



# Cytoscape Network Visualization and Analysis Workshop

Welcome and Introduction

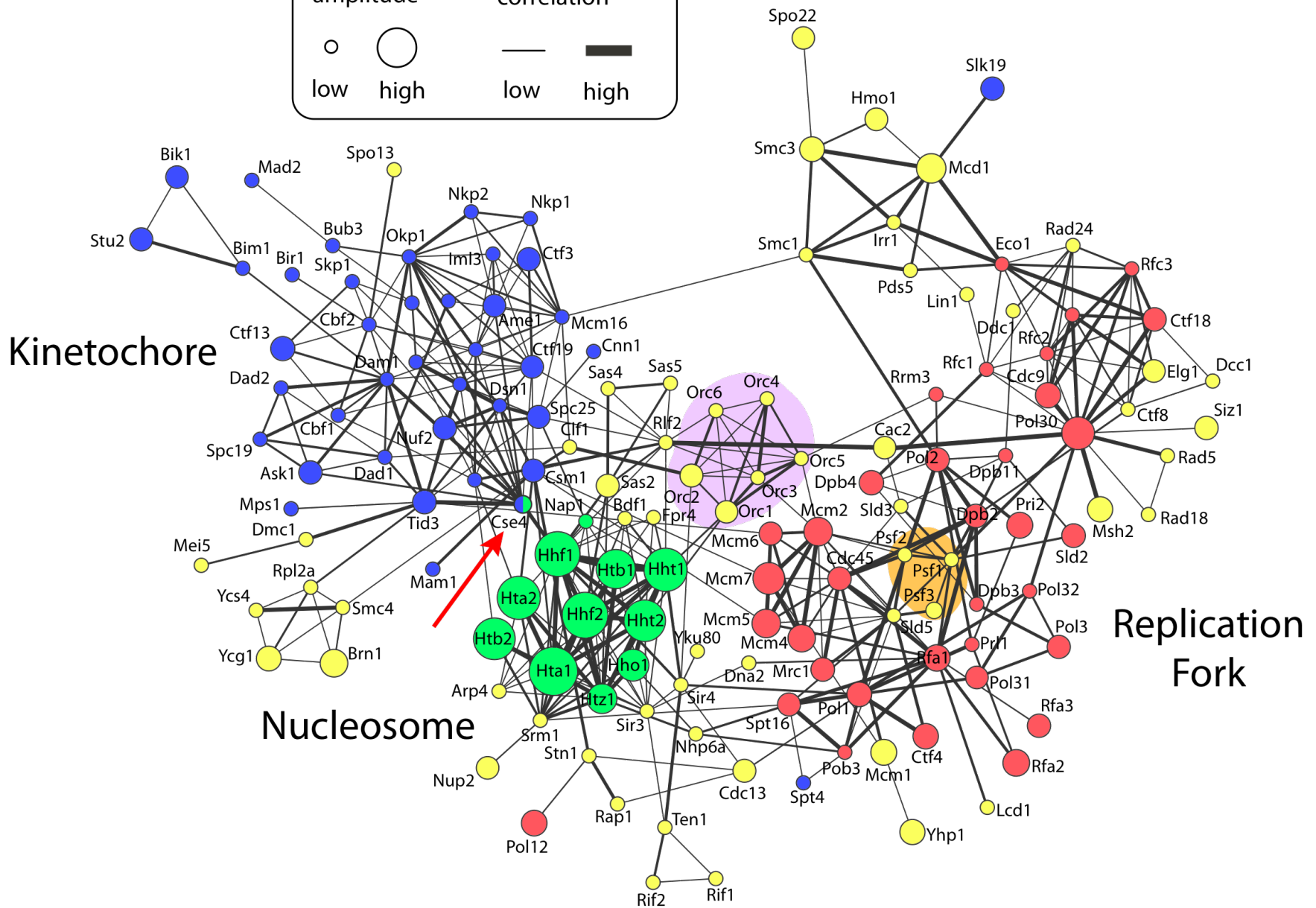
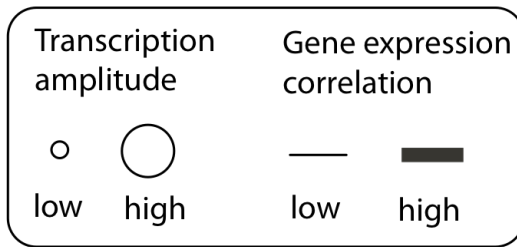
Sep.30-Oct.1, 2010

EMBL, Heidelberg



# Course Outline

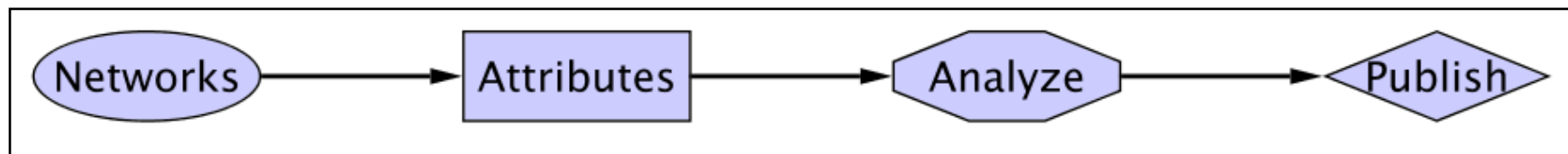
- Collaborative course notes
  - Wiki page: <http://baderlab.org/EMBLHeidelbergCytoscapeCourse>
  - Google document for collaborative course notes
- Introduction to gene lists, network visualization and analysis using Cytoscape
- Workflow: loading a network to publication
- Popular analysis scenarios
- In depth application: pathway analysis of a gene list
- Each module composed of lectures and labs



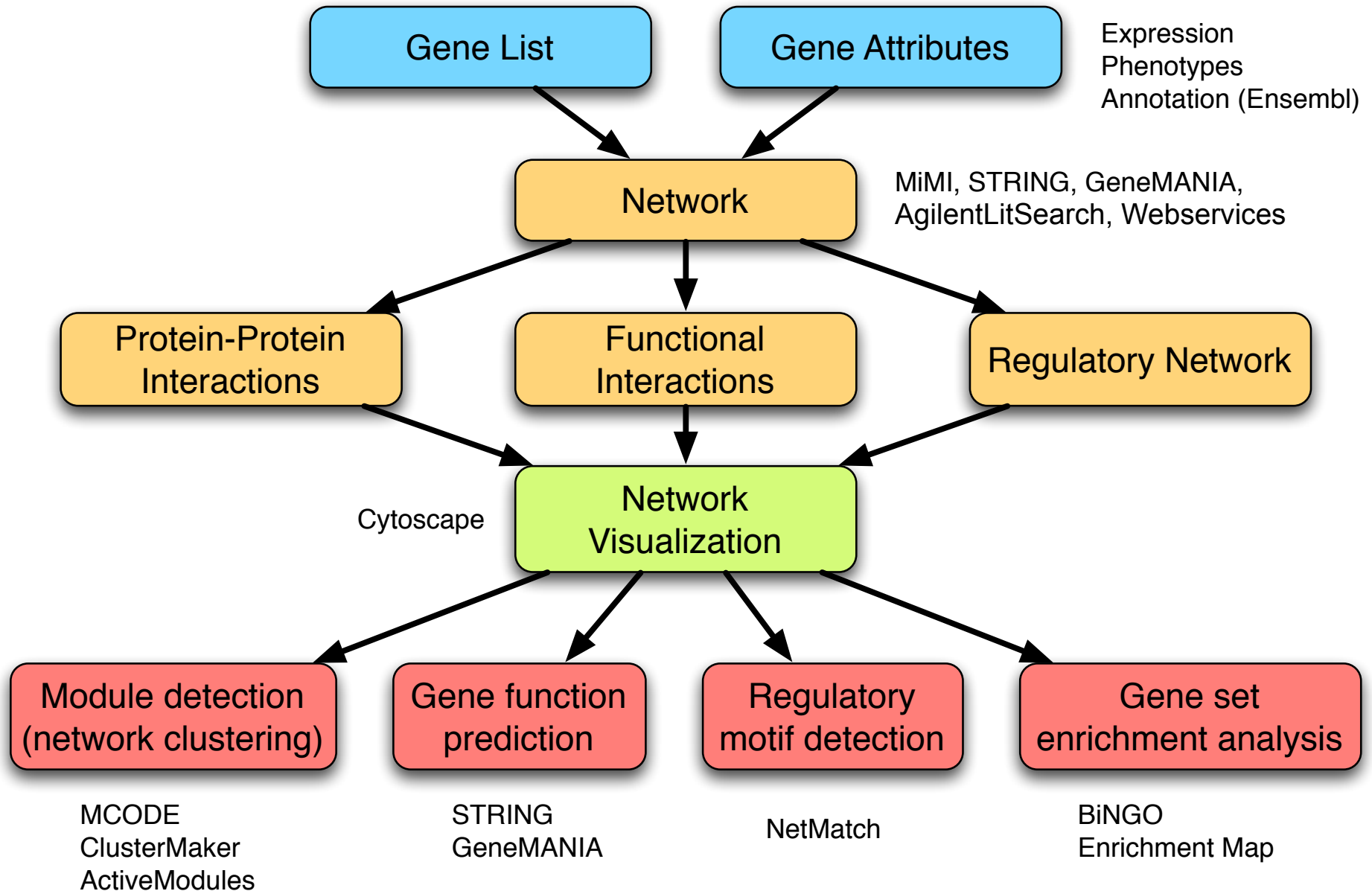


# Cytoscape Workflow

1. Load Networks (Import network data into Cytoscape)
  2. Load Attributes (Get data about networks into Cytoscape)
  3. Analyze and Visualize Networks
  4. Prepare for Publication
- A specific example of this workflow:
    - Cline, et al. “Integration of biological networks and gene expression data using Cytoscape”, Nature Protocols, 2, 2366-2382 (2007).



# Gene List and Network Analysis Overview





# Schedule

- Day 1
  - Course introduction
  - Introduction to Cytoscape
  - Public seminar
  - Cytoscape lectures and labs
- Day 2
  - Pathway analysis of gene lists lecture
  - Cytoscape labs
  - Developing Cytoscape plugins



# Instructor Introductions



**Gary Bader**  
CCBR, University of Toronto  
<http://baderlab.org>

Network visualization and  
analysis using Cytoscape  
lectures

**Piet Molenaar**  
AMC Human Genetics,  
Amsterdam, The Netherlands  
<http://humangenetics-amc.nl/>

Network visualization and  
analysis using Cytoscape  
Developing Cytoscape plugins in  
Java



**Aidan Budd**  
Computational Biologist, Gibson Team, EMBL Heidelberg  
<http://www.embl.de/~budd/>

Course coordinator/organizer



**Pablo Minguez**  
Postdoctoral Fellow, Bork Group  
EMBL Heidelberg  
[http://www.embl.de/research/units/scb/bork/members/?s\\_personId=6034](http://www.embl.de/research/units/scb/bork/members/?s_personId=6034)

Assisting with labs, Cytoscape user



# Participant Introductions

- Where are you from?
- What research do you do?
- Where does your gene list or network come from? (i.e. what type of experiment?)

1 Minute Each







# Logistics

- Contact information
- Housekeeping
- Meals
- Take advantage of class & break times to peer network!
- Questions: Just ask any one of us, we will all be here for the two days.

This page is available in the following languages:

Afrikaans বাংলা Azərbaycanca Català Dansk Deutsch Ελληνικά English English (CA) English (GB) English (US) Esperanto  
Español Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE)  
Euskara Suomi français français (CA) Galego עברית hrvatski Magyar Italiano 日本語 한국어 Macedonian Malayu  
Nederlands Norsk Sesotho sa Leboa polski Português română slovenski jezik српски (latinica) Sotho svenska  
中文 華語 (台灣) isiZulu



### Attribution-Share Alike 2.5 Canada

#### You are free:



**to Share** — to copy, distribute and transmit the work



**to Remix** — to adapt the work



#### Under the following conditions:



**Attribution.** You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).



**Share Alike.** If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar licence to this one.

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- The author's moral rights are retained in this licence.

[Disclaimer](#)

Your fair dealing and other rights are in no way affected by the above.  
This is a human-readable summary of the Legal Code (the full licence) available in the following languages:  
[English](#) [French](#)

Scooter Morris  
Piet Molenaar  
Gary Bader  
Tero Aittokallio  
Boris Steipe  
Canadian Bioinformatics Workshops

[Learn how to distribute your work using this licence](#)



# Introduction to Gene Lists and Networks

Gary Bader



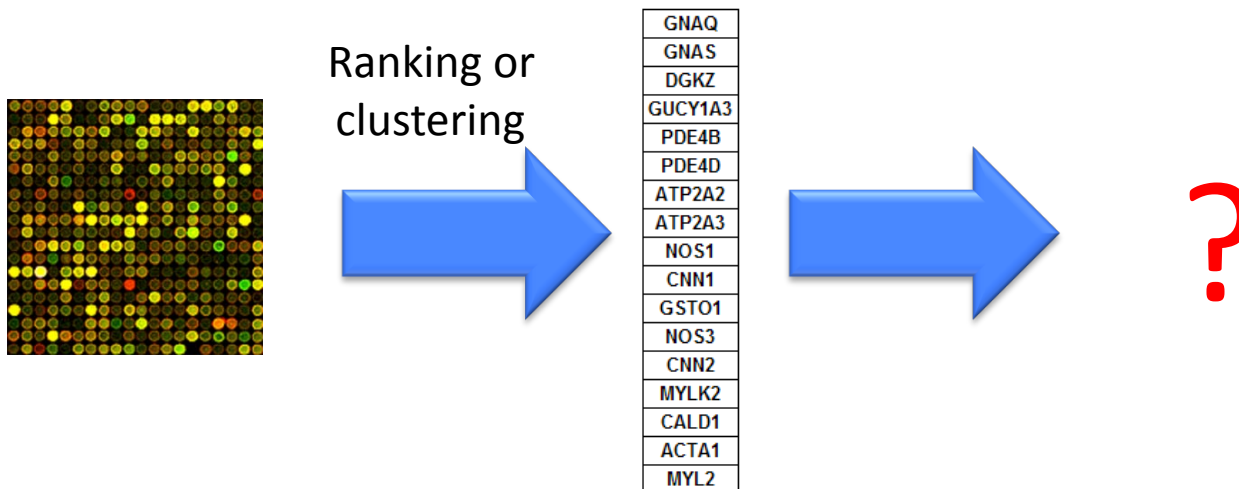
# Gene Lists and Networks Outline

- Gene lists
  - Where they come from
  - What do they mean
- Networks
  - What are they
  - Analysis
  - Use in Biology
  - Biological questions/applications



# Interpreting Gene Lists

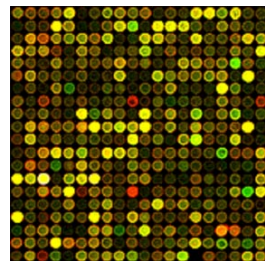
- My cool new screen worked and produced 1000 hits! ...Now what?
- Genome-Scale Analysis (Omics)
  - Genomics, Proteomics
- Tell me what's interesting about these genes





# Interpreting Gene Lists

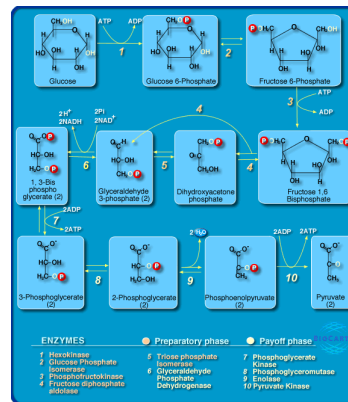
- My cool new screen worked and produced 1000 hits! ...Now what?
- Genome-Scale Analysis (Omics)
  - Genomics, Proteomics
- Tell me what's interesting about these genes
  - Are they enriched in known pathways, complexes, functions



Ranking or clustering

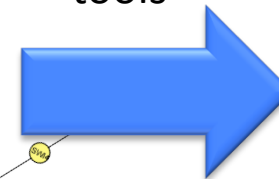


GNAQ
GNAS
DGKZ
GUCY1A3
PDE4B
PDE4D
ATP2A2
ATP2A3
NOS1
CNN1
GSTO1
NOS3
CNN2
MYLK2
CALD1
ACTA1
MYL2



Prior knowledge about cellular processes

Analysis tools



Eureka! New heart disease gene!



# Where Do Gene Lists Come From?

- Molecular profiling e.g. mRNA, protein
  - Identification → Gene list
  - Quantification → Gene list + values
  - Ranking, Clustering (biostatistics)
- Interactions: Protein interactions, microRNA targets, transcription factor binding sites (ChIP)
- Genetic screen e.g. of knock out library
- Association studies (Genome-wide)
  - Single nucleotide polymorphisms (SNPs)
  - Copy number variants (CNVs)



# What Do Gene Lists Mean?

- Biological system: complex, pathway, physical interactors
- Similar gene function e.g. protein kinase
- Similar cell or tissue location
- Chromosomal location (linkage, CNVs)





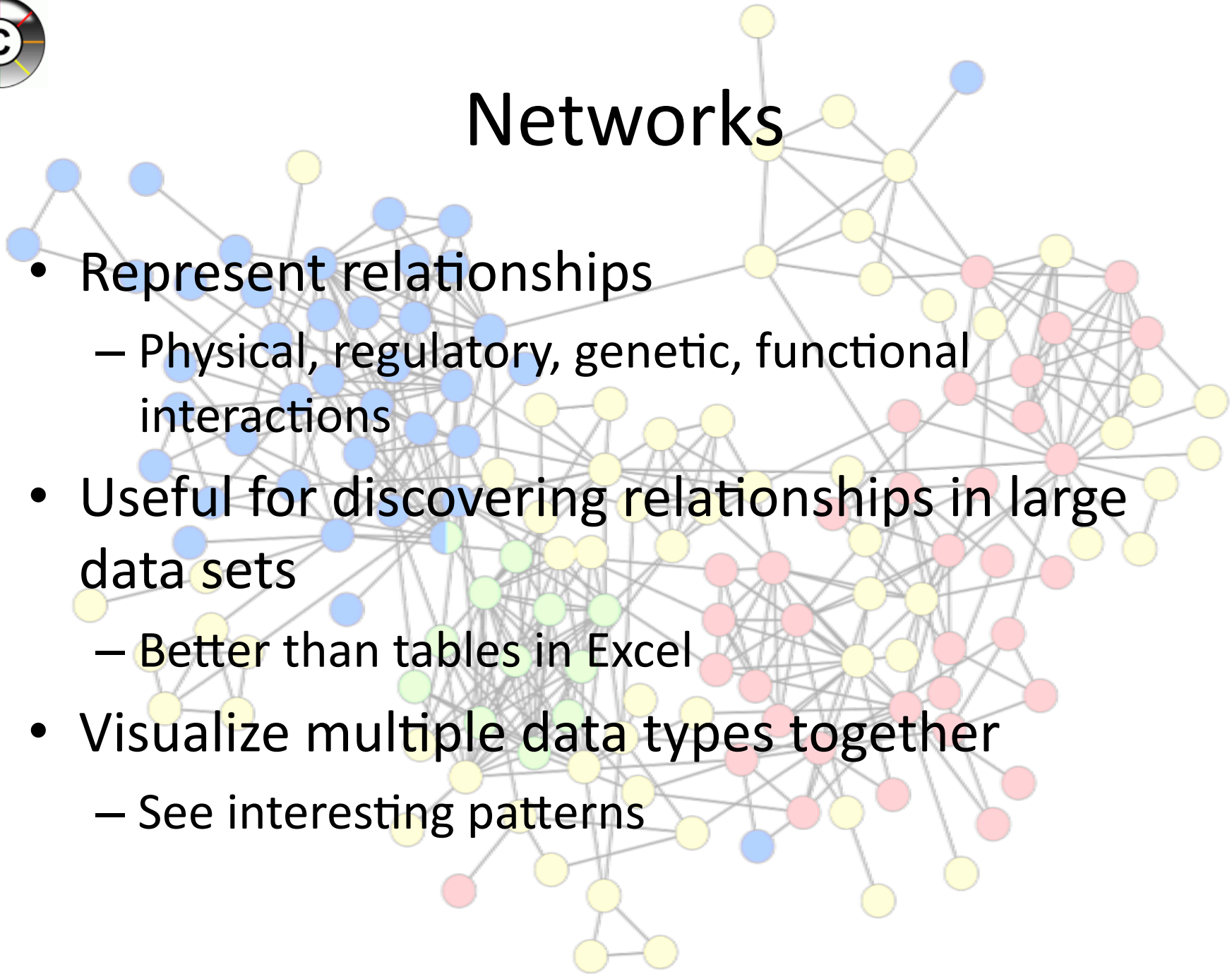
# Gene Lists and Networks Outline

- Gene lists
  - Where they come from
  - What do they mean
- Networks
  - What are they
  - Analysis
  - Use in Biology
  - Biological questions/applications



# Networks

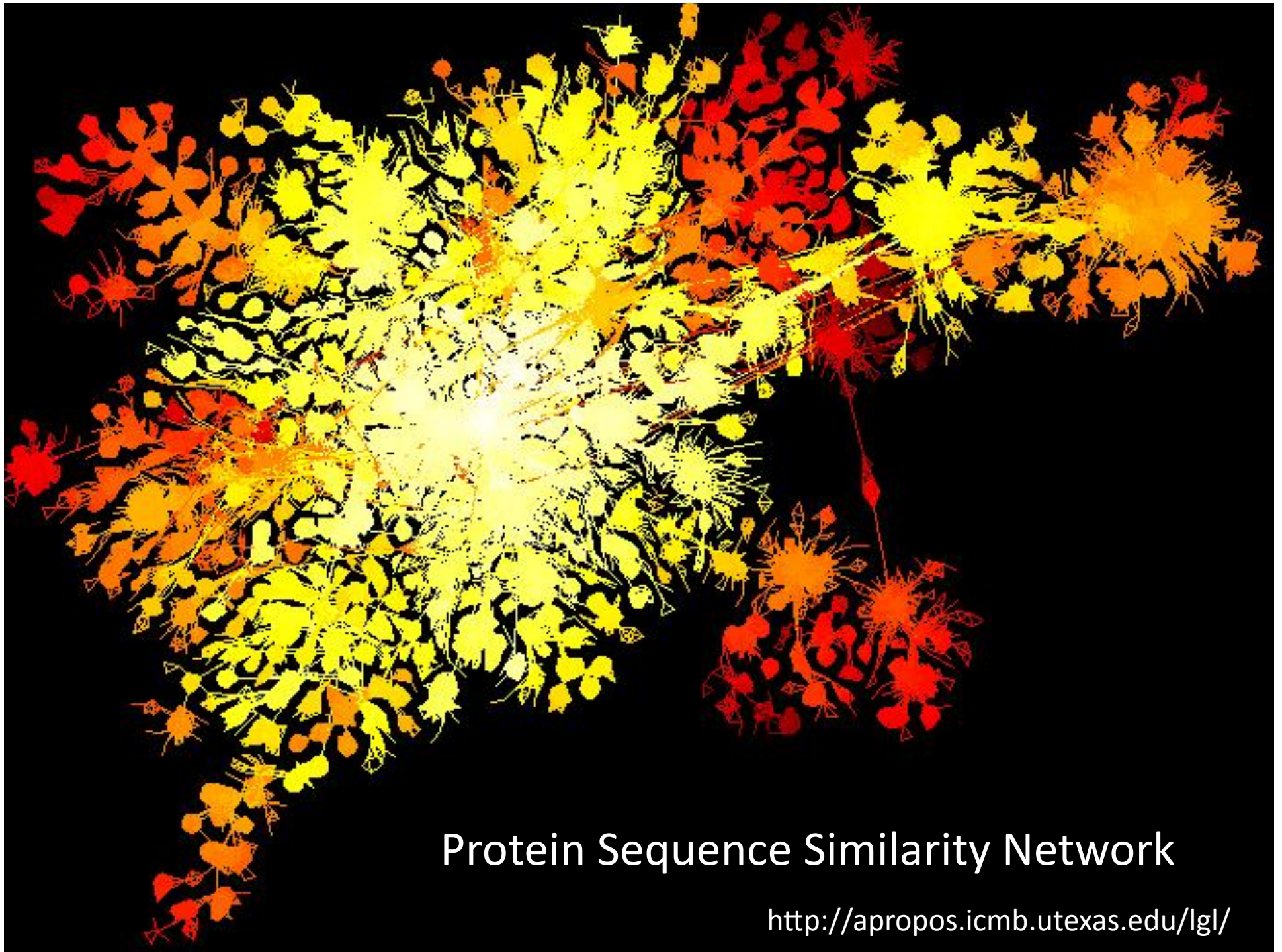
- Represent relationships
  - Physical, regulatory, genetic, functional interactions
- Useful for discovering relationships in large data sets
  - Better than tables in Excel
- Visualize multiple data types together
  - See interesting patterns





# Mapping Biology to a Network

- A simple mapping
  - one compound/node, one interaction/edge
- A more realistic mapping
  - Cell localization, cell cycle, cell type, taxonomy
  - Only represent physiologically relevant interaction networks
- Edges can represent other relationships
- **Critical**: understand the mapping for network analysis



# Protein Sequence Similarity Network

<http://apropos.icmb.utexas.edu/lgl/>



# Six Degrees of Separation

- Everyone in the world is connected by at most six links
- Which path should we take?
- Shortest path by breadth first search
  - If two nodes are connected, will find the shortest path between them
- Are two proteins connected? If so, how?
- Biologically relevant?





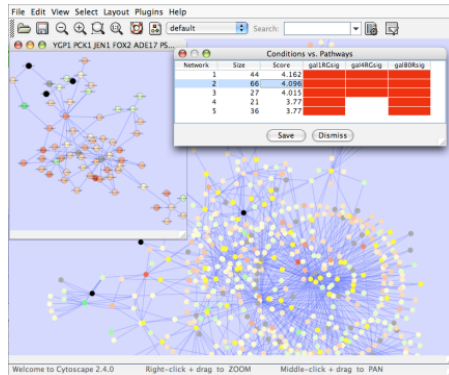
# Biological Questions

- Step 1: What do you want to accomplish with your list or network (hopefully part of experiment design! 😊 )
  - Summarize biological processes or other aspects of gene function
  - Find a controller for a process (TF, miRNA)
  - Find new pathways or new pathway members
  - Discover new gene function
  - Correlate with a disease or phenotype (candidate gene prioritization)
  - Perform differential analysis – what's different between samples?

Other  
Questions?

