

Data Integration: hands-on session

July 21, 2017

Karan Uppal, PhD

Assistant Professor of Medicine

Director of Computational Metabolomics & Integrative Omics,
Clinical Biomarkers Laboratory

Emory University

1. MetaboAnalyst: Integrated Pathway Analysis

URL: <http://www.metaboanalyst.ca/>



The screenshot shows the MetaboAnalyst 3.0 website. The header features the title "MetaboAnalyst 3.0" and the subtitle "a comprehensive tool suite for metabolomic data analysis". A navigation menu on the left includes links for Home, Overview, Data Formats, FAQs, Tutorials, Troubleshooting, Resources, Update History, User Stats, and About. The main content area displays a "Welcome" message with a link to "click here to start". Below this is a "News & Updates" section listing various updates and fixes, such as support for correlation analysis, heatmap generation, and microbiome data analysis. At the bottom, there is a "Please Cite:" section with references to publications by Xia, J. and Wishart, D.S. The website also features logos for Genome Canada and Genome Québec.

MetaboAnalyst 3.0
– a comprehensive tool suite for metabolomic data analysis

Home
Overview
Data Formats
FAQs
Tutorials
Troubleshooting
Resources
Update History
User Stats
About

GenomeCanada
GenomeQuébec

Welcome >> [click here to start](#) <<

News & Updates

- More than **1.6 million jobs** have been processed over the last 12 months (06/08/2017);
- Added support for correlation analysis on either samples or features (06/07/2017); **NEW**
- Added support for generating heatmap on group averages (06/05/2017); **NEW**
- Check out our [MicrobiomeAnalyst](#) tool for comprehensive analysis of microbiome data (04/28/2017); **NEW**
- Fixed issues with PDF report generation (03/15/2017); **NEW**
- Added support for peak filtering based on QC samples for untargeted metabolomics (03/10/2017); **NEW**
- Added support for "flipping" PCA for cross-study comparison (02/09/2017);
- Added support for network summary of enrichment analysis result (02/06/2017);
- Fixed the bug in feature table display in Biomarker Tester module (01/05/2017);
- Updated the pathway result table to show all/matched compounds (11/25/2016);
- Enhanced Normalization and Data Editor for better user experience (11/15/2016);
- Added support for sparse PLS-DA (sPLS-DA) analysis (10/28/2016);
- Added support for quantile normalization (08/29/2016);
- Improved name mapping functions for common metabolite names (08/18/2016);

[Read more](#)

Please Cite:

Xia, J. and Wishart, D.S. (2016) [Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis](#) Current Protocols in Bioinformatics, 55:14.10.1-14.10.91.

Xia, J., Sineelnikov, I., Han, B., and Wishart, D.S. (2015) [MetaboAnalyst 3.0 - making metabolomics more meaningful](#). Nucl. Acids Res. 43, W251-257.

Xia, J., Mandal, R., Sineelnikov, I., Broadhurst, D., and Wishart, D.S. (2012) [MetaboAnalyst 2.0 - a comprehensive server for metabolomic](#)

Click on “Click here to start”

www.metaboanalyst.ca

MetaboAnalyst 3.0

- a comprehensive tool suite for metabolomic data analysis

[Home](#)
[Overview](#)
[Data Formats](#)
[FAQs](#)
[Tutorials](#)
[Troubleshooting](#)
[Resources](#)
[Update History](#)
[User Stats](#)
[About](#)

GenomeCanada
GenomeQuébec

Welcome [>> click here to start <<](#)

News & Updates

- More than [1.6 million jobs](#) have been processed over the last 12 months (06/08/2017);
- Added support for correlation analysis on either samples or features (06/07/2017); **NEW**
- Added support for generating heatmap on group averages (06/05/2017); **NEW**
- Check out our [MicrobiomeAnalyst](#) tool for comprehensive analysis of microbiome data (04/28/2017); **NEW**
- Fixed issues with PDF report generation (03/15/2017); **NEW**
- Added support for peak filtering based on QC samples for untargeted metabolomics (03/10/2017); **NEW**
- Added support for “flipping” PCA for cross-study comparison (02/09/2017);
- Added support for network summary of enrichment analysis result (02/06/2017);
- Fixed the bug in feature table display in Biomarker Tester module (01/05/2017);
- Updated the pathway result table to show all/matched compounds (11/25/2016);
- Enhanced Normalization and Data Editor for better user experience (11/15/2016);
- Added support for sparse PLS-DA (sPLS-DA) analysis (10/28/2016);
- Added support for quantile normalization (08/29/2016);
- Improved name mapping functions for common metabolite names (08/18/2016);

[Read more.....](#)

Please Cite:

Xia, J. and Wishart, D.S. (2016) [Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis](#) Current Protocols in Bioinformatics, 55:14.10.1-14.10.91.

Xia, J., Sinelnikov, I., Han, B., and Wishart, D.S. (2015) [MetaboAnalyst 3.0 - making metabolomics more meaningful](#). Nucl. Acids Res. 43, W251-257.

Xia, J., Mandal, R., Sinelnikov, I., Broadhurst, D., and Wishart, D.S. (2012) [MetaboAnalyst 2.0 - a comprehensive server for metabolomic](#)

Select the “Integrated Pathway Analysis” module

The screenshot shows the MetaboAnalyst 3.0 web interface. The browser address bar displays www.metaboanalyst.ca/faces/ModuleView.xhtml. The page features a sidebar on the left with navigation links: Home, Overview, Data Formats, FAQs, Tutorials, Troubleshooting, Resources, Update History, User Stats, and About. Below these links are logos for GenomeCanada, GenomeQuebec, NSERC CRNG, and TMIC. The main content area is titled "Please choose a functional module to proceed:" and contains eight modules, each with a brief description. The "Integrated Pathway Analysis" module is highlighted with a red rectangular box. The footer of the page reads "Xia Luo & McGill (last updated 2017-06-20)".

MetaboAnalyst 3.0
— a comprehensive tool suite for metabolomic data analysis

Please choose a functional module to proceed:

- Statistical Analysis**
This module offers various commonly used statistical and machine learning methods including t-tests, ANOVA, PCA, PLS-DA and Orthogonal PLS-DA. It also provides clustering and visualization tools to create dendrograms and heatmaps as well as to classify based on random forests and SVM.
- Enrichment Analysis**
This module performs metabolite set enrichment analysis (MSEA) for human and mammalian species based on several libraries containing ~6300 groups of metabolite sets. Users can upload either 1) a list of compounds, 2) a list of compounds with concentrations, or 3) a concentration table.
- Pathway Analysis**
This module supports pathway analysis (integrating enrichment analysis and pathway topology analysis) and visualization for 21 model organisms, including Human, Mouse, Rat, Cow, Chicken, Zebrafish, Arabidopsis thaliana, Rice, Drosophila, Malaria, S. cerevisiae, E.coli. and others, with a total of ~1600 metabolic pathways.
- Time-series/Two-factor Design**
This module supports temporal and two-factor data analysis including data overview, two-way ANOVA, and empirical Bayes time-series analysis for detecting distinctive temporal profiles. It also supports ANOVA-simultaneous component analysis (ASCA) to identify major patterns associated with each experimental factor.
- Power Analysis**
This module uses pilot data to calculate the minimum number of samples required to detect a statistically significant difference between two populations with a given degree of confidence (called Power Analysis).
- Biomarker Analysis**
This module performs various ROC curve based biomarker analyses for a single or multiple biomarkers. It also allows users to manually specify biomarker models as well as new sample prediction.
- Integrated Pathway Analysis**
This module performs integrated metabolic pathway analysis on results obtained from combined metabolomics and gene expression studies conducted under the same experimental conditions.
- Other Utilities**
This module contains several common utility functions. At this moment, compound ID conversion, batch effect correction and lipidomics data analysis are available.

GenomeCanada
GenomeQuebec
NSERC CRNG
TMIC

Xia Luo & McGill (last updated 2017-06-20)

Data upload – select the “Use our example data” option

The interface features a sidebar on the left with a home icon and a menu containing 'Upload' (highlighted), 'Integrative Analysis', 'Download', and 'Exit'. The main area is divided into two panels: 'Gene List' and 'Metabolite List'. Each panel contains a text area with example data, an 'ID Type' dropdown menu, and a 'Specify organism' dropdown menu at the bottom. A red box highlights the 'Use our example data' checkbox, and a 'Submit' button is located at the bottom center.

