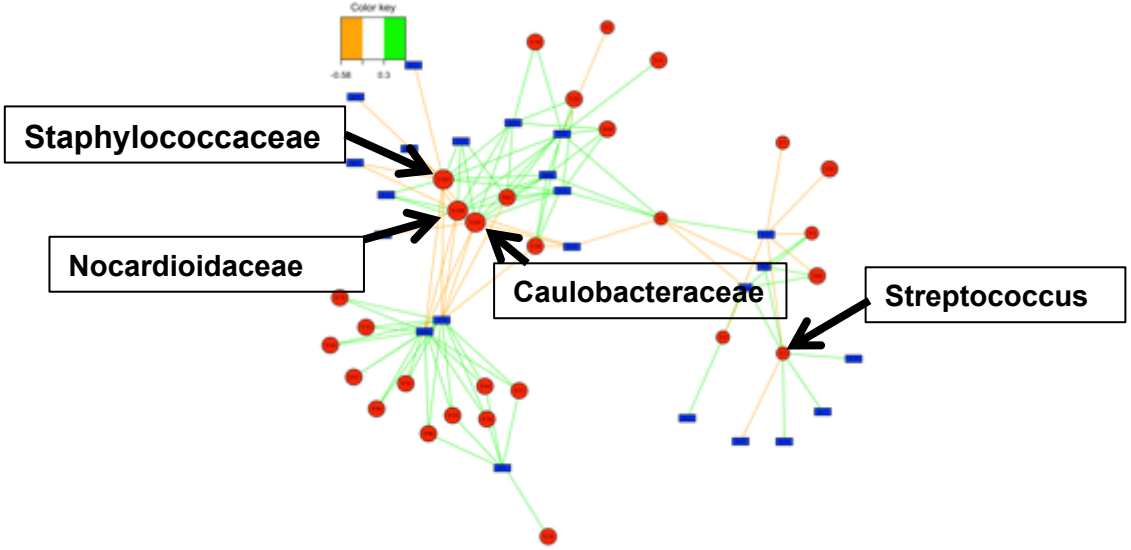
A. Association threshold: 0.3



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)

Integrating more than two datasets

Proteomics data
(n subjects X s proteins)

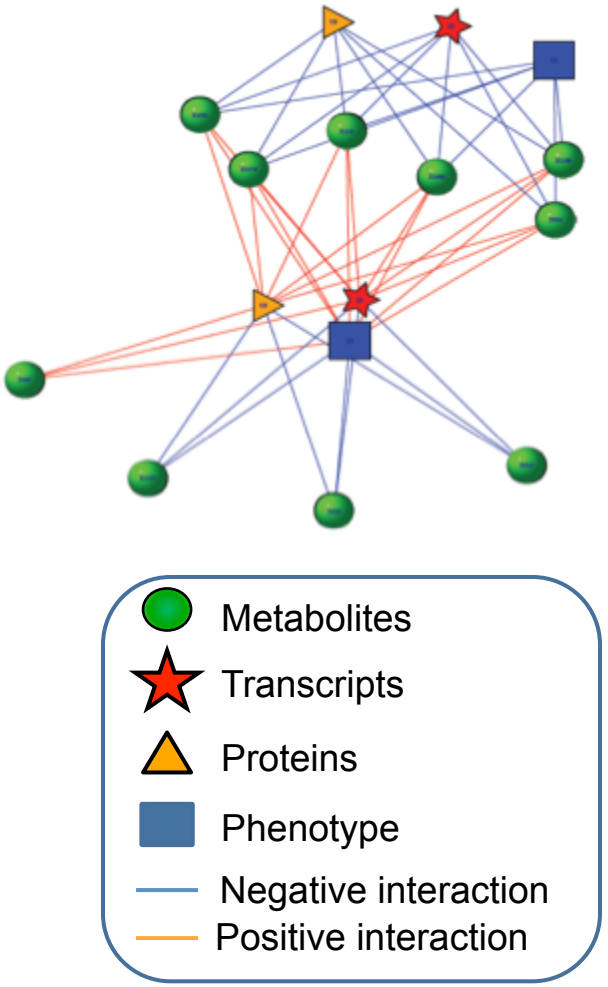
	E1	E2	-	Es
Subject1	199	19	-	100
Subject2	10	40		90
-	-	-		-
SubjectN	50	30	-	20

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



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

A.

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

↓

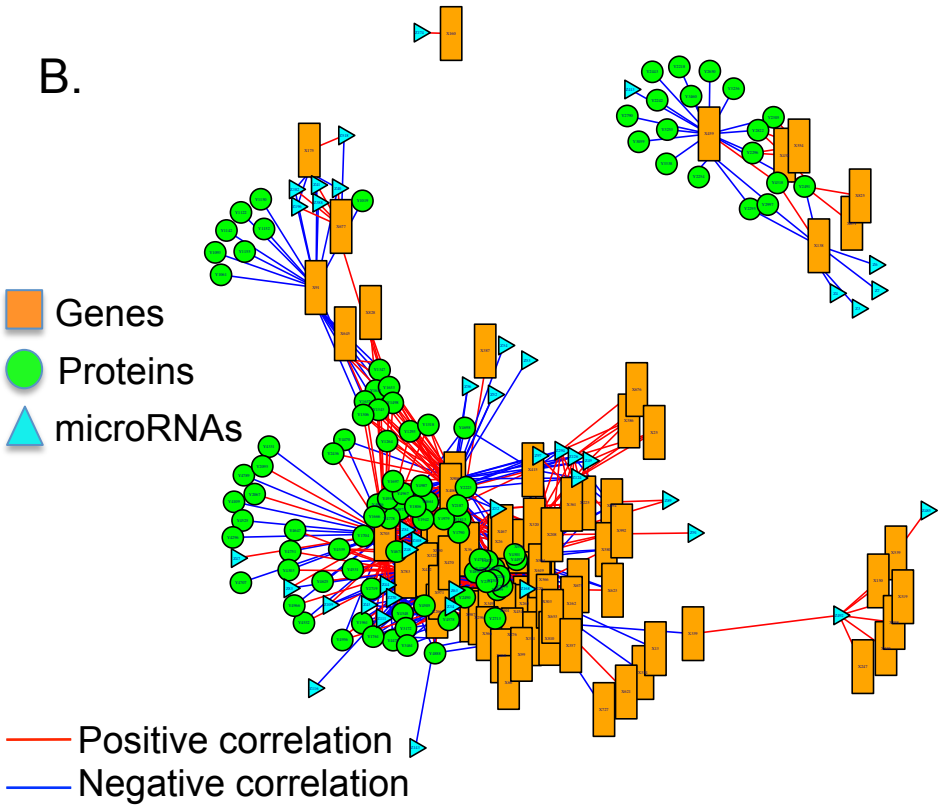
Generate a k-partite graph, G, using the union of edge list matrices

↓

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

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

B.



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

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

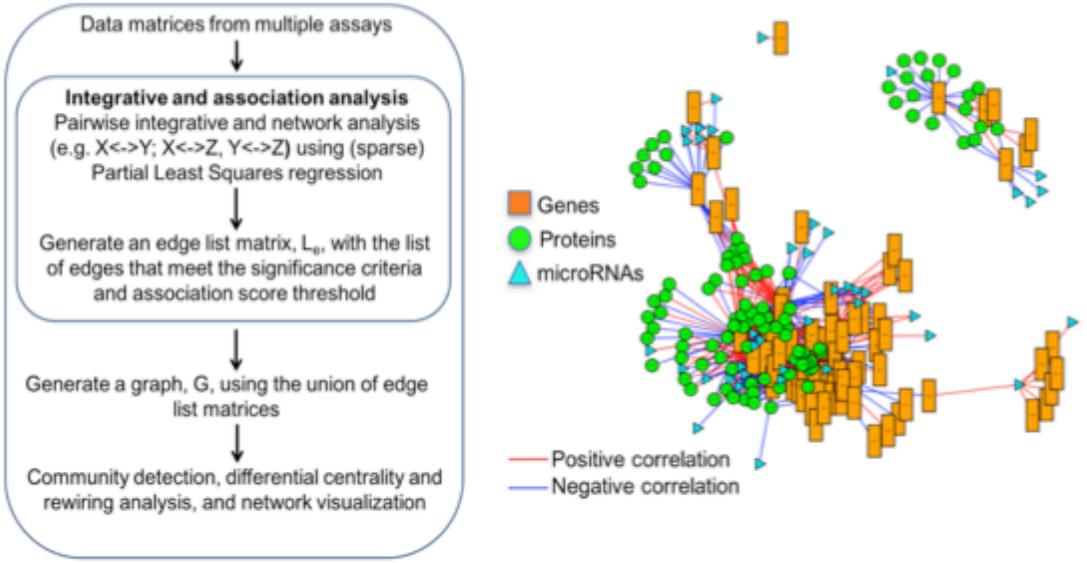
xMWAS - a data-driven integration and network analysis tool (v0.54)

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:

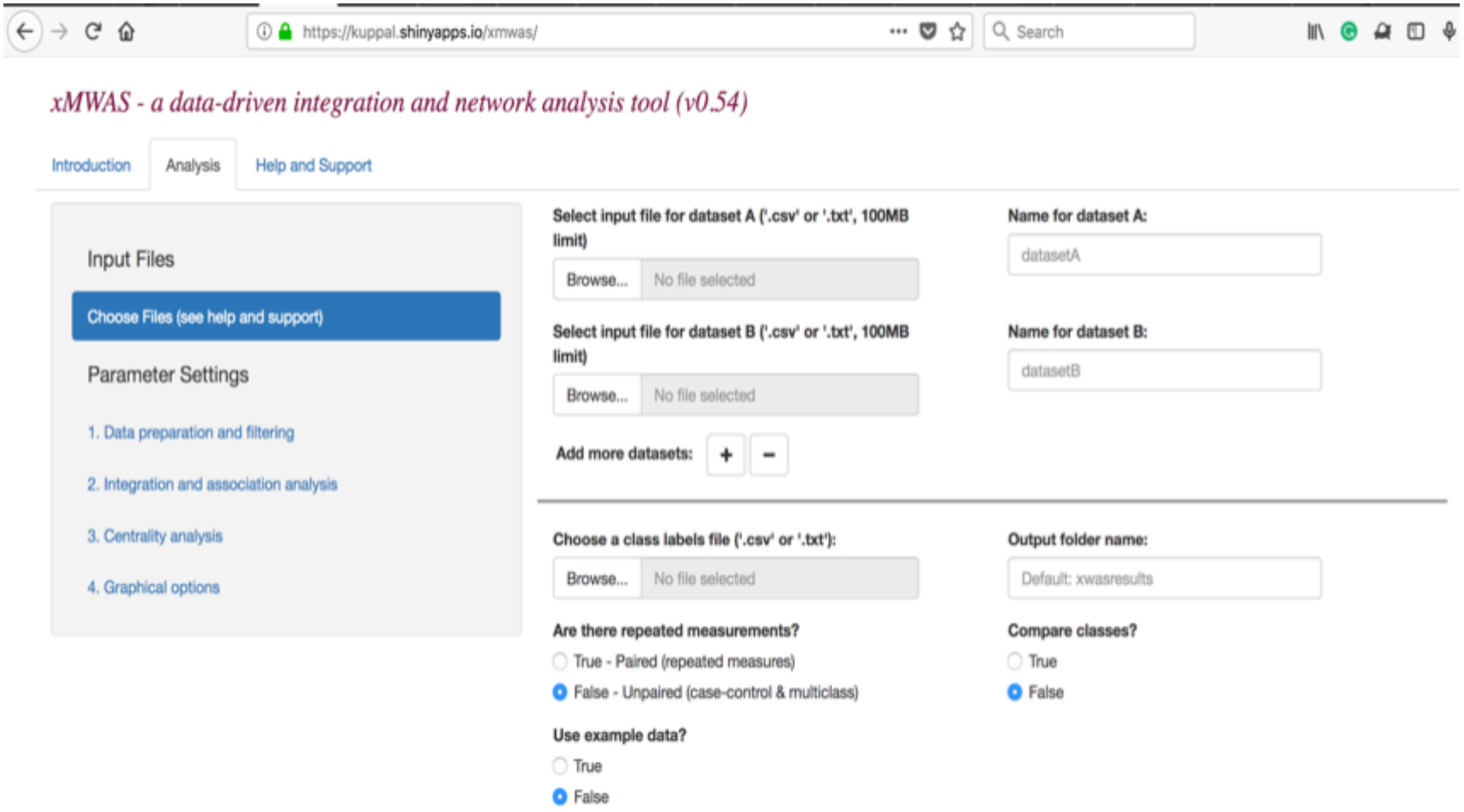
`library(devtools);install_github("kuppal2/xMWAS")`



