From this page the supplemental Material begins

OMICS: A Journal For Integrative Biology (2013) in press

SUPPLEMENTAL TUTORIAL

Graphical identification of cancer-associated gene subnetworks based on small proteomics data sets Tutorial

Karim Mezhoud

UR04CNSTN01- Bio-computing Unit, Life Science Department, National Center for Nuclear

Sciences and Technologies. Sidi Thabet BioTech Park, 2020, Ariana, Tunisia.

Email: Karim mezhoud: kmezhoud@bioinformatics.tn

This tutorial describes the main steps used to construct an enriched network from a list of genes with their expression values. The network includes the fold change, the False discovery rate (FDR), the gene expression correlation, the most often involved biological processes, and the most likely corresponding diseases. All the main steps are illustrated by screen shots. The "Tools & Data" section lists all the tools and data required. The "Data computing" section shows you how to compute the data. The Experimental procedure section lists briefly all the steps, and the Walk-through example section illustrates these steps. The final section lists all the software and packages that were tested during this study, but not selected for inclusion in this Note.

a video version (106 Mo) of this tutorial is available on supplemental material at

https://wiki.ubuntu.com/kmezhoud/supplemental_material .

Tools & Data

For this tutorial you will need:

1. Bioinformatic tools

   Cytoscape (Shannon et al., 2003; Christmas et al., 2005; Cline et al., 2007) software implemented by Reactome FIs (Wu et al., 2010), Advanced Network merge, NetworkAnalyzer, and ClueGO (Bindea et al., 2009) plug-ins.

2. **Experimental Data**

   To find out more about the file format, download the corresponding supplemental file indicated between parentheses.

   1. The list of modulated protein saved in one column in text file (only_modulated_symbol.txt). Convert the protein ID on gene symbol (HGNC or HUGO symbol) using ID Mapping service, such as:

      - BioMart (http://www.ensembl.org/info/data/biomart.html)
      - Uniprot (http://www.uniprot.org/)
      - Idconverter (Alibés et al., 2007) (http://idconverter.bioinfo.cnio.es/)
      - You can use other CRAN package as GAGE (Luo et al., 2009) with the "eg2sym"command.

   2. The protein expression values. The following example of gene expression is divided into two groups and 3 replicates. (proteinEXP.txt)

Node Attributes

      - Fold change of the protein identified (Fold_Change.NA)
      - False discovery rate of the z score or some other statistical test (FDR.NA)
      - Correlated protein expression by statistical analysis (Correlation_Network.sif)
      - Cancer-protein network (Cancer_Network.sif)

   3. Edge attributes

      - The correlation rate of Protein expression (Correlation+Cancer.EA)
      - The gene expression profile overlap between our experiments and clinical data from the cBio Cancer Genomics Portal (Cerami et al., 2012) (http://www.cbioportal.org/) (Correlation+Cancer.EA)

Data Computing

1. Gene list

We assume that the protein list with quantitative analysis has already been compiled . Supplemental **proteinEXP.txt** file gives an example of the quantitative data which we will be using.

Pre-process the expression data into a format suitable for network analysis, clean the data by removing obvious outlier samples or excessive numbers of missing entries, and

then load the data. The gene list is selected after statistical analysis.

The Fold change and False discovery rate are then computed and saved in the corresponding files: F**old_Change.NA** and **FDR.NA**.

2. **Multiple Correlation analysis of protein expression**

Two genes are said to be connected by an edge if their expression values are closely correlated (statistical significance).

We used CORREP package (Zhu et al., 2007) to estimate the gene expression correlation. See the bioconductor page (http://www.bioconductor.org/packages/release/bioc/html/CORREP.html) or see below for how to use the CORREP package.

This program assumes that data are arranged with the variables (genes) in rows and conditions (persons) in columns . It is strongly suggested that factor names be included in the data file, as this information is used in the results output.

We specified that the proteins (or genes) are the variables, and the people correspond to the different conditions. In our case we have 235 variables, 6 conditions, and one replicate. It is not possible to carry out multiple correlation analysis using only two conditions and 3 replicates.

Following R and shell codes makes it possible to compute and extract the correlation rate between gene expression levels.

```
###CORREP install###

R          #start R

source("http://bioconductor.org/biocLite.R")

biocLite("CORREP")

library(CORREP)

####Data computing#####

proteinEXP <- read.table("proteinEXP", header=T, row.name=1) # import
expression data

proteinEXP.std <- apply(proteinEXP,1, function(x) x/sd(x))    #
standardization of data

cor.balance(t(proteinEXP.std), m=1, G=235)                  # run CORREP

cor.proteinEXP.std <- cor.balance(t(proteinEXP.std), m=1, G=235) #
save matrix as "cor.proteinEXP.std"

write.table(cor.proteinEXP.std.X, file="cor.proteinEXP.std") # write
matrix as table named "cor.proteinEXP.std"

####filter the minimal threshold of correlation for N=6## r >0.811 (p
0.05)

sed /-?0\.[0-7][0-9]*/_/ cor.proteinEXP.std

sed /-?0\.80[0-9]*/_/ cor.proteinEXP.std.

sed /-?0.810[0-9]*/_/ cor.proteinEXP.std >
cor.proteinEXP.std.X.filter

sed /^\d\d?\d?\t// cor.proteinEXP.std.filter

####grouping gene with r as gene1~cor~gene2

sed s/^([^\t]+)([^#\n]*)#([^0\n]*)(0\.[0-9]*)/\1\2\3\1~\4#/g
cor.proteinEXP.std.filter

sed s/#// cor.proteinEXP.std.filter

R> cor.proteinEXP.std.filter <-
read.table("cor.proteinEXP.std.filter", header= F)

R> tcor.proteinEXP.std.filter <- t(cor.proteinEXP.std.filter)

R> write.table(tcor.proteinEXP.std.filter,
file="tcor.proteinEXP.std.filter")

sed s/^([^\t]*)([^A-Z]*)([A-Z0-9]*~-?0\.[0-9]*)/\1\2\3~\1#/
tcor.proteinEXP.std.filter

sed s/^([^\t]+)([^#\n]*)#([^0\n]*)([A-Z0-9]*~-?0\.[0-
9]*)/\1\2\3\4~\1#/g tcor.proteinEXP.std.filter

sed s/#// tcor.proteinEXP.std.filter

sed s/\t1\t/\t_\t/g tcor.proteinEXP.std.filter

sed s/_\t//g tcor.proteinEXP.std.filter

sed s/_\n// tcor.proteinEXP.std.filter

sed s/\t/\n/ tcor.proteinEXP.std.filter

sed s/~/\t/ tcor.proteinEXP.std.filter
```

Use OpenOffice Calc to convert the tcor.proteinEXP.std.filter file to:

gene1+%28pp%29+gene2 = correlation value

…..

Save the table as a text file for Edge attribute (**Correlation+Cancer.EA).**

Replace "+%28pp%29+" by " pp " and save it as **Correlation_Network.sif** file for the correlated protein.

3. Cancer-protein Network Construction: overlapping gene expression between experimental data and cancer cases

This network adds the overlapping of gene expression between our experimental data and clinical data from the cBio Cancer genomic portal.

The gene list was submitted to the cBio Cancer genomic portal   (Cerami et al., 2012) (http://www.cbioportal.org/public-portal/index.do) with the "all cancer studies" option. This first request told us how much the genes were modulated in twenty cancer studies **(supplemental Sheet 1 in Edges_Disease.xls file)**. The 20 cancers were then ranked in decreasing order of the overlap of their gene expression profiles (**Cancer-frequency.NA**). In the second step, the search was refined by specifying the cancers (one by one), which involved a wider range of gene lists. We used only the "mRNA Expression Z-Scores vs Normals" option. We specified the direction of the modulation using Advanced: Onco Query Language (OQL). For example, with PZP: EXP > 1, the search selects only cancer cases that display up-regulation of PZP. Based on the Fold_Change.NA file, we used OQL as explained in supplemental **only-modulatedEXP.txt**.

Figure 2: Example of a cBio Cancer genomics portal output. Only-modulatedEXP.txt was submitted to the cBio portal (Prostate cancer and mRNA Expression vs Normal options).

Copy the results from Download panel, and paste them into the spreadsheet. Transform the results as follows, and then save them at the end of the **Correlation+Cancer.EA** file:

LGALS3BP+%28pp%29+Prostate_C = 0.6

ACTN1+%28pp%29+Prostate_C = 0.39

ACTB+%28pp%29+Prostate_C = 0.21

HSPG2+%28pp%29+Prostate_C = 0.16

TFRC14+%28pp%29+Prostate_C = 0.16

HBA1+%28pp%29+Prostate_C = 0.15

CFHR2+%28pp%29+Prostate_C = 0.14

The cancer-protein network could be constructed by converting "+%28pp%29+" by " pp". "= value" could be deleted. Save the cancer-protein network as **Cancer_Network.sif**.

Experimental Procedure

This paragraph list briefly all steps that can be viewed in supplemental video.

