Computational methods for data integration

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

Disclosure of Conflict of Interest

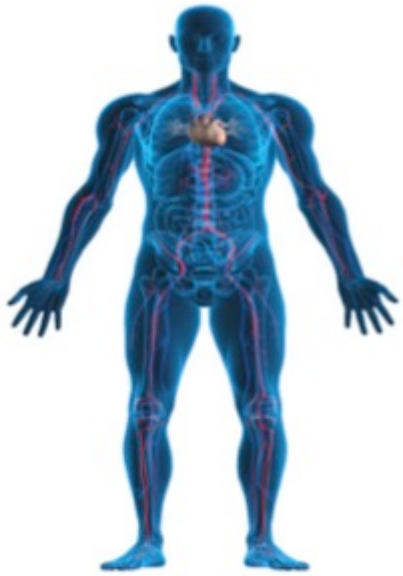
- I have no actual or potential conflict of interest in relation to this presentation.

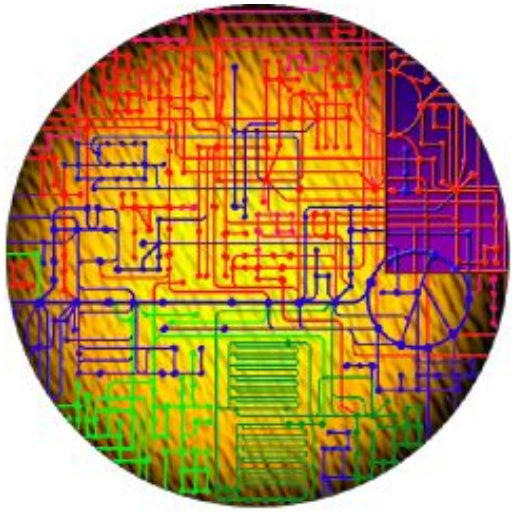
Learning Objectives

- 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

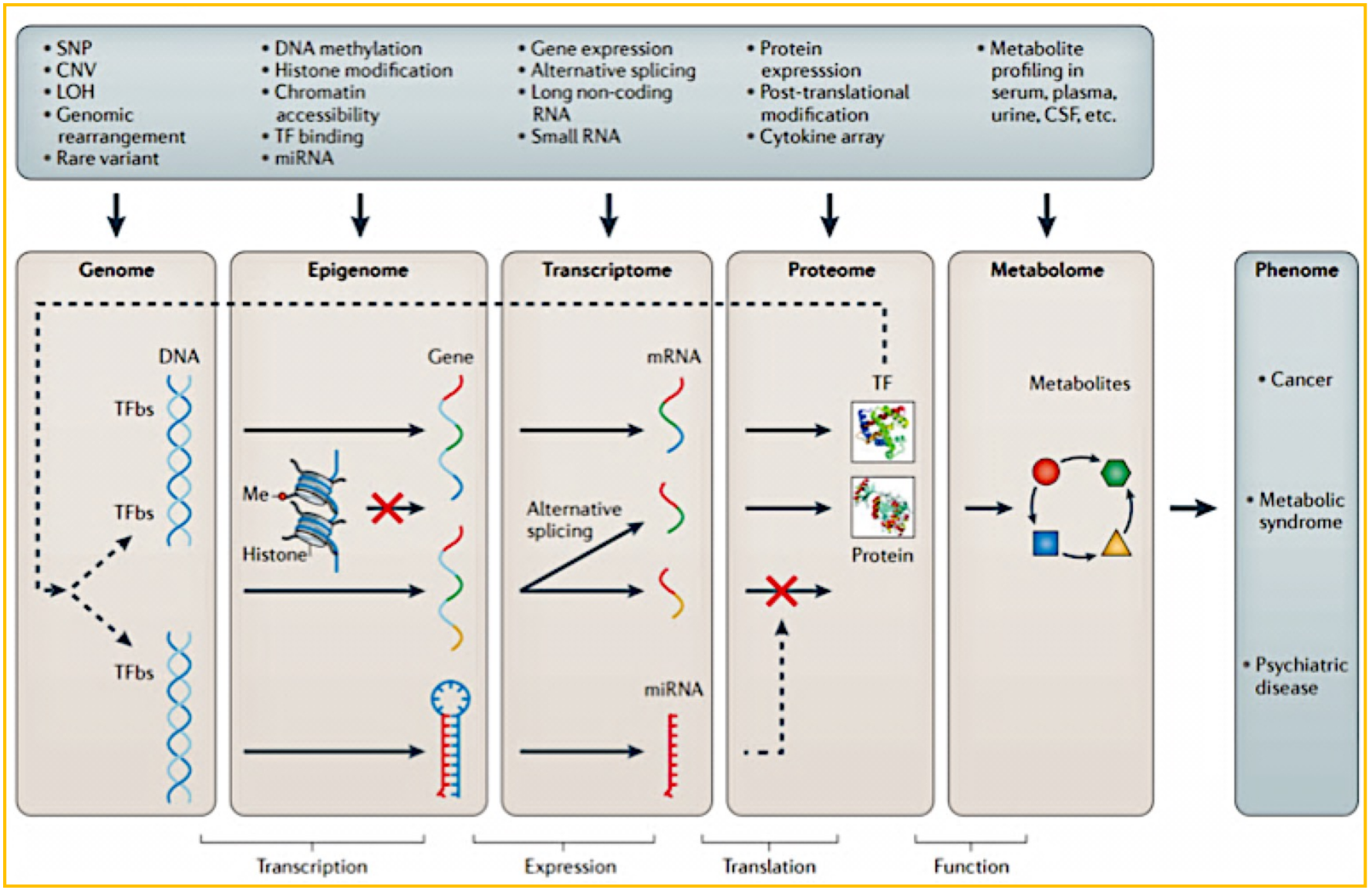




Exposures
Internal measurements
Disease states

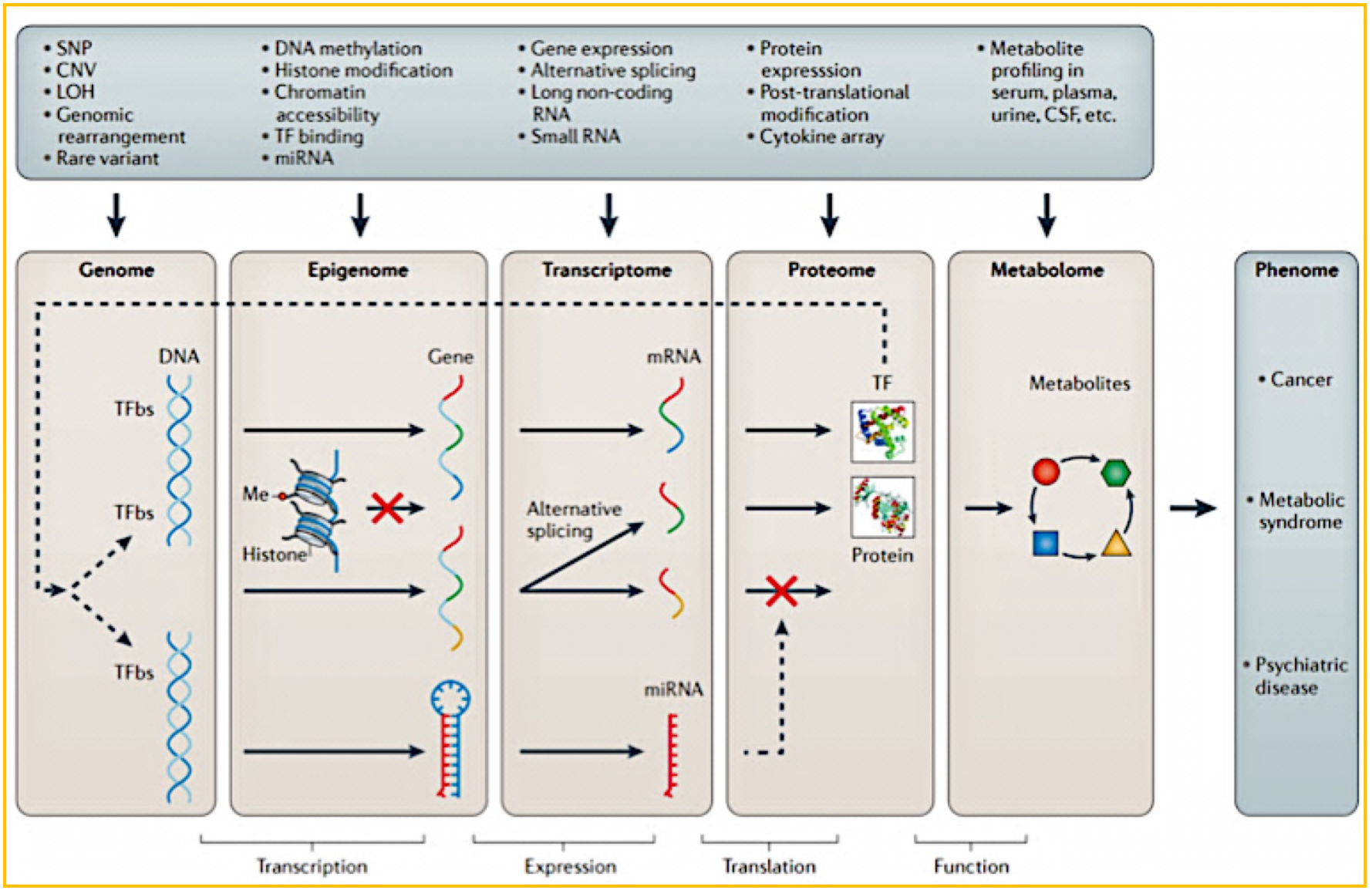
Systems Biology
"*Integrative approach in which scientists study pathways and networks will touch all areas of biology, including drug discovery*"
C. Henry and C. Washington

Integrative omics: 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.*

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



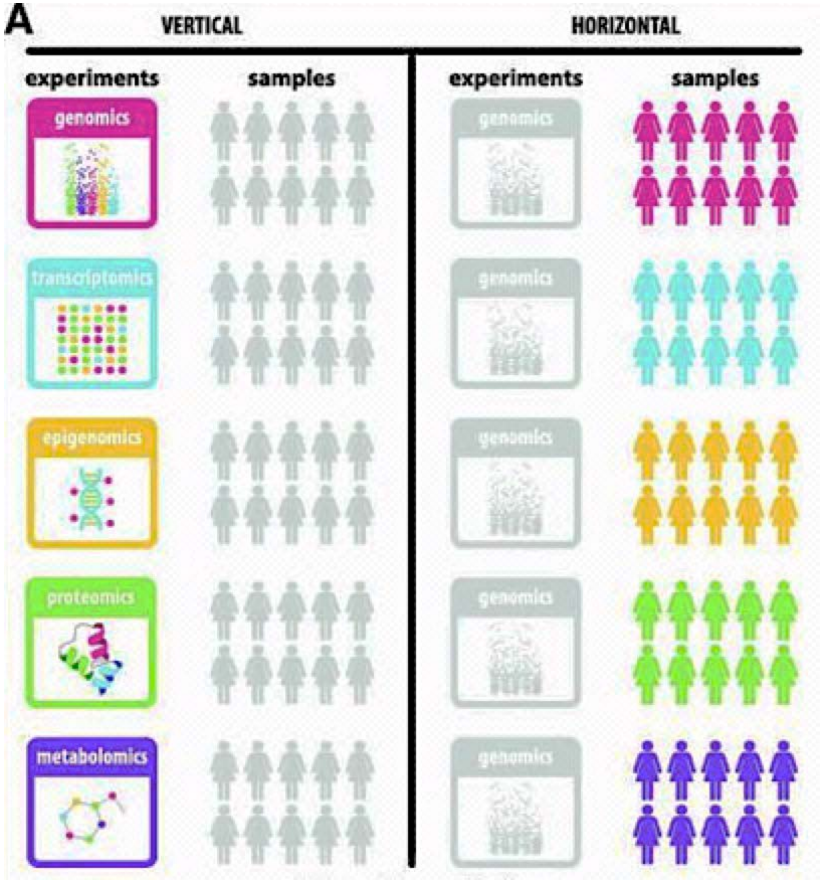
Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. 2015. Methods of integrating data to uncover genotype-phenotype interactions. Nat Rev Genet *16:85-97.*

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 collected on the same subjects 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

Data integration study designs

- Paired or vertical integrative analysis
  - Integrative analysis of **multiple omics datasets** from the **same N subjects**
  - Discover networks of associations or correlated variables (e.g. genes, proteins, metabolites, microbiome, epigenetic alterations, clinical variables)
    - Univariate or multivariate regression
    - Example: explaining protein abundance with respect to gene expression
- Horizontal integrative analysis
  - Meta-analysis of **multiple studies/cohorts** looking at the same type of data
  - Cross-laboratory or cross-platform comparisons



Eidem 2018, BMC Med Genomics
Ref: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6245874/

Main approaches for data integration

- Pathway-based integration
  - Pathway information from KEGG or other databases
  - 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
  - Interpretation using pathway analysis tools
  - Examples: 3Omics, mixOmics, xMWAS
- Using literature-derived associations for integration
  - Using co-occurrence criteria for establishing relationship
  - Examples: HiPub, CoPub, ArrowSmith

Main approaches for data integration

- **Pathway-based integration**
  - **Pathway information from KEGG or other databases**
  - **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
  - Interpretation using pathway analysis tools
  - Examples: 3Omics, mixOmics, xMWAS
- Using literature-derived associations for integration
  - Using co-occurrence criteria for establishing relationship
  - Examples: HiPub, CoPub, ArrowSmith

Pathway-based data integration - I

Metabolomics data
(n subjects X p metabolites)

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

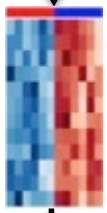
Differentially expressed metabolites

Over-represented pathways

Rank	Pathway ID (hsa:)	Pathway Title
1	04974	Protein digestion and absorption
2	02010	ABC transporters
3	00250	Alanine, aspartate and glutamate metabolism
4	00330	Arginine and proline metabolism
5	00480	Glutathione metabolism
6	00260	Glycine, serine and threonine metabolism
7	00910	Nitrogen metabolism
8	00460	Cyanoamino acid metabolism
9	00270	Cysteine and methionine metabolism
10	00770	Pantothenate and CoA biosynthesis

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

Over-represented pathways

Rank	Pathway ID (hsa:)	Pathway Title
1	00260	Glycine, serine and threonine metabolism
2	00340	Histidine metabolism
3	00480	Glutathione metabolism
4	00450	Selenoamino acid metabolism
5	00360	Phenylalanine metabolism
6	00071	Fatty acid metabolism
7	00330	Arginine and proline metabolism
8	00561	Glycerolipid metabolism
9	00380	Tryptophan metabolism
10	00250	Alanine, aspartate and glutamate metabolism

Pathway-based data integration - I

Metabolomics data
(n subjects X p metabolites)

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

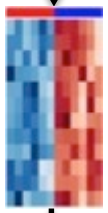
Differentially expressed metabolites

Over-represented pathways

Rank	Pathway ID (hsa:)	Pathway Title
1	04974	Protein digestion and absorption
2	02010	ABC transporters
3	00250	Alanine, aspartate and glutamate metabolism
4	00330	Arginine and proline metabolism
5	00480	Glutathione metabolism
6	00260	Glycine, serine and threonine metabolism
7	00910	Nitrogen metabolism
8	00460	Cyanoamino acid metabolism
9	00270	Cysteine and methionine metabolism
10	00770	Pantothenate and CoA biosynthesis