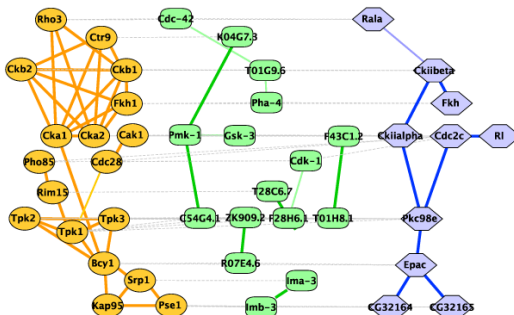


# Applications of Network Biology



jActiveModules, UCSD

[b] Phosphorus metabolism  
Complexes 32, 296, 728, 822, 894, 927

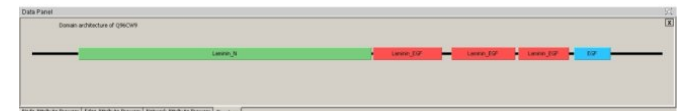
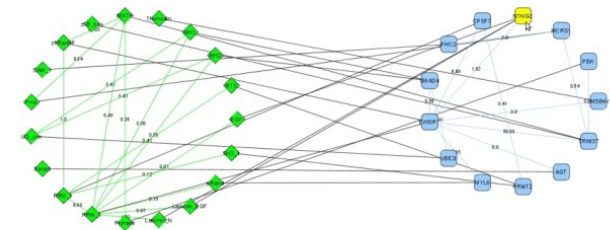


PathBlast, UCSD

- **Gene Function Prediction** – shows connections to sets of genes/proteins involved in same biological process
- **Detection of protein complexes/ other modular structures** – discover modularity & higher order organization (motifs, feedback loops)
- **Network evolution** – biological process(es) conservation across species
- **Prediction of new interactions and functional associations** – Statistically significant domain-domain correlations in protein interaction network to predict protein-protein or genetic interaction

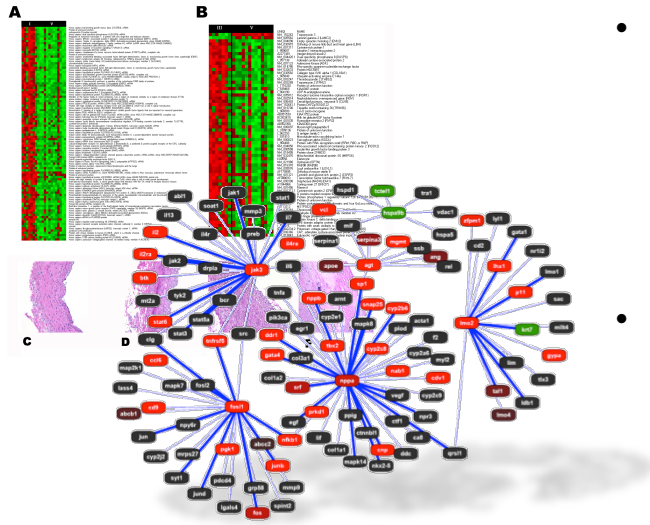


MCODE, University of Toronto

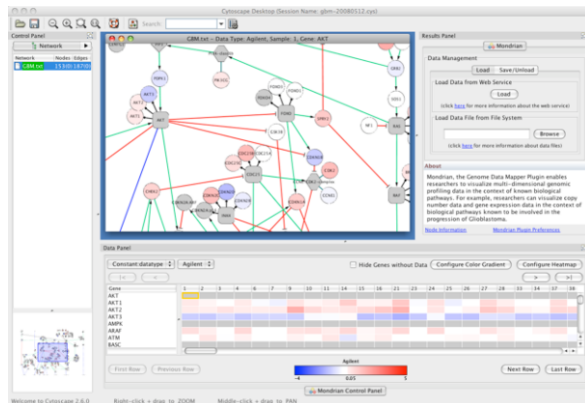


DomainGraph, Max Planck Institute

# Applications of Network Informatics in Disease

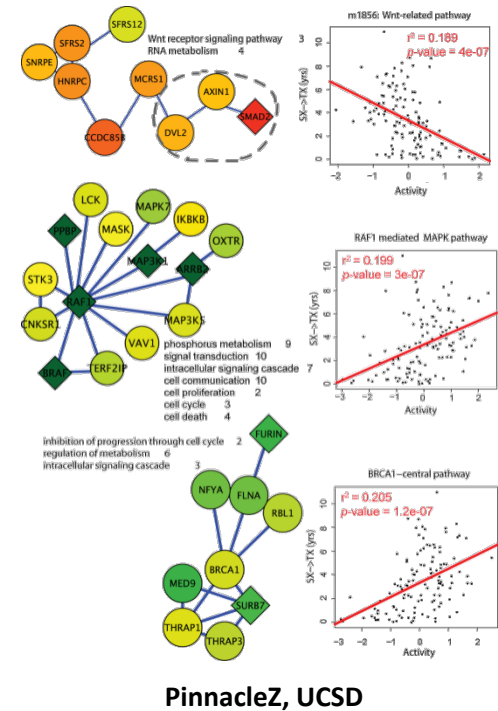


Agilent Literature Search



Mondrian, MSKCC

- **Identification of disease subnetworks** – identification of disease network subnetworks that are transcriptionally active in disease.
- **Subnetwork-based diagnosis** – source of biomarkers for disease classification, identify interconnected genes whose aggregate expression levels are predictive of disease state
- **Subnetwork-based gene association** – map common pathway mechanisms affected by collection of genotypes



PinnacleZ, UCSD





# What Have We Learned?

- Gene lists can be interpreted in the context of known biological processes
- Networks are useful for seeing relationships in large data sets
- Important to understand what the nodes and edges mean
- Important to define the biological question - know what you want to do with your gene list or network
- Many methods available for gene list and network analysis



# Introduction to Network Visualization

Gary Bader



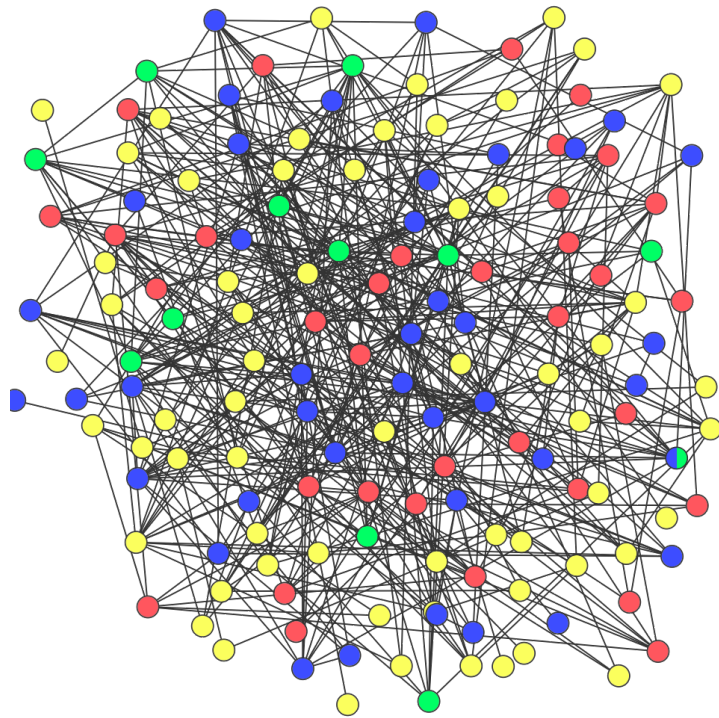
# Network Visualization Outline

- Automatic network layout
- Visual features
- Visually interpreting a network

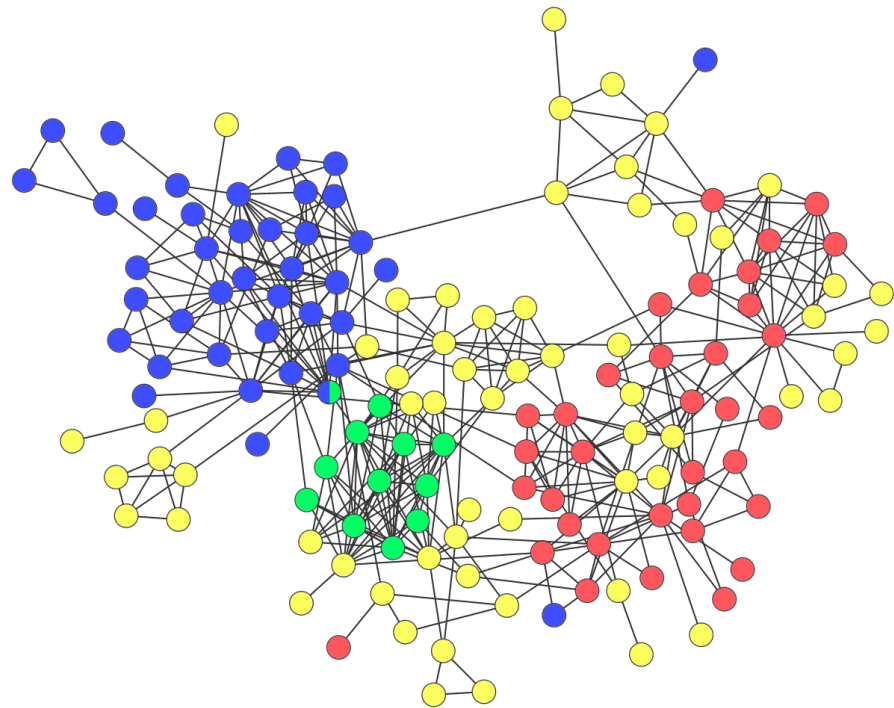


# Automatic network layout

Before layout



After layout



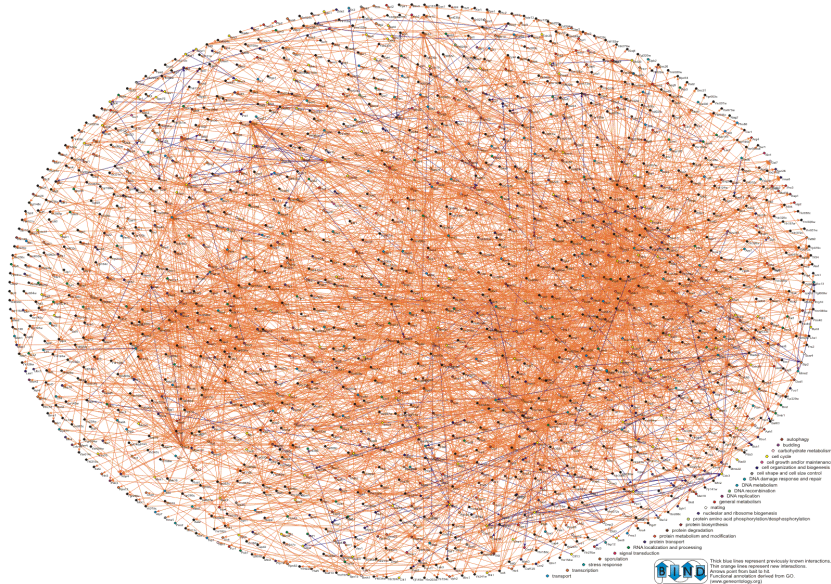


# Automatic network layout

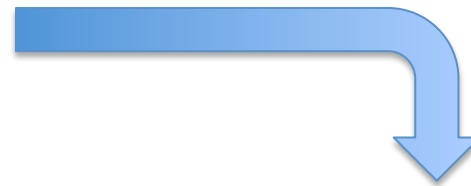
- Force-directed: nodes repel and edges pull
- Good for up to 500 nodes
  - Bigger networks give hairballs - Reduce number of edges
- Advice: try force directed first, or hierarchical for tree-like networks
- Tips for better looking networks
  - Manually adjust layout
  - Load network into a drawing program (e.g. Illustrator) and adjust labels

# Overview

Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry



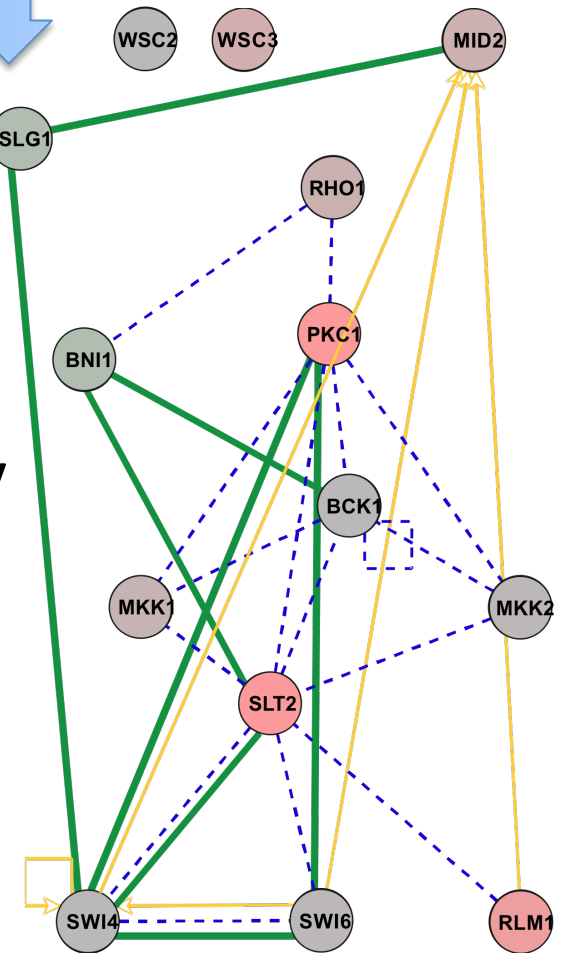
## Zoom



## Focus

### PKC Cell Wall Integrity

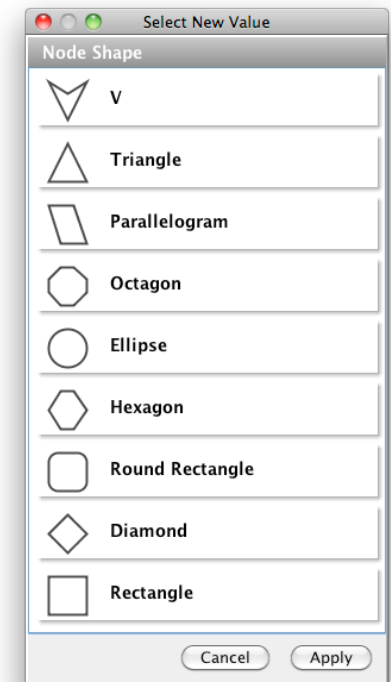
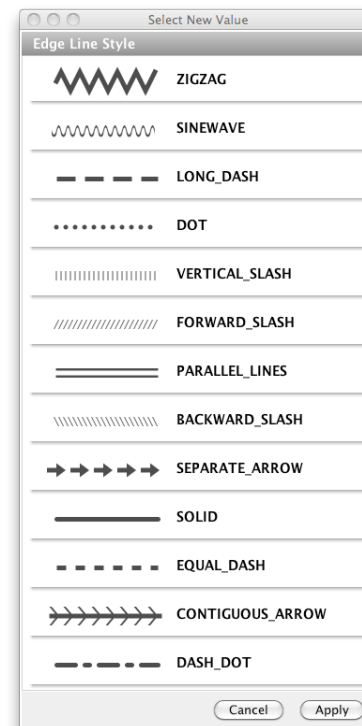
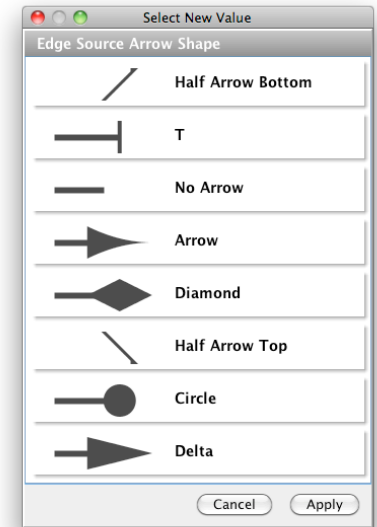
-  Synthetic Lethal
-  Transcription Factor Regulation
-  Protein-Protein Interaction
-  Up Regulated Gene Expression
-  Down Regulated Gene Expression





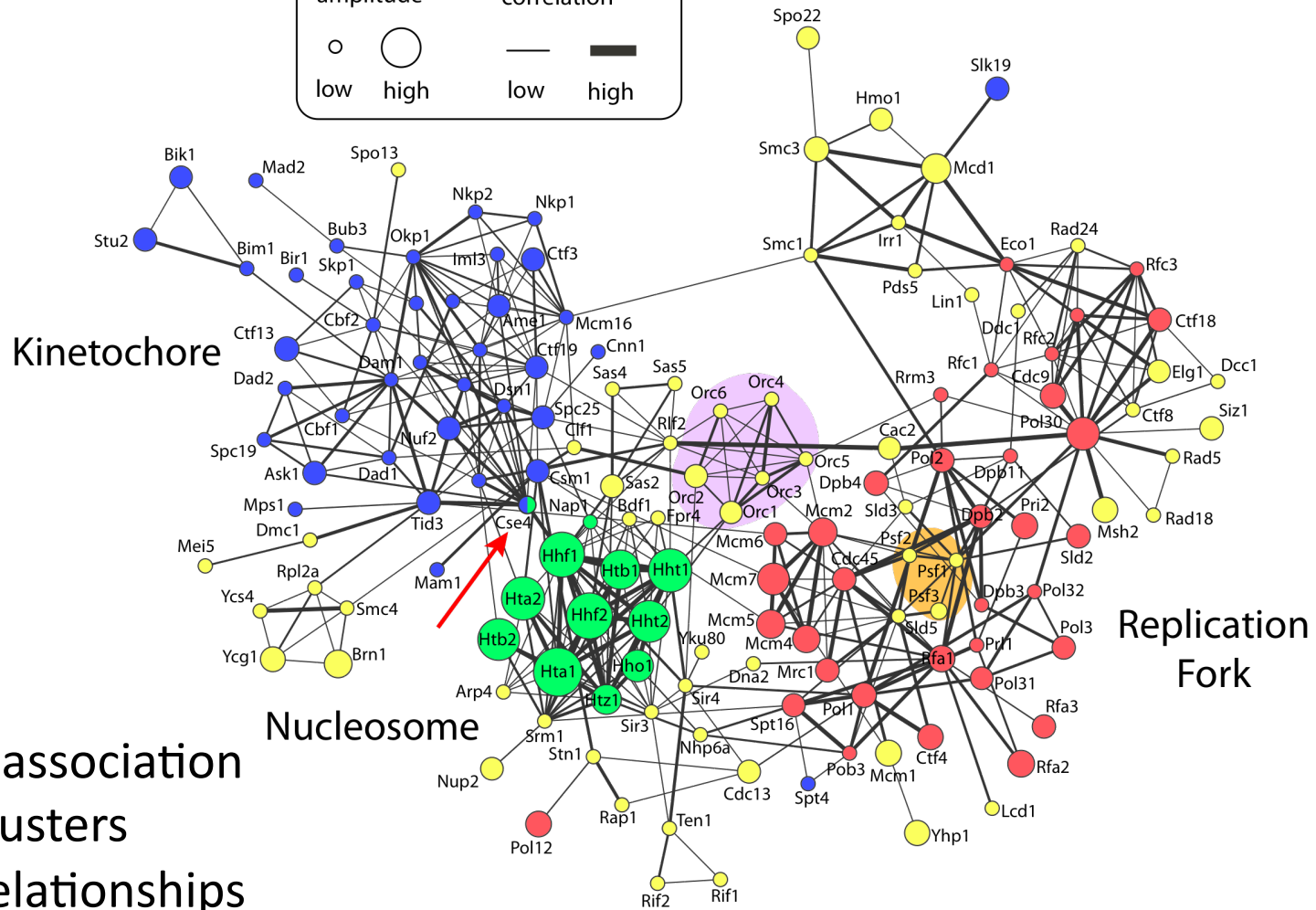
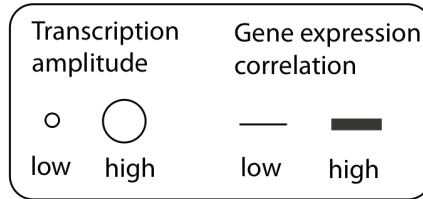
# Visual Features

- Node and edge attributes
  - String, integer, float, Boolean, list
  - E.g. represent gene, interaction attributes
- Visual attributes
  - Node, edge visual properties
  - Colour, shape, size, borders, opacity...





# Visual Features

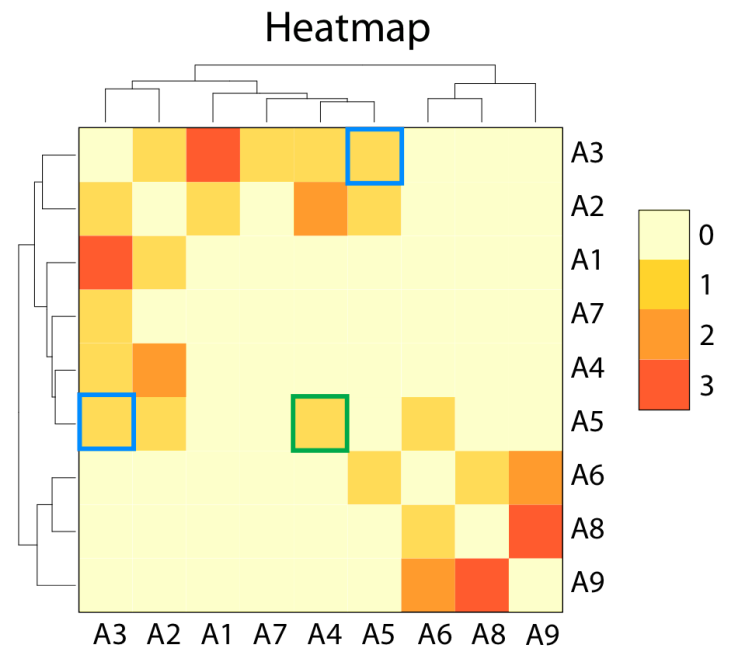
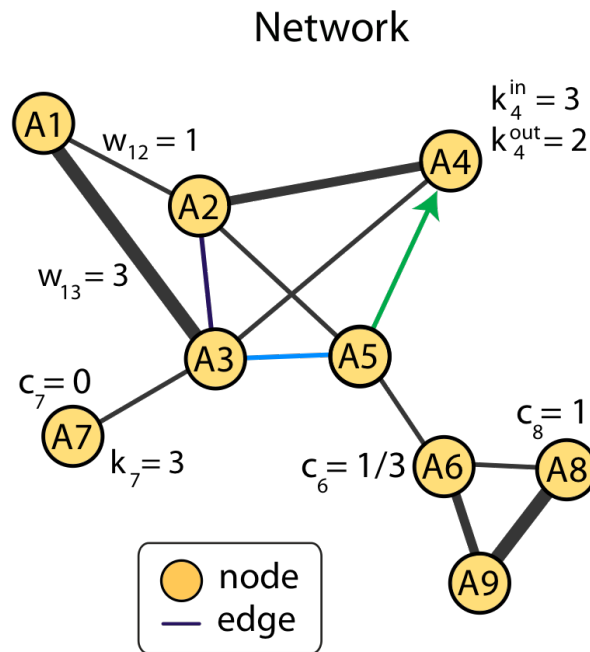






# Network Representations

Relationships	Optional weight
A1 ↔ A2	1
A1 ↔ A3	3
A2 ↔ A3	1
A2 ↔ A4	2
A2 ↔ A5	1
A3 ↔ A4	1
A3 ↔ A5	1
A3 ↔ A7	1
A5 → A4	1
A5 ↔ A6	1
A6 ↔ A8	1
A6 ↔ A9	2
A8 ↔ A9	3





# What Have We Learned?

- Automatic layout is required to visualize networks
- Networks help you visualize interesting relationships in your data
- Avoid hairballs by focusing analysis
- Visual attributes enable multiple types of data to be shown at once – useful to see their relationships



# Network Visualization/Analysis using Cytoscape

- Network visualization and analysis using Cytoscape software
- Cytoscape basics
- Cytoscape network analysis examples

<http://cytoscape.org>

# Network visualization and analysis

Pathway comparison  
Literature mining  
Gene Ontology analysis  
Active modules  
Complex detection  
Network motif search

UCSD, ISB, Agilent,  
MSKCC, Pasteur, UCSF,  
Unilever, UToronto, U  
Texas

The screenshot displays the Cytoscape Desktop interface. The main window shows a network graph with nodes of varying sizes and colors (yellow, green, red, grey) connected by edges. The interface includes several panels:

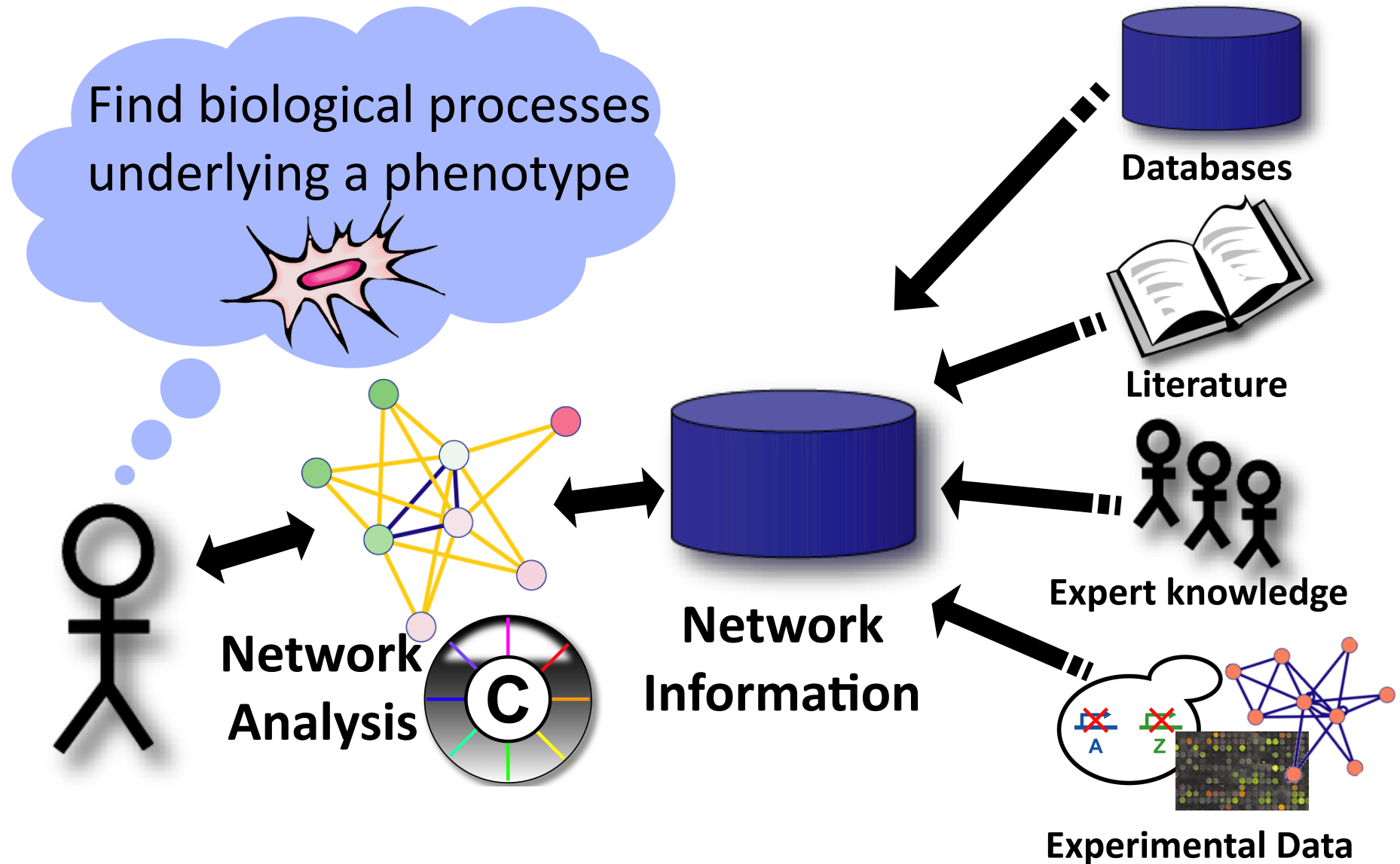
- Control Panel:** Shows the current visual style (Sample3), defaults (Source-Target), and visual mapping browser. The mapping browser shows continuous mappings for Node Color (gal1RGexp) and Edge Color (interaction).
- Tool Panel:** Contains tools for Rotate, Scale, and Align and Distribute.
- Data Panel:** Displays a table of node attributes.

ID	Degree	annotation.GO BIOLOGICAL_PROCESS	gal1RGexp
YER054C	2	[glycogen metabolic process, protein amino acid dep...	0.057
YJL219W	2	[hexose transport]	0.298
YBR045C	3	[regulation of phosphoprotein phosphatase activity, s...	0.786
YGR058W	2	[biological_process]	0.045
YPL149W	2	[autophagy, protein targeting to vacuole]	0.033
VUD171W	3	[C-terminal protein ligation, autophagy, protein mo...	-0.134

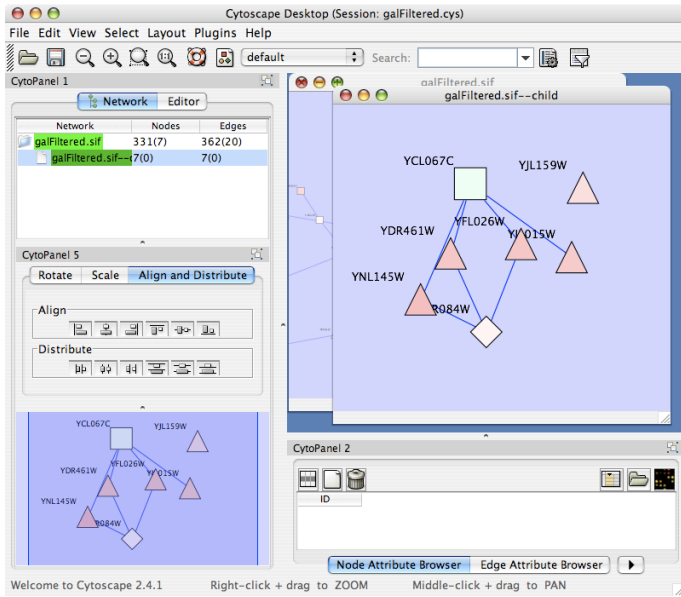
At the bottom, there are buttons for Node Attribute Browser, Edge Attribute Browser, and Network Attribute Browser. The status bar at the very bottom reads: "Welcome to Cytoscape 2.5 Right-click + drag to ZOOM Middle-click + drag to PAN".



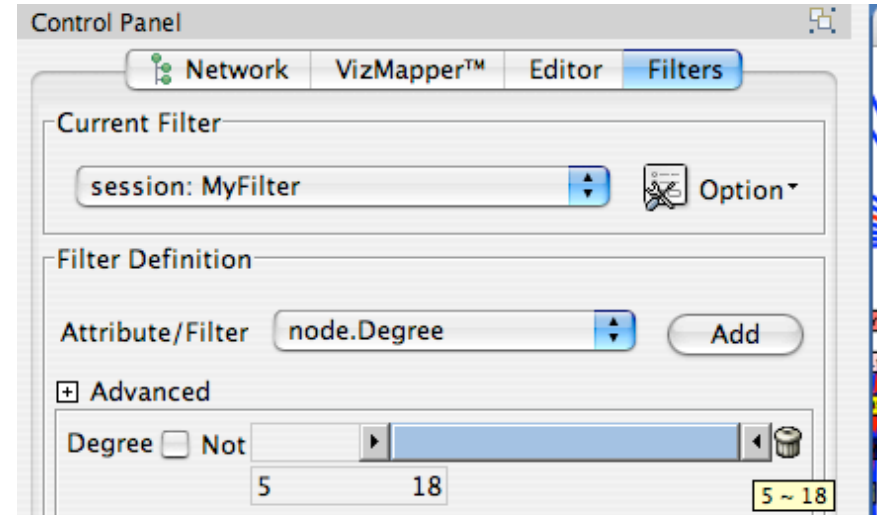
# Network Analysis using Cytoscape



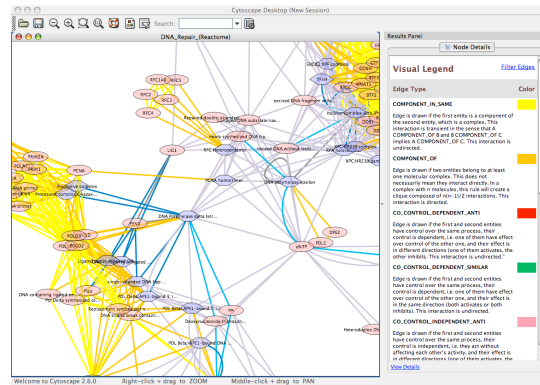
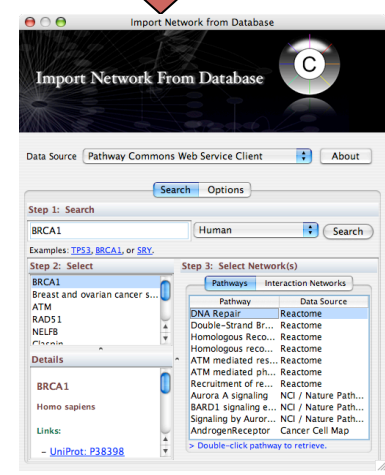
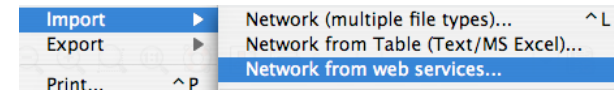
# Manipulate Networks



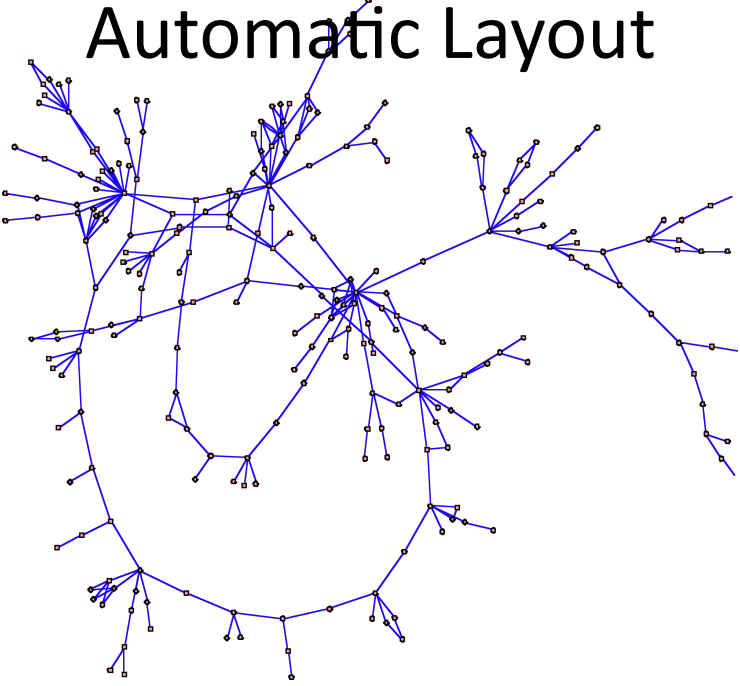
# Filter/Query



# Interaction Database Search



# Automatic Layout





# Active Community

<http://www.cytoscape.org>

- Help
  - 8 tutorials, >10 case studies
  - Mailing lists for discussion
  - Documentation, data sets
- Annual Conferences
- 10,000s users, 2500 downloads/month
- >80 Plugins Extend Functionality
  - Build your own, requires programming

Cline MS et al. Integration of biological networks and gene expression data using Cytoscape  
Nat Protoc. 2007;2(10):2366-82



# Cytoscape Demo

Gary Bader

Version 2.7

[www.cytoscape.org](http://www.cytoscape.org)





# Working with Data

- Loading Networks
- Loading and Manipulating Attributes



# Loading Networks

The screenshot displays the Cytoscape Desktop application interface. The main window shows a network graph with four nodes (node0, node1, node2, node3) and their connections. The nodes are represented as red circles, and the edges are blue lines. The graph is titled "Network.txt".