Network Construction

Construct the protein interaction network using modulated proteins (Network 1).

Construct the correlation network using the CORREP matrix (Network 2).

Merge networks 1 and 2 (Network 3).

Organize the network using organic Layout.

Attribute Reactome Function Interaction

Fetch Function Interaction (FI) annotations for node

Analyze the network functions.

Cluster the Function Interaction Network

Analyze module functions

Optimize the network view using the tool panel (scaling and rotating) and manual organization

Attribute Experimental data

- Nodes attributes

  Load Fold_change.NA

  Load FDR.NA

- Edges Attribute

  Load Correlation.EA

VizMapper Setting

1. Reactome FI attributes

   - Node Color: Cancer Module (from Reactome FI visual style)

   - Node Label: canonical name (from Reactome FI visual style)

   - Node Shape: IsLinker (from Reactome FI visual style)

   - Edge Line style: FI annotation (from Reactome FI visual style)

   - Edges Source Arrow: FI Direction (from Reactome FI visual style)

   - Edges Target Arrow: FI Direction (from Reactome FI visual style)

2. Experimental Data attributes

- Node border Color: Fold change (From Fold_Change.NA file)
- Node line width: False Discovery Rate (FDR) (from FDR.NA file)
- Edge line width: Multiple Correlation Analysis and Cancer cases overlapping (from Correlation+Cancer.EA file)
- Edge line color: Multiple Correlation Analysis (from Correlation+Cancer.EA file)

Cancer-protein Network Construction

Import the cancer-protein network from the Cancer_Network.sif file.

Merge Cancer_Network and Network 3.

attribute the Cancer-frequency to node size.

Optimize the view using the organic Layout and Tool panel.

Biological Process Enrichment : ClueGO setting
- Launch the ClueGO cytoscape plug-in.
- Set the type of analysis: single cluster
- Select the organism: *Homo sapiens,* and the type of IDs used (symbol)
- Load sample gene lists from file (only-modulated_symbol.txt).
- Select the Ontologies:
  - GO_BiologicalProcess (Evidence codes:EXP, IDA, IMP)
  - KEGG_Pathways (not necessary)
  - REACTOME_BioCarta (not necessary)
- Select the statistical test: Enrichment/Depletion (Two-sided hypergeometric test)
- Select the correlation method: Bonferroni
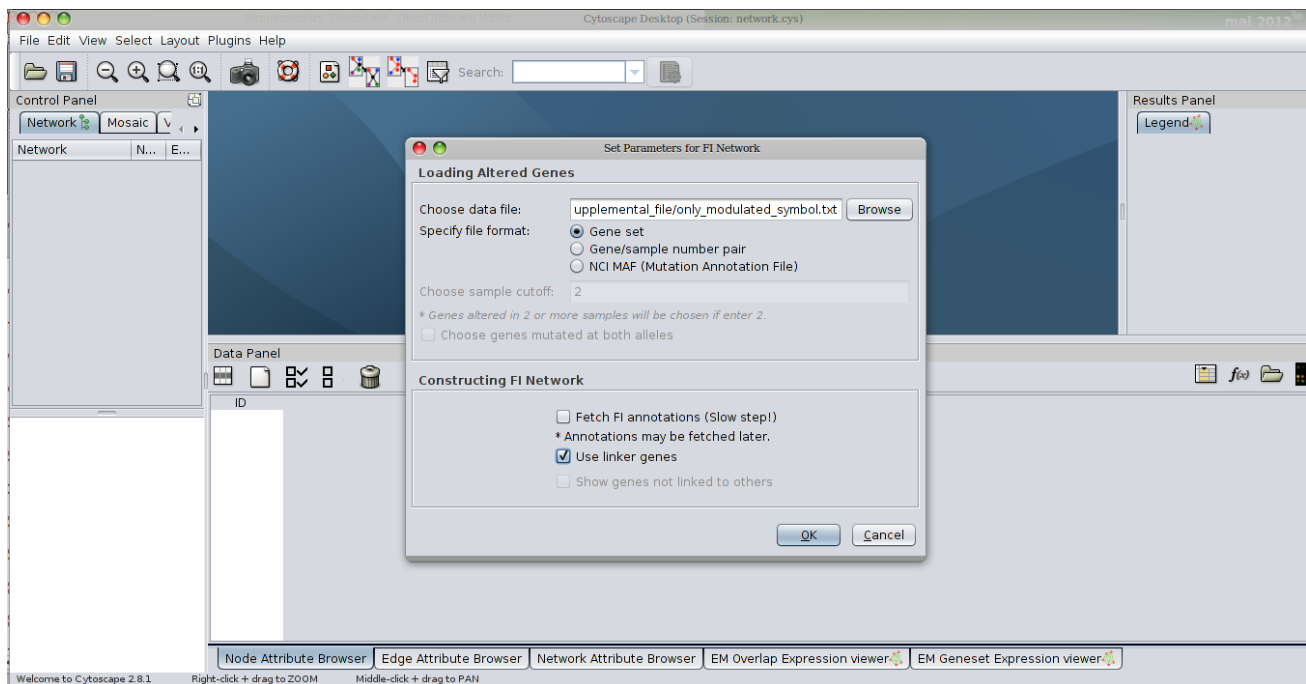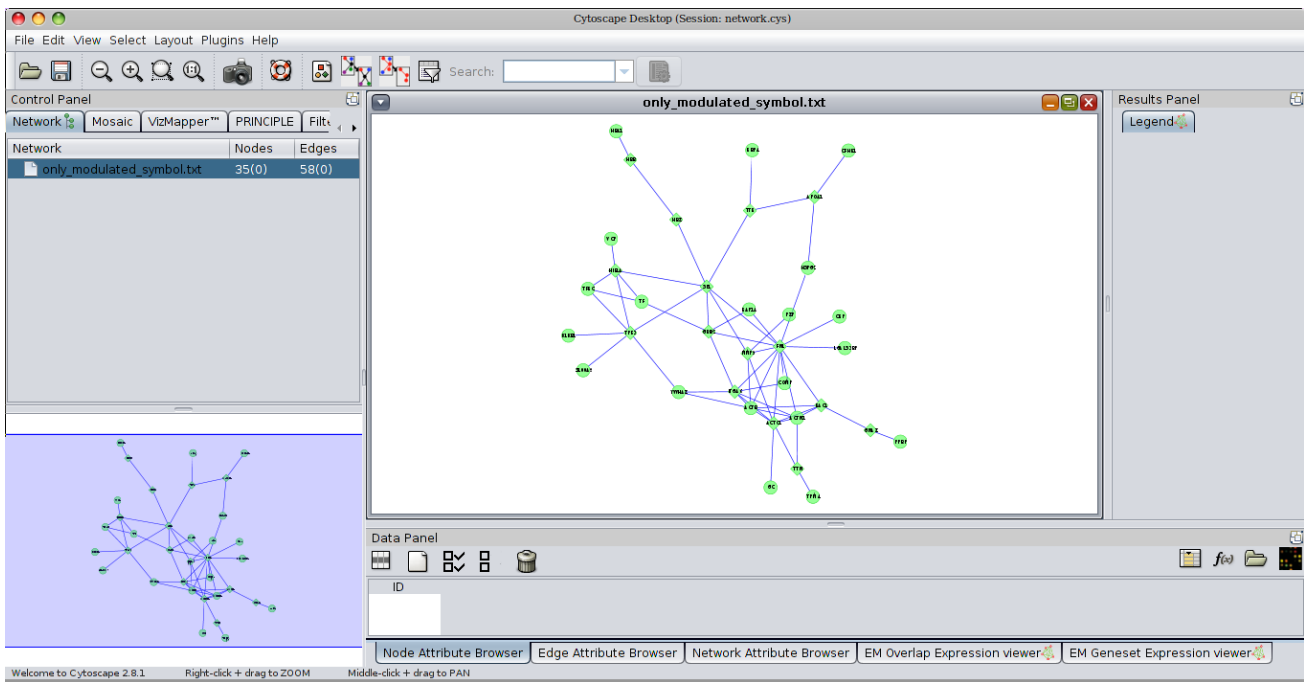- Use default advanced settings
- Start.