Gene List

Gene list with optional fold changes

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

ID Type: Entrez ID

Metabolite List

Compound list with optional fold changes

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

ID Type: KEGG ID

Specify organism: Homo sapiens (human)

☒ Use our example data

Submit

Enrichment Analysis

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

- ☒ Hypergeometric Test
- ☐ Fisher's Exact Test

Topology Analysis

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

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

Pathway Databases

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

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

→ Submit

Set Parameters



Upload

Integrative Analysis

ID map

Set parameter

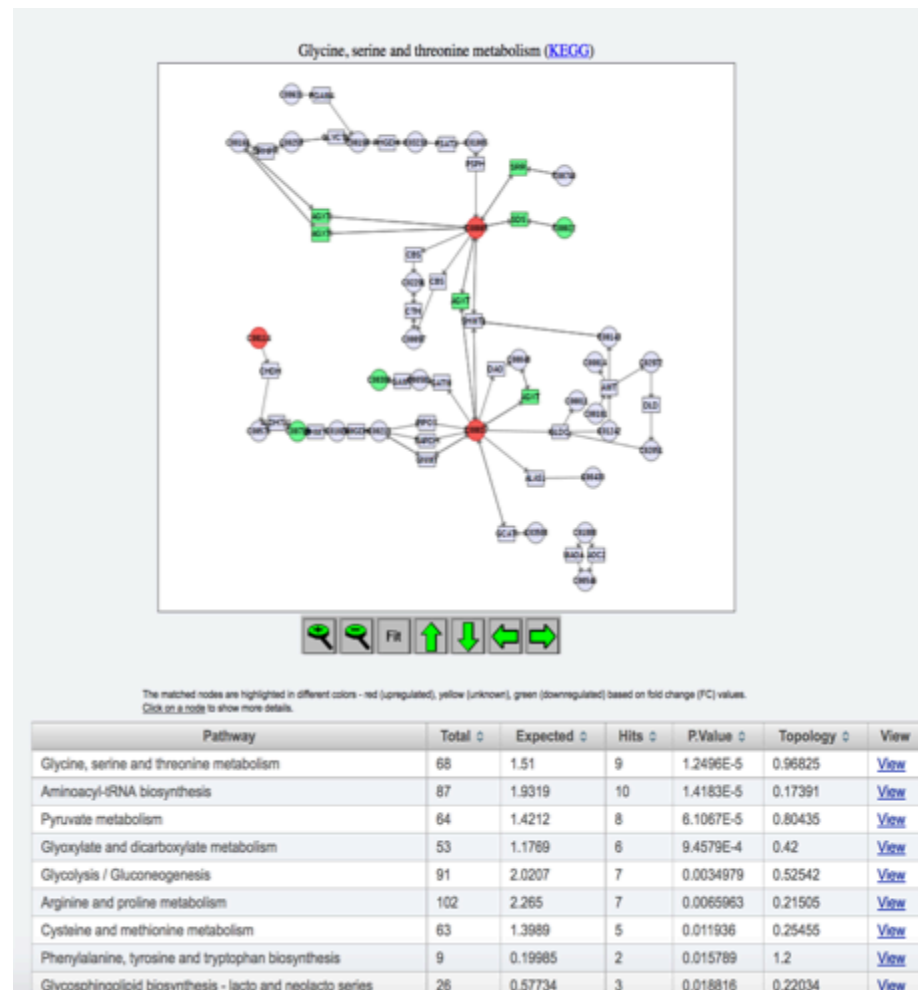
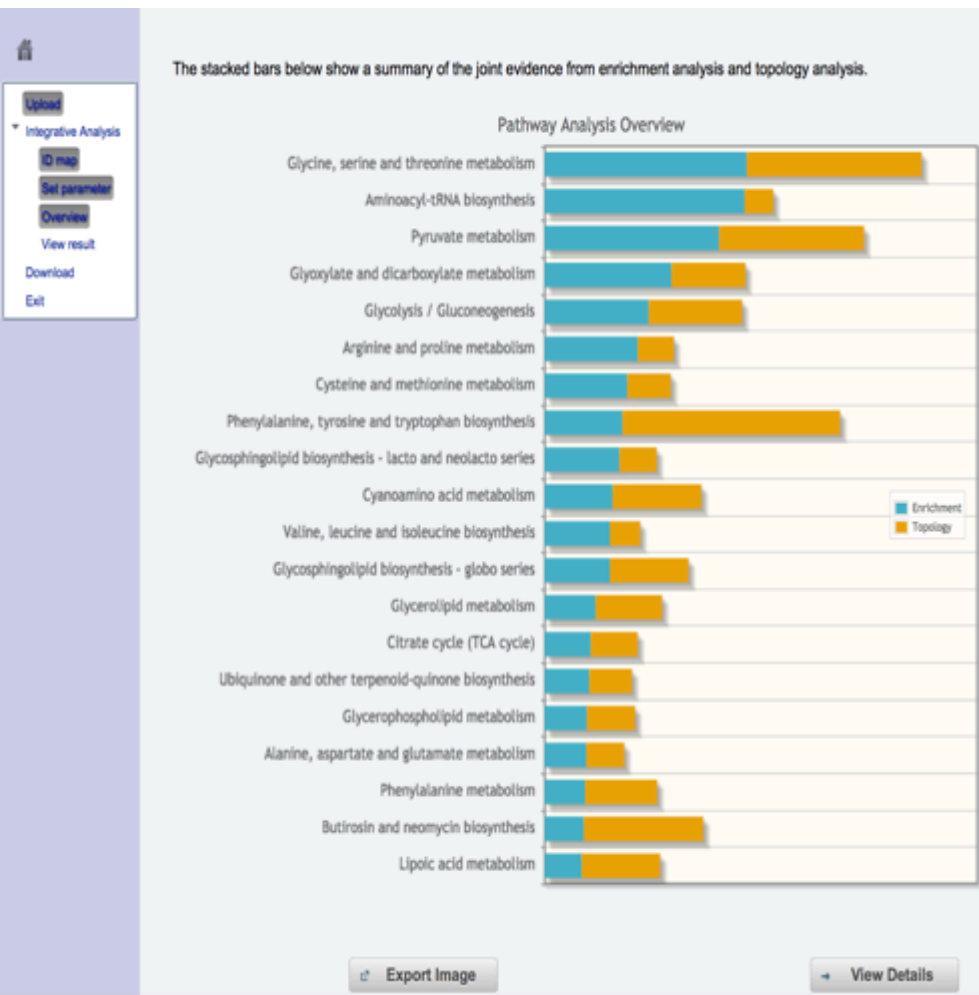
Overview

View result

Download

Exit

Results



2. 3Omics - homepage



The screenshot shows the 3Omics homepage in a web browser. The browser's address bar displays "3omics.cmdm.tw". The page features a navigation menu on the left with links for "Project Features", "Overview", "Name-ID Converter", "Help", and "Contact Us". Below the menu are logos for "CMDML" and "NATIONAL TAIWAN UNIVERSITY". The main content area is titled "Overview" and describes 3Omics as a web-based systems biology visualization tool for integrating human transcriptomic, proteomic, and metabolomic data. It lists five commonly used analyses: correlation network, co-expression, phenotype generation, KEGG/HumanCyc pathway enrichment, and GO enrichment. A section titled "Please select the desired analysis:" offers three options: "1. Transcriptomics-Proteomics-Metabolomics", "2. Transcriptomics-Proteomics", and "3. Transcriptomics only". Below this, a legend identifies the symbols used in the diagrams: green squares for Transcripts (T), red triangles for Proteins (P), and blue circles for Metabolites (M). The legend also defines the line types: a solid line for "Correlation" and a dashed line for "Literature-derived relationship". Five diagrams (a) through (e) illustrate different integration scenarios: (a) shows a linear path T-P-M with solid lines; (b) shows T-P-M with solid lines and a dashed line from T to P; (c) shows T-P-M with solid lines and a dashed line from P to M; (d) shows T-P-M with solid lines and a dashed line from T to M; (e) shows T-P-M with solid lines and a dashed line from T to P and a dashed line from P to M. The footer contains copyright information: "Copyright (c) 2006-2012 Computational Molecular Design & Metabolomics Laboratory | BEB | NTU | Contact us" and a recommendation to use the latest version of Google Chrome or Mozilla Firefox.

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. [?](#)

Example: 3omics.cmdm.tw/example/T_example.csv

Example: 3omics.cmdm.tw/example/P_example.csv

Example: 3omics.cmdm.tw/example/M_example_pubchem.csv

Data format

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

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

Select checkbox next to “Use example data”

3omics.cmdm.tw

MDD LAB
NATIONAL TAIWAN UNIVERSITY

KEGG/HumanCyc pathway enrichment, and GO enrichment.

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.