The interface includes a Control Panel on the left, a Data Panel at the bottom, and a Visual Mapping Browser. The Control Panel shows the current visual style is "default" and includes a Defaults section with a preview of two nodes connected by an edge. The Visual Mapping Browser shows the "Node Label" property mapped to "ID" using "Passthrough Mapping".

The Data Panel shows a table with the following data:

ID

At the bottom of the window, there are three tabs: "Node Attribute Browser", "Edge Attribute Browser", and "Network Attribute Browser". The "Node Attribute Browser" tab is currently selected.

At the bottom of the window, there is a welcome message: "Welcome to Cytoscape 2.8.0" and instructions: "Right-click + drag to ZOOM" and "Middle-click + drag to PAN".



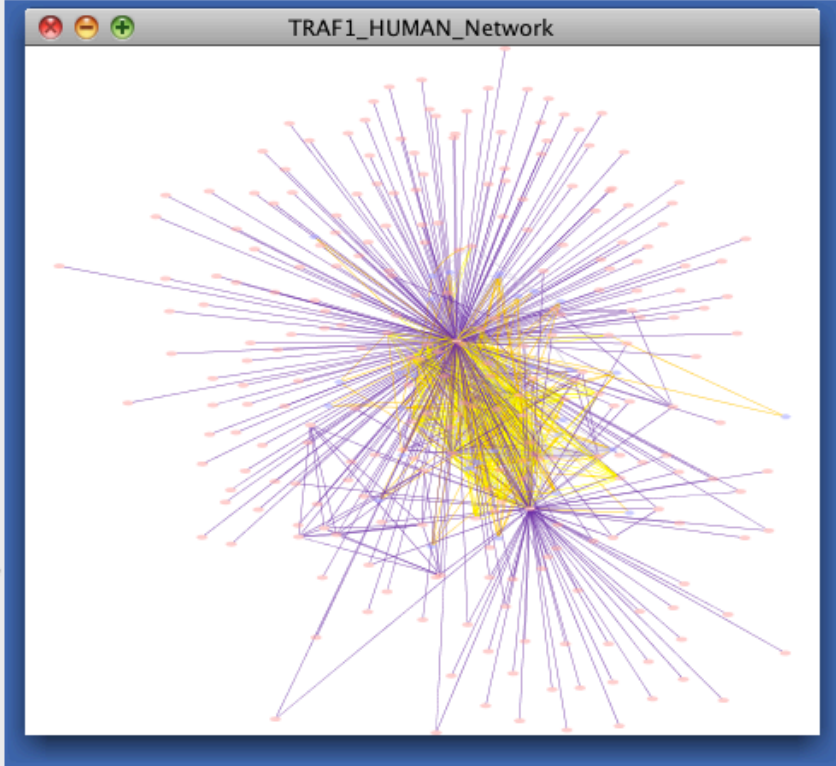
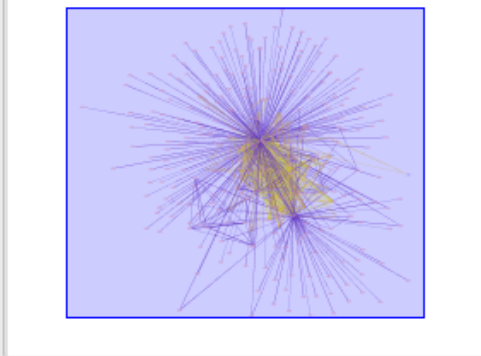
Cytoscape File Edit View Select Layout Plugins Help

Cytoscape Desktop (New Session)

Control Panel Search:

Control Panel Network VizMapper™

Network	Nodes	Edges
TRAF1_HUMAN_Network	288(0)	952(0)



Results Panel

Node Details

Visual Legend [Filter Edges](#)

Edge Type

**IN\_SAME\_COMPONENT**

Edge is drawn if the first component of the second complex. This interaction sense that A COMPONENT\_OF C implies COMPONENT\_OF C. This is undirected.

**COMPONENT\_OF**

Edge is drawn if two entities share at least one molecular component. This necessarily means they interact. A complex with n molecules creates a clique component.

[View Details](#)

Tool Panel Rotate Scale Align and Distribute

Scale:

1/8 1/4 1/2 1 2 4 8

Reset scale bar

Scale Selected Nodes Only

Data Panel

ID	Symbol
----	--------

Node Attribute Browser Edge Attribute Browser Network Attribute Browser



### Import Annotation File

## Import Attribute from Table

**Data Sources**

Attributes:  Node  Edge  Network

Input File:

**Advanced**

Show Mapping Options  Show Text File Import Options  Import everything (Key is always ID)  Case Sensitive

**Annotation File to Attribute Mapping**

Key Column in Annotation File

Primary Key:  Key Attribute for Network

Alias?	Column (Attribute Name)	Data Type
<input type="checkbox"/>	Source	a String
<input type="checkbox"/>	<b>Target</b>	a String
<input type="checkbox"/>	Source Symbol	a String
<input type="checkbox"/>	Symbol	a String
<input type="checkbox"/>	Interaction	a String

**Preview**

Text File Left Click: Enable/Disable Column, Right Click: Edit Column Legend: Key Alias

<input checked="" type="checkbox"/> Source	<input checked="" type="checkbox"/> Target	<input checked="" type="checkbox"/> Source Symbol	<input checked="" type="checkbox"/> Symbol	<input checked="" type="checkbox"/> Interaction
node0	node1	Node 0	Node 1	inhibits
node0	node3	Node 0	Node 3	activates
node0	node2	Node 0	Node 2	interacts
node2	node3	Node 2	Node 3	interacts

Key Attributes

- node0
- node1
- node2
- node3

Key Matched: 4 | First 100 entries are loaded for preview. | File Size: 213 Bytes



**Cytoscape** File Edit View Select Layout Plugins Help  
Cytoscape Desktop (Session Name: Network.cys)

Control Panel

Network Nodes Edges  
Network.txt 4(4) 4(4)

Network.txt

```
graph TD; node2 --- node3; node2 --- node0; node3 --- node0; node0 --- node1;
```

Data Panel

ID	Symbol
node2	Node 2
node0	Node 0
node1	Node 1
node3	Node 3

Node Attribute Browser | Edge Attribute Browser | Network Attribute Browser

Welcome to Cytoscape 2.7.0 Right-click + drag to ZOOM Middle-click + drag to PAN

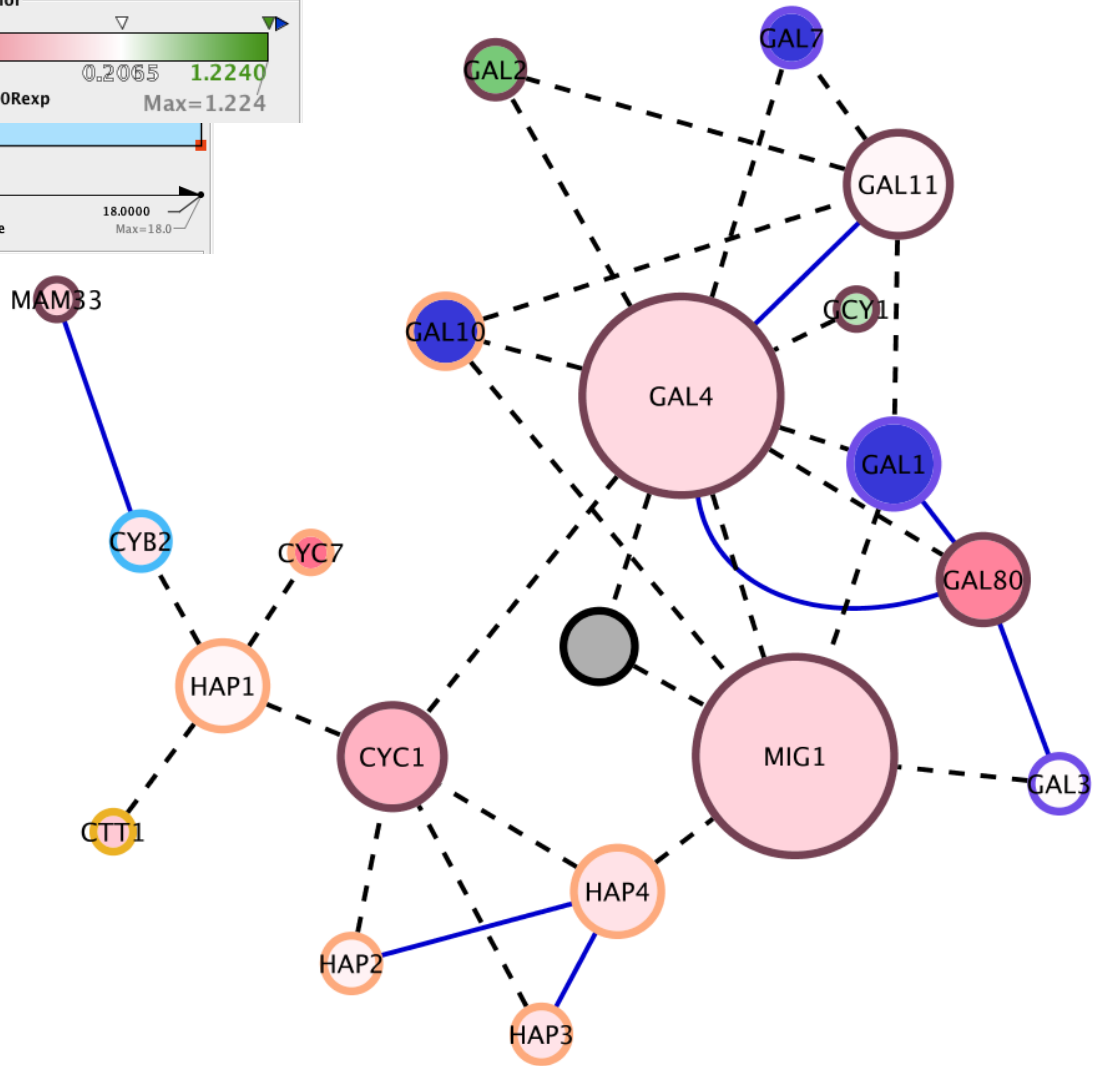
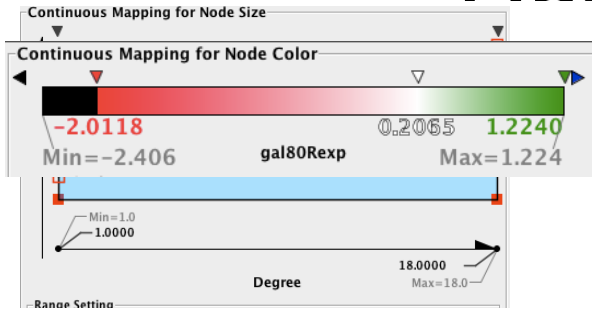


# Visualizing data

- Data mapping
- Layouts



# Data mapping





# Data mapping

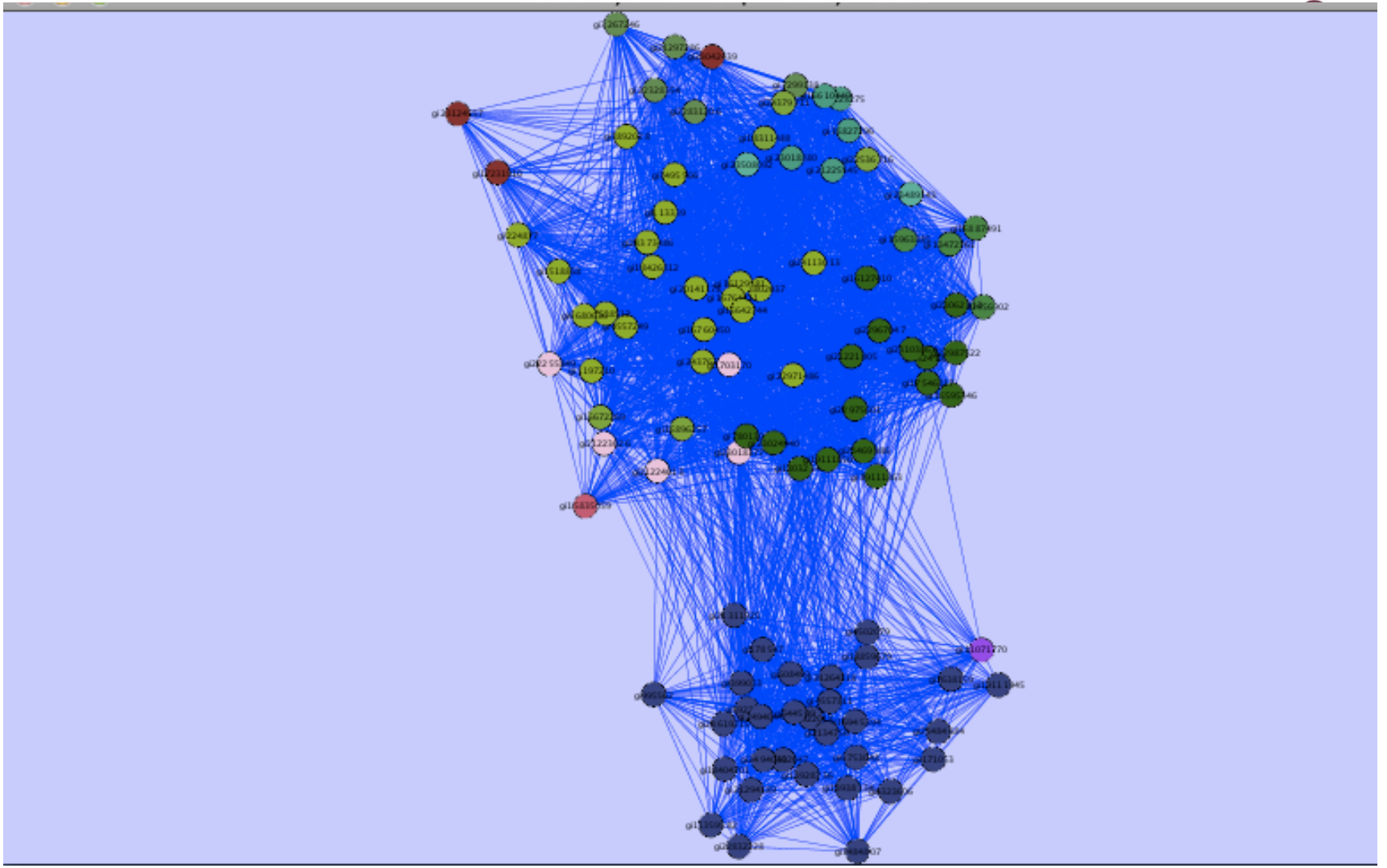
- Avoid cluttering your visualization with too much data
  - Map the data you are specifically interested in to call out meaningful differences
  - Mapping too much data to visual attributes may just confuse the viewer
  - Can always create multiple networks and map different values







# Network





# Layouts

- Use layouts to convey the relationships between the nodes
- Layout algorithms may need to be “tuned” to fit your network
  - Layouts → Settings... menu
- Lots of parameters to change layout algorithm behavior
- Can also consider laying out portions of your network



# Lab Time

- Try out Cytoscape
- [http://opentutorials.rbvi.ucsf.edu/index.php/Tutorial:Introduction\\_to\\_Cytoscape](http://opentutorials.rbvi.ucsf.edu/index.php/Tutorial:Introduction_to_Cytoscape)
- Survey – Developers
  - Beginner, have already developed plugin?



FYI

Cytoscape Desktop (Session: galFiltered.cys)

galFiltered.sif

Control Panel

Network

Network Nodes Edges

galFiltered.sif 331(4) 362(0)

Desktop

Canvas

Attribute browser

Data Panel

ID	Degree	gal1RGexp	gal1RGsig	gal4RGexp	gal4RGsig	gal80Rexp	gal80Rsig
YAL038W	3	-0.652	1.3173E-10	0.123	0.10377	-0.453	1.5489E-7
YAL003W	2	-0.157	0.0018696	-0.2	6.2814E-4	-0.146	0.013062
YAL030W	2	-0.05	0.18782	0.027	0.55374	0.0030	0.95396

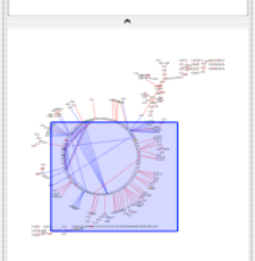
Node Attribute Browser | Edge Attribute Browser | Network Attribute Browser

Welcome to Cytoscape 2.6.0 Right-click + drag to ZOOM Middle-click + drag to PAN

Network manager

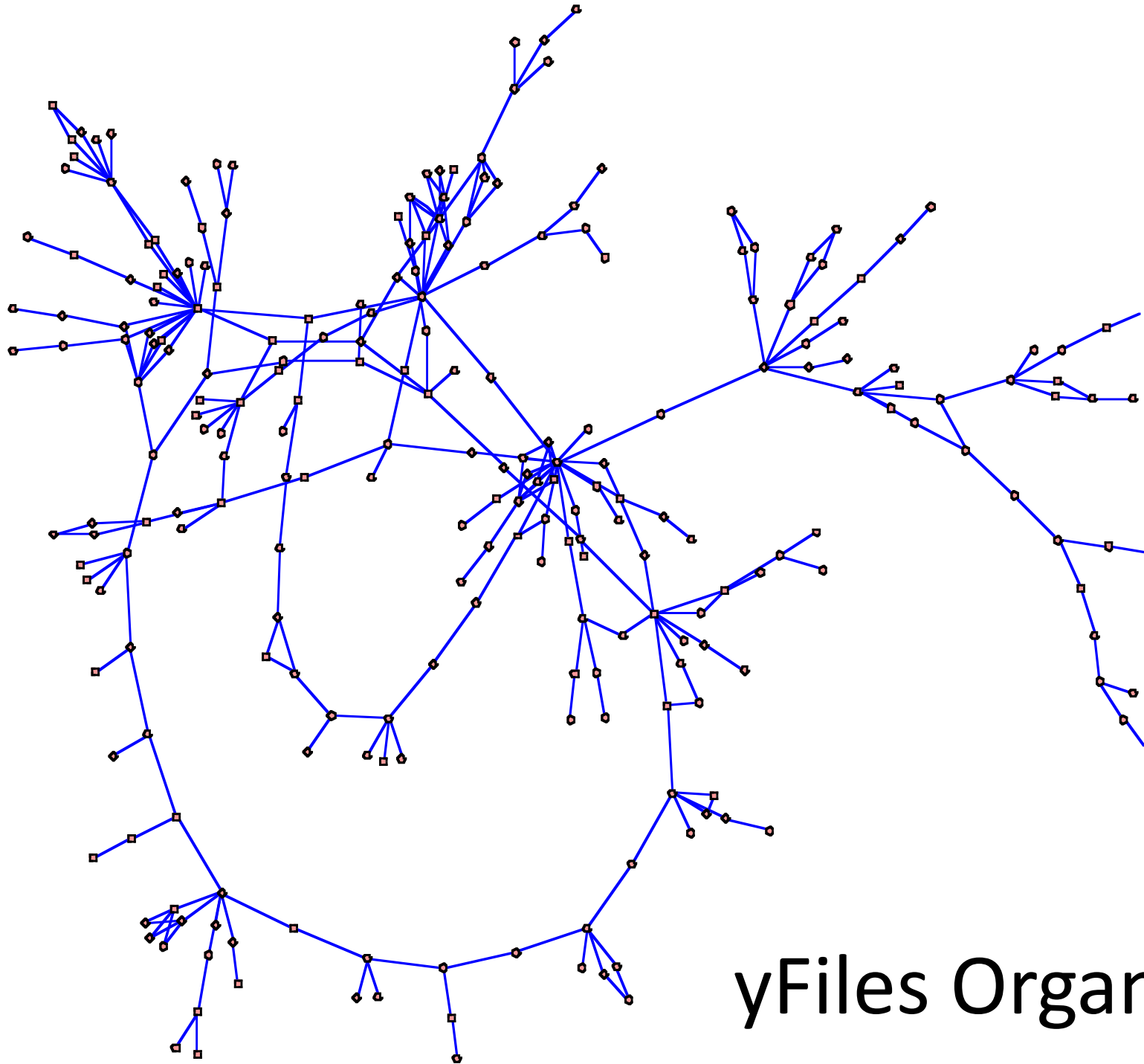
CytoPanels

Network overview





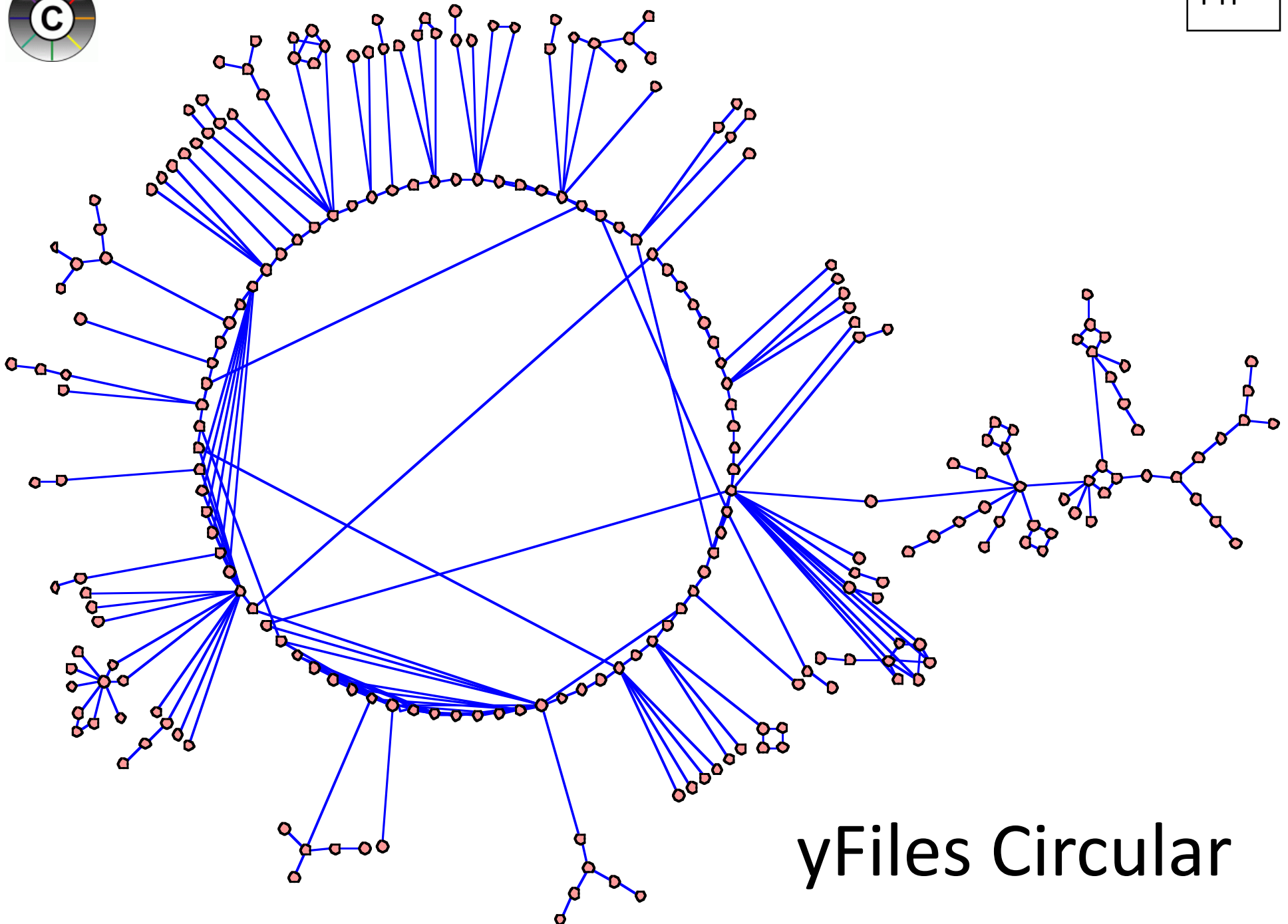
FYI



yFiles Organic



FYI

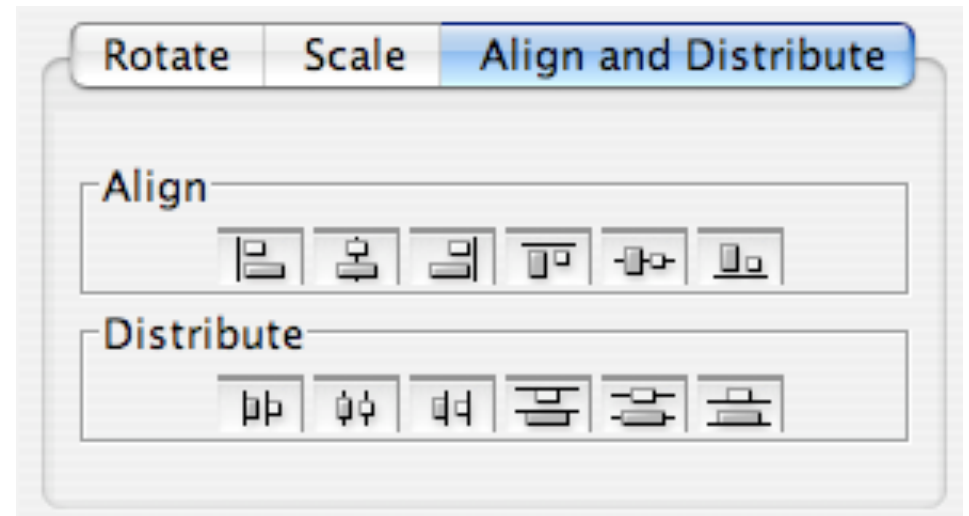
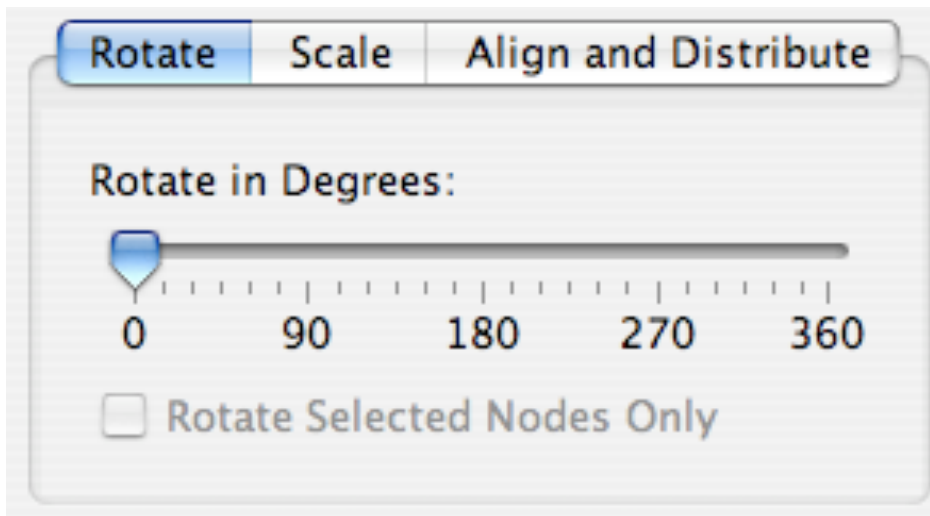


yFiles Circular



# Network Layout

- 15 algorithms available through plugins
- Demo: Move, zoom/pan, rotate, scale, align







# Create Subnetwork

FYI

Cytoscape Desktop (Session: galFiltered.cys)

File Edit View Select Layout Plugins Help

New Network From selected nodes, all edges ^N  
Open ^O Session From selected nodes, selected edges ^⇧N  
Save ^S Clone current network  
Save As... ^⇧S Network Editor Empty Network  
Import Construct network using cPath...  
Export network  
Print... ^P Nodes  
Quit ^Q Edges

331(7) 362(20)

CytoPanel 5

Rotate Scale **Align and Distribute**

Align

Distribute

CytoPanel 2

ID  
YFL026W  
YDR461W  
YNL145W

Node Attribute Browser Edge Attribute Browser

Welcome to Cytoscape 2.4.1 Right-click + drag to ZOOM Middle-click + drag to PAN



# Create Subnetwork

FYI

Cytoscape Desktop (Session: galFiltered.cys)

File Edit View Select Layout Plugins Help

default Search:

CytoPanel 1

Network	Nodes	Edges
galFiltered.sif	331(7)	362(20)
galFiltered.sif--(7(0))	7(0)	7(0)

CytoPanel 5

Rotate Scale **Align and Distribute**

Align

Distribute

galFiltered.sif galFiltered.sif--child

CytoPanel 2

ID

Node Attribute Browser Edge Attribute Browser

Welcome to Cytoscape 2.4.1 Right-click + drag to ZOOM Middle-click + drag to PAN



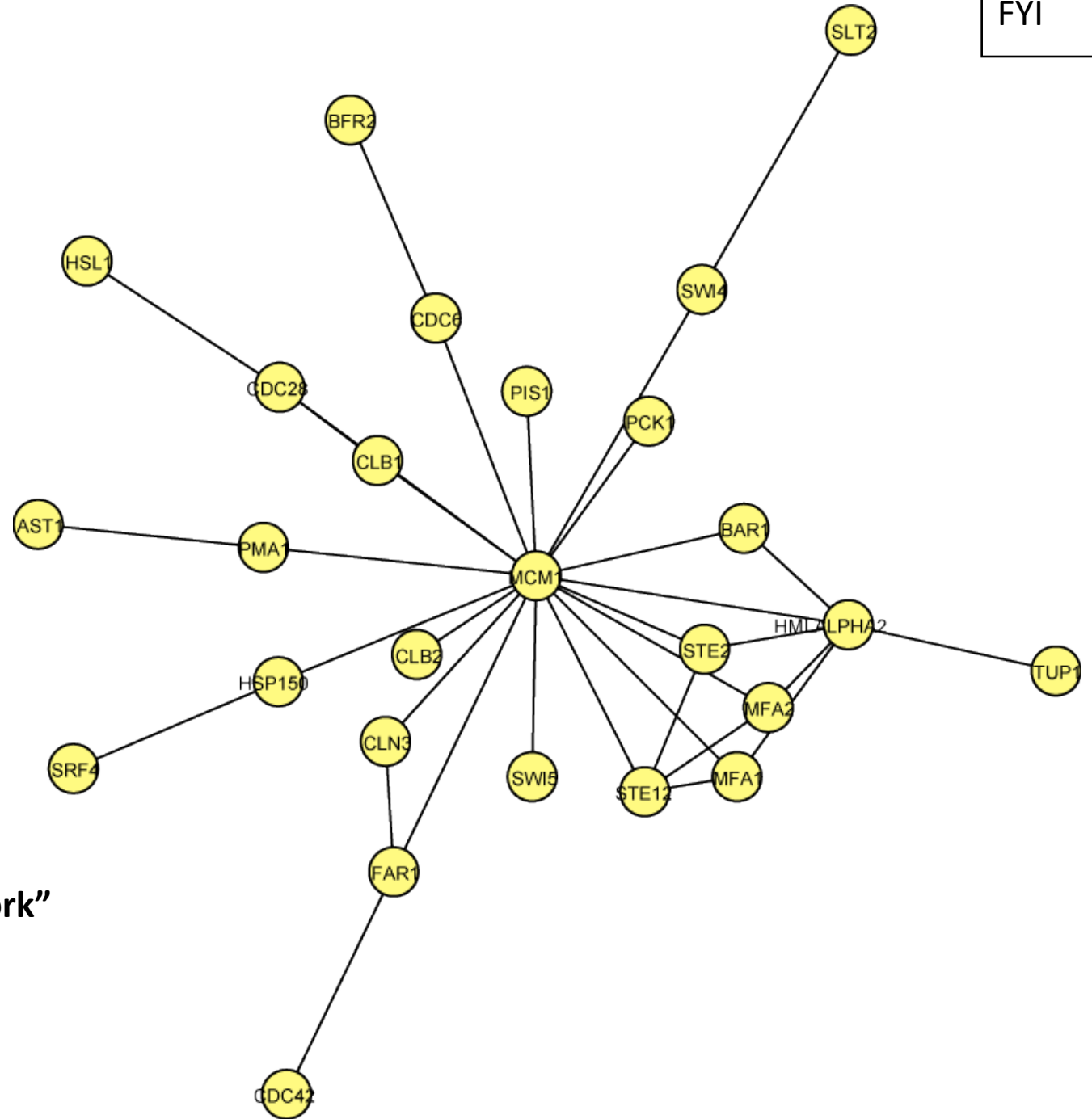
# Visual Style

- Customized views of experimental data in a network context
- Network has node and edge attributes
  - E.g. expression data, GO function, interaction type
- Mapped to visual attributes
  - E.g. node/edge size, shape, colour...
- E.g. Visualize gene expression data as node colour gradient on the network



# Visual Style

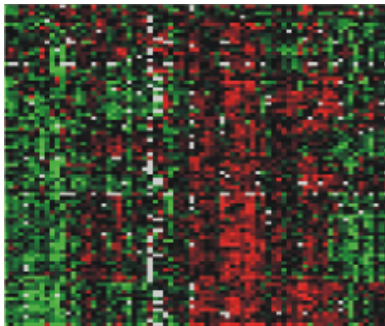
FYI



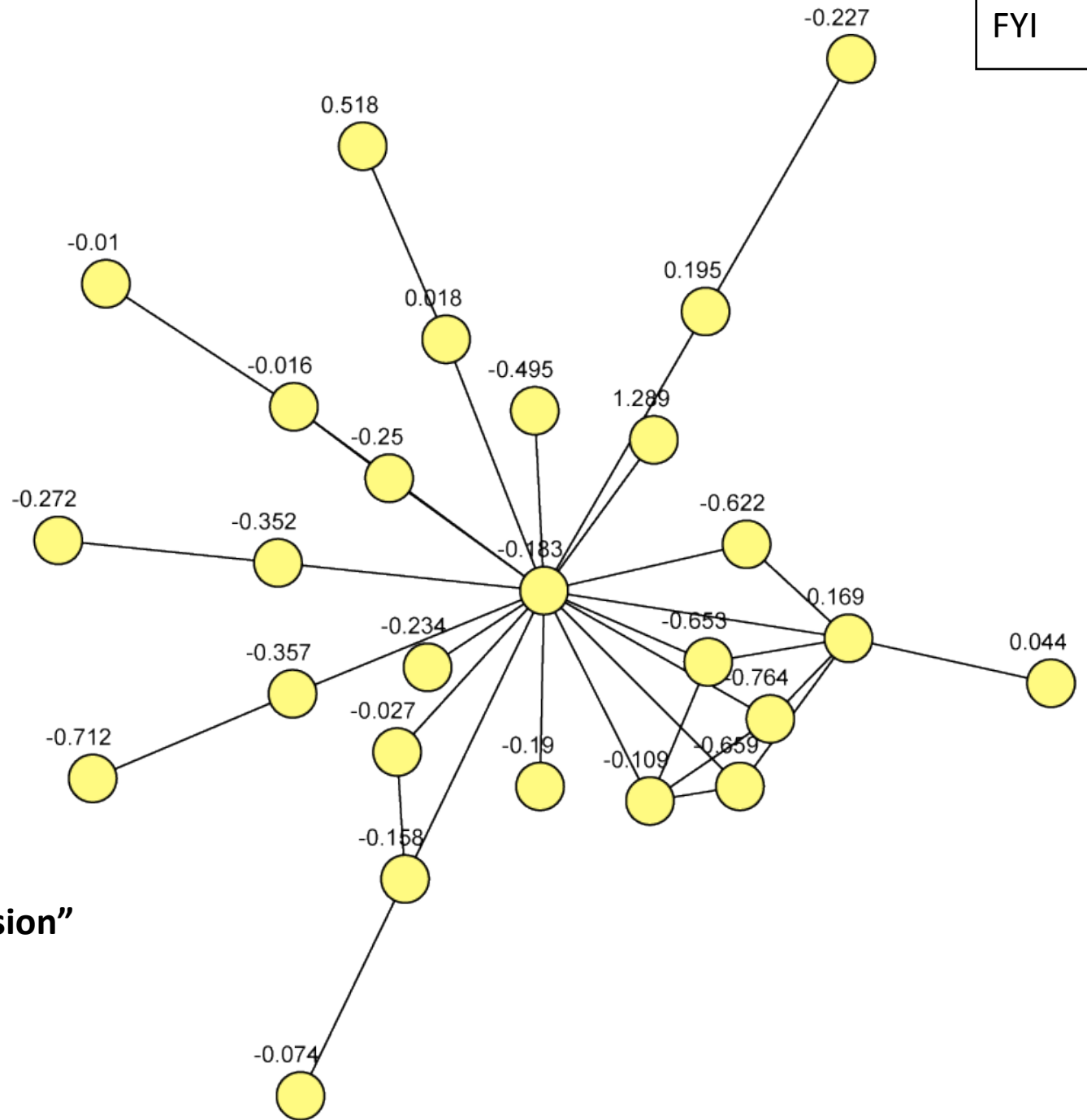
Load "Your Favorite Network"



# Visual Style



Load "Your Favorite Expression"  
Dataset



FYI



# Visual Style

FYI

Cytoscape Desktop (Session: galFiltered.cys)


Control Panel

Network VizMapper™ Editor Filters

Current Visual Style: Sample3

Defaults

Visual Mapping Browser

- Edge Visual Mapping
  - Edge Color: interaction
- Node Visual Mapping
  - Node Tooltip: gal4RGexp
  - Node Label: ID
  - Node Color: gal4RGexp
    - Mapping Type: Continuous Mapping
    - Graphical View:  -2.41 to 1.22
- Unused Properties
  - Node Border Color: Double-Click to create...
  - Node Shape: Double-Click to create...
  - Node Width: Double-Click to create...
  - Node Height: Double-Click to create...

galFiltered.sif

Gradient Editor for Node Color

Continuous Mapping for Node Color

Min = -2.406 Max = 1.224

gal4RGexp

Range Setting:

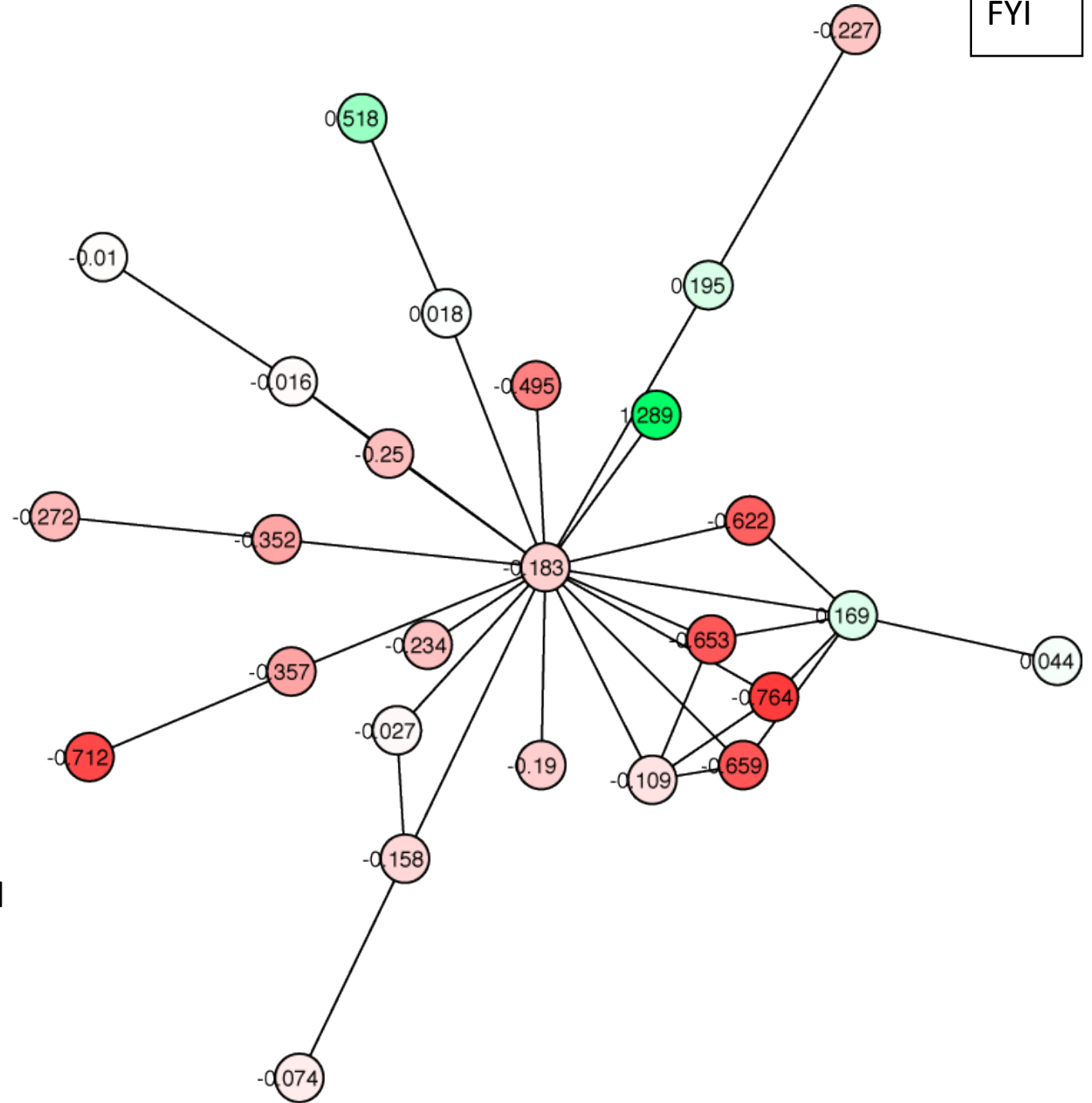
Min/Max Add Delete

Map expression values to node colours using a continuous mapper



# Visual Style

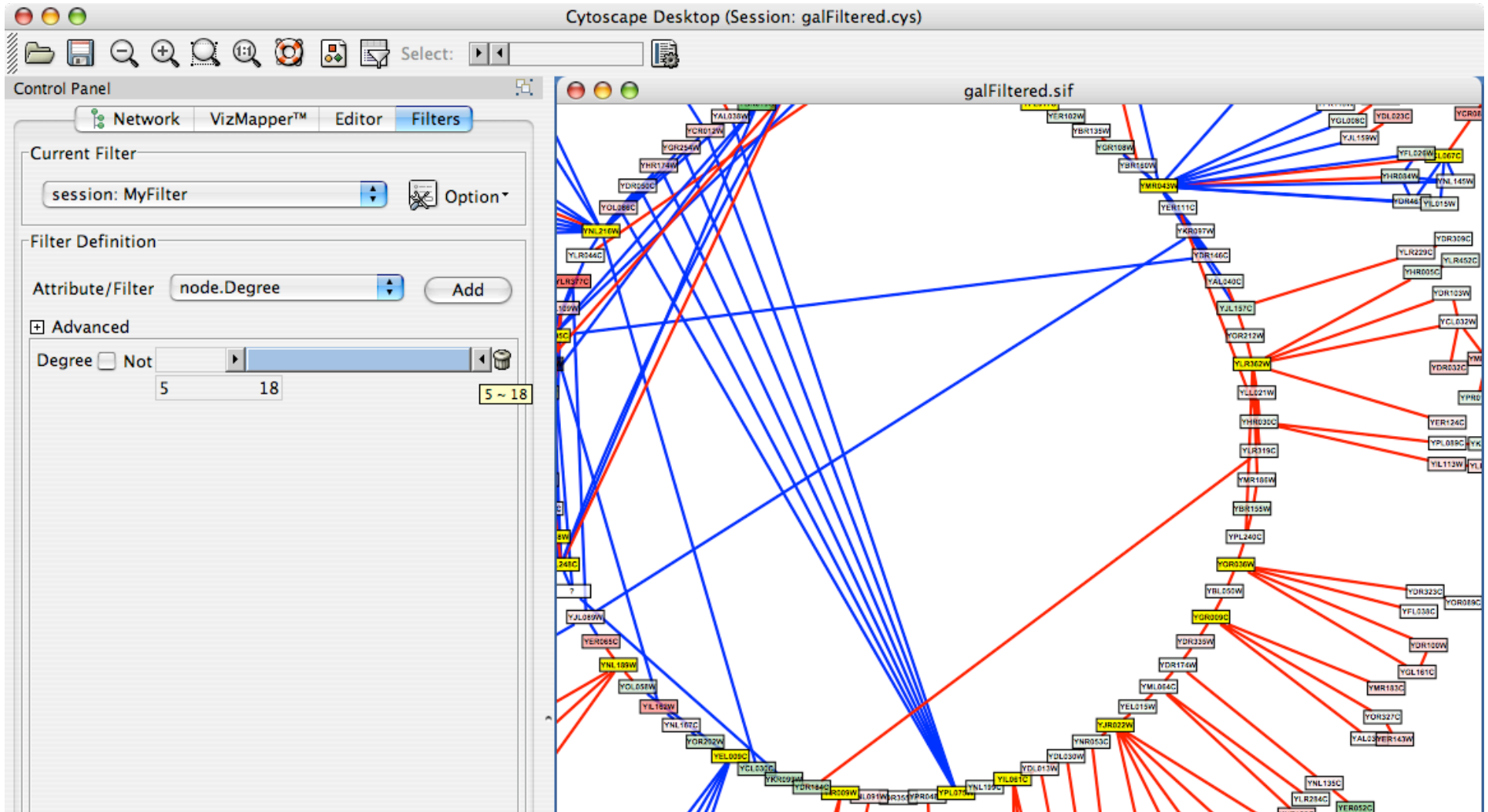
FYI





FYI

# Network Filtering

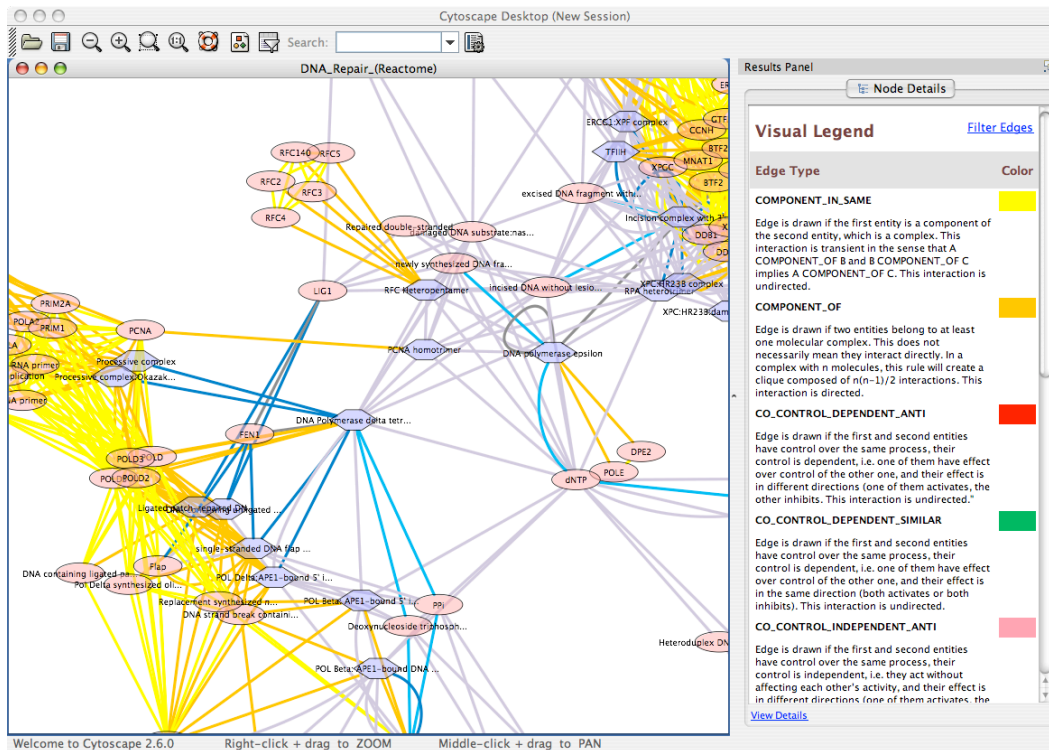
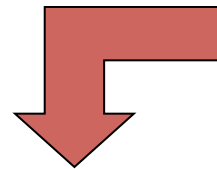
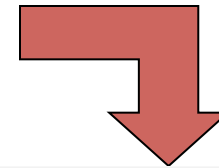
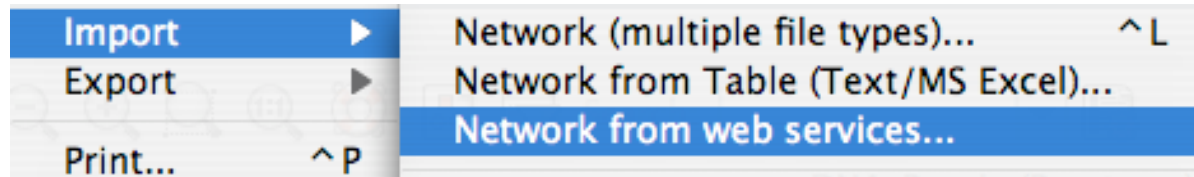






# Interaction Database Search

FYI



Data Source: Pathway Commons Web Service Client

Search: BRCA1 Human

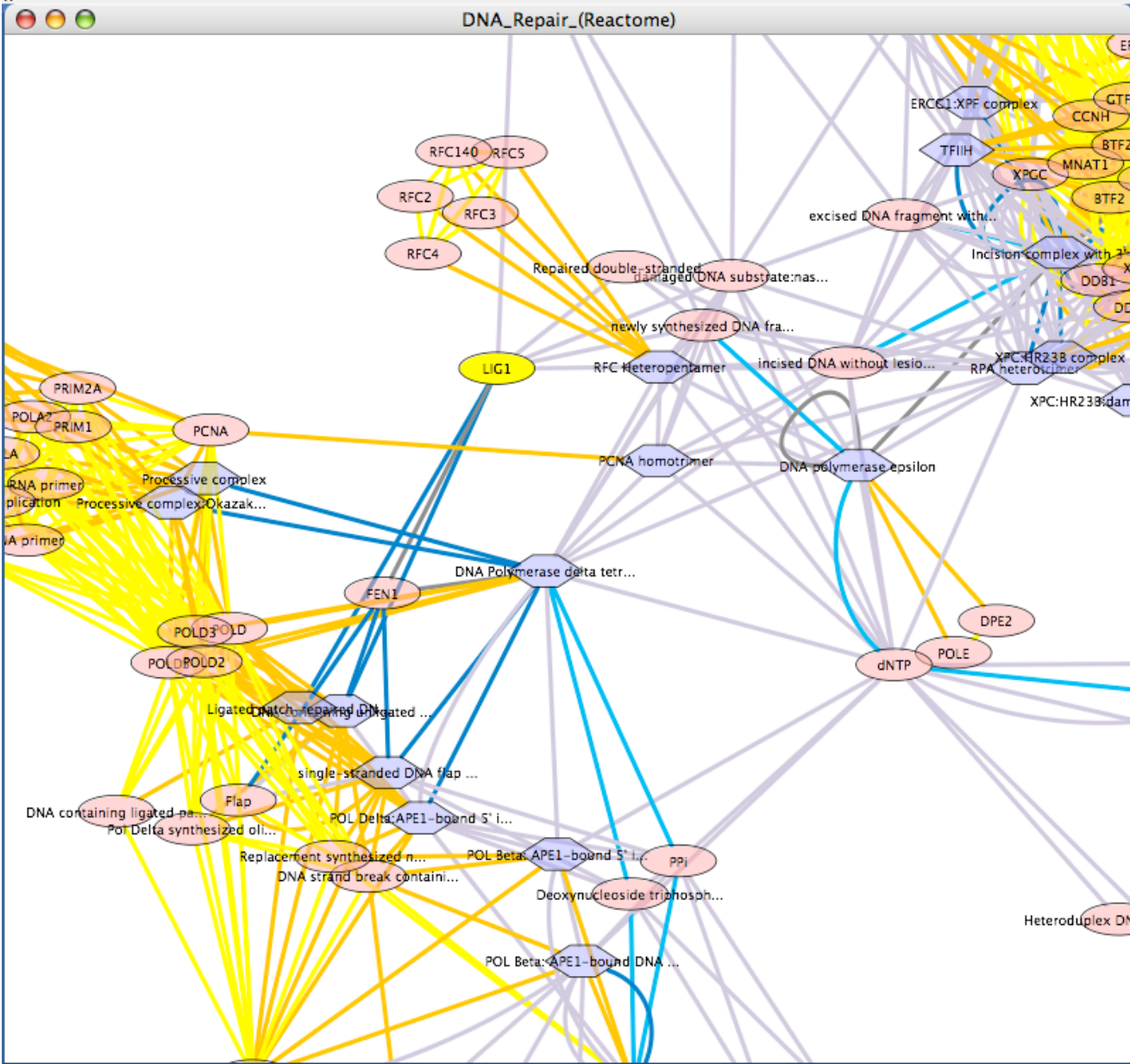
Examples: TP53, BRCA1, or SRY.

Step 2: Select

Step 3: Select Network(s)

Pathway	Data Source
DNA Repair	Reactome
Double-Strand Br...	Reactome
Homologous Reco...	Reactome
Homologous reco...	Reactome
ATM mediated res...	Reactome
ATM mediated ph...	Reactome
Recruitment of re...	Reactome
Aurora A signaling	NCI / Nature Path...
BARD1 signaling e...	NCI / Nature Path...
Signaling by Auror...	NCI / Nature Path...
AndrogenReceptor	Cancer Cell Map

DNA\_Repair\_(Reactome)



Results Panel

Node Details

LIG1

Protein

Homo sapiens

[Pathway Commons: 6311](#)

Synonyms:

- LIG1

Links:

- [UNIPROT: P18858](#)
- [UNIPROT: Q32P23](#)
- [REF\\_SEQ: NP\\_000225](#)
- [Search iHOP](#)

[Visual Legend](#)

### Processive complex:Okazaki fragment complex

Complex

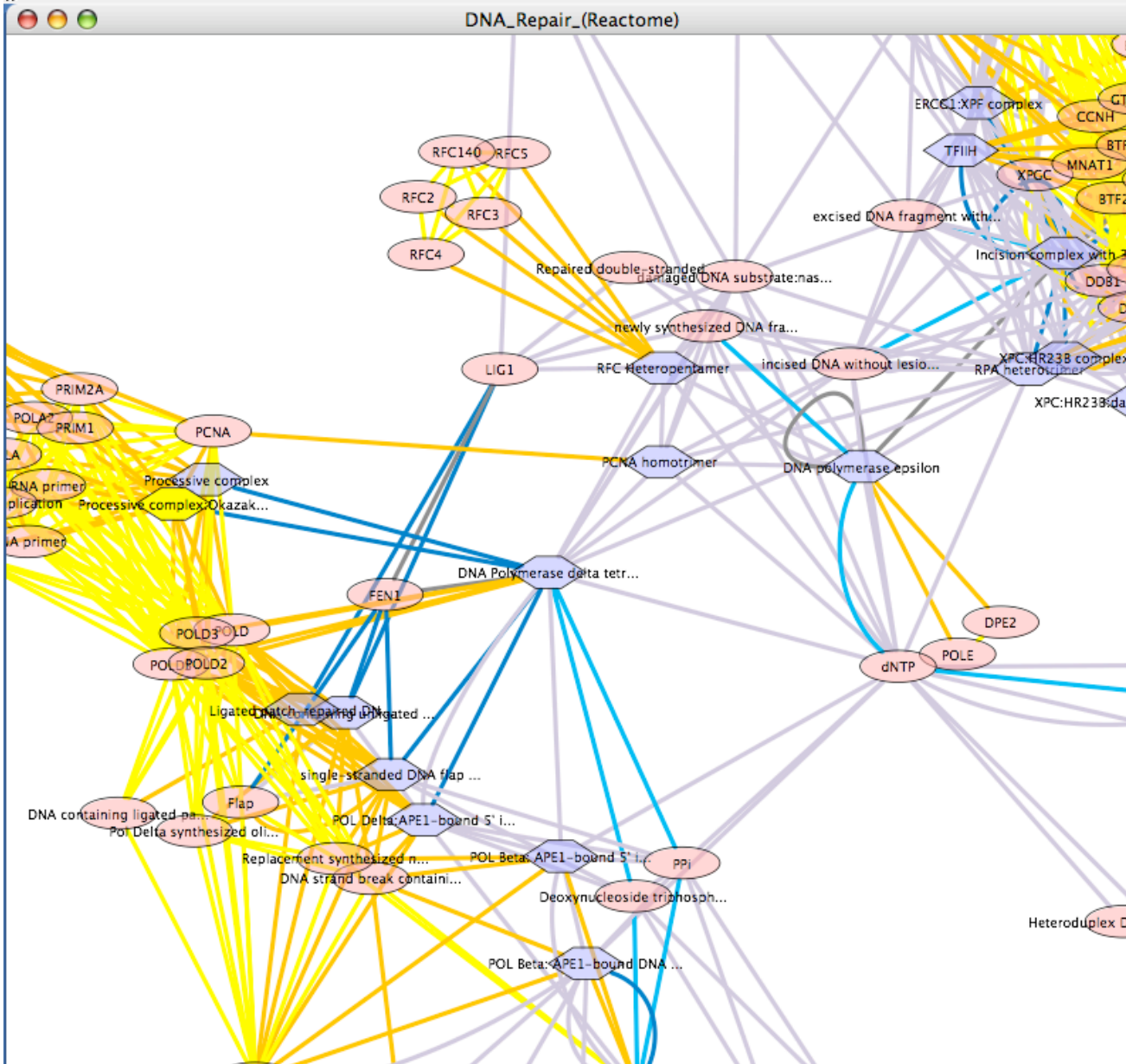
Homo sapiens

[Pathway Commons: 5256](#)

Links:

- [REACTOME: 68453](#)
- [REACTOME: REACT\\_5537](#)

[Visual Legend](#)





# Public Seminar

Break 10:30-11:00

Seminar: 11:00-12:00

Lunch: 12:00-13:00

Lab time: 13:00-13:30



# Lab Time

- Try out Cytoscape
- [http://opentutorials.rbvi.ucsf.edu/index.php/Tutorial:Introduction\\_to\\_Cytoscape](http://opentutorials.rbvi.ucsf.edu/index.php/Tutorial:Introduction_to_Cytoscape)
- Timing: until 13:45



# Tips and Tricks

Gary Bader



# Tips & Tricks

- “Root graph”
  - “There is one graph to rule them all....”
  - The networks in Cytoscape are all “views” on a single graph.
  - Changing the attribute for a node in one network *will* also change that attribute for a node with the same ID in all other loaded networks
  - There is no way to “copy” a node and keep the same ID
  - Make a copy of the session



# Tips & Tricks

- Network views
  - When you open a large network, you will not get a view by default
  - To improve interactive performance, Cytoscape has the concept of “Levels of Detail”
    - Some visual attributes will only be apparent when you zoom in
    - The level of detail for various attributes can be changed in the preferences
    - To see what things will look like at full detail:
      - View → Show Graphics Details





# Tips & Tricks

- Sessions
  - Sessions save pretty much everything:
    - Networks
    - Properties
    - Visual styles
    - Screen sizes
  - Saving a session on a large screen may require some resizing when opened on your laptop



# Tips & Tricks

- Logging
  - By default, Cytoscape writes its logs to the Error Dialog: Help → Error Dialog
  - Can change a preference to write it to the console
    - Edit → Preferences → Properties...
    - Set `logger.console` to true
    - Don't forget to save your preferences
    - Restart Cytoscape
  - (can also turn on debugging: `cytoscape.debug`, but I don't recommend it)



# Tips & Tricks

- Memory
  - Cytoscape uses lots of it
  - Doesn't like to let go of it
  - An occasional restart when working with large networks is a good thing
  - Destroy views when you don't need them
  - Java doesn't give us a good way to get the memory right at start time
    - Cytoscape 2.7 does a much better job at “guessing” good default memory sizes than previous versions



# Tips & Tricks

- .cytoscape directory
  - Your defaults and any plugins downloaded from the plugin manager will go here
  - Sometimes, if things get really messed up, deleting (or renaming) this directory can give you a “clean slate”
- Plugin manager
  - “Outdated” doesn’t necessarily mean “won’t work”
  - Plugin authors don’t always update their plugins immediately after new releases
  - Click on “Show outdated plugins” to see the entire list of plugins.