Walk-through example

Open a new Cytoscape session and save it as networks.cys

Network Construction

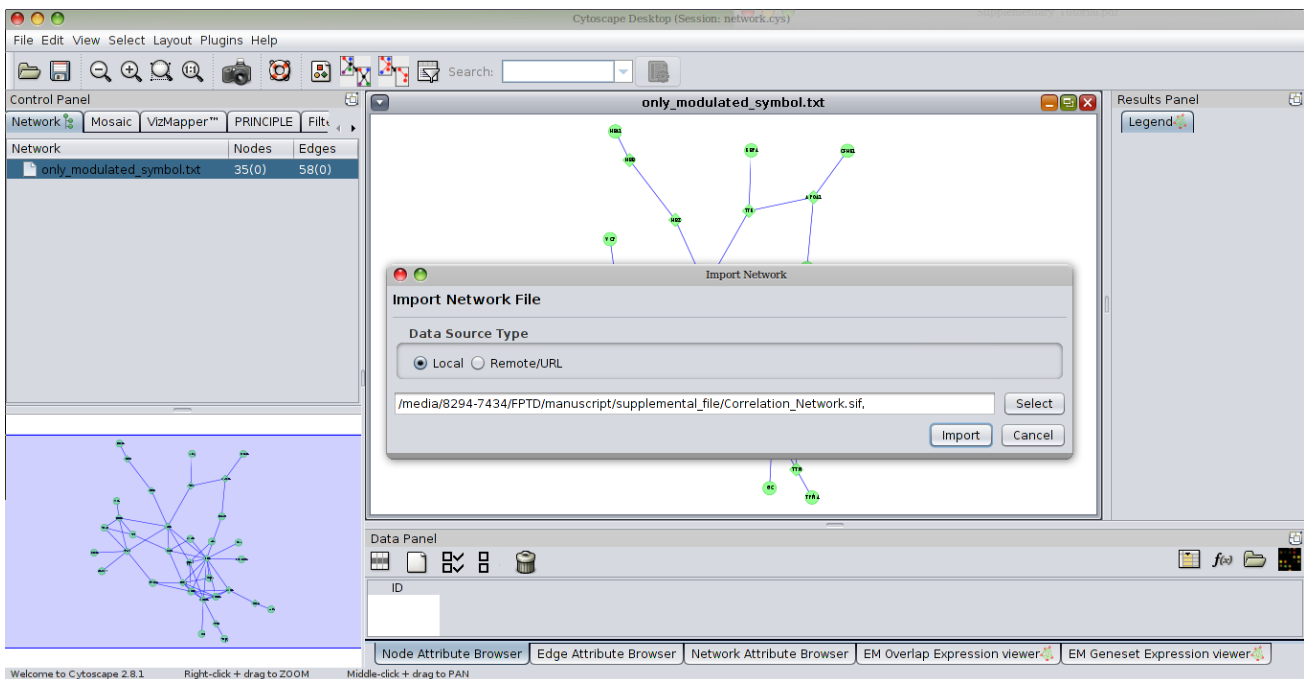Load the gene list (only-modulated_symbol.txt) through the Reactome FI plug-in, and use linker genes. Click OK.

Caption 1: Cytoscape FI setting

**Caption 2:** Network constructed by Cytoscape FI with the only-modulated-symbol.txt file

Import correlation network from Correlation_Network.sif file.

Caption 3: Load the correlation network from the Correlation_Network.sif file

Remove duplicated edges using the "Network Modifications" plug-in.

Merge the two networks (only-modulated-symbol.txt and Correlation_Network.sif) using the "Advanced Network Merge" plug-in.
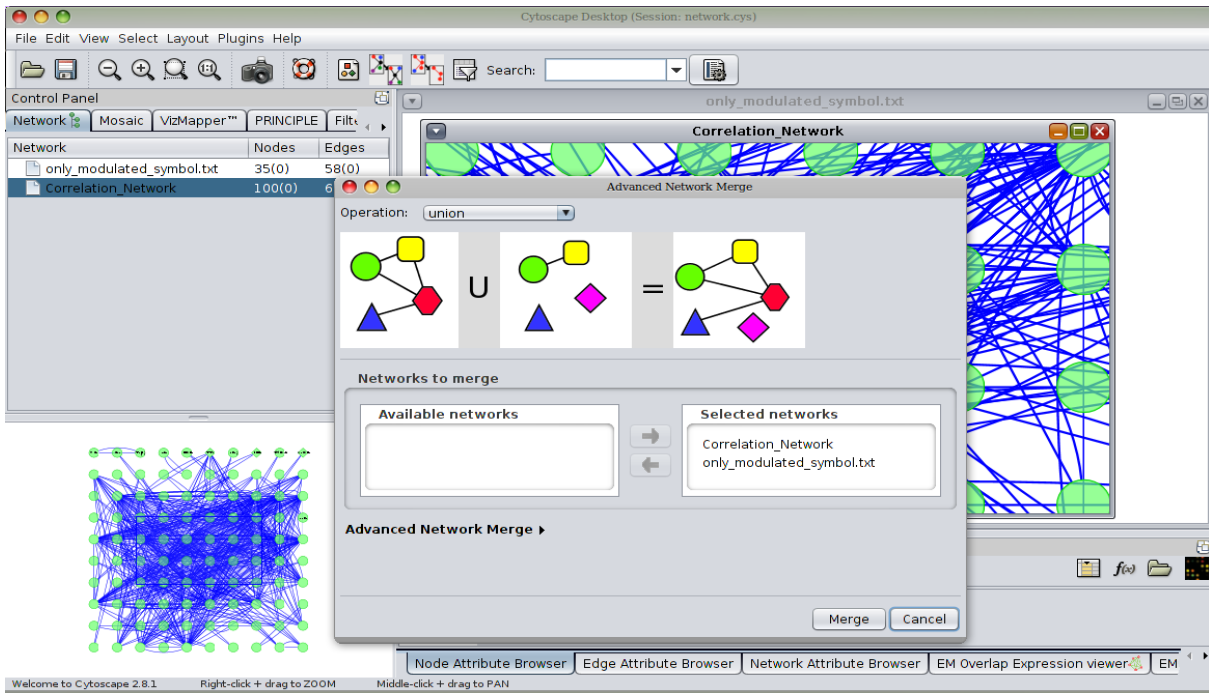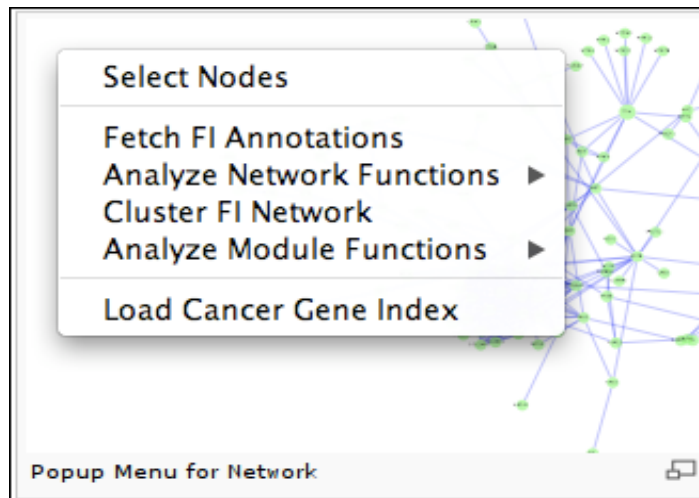
Caption 4: Merging only-modulated-symbol.txt and Correlation_Network.sif
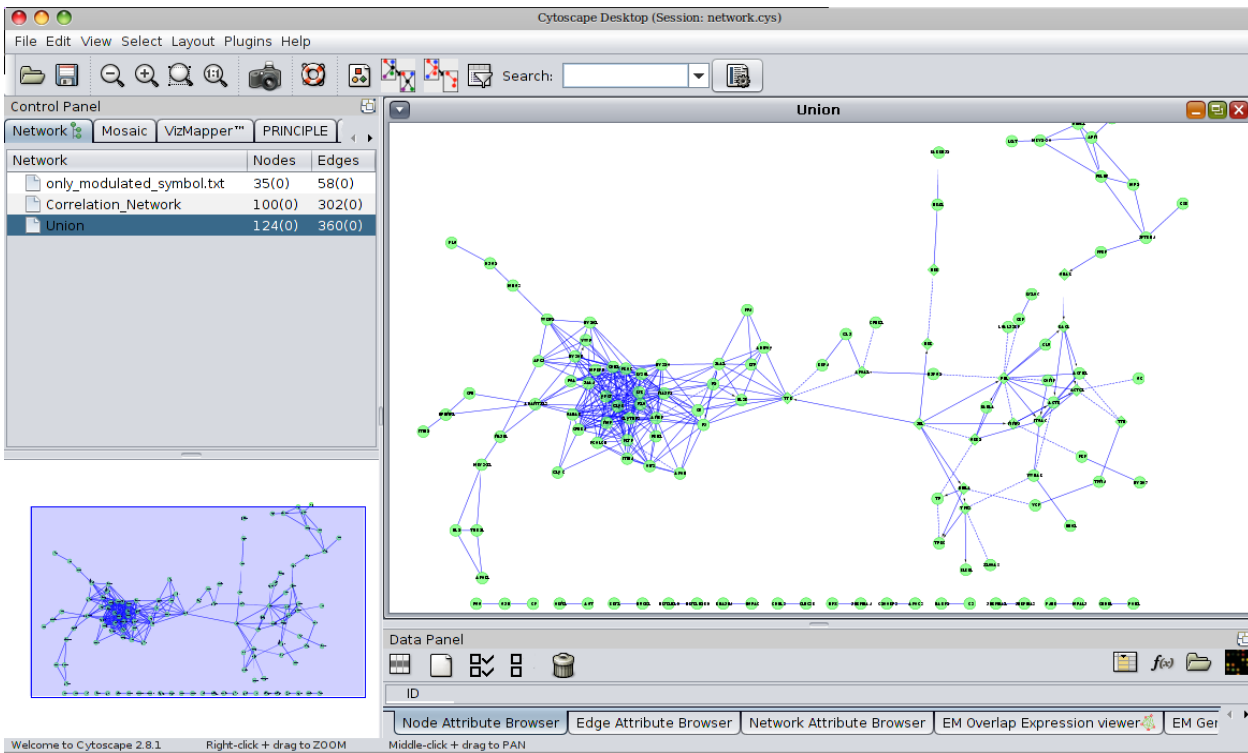
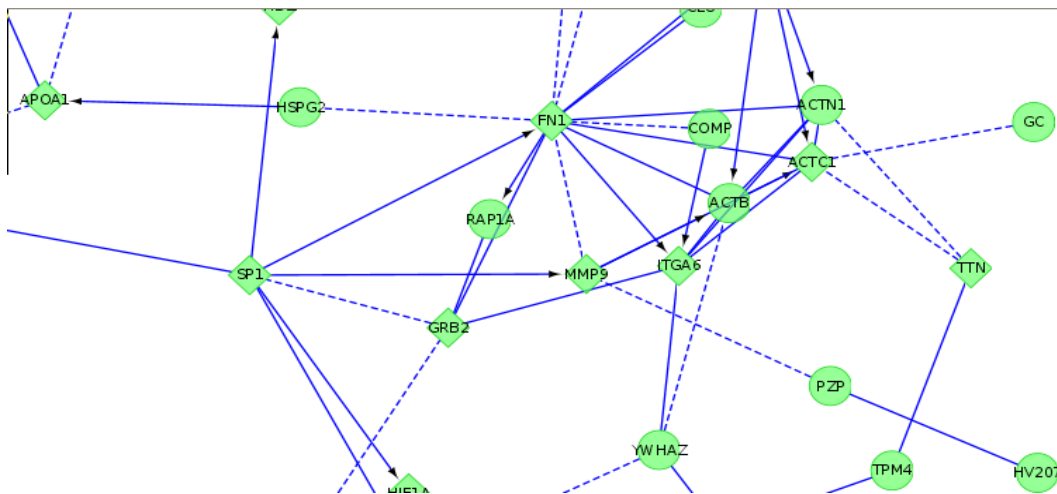**Caption 5:** The Layout setting (organic Layout)