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) and Karan Uppal (kuppal2@emory.edu) at Clinical Biomarkers Laboratory , Emory University, Atlanta, GA, USA

Step 1. Upload data files



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

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

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):

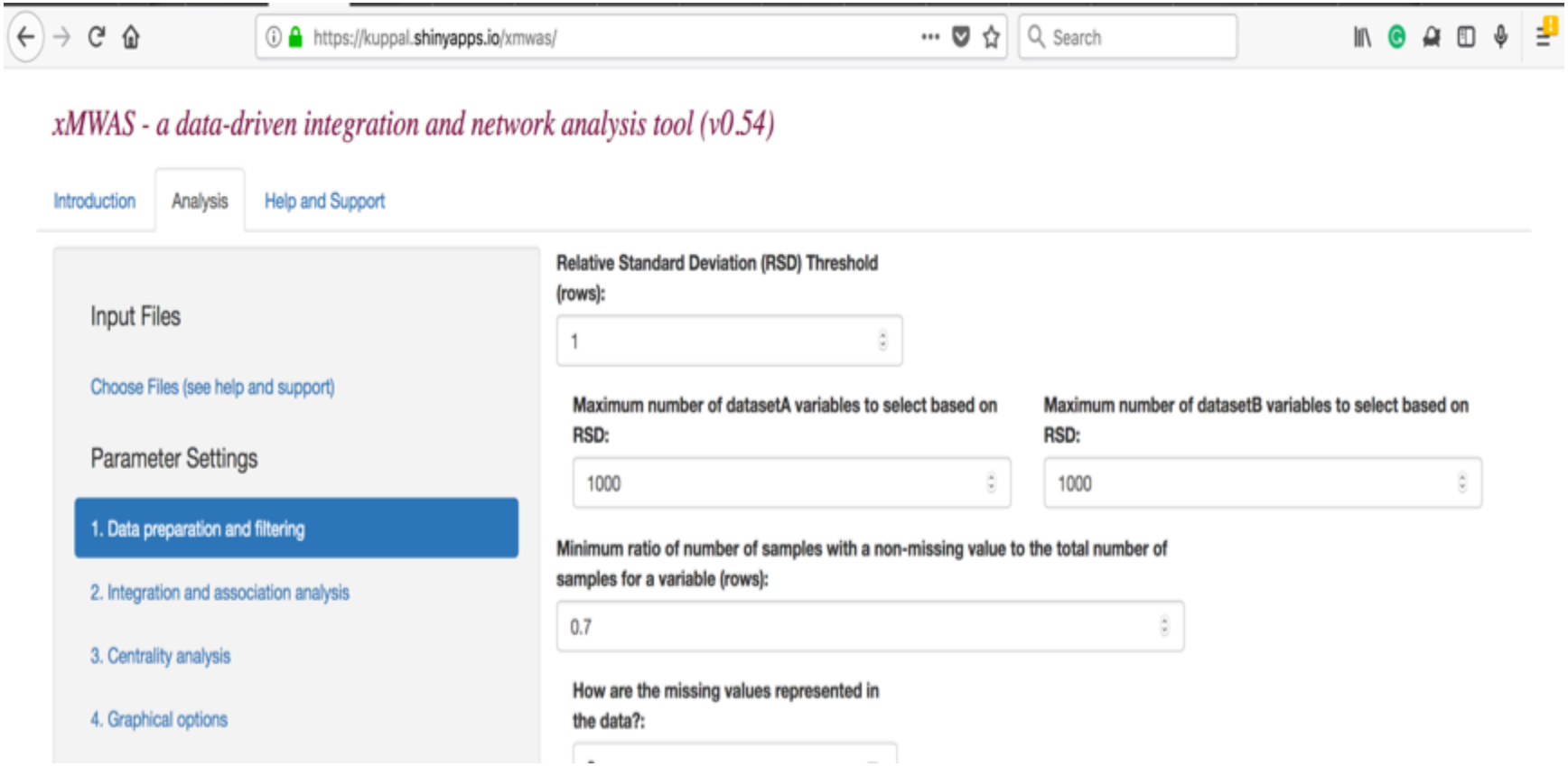
FileName	Class
CNS.SF_268	CNS
CNS.SF_295	CNS
CNS.SF_539	CNS
CNS.SNB_19	CNS
CNS.SNB_75	CNS
CNS.U251	CNS
LE.CCRF_CEM	LE
LE.HL_60	LE
LE.K_562	LE
LE.MOLT_4	LE
LE.RPMI_8226	LE
LE.SR	LE

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):

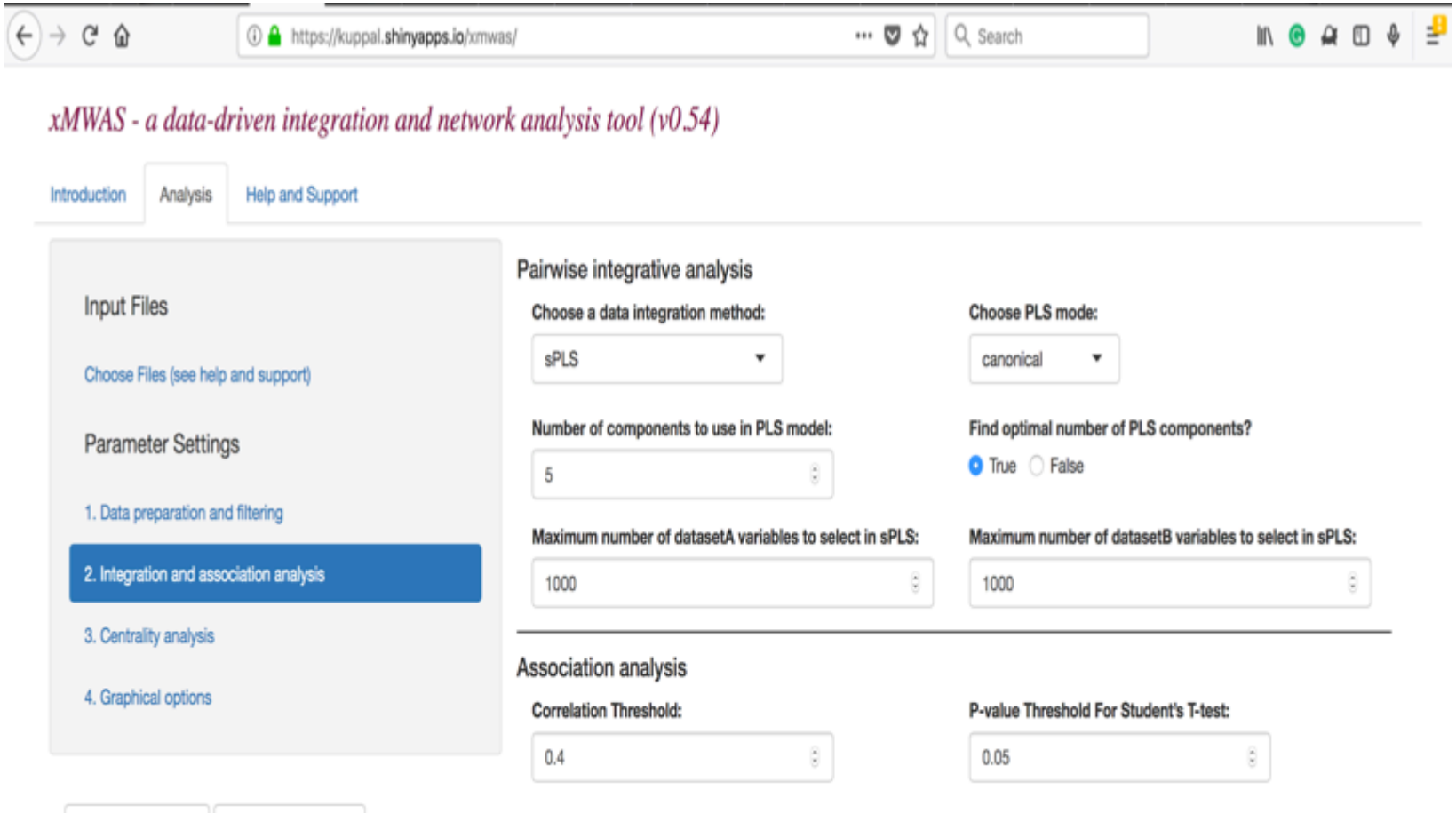
FileName	Subject	Factor1
S1	Subject001	TP1
S2	Subject002	TP1
S3	Subject003	TP1
S4	Subject004	TP1
S5	Subject005	TP1

Step 2. Data preprocessing and filtering



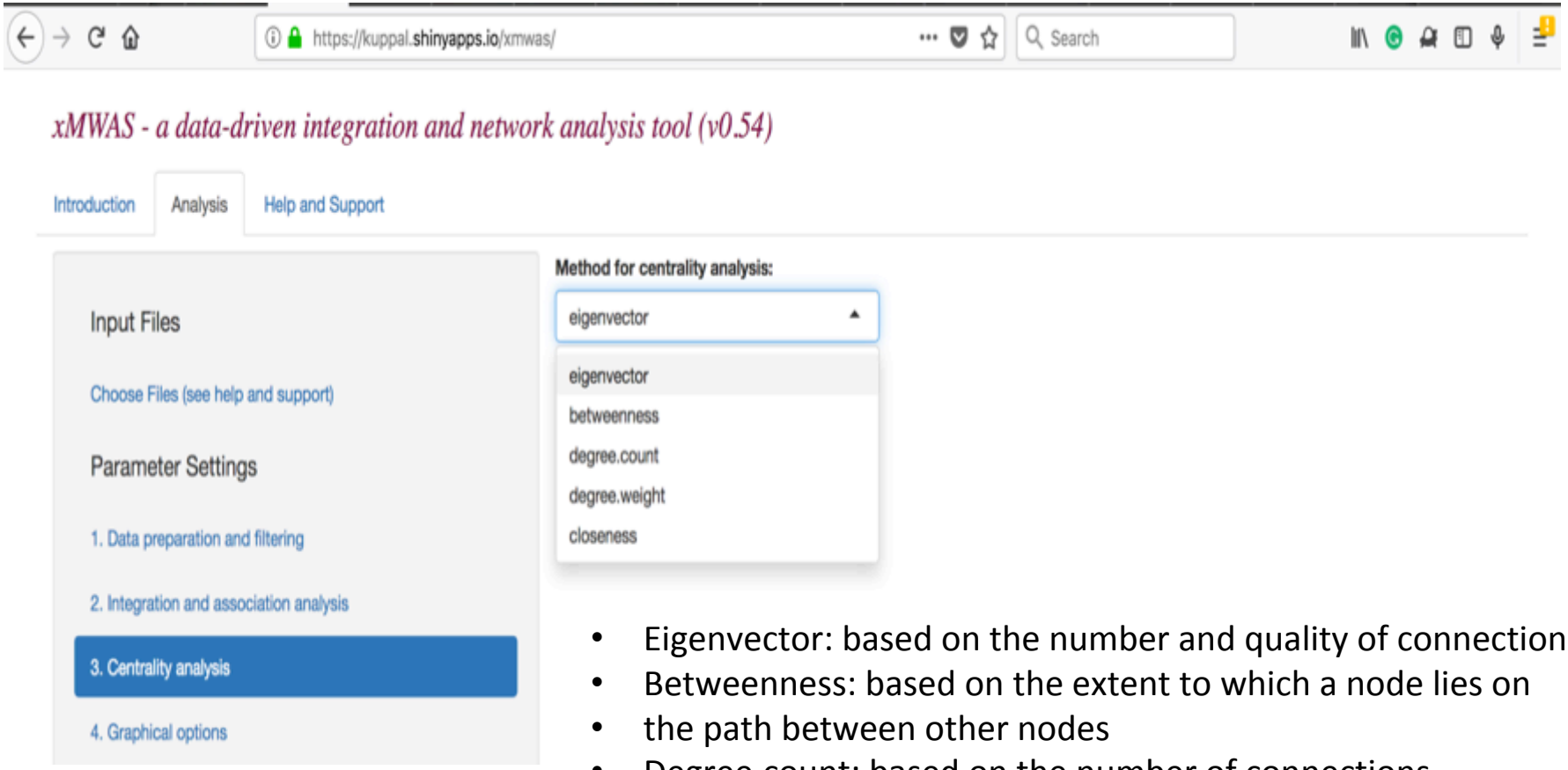
https://github.com/kuppal2/xMWAS/blob/master/example_manual_tutorial/xMWAS-manual.pdf