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

Over-represented pathways

Rank	Pathway ID (hsa:)	Pathway Title
1	00260	Glycine, serine and threonine metabolism
2	00340	Histidine metabolism
3	00480	Glutathione metabolism
4	00450	Selenoamino acid metabolism
5	00360	Phenylalanine metabolism
6	00071	Fatty acid metabolism
7	00330	Arginine and proline metabolism
8	00561	Glycerolipid metabolism
9	00380	Tryptophan metabolism
10	00250	Alanine, aspartate and glutamate metabolism

Pathway rank aggregation

Rank	Pathway ID (hsa:)	Pathway Title
1	00260	Glycine, serine and threonine metabolism
2	00330	Arginine and proline metabolism
3	00480	Glutathione metabolism

iPEAP (Sun 2013)

Pathway-based data integration - II

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)

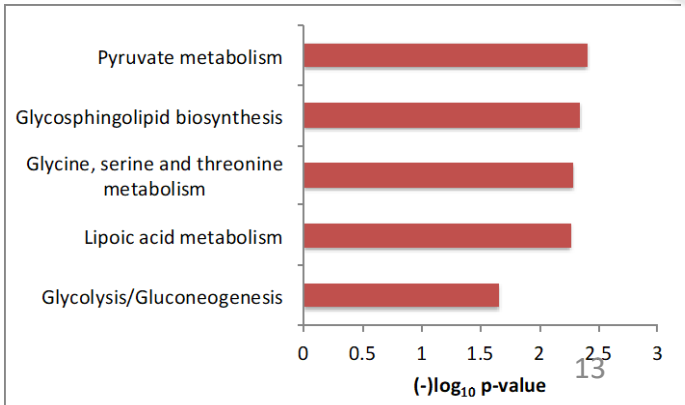
	G1	G2	-	Gq
Subject1	19	19	-	100
Subject2	10	40	-	90
-	-	-	-	-
SubjectN	10	40	-	50

Differentially expressed metabolites

Differentially expressed genes

MetaboAnalyst4.0 – Joint Pathway Analysis module

Over-representation analysis in KEGG using gene and metabolite IDs

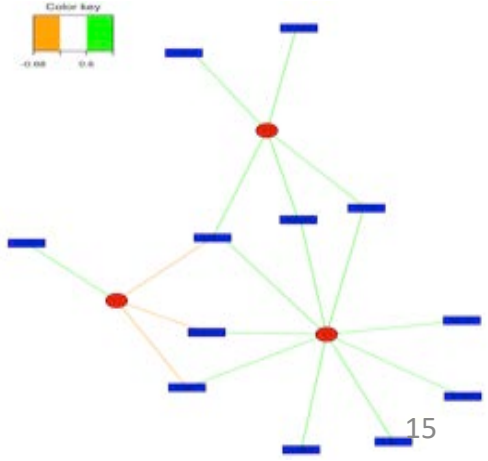


Main approaches for data integration

- Pathway-based integration
  - Pathway information from KEGG or other databases
  - 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**
  - **Interpretation using pathway analysis tools**
  - **Examples: 3Omics, mixOmics, xMWAS**
- Using literature-derived associations for integration
  - Using co-occurrence criteria for establishing relationship
  - Examples: HiPub, CoPub, ArrowSmith

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: genes
Rectangles: metabolites

Methods for generating relevance networks

- Univariate
  - 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): a data-driven integration and differential network analysis
    - Availability: https://kuppal.shinyapps.io/xmwas (Online) and https://github.com/kuppal2/xMWAS/ (R)

3Omics: a web-based tool for analysis, integration and visualization of human transcriptome, proteome and metabolome data (Kuo 2013, BMC Systems Biology)

- Web-based tool
- Correlation analysis and network visualization
- Additional features:
  - Metabolic pathway analysis
  - Gene ontology enrichment analysis
  - Hierarchical clustering analysis

URL: http://3omics.cmdm.tw/

Transcriptomics

Gene	Sample 1	Sample 2	Sample 3
akap9	-0.24	-0.6	-0.47
macf1	-0.3	-0.3	0.48
RNPEP	0.24	0.85	0.15
SDHA	0.1	0.37	0.18

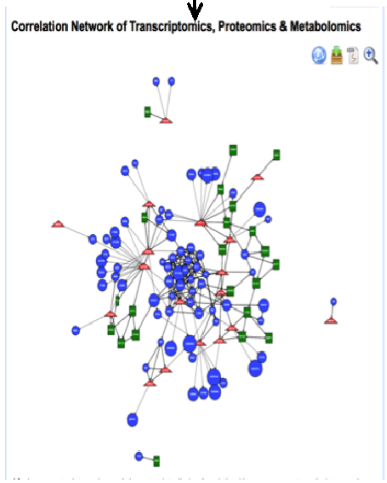
Metabolomics

Metabolite	Sample 1	Sample 2	Sample 3
4277439	-0.3109937	-0.2792995	-0.2548517
441	-0.2967872	-0.2895908	-0.2674823
69362	-0.3183692	-0.2828533	-0.272917
10258	-0.0614116	-0.1180467	-0.1231662

Proteomics

Protein	Sample 1	Sample 2	Sample 3
P14060	2.06	1.2	1.61
P26439	1.8	3.57	2.04
P29372	-0.64	-0.71	-0.21
Q96J02	-0.52	-1.34	-0.15

Pairwise Pearson correlation analysis



Metabolites
Genes
Proteins
Data-derived correlation
Literature-derived association

xMWAS: a data-driven integration and differential network analysis (Uppal 2018, Bioinformatics)

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

R package: https://github.com/kuppal2/xMWAS

Transcriptomics

Gene	Sample 1	Sample 2	Sample 3
akap9	-0.24	-0.6	-0.47
macf1	-0.3	-0.3	0.48
RNPEP	0.24	0.85	0.15
SDHA	0.1	0.37	0.18

Metabolomics

Metabolite	Sample 1	Sample 2	Sample 3
4277439	-0.3109937	-0.2792995	-0.2548517
441	-0.2967872	-0.2895908	-0.2674823
69362	-0.3183692	-0.2828533	-0.272917
10258	-0.0614116	-0.1180467	-0.1231662

Proteomics

Protein	Sample 1	Sample 2	Sample 3
P14060	2.06	1.2	1.61
P26439	1.8	3.57	2.04
P29372	-0.64	-0.71	-0.21
Q96J02	-0.52	-1.34	-0.15

↓

Pairwise (sparse) Partial Least Squares regression for data integration (Cao 2009)

↓

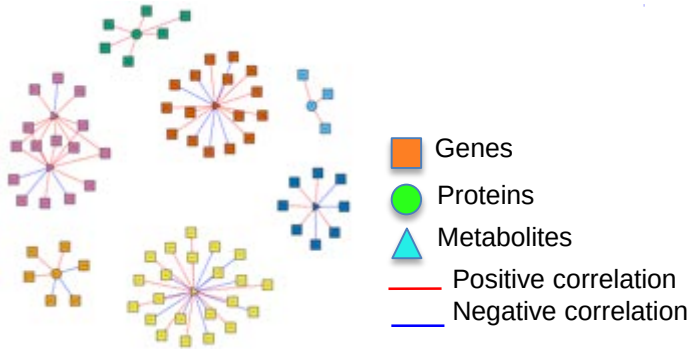
Approximation of Pearson correlation using PLS components

↓

Filtering based on |r|>threshold and p-value<alpha criteria

↓

Community (clusters) detection and centrality (importance) analysis



Sparse Partial Least Squares (PLS) regression method (Cao 2009, Liquet 2012)

- 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
- Multilevel sparse PLS – accounts for repeated measures
- 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

Association matrix using the PLS components

→

	X1	X2	-	Xn
Y1	0.4	0.9	-	0.3
Y2	0.7	0.1	-	0.5
Y3	0.1	0.6		0.8

Community detection

- Community: set of densely connected nodes that have more connections with the nodes in the same community as compared to nodes in other communities
- Multilevel community detection: a multi-step procedure

  1) each node is assigned to a different community

  2) each node is moved to a community with which it achieves the highest positive contribution to modularity

  3) Step 2 is repeated for all nodes until no improvement can be achieved

  4) Each community after step 3 is now considered a node and step 2 is repeated until there is a single node left or the modularity can no longer be improved

Centrality analysis

- Centrality: measure of importance of a node in the network
- Common centrality measures
  - 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
- Differential centrality analysis: delta centrality between two conditions (e.g. $|centrality_{exposed} – centrality_{control}|$

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

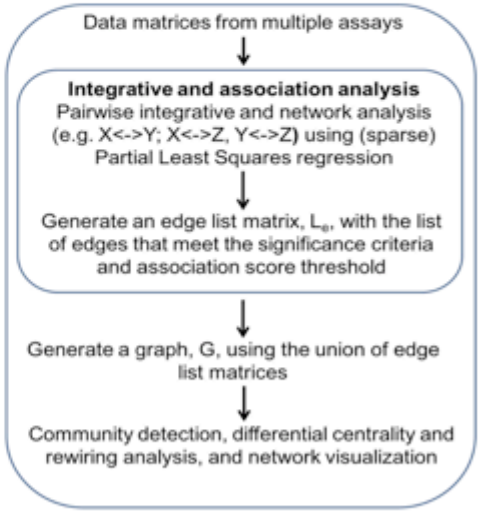
xMWAS - a data-driven integration and network analysis tool (v0.54)

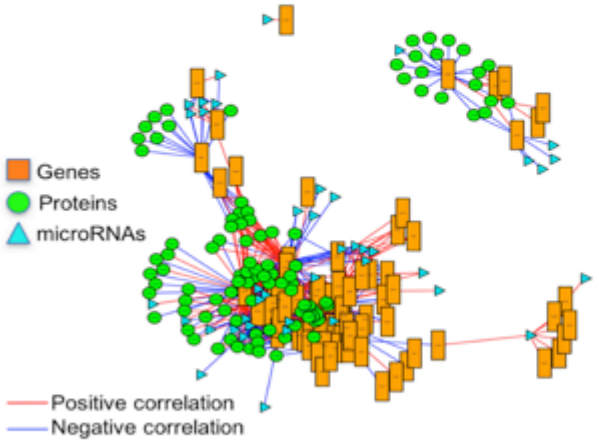
Introduction | Analysis | Help and Support

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")`

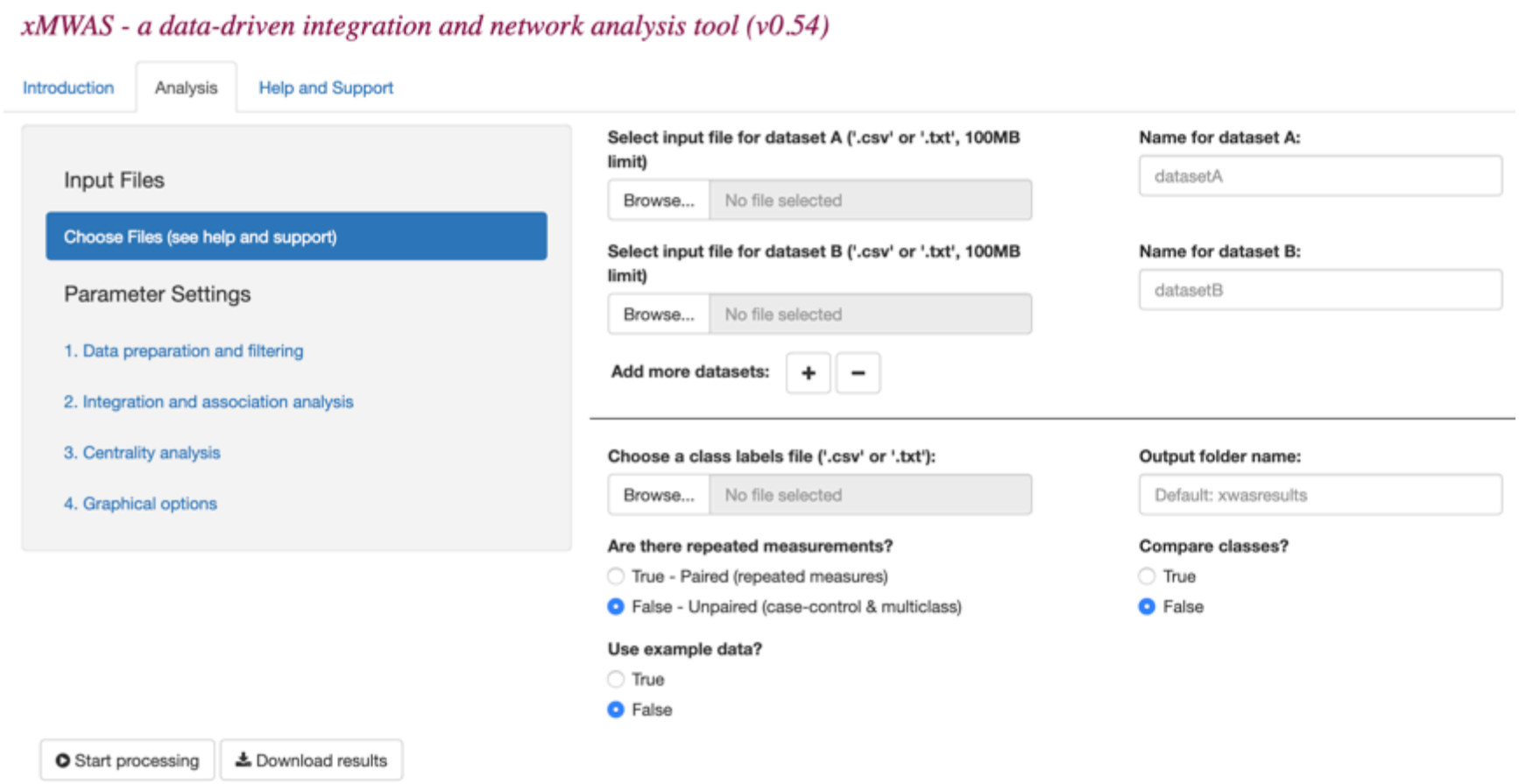




Citation: Uppal K, Ma C, Go YM, Jones DP. xMWAS: a data-driven integration and differential network analysis tool. *Bioinformatics*. 2017 Oct 23. PMID: 29069296

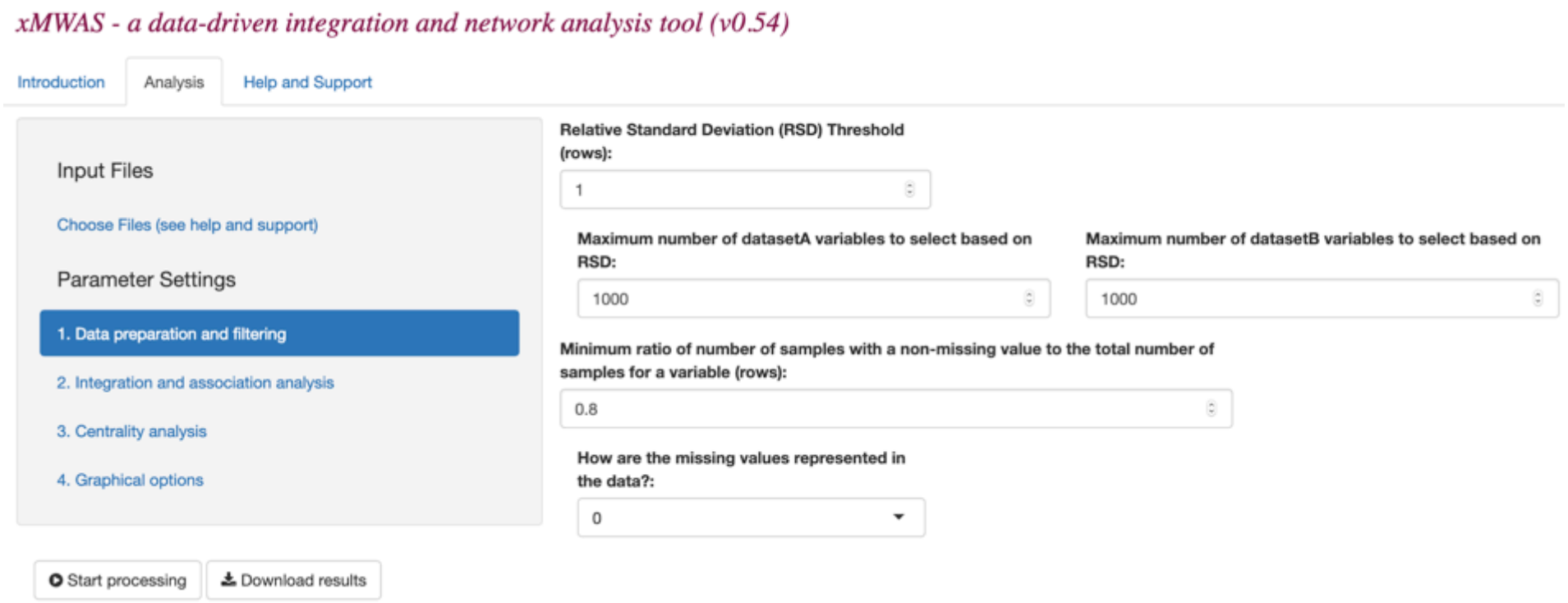
Maintained by Chunyu Ma (chunyu.ma@emory.edu) and Karan Uppal (kuppal2@emory.edu) at Clinical Biomarkers Laboratory , Emory University, Atlanta, GA, USA