 (a)

[← Back](#)

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

☒ Use example data [?](#)

Transcriptomics

Use example data as input.
Please un-check "Use example data" above for uploading your data.
GenBank ID: e.g. [NAT1](#), [ABL1](#)

Proteomics

Use example data as input.
Please un-check "Use example data" above for uploading your data.
Uniprot Accession: e.g. [P31946](#), [P62258](#)

Metabolomics

Use example data as input.
Please un-check "Use example data" above for uploading your data.

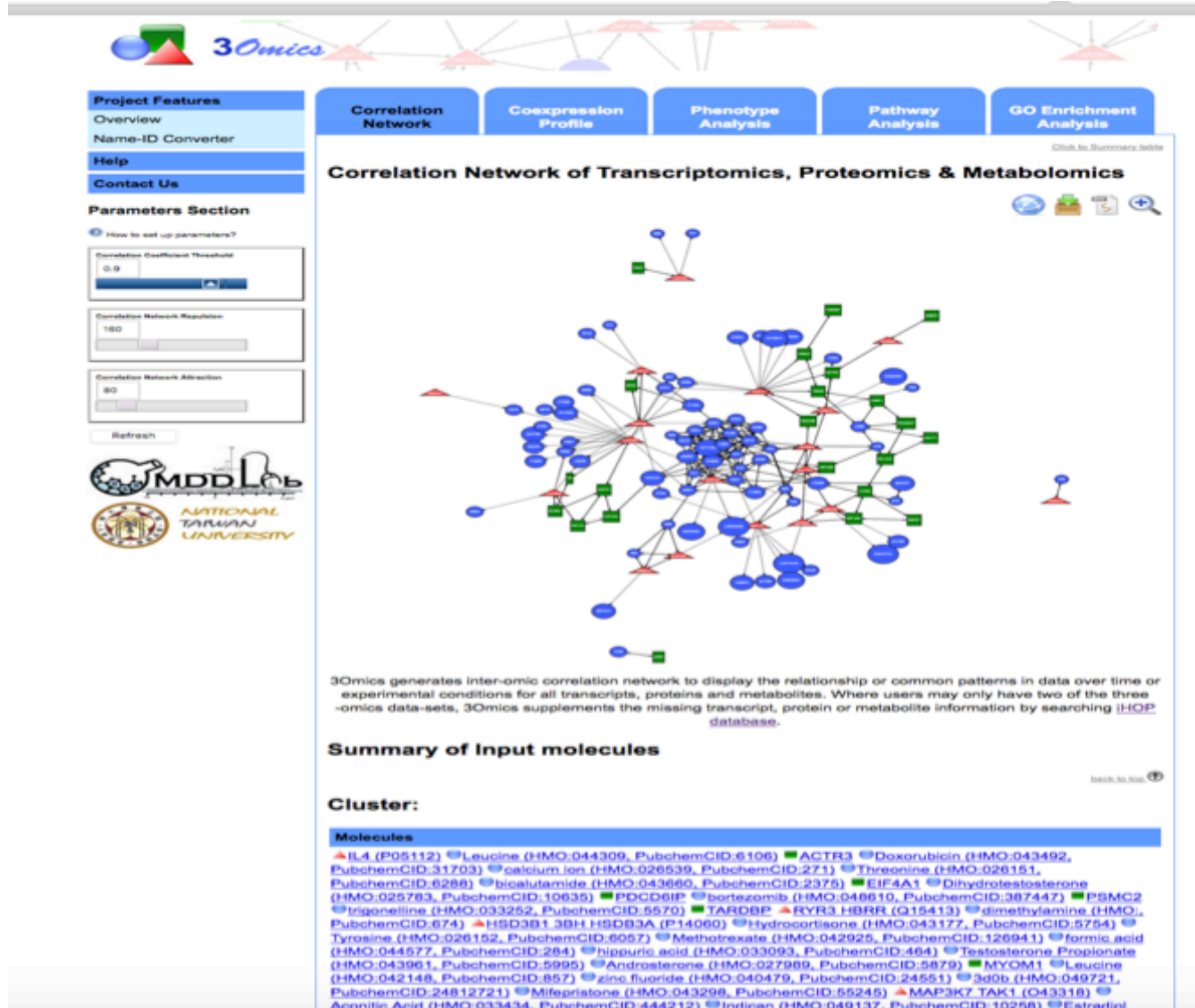
☐ Enrichment Test
Metabolite PubChem CID: e.g. [6971019](#), [5754](#)

 Transcripts  Proteins  Metabolites  Correlation  Literature-derived relationship

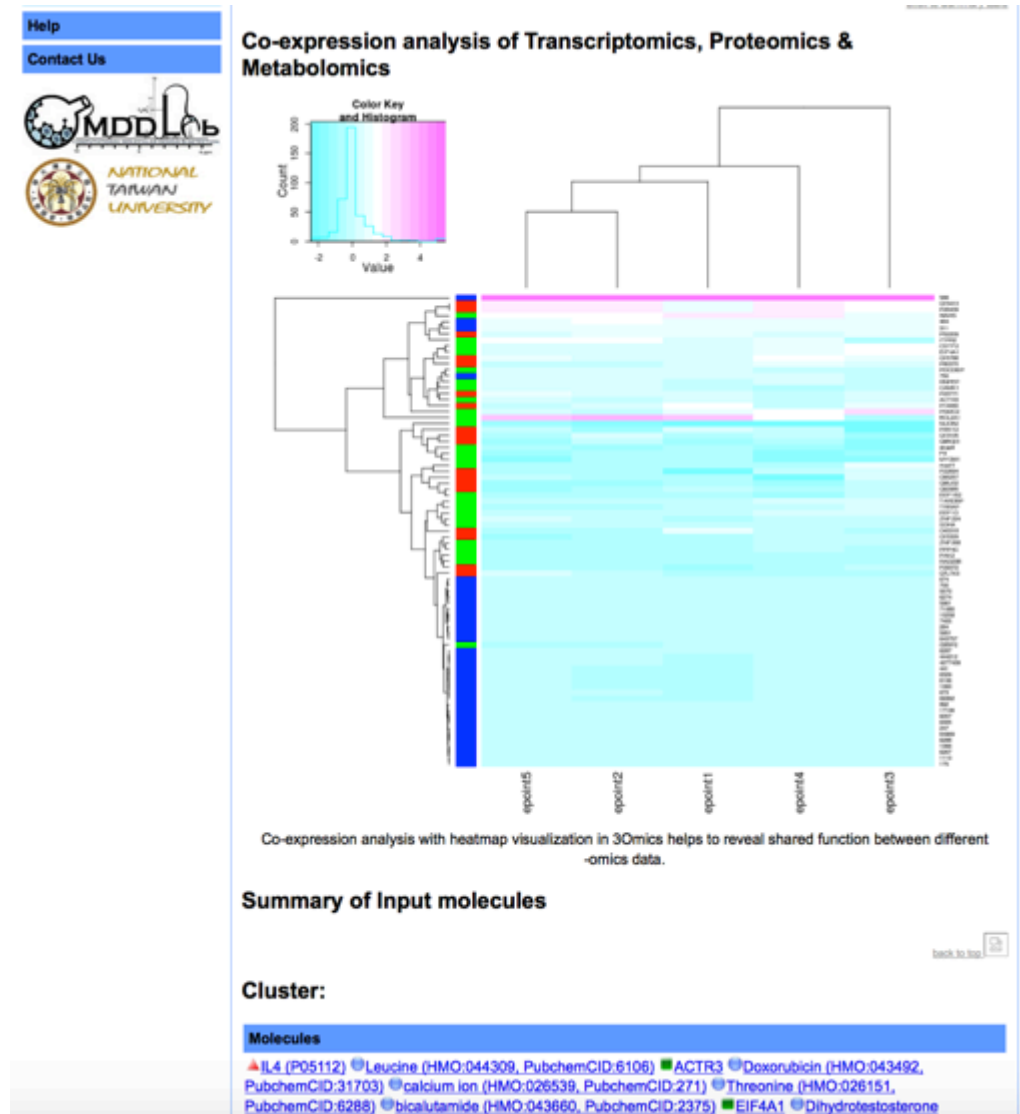
Wait for upload to complete

The screenshot displays the 3Omics web application interface. The browser address bar shows the URL `3omics.cmdm.tw/result.php`. The page features a navigation menu on the left with sections: Project Features (Overview, Name-ID Converter, Help, Contact Us), Parameters Section (How to set up parameters?, Correlation Coefficient Threshold: 0.9, Correlation Network Repulsion: 160, Correlation Network Attraction: 80, Refresh), and logos for MDD Lab and the National Taiwan University. The main content area has five tabs: Correlation Network, Coexpression Profile, Phenotype Analysis, Pathway Analysis, and GO Enrichment Analysis. The 'Summary of Input molecules' section is currently active, displaying a 'Loading...' message and a 'Click to Summary table' link. A 'back to top' link is also visible.