Attribute Reactome Function Interaction

The main features of the Reactome FI plug-in should be selected from a popup menu, which can be displayed by right clicking on an empty space in the network view panel.


Caption 6: Fetch FI Annotations

Fetch FI Annotations

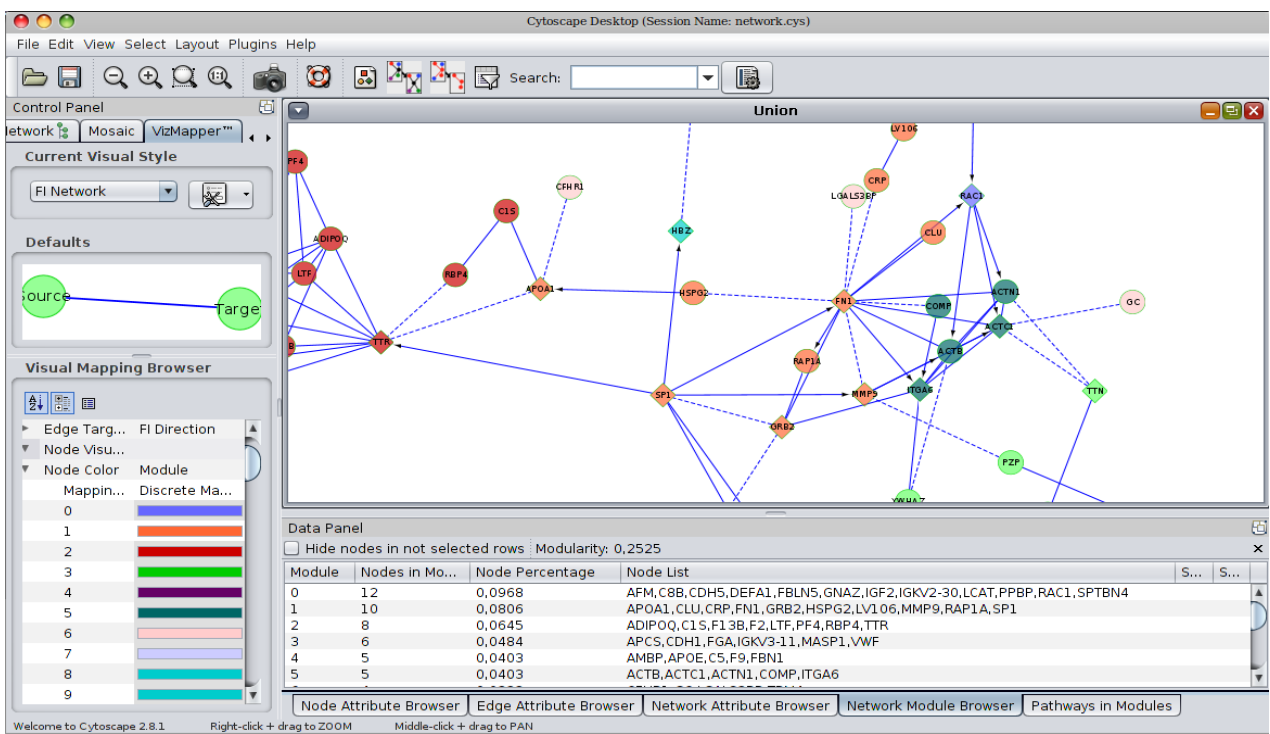**Caption 7:** Edges will be displayed on the basis of the FI direction attribute values. In the following screen shot, "->" for activating/catalyzing, "-|" for inhibition, "-" for FIs extracted from complexes or inputs,"---" for predicted FIs, Node Label: canonical name or ID, Node Shape: IsLinker. See the "VizMapper" tab, Node source Shape, Edge Source Arrow Shape and Edge Target Arrow Shape values for details.

Analyse Network Functions
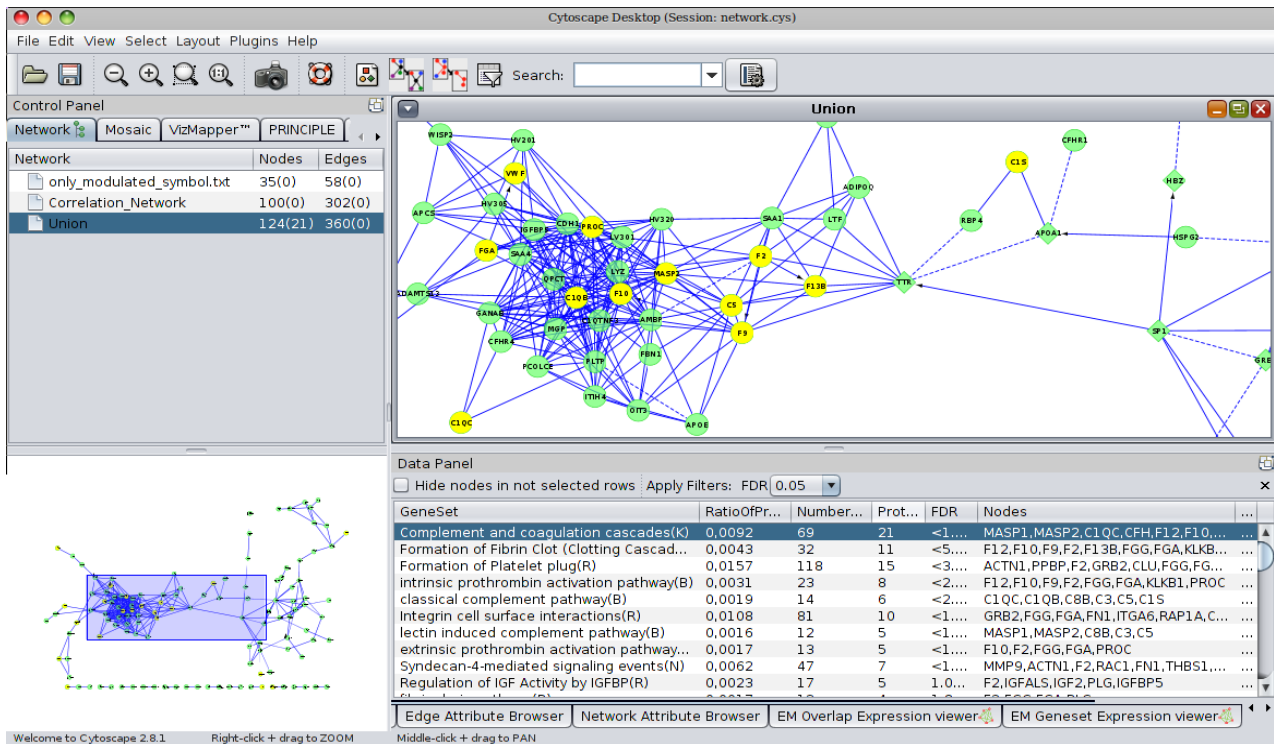
Caption 8: Pathway or GO term enrichment analysis for the network displayed. You can choose to filter the enrichment results by an FDR cutoff value.