Step 3. Set parameters for integration and association analysis



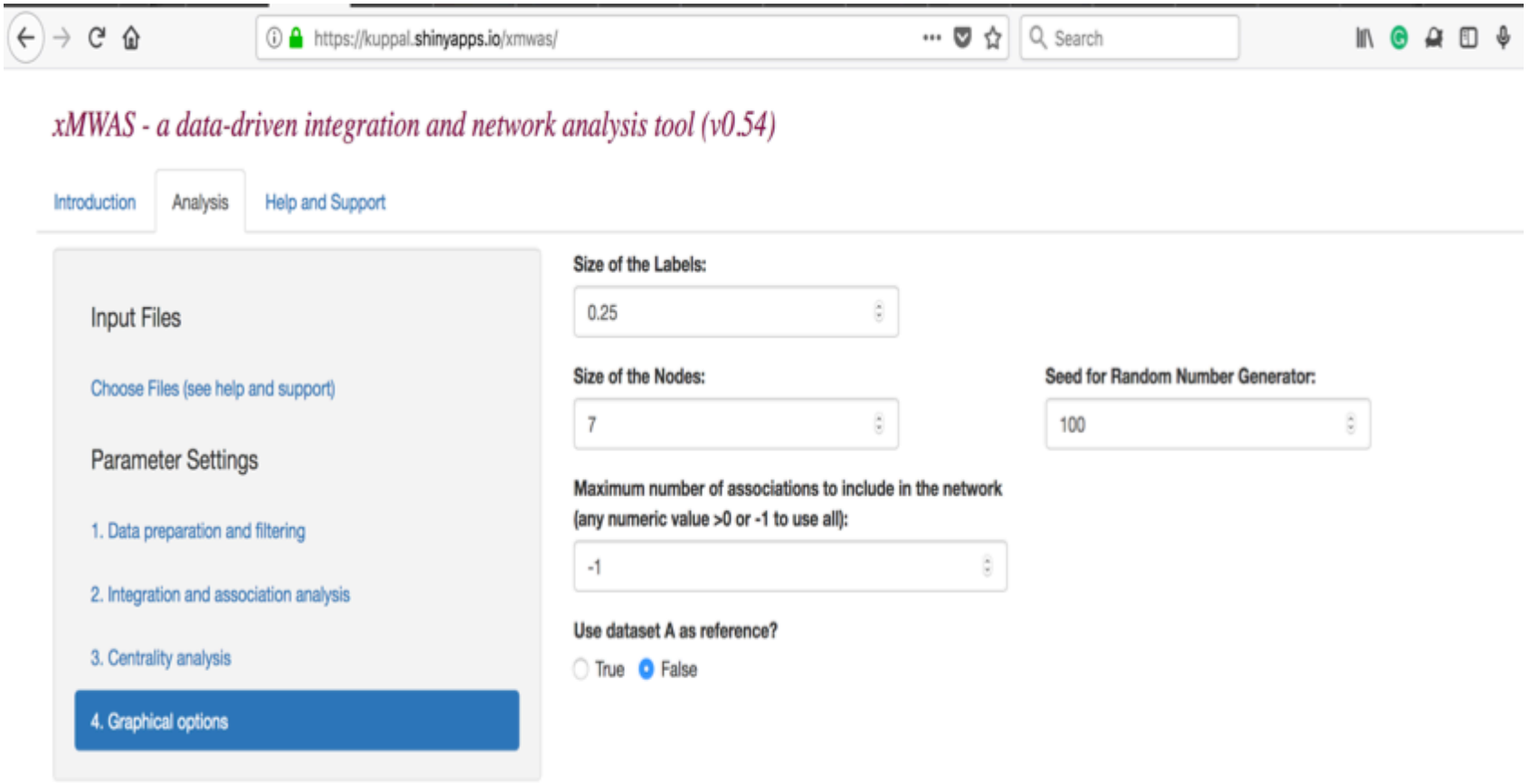
https://github.com/kuppal2/xMWAS/blob/master/example_manual_tutorial/xMWAS-manual.pdf

Step 4. Select method for centrality analysis



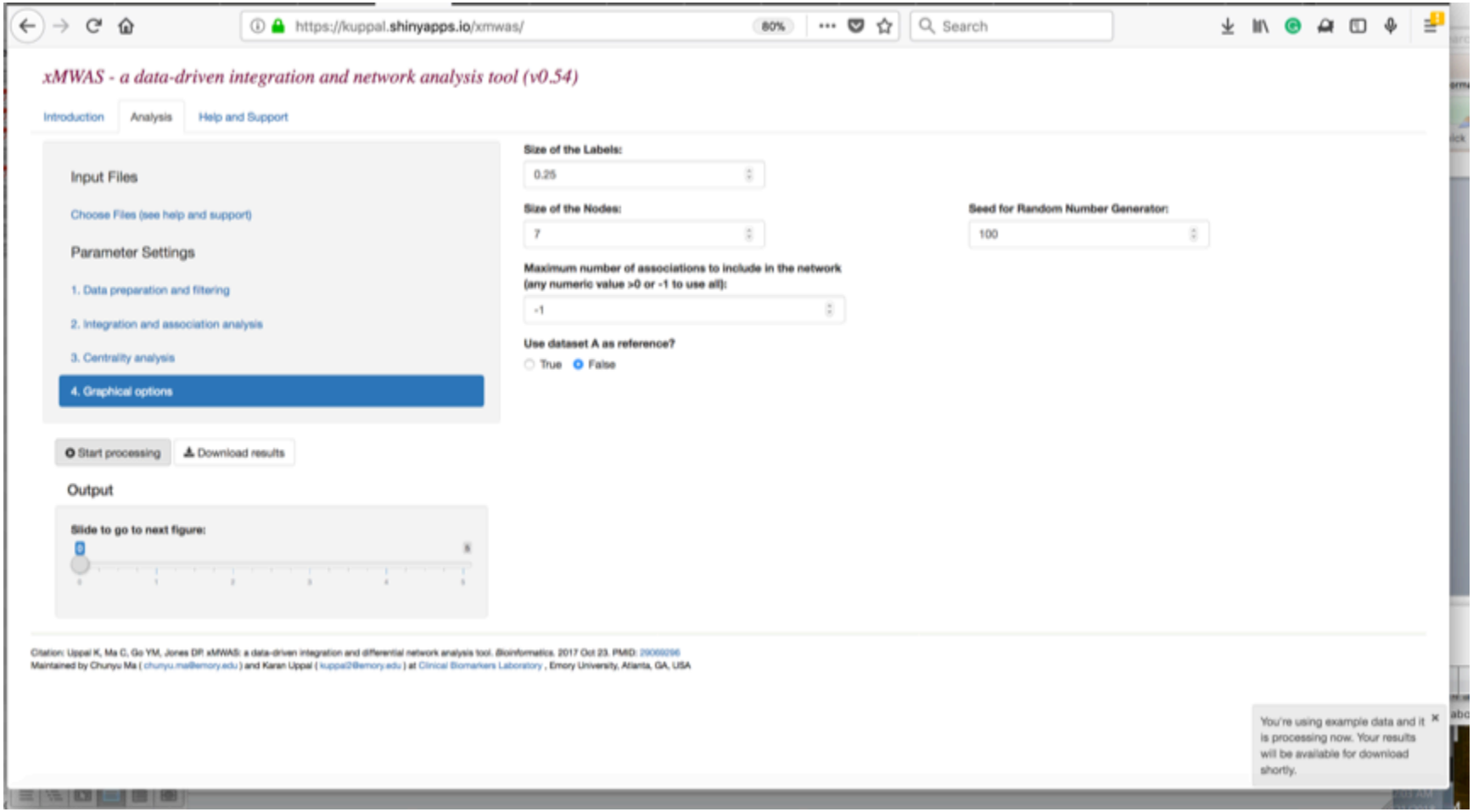
- Eigenvector: based on the number and quality of connections
- Betweenness: based on the extent to which a node lies on
- the path between other nodes
- Degree.count: based on the number of connections
- Degree.weight: based on the magnitude of edges (association
- scores)
- Closeness: based on the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

Step 5. Set graphical options

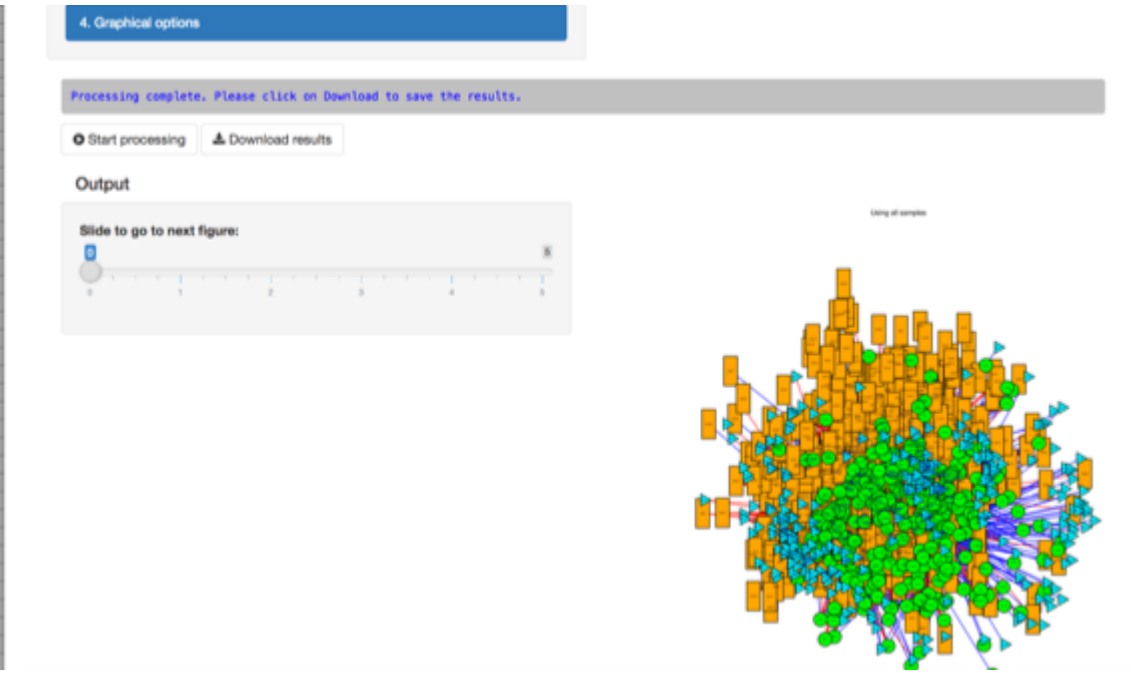


https://github.com/kuppal2/xMWAS/blob/master/example_manual_tutorial/xMWAS-manual.pdf

Step 6. Click on “Start processing”