View Results – Correlation Network

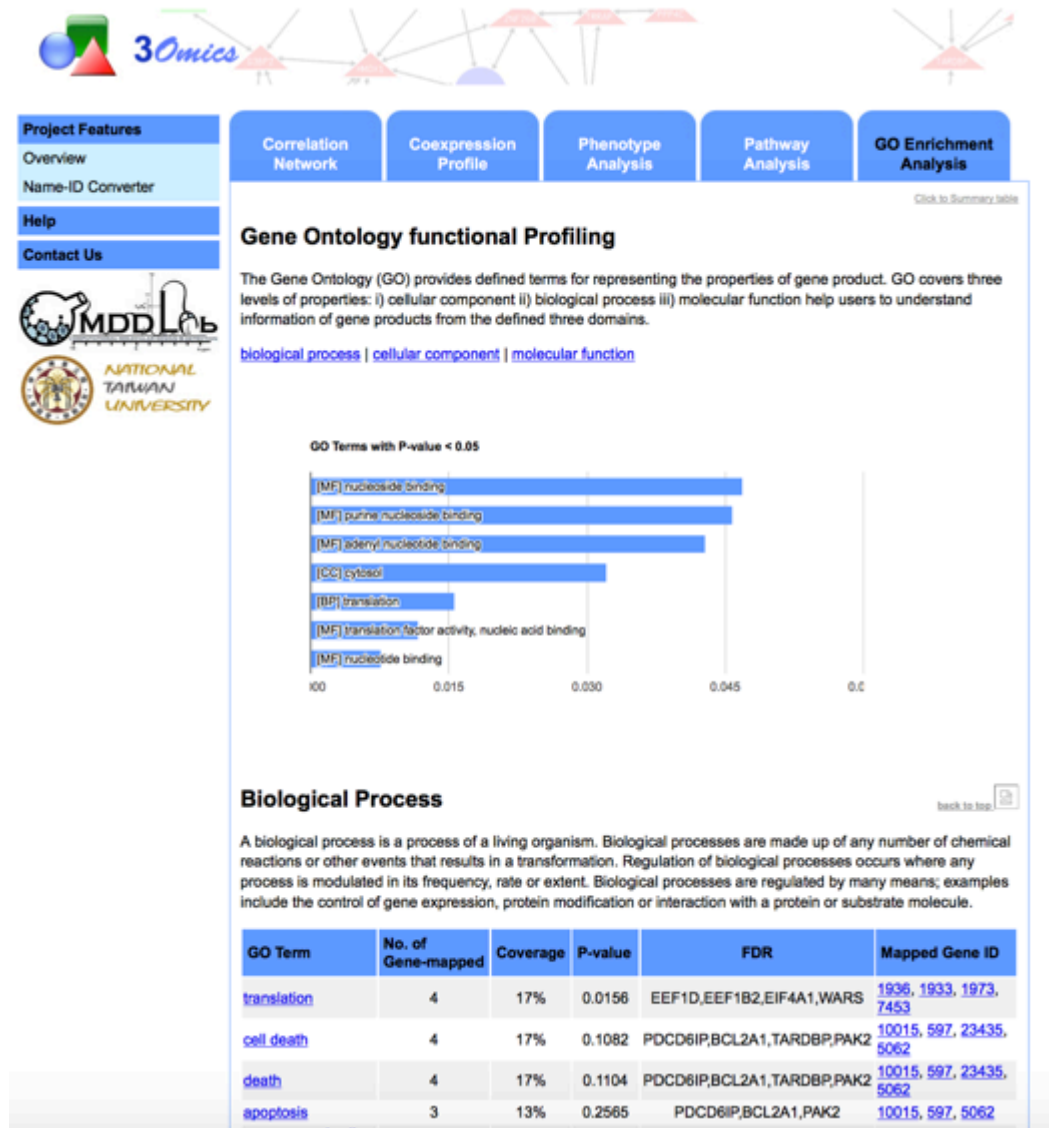


View Results – Co-expression analysis



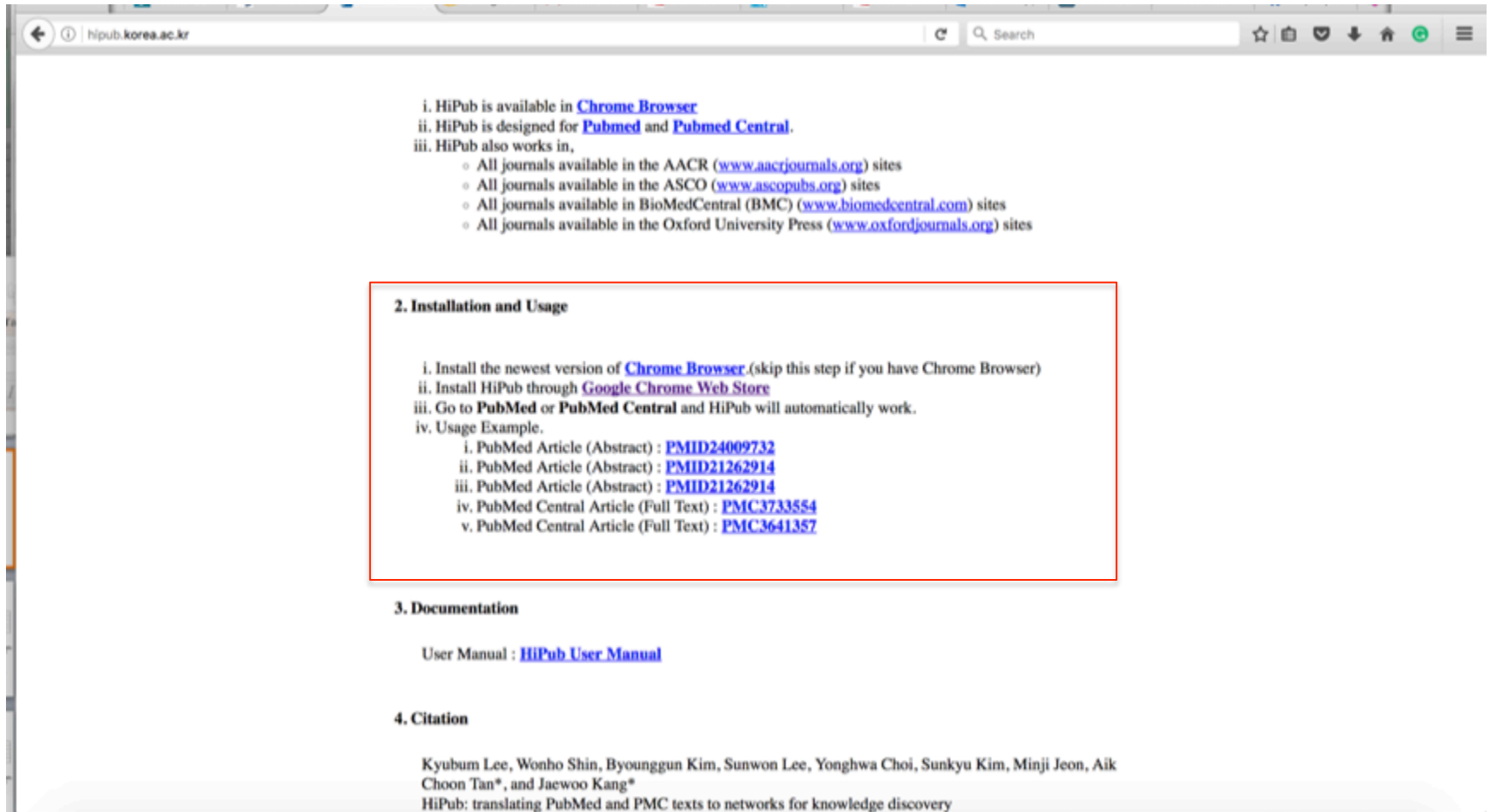
Rows: Variables
Columns: Samples

View Results - GO Enrichment Analysis



3. HiPub

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



The screenshot shows the HiPub website interface. The browser address bar displays 'hipub.korea.ac.kr'. The main content area lists instructions for using HiPub, including compatibility with Chrome Browser, PubMed, and PubMed Central, and a list of journals available. A red rectangular box highlights the '2. Installation and Usage' section, which includes steps for installing the newest version of Chrome Browser, installing HiPub through the Google Chrome Web Store, and providing usage examples with PubMed and PubMed Central article links. Below the box, the '3. Documentation' section links to the 'HiPub User Manual', and the '4. Citation' section lists the authors and the purpose of the project.