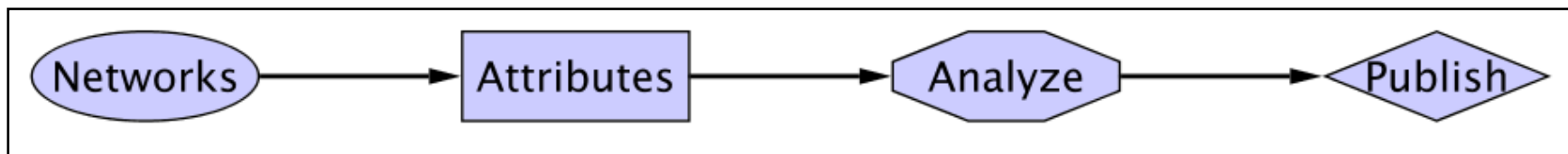
# Cytoscape Workflow

Piet Molenaar



# Cytoscape Workflow

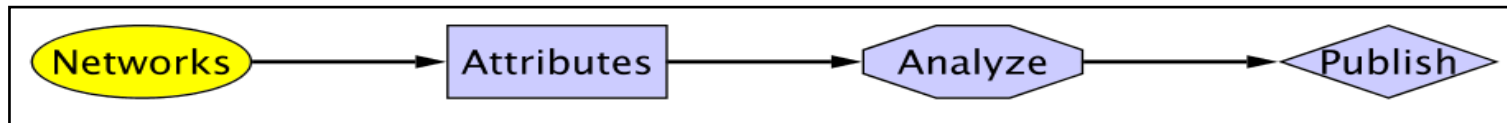
1. Load Networks (Import network data into Cytoscape)
  2. Load Attributes (Get data about networks into Cytoscape)
  3. Analyze and Visualize Networks
  4. Prepare for Publication
- A specific example of this workflow:
    - Cline, et al. “Integration of biological networks and gene expression data using Cytoscape”, Nature Protocols, 2, 2366-2382 (2007).





# All kinds of network data...

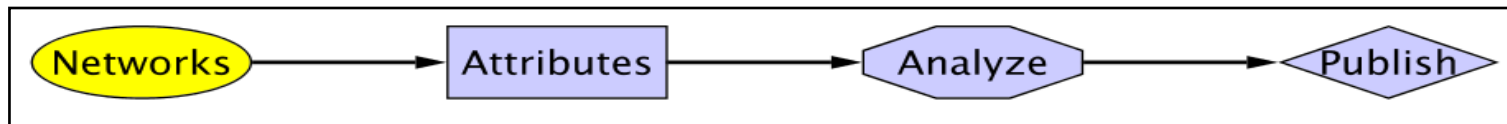
- Physical interactions
  - Protein – Protein interactions
  - Protein – DNA interactions
  - Metabolic interactions
- Functional interactions
  - Co-expression relations
  - Genetic interactions
  - Knockout/siRNA – targets





# Pre-formatted Network Files

- Cytoscape supports many popular file formats:
  - SIF (Simple Interaction Format)
  - GML (Graph Markup Language)
  - XGMML (eXtensible Graph Markup and Modeling Language)
  - BioPAX (Biological Pathway Data)
  - PSI-MI 1 & 2.5 (Protein Standards Initiative)
  - SBML Level 2 (Systems Biology Markup Language)
  - KGML (KEGG Markup Language)
- Available for download from data sources (URLs, web-services, formatted table files)

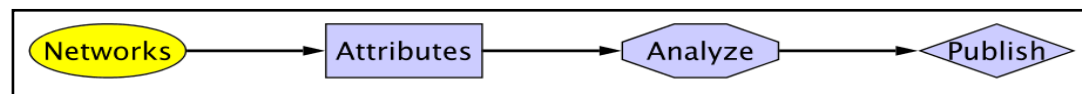
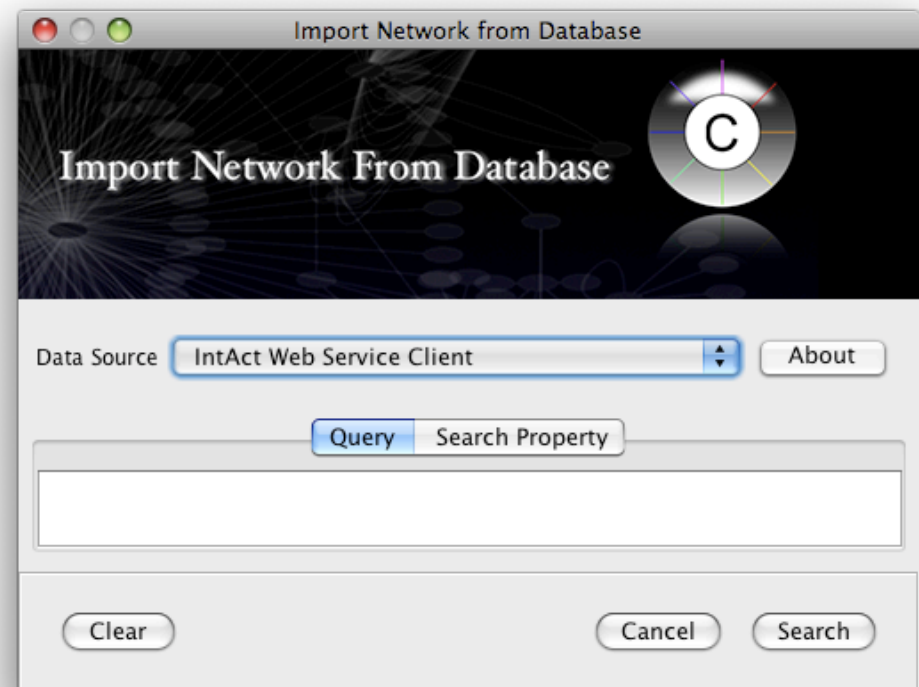






# Internet Databases

- Cytoscape version 2.6
  - web service clients: import networks directly from several trusted internet resources
    - IntAct (EMBL-EBI)
    - PathwayCommons (collection of data resources)
    - NCBI Entrez Gene
    - Many more will be included...





# Text Mining

- Computationally extract gene relationships from text, usually PubMed abstracts
- Literature search tool, lots of network data
- BUT not perfect
  - Problems recognizing gene names
  - Natural language processing not perfect
- Agilent Literature Search Cytoscape plugin
- Others: E.g. iHOP
  - [www.ihop-net.org/UniPub/iHOP/](http://www.ihop-net.org/UniPub/iHOP/)

**Agilent Literature Search 1.0.4**

Edit View Help

**Terms**  
 CSF2RB  
 EDN1  
 EGFR  
 LMNA  
 PDK2  
 TRAF1  
 WBSR14

**Context**  
 atherosclerosis

**Match Controls**  
 Max Engine Matches: 10 | Organism: Homo sapiens

**Query Controls**  
 Use Aliases:  Use Context:

**Extraction Controls**  
 Interaction Lexicon: limited

**Query Editor**  
 ((csf2rb OR if5rb OR cd131 OR cdw131 OR if3rb)) AND atherosclerosis  
 ((edn1 OR et1)) AND atherosclerosis  
 ((egfr OR mena OR erbb OR erbb1)) AND atherosclerosis  
 ((lmna OR lmnc OR cmt2b1 OR fpl OR ifp OR hgps OR emd2 OR ldp1 OR lmn1 OR fpld)) AND atherosclerosis  
 (PDK2) AND atherosclerosis  
 ((traf1 OR mgc:10353 OR ebi6)) AND atherosclerosis  
 ((wbscr14 OR ws-bhlh OR chrebp OR mondob OR mio)) AND atherosclerosis

**Query Matches**



**Cytoscape Desktop**

File Edit Data Select Layout Visualization Plugins Help Filters

Network Nodes Edges  
 1 46(0) 77(0)

Nodes: 46 (0 selected) Edges: 77 (0 selected)



Use Aliases:  Use Context:  Interaction Lexicon: limited

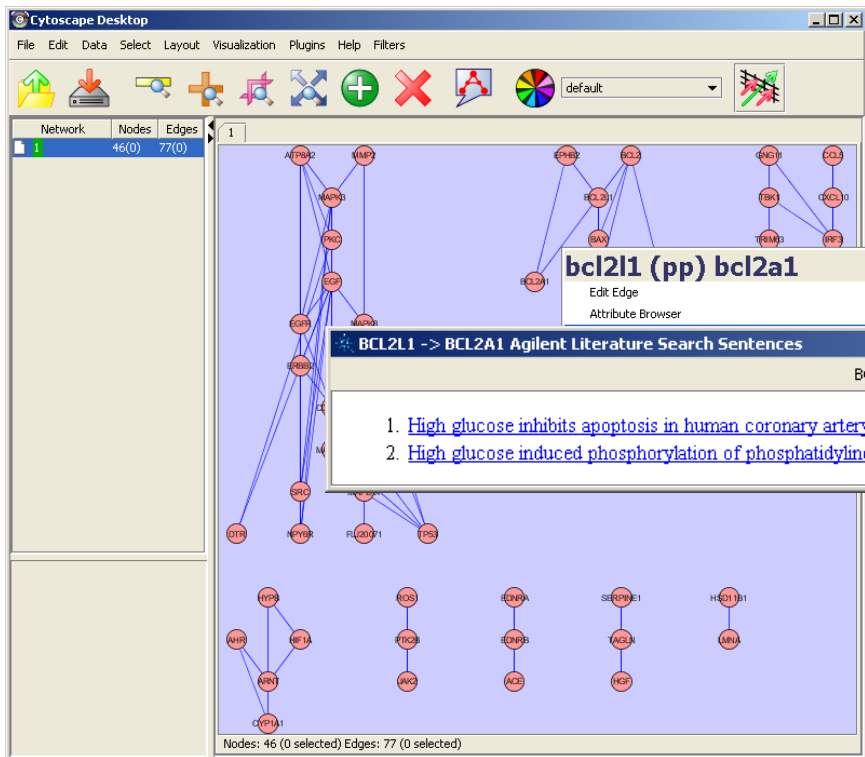
**Query Editor**  
 ((csf2rb OR if5rb OR cd131 OR cdw131 OR if3rb)) AND atherosclerosis  
 ((CRKL) AND atherosclerosis  
 ((csf2rb OR if5rb OR cd131 OR cdw131 OR if3rb)) AND atherosclerosis  
 ((edn1 OR et1)) AND atherosclerosis  
 ((egfr OR mena OR erbb OR erbb1)) AND atherosclerosis  
 ((lmna OR lmnc OR cmt2b1 OR fpl OR ifp OR hgps OR emd2 OR ldp1 OR lmn1 OR fpld)) AND atherosclerosis  
 (PDK2) AND atherosclerosis  
 ((traf1 OR mgc:10353 OR ebi6)) AND atherosclerosis  
 ((wbscr14 OR ws-bhlh OR chrebp OR mondob OR mio)) AND atherosclerosis

**Query Matches**

**Results**

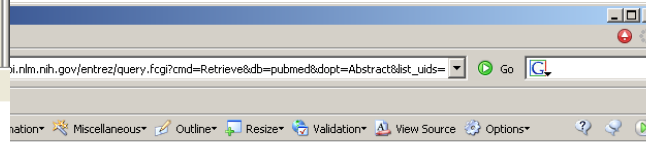
- [Association between the eNOS \(Glu298Asp\) and the RAS genes polymorphisms and premature coronary artery disease in a Turkish population \(by Berdeli A, Sekuri C, Sirri Can F, Ercan E, Sagcan A, Tengiz I, Eser E, Akim M\).](#)  
 BACKGROUND: The renin-angiotensin system (RAS) and endothelial nitric oxide (NO) affect the pathogen...  
 Source:  
 [PubMed]http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list\_uids=15563875

# Cytoscape Network produced by Literature Search.



Abstract from the scientific literature

Sentences for an edge



BCL2L1 -> BCL2A1 Agilent Literature Search Sentences

BCL2L1 -> BCL2A1 Agilent Literature Search Sentences

1. [High glucose inhibits apoptosis in human coronary artery smooth muscle cells by increasing bcl-xL and bfl-1/A1.](#)
2. [High glucose induced phosphorylation of phosphatidylinositol 3-kinase \(PI 3-K\) and extracellular signal-regulated kinase \(ERK\)1/2 along with bcl-xL and bfl-1/A1 upregulation.](#)

Physiol. 2002 Aug;283(2):C422-8. Related Articles, Links

**High glucose inhibits apoptosis in human coronary artery smooth muscle cells by increasing bcl-xL and bfl-1/A1.**

Okumura M, Okumura M, Kojima T, Maruyama T, Yasuda K.

Department of Internal Medicine, Gifu University School of Medicine, Gifu 500-8705, Japan.

- Clinical Queries
- LinkOut
- My NCBI (Cubby)
- Related Resources
- Order Documents
- NLM Catalog
- NLM Gateway
- TOXNET
- Consumer Health
- Clinical Alerts
- ClinicalTrials.gov
- PubMed Central

Cardiovascular disease is a serious complication in diabetic patients. To elucidate the precise mechanisms of atherosclerosis in diabetic patients, the effects of high glucose concentration (25 mM) on apoptosis regulation and bcl-2 family protein expression in human coronary artery smooth muscle cells (CASMC) were examined. Treatment with a high level of glucose (25 mM) caused a significant decrease in apoptosis in CASMC compared with the same cells treated with a physiologically normal glucose concentration (5.5 mM) (23.9 +/- 2.4% vs. 16.5 +/- 1.8%, P < 0.01). With respect to apoptosis regulation, treatment of CASMC with high glucose concentration markedly increased mRNA expressions of bcl-xL and bfl-1/A1 compared with cells treated with normal glucose. High glucose induced phosphorylation of phosphatidylinositol 3-kinase (PI 3-K) and extracellular signal-regulated kinase (ERK)1/2 along with bcl-xL and bfl-1/A1 upregulation. These results suggest that high glucose suppresses apoptosis via upregulation of bcl-xL and bfl-1/A1 levels through PI 3-K and ERK 1/2 pathways in CASMC. High glucose-induced increase in the expression of antiapoptotic proteins may be important in the development of atherosclerosis in diabetic patients.

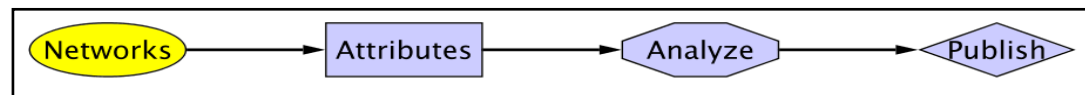
PMID: 12107051 [PubMed - indexed for MEDLINE]

Display: Abstract Show: 20 Sort by: Send to:

[Write to the Help Desk](#)  
[NCBI | NLM | NIH](#)  
[Department of Health & Human Services](#)  
[Privacy Statement](#) | [Freedom of Information Act](#) | [Disclaimer](#)



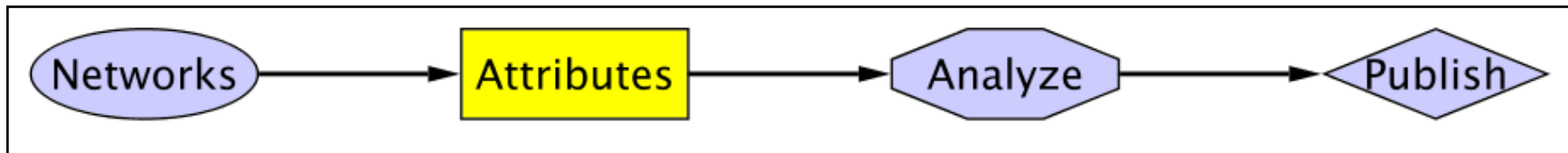
# Demo Creating Network From Internet Database





# Cytoscape Workflow

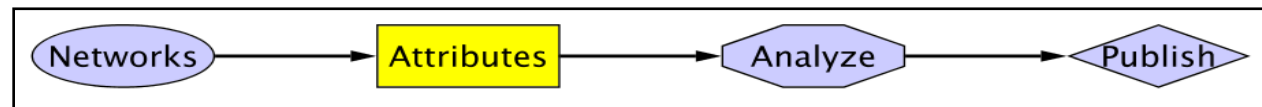
1. Load Networks (Get network data into Cytoscape)
2. Load Attributes (Get data about networks into Cytoscape)
3. Analyze and Visualize Networks
4. Prepare for Publication





# What are Attributes?

- Any data that describes or provides details about the nodes and edges in the network
  - Gene Expression Data
  - Mass Spectrometry Data
  - Protein Structure Information
  - Gene Ontology (GO) terms
  - Interaction Confidence Values, etc
- Cytoscape support multiple data types
  - Numbers (integer, float)
  - Text (string)
  - Logical (Boolean)
  - Lists...





# Attribute Management

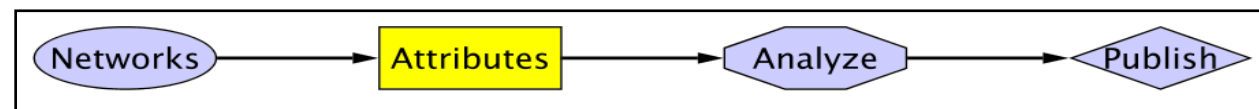
Select Attributes for Display

Node or Edge ID

Strings and floating type of attributes

ID	annotation.GO BIOLOGICAL_PROCESS	gal1RGexp	gal1RGsig	gal4RGexp
YGR136W	[biological_process]	-0.167	2.4958E-4	-0.163
YOR355W	[aerobic respiration]	-0.176	1.6613E-4	-0.044
YNR053C	[ribosomal large subunit export from nucle...]	0.352	2.1301E-7	-0.238
YPR041W	[mature ribosome assembly, regulation of ...]	-0.059	0.11203	-0.243
YER110C	[protein import into nucleus]	0.05	0.26052	-0.233
YPR035W	[glutamine biosynthetic process, nitrogen c...]	-0.197	2.3885E-5	-1.06
YGL208W	[cell aging, cellular response to glucose sta...]	0.354	1.7995E-6	0.406
YER133W	[35S primary transcript processing, cell bu...]	0.051	0.20733	-0.085
YLR377C	[gluconeogenesis]	0.873	2.1938E-10	1.067
YJR060W	[chromatin assembly or disassembly, chro...]	0.165	0.0013953	-0.306
YDL215C	[nitrogen compound metabolic process]	0.485	9.0717E-9	0.242
YJL159W	[cell wall organization and biogenesis]	-0.357	6.8879E-8	0.111
YPR167C	[methionine metabolic process, sulfate ass...]	-0.066	0.17278	-1.034
YKR099W	[histidine biosynthetic process, purine bas...]	0.466	6.1231E-6	-0.936
YBL079W	[NLS-bearing substrate import into nucleu...]	-0.186	2.5668E-4	-0.032
YNL236W	[transcription from RNA polymerase II pro...]	-0.146	0.018347	-0.218
YMR183C	[Golgi to plasma membrane transport, ves...]	-0.822	2.1741E-11	0.256
YLL021W	[Rho protein signal transduction, actin fila...]	-0.155	3.4013E-4	0.05
YNL145W	[pheromone-dependent signal transductio...]	-0.764	3.148E-11	-0.098

Specific Attribute Tabs

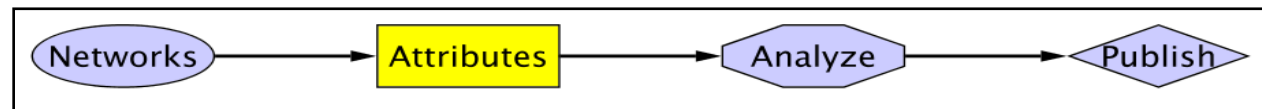






# Load Attributes: Import Attribute Files

- Map data about Networks onto Networks.
- Attributes can be loaded in many of the same ways as networks.
  - Import pre-formatted attribute files
  - Import formatted text or Excel files
  - Create attributes manually in attribute editor
  - Load attributes from web services
  - ID mapping though node attributes





# Public Sources of Gene Attributes

- Ensembl BioMart (eukaryotes)
  - <http://www.ensembl.org>
- Entrez Gene (general)
  - <http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>
- Model organism databases
  - E.g. SGD: <http://www.yeastgenome.org/>
- Many others: discuss during lab



# Ensembl BioMart

- Convenient access to gene list annotation

The screenshot displays the Ensembl BioMart interface. At the top, the 'Dataset' section is set to 'Ensembl Genes 58' and 'Homo sapiens genes (GRCh37)'. Below this, the 'Attributes' section is expanded to show various categories like REGION, GENE, TRANSCRIPT EVENT, etc. On the right, a list of attributes to download is shown, including Features, Structures, Transcript Event, Homologs, Variations, and Sequences. Blue arrows point from the 'Select genome' text to the dataset dropdown, from 'Select filters' to the filter options, and from 'Select attributes to download' to the attribute list.

**Dataset**  
Homo sapiens genes (GRCh37)  
**Filters**  
[None selected]  
**Attributes**  
Ensembl Gene ID  
Ensembl Transcript

Ensembl Genes 58  
Homo sapiens genes (GRCh37) Select genome

REGION:  
GENE:  
TRANSCRIPT EVENT:  
GENE ONTOLOGY:  
EXPRESSION:  
MULTI SPECIES COMPARISONS:  
PROTEIN DOMAINS:  
 Limit to genes ... with Protein feature scanprosite ID(s)  Only  Excluded  
 Limit to genes with these family or domain IDs: Ensembl Protein Family ID(s) [e.g. ENSFM00250000000002]  
 Transmembrane domains  Only  Excluded  
 Signal domains  Only  Excluded  
VARIATIONS:

Select filters

Select attributes to download

Features  Homologs  
 Structures  Variations  
 Transcript Event  Sequences

GENE:  
EXTERNAL:  
EXPRESSION:  
PROTEIN DOMAINS:

Cytoscape Desktop (New Session)

File Edit View Select Layout Plugins Help

Control Panel

Network VizMapper™ Editor Filters

Network	Nodes	Edges
Wnt_signaling_(NCI)_	76(0)	363(0)

Wnt\_signaling\_(NCI)

CIBP1/3BP/TC/PC/UB1/HPD/0/8K  
 FRAT1\_HUMAN/HUC1A/H  
 Beta-Catenin/WF2/ABP1/6/MBD  
 PKC/CBP1/3BP/1/5  
 DKK2/LRP6/KIB2/HPC/0/5K  
 LRP6/FZD2/HHDA01  
 GDP/Gan/G14/0/1  
 CSK21\_HUMAN/DAG/0/31  
 WNT1/FRP1/CSK/MCF10  
 TLE1\_HUMAN/prot/ESVN

Data Panel

ID	Affy HG U133-P

Node Attribute Browser | Edge Attribute Browser | Network Attribute Browser

Biomart Web Service Client

Query

bio·mart

Data Source  
 ENSEMBL 54 GENES (SANGER UK) - Homo sapiens genes (NCBI36)

Key Attribute  
 Attribute: ID  
 Data Type: EntrezGene ID(s)

Available attributes

- 5' UTR Start (5\_utr\_start)
- Aedes Chromosome (aedes\_chromosome)
- Aedes Chromosome End (bp) (aedes\_chrom\_end)
- Aedes Chromosome Start (bp) (aedes\_chrom\_start)
- Aedes Ensembl Gene ID (aedes\_ensembl\_gene)
- Aedes Ensembl Protein ID (aedes\_homolog\_ensembl\_peptide)
- Affy HC G110 (affy\_hc\_g110)
- Affy HG FOCUS (affy\_hg\_focus)
- Affy HG U133-PLUS-2 (affy\_hg\_u133\_plus\_2)
- Affy HG U133A (affy\_hg\_u133a)
- Affy HG U133A\_2 (affy\_hg\_u133a\_2)
- Affy HG U133B (affy\_hg\_u133b)
- Affy HG U95A (affy\_hg\_u95a)
- Affy HG U95AV2 (affy\_hg\_u95av2)
- Affy HG U95B (affy\_hg\_u95b)
- Affy HG U95C (affy\_hg\_u95c)
- Affy HG U95D (affy\_hg\_u95d)

Reset Cancel Import



# Gene and Protein Identifiers

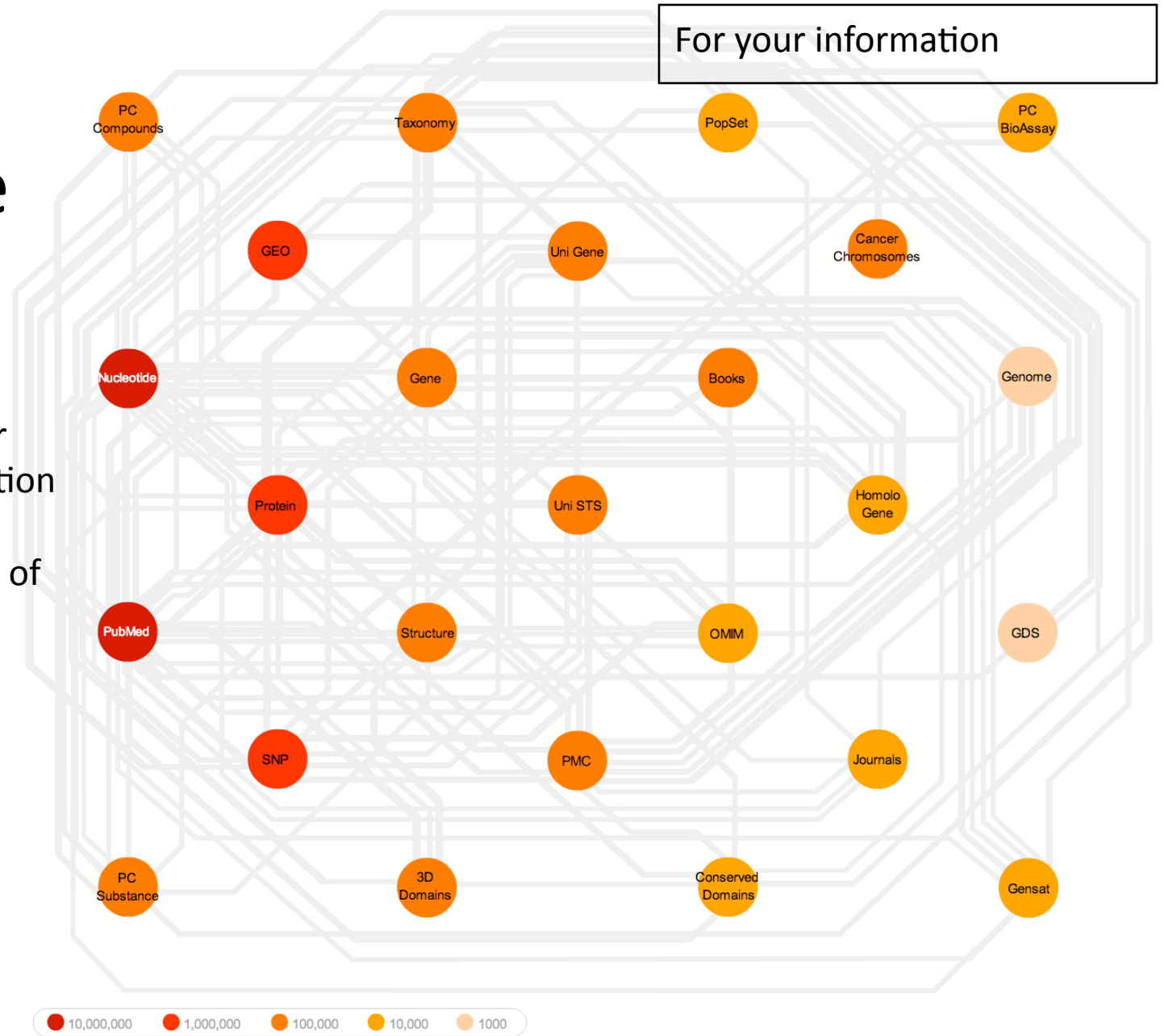
- Identifiers (IDs) are ideally unique, stable names or numbers that help track database records
  - E.g. Social Insurance Number, Entrez Gene ID 41232
- Gene and protein information stored in many databases
  - → Genes have many IDs
- Records for: Gene, DNA, RNA, Protein
  - Important to recognize the correct record type
  - E.g. Entrez Gene records don't store sequence. They link to DNA regions, RNA transcripts and proteins e.g. in RefSeq, which stores sequence.