Step 1. Upload data files



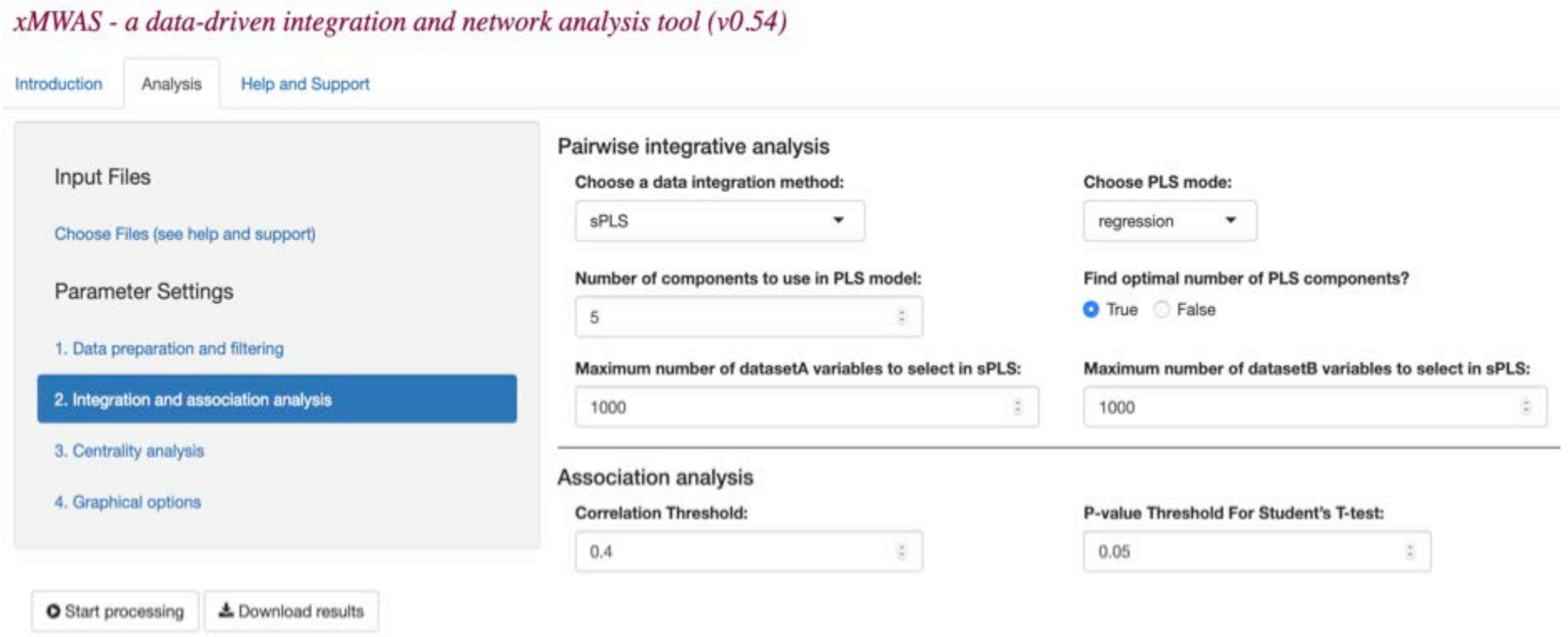
https://kuppal.shinyapps.io/xmwas/
(See: Help & Support)

Step 2. Data preprocessing and filtering



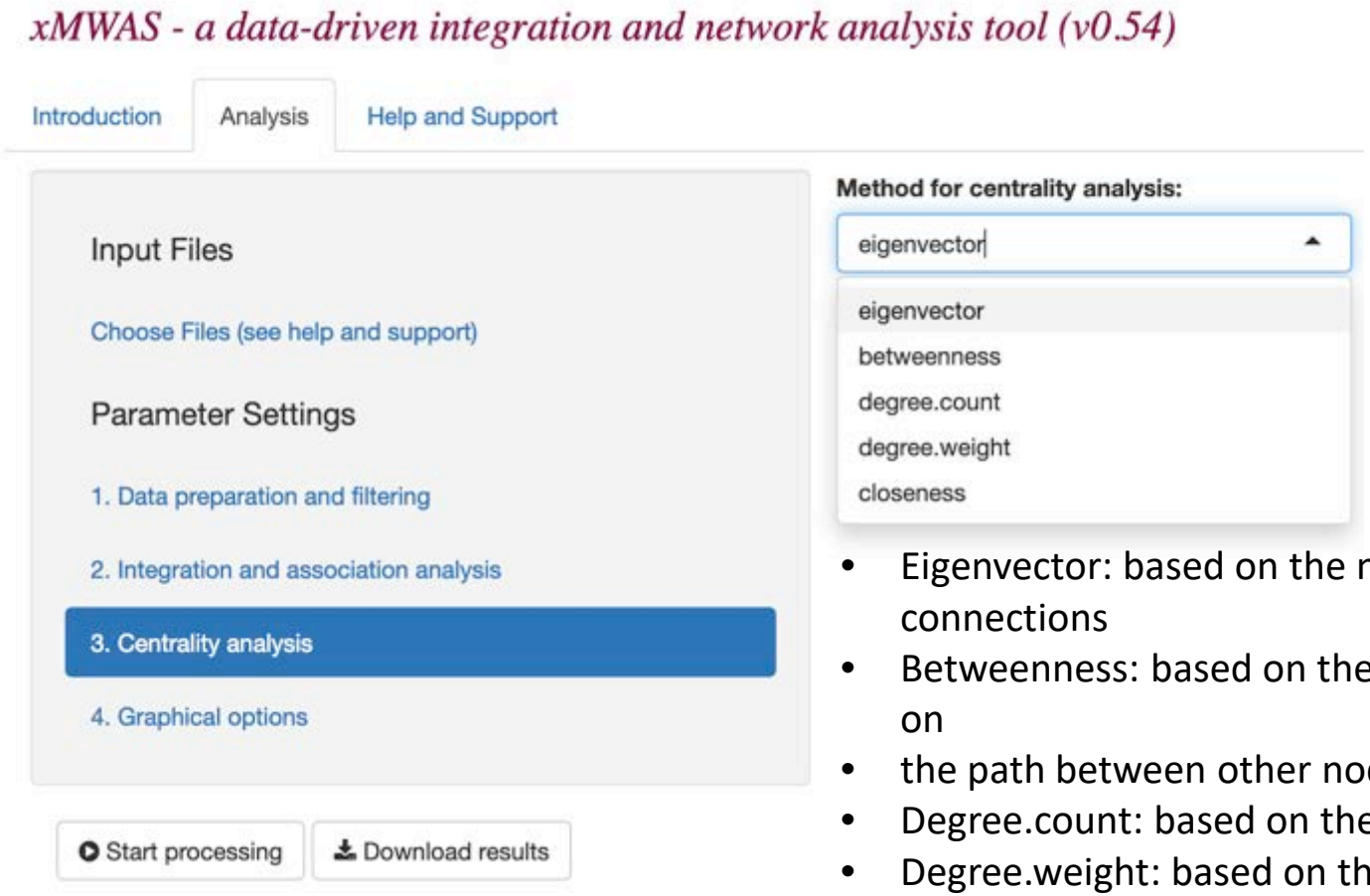
https://github.com/kuppal2/xMWAS/blob/master/example_manual_tutorial/xMWAS-manual.pdf

Step 3. Set parameters for integration and association analysis



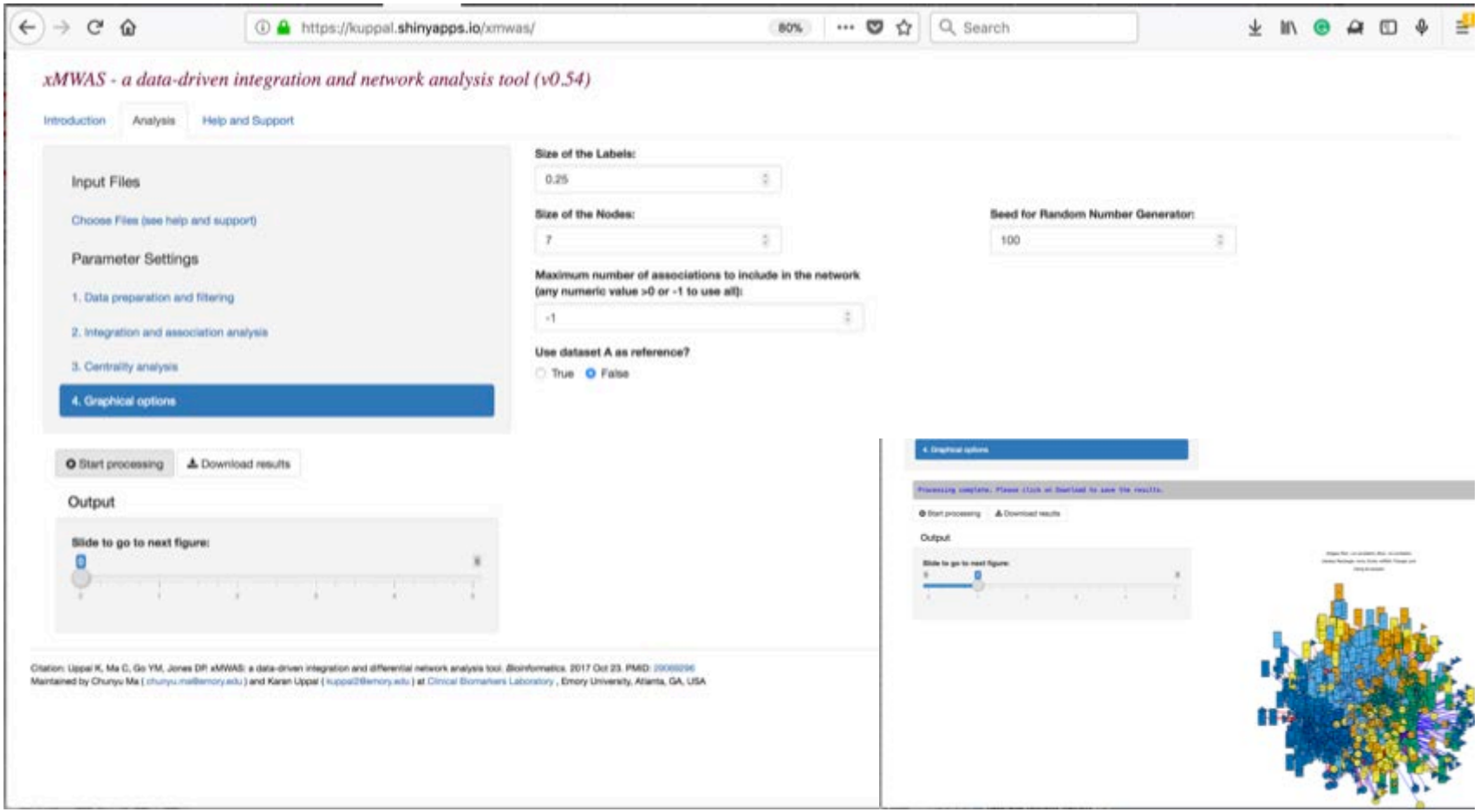
https://github.com/kuppal2/xMWAS/blob/master/example_manual_tutorial/xMWAS-manual.pdf

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. Click on “Start processing”



Additional methods

- Additional methods
  - MINT and Diablo in mixOmics R packages for horizontal and vertical data integration
  - Recent review article by Meng et al. in Bioinformatics reviewed over 20 dimensionality reduction methods for data integration

Comparison of 5 methods for assessing **pairwise associations** between variables implemented in MetabNet, 3Omics, xMWAS, and mixOmics

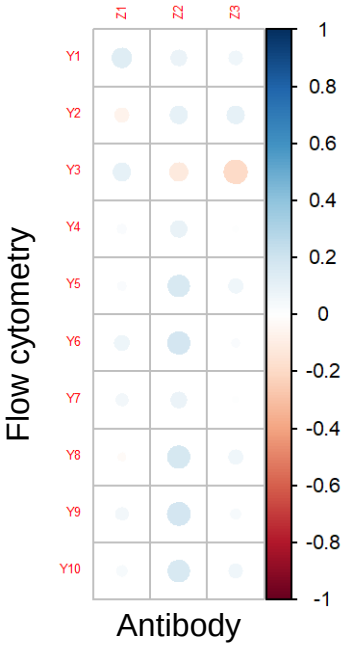
	3Omics	MetabNet	xMWAS	mixOmics
Pearson correlation	x	x	x	x
PLS (regression)			x	x
PLS (canonical)			x	x
Regularized canonical correlation analysis (RCC)				X
sparse generalized canonical correlation analysis (sGCCA)				x

T cell responses to H1N1v and a longitudinal study of seasonal influenza vaccination (TIV) - 2011

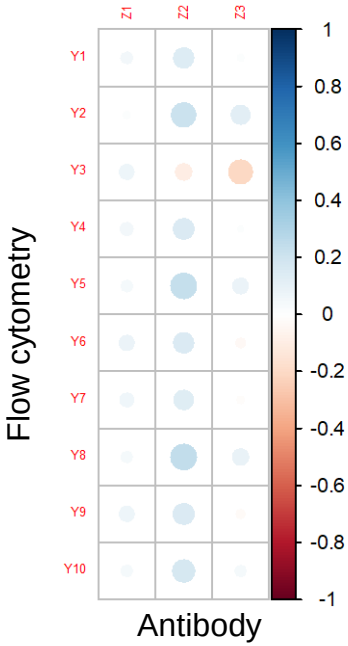
(https://www.immport.org/shared/study/SDY112)

- Transcriptomics at day 0: 18,867 genes
- Flow cytometry at day 0: 24 cells
- Antibody at day 0 and 28: 3 antibodies (California, Perth, and Brisbane)
- Number of subjects:
  - Total: 89
  - 85 had matching gene expression, flow cytometry, and antibody data at day 0