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

Installation

- Install Chrome:

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

- Install the HiPub plugin:

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

Open the following URL in Google Chrome:
<https://www.ncbi.nlm.nih.gov/pubmed/24009732>

The screenshot displays the PubMed web interface. At the top, the browser address bar shows the URL <https://www.ncbi.nlm.nih.gov/pubmed/24009732>. The PubMed logo and navigation links are visible. The main content area shows the abstract for a 2013 PLoS One article. The title is "The conformational control inhibitor of tyrosine kinases DCC-2036 is effective for imatinib-resistant cells expressing T674I FIP1L1-PDGFRα". The authors listed are Shen Y¹, Shi X, and Pan J. The abstract text describes the study of DCC-2036 as a third-generation TKI against imatinib-resistant cells. On the right side, there are sections for "Full text links" (PLOS, PMC), "Save items" (Add to Favorites), "Similar articles" (Antitumor activity of S116836, Cyclin-dependent kinase 7/9 inhibitor SNS-032, Ponatinib), "Cited by 5 PubMed Central articles" (Verteporfin induces apoptosis), and "Review" links. At the bottom, there is a section for "Images from this publication" showing several thumbnail images of graphs and diagrams.

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-PDGFRα.

Shen Y¹, 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 hyper eosinophilic 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](#) [PMC Full text](#)

Save items

[Add to Favorites](#)

Similar articles

Antitumor activity of S116836, a novel tyrosine kinase inhibitor, against imatinib [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 [Mol Cancer. 2014]

[Review](#) [FIP1L1-PDGFRα positive chronic eosinophilic [Zhonghua Xue Ye Xue Za Zhi. 2013]

[Review](#) The FIP1L1-PDGFRα fusion tyrosine kinase in hyper eosinophilic [Blood. 2004]

[See reviews...](#)

[See all...](#)

Cited by 5 PubMed Central articles

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

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

[HiPub](#)

4. xMWAS: R package

URL: <https://sourceforge.net/projects/xmwas/files/?source=navbar>

A) Install R: <https://mirrors.nics.utk.edu/cran/>

B) R commands for installing dependencies:

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

C) Install xMWAS: <https://sourceforge.net/projects/xmwas/files/?source=navbar>

Installation on Windows:

- 1) Download the "xMWAS_*.zip" file (<https://sourceforge.net/projects/xmwas/files/?source=navbar>)
- 2) Open R
- 3) Click on "Packages" under the menu bar
- 4) Click on "Install packages from local zip files"
- 5) Browse to the download location of the "xMWAS_*.zip" file
- 6) Double click on the file and installation should begin
- 7) Run "library(xMWAS)" command from within R to make sure the package is successfully installed

xMWAS installation on Mac:

- 1) Go to "Applications"
- 2) Open R
- 3) Go to "Packages & Data"
- 4) Select "Local Source Package" option from the drop down menu
- 5) Click on "Install"
- 6) Browse to the download location
- 7) Click "Open"

R code

```
library(xMWAS)
```

```
#example dataset that includes mRNA, protein, and miRNA data
```

```
data(exnci60)
```

```
data(classlabels_casecontrol) #example classlabels file for case vs control design
```

```
data(classlabels_repeatmeasures) #example classlabels file for repeat measures design
```

```
xMat<-exnci60$mrna
```

```
yMat<-exnci60$miRNA
```

```
zMat<-exnci60$prot
```

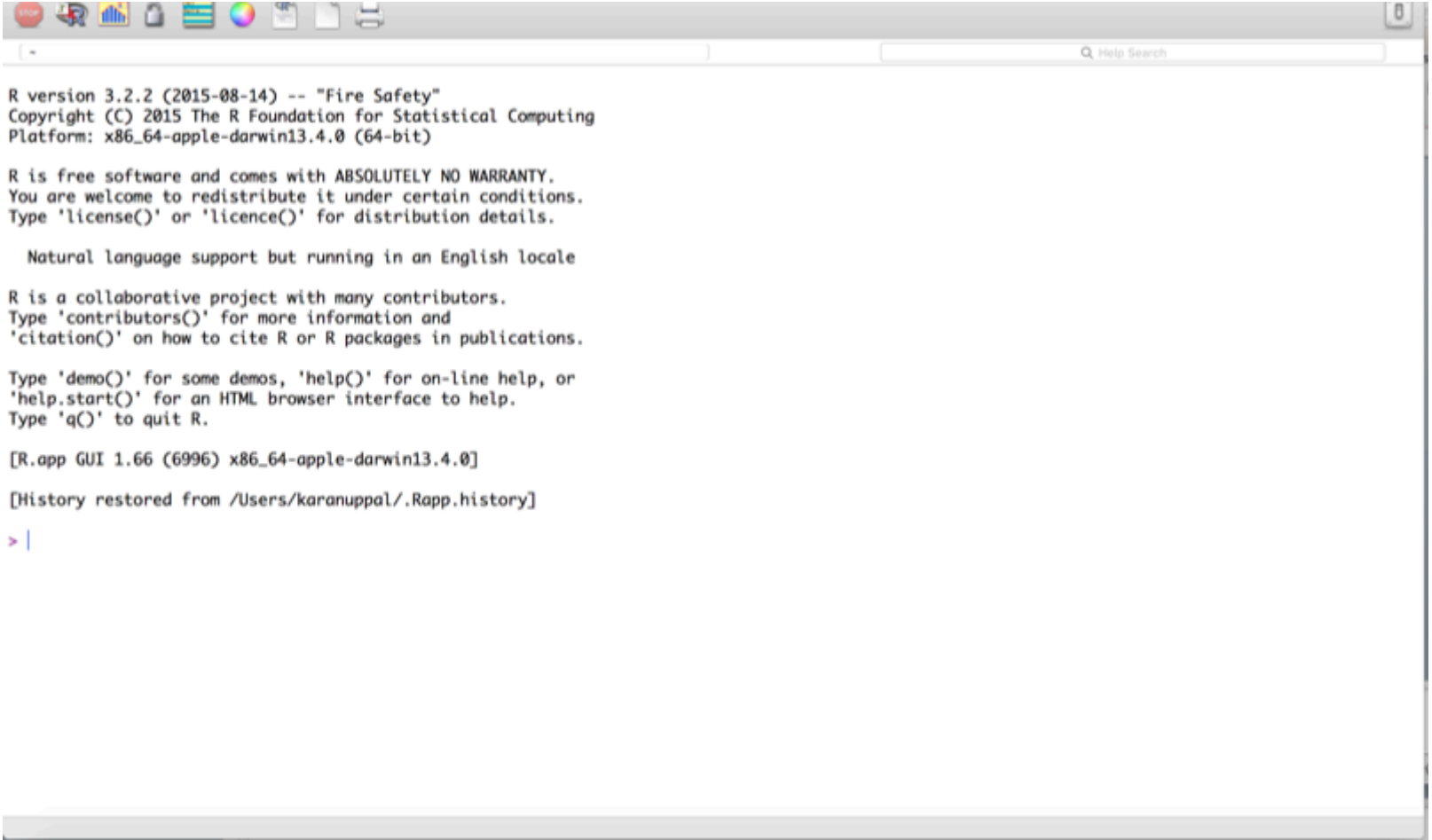
```
classlabels<-exnci60$classlabels
```

```
output<-"/Users/karanuppall/xMWASoutput/" #change for your computer
```

```
xmwas_res<-run_xmwas(Xome_data=xMat,Yome_data=yMat,Zome_data=zMat,Wome_data=NA,outloc=output,  
classlabels=classlabels,class_fname=NA,xmwasmethod="spl",plsmode="canonical",max_xvar=1000,max_yvar=1000,  
max_zvar=1000,max_wvar=1000,rsd.filt.thresh=1,corthresh=0.8,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","gold"),Xname="X",Yname="Y",Zname="Z",Wname="W",  
net_node_shape=c("rectangle","circle","triangle","star"),all.missing.thresh=0.3,maxnodesperclass=100,  
seednum=100,label.cex=0.2,vertex.size=6,graphclustering=TRUE,interactive=FALSE,max_connections=2000,  
centrality_method="eigenvector",use.X.reference=FALSE,removeRda=TRUE)
```

```
suppressWarnings(try(sink(file=NULL),silent=TRUE))
```