Cluster Function Interactions network

Caption 9: Nodes in different network modules will be shown in different colors

Analyse of the module functions : Pathway enrichment

Caption 10: Pathway or GO term enrichment analysis.

VizMapper setting from Reactome FI visual style

Node Label: canonical name or ID

Node Shape: IsLinker (from Reactome FI visual style)

Node Color: Module(from Reactome FI visual style)

Edges Target Arrow: FI Direction (from Reactome FI visual style)

Caption 11: VizMapper setting from Reactome FI
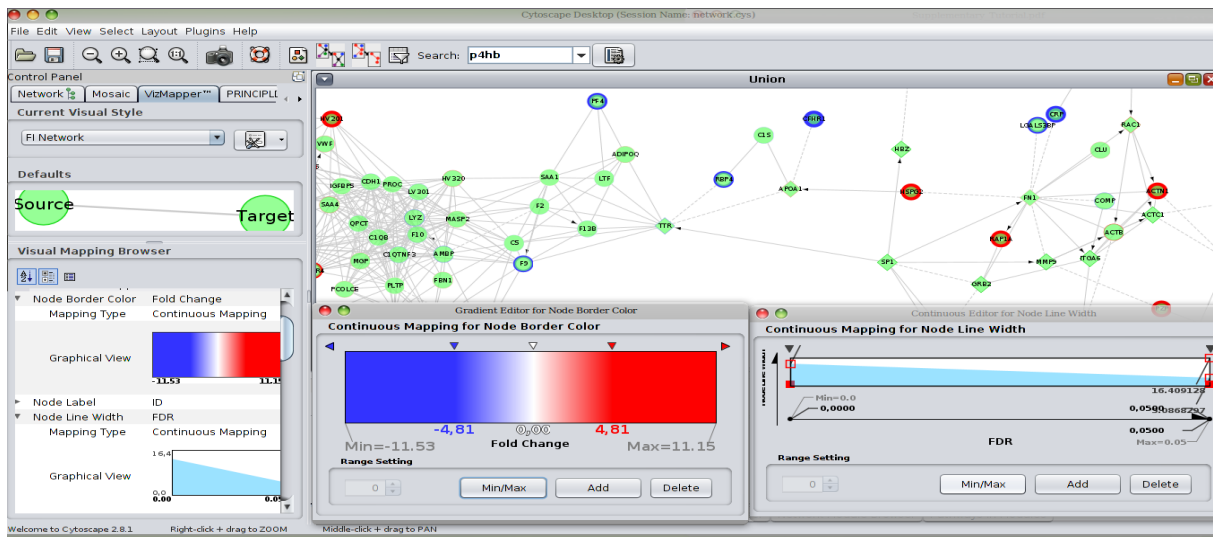
14

Attribute Experimental data for nodes

Import Node attributes files (Fold_Change.NA and FDR.NA).

Double click on "Node Border color" and "Node Line Width" in the VizMapper panel (unused properties).

Select the value of the Node Border color to "Fold Change"

Select the value of the Node line width to "FDR"

Set the mapping Type to "Continuous Mapping"

Caption 12: Attribute Node line width and Node line color as "Fold Change" and "FDR", respectively. Negative and Positive fold changes are indicated by blue and red, respectively. The intensity of the color depends on the fold change value. The node width line is limited by the FDR interval [0-0.05]. The node color module has been omitted to make it easier to view the node line.

Attribute Experimental data for edges

Import Correlation+Cancer.EA file

Double click on "Edge Line width" and "Edge Color" from VizMapper panel (unused properties)
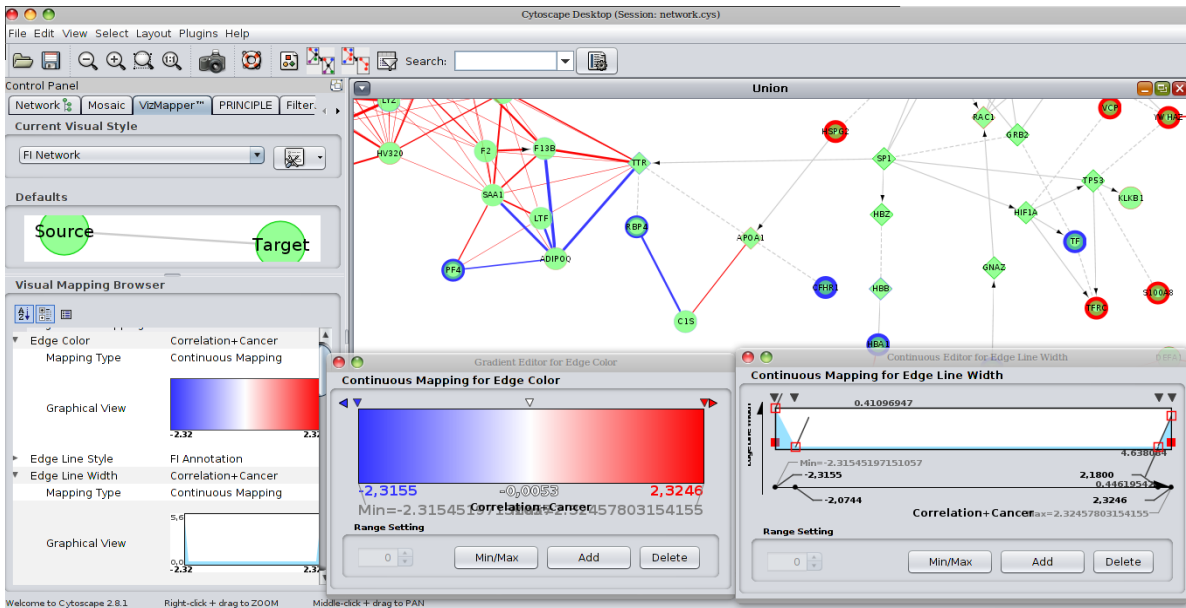
Caption 13: Attribute edge line width and edge color using the Correlation+Cancer.EA file. The correlation coefficients have been multiplied by 2.5 to distinguish in the graph between the correlation values and the cancer values. The edge line width was set to r<-2.07 and r>2.18, corresponding to r<-0.807 and r>0.818, respectively. Negative and the Positive correlations are indicated by blue and red, respectively.
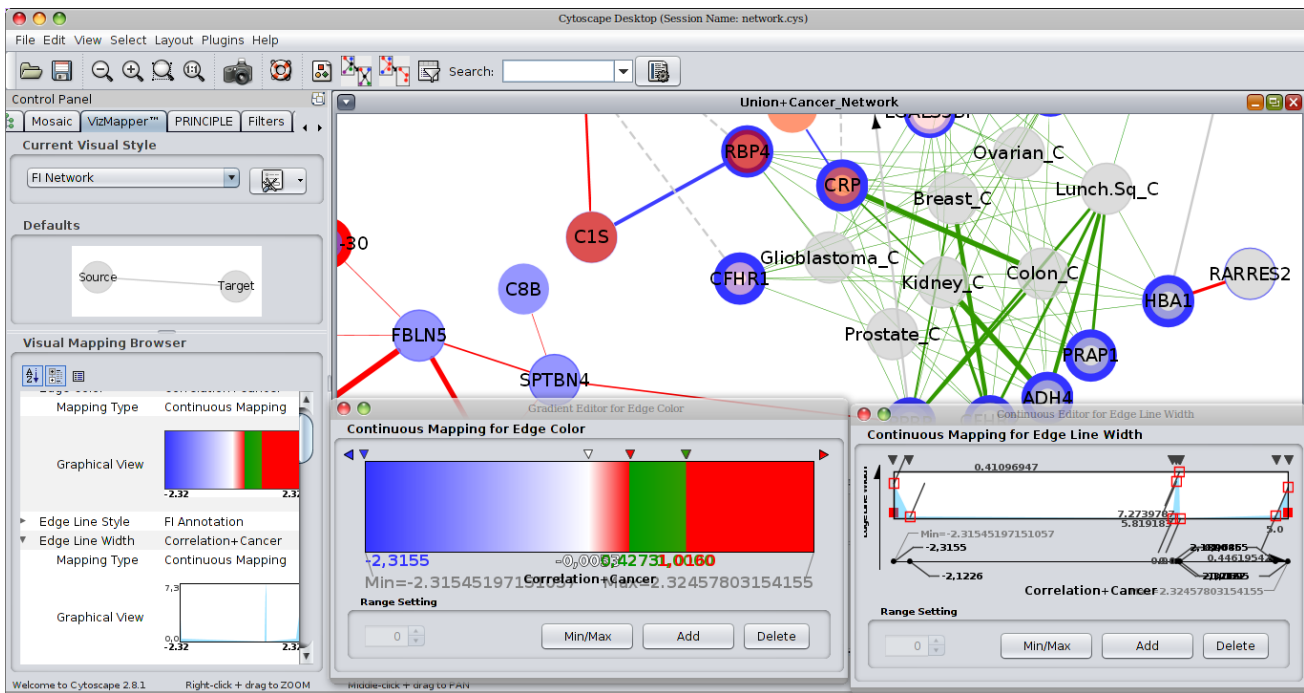
Cancer-protein Network Construction

Import Cancer network from the Cancer_Network.sif file

Optimize the view using organic Layout and manual organization.