Results

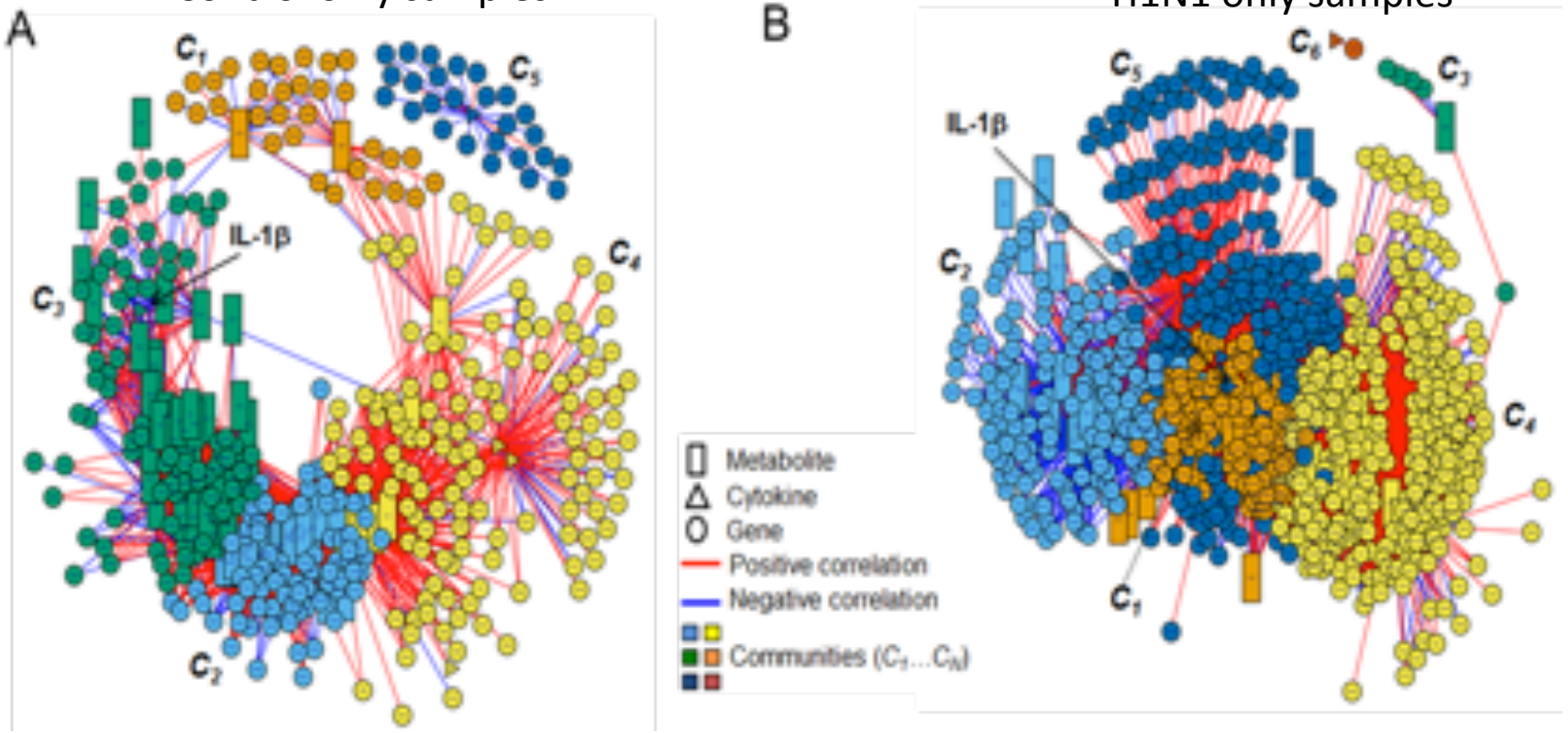

4. Graphical options
Processing complete. Please click on Download to save the results.
Start processing
Download results
Output
Slide to go to next figure:

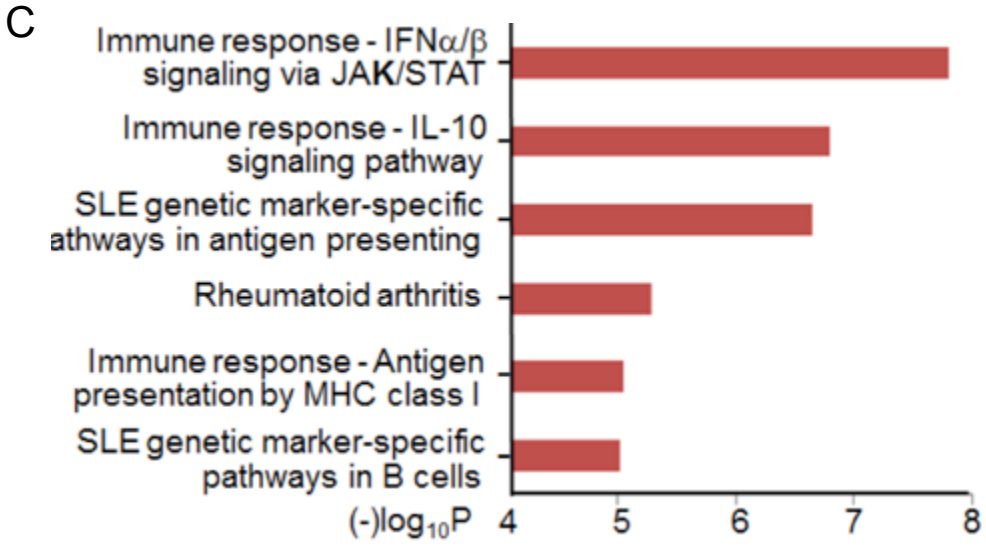
Case Study: Application of xMWAS for integrative –and differential network analysis of more than 2 dataests.

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



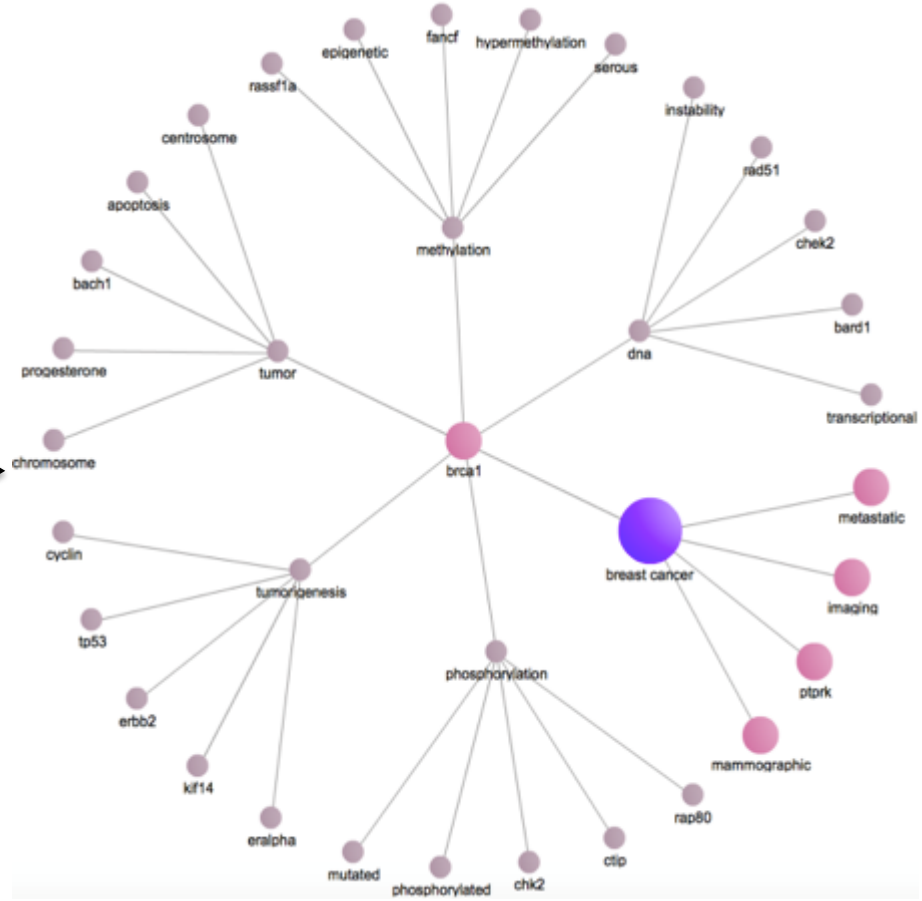


Uppal 2017, Bioinformatics

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**

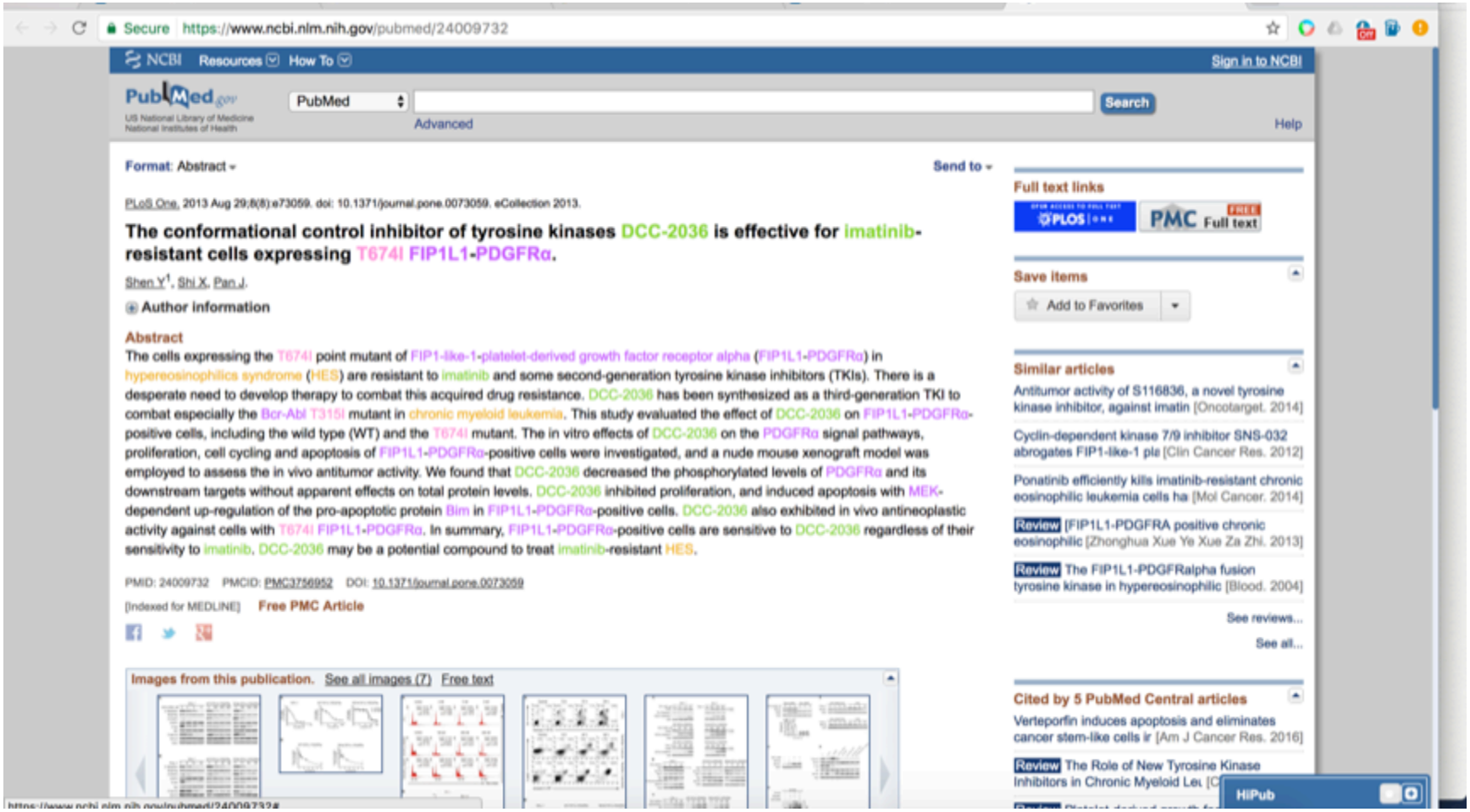
Text mining tools for literature-based relation discovery biomedical text



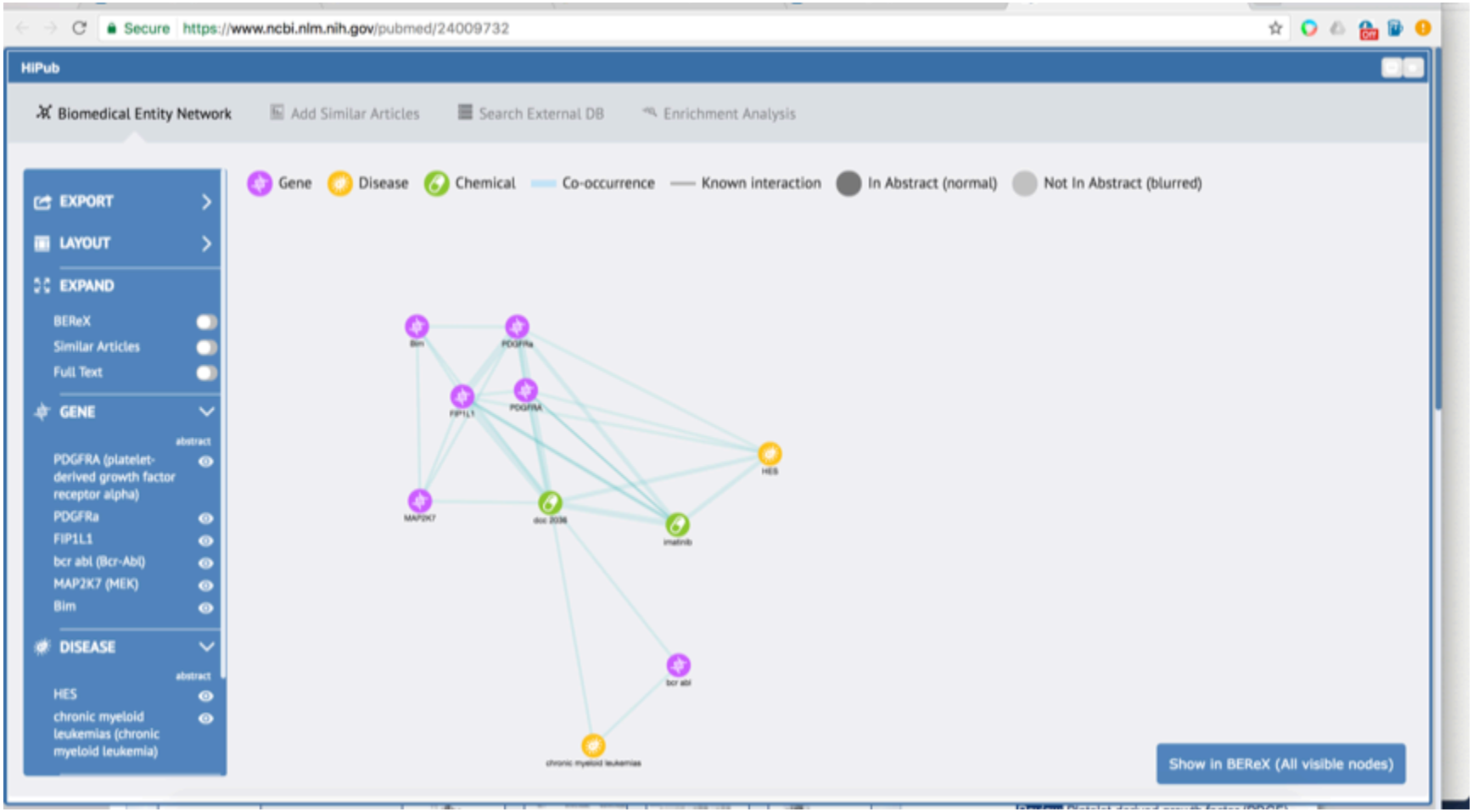
Zhiyong Lu, "PubMed and beyond: a survey of web tools for searching biomedical literature", Database Vol. 2011