Open R



The screenshot shows the R console window on a macOS system. The window has a title bar with standard macOS window controls and a menu bar. The main content area displays the R startup sequence, including version information, copyright notice, and usage instructions. The prompt is a red greater-than sign followed by a vertical bar.

```
R version 3.2.2 (2015-08-14) -- "Fire Safety"
Copyright (C) 2015 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.4.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

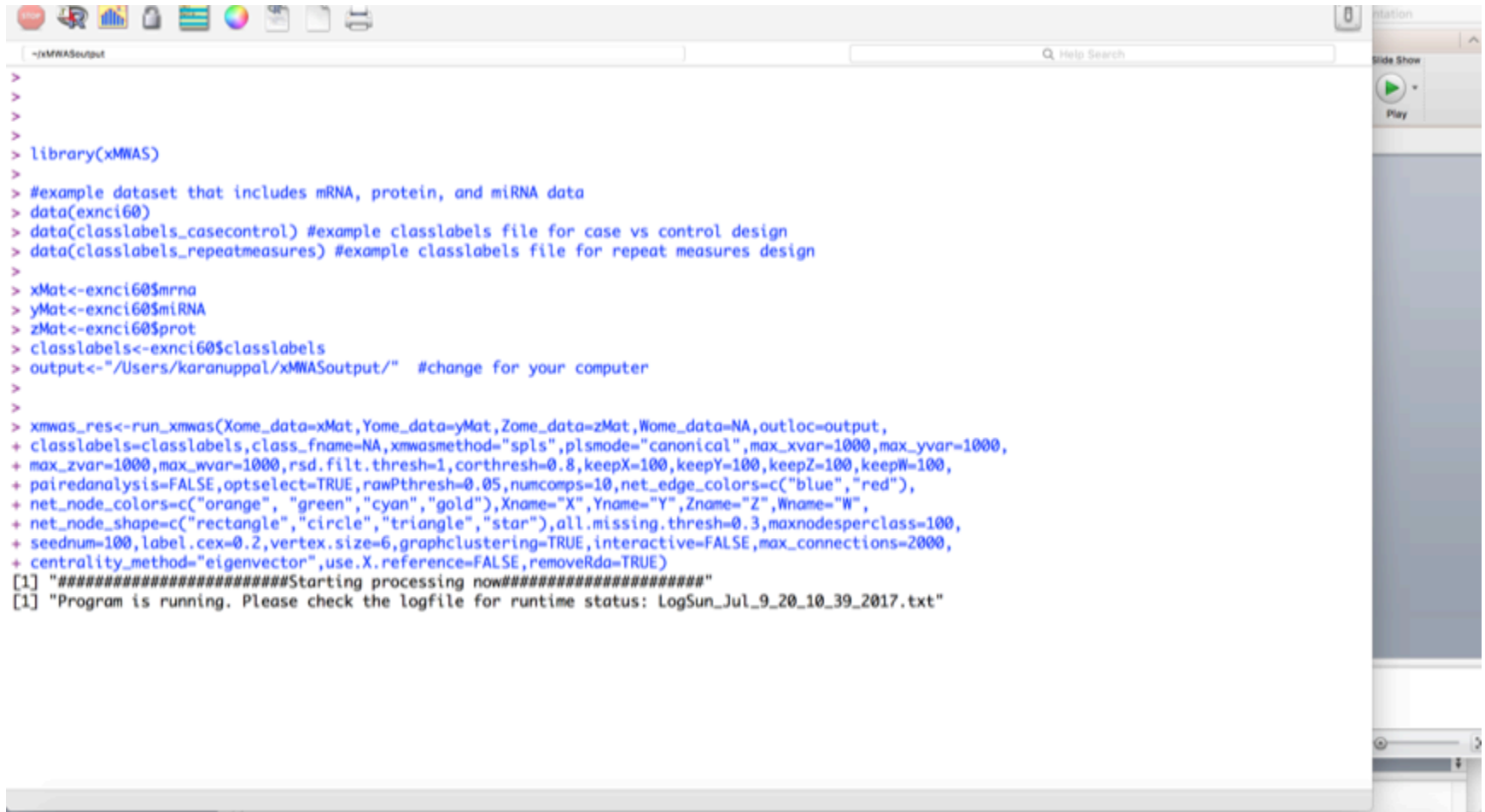
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.66 (6996) x86_64-apple-darwin13.4.0]

[History restored from /Users/karanuppal/.Rapp.history]

> |
```

Copy and paste the R code on slide 18 (after changing the output location)



```
>
>
>
> library(xMWAS)
>
> #example dataset that includes mRNA, protein, and miRNA data
> data(exnci60)
> data(classlabels_casecontrol) #example classlabels file for case vs control design
> data(classlabels_repeatmeasures) #example classlabels file for repeat measures design
>
> xMat<-exnci60$mrna
> yMat<-exnci60$miRNA
> zMat<-exnci60$prot
> classlabels<-exnci60$classlabels
> output<-" /Users/karanuppal/xMWASOutput/" #change for your computer
>
>
> xmwas_res<-run_xmwas(Xome_data=xMat,Yome_data=yMat,Zome_data=zMat,Wome_data=NA,outloc=output,
+ classlabels=classlabels,class_fname=NA,xmwasmethod="spl",plsmode="canonical",max_xvar=1000,max_yvar=1000,
+ max_zvar=1000,max_wvar=1000,rsd.filt.thresh=1,corthresh=0.8,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","gold"),Xname="X",Yname="Y",Zname="Z",Wname="W",
+ net_node_shape=c("rectangle","circle","triangle","star"),all.missing.thresh=0.3,maxnodesperclass=100,
+ seednum=100,label.cex=0.2,vertex.size=6,graphclustering=TRUE,interactive=FALSE,max_connections=2000,
+ centrality_method="eigenvector",use.X.reference=FALSE,removeRda=TRUE)
[1] "#####Starting processing now#####"
[1] "Program is running. Please check the logfile for runtime status: LogSun_Jul_9_20_10_39_2017.txt"
```


xMWAS – processing complete



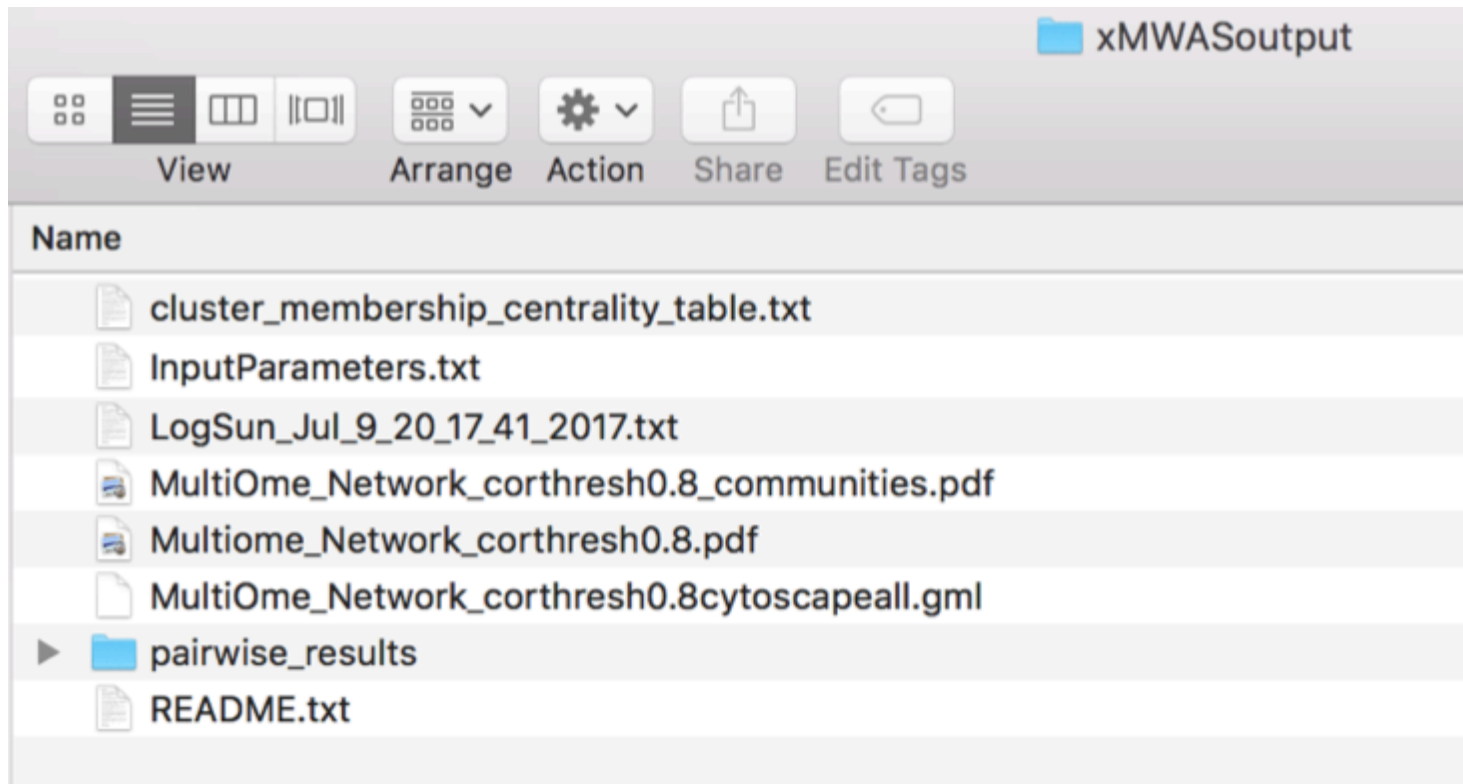
The following objects are masked from 'package:snow':

```
clusterApply, clusterApplyLB, clusterCall, clusterEvalQ, clusterExport, clusterMap, clusterSplit, makeCluster, parApply,
parCapply, parLapply, parRapply, parSapply, splitIndices, stopCluster
```