Reset the edge line width and edge color as following in order to distinguish the cancer edges.

Caption 14: The node color was reset using the cluster module color. The cancer nodes are in the default color (grey). The edge color mapping has been optimized so that cancer node interactions stand out in green. The edge width line has been optimized so as to highlight the genes most frequently regulated in cancers.

Attributes the cancer frequency to node size.
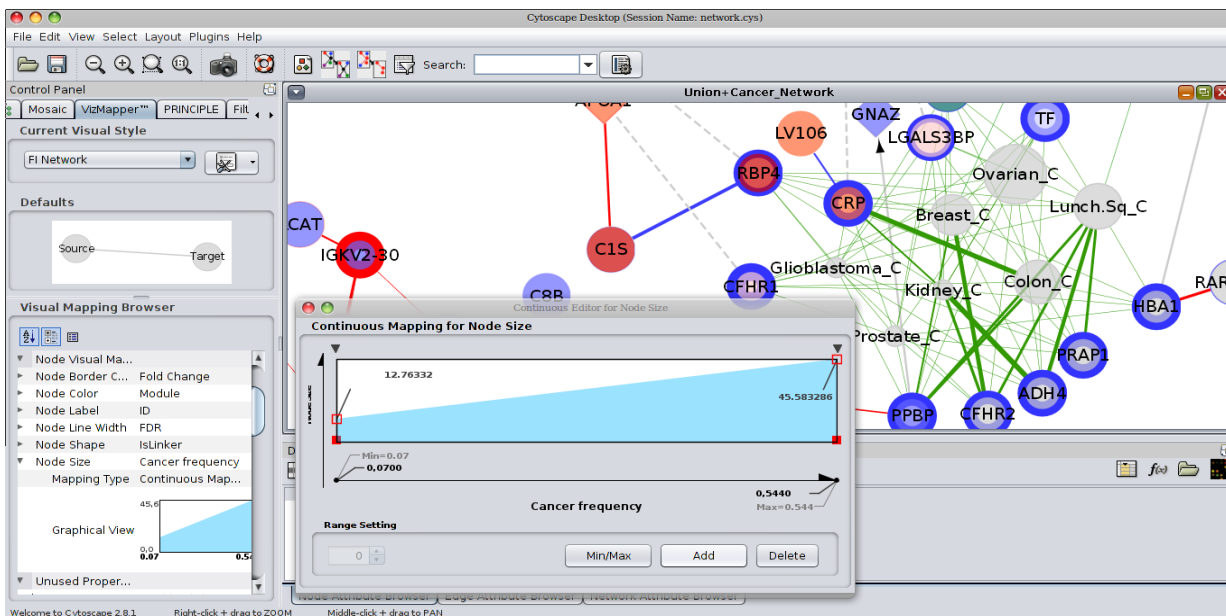
Caption15: The most probable cancers are distinguished by node size. In this case clinical Ovarian carcinoma data display the most gene expression overlap with our data.

Biological Process enrichment – ClueGO

Depending on the context, ClueGO user could include gene sets other than those provided by the ClueGO plug-in, such as the MsigDB gene sets. See the ClueGO documentation for the format of files.

Our example uses a gene list from a study looking at human plasma exposed to ionizing radiation. Enrichment Map was run using the default settings. The gene list was selected from the only-modulated-symbol network and submitted to the GO_Biological_Process_03.05.2012 database with three GO evidence codes selected (EXP, IDA, IMP) (see ClueGO settings).
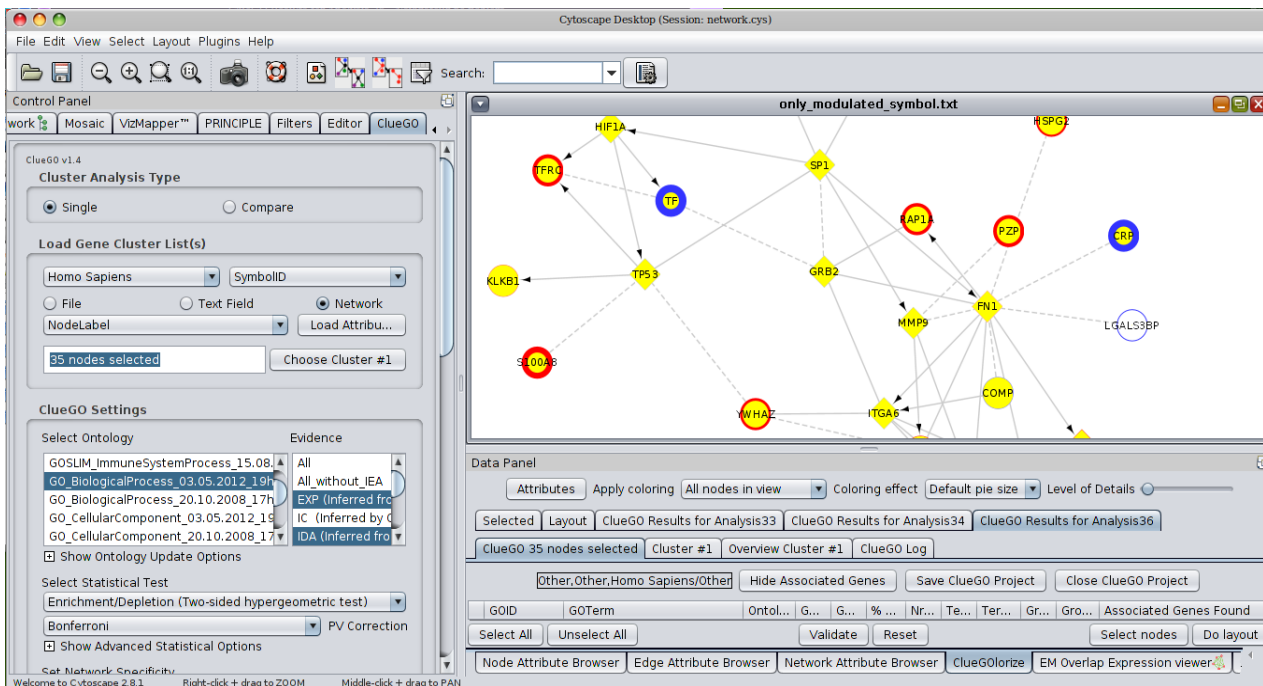
Caption 16: ClueGO Biological process [EXP, IDA, IMP] using nodes from Reactome Function interactions prediction (Network 1). EXP, Inferred from Experiment; IDA, Inferred from Direct Assay; IMP, Inferred from Mutant Phenotype.

ClueGO setting

All Results were created with ClueGO v1.4

Evidence Codes used: [EXP,IDA,IMP]

#Genes in GO_BiologicalProcess_03.05.2012_19h02 : 4951

#All unique Genes: 4951

#All Genes found from initial Cluster#1 (35): 19 (54,29%)

#Genes found from Cluster#1 after selection: 5 (14,29%)

KappaScore Grouping:

Iteration: 0 with 0 groups

Final KappaScore groups = 0

Terms not grouped = 5

#GO All Non Redundant Terms Specific for Cluster #1: 5

Statistical Test Used = Enrichment/Depletion (Two-sided hypergeometric test)