Association mining based on co-occurrence

HiPub (Lee 2016): http://hipub.korea.ac.kr/



HiPub (Lee 2016):

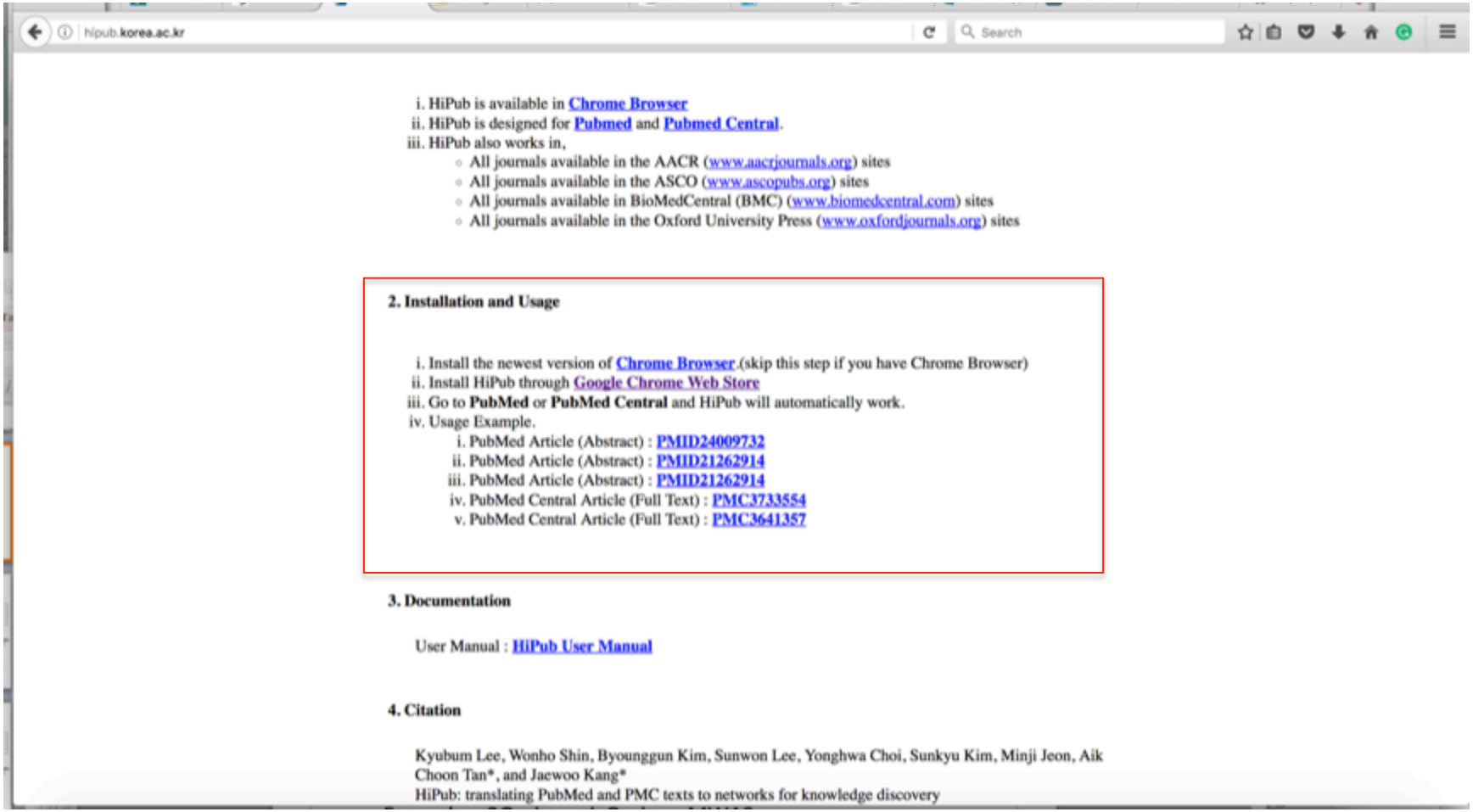


Summary

- 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/



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/jlbmiklekmigmbmcodhjgdpooldjcjam
- 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.gov
US National Library of Medicine
National Institutes of Health

PubMed Search

Advanced 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-PDGFRα.

Shen Y[1], Shi X, Pan J.

Author information

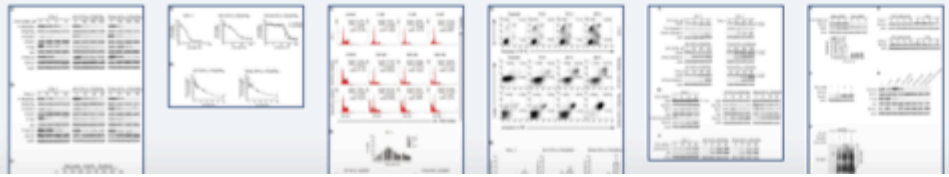
Abstract

The cells expressing the T674I point mutant of FIP1-like-1-platelet-derived growth factor receptor alpha (FIP1L1-PDGFRα) in hypereosinophilics 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-PDGFRα-positive cells, including the wild type (WT) and the T674I mutant. The in vitro effects of DCC-2036 on the PDGFRα signal pathways, proliferation, cell cycling and apoptosis of FIP1L1-PDGFRα-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 PDGFRα 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-PDGFRα-positive cells. DCC-2036 also exhibited in vivo antineoplastic activity against cells with T674I FIP1L1-PDGFRα. In summary, FIP1L1-PDGFRα-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



Full text links

PLOS ONE | 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 pla [Clin Cancer Res. 2012]

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

Review [FIP1L1-PDGFRA positive chronic eosinophilic [Zhonghua Xue Ye Xue Za Zhi. 2013]

Review The FIP1L1-PDGFRalpha fusion tyrosine kinase in hypereosinophilic [Blood. 2004]

See reviews...

See all...

Cited by 5 PubMed Central articles

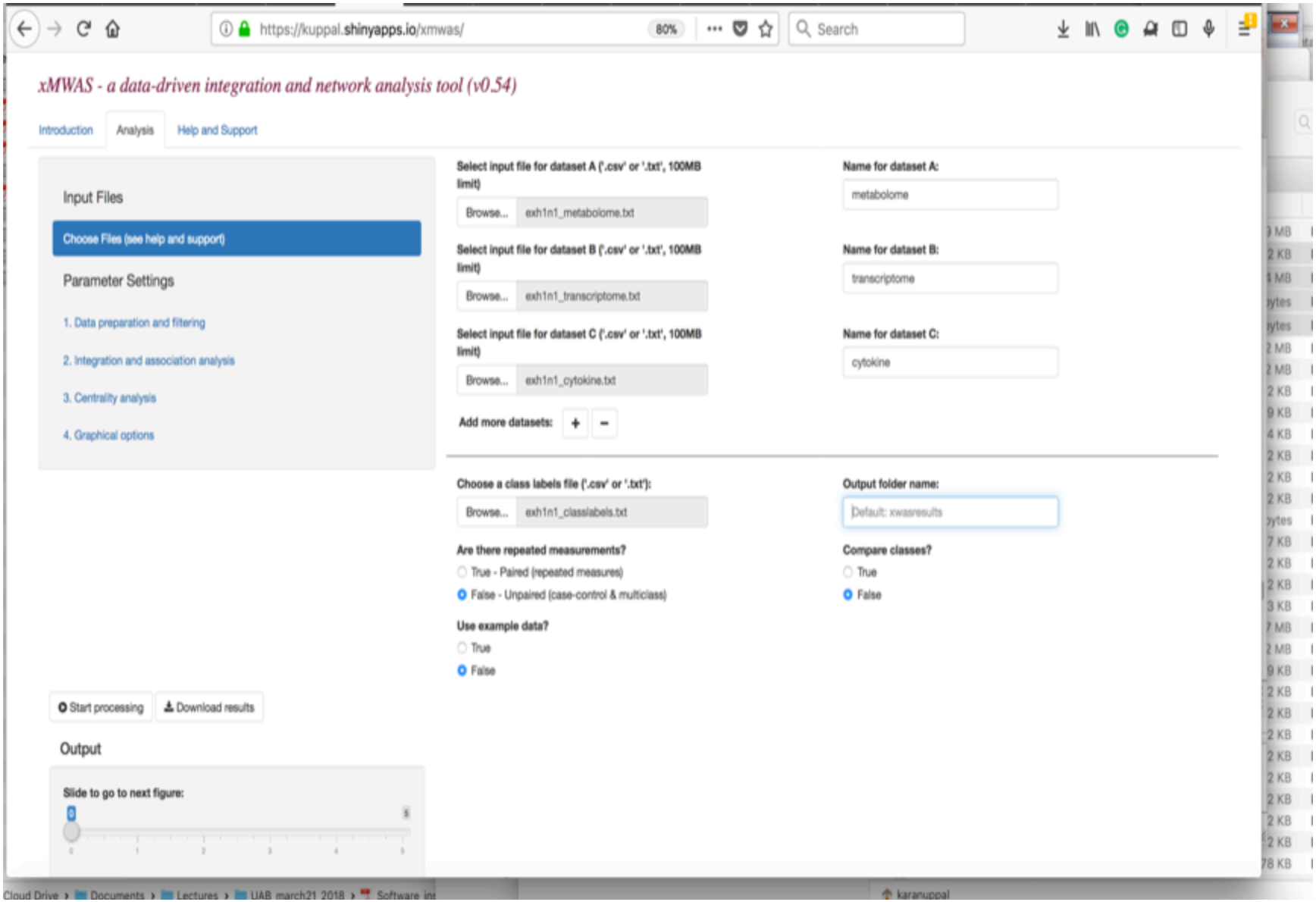
Verteporfin induces apoptosis and eliminates cancer stem-like cells ir [Am J Cancer Res. 2016]