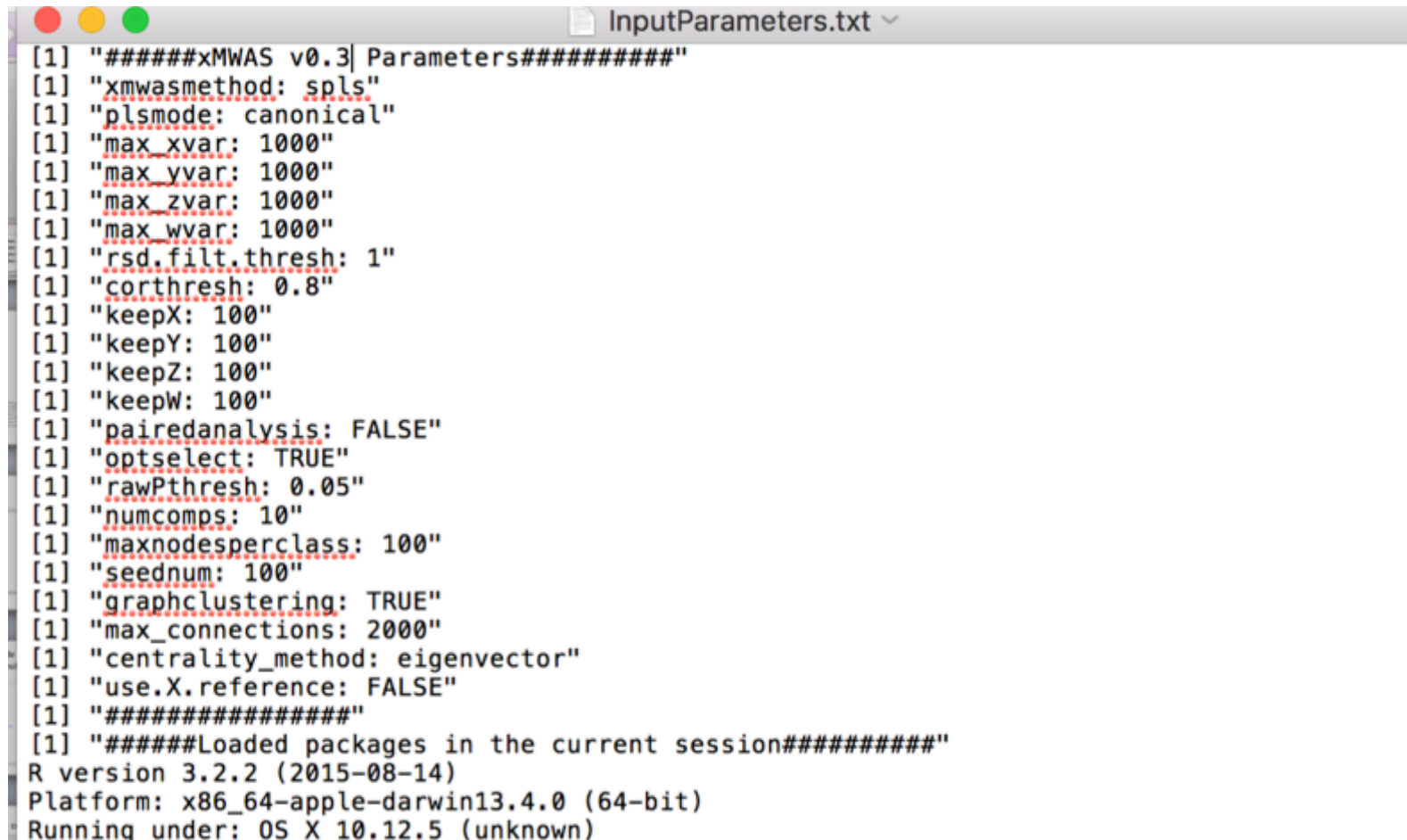
There were 12 warnings (use warnings() to see them)

```
>
> #example dataset that includes mRNA, protein, and miRNA data
> data(exnci60)
> data(classlabels_casecontrol) #example classlabels file for case vs control design
> data(classlabels_repeatmeasures) #example classlabels file for repeat measures design
>
> xMat<-exnci60$mrna
> yMat<-exnci60$miRNA
> zMat<-exnci60$prot
> classlabels<-exnci60$classlabels
> output<-"~/Users/karanuppai/xMWASoutput/" #change for your computer
>
>
> xmw_res<-run_xmw(Xome_data=xMat,Yome_data=yMat,Zome_data=zMat,Wome_data=NA,outloc=output,
+ classlabels=classlabels,class_fname=NA,xmwsmethod="spl",plsmode="canonical",max_xvar=1000,max_yvar=1000,
+ max_zvar=1000,max_wvar=1000,rsd.filt.thresh=1,corthresh=0.8,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","gold"),Xname="X",Yname="Y",Zname="Z",Wname="W",
+ net_node_shape=c("rectangle","circle","triangle","star"),all.missing.thresh=0.3,maxnodesperclass=100,
+ seednum=100,label.cex=0.2,vertex.size=6,graphclustering=TRUE,interactive=FALSE,max_connections=2000,
+ centrality_method="eigenvector",use.X.reference=FALSE,removeRda=TRUE)
[1] "#####Starting processing now#####"
[1] "Program is running. Please check the logfile for runtime status: LogSun_Jul_9_20_17_41_2017.txt"
[1] "Processing complete!"
>
> suppressWarnings(try(sink(file=NULL),silent=TRUE))
>
>
```

xMWAS output



Input Parameters



```
InputParameters.txt
[1] "#####xMWAS v0.3| Parameters#####"
[1] "xmwasmethod: spls"
[1] "plsmode: canonical"
[1] "max_xvar: 1000"
[1] "max_yvar: 1000"
[1] "max_zvar: 1000"
[1] "max_wvar: 1000"
[1] "rsd_filt_thresh: 1"
[1] "corthresh: 0.8"
[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] "maxnodesperclass: 100"
[1] "seednum: 100"
[1] "graphclustering: TRUE"
[1] "max_connections: 2000"
[1] "centrality_method: eigenvector"
[1] "use.X.reference: FALSE"
[1] "#####"
[1] "#####Loaded packages in the current session#####"
R version 3.2.2 (2015-08-14)
Platform: x86_64-apple-darwin13.4.0 (64-bit)
Running under: OS X 10.12.5 (unknown)
```

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 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)."