# NCBI Database Links

NCBI:  
U.S. National Center for  
Biotechnology Information

Part of National Library of  
Medicine (NLM)



[http://www.ncbi.nlm.nih.gov/Database/datamodel/data\\_nodes.swf](http://www.ncbi.nlm.nih.gov/Database/datamodel/data_nodes.swf)

For your information

# Common Identifiers

## Gene

[Ensembl](#) [ENSG00000139618](#)

[Entrez Gene](#) [675](#)

Unigene [Hs.34012](#)

## RNA transcript

GenBank [BC026160.1](#)

[RefSeq](#) [NM\\_000059](#)

Ensembl [ENST00000380152](#)

## Protein

Ensembl [ENSP00000369497](#)

[RefSeq](#) [NP\\_000050.2](#)

[UniProt](#) [BRCA2\\_HUMAN](#) or

[A1YBP1\\_HUMAN](#)

IPI [IPI00412408.1](#)

EMBL [AF309413](#)

PDB [1MIU](#)

## Species-specific

HUGO HGNC [BRCA2](#)

MGI [MGI:109337](#)

RGD [2219](#)

ZFIN [ZDB-GENE-060510-3](#)

FlyBase [CG9097](#)

WormBase [WBGene00002299](#) or [ZK1067.1](#)

SGD [S000002187](#) or [YDL029W](#)

## Annotations

InterPro [IPR015252](#)

OMIM [600185](#)

Pfam [PF09104](#)

Gene Ontology [GO:0000724](#)

SNPs [rs28897757](#)

## Experimental Platform

Affymetrix [208368\\_3p\\_s\\_at](#)

Agilent [A\\_23\\_P99452](#)

CodeLink [GE60169](#)

Illumina [GI\\_4502450-S](#)

Red = Recommended



# Identifier Mapping

- So many IDs!
  - Mapping (conversion) is a headache
- Four main uses
  - Searching for a favorite gene name
  - Link to related resources
  - Identifier translation
    - E.g. Genes to proteins, Entrez Gene to Affy
  - Unification during dataset merging
    - Equivalent records





# ID Mapping Services

## THE SYNERGIZER

The Synergizer database is a growing repository of gene and protein identifier synonym relationships. This tool facilitates the conversion of identifiers from one naming scheme (a.k.a "namespace") to another.

load sample inputs

Select species:

Select authority:

Select "FROM" namespace:

Select "TO" namespace:

*(NB: The strings in [brackets] are representative IDs in the corresponding namespaces.)*

File containing IDs to translate:

and/or

IDs to translate:

Output as spreadsheet:



*	entrezgene
YIL062C	854748
YLR370C	851085
YKL013C	853856
YNR035C	855771
YBR234C	852536

- Synergizer
  - <http://llama.med.harvard.edu/synergizer/translate/>
- Ensembl BioMart
  - <http://www.ensembl.org>
- PICR (proteins only)
  - <http://www.ebi.ac.uk/Tools/picr/>



# ID Mapping Challenges

- Avoid errors: map IDs correctly
- Gene name ambiguity – not a good ID
  - e.g. FLJ92943, LFS1, TRP53, p53
  - Better to use the standard gene symbol: TP53
- Excel error-introduction
  - OCT4 is changed to October-4
- Problems reaching 100% coverage
  - E.g. due to version issues
  - Use multiple sources to increase coverage

Zeeberg BR et al. Mistaken identifiers: gene name errors can be introduced inadvertently when using Excel in bioinformatics BMC Bioinformatics. 2004 Jun 23;5:80



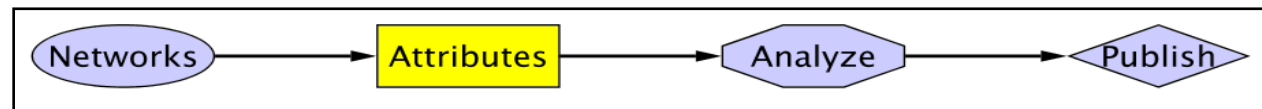
# Recommendations

- For proteins and genes
  - (doesn't consider splice forms)
- Map everything to Entrez Gene IDs using a spreadsheet
- If 100% coverage desired, manually curate missing mappings
- Be careful of Excel auto conversions – especially when pasting large gene lists!



Demo:

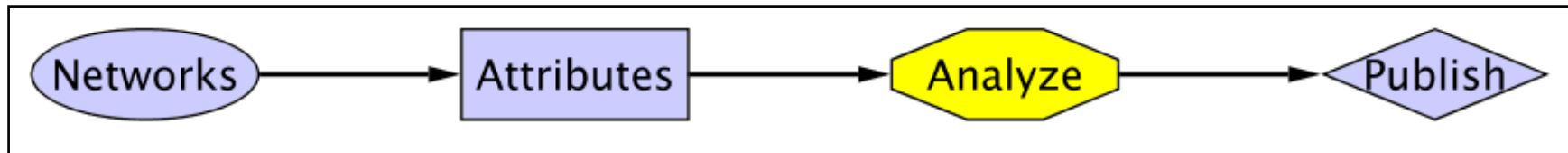
- (1) Mapping ids
- (2) Annotating Network  
with Array Data





# Cytoscape Workflow

1. Load Networks (Get network data into Cytoscape)
2. Load Attributes (Get data about networks into Cytoscape)
3. Analyze and Visualize Networks
4. Prepare for Publication



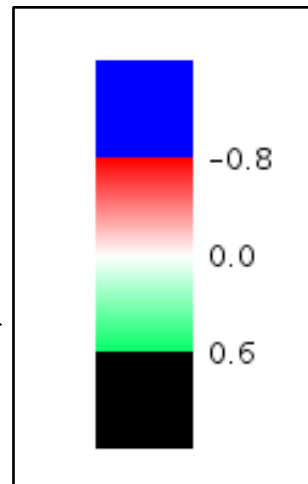


# Visual Data Integration

## 1. Network Data

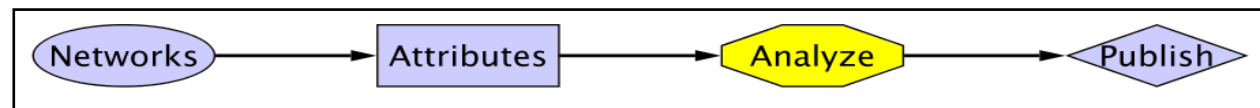
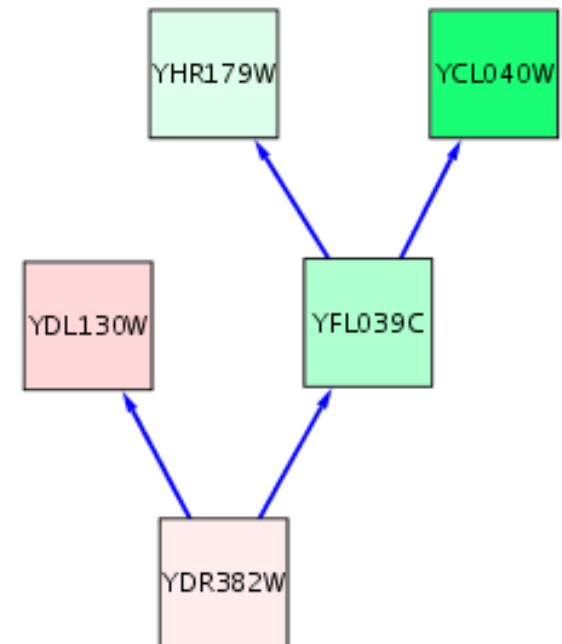
```
YDR382W pp YDL130W  
YDR382W pp YFL039C  
YFL039C pp YCL040W  
YFL039C pp YHR179W
```

VizMapper



## 2. Attribute Data

```
ExpressionValue  
YCL040W = 0.542  
YDL130W = -0.123  
YDR382W = -0.058  
YFL039C = 0.192  
YHR179W = 0.078
```





# VizMapper

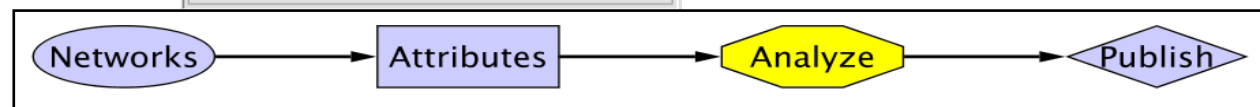
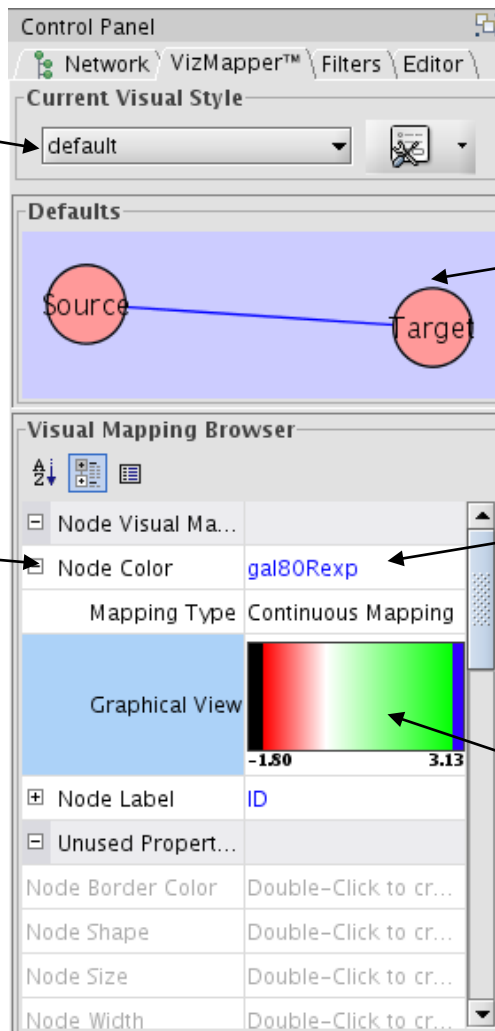
List of Visual Styles

List of Visual Attributes

Default Visual Style Editor

List of Data Attributes

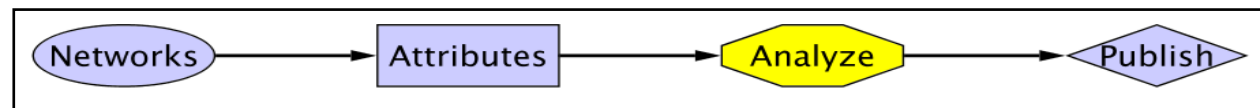
Mapping definition





# Types of mappings

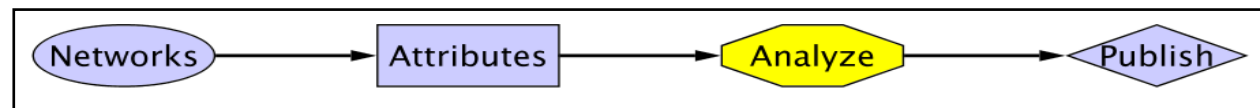
- Continuous Data mapped to Continuous Visual Attributes (e.g. gene expression levels mapped to node color)
- Continuous Data mapped to Discrete Visual Attributes (e.g. p-value categories mapped to node shape)
- Discrete (categorical) Data to Discrete Visual Attributes (e.g. GO annotation mapped to node shape)
- Discrete Data mapped to Continuous Visual Attributes (e.g. multiple GO terms mapped to pie coloring)







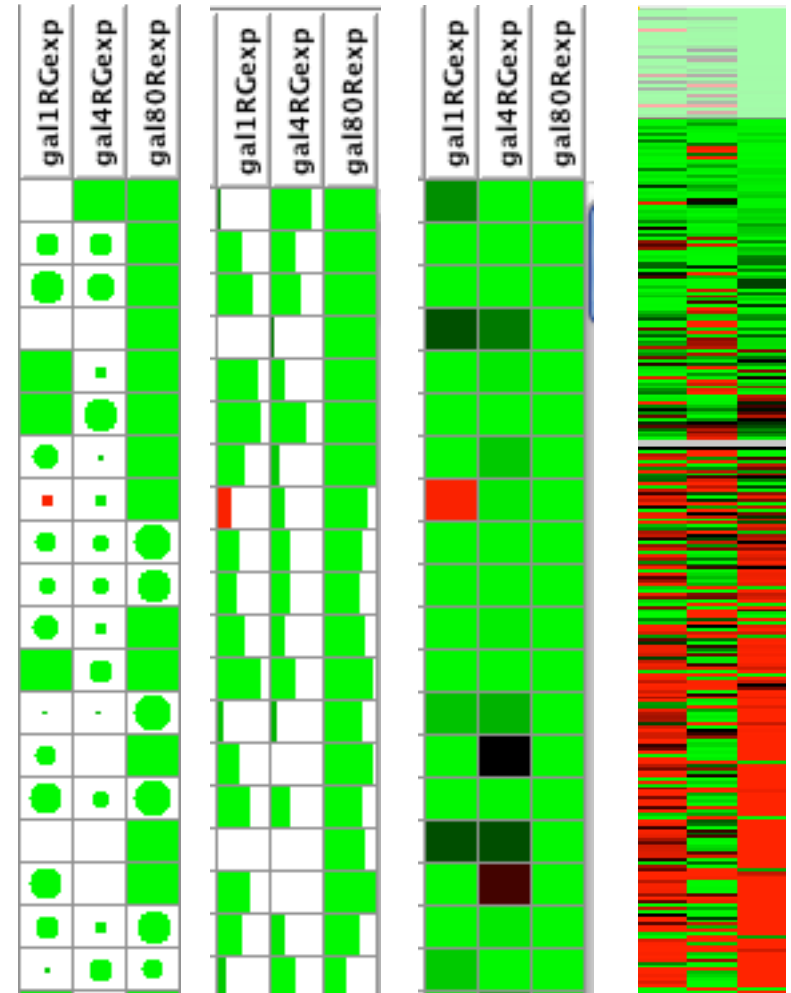
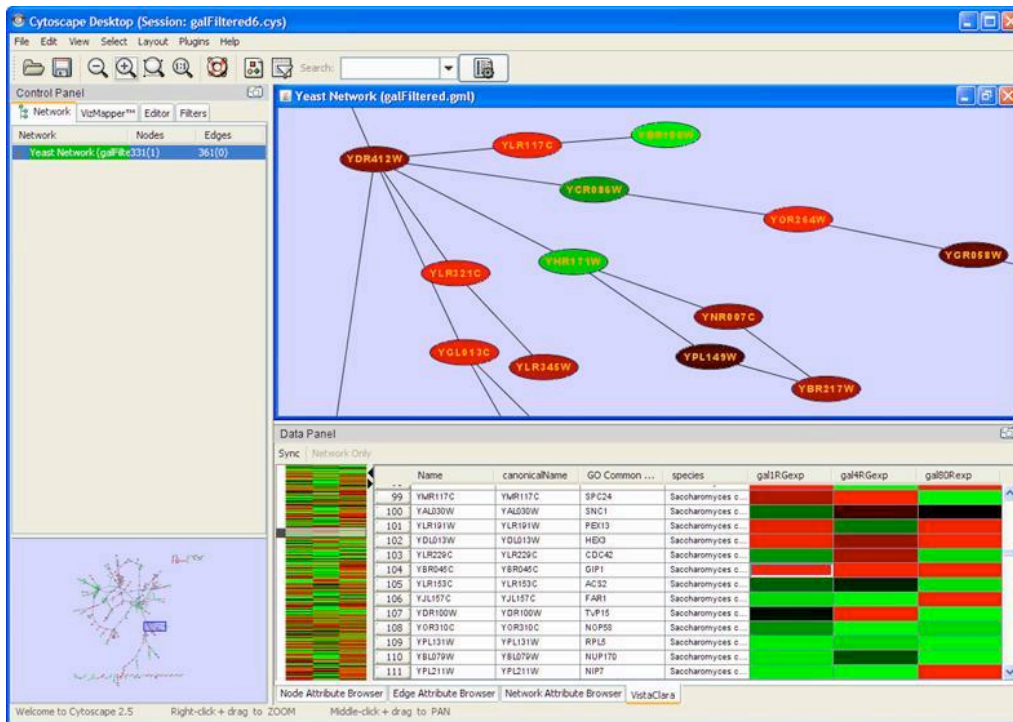
# Demo Applying Vizmapping





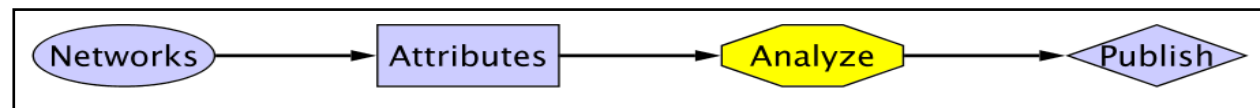
# VistaClara

- Visualization for gene expression data
- Heat maps, sorting, animation



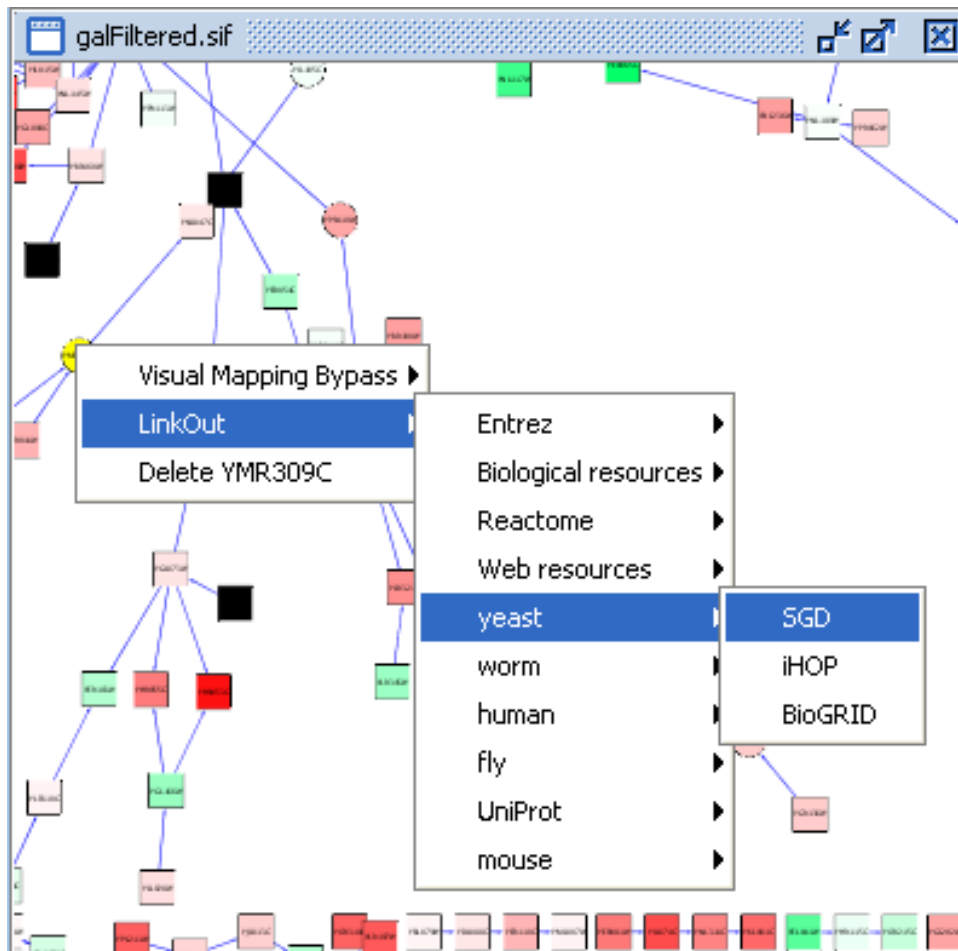


# Demo network filtering and layout

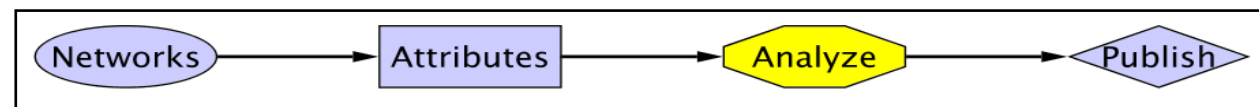




# Linkout



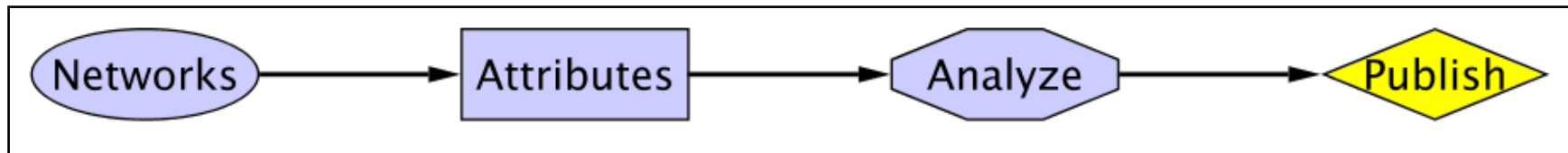
- Nodes and Edges act as hyperlinks to external databases.
- User-configurable URLs
- Collection of the biological results for the publication





# Cytoscape Workflow

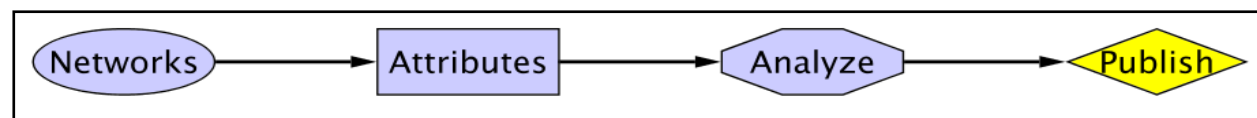
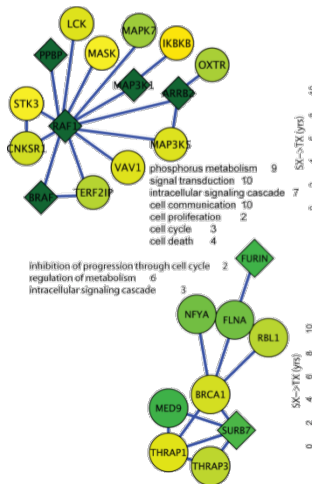
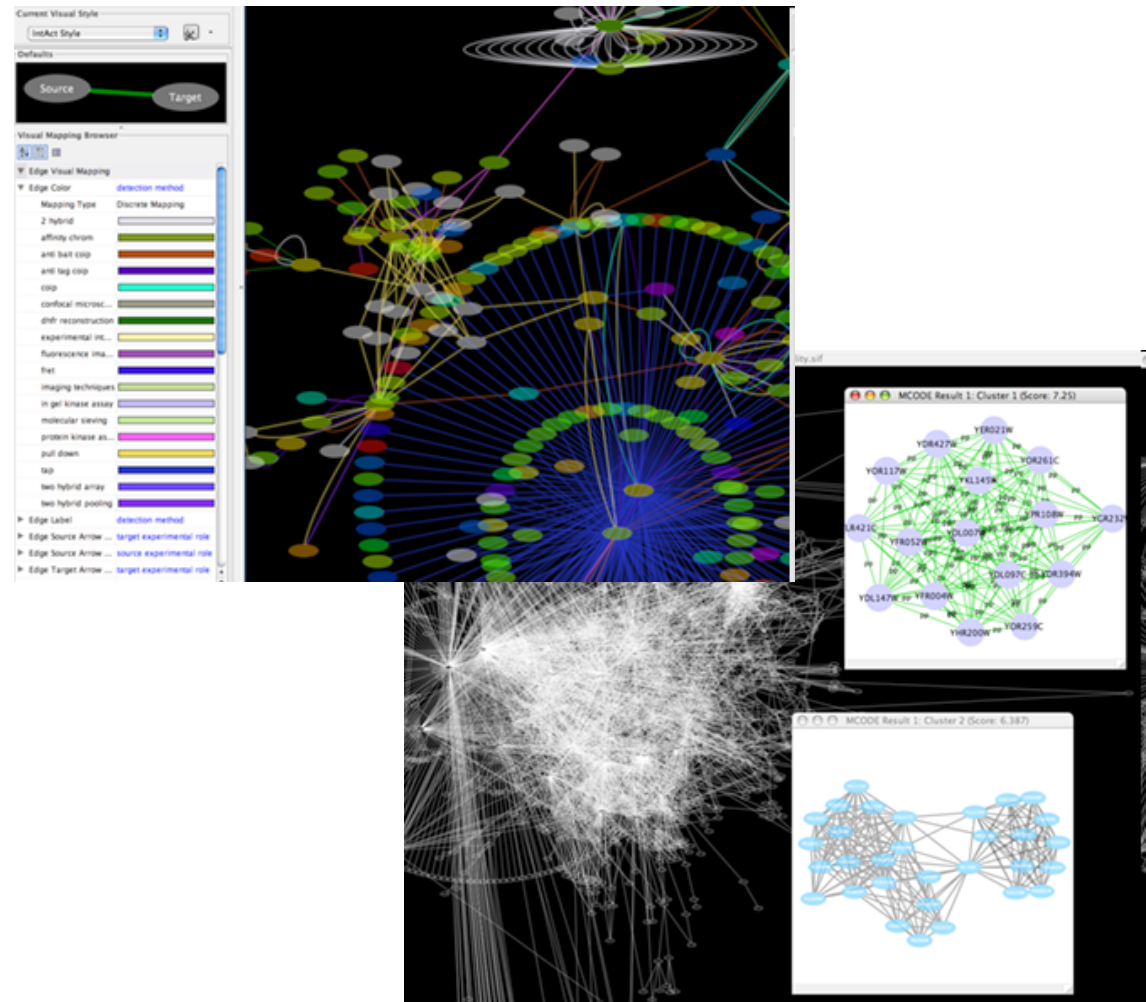
1. Load Networks (Get network data into Cytoscape)
2. Load Attributes (Get data about networks into Cytoscape)
3. Analyze and Visualize Networks
4. Prepare for Publication





# Publication quality figures

- Publication Quality Graphics in several formats
  - PDF, EPS, SVG, PNG, JPEG, and BMP
- Export Session to HTML for Web





# Break

14:45-15:15

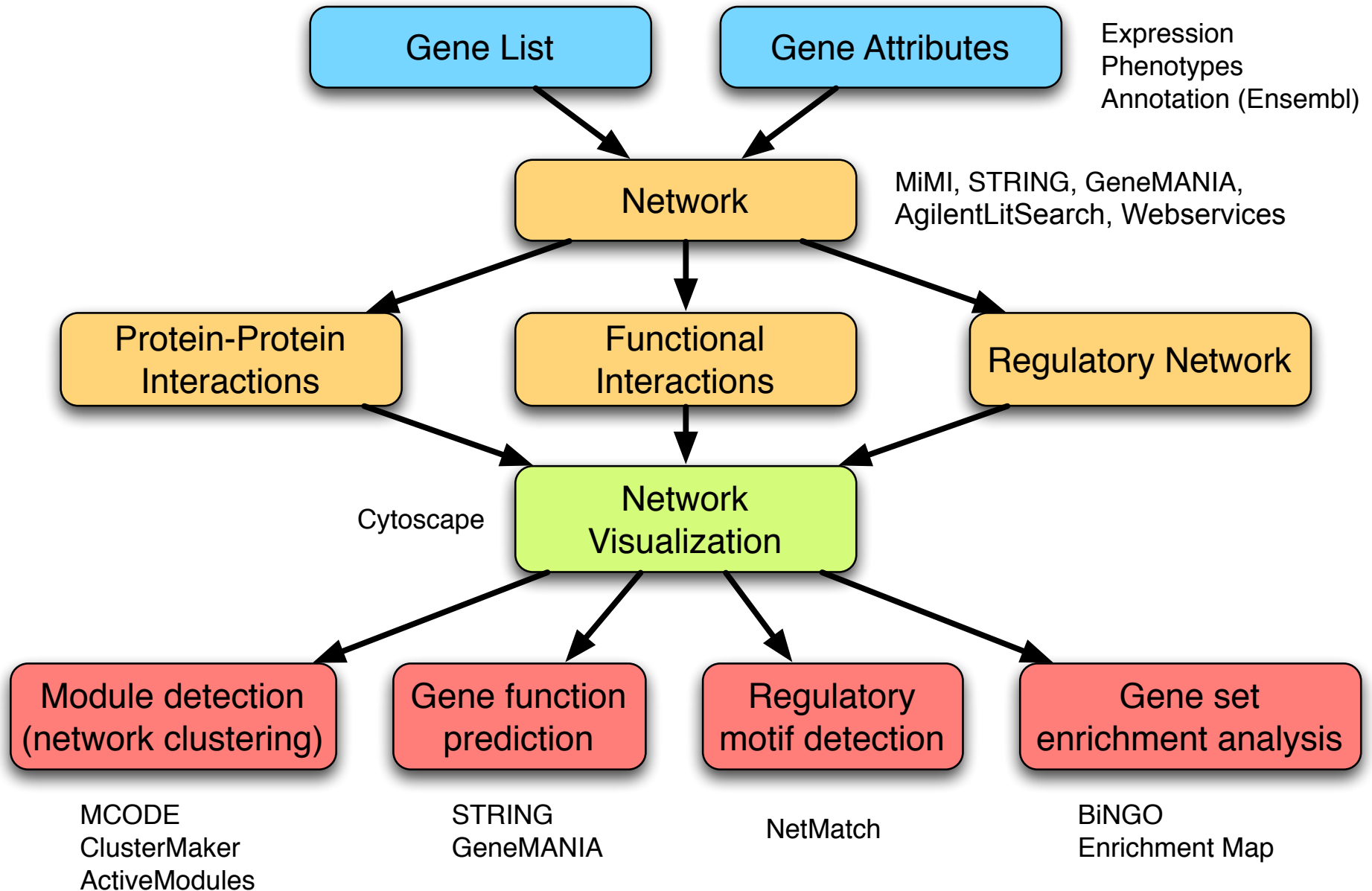


# Lab Time

- Try out workflow
  - Agilent Literature Search, VistaClara
  - ID mapping services
  - Use your own data or sample data that comes with Cytoscape
- Resource:
  - [http://opentutorials.rbvi.ucsf.edu/index.php/Tutorial:Introduction\\_to\\_Cytoscape](http://opentutorials.rbvi.ucsf.edu/index.php/Tutorial:Introduction_to_Cytoscape)
- Timing: 15:15-16:15