Correction Method Used = Bonferroni

SampleFile1 = only_modulated_symbol.txt

MinLevel = 3

MaxLevel = 8

EntireLevel = false

NoGenes1 = 2

GetAllGenes1 = false

MinPercentage1 = 4.0

GetAllPercentage1 = false

GOFusion = true
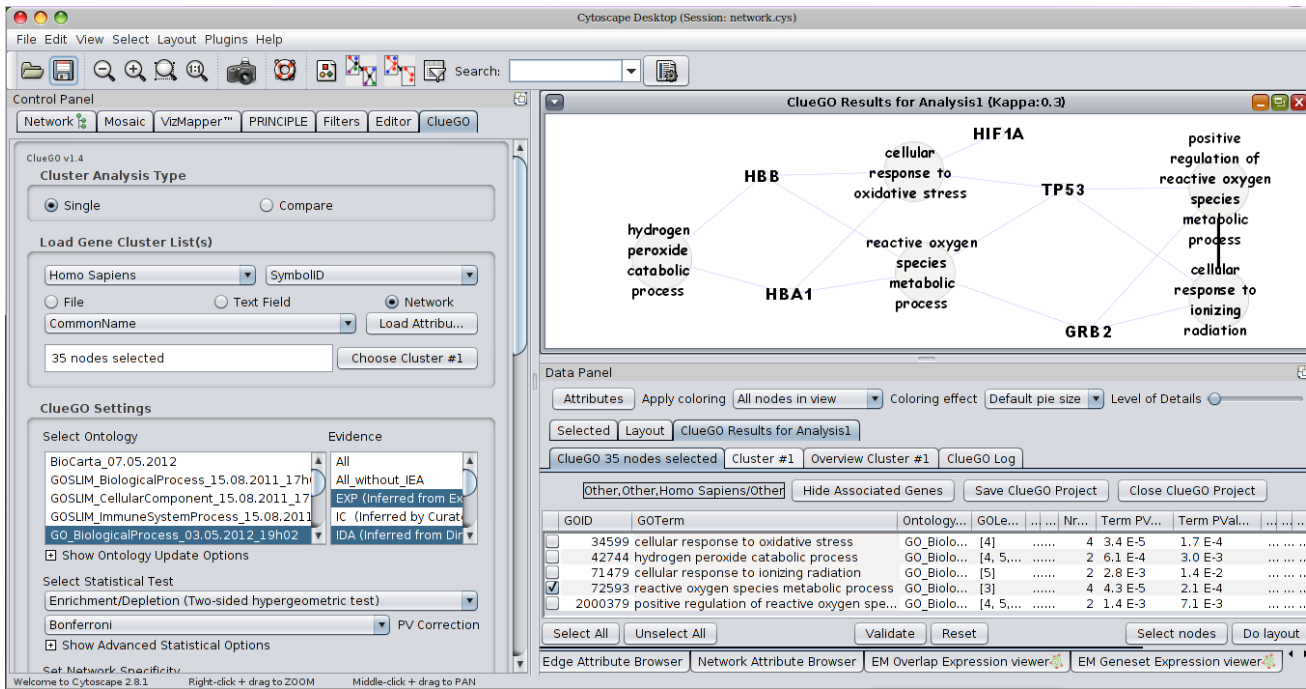
GOGroup = true

KappaScoreThreshold = 0.3

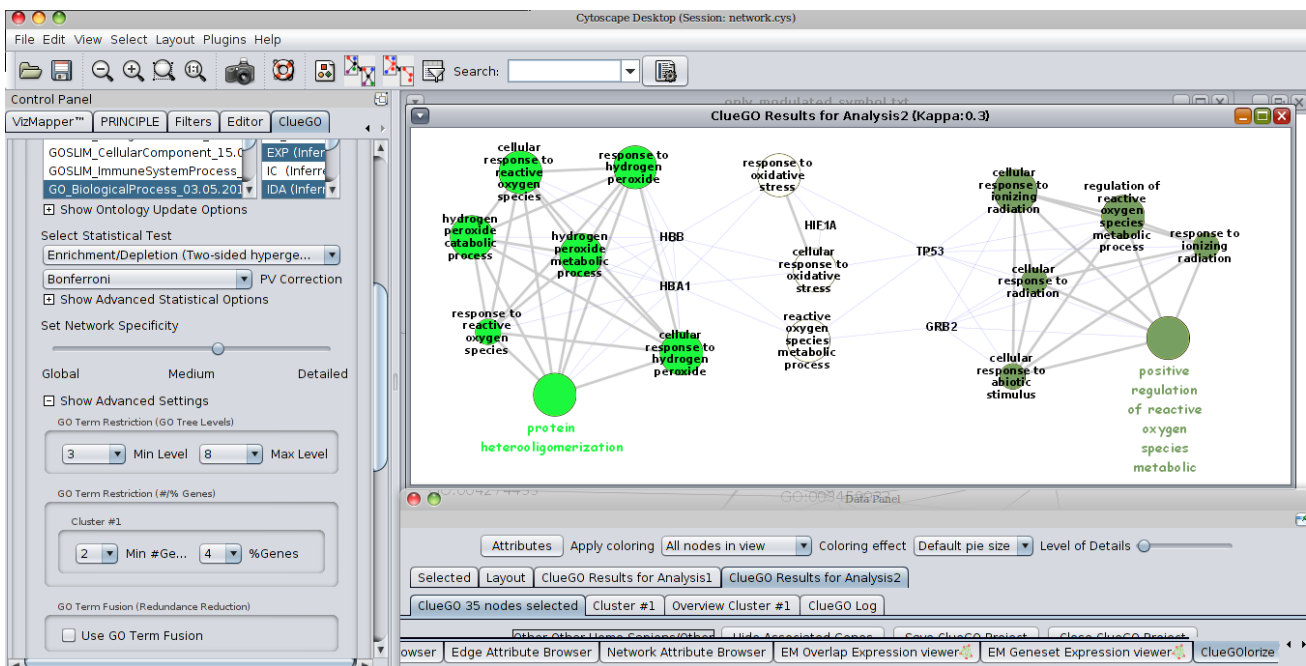OverViewTerm = SmallestPValue

GroupByKappaStat = true

InitialGroupSize = 3

SharingGroupPercentage = 50

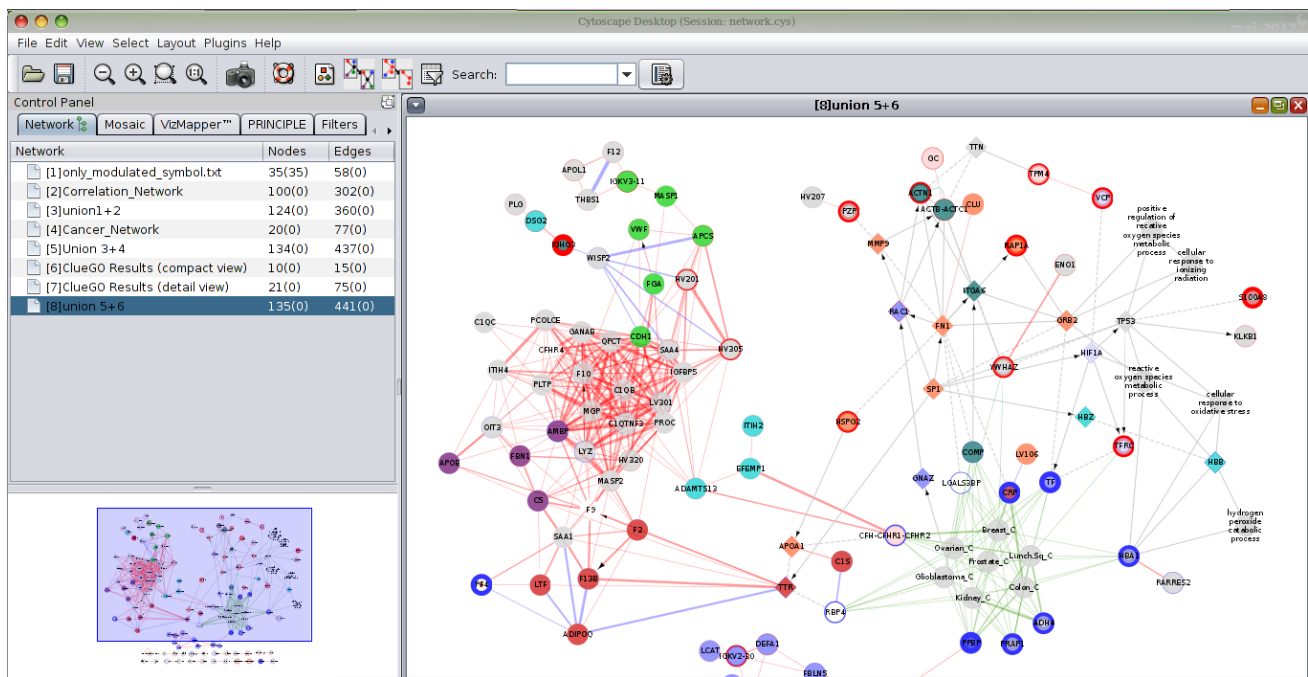Correction Method Used = Bonferroni
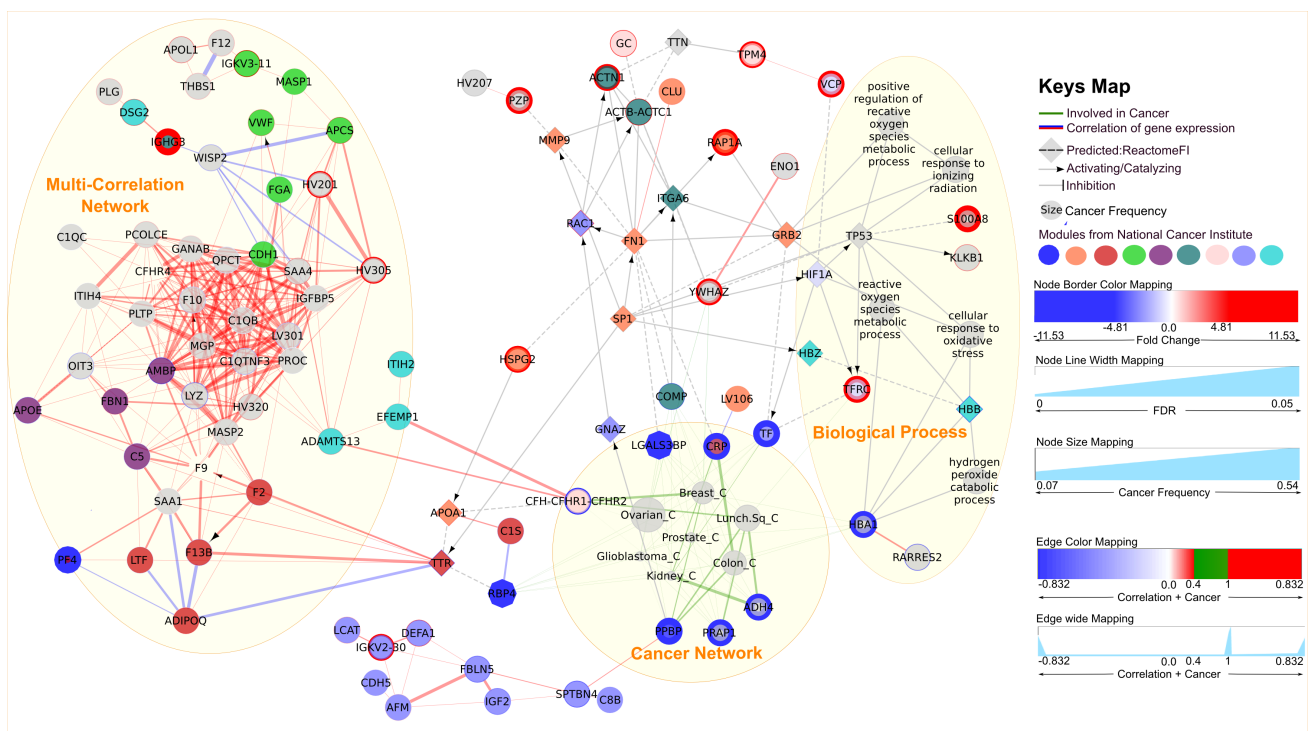
Caption 17: ClueGO Biological process [EXP,IDA,IMP] result with Fusion GOTerms