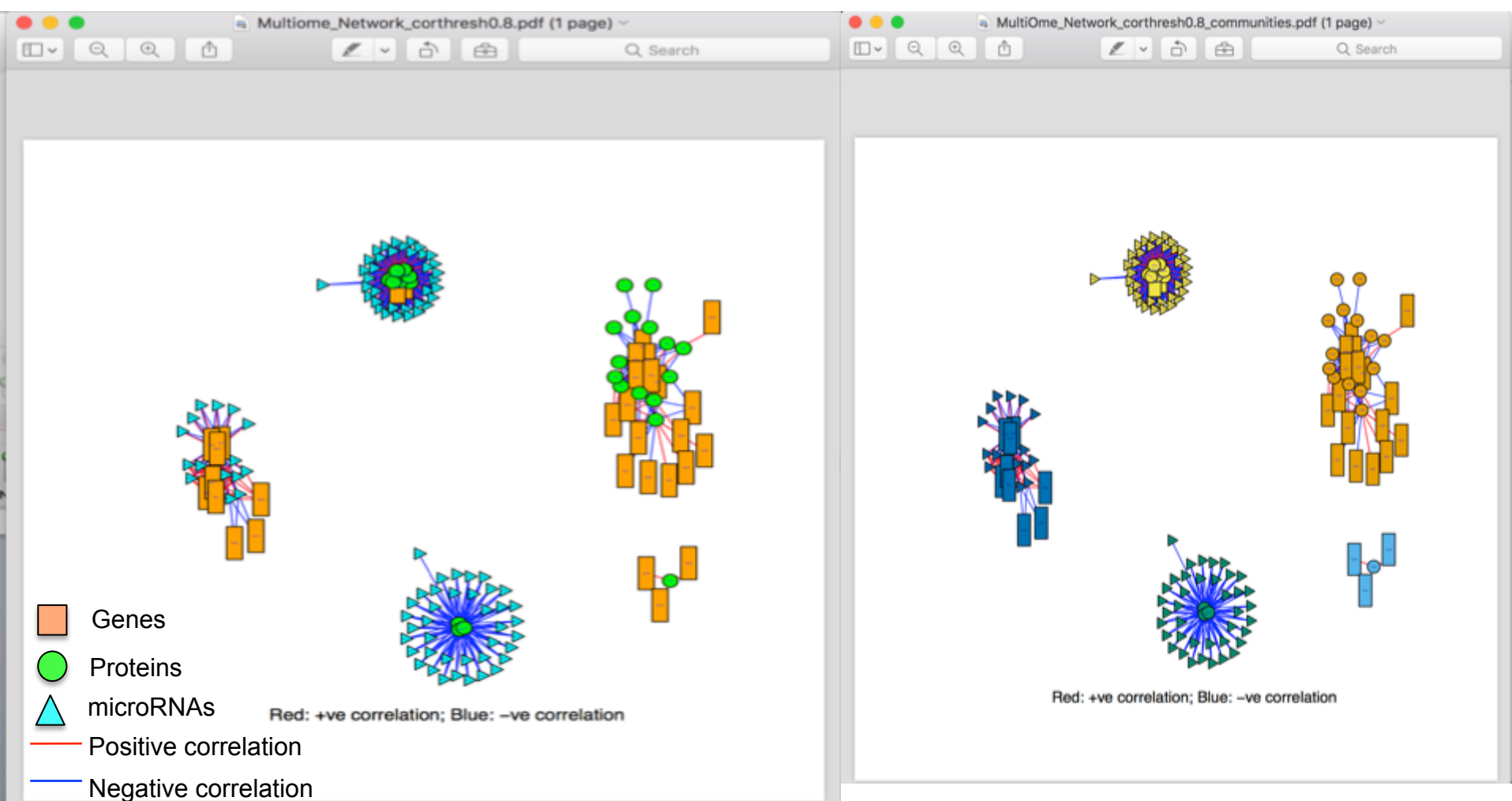
"5: MultiOme_Network_corthreshxcytoscape.gml: GML file for all significantly associated variables that can be uploaded to Cytoscape"

"6: The cluster_membership_centralty_mapped.txt file includes community detection results using the multilevel community detection algorithm and the centrality measures."

"7: The matrix_centralty.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



Community detection and centrality analysis

cluster_membership_centralty_table.txt

100%

Home Layout Tables Charts SmartArt Formulas Data Review

Edit Font Alignment Number

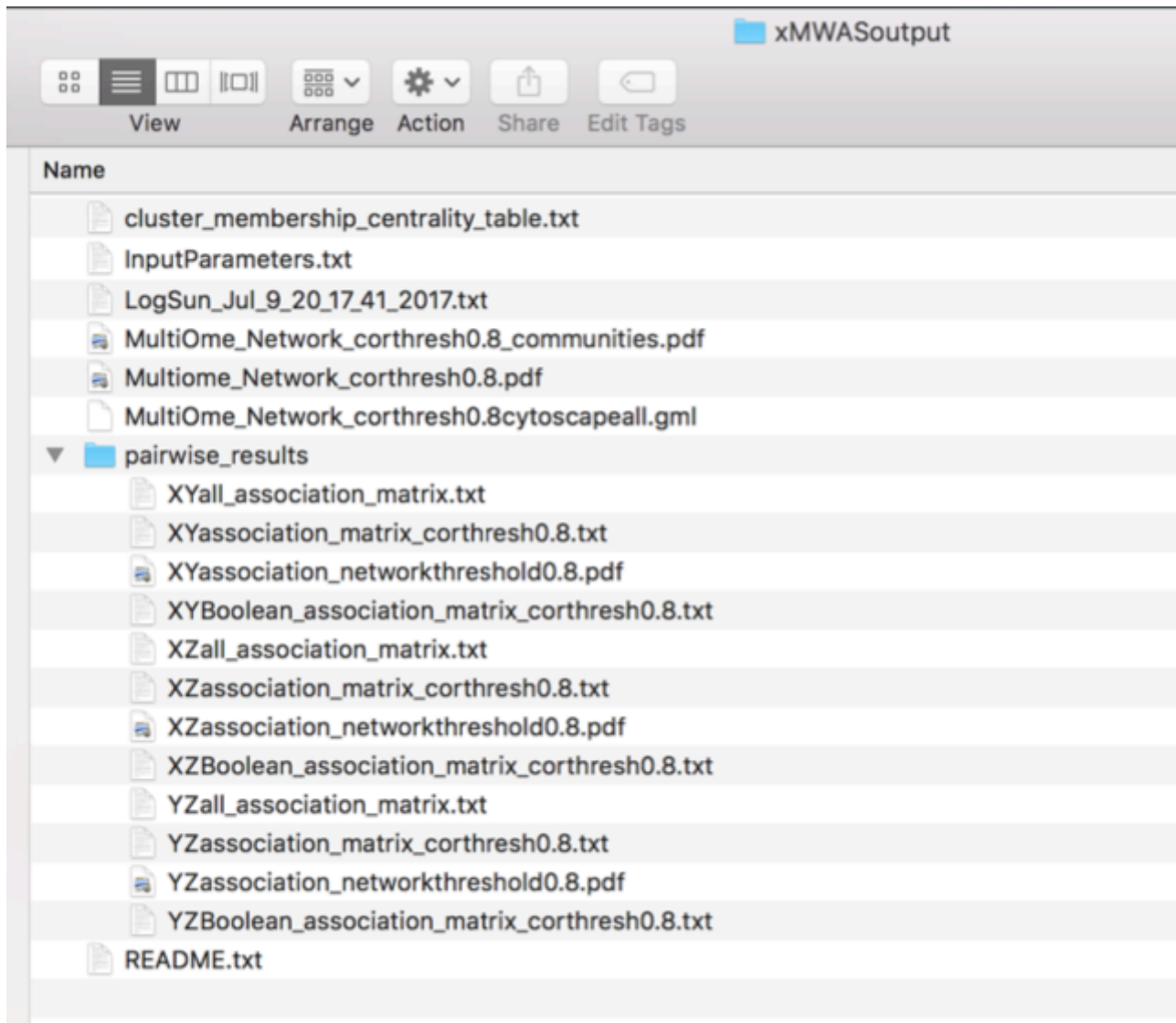
Paste Fill Calibri (Body) 12 A A abc Wrap Text General

B I U Merge

C10 fx MI0000060_miRNA 1

	A	B	C	D	E	F	G	H	I
	Node	Cluster	Name	centrality_vec					
1	Y8		4 MI0000065_miRNA 8	1					
2	Y14		4 MI0000068_miRNA 14	0.972					
3	X321		4 HLA-Z_4158_mrna	0.971					
4	Y198		4 MI0000269_miRNA 203	0.936					
5	X689		4 RBBP9_9004_mrna	0.93					
6	Y3		4 MI0000061_miRNA 3	0.91					
7	Y4		4 MI0000062_miRNA 4	0.909					
8	Y7		4 MI0000065_miRNA 7	0.908					
9	Y1		4 MI0000060_miRNA 1	0.906					
10	Y297		4 MI0000297_miRNA 304	0.898					
11	Y204		4 MI0000289_miRNA 209	0.897					
12	Y6		4 MI0000064_miRNA 6	0.893					
13	X105		5 WDR35_12243_mrna	0					
14	X175		5 TBK1_11191_mrna	0					
15	X21		5 GP5_3824_mrna	0					
16	X218		2 TGFB3_11288_mrna	0					
17	X265		1 CDT1_1793_mrna	0					
18	X266		1 KIRREL_4812_mrna	0					

Pairwise results – $X \leftrightarrow Y$, $X \leftrightarrow Z$, $Y \leftrightarrow Z$



Questions/Comments?