# Gene List and Network Analysis Overview

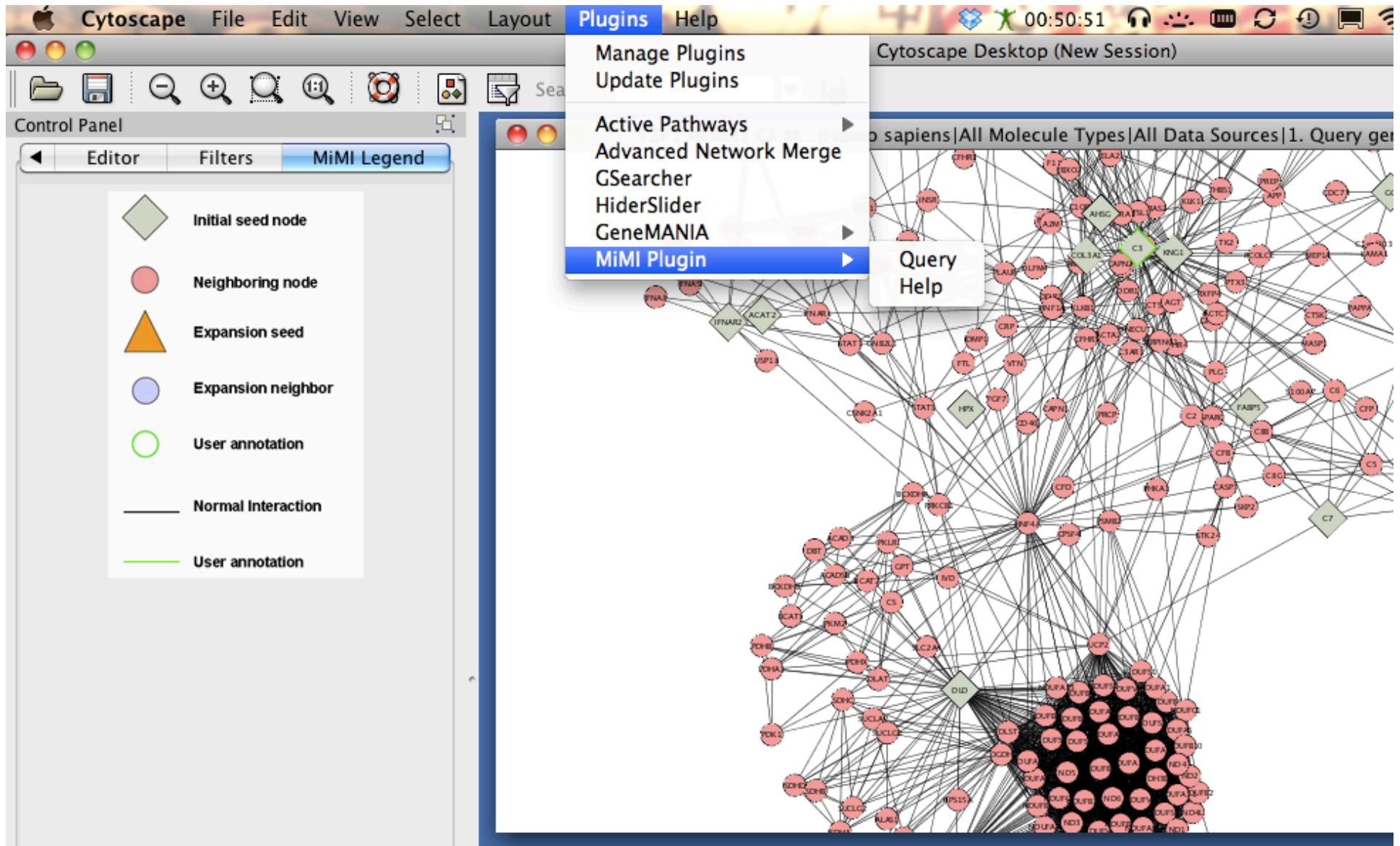


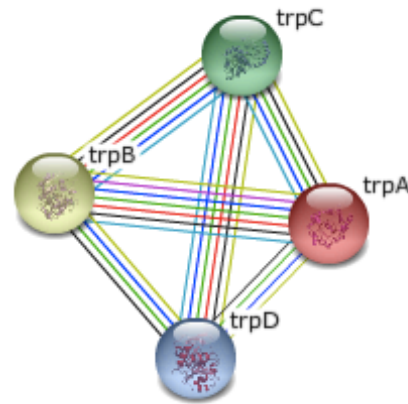


# Gene List to Network Lab

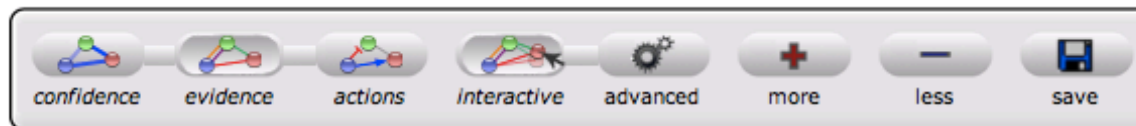
- Start with a gene list and find a network
  - MIMI – Protein-protein interactions (PPI)
  - STRING, GeneMANIA – Functional interactions
  - AgilentLitSearch – text mined interactions
  - BisoGeNet – another PPI source
- Gene function prediction with STRING and GeneMANIA

# MiMI: Protein interactions





This is the **evidence view**. Different line colors represent the types of evidence for the association.



(requires Flash player 10 or better)

## Your Input:

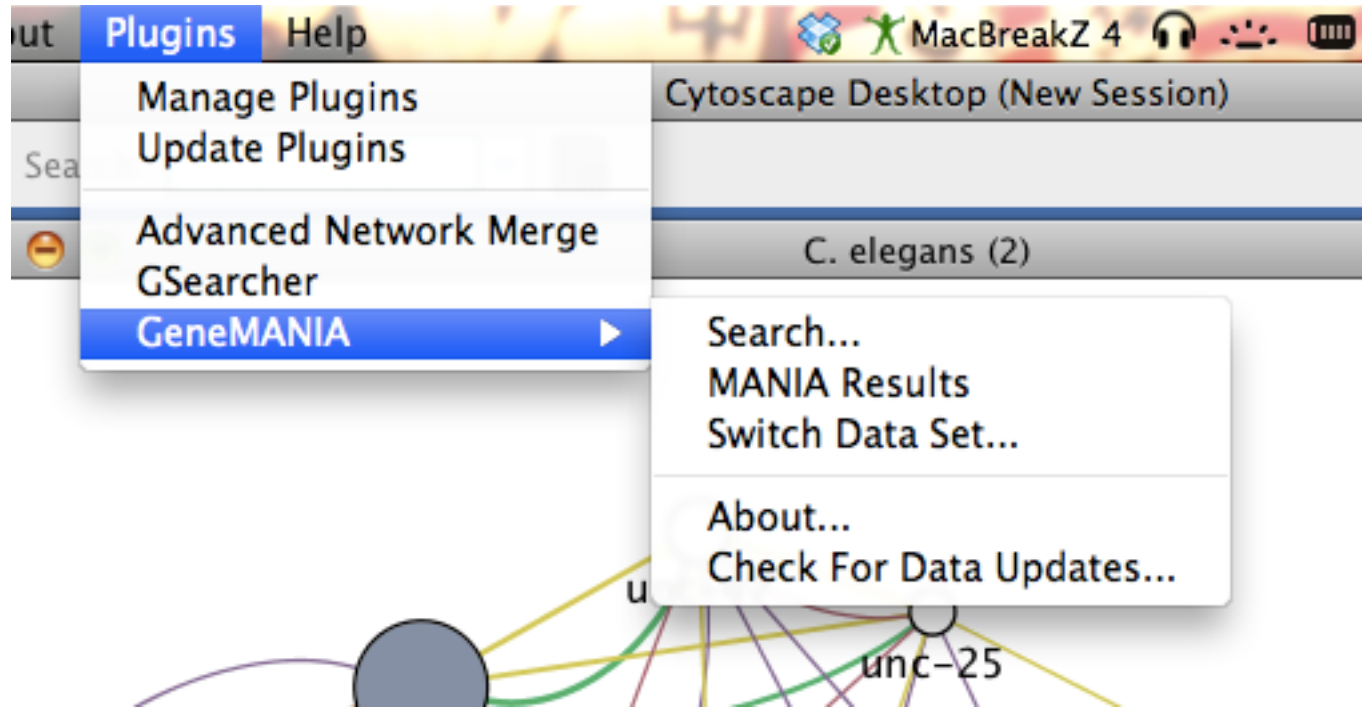
- trpA Tryptophan synthase, alpha subunit; The alpha subunit is responsible for the aldol cleavage of indoleglycerol phosphate to indole and glyceraldehyde 3- phosphate (268 aa)
- trpB Tryptophan synthase beta subunit (Tryptophan synthase subunit B); The beta subunit is responsible for the synthesis of L- tryptophan from indole and L-serine (397 aa)
- trpC N-(5-phosphoribosyl)anthranilate isomerase and indole-3-glycerolphosphate synthetase; Bifunctional enzyme that catalyzes two sequential steps of tryptophan biosynthetic pathway. The first reaction is catalyzed by the isomerase, coded by the trpF domain; the second reaction is catalyzed by the synthase, coded by the trpC domain (452 aa)
- trpD fused glutamine amidotransferase (component II) of anthranilate synthase/anthranilate phosphoribosyl transferase (531 aa)  
(*Escherichia coli* K12)

## Predicted Functional Partners:

Neighborhood  
Gene Fusion  
Cooccurrence  
Coexpression  
Experiments  
Databases  
Textmining  
[Homology]  
Score

# http://string.embl.de

<a href="#"><u>tabdelimited.zSVr2AGatnE .txt</u></a>	Text Summary (TXT - simple tab delimited flatfile)
<a href="#"><u>xml_summary.zSVr2AGatnE .xml</u></a>	XML Summary (PSI - Proteomics Standards Initiative)
<a href="#"><u>network_medusa.zSVr2AGatnE .dat</u></a>	Graph Layout (Data for the 'Medusa' Network Viewer)
<a href="#"><u>protein_sequences.zSVr2AGatnE .fa</u></a>	Network Proteins / Amino Acid Sequences (Multi-Sequence File; FASTA format)
<a href="#"><u>proteins_desc.zSVr2AGatnE .txt</u></a>	Network proteins description (TXT - simple tab delimited flatfile)



<http://www.genemania.org/plugin/>



# GeneMANIA

## Available Data

**Organisms** 1    **Networks** 76    **Genes** 20247    **Interactions** 9394174    **Version** 2010-04-28

[Manage Data](#)

## Choose Query Genes

Organism:

Name	Description
unc-18 (UNC18_CAEEL)	unc-18 encodes the C. elegans ortholog of Saccharomyces cerevisiae SEC1 and mammalian Munc18 proteins. U
unc-30 (UNC30_CAEEL)	unc-30 encodes a homeodomain-containing protein that is orthologous to the Pitx family of homeodomain tra
unc-4 (UNC4_CAEEL)	The unc-4 gene encodes a paired-class homeodomain protein with homologs in Drosophila and vertebrates. I
unc-5 (UNC5_CAEEL)	unc-5 encodes a netrin receptor. unc-5 activity is required cell autonomously for dorsalward cell and pioneer

[Remove](#)

[Remove All](#)

## Choose Interaction Networks

Select: [all](#), [none](#), [default](#).

- Co-expression (3/10)
- Co-localization (1/1)
- Genetic interactions (2/4)
- Other (0/1)
- Physical interactions (4/8)
- Predicted (0/50)
- Shared protein domains (0/2)

- Baugh-Hunter-2005
- Fox-Miller-2007 A
- Fox-Miller-2007 B
- Kirienko-Fay-2007
- Lee-Marcotte-2008 Co-expressi
- Lewis-Jackson-2009
- McElwee-Gems-2004
- Stuart-Kim-2003
- Troemel-Kim-2006

Find the top

related genes using

weighting.

[Start](#)



Cytoscape Desktop (New Session)

Control Panel

Network VizMapper™

Network	Nodes	Edges
C. elegans (1)	14(0)	18(0)
C. elegans (2)	14(0)	53(0)

Search:

C. elegans (2)

Results Panel

MANIA Results

Organism: C. elegans

Networks Genes

Sort by: [name](#), [per cent weight](#)

Expand: [all](#), [top-level](#), [none](#)

Enable: [all](#), [none](#)

- Predicted 84.
- Other 4.
- Genetic interactions 3.
- Co-expression 3.
- Shared protein domains 2.
- Physical interactions 0.
- Co-localization 0.

Export results...  
Hide results

Data Panel

ID

Node Attribute Browser Edge Attribute Browser Network Attribute Browser

Welcome to Cytoscape 2.7.0 Right-click + drag to ZOOM Middle-click + drag to PAN





# http://www.genemania.org

**GENEMANIA**  
fast gene function predictions

Find genes in



related to

[Show advanced options](#)

Save ▾ Actions ▾ Networks legend Functions legend

Networks Genes Functions Help

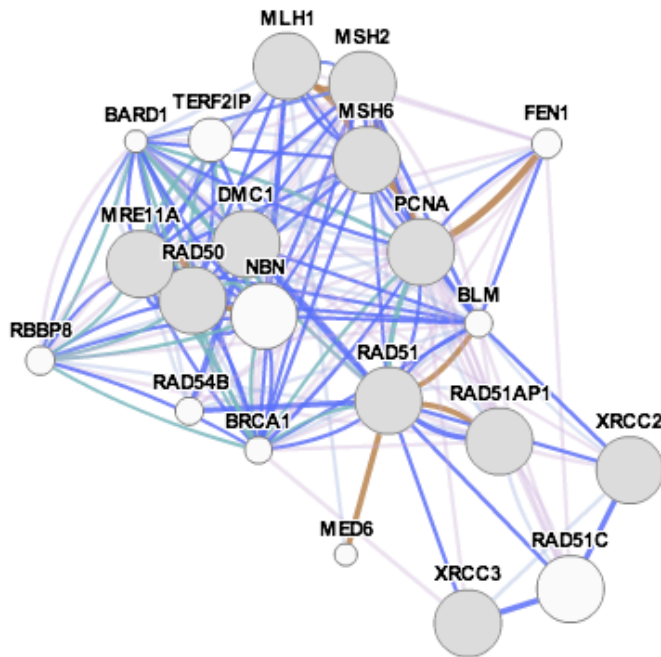


Sort by: [name](#), [per cent weight](#)

Expand: [all](#), [only top level](#), [none](#)

Enable: [all](#), [none](#)

<input checked="" type="checkbox"/>	<b>Predicted</b>	39.22 %
<input checked="" type="checkbox"/>	<b>Co-expression</b>	26.78 %
<input checked="" type="checkbox"/>	<b>Pathway</b>	12.46 %
<input checked="" type="checkbox"/>	<b>Physical interactions</b>	11.65 %
<input checked="" type="checkbox"/>	<b>Co-localization</b>	9.89 %





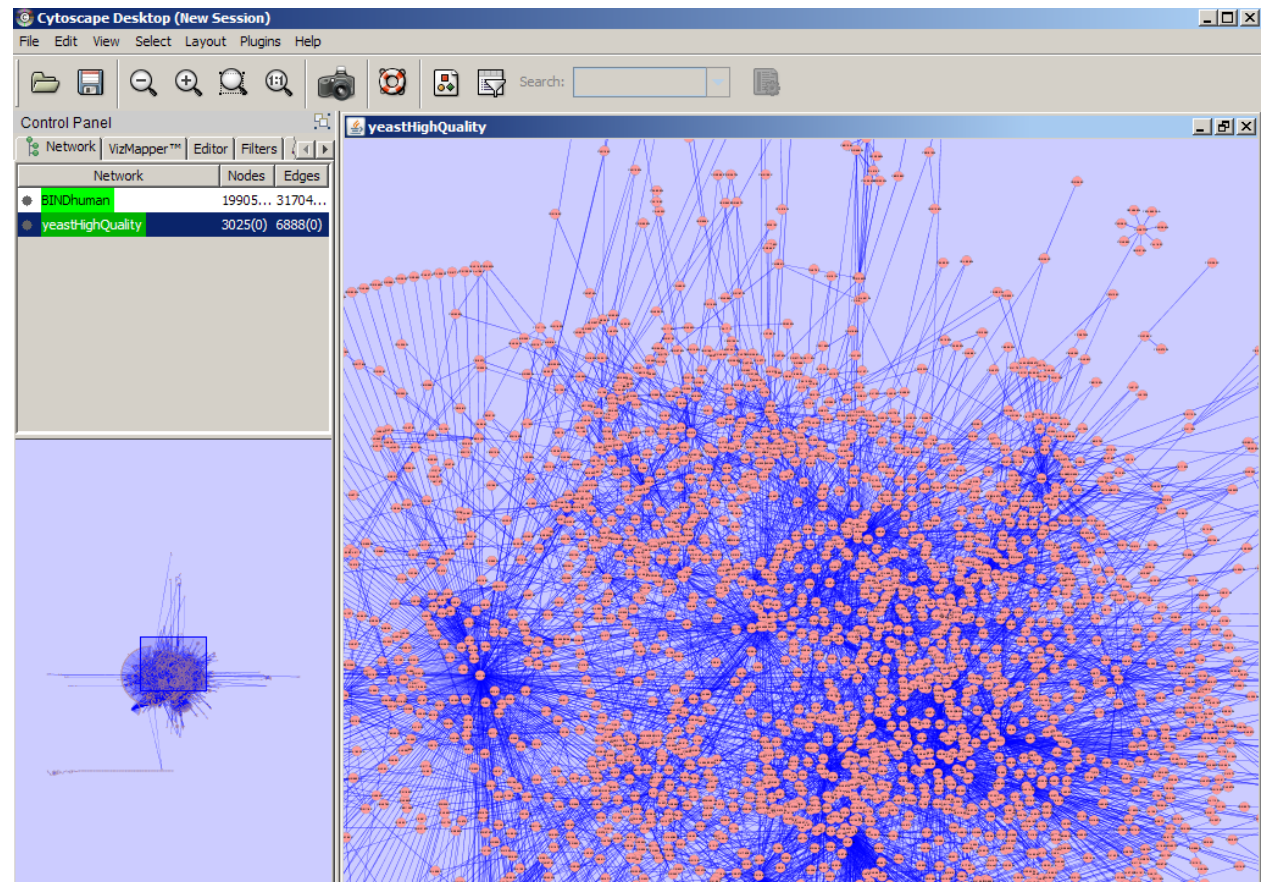
# Gene List to Network Lab

- Enter gene list into STRING and GeneMANIA websites
- Save results as text and load into Cytoscape
- Try MIMI plugin
- Gene function prediction: input a list of genes known to be in a given function. Ask STRING or GeneMANIA to find more genes like those (guilt by association)



# PPI network analysis Lab

- Load protein-protein interaction network (e.g. YeastHighQuality: 7000 Yeast interactions among 3000 proteins)
- Visualize





# PPI network analysis

- Load protein-protein interaction network (eg YeastHighQuality)
- Visualize
  - Large dataset; hairball
  - Layouts don't help
- Cluster - MCODE, ClusterMaker, ActiveModules



# Analysis

## Find Network Clusters - MCODE Plugin

- Network clusters are highly interconnected sub-networks that may be also partly overlapping
- Clusters in a protein-protein interaction network have been shown to represent protein complexes and parts of biological pathways
- Clusters in a protein similarity network represent protein families
- Network clustering is available through the MCODE Cytoscape plugin



# MCODE plug-in demo



Rank	Score	Size	Names	Complex
1	7.25	16,116	YGR232W, YDL007W, YKL145W, YFR052W, YFR004W, YLR421C, YOR261C, YDL147W, YDR427W, YHR200W, YER021W, YOR117W, YDL097C, YOR259C, YPR108W, YDR394W	
2	6.387	31,198	YPL093W, YBL004W, YOR272W, YNL110C, YKL009W, YFL002C, YOL077C, YPL126W, YIL035C, YLR409C, YLR129W, YOR061W, YKR060W, YCR057C, YDR449C, YOR039W, YJL109C, YPL012W, YGR103W, YLR449W, YOR206W, YKL014C, YLL008W, YKL172W, YNL002C, YLR002C, YGL111W, YOL041C, YGL019W, YOR145C, YPR016C	
3	5.417	12,65	YGL011C, YOL038W, YPR103W, YMR314W, YBL041W, YOR362C, YER012W, YJL001W, YML092C, YGR253C, YER094C, YGR135W	
4	5	15,75	YPL043W, YMR290C, YER006W, YKR081C, YDR496C, YDL031W, YNL061W, YNL132W, YLR222C, YLR197W, YMR049C, YHR052W, YJL069C, YKL099C, YDL014W	
5	5	12,60	YPR187W, YPR010C, YPR110C, YNL248C, YOR341W, YNR003C, YKL144C, YOR207C, YPR190C, YNL113W, YOR116C, YBR154C	

Create a new child network.



Rank	Score	Size	Names	Complex
1 <b>Proteasome 26S</b>	7.25	16,116	YGR232W, YDL007W, YKL145W, YFR052W, YFR004W, YLR421C, YOR261C, YDL147W, YDR427W, YHR200W, YER021W, YOR117W, YDL097C, YOR259C, YPR108W, YDR394W	
2 <b>Ribosome</b>	6.387	31,198	YPL093W, YBL004W, YOR272W, YNL110C, YKL009W, YFL002C, YOL077C, YPL126W, YIL035C, YLR409C, YLR129W, YOR061W, YKR060W, YCR057C, YDR449C, YOR039W, YJL109C, YPL012W, YGR103W, YLR449W, YOR206W, YKL014C, YLL008W, YKL172W, YNL002C, YLR002C, YGL111W, YOL041C, YGL019W, YOR145C, YPR016C	
3 <b>Proteasome 20S</b>	5.417	12,65	YGL011C, YOL038W, YPR103W, YMR314W, YBL041W, YOR362C, YER012W, YJL001W, YML092C, YGR253C, YER094C, YGR135W	
4 <b>RNA Splicing</b>	5	15,75	YPL043W, YMR290C, YER006W, YKR081C, YDR496C, YDL031W, YNL061W, YNL132W, YLR222C, YLR197W, YMR049C, YHR052W, YJL069C, YKL099C, YDL014W	
5 <b>RNA Pol core</b>	5	12,60	YPR187W, YPR010C, YPR110C, YNL248C, YOR341W, YNR003C, YKL144C, YOR207C, YPR190C, YNL113W, YOR116C, YBR154C	

Create a new child network.





# Lab Time

- Try out MCODE with your dataset of choice
- Or use one of the other sets available
- If you find something interesting please share!





# Gene Ontology analysis Lab

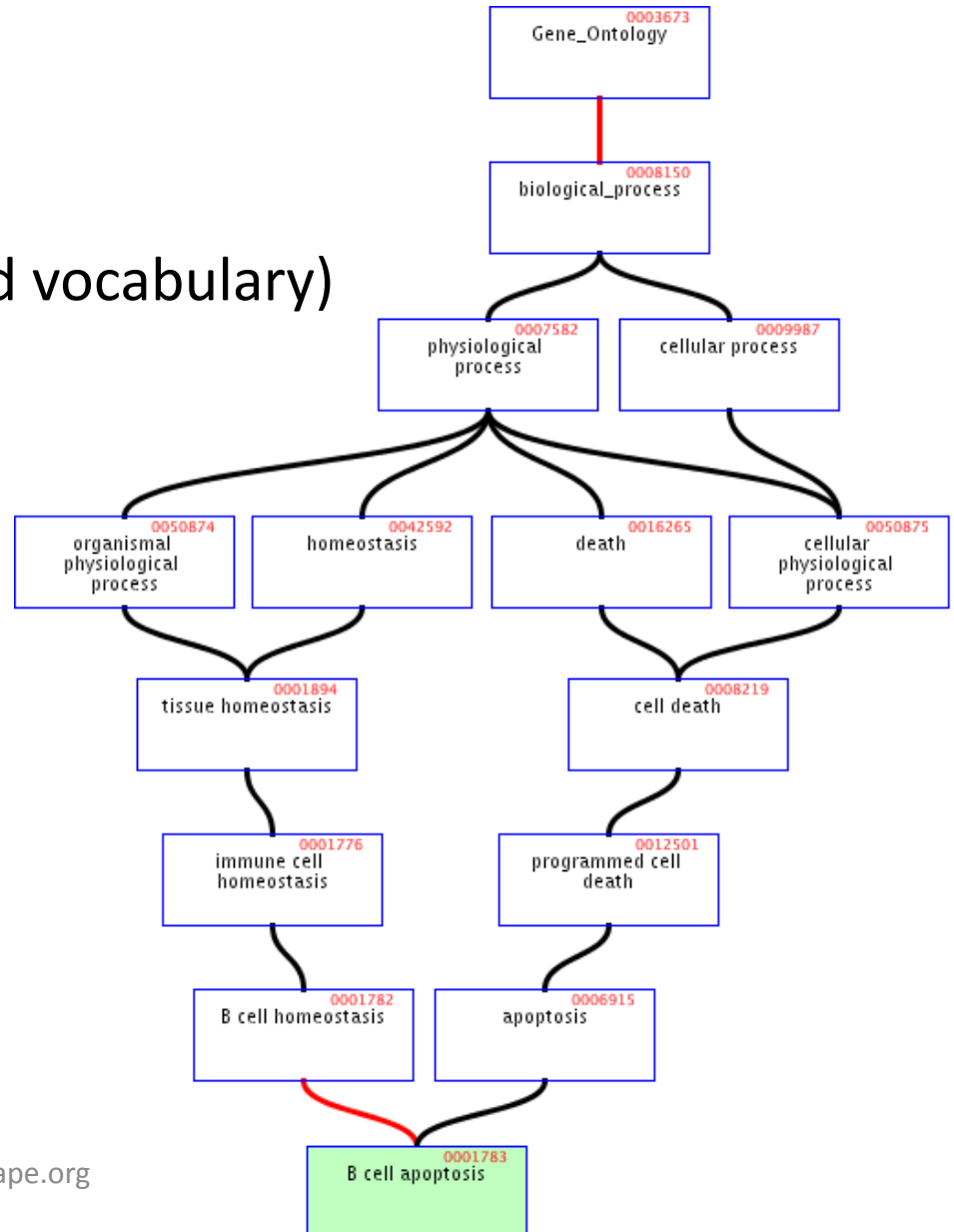
➤ Describes gene function

## 1. Agreed upon terms (controlled vocabulary)

- Biological process
- Cellular component
- Molecular function

## 2. Genome annotation

[www.geneontology.org](http://www.geneontology.org)





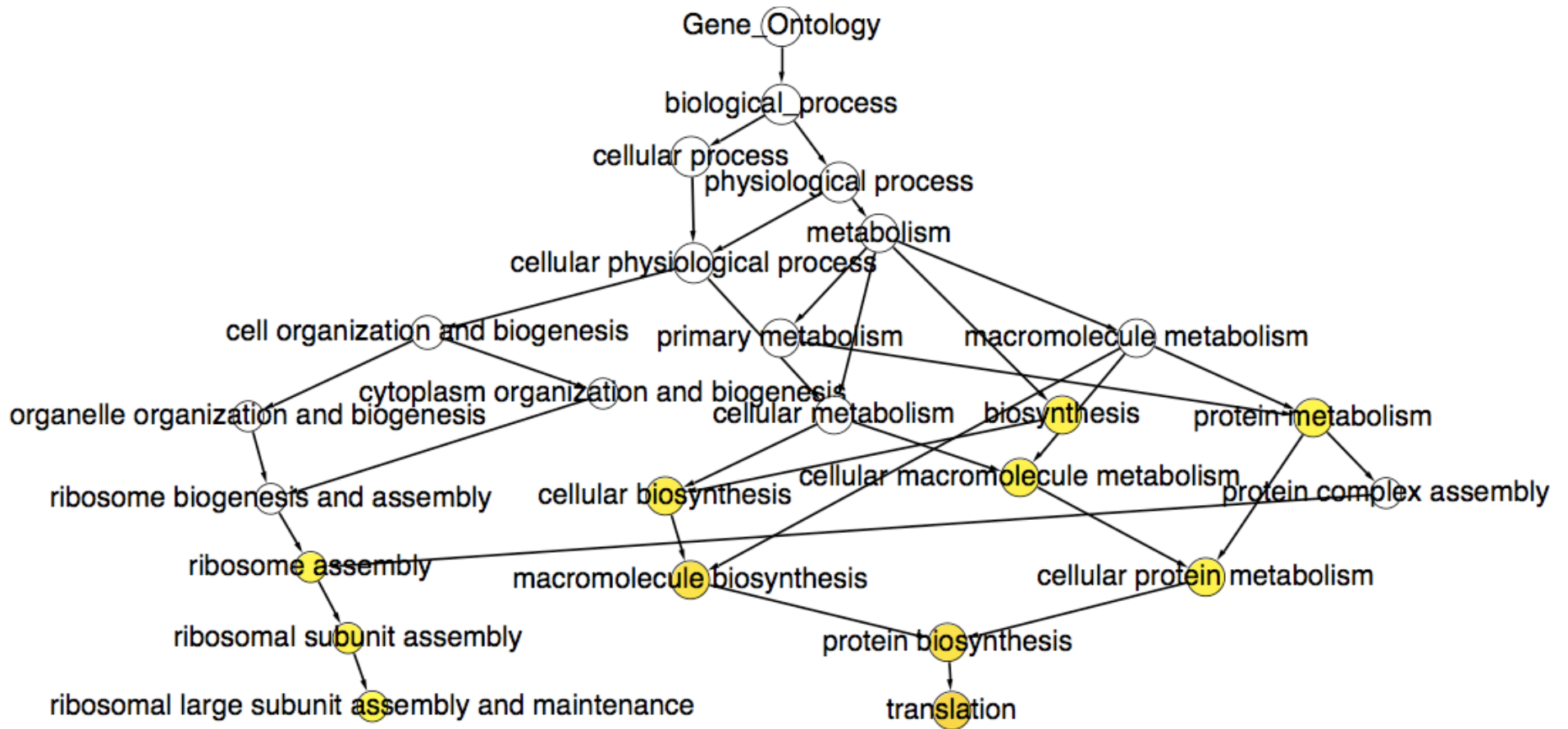
# BinGO plugin

- Calculates over-representation of a subset of genes with respect to a background set in a specific GO category
- Input: subnetwork, or list
  - Background set by user
- Output: tree with nodes color reflecting overrepresentation; also as lists
- Caveats: Gene identifiers must match; low GO term coverage, GO bias, Background determining