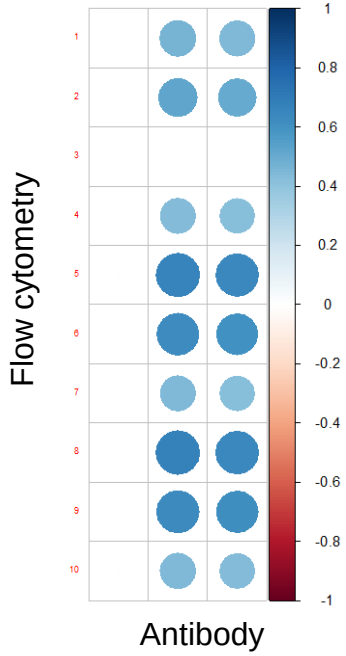
A. Pearson Correlation



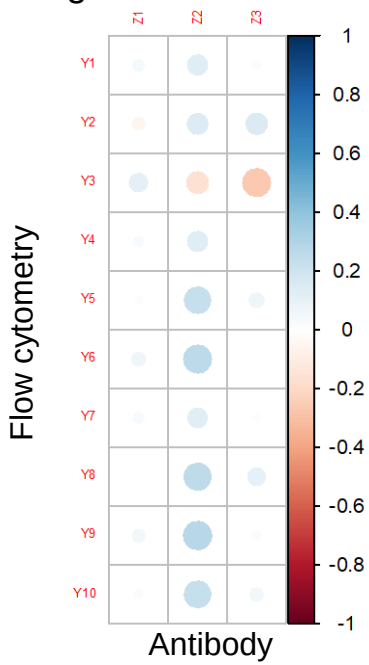
B. PLS – regression mode



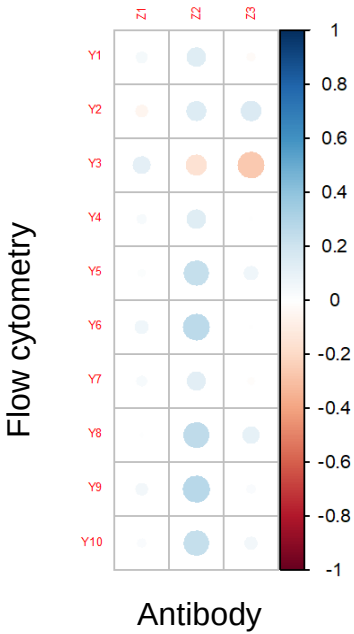
E. Sparse generalized canonical correlation analysis (sGCCA)



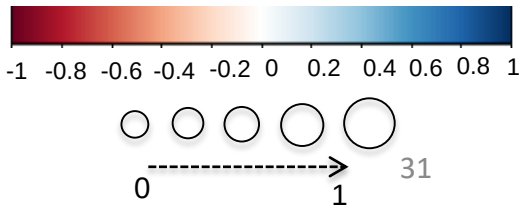
C. Regularized canonical correlation



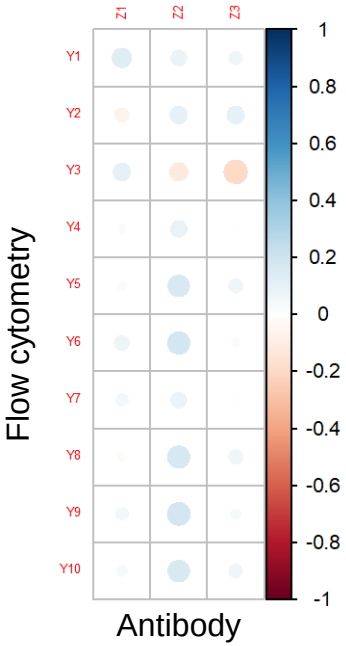
D. PLS - canonical mode



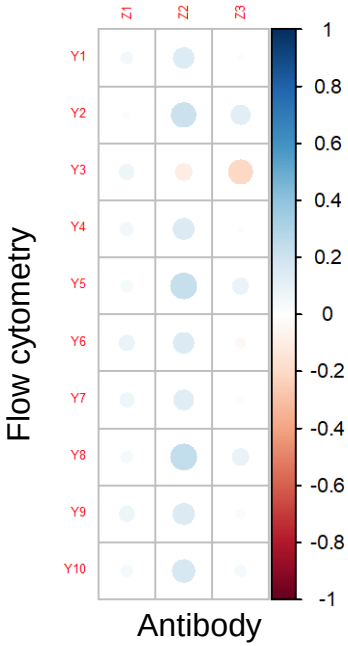
Association score



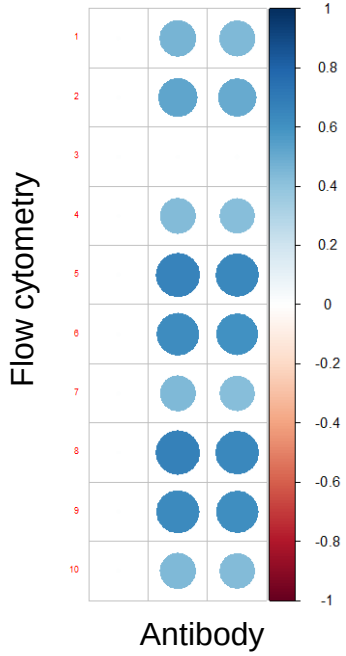
A. Pearson Correlation



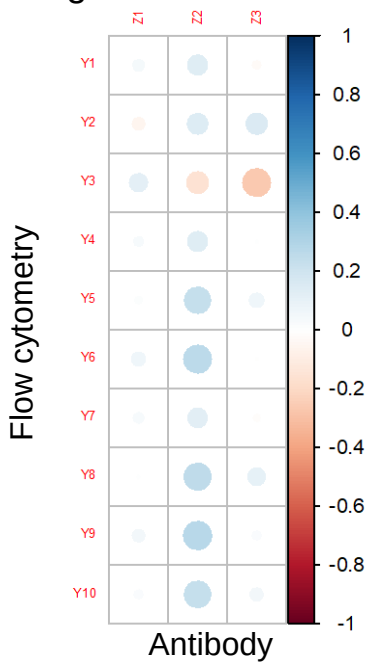
B. PLS – regression mode



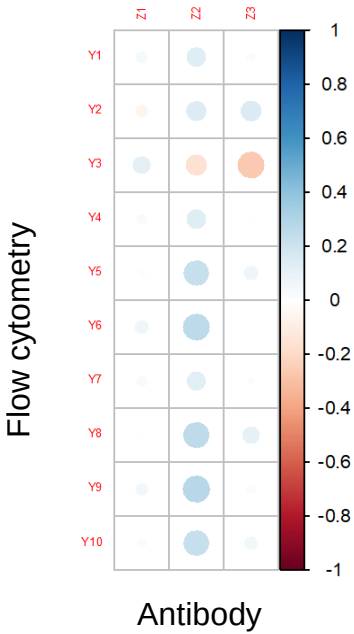
E. Sparse generalized canonical correlation analysis (sGCCA)



C. Regularized canonical correlation

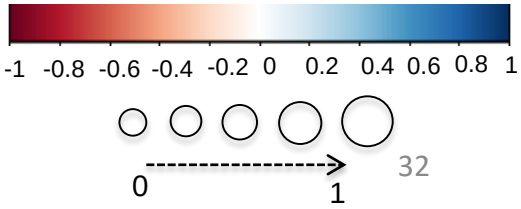


D. PLS - canonical mode



Inflated association scores using methods A and D – interpretation challenges!

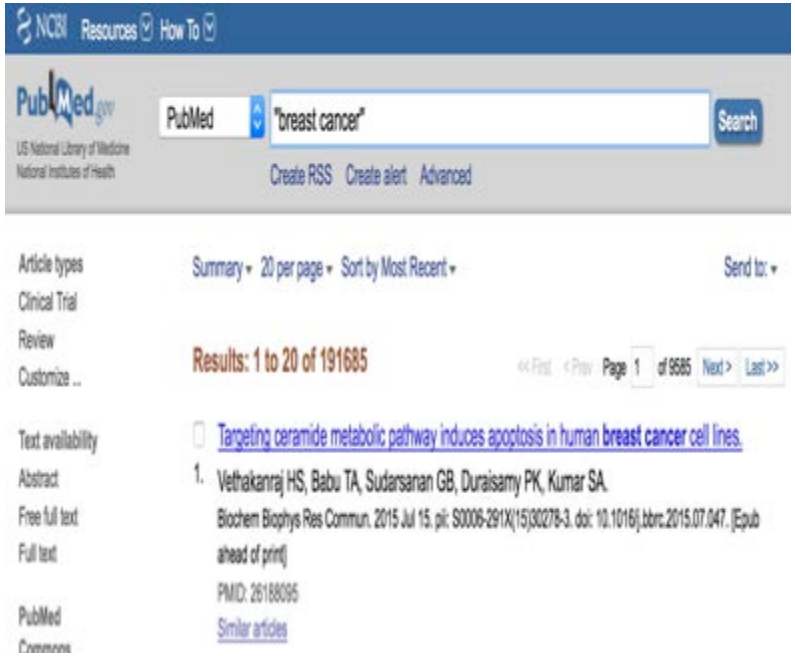
Association score

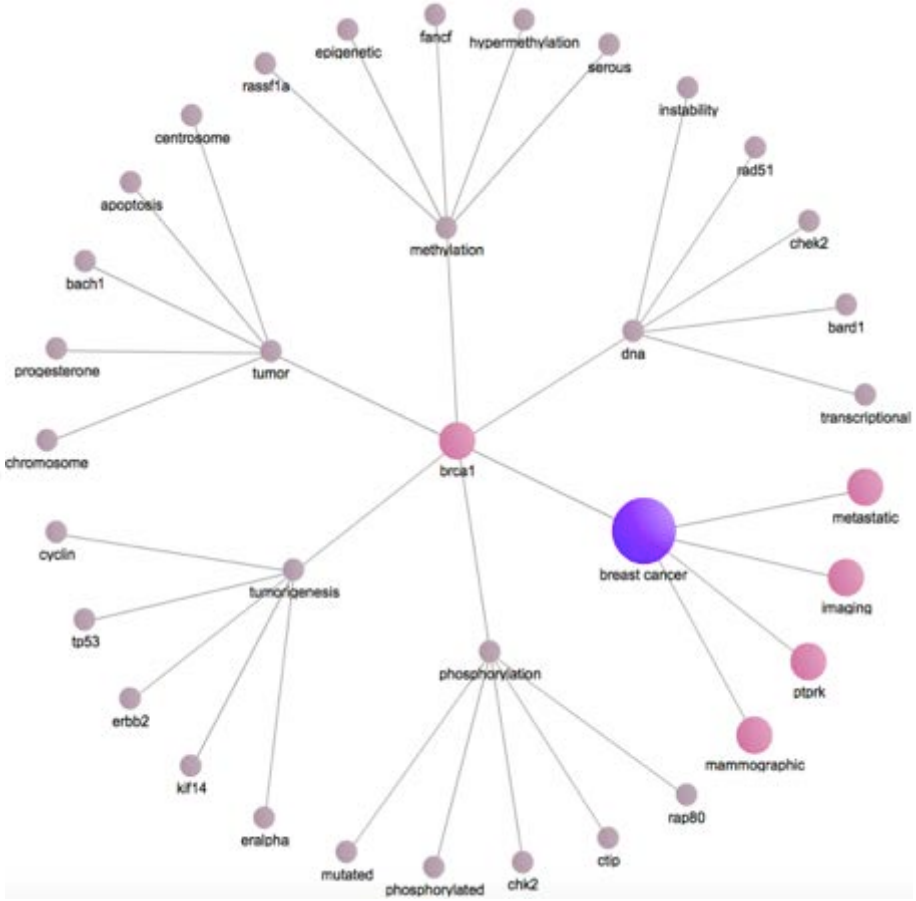


Main approaches for data integration

- Pathway-based integration
  - Pathway information from KEGG or other databases
  - 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
  - Interpretation using pathway analysis tools
  - 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