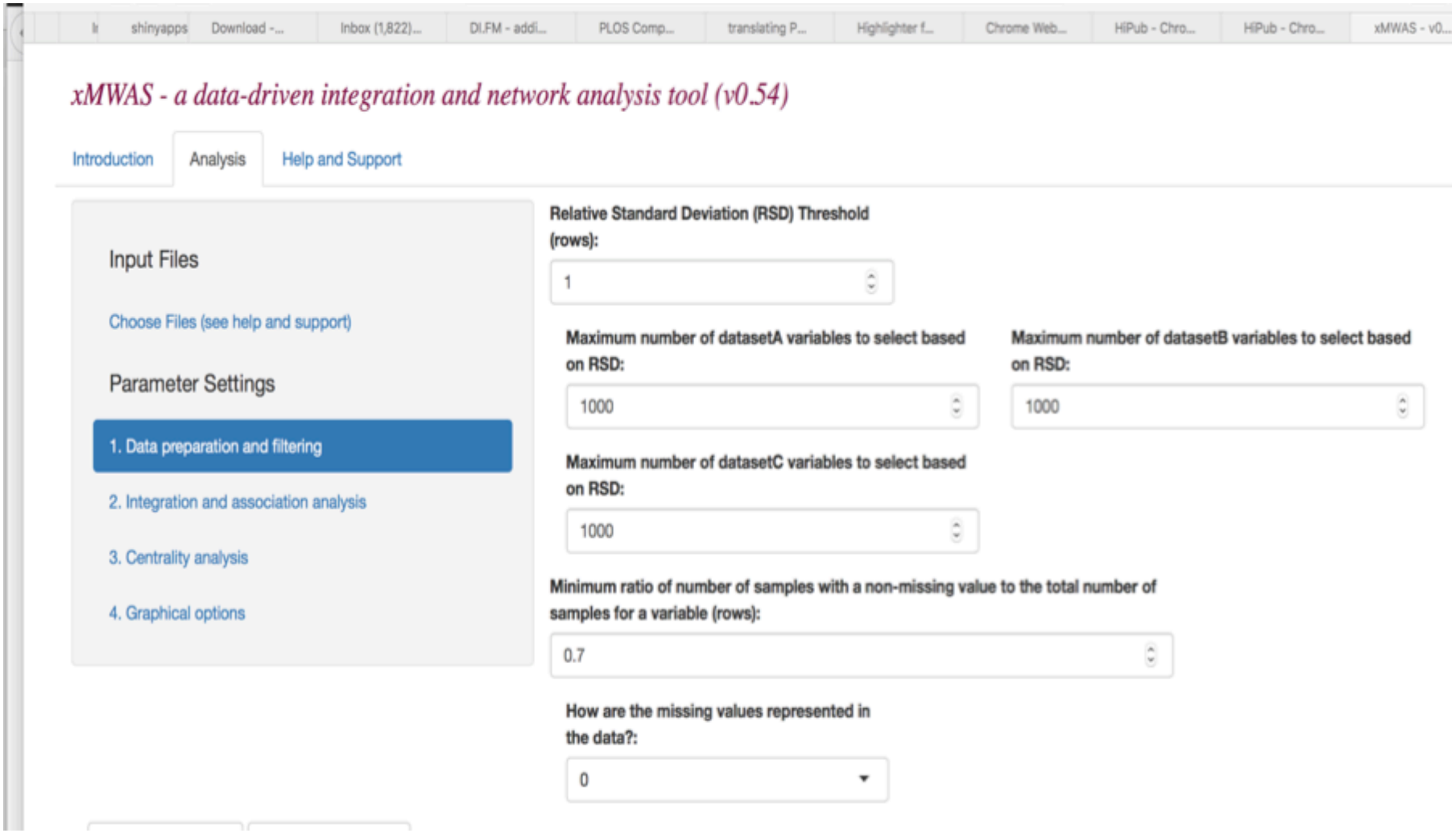
Review The Role of New Tyrosine Kinase Inhibitors in Chronic Myeloid Leu [C

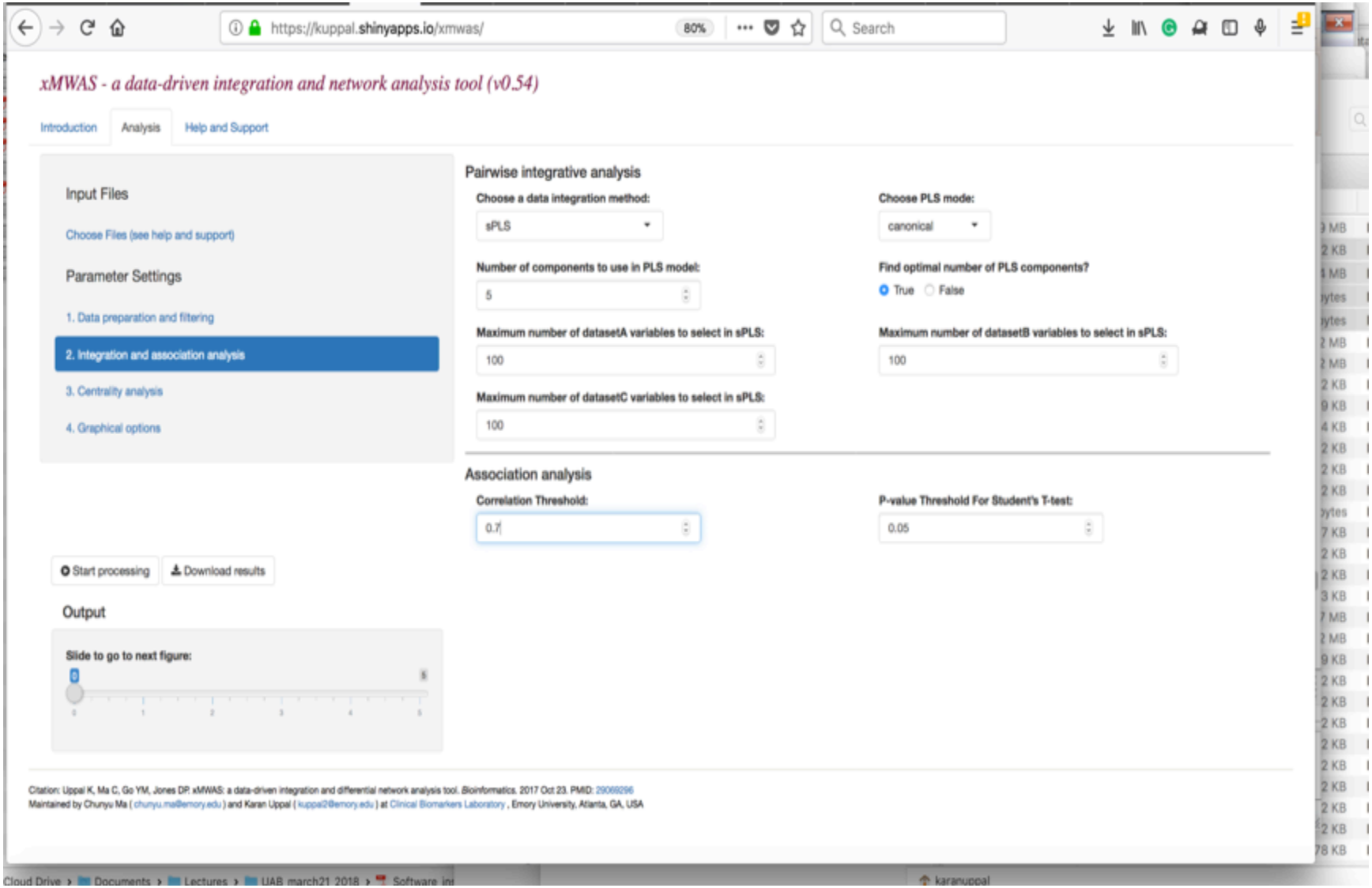
HiPub

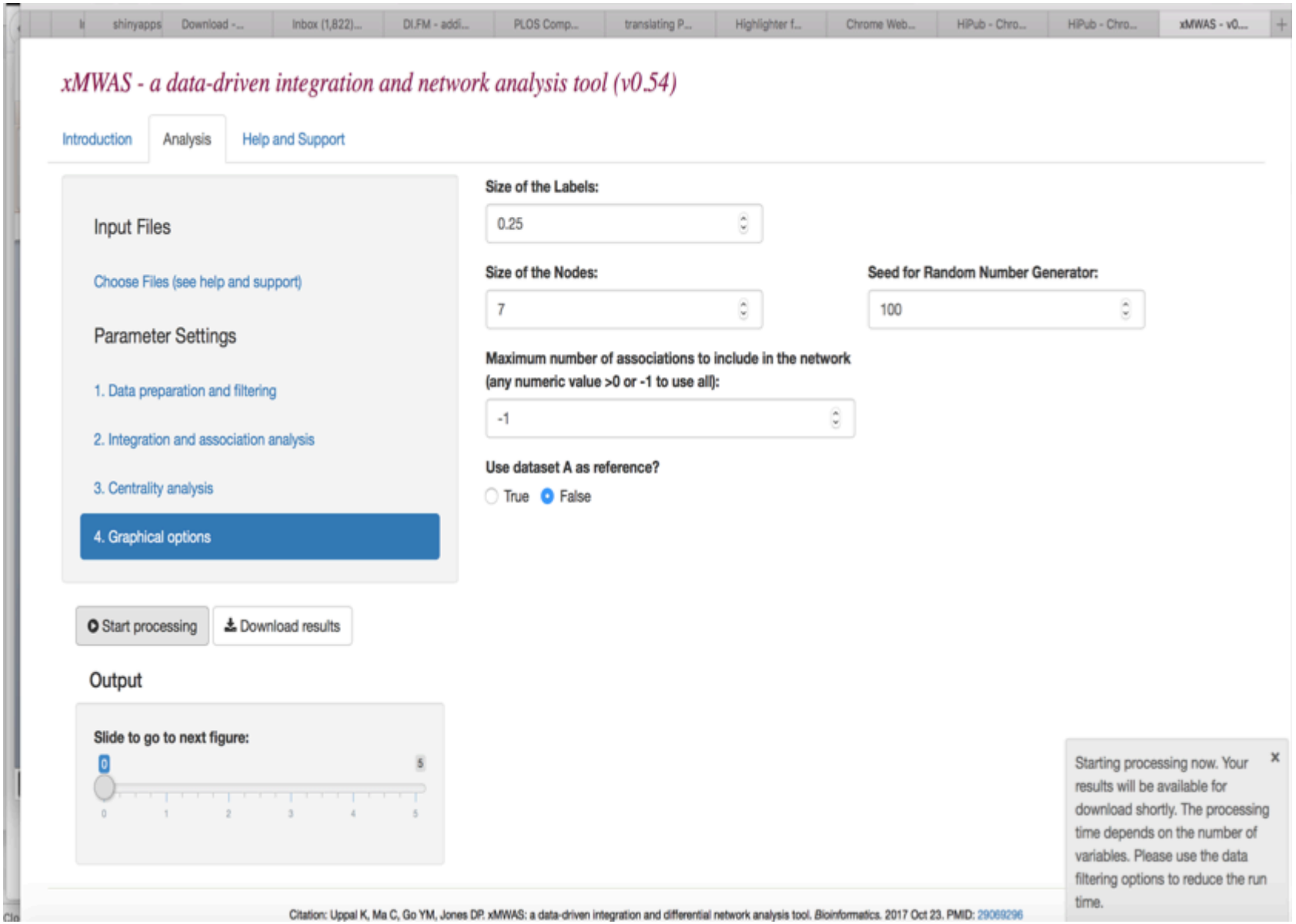
2. xMWAS: Web version

- URL: https://kuppal.shinyapps.io/xmwas/
- Input files: 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
Download results
Output
Slide to go to next figure:


shinyapps
Download -...
Inbox (1,822)...
DI.FM - addi...
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
Relative Standard Deviation (RSD) Threshold (rows):
1
Maximum number of datasetA variables to select based on RSD:
1000
Maximum number of datasetB variables to select based on RSD:
1000
Maximum number of datasetC variables to select based on RSD:
1000
Minimum ratio of number of samples with a non-missing value to the total number of samples for a variable (rows):
0.7
How are the missing values represented in the data?:
0