Caption 18: ClueGO Biological process [EXP,IDA,IMP] result without Fusion GOTerms

Caption 19: The Merge of ClueGO results (compact, Network 6) and the Network with cancer nodes (Network 5)

Capture 20: A, Work-flow of Interactive Enrichment Network; B, Enriched Network.

Supplemental References for Supplemental Tutorial

Adrian Alexa, 2012. Bioconductor - topGO . URL http://www.bioconductor.org/packages/release/bioc/html/topGO.html

Alibés, A., Yankilevich, P., Cañada, A., Díaz-Uriarte, R., 2007. IDconverter and IDClight: conversion and annotation of gene and protein IDs. BMC bioinformatics 8, 9.

Beissbarth, T., Speed, T.P., 2004. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. Bioinformatics (Oxford, England) 20, 1464–5.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z., Galon, J., 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics (Oxford, England) 25, 1091–3.

Bioconductor - Cormotif , 2012. . URL http://www.bioconductor.org/packages/release/bioc/html/Cormotif.html

Boorsma, A., Foat, B.C., Vis, D., Klis, F., Bussemaker, H.J., 2005. T-profiler: scoring the activity of predefined groups of genes using gene expression data. Nucleic acids research 33, W592–5.

Castelo, R., Guinney, J., Sonja, H., 2011. BIOINFORMATICS GSVA : Gene Set Variation Analysis. Doctor 1–7.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A.P., Sander, C., Schultz, N., 2012. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery 2, 401–404.

Christmas, R., Avila-Campillo, I., Bolouri, H., Schwikowski, B., Anderson, M., Kelley, R., Landys, N., Workman, C., Ideker, T., Cerami, E., Sheridan, R., Bader, G.D., Sander, C., 2005. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. AACR Education Book 2005, 12.

Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., Hanspers, K., Isserlin, R., Kelley, R., Killcoyne, S., Lotia, S., Maere, S., Morris, J., Ono, K., Pavlovic, V., Pico, A.R., Vailaya, A., Wang, P.-L., Adler, A., Conklin, B.R., Hood, L., Kuiper, M., Sander, C., Schmulevich, I., Schwikowski, B., Warner, G.J., Ideker, T., Bader, G.D., 2007. Integration of biological networks and gene expression data using Cytoscape. Nature protocols 2, 2366–82.

CRAN - Package CCA , 2012. . URL http://cran.r-project.org/web/packages/CCA/

De, S., Zhang, Y., Garner, J.R., Wang, S.A., Becker, K.G., 2010. Disease and phenotype gene set analysis of disease-based gene expression in mouse and human. Physiological genomics 42A, 162–7.

Dinu, I., Potter, J.D., Mueller, T., Liu, Q., Adewale, A.J., Jhangri, G.S., Einecke, G., Famulski, K.S., Halloran, P., Yasui, Y., 2009. Gene-set analysis and reduction. Briefings in Bioinformatics 10, 24–34.

Draghici, S., Khatri, P., Tarca, A.L., Amin, K., Done, A., Voichita, C., Georgescu, C., Romero, R., 2007. A systems biology approach for pathway level analysis. Genome research 17, 1537–45.

Feng, G., Du, P., Krett, N.L., Tessel, M., Rosen, S., Kibbe, W.A., Lin, S.M., 2010. A

collection of bioconductor methods to visualize gene-list annotations. BMC research notes 3, 10.

Grossmann, S., Bauer, S., Robinson, P.N., Vingron, M., 2007. Improved detection of overrepresentation of Gene-Ontology annotations with parent child analysis. Bioinformatics (Oxford, England) 23, 3024–31.

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research 37, 1–13.

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols 4, 44–57.

Ignacio González, Sébastien Déjean, Pascal G. P. Martin, Alain Baccini, 2008. CCA: An R Package to Extend Canonical Correlation Analysis. Journal of statistic software 23.

Jari Oksanen, F., Guillaume, Blanchet, Roeland, K., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Peter, S., Stevens, H.H., Wagner, H., n.d. Canonical Correlation Analysis (CCorA). URL http://cran.r-project.org/web/packages/vegan/index.html

Kelder, T., van Iersel, M.P., Hanspers, K., Kutmon, M., Conklin, B.R., Evelo, C.T., Pico, A.R., 2012. WikiPathways: building research communities on biological pathways. Nucleic acids research 40, D1301–7.

Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics 9, 559.

Luo, W., Friedman, M.S., Shedden, K., Hankenson, K.D., Woolf, P.J., 2009. GAGE: generally applicable gene set enrichment for pathway analysis. BMC bioinformatics 10, 161.

Merico, D., Isserlin, R., Stueker, O., Emili, A., Bader, G.D., 2010. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PloS one 5, e13984.

Montojo, J., Zuberi, K., Rodriguez, H., Kazi, F., Wright, G., Donaldson, S.L., Morris, Q., Bader, G.D., 2010. GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. Bioinformatics (Oxford, England) 26, 2927–8.

Nyangoma, S., 2012. Clippda: A package for clinical proteomic profiling data analysis. submitted.

Research, T.C.G.A., 2011. Integrated genomic analyses of ovarian carcinoma. Nature 474, 609–15.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research 13, 2498–504.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America 102, 15545–50.

Tian, L., Greenberg, S.A., Kong, S.W., Altschuler, J., Kohane, I.S., Park, P.J., 2005. Discovering statistically significant pathways in expression profiling studies. Proc Natl Acad Sci U S A 102, 13544–13549.

Tomfohr, J., Lu, J., Kepler, T.B., 2005. Pathway level analysis of gene expression using singular value decomposition. BMC bioinformatics 6, 225.

Wu, G., Feng, X., Stein, L., 2010. A human functional protein interaction network and its application to cancer data analysis. Genome biology 11, R53.

Yoder, P.J., Blackford, J.U., Waller, N.G., Kim, G., 2004. Enhancing power while controlling family-wise error: an illustration of the issues using electrocortical studies. Journal of clinical and experimental neuropsychology 26, 320–31.

Yu, G., Wang, L.-G., Han, Y., He, Q.-Y., 2012. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. Omics : a journal of integrative biology 16, 284–7.

Yu, Y., Tu, K., Zheng, S., Li, Y., Ding, G., Ping, J., Hao, P., Li, Y., 2009. GEOGLE: context mining tool for the correlation between gene expression and the phenotypic distinction. BMC bioinformatics 10, 264.

Zhu, D., Li, Y., Li, H., 2007. Multivariate correlation estimator for inferring functional

relationships from replicated genome-wide data. Bioinformatics (Oxford, England) 23,

2298–305.