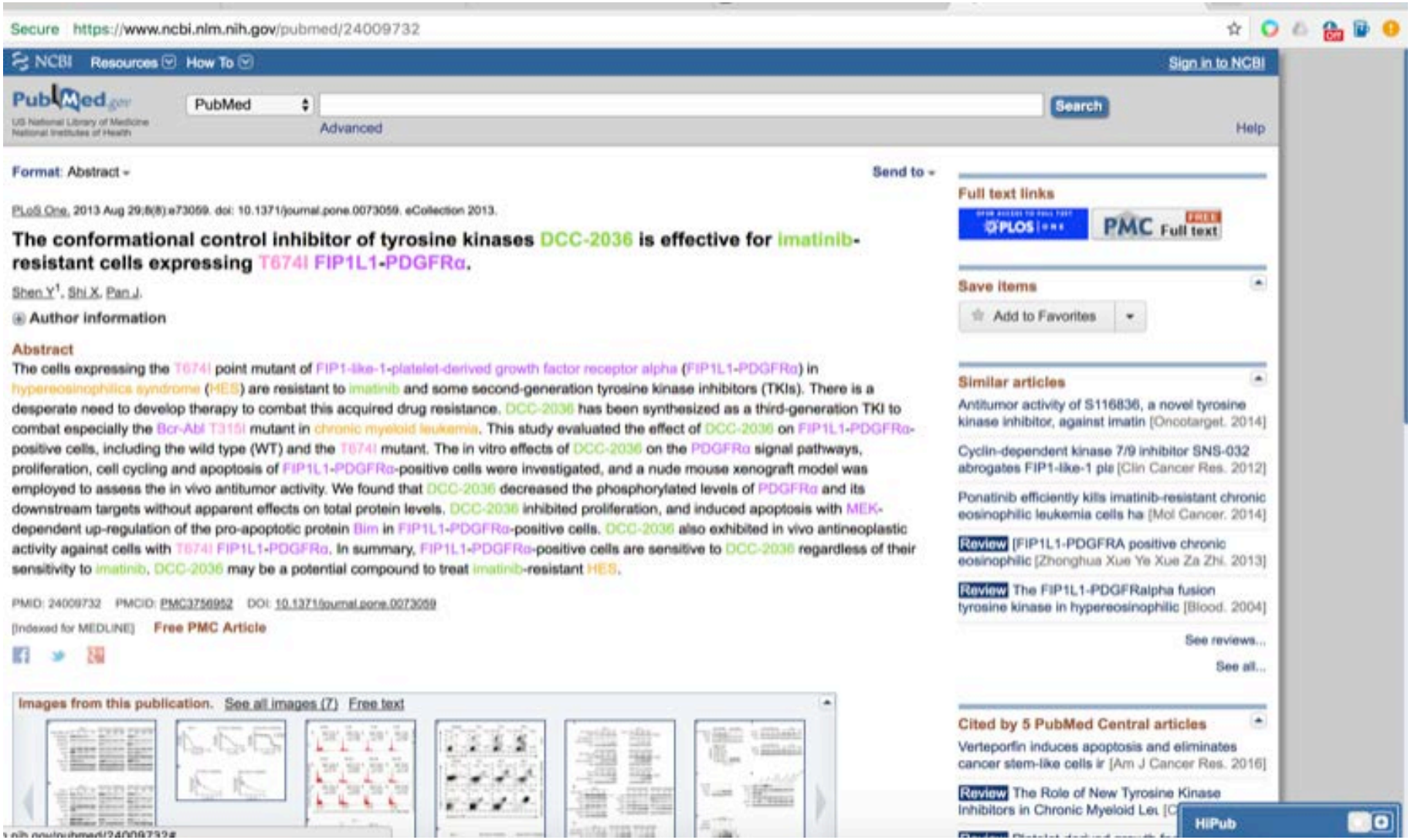




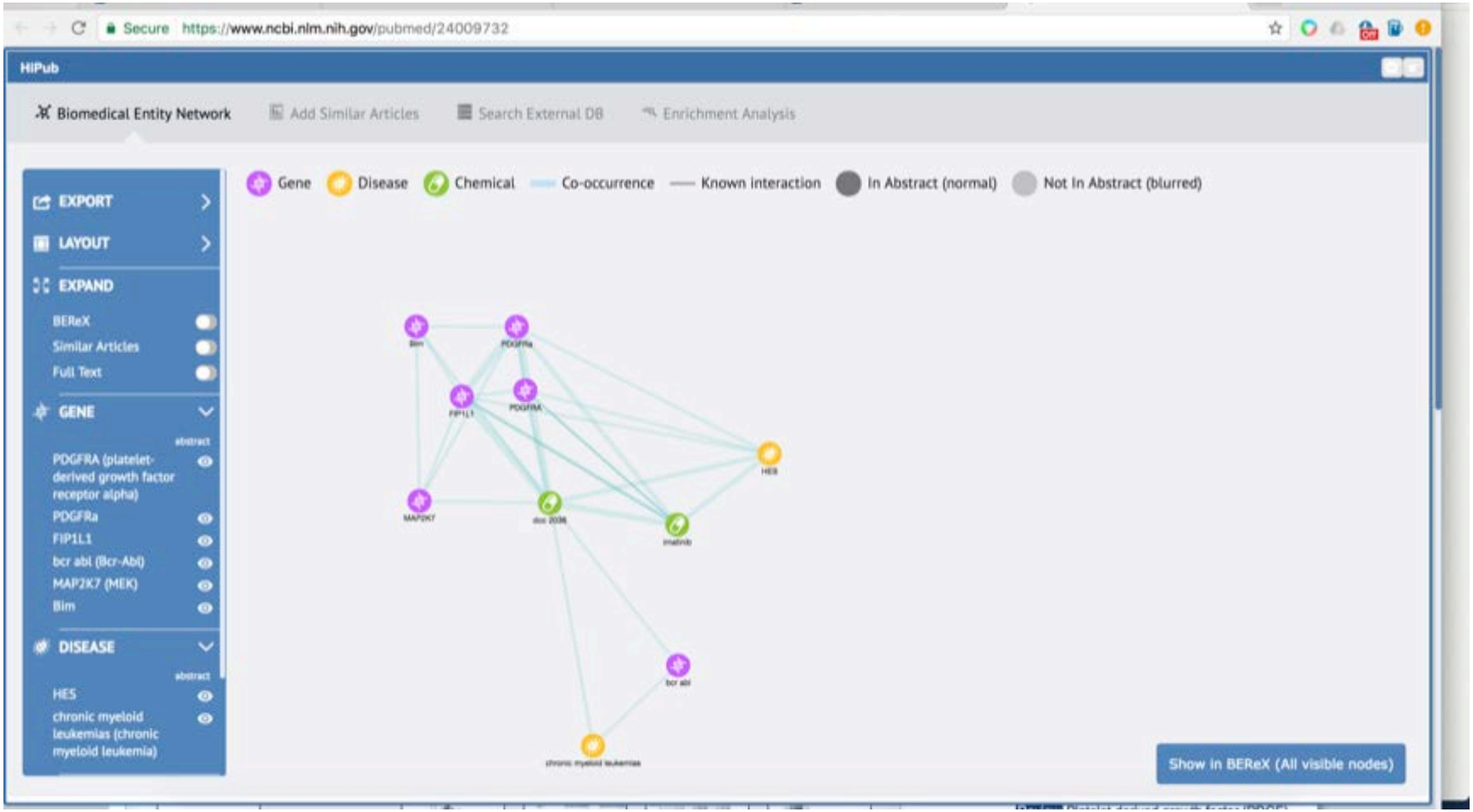
Association mining based on co-occurrence

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

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



HiPub (Lee 2016):



Case Study: Integrative analysis of platelet metabolome with mitochondrial bioenergetics

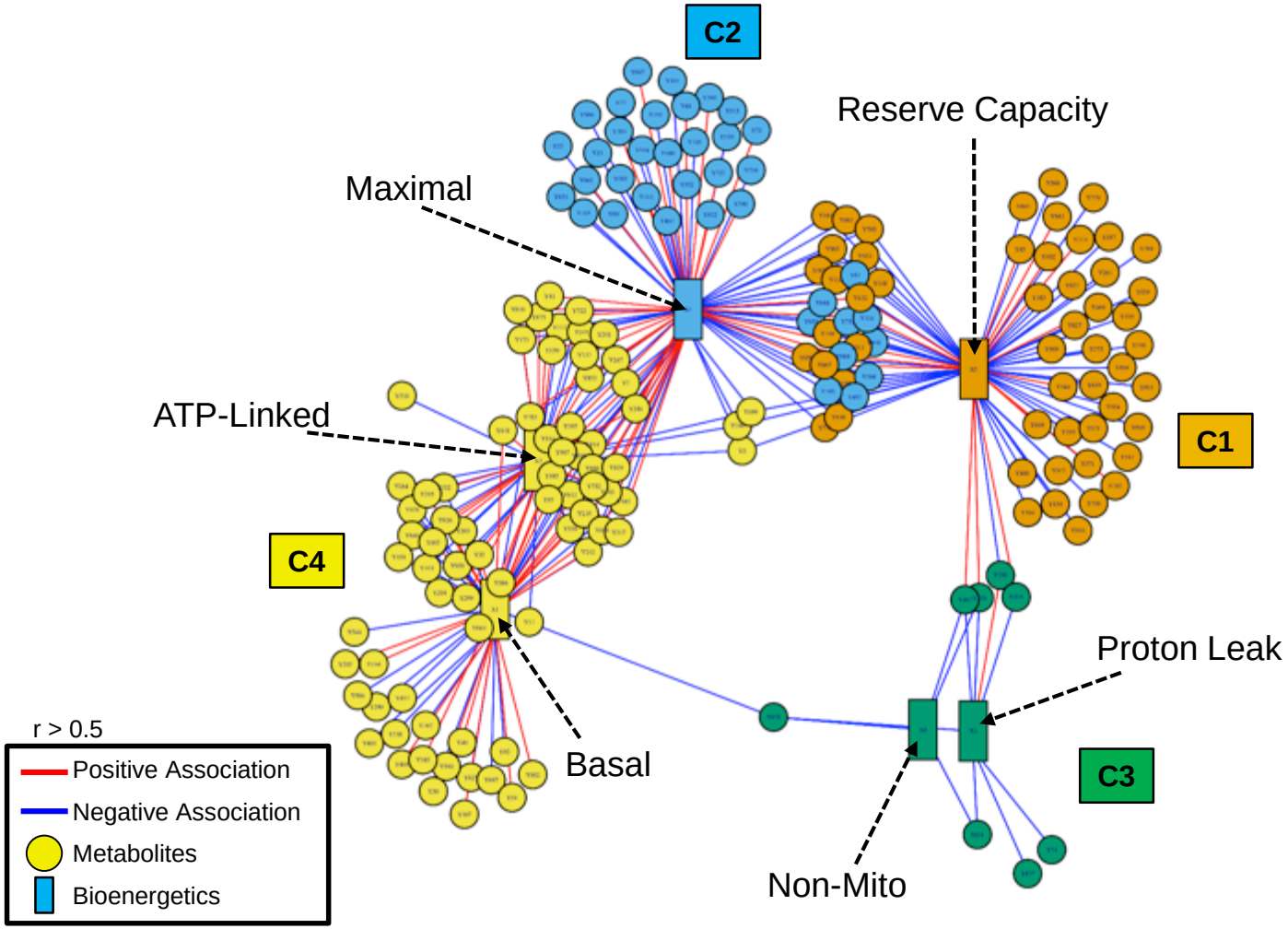
N=13 healthy volunteers

Mitochondrial bioenergetics: 6 energetic parameters (ATP-linked, basal, proton leak, maximal, non-mitochondrial, and reserve capacity OCR)

High-resolution metabolomics: 2,705 metabolic features

Chacko BK, Smith MR, Johnson MS, Benavides G, Culp ML, Pilli J, Shiva S, Uppal K, Go YM, Jones DP, Darley-Usmar VM. Mitochondria in precision medicine; linking bioenergetics and metabolomics in platelets. **Redox Biol. 2019**

Collaboration between Emory and UAB


C2
Reserve Capacity
Maximal
ATP-Linked
C1
C4
Proton Leak
Basal
C3
Non-Mito
r > 0.5
Positive Association
Negative Association
Metabolites
Bioenergetics

Case Study: Application of xMWAS for integrative network analysis of metabolome and metallome datasets from the Strong Heart study

N=145 (12 American Indian communities; free of type-2 diabetes)
Metallome: urinary concentrations of 12 metals
Metabolome: High-resolution metabolomics data for 8,810 features

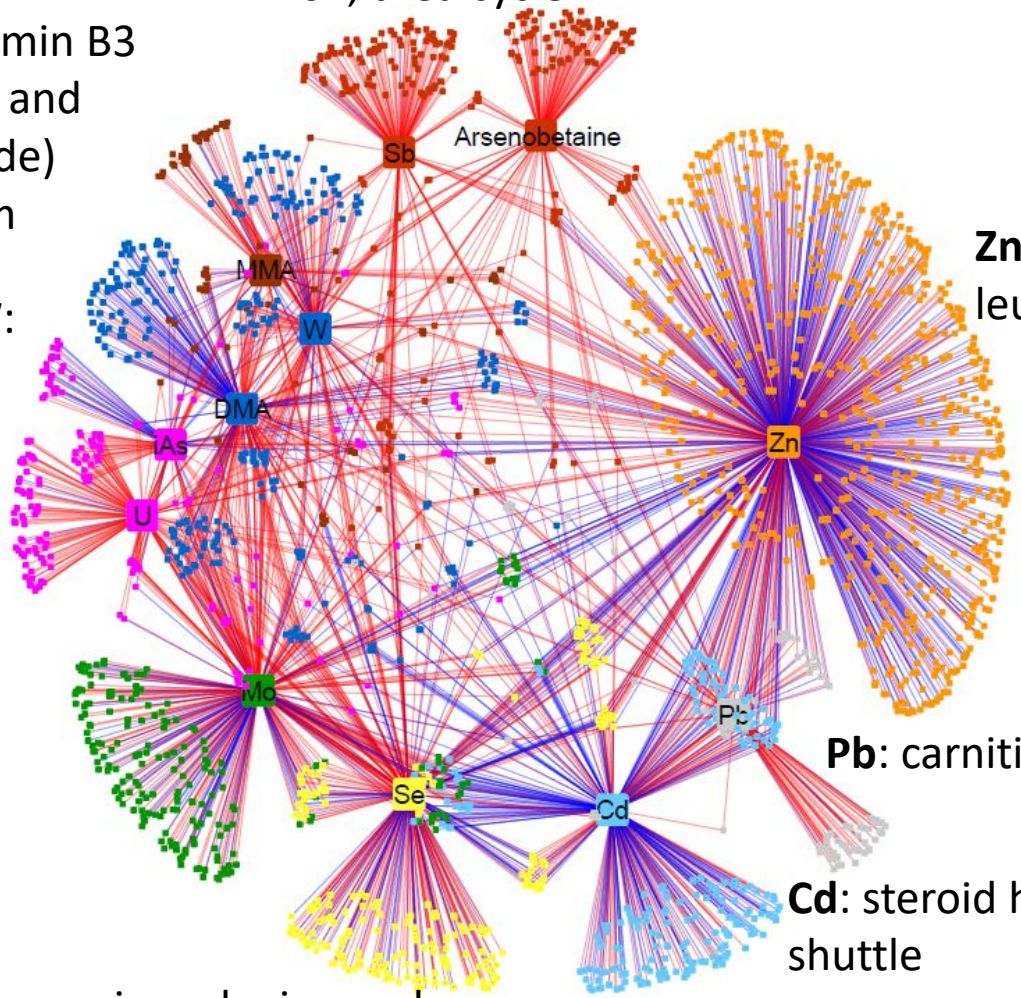
Collaboration between Emory and Columbia Universities

Metallome-metabolome integrative network using xMWAS

AsBe and Sb: Asp and Asn; urea cycle

MMA: Vitamin B3 (nicotinate and nicotinamide) metabolism

DMA and W: Arg and Pro; linoleate; Ala and Asp



Zn: bile acid; prostaglandin; leukotriene; glycosphingolipid

Mo: Linoleate metabolism

Pb: carnitine shuttle

Cd: steroid hormone; carnitine shuttle

Se: Glycine, serine, alanine and threonine metabolism

Case Study: Integrative network analysis of clinical, biomolecular (metabolites, microRNAs, plasma protein markers, and cytokines), and environmental exposure data from a dataset of 66 service personnel post-deployment

Collaboration between Emory, Rochester, and Department of Defense

Input data for xMWAS

1. Molecular data: metabolites, miRNAs, cytokines, and proteins (3,274 molecular variables x 66 subjects)

	Subject1	Subject2	-	Subject N
Metabolite 1	199	19	-	100
-	-	-		-
miRNA 1	50	30	-	20
-	-	-		-
Cytokine 1	33	12	-	39

2. Environmental chemicals (5 variables x 66 subjects)

	Subject1	Subject2	-	Subject N
chemical 1	3	2.5	-	13
chemical 2	1	4	-	9
-	-	-		-
chemical s	5	3	-	2

ICD-9; (49 any cardiopulmonary ICD-9 codes x 66 subjects)

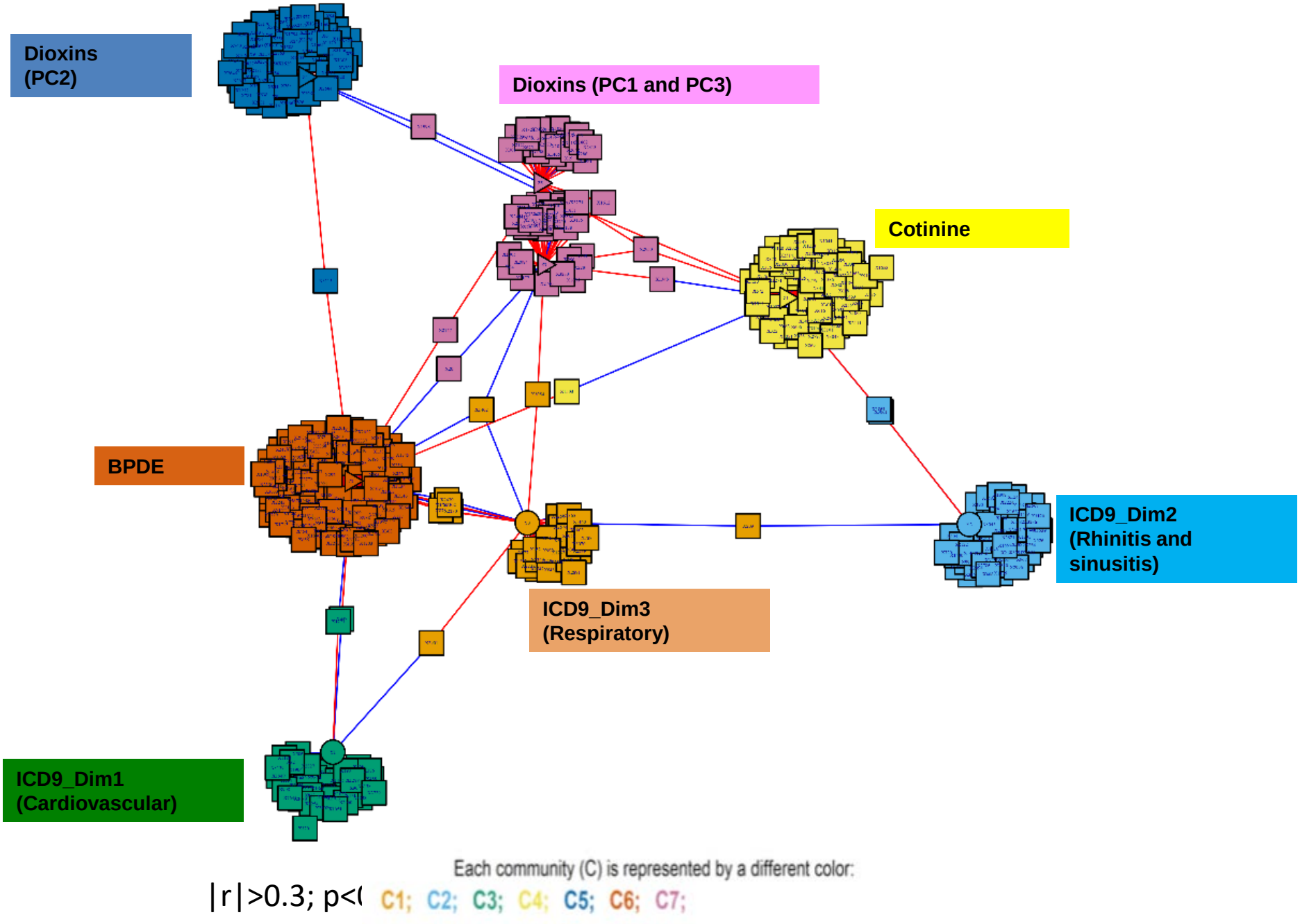
	Subject1	Subject2	-	Subject N
4019	0	1	-	0
4011	1	1	-	0
-	-	-	-	-
49301	1	0	-	0