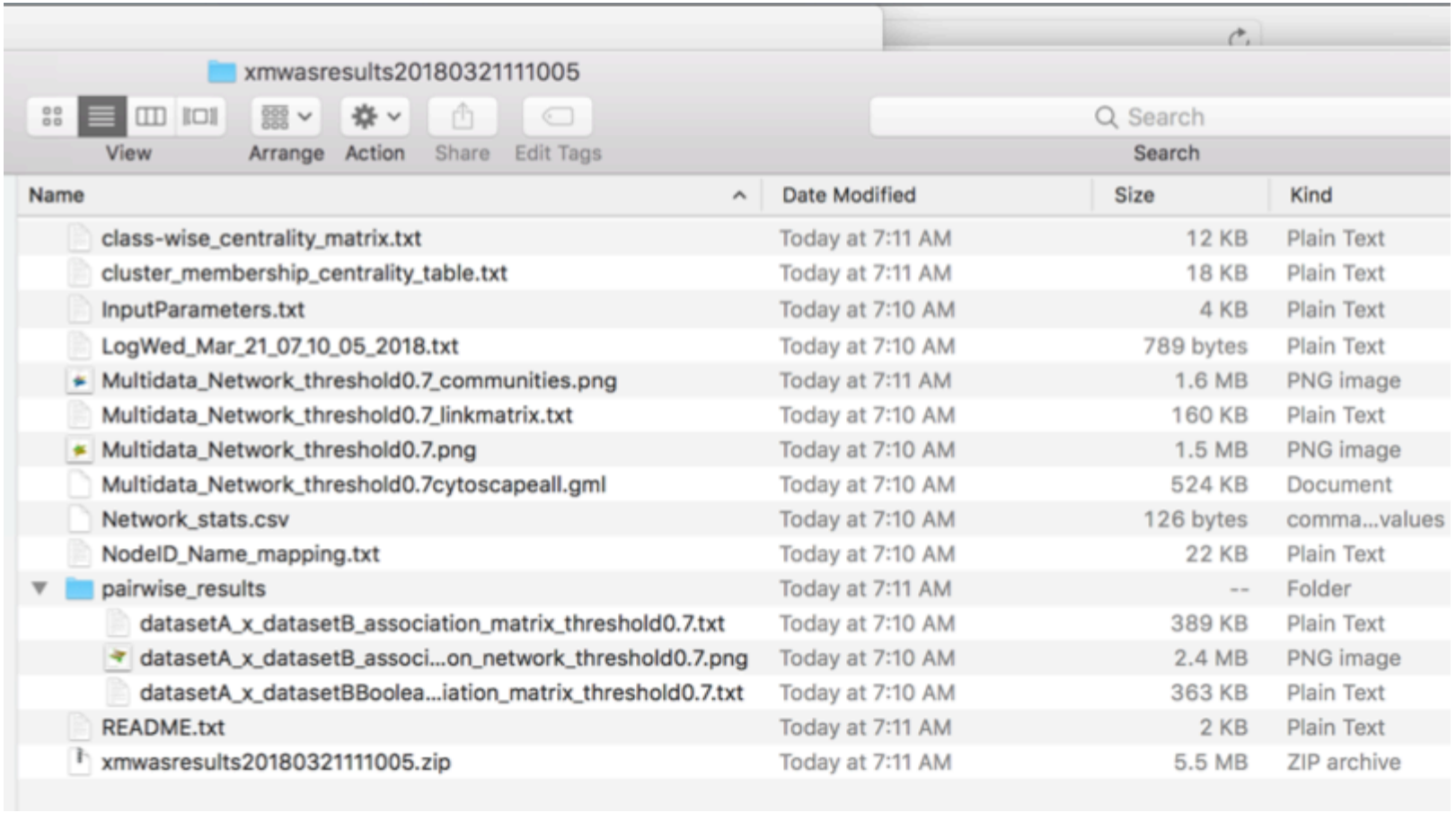

https://kuppal.shinyapps.io/xmwas/
80%
Search
xMWAS - a data-driven integration and network analysis tool (v0.54)
Introduction
Analysis
Help and Support
Input Files
Choose Files (see help and support)
Parameter Settings
1. Data preparation and filtering
2. Integration and association analysis
3. Centrality analysis
4. Graphical options
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
False
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
Download results
Output
Slide to go to next figure:
0
5
0
1
2
3
4
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: 29069296
Maintained by Chunyu Ma (chunyu.ma@emory.edu) and Karan Uppal (kuppal2@emory.edu) at Clinical Biomarkers Laboratory , Emory University, Atlanta, GA, USA


shinyapps
Download -...
Inbox (1,822)...
DI.FM - addi...
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:
0
5
0
1
2
3
4
5
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


0.7
0.05
Processing complete. Please click on Download to save the results.
Start processing
Download results
Output
Slide to go to next figure:
0
3
0
1
2
3
(Edges) Red: +ve correlation; Blue: -ve correlation
(Nodes) Rectangle: datasetA; Circle: datasetB
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.

Download results



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_associ...on_network_threshold0.7.png	Today at 7:10 AM	2.4 MB	PNG image
datasetA_x_datasetBBoolea...iation_matrix_threshold0.7.txt	Today at 7:10 AM	363 KB	Plain Text
README.txt	Today at 7:11 AM	2 KB	Plain Text
xmwasresults20180321111005.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)

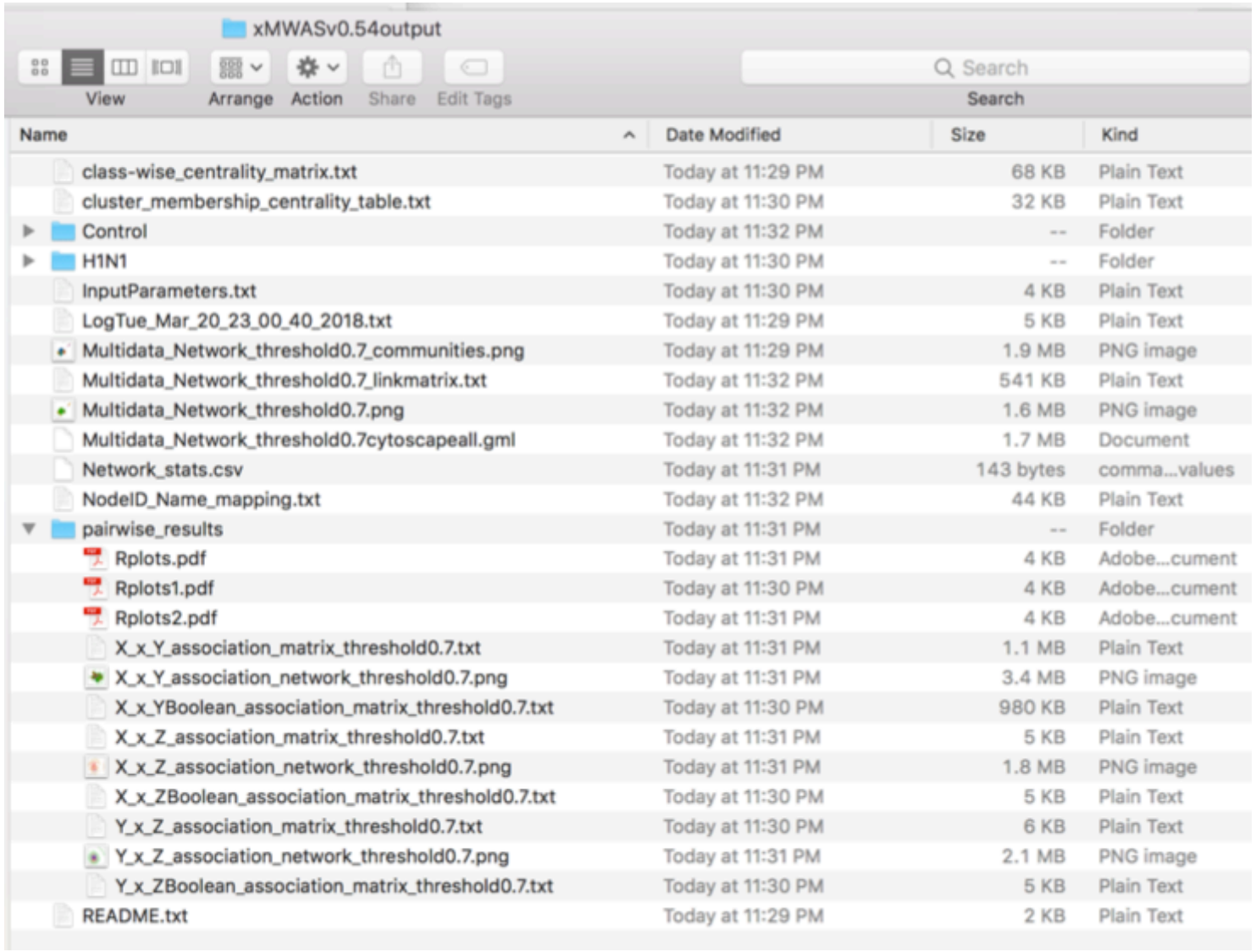
```
#load package
library(xMWAS)

#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$metabolome
yMat<-exh1n1$transcrlptome
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



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...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

Input Parameters

InputParameters.txt