→

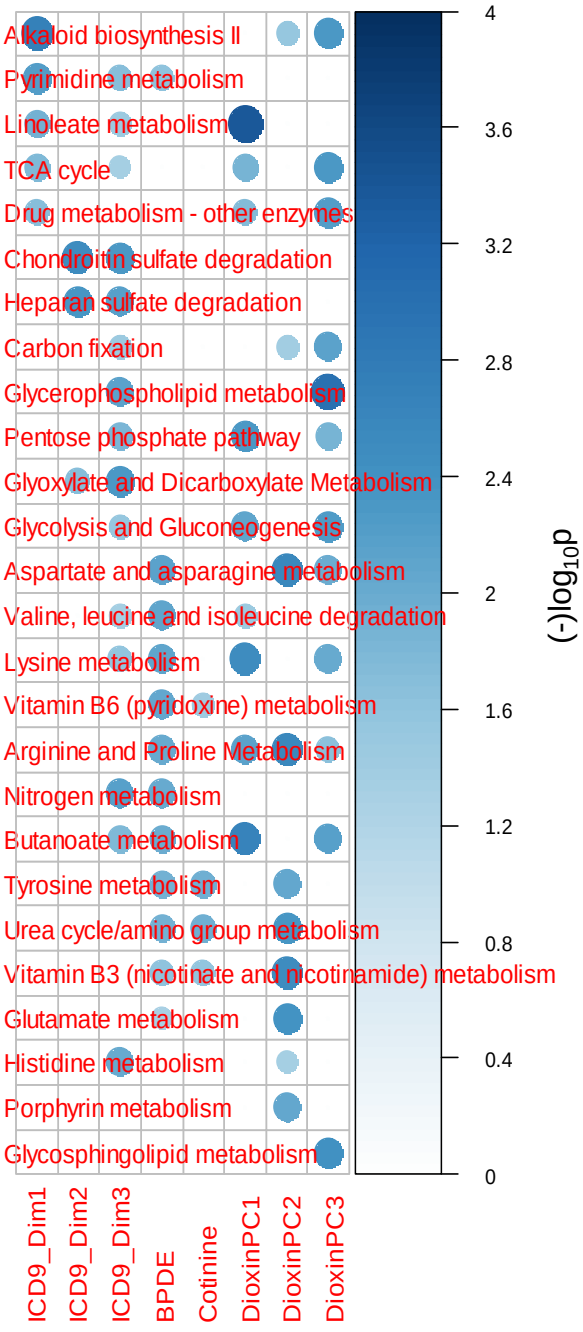
3. Multiple correspondence analysis (8 dimensions x 66 subjects)

(Only dimensions with >5% variance explained were included)

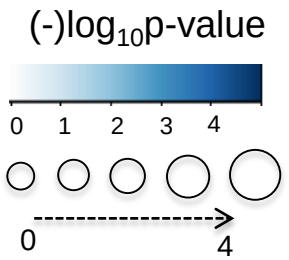
	Subject1	Subject2	-	Subject N
Dim 1	199	19	-	100
Dim 2	10	40		90
-	-	-		-
Dim r	50	30	-	20


Dioxins (PC2)
Dioxins (PC1 and PC3)
Cotinine
BPDE
ICD9_Dim2 (Rhinitis and sinusitis)
ICD9_Dim3 (Respiratory)
ICD9_Dim1 (Cardiovascular)
Each community (C) is represented by a different color:
C1; C2; C3; C4; C5; C6; C7;

|r|>0.3; p<(

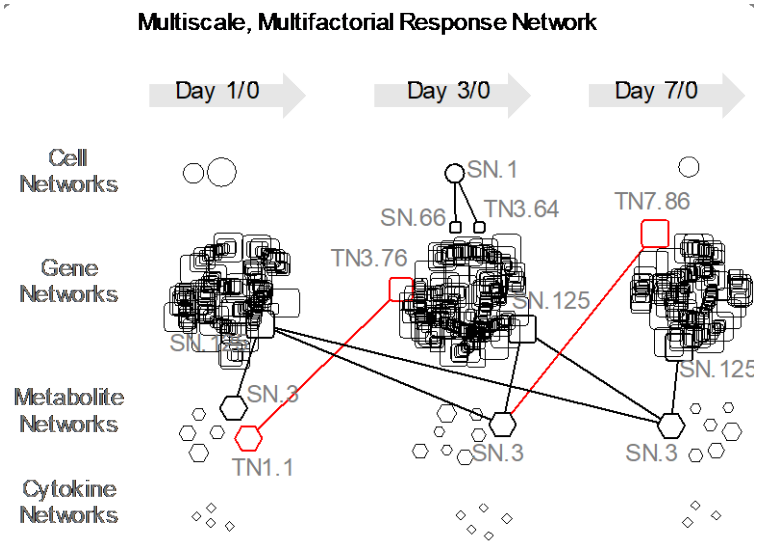


- Bubble plot showing metabolic pathways associated with clinical and environmental exposures data
- Metabolic pathway analysis performed using Mummichog



Current challenges and future work

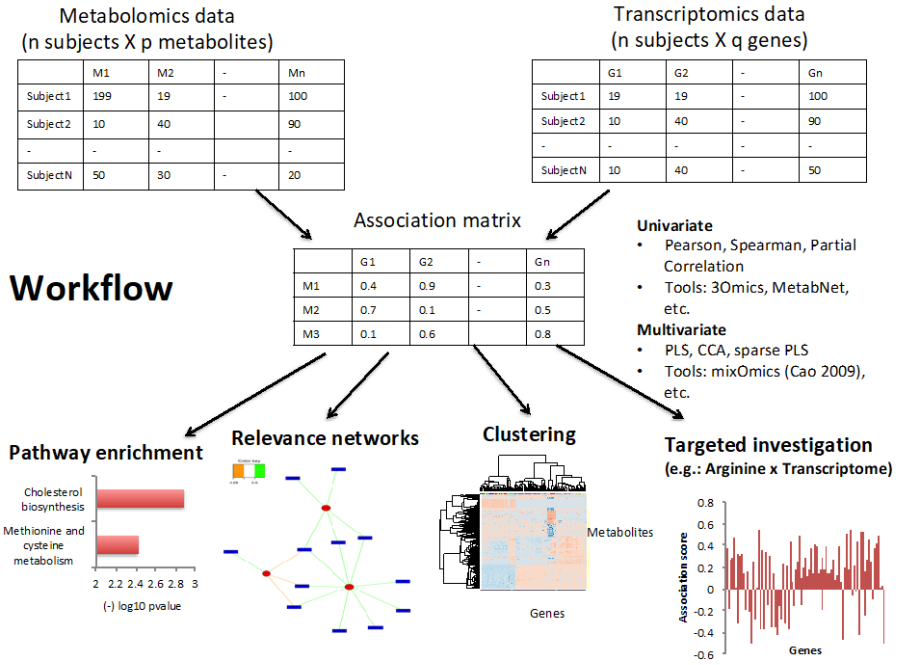
- Development of hybrid methods
  - combined knowledge-based and data-driven approaches
  - incorporation of literature-derived associations in xMWAS
    - Using co-occurrence criteria for establishing relationship (PolySearch2.0)
- Improving scalability
  - Ability to handle >100,000 variables
  - Performing integrative analysis at communities, clusters, or eigenvariables (first PCs) level



Li et al., 2017. *Cell* 169, 862–877

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



Acknowledgements

Emory

Dean Jones, Young-Mi Go, Shuzhao Li, Chunyu Ma, Ken Liu, Kristine Dennis, ViLinh Tran, Michael Orr, Ryan Smith, Xin Hu, Jolyn Fernandes, Bill Liang, Yating Wang

Collaborators:
Victor M. Darley-Usmar, Ana Navas-Acien, Tiffany R. Sanchez, Nancy Loiacono, Jinying Zhao, Timothy Mallon, Mark Utell, Juilee Thakar, Gary Miller, Douglas Walker, Milam Brantley, Ihab Hajjar, Arshed Quyyumi, John Roback, MoTrPAC, and others

Funding
NIEHS, NIA, NCI, NHLBI, NIDDK, NIAAA, NIAID, Woodruff Foundation, Emory Dept of Medicine, Georgia Research Alliance Exposome Center
Integrated Health Science and Facilities Core
NIEHS P30 ES019776
DK112341 (MoTrPAC)
AG057470
AA026928
EY022618
ES026071
DK117246-01

Questions?

Email: kuppal2@emory.edu

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

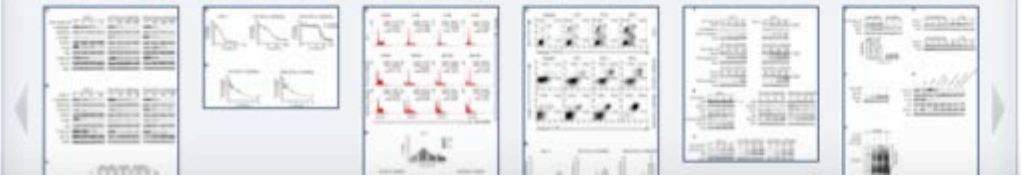
Abstract

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

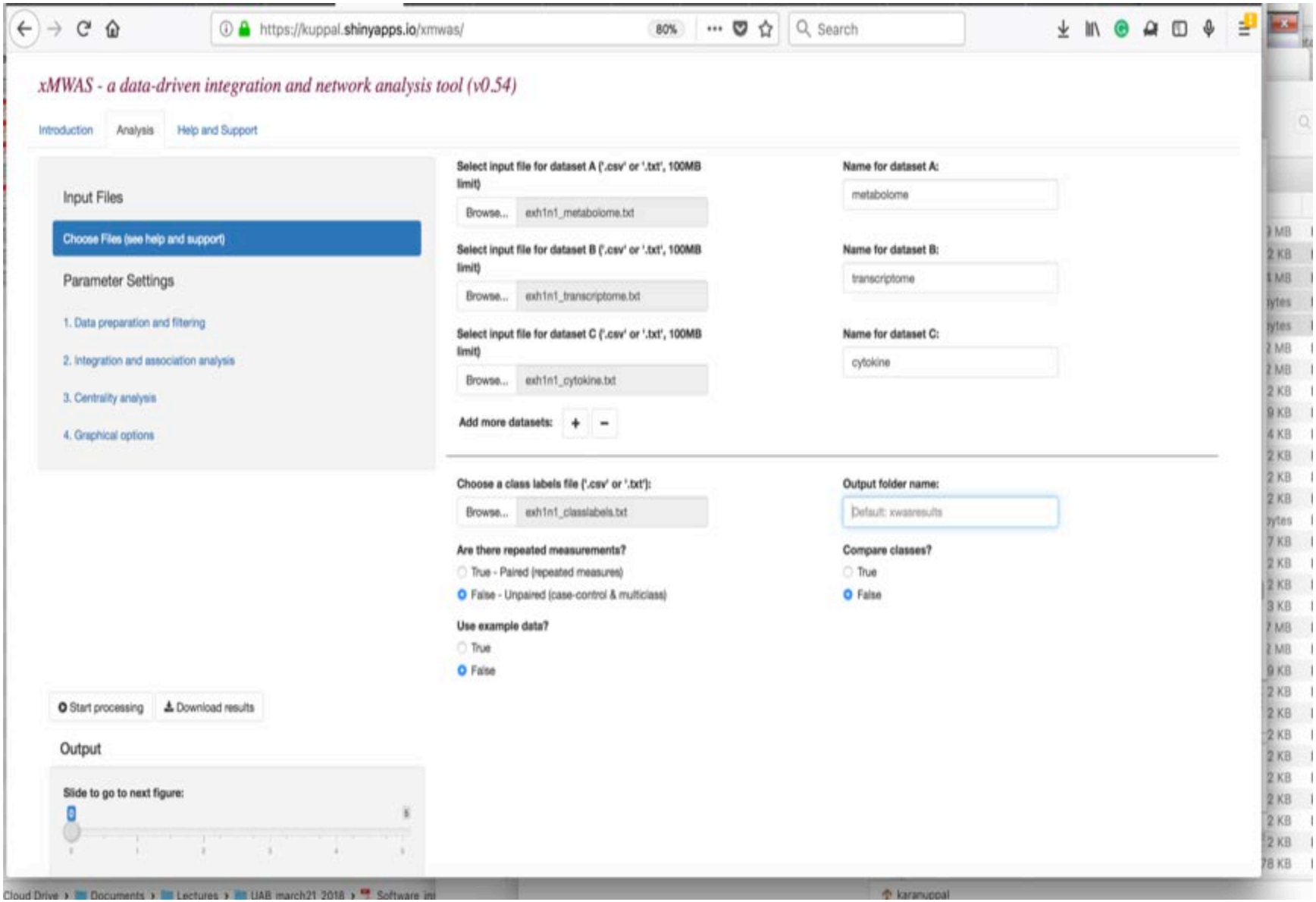
Verteporfin induces apoptosis and eliminates cancer stem-like cells ir [Am J Cancer Res. 2016]

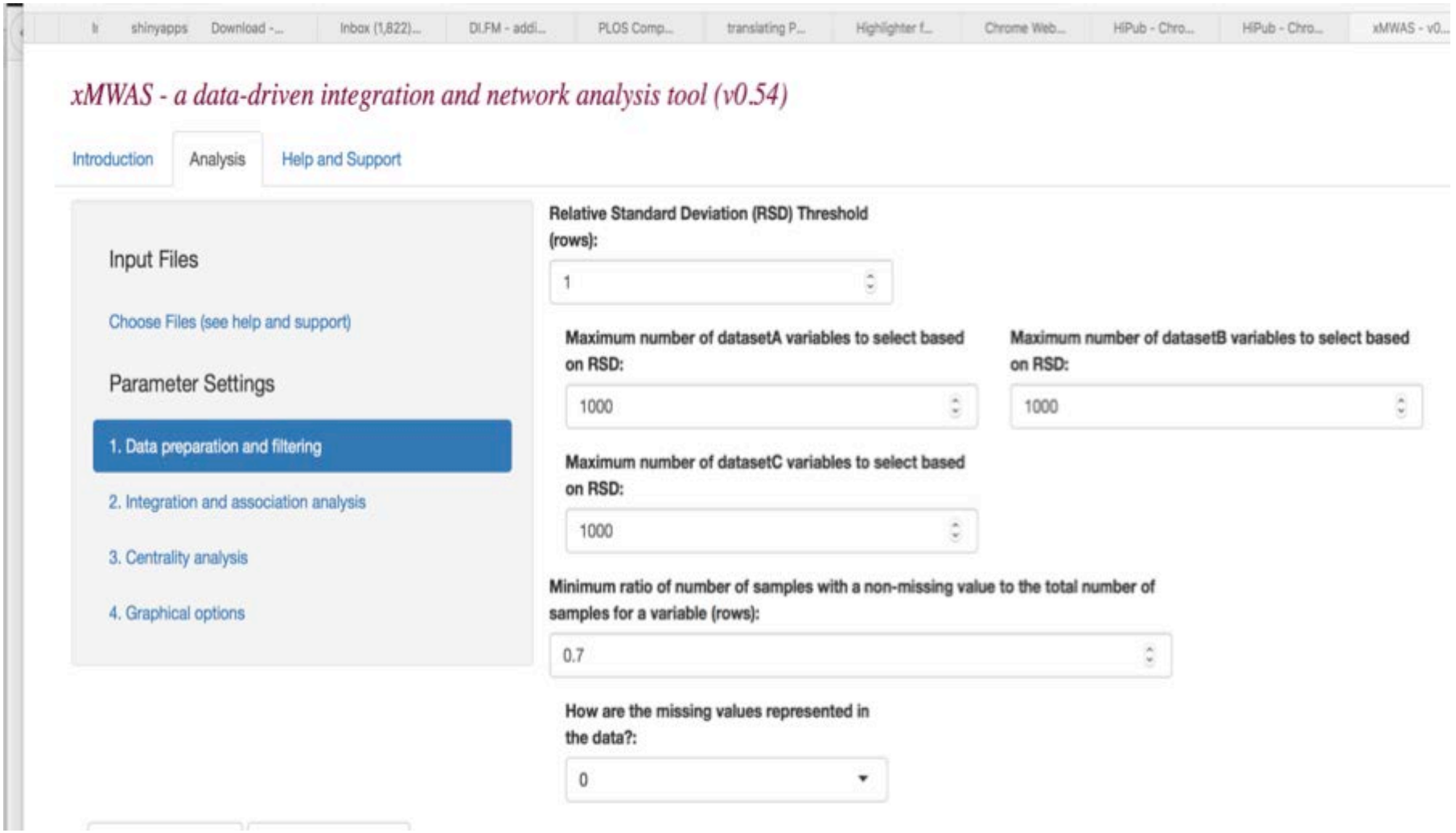
Review The Role of New Tyrosine Kinase Inhibitors in Chronic Myeloid Leu [C

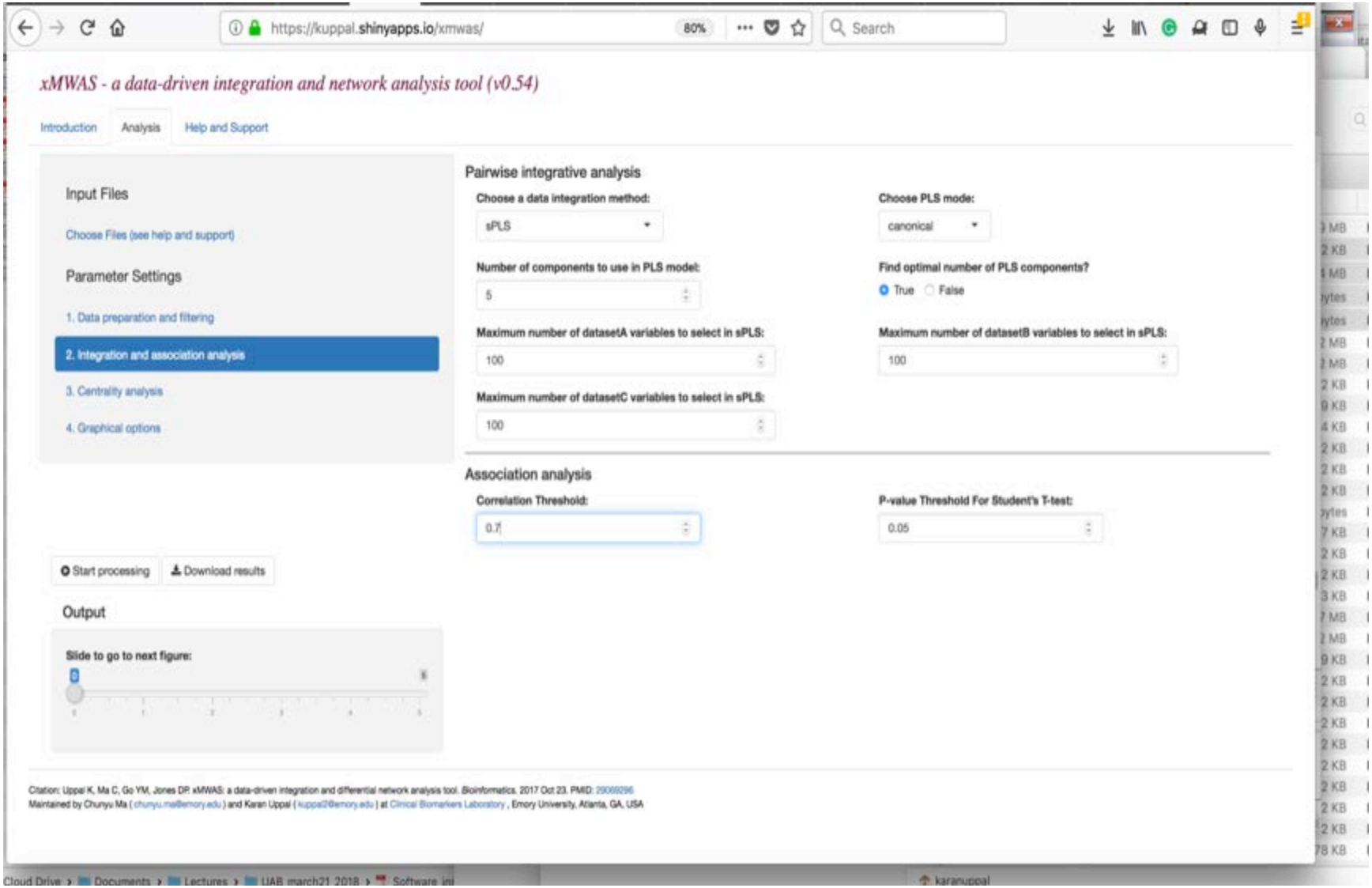
HiPub

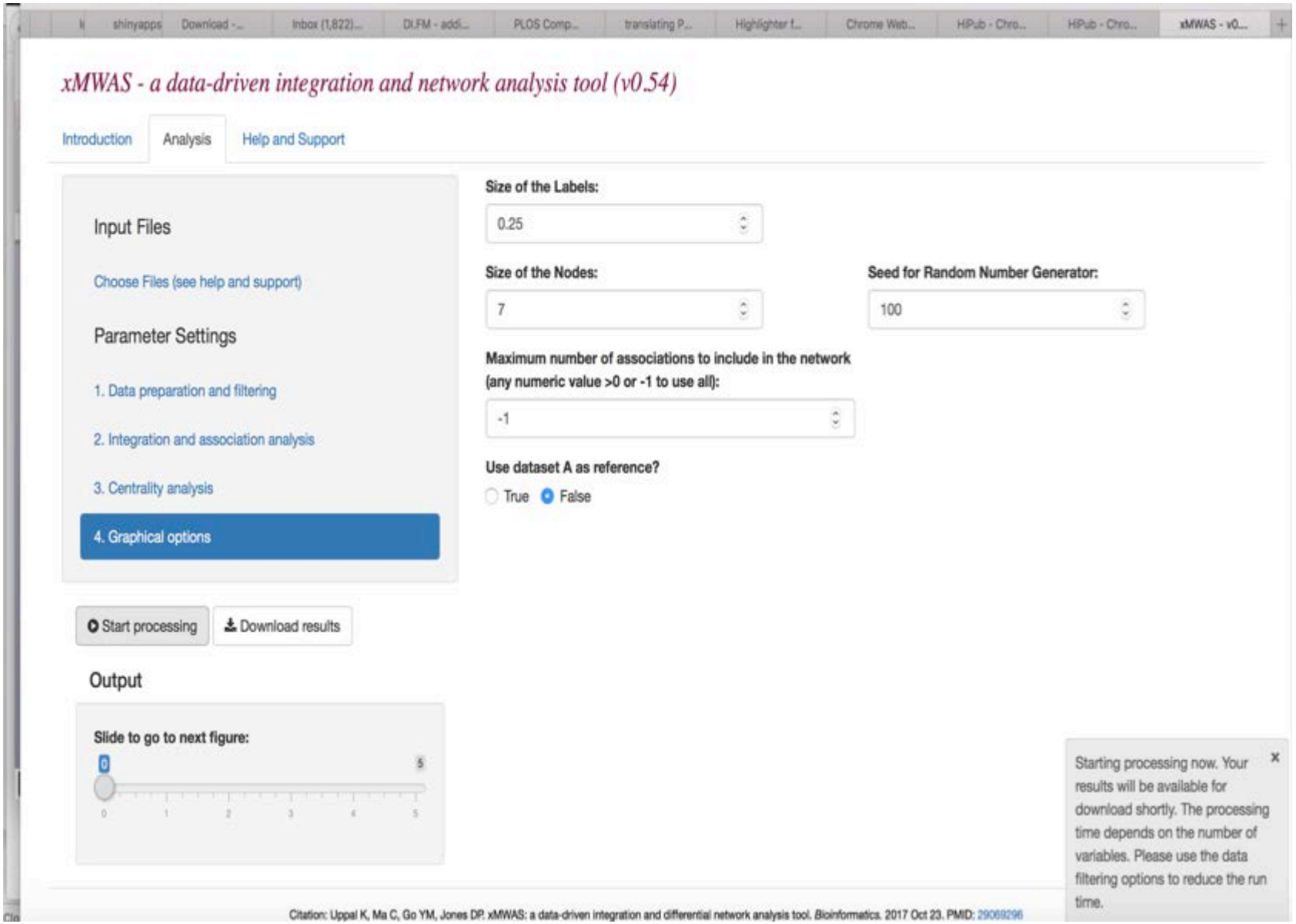
2. xMWAS: Web version

- 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

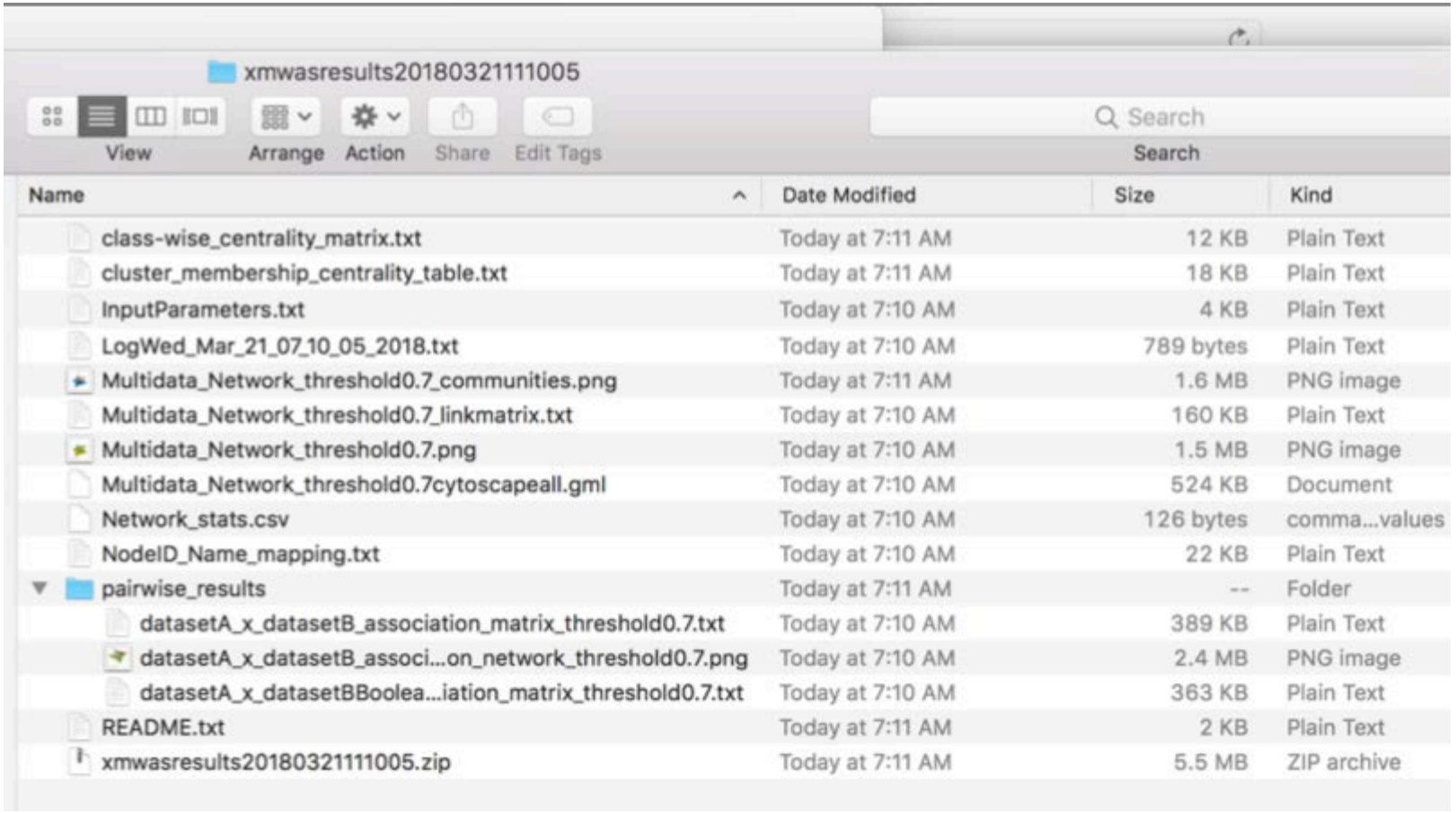

https://kuppal.shinyapps.io/xmwas/
80%
Search
xMWAS - a data-driven integration and network analysis tool (v0.54)
Introduction
Analysis
Help and Support
Input Files
Choose Files (see help and support)
Parameter Settings
1. Data preparation and filtering
2. Integration and association analysis
3. Centrality 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:
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)...
DLFM - 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
(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)