```
[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
```

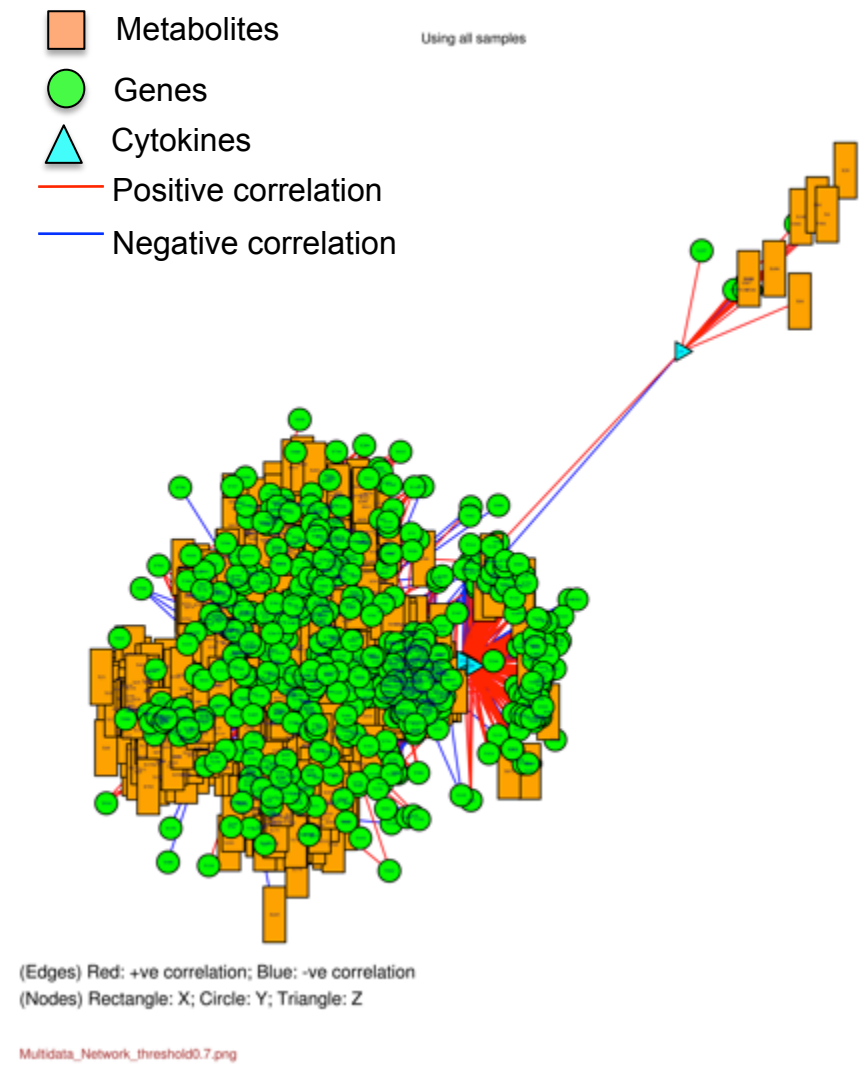
Description of output: Readme.txt

README.txt

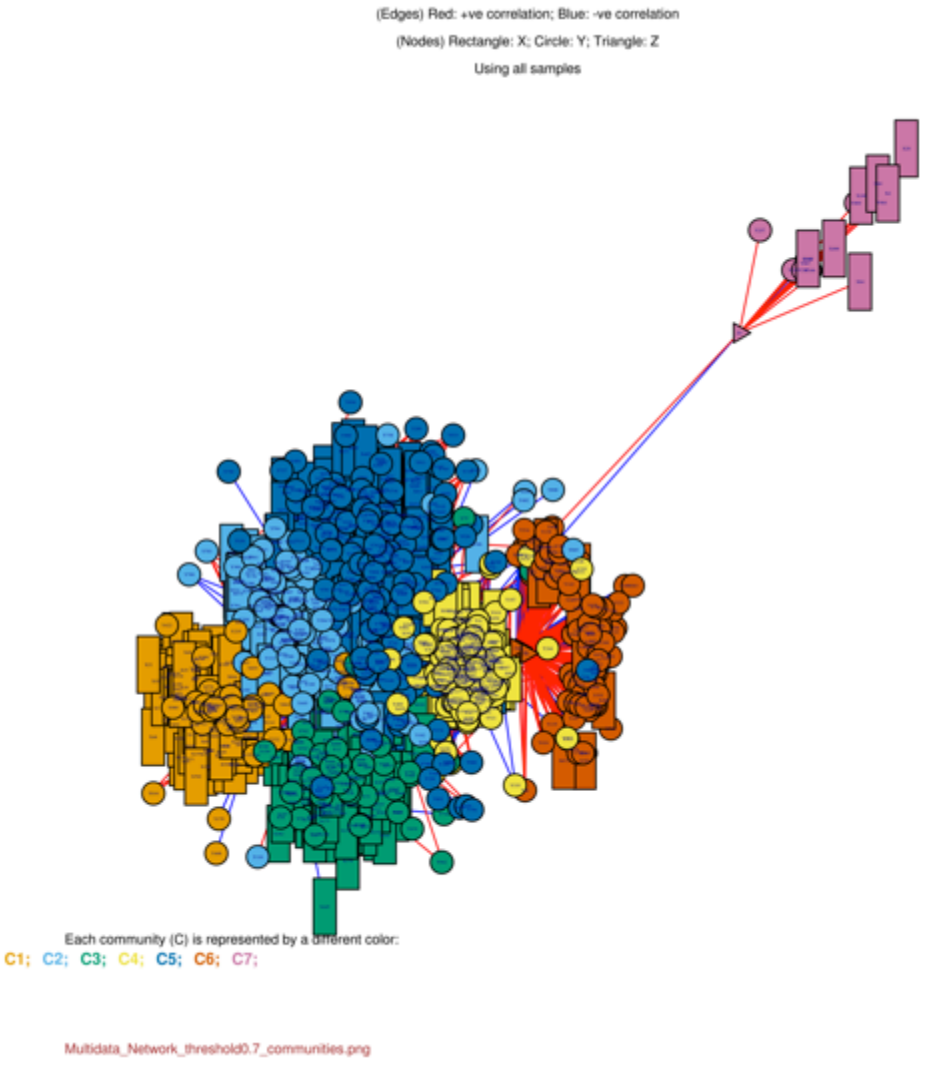
"Description of files"
"1: X labels correspond to xome_fname data, Y labels correspond to yome_fname data, Z labels correspond to zome_fname data, W labels correspond to wome_fname data"
"2: Pairwise integrative analysis results are under pairwise_results. The files corresponding to each pairwise comparison (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 bu 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 detectic 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 ar written in individual subfolders."

Network graphs

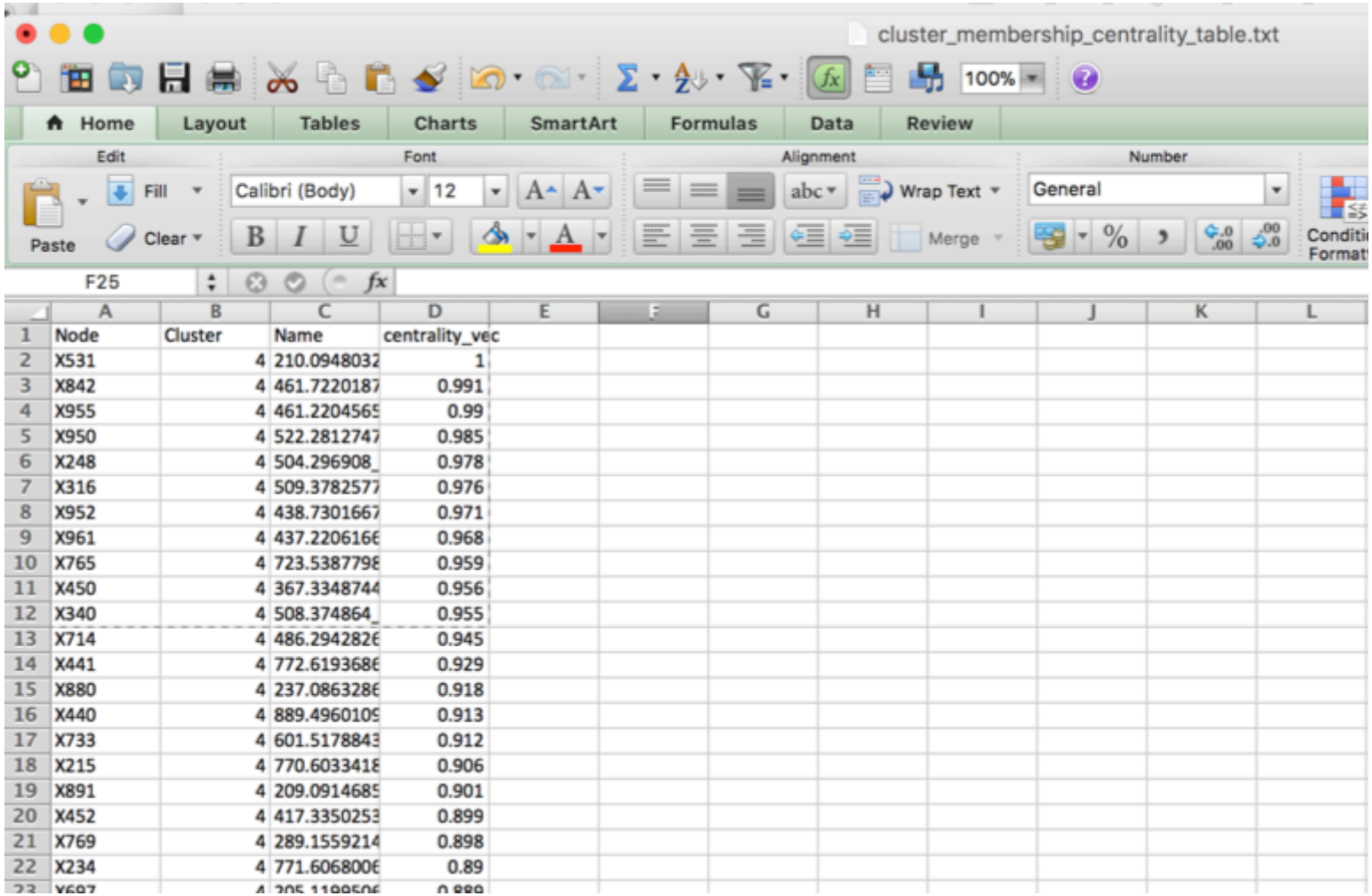
A. Colored by data type



B. Colored by community membership



Community detection and centrality analysis



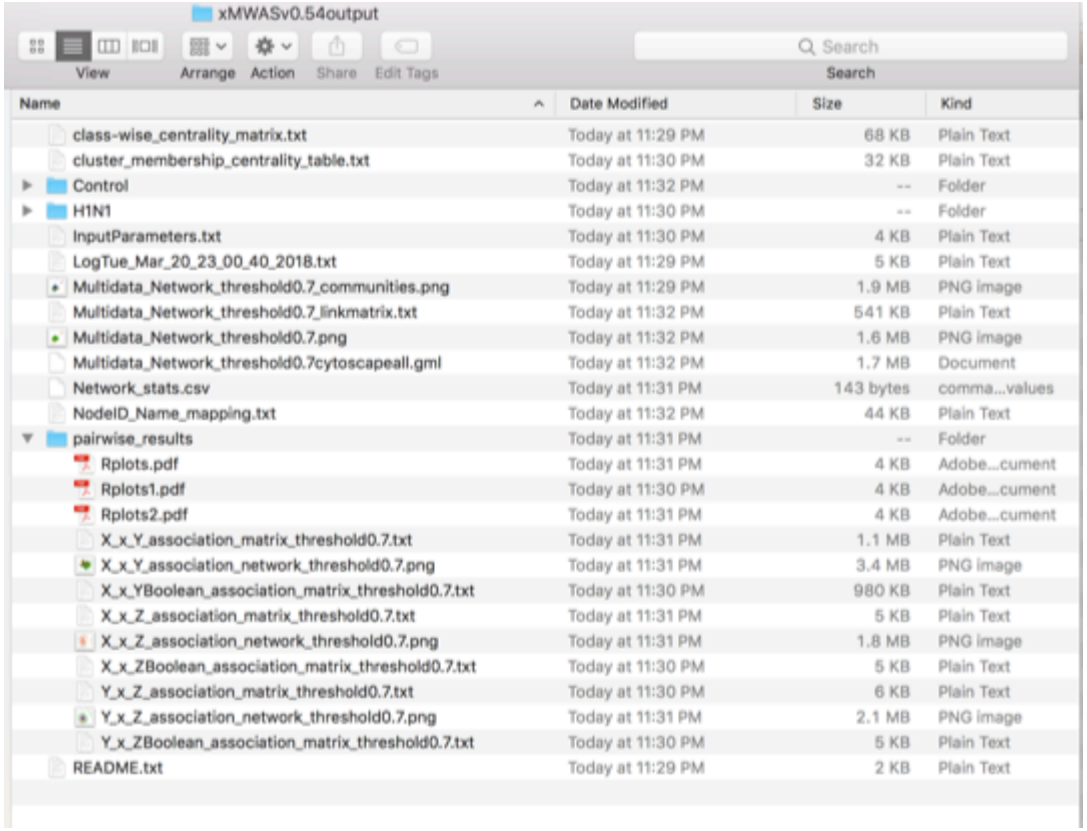
cluster_membership_centrality_table.txt

Home | Layout | Tables | Charts | SmartArt | Formulas | Data | Review

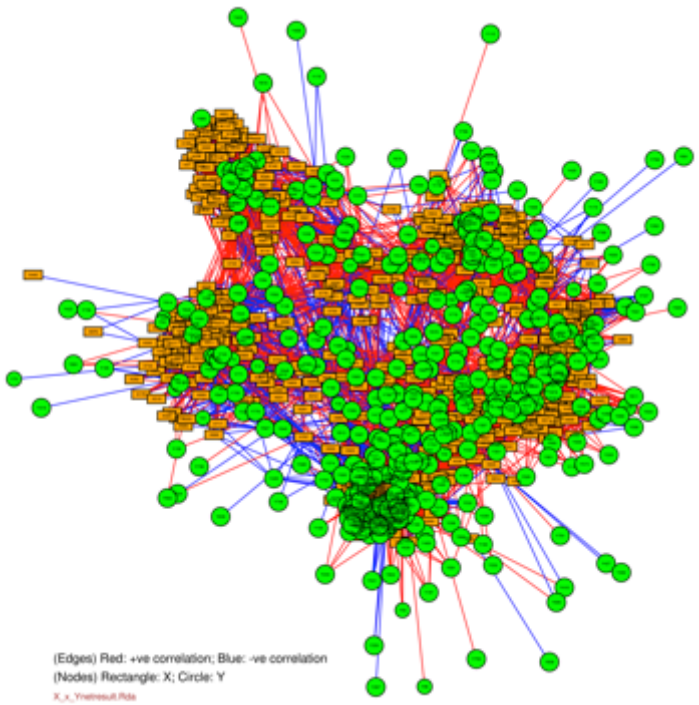
F25

	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								

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

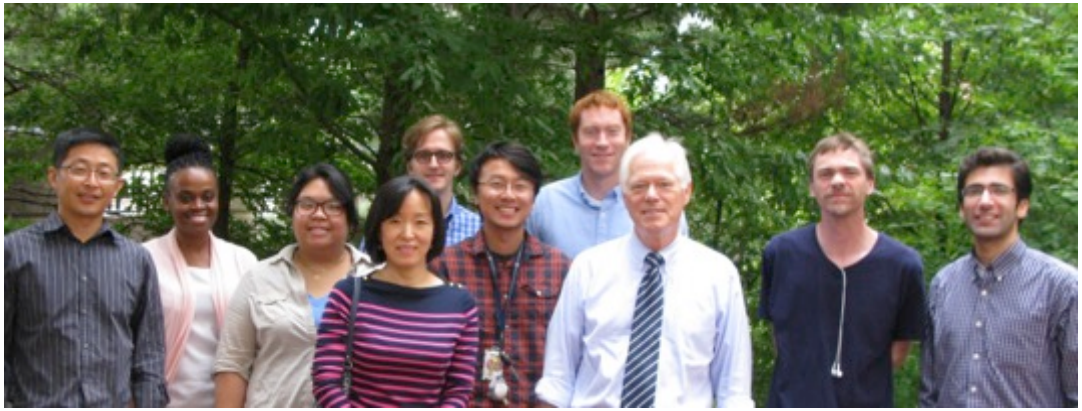


Name	Date Modified	Size	Kind
class-wise_centrality_matrix.txt	Today at 11:29 PM	68 KB	Plain Text
cluster_membership_centrality_table.txt	Today at 11:30 PM	32 KB	Plain Text
Control	Today at 11:32 PM	--	Folder
H1N1	Today at 11:30 PM	--	Folder
InputParameters.txt	Today at 11:30 PM	4 KB	Plain Text
LogTue_Mar_20_23_00_40_2018.txt	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...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

Email: kuppal2@emory.edu

Questions?