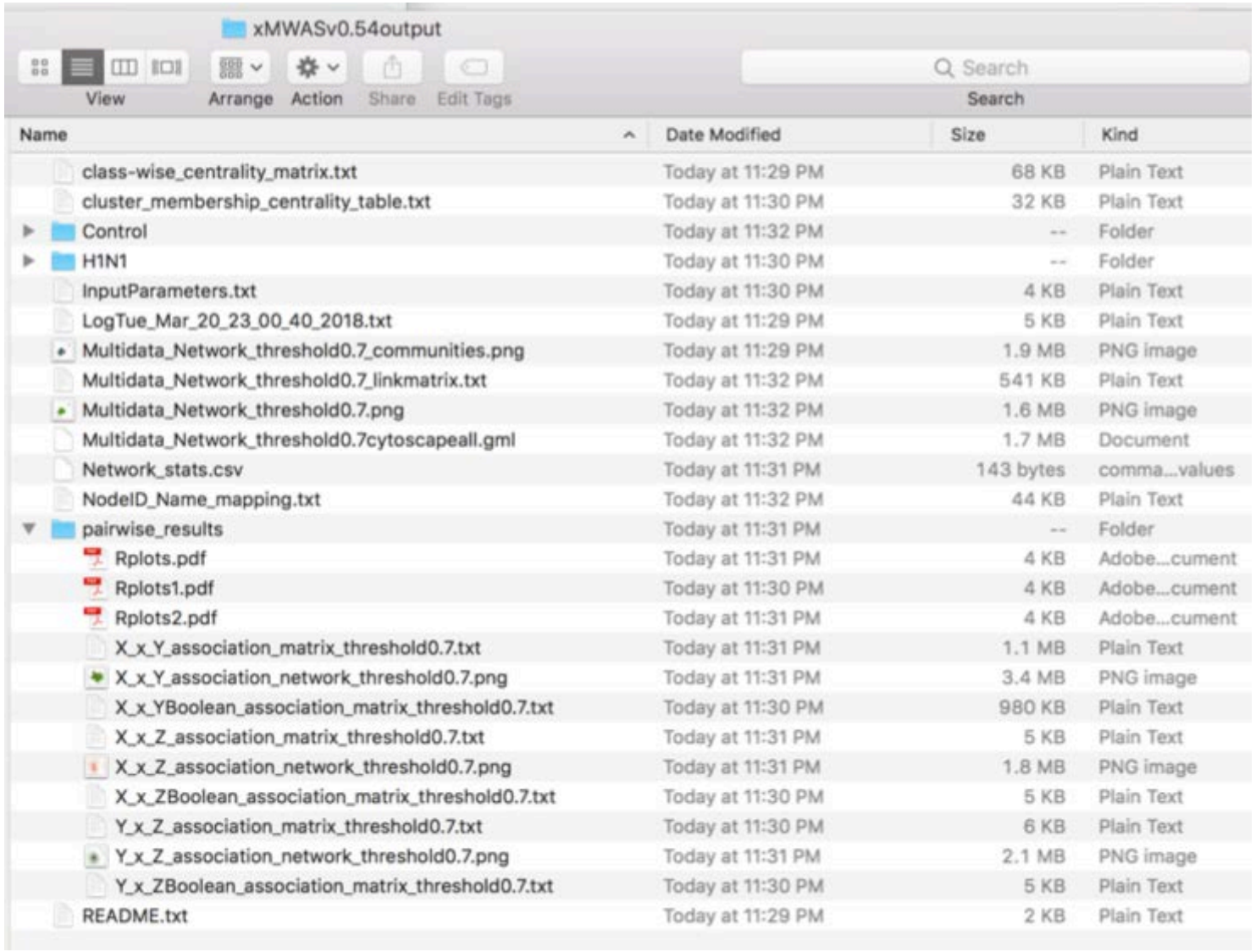
```
#load package
library(xMWAS)

#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
```

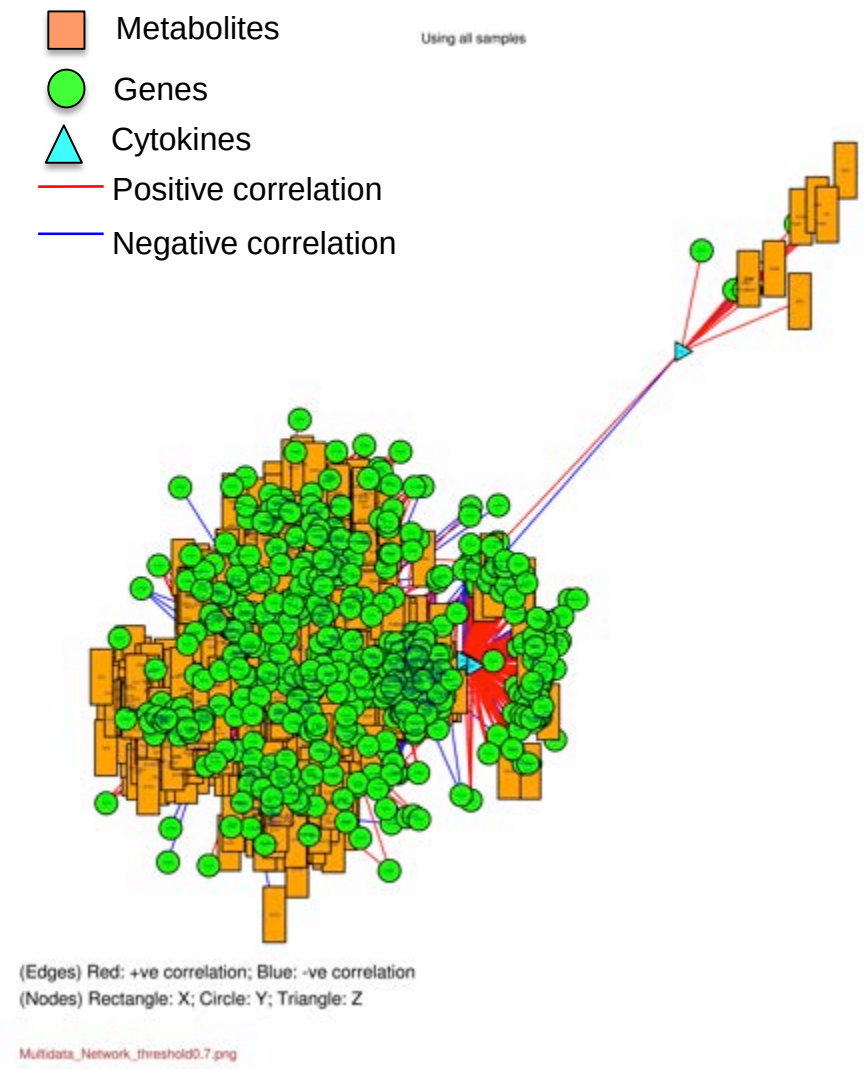
Description of output: Readme.txt

README.txt

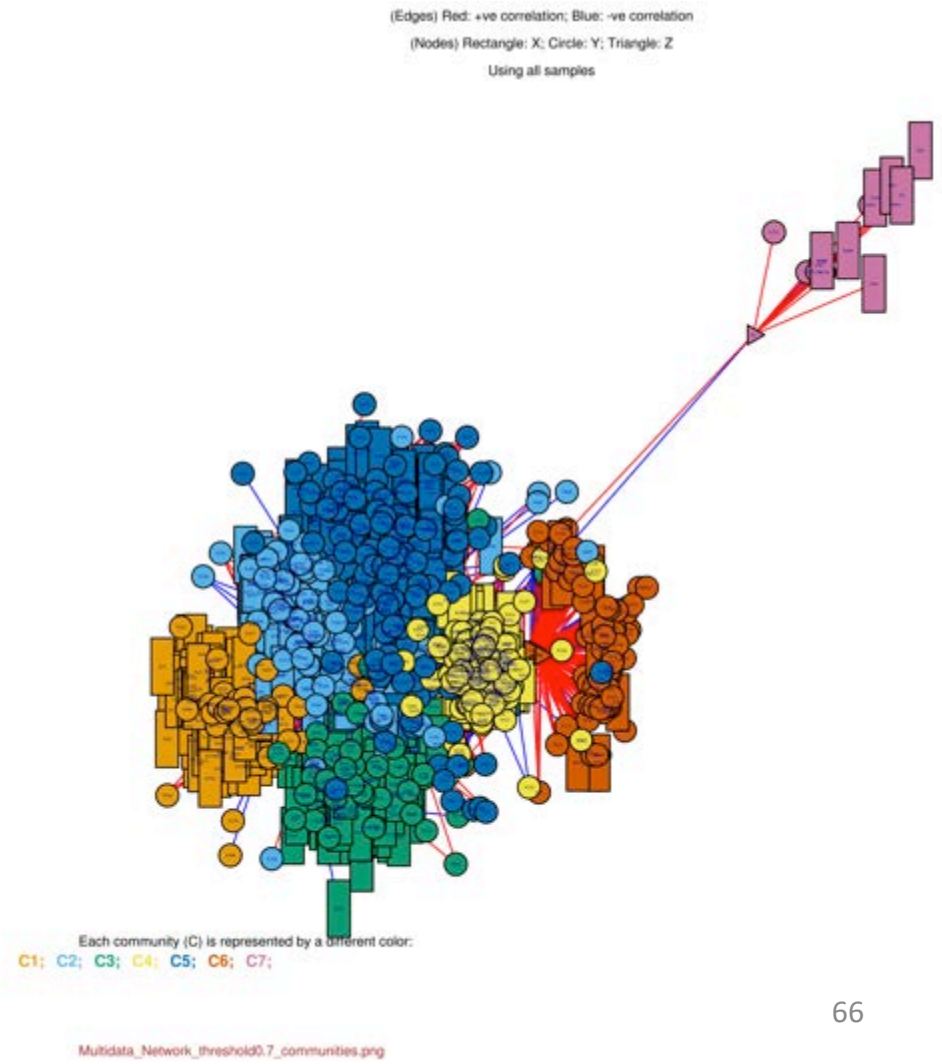
"Description of files"
"1: X labels correspond to xome_fname data, Y labels correspond to yome_fname data, Z labels correspond to zome_fname data, W labels correspond to wome_fname data"
"2: Pairwise integrative analysis results are under pairwise_results. The files corresponding to each pairwise comparison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

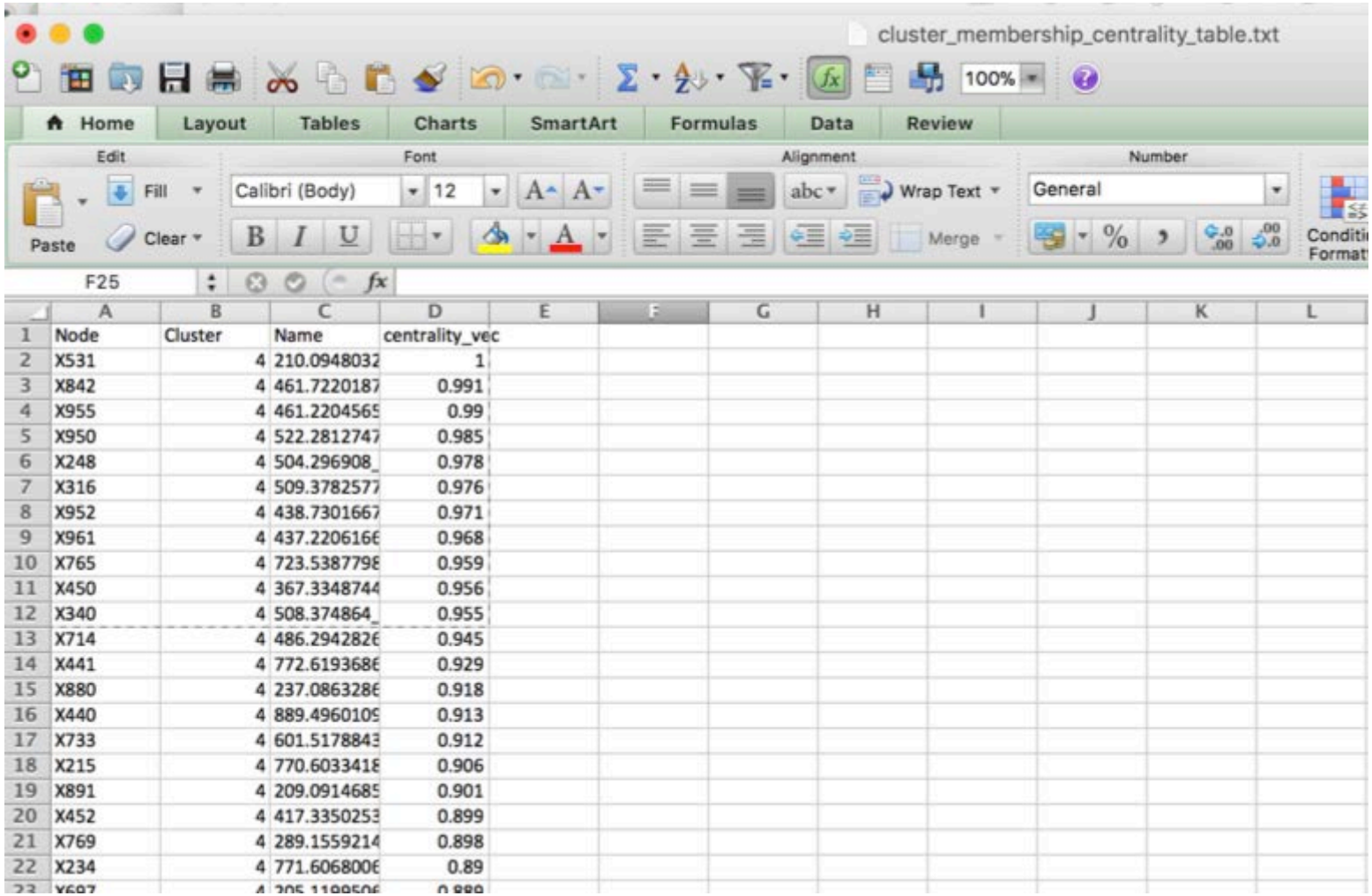
A. Colored by data type



B. Colored by community membership

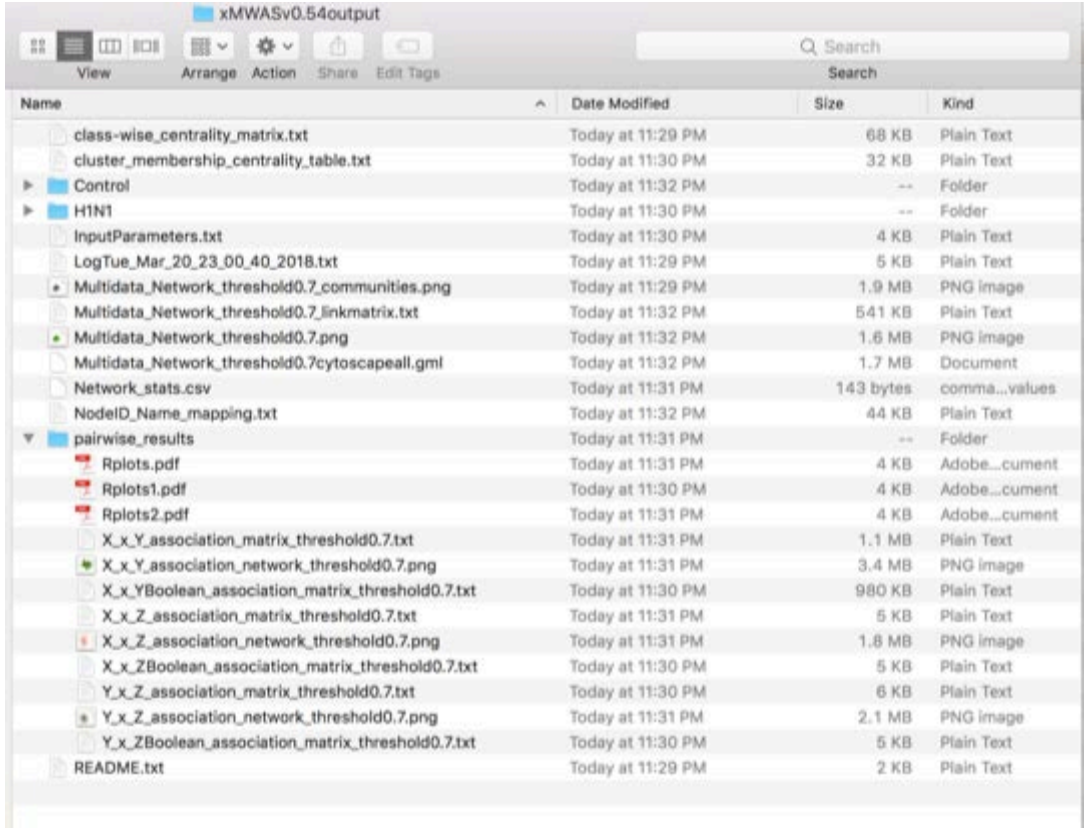


Community detection and centrality analysis



	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



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