# BiNGO

Hypergeometric p-value  
Multiple testing correction  
(Benjamini-Hochberg FDR)





# Lab time

- Try BINGO with subsets from e.g. MCODE



# Analysis Lab

## Find Active Subnetworks Lab

- Active modules are sub-networks that show differential expression over user-specified conditions or time-points
  - Microarray gene-expression attributes
  - Mass-spectrometry protein abundance
- Method
  - Calculate z-score/node, ZA score/subgraph, correct for random expression data sampling
  - Score over multiple experimental conditions
  - Simulated annealing-based search method is used to find the high scoring networks



# Analysis Lab

## Find Active Subnetworks Lab

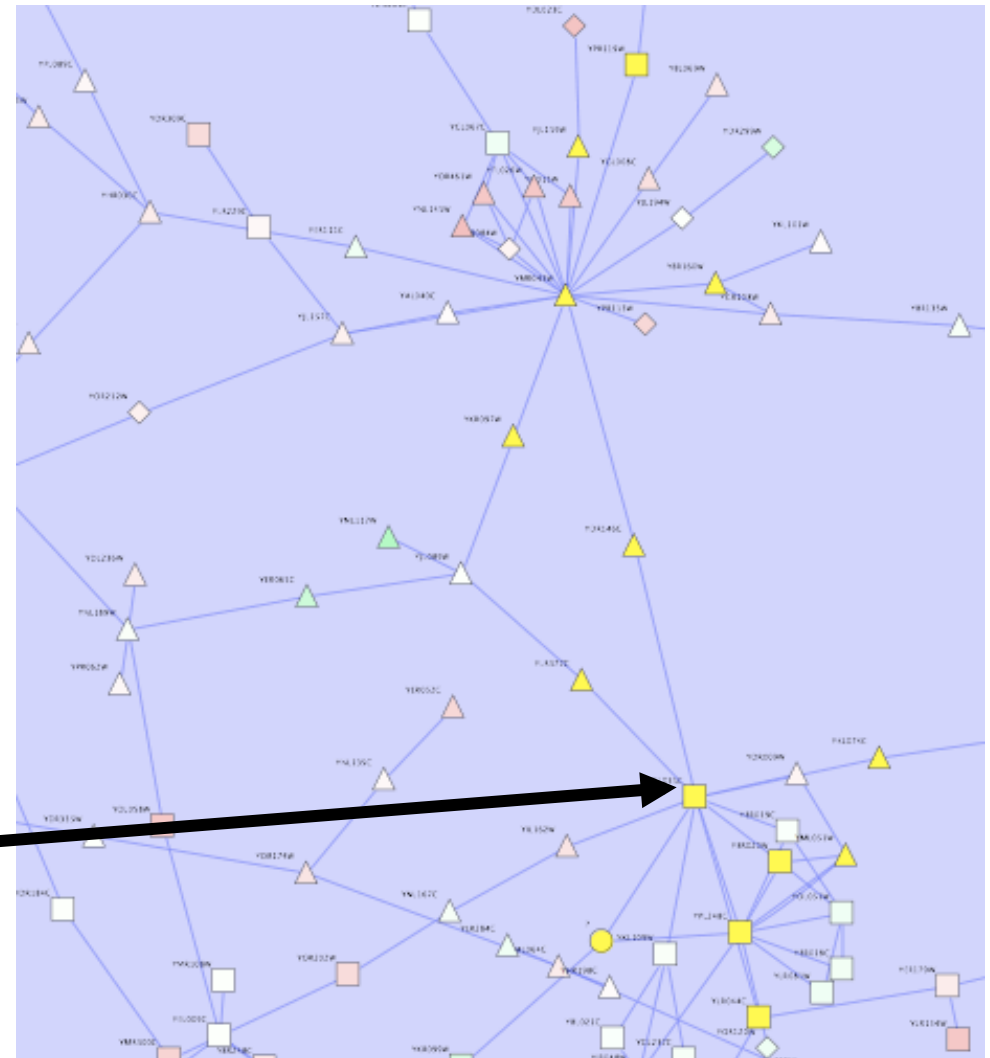
### jActiveModules plug-in

**Input:** interaction network and p-values for gene expression values over several conditions

**Output:** significant sub-networks that show differential expression over one or several conditions

Conditions vs. Pathways					
Network	Size	Score	gal1Rgsig	gal4Rgsig	gal80Rsig
1	14	3.78			
2	26	3.584			
3	10	2.994			
4	7	2.934			
5	4	2.636			

Save Dismiss





# Lab Time

- Try out jActiveModules
- Use the gal expression dataset





# Analysis Lab

## Find Network Motifs - Netmatch plugin

- Network motif is a sub-network that occurs significantly more often than by chance alone
- Input: query and target networks, optional node/edge labels
- Output: topological query matches as subgraphs of target network
- Supports: subgraph matching, node/edge labels, label wildcards, approximate paths
- <http://alpha.dmi.unict.it/~ctnyu/netmatch.html>



# Finding specific biological relevant TF-PPI sub-networks

The screenshot displays the NetMatch V1.0.1 software interface, divided into two main windows: the Query Editor and the Results window.

**Query Editor (Left Window):**

- Title:** NetMatch Query Editor - new query\*
- Query Edit:** Contains a toolbar with icons for file operations and a palette of motifs. A "Feed Forward Loop" motif is highlighted in the palette.
- Info:** A text area for additional information.
- Query:** A central workspace showing a network diagram with three nodes (represented by red circles with question marks) and three directed edges. A label "Query" is overlaid on this diagram.
- Bottom Status:** Nodes: 6 Edges: 6 Paths: 0 Loops: 0

**Results Window (Right Window):**

- Title:** NetMatch V1.0.1
- File Query Wizard Help:** Contains configuration options for graph and query properties.
- Graph Properties:** Labeled (checked), Directed (checked).
- Query Properties:** Query: Draw a query...; Selected: QE-FFL.
- Query Node Attributes:** QE-FFL - Nodes Attributes.
- Query Edge Attributes:** QE-FFL - Edges Attributes.
- Network Properties:** Network: 1-galFiltered.sif; Network Node Attributes: annotation.GO BIOLOGIC...; Network Edge Attributes: TextSourceInfo.
- Options:** Acquire Data, Go, Reset buttons.
- Match List:** A table with columns "Match Number", "Nodes", and "Image".
- Match 1:** Nodes: YMR309C, YOR361C, YPR041W. Image shows a network diagram with three nodes and two edges.
- Match 2:** Nodes: YOR310C, YDL014W, YLR197W. Image shows a network diagram with three nodes and two edges.
- Match 3:** Nodes: YDR100W, YGL161C, YOR036W. Image shows a network diagram with three nodes and two edges.
- Match 4:** Nodes: YIL015W, YMR043W, YCL067C. Image shows a network diagram with three nodes and two edges.
- Buttons:** "Create a new child network." (unchecked), "Save".
- Results List:** A scrollable list of matches with counts, such as "1 matches YBR020W", "2 matches YGL035C", etc. A label "Results" is overlaid on this list.

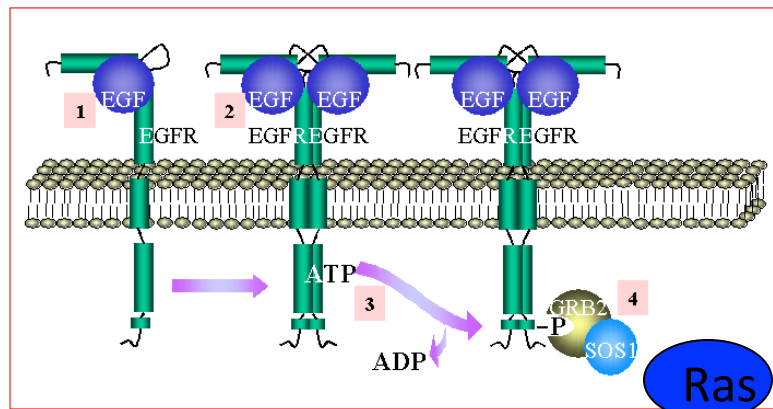


# Find Signaling Pathways

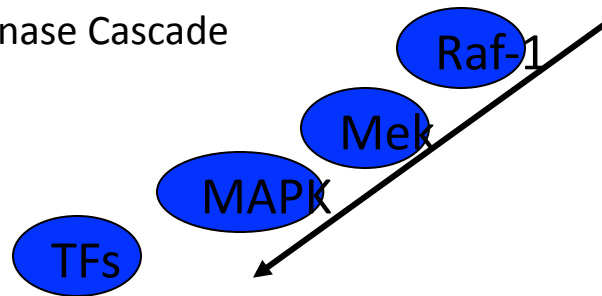
- Potential signaling pathways from plasma membrane to nucleus via cytoplasm

NetMatch Results

Signaling pathway example

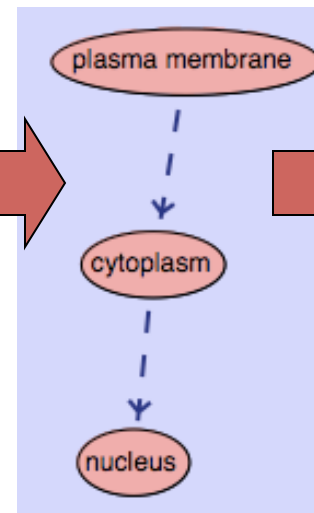


MAP Kinase Cascade



Nucleus - Growth Control  
Mitogenesis

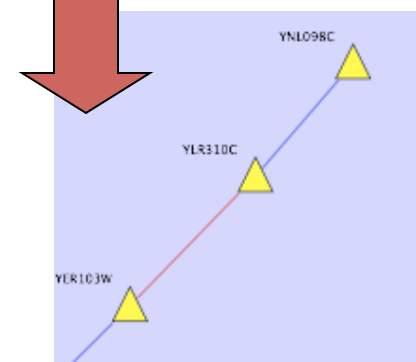
NetMatch query



Shortest path between subgraph matches

cytoscape.org

Match Number	Nodes	Image
	YGL008C	
4	YJL157C, YMR043W, YLR229C	
5	YJL157C, YAL040C, YLR229C	
6	YLR310C, YER103W, YNL098C	

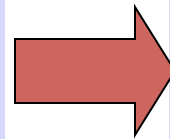
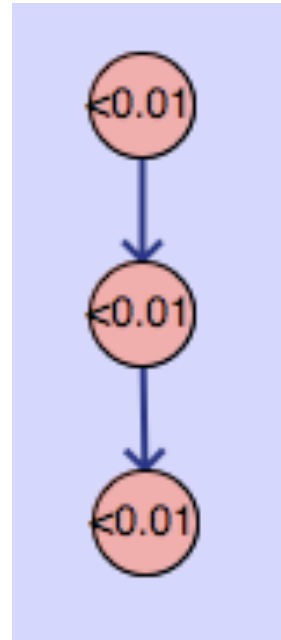




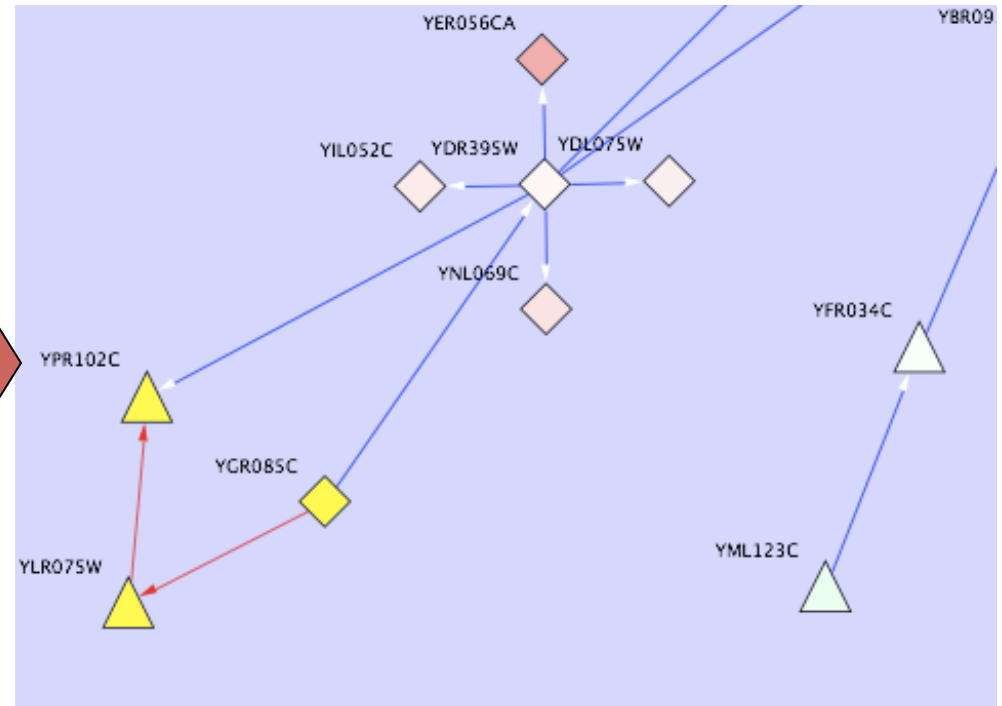
# Find Expressed Motifs

Find specific subgraphs where certain nodes are significantly differentially expressed

NetMatch query



NetMatch Results



Protein	Differential Expression Significance
YLR075W	1.7255E-4
YGR085C	2.639E-4
YPR102C	3.7183E-4



# Lab Time

## Find motifs with Netmatch

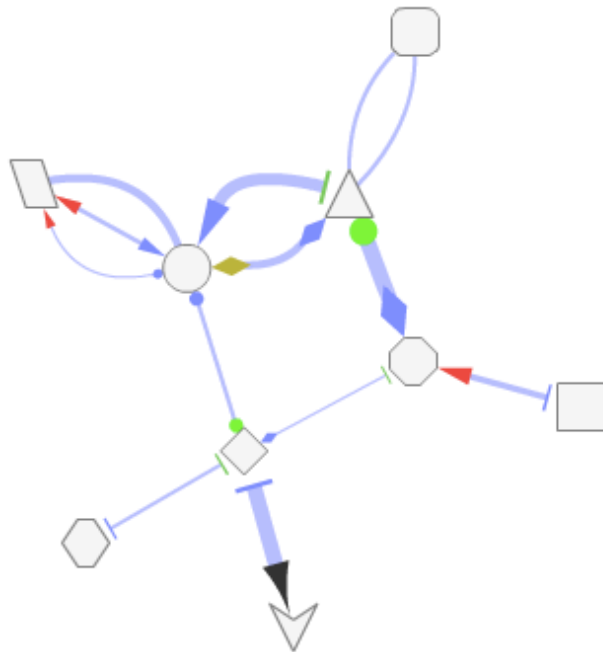
- Use the provided dataset (not yet in sampleData)



# Cytoscape 2.8

- Cytoscape 2.8 beta out now
  - [http://chianti.ucsd.edu/Cyto-2\\_8-beta/](http://chianti.ucsd.edu/Cyto-2_8-beta/)
  - Equations in attributes
  - Custom graphics
- Demo of new features

Save file    Open file    Style ▾    Layout ▾



Examples    Visual style    Filter    Properties

Shapes example



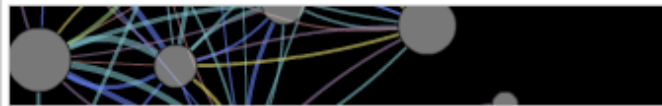
A graph that contains all possible shapes for nodes and arrows

Petersen example



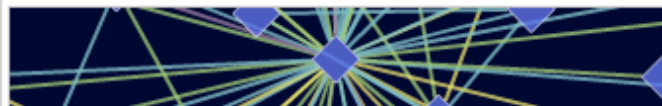
The Petersen graph

Disconnected example



A graph that contains several, disconnected components

Genetics example



A modified graph from GeneMANIA with different visual styles



Day 2



# Visualizing gene set enrichment analysis results

Gary Bader



**Donnelly Centre**  
for Cellular + Biomolecular Research



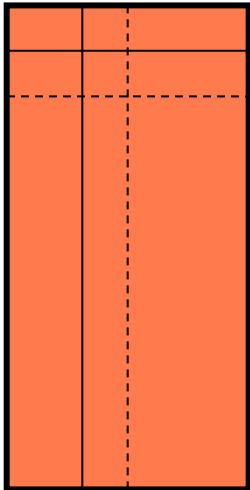
UNIVERSITY OF  
**TORONTO**

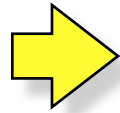


# Enrichment Test: General Framework

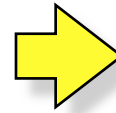
Experimental

Data

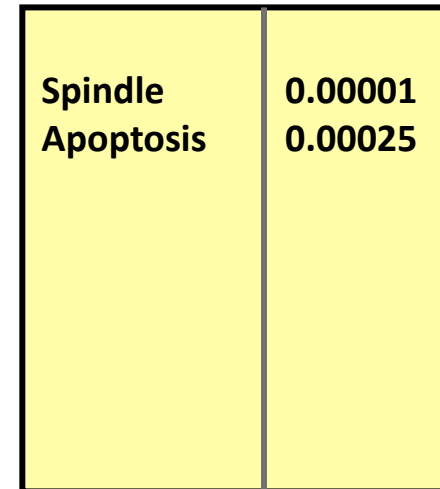


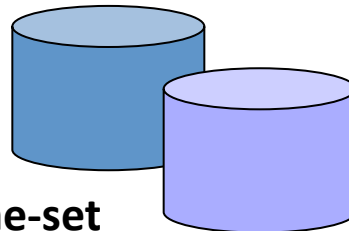
ENRICHMENT  
TEST



Enrichment Table



Spindle	0.00001
Apoptosis	0.00025



Gene-set  
Databases

- Excellent idea used to interpret data in hundreds of papers

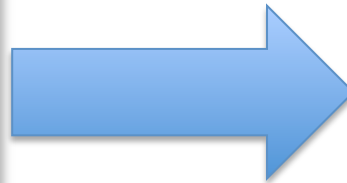
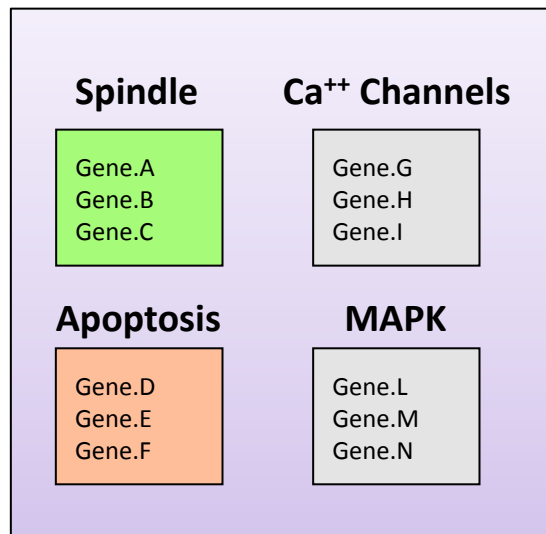
GO.id	GO.name	p.value	cover	cover.rat	Deg.mdn	Deg.iqr
GO:0042330	taxis	2.18E-06	23	0.056930693	54.94499375	9.139238998
<b>GO:0006935</b>	<b>chemotaxis</b>	<b>2.18E-06</b>	<b>23</b>	<b>0.060209424</b>	<b>54.94499375</b>	<b>9.139238998</b>
GO:0002460	adaptive immune response based on somatic recombination	7.10E-05	25	0.111111111	57.32306955	16.97054864
<b>GO:0002250</b>	<b>adaptive immune response</b>	<b>7.10E-05</b>	<b>25</b>	<b>0.111111111</b>	<b>57.32306955</b>	<b>16.97054864</b>
GO:0002443	leukocyte mediated immunity	0.000419328	23	0.097046414	58.27890582	15.58333739
<b>GO:0019724</b>	<b>B cell mediated immunity</b>	<b>0.000683758</b>	<b>20</b>	<b>0.114285714</b>	<b>57.84161096</b>	<b>15.03496347</b>
<b>GO:0030099</b>	<b>myeloid cell differentiation</b>	<b>0.000691589</b>	<b>24</b>	<b>0.089219331</b>	<b>62.22171598</b>	<b>10.35284833</b>
GO:0002252	immune effector process	0.000775626	31	0.090116279	58.27890582	23.86214773
GO:0050764	regulation of phagocytosis	0.000792138	8	0.2	53.54786293	5.742849971
GO:0050766	positive regulation of phagocytosis	0.000792138	8	0.216216216	53.54786293	5.742849971
GO:0002449	lymphocyte mediated immunity	0.00087216	22	0.101851852	57.84161096	16.13171132
<b>GO:0019838</b>	<b>growth factor binding</b>	<b>0.000913285</b>	<b>15</b>	<b>0.068181818</b>	<b>83.0405088</b>	<b>10.58734852</b>
GO:0051258	protein polymerization	0.00108876	17	0.080952381	57.97543252	17.31639968
<b>GO:0005789</b>	<b>endoplasmic reticulum membrane</b>	<b>0.001178198</b>	<b>18</b>	<b>0.036072144</b>	<b>64.02284752</b>	<b>12.05209158</b>
GO:0016064	immunoglobulin mediated immune response	0.001444464	19	0.113095238	58.27890582	15.58333739
<b>GO:0007507</b>	<b>heart development</b>	<b>0.001991562</b>	<b>26</b>	<b>0.052313883</b>	<b>84.02538284</b>	<b>18.60761304</b>
GO:0009617	response to bacterium	0.002552999	10	0.027173913	52.75249873	23.23104637
GO:0030100	regulation of endocytosis	0.002658555	11	0.099099099	56.38041132	16.02486889
GO:0002526	acute inflammatory response	0.002660742	24	0.103004292	57.80098769	24.94311116
GO:0045807	positive regulation of endocytosis	0.002903401	9	0.147540984	54.94499375	6.769909171
GO:0002274	myeloid leukocyte activation	0.002969661	7	0.077777778	54.94499375	16.07042339
<b>GO:0008652</b>	<b>amino acid biosynthetic process</b>	<b>0.003502921</b>	<b>7</b>	<b>0.017241379</b>	<b>45.19797271</b>	<b>31.18248579</b>
<b>GO:0050727</b>	<b>regulation of inflammatory response</b>	<b>0.004999055</b>	<b>7</b>	<b>0.084337349</b>	<b>54.94499375</b>	<b>7.737346076</b>
GO:0002253	activation of immune response	0.00500146	23	0.116161616	60.29679989	18.41103376
GO:0002684	positive regulation of immune system process	0.006581245	27	0.111570248	60.29679989	22.05051447
GO:0050778	positive regulation of immune response	0.006581245	27	0.113924051	60.29679989	22.05051447
GO:0019882	antigen processing and presentation	0.007244488	7	0.029661017	54.94499375	16.58797889
GO:0002682	regulation of immune system process	0.007252134	29	0.099656357	61.05645008	22.65935206
GO:0050776	regulation of immune response	0.007252134	29	0.102112676	61.05645008	22.65935206
GO:0043086	negative regulation of enzyme activity	0.008017022	9	0.040723982	53.28031076	17.48904224
GO:0006909	phagocytosis	0.008106069	10	0.080645161	55.66270253	12.47536747
GO:0002573	myeloid leukocyte differentiation	0.008174948	10	0.092592593	62.86577216	9.401887596
GO:0006959	humoral immune response	0.008396095	16	0.044568245	55.05654091	18.94209565
GO:0046649	lymphocyte activation	0.009044401	29	0.059917355	61.92213317	21.03553355
GO:0030595	leukocyte chemotaxis	0.009707319	7	0.101449275	56.33116709	6.945510559
GO:0006469	negative regulation of protein kinase activity	0.010782155	7	0.046357616	52.22863516	12.58524145
GO:0051348	negative regulation of transferase activity	0.010782155	7	0.04516129	52.22863516	12.58524145
<b>GO:0007179</b>	<b>transforming growth factor beta receptor signaling pathw</b>	<b>0.012630825</b>	<b>13</b>	<b>0.071038251</b>	<b>83.49440788</b>	<b>12.63256309</b>
<b>GO:0005520</b>	<b>insulin-like growth factor binding</b>	<b>0.012950071</b>	<b>9</b>	<b>0.097826087</b>	<b>81.41963394</b>	<b>7.528247832</b>
GO:0042110	T cell activation	0.013410548	20	0.064516129	59.77891783	26.06174863
GO:0002455	humoral immune response mediated by circulating immunogl	0.016780163	10	0.125	54.70766244	14.2572143
<b>GO:0005830</b>	<b>cytosolic ribosome (sensu Eukaryota)</b>	<b>0.016907351</b>	<b>8</b>	<b>0.01843318</b>	<b>61.68933284</b>	<b>7.814673781</b>

- Excellent idea used to interpret data in hundreds of papers
  - But! Major cognitive burden relating overlapping gene sets

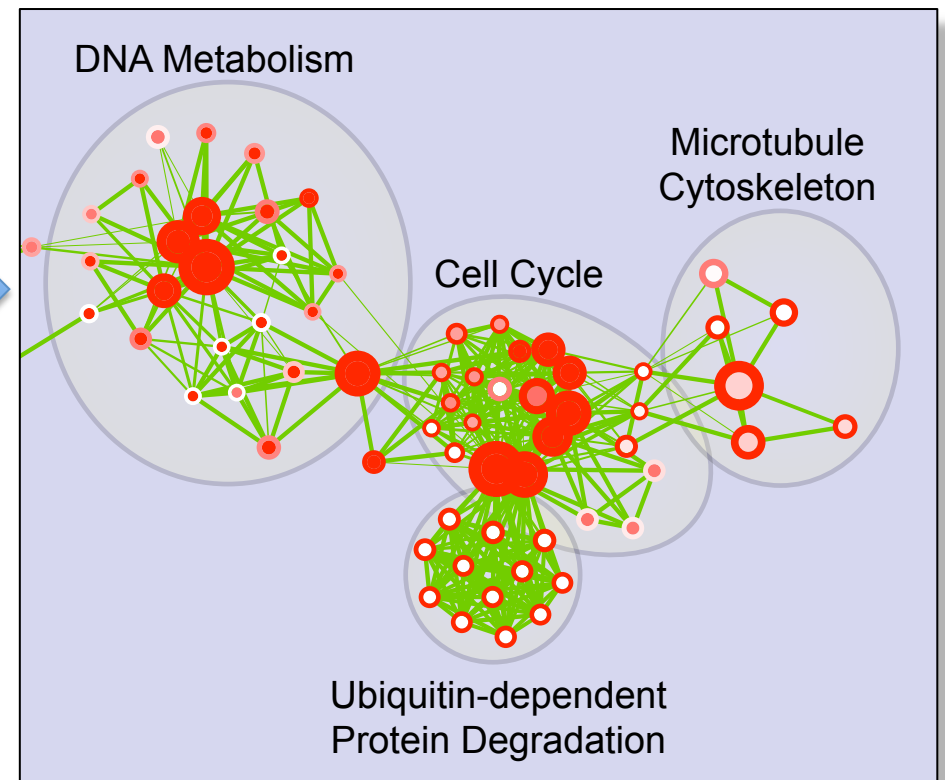
GO.id	GO.name	p.value	cover	cover.rat	Deg.mdn	Deg.iqr
GO:0042330	taxis	2.18E-06	23	0.056930693	54.94499375	9.139238998
GO:0006935	chemotaxis	2.18E-06	23	0.060209424	54.94499375	9.139238998
GO:0002460	adaptive immune response based on somatic recombination	7.10E-05	25	0.111111111	57.32306955	16.97054864
GO:0002250	adaptive immune response	7.10E-05	25	0.111111111	57.32306955	16.97054864
GO:0002443	leukocyte mediated immunity	0.000419328	23	0.097046414	58.27890582	15.58333739
GO:0019724	B cell mediated immunity	0.000683758	20	0.114285714	57.84161096	15.03496347
GO:0030099	myeloid cell differentiation	0.000691589	24	0.089219331	62.22171598	10.35284833
GO:0002252	immune effector process	0.000775626	31	0.090116279	58.27890582	23.86214773
GO:0050764	regulation of phagocytosis	0.000792138	8	0.2	53.54786293	5.742849971
GO:0050766	positive regulation of phagocytosis	0.000792138	8	0.216216216	53.54786293	5.742849971
GO:0002449	lymphocyte mediated immunity	0.00087216	22	0.101851852	57.84161096	16.13171132
GO:0019838	growth factor binding	0.000913285	15	0.068181818	83.0405088	10.58734852
GO:0051258	protein polymerization	0.00108876	17	0.080952381	57.97543252	17.31639968
GO:0005789	endoplasmic reticulum membrane	0.001178198	18	0.036072144	64.02284752	12.05209158
GO:0016064	immunoglobulin mediated immune response	0.001444464	19	0.113095238	58.27890582	15.58333739
GO:0007507	heart development	0.001991562	26	0.052313883	84.02538284	18.60761304
GO:0009617	response to bacterium	0.002552999	10	0.027173913	52.75249873	23.23104637
GO:0030100	regulation of endocytosis	0.002658555	11	0.099099099	56.38041132	16.02486889
GO:0002526	acute inflammatory response	0.002660742	24	0.103004292	57.80098769	24.94311116
GO:0045807	positive regulation of endocytosis	0.002903401	9	0.147540984	54.94499375	6.769909171
GO:0002274	myeloid leukocyte activation	0.002969661	7	0.077777778	54.94499375	16.07042339
GO:0008652	amino acid biosynthetic process	0.003502921	7	0.017241379	45.19797271	31.18248579
GO:0050727	regulation of inflammatory response	0.004999055	7	0.084337349	54.94499375	7.737346076
GO:0002253	activation of immune response	0.00500146	23	0.116161616	60.29679989	18.41103376
GO:0002684	positive regulation of immune system process	0.006581245	27	0.111570248	60.29679989	22.05051447
GO:0050778	positive regulation of immune response	0.006581245	27	0.113924051	60.29679989	22.05051447
GO:0019882	antigen processing and presentation	0.007244488	7	0.029661017	54.94499375	16.58797889
GO:0002682	regulation of immune system process	0.007252134	29	0.099656357	61.05645008	22.65935206
GO:0050776	regulation of immune response	0.007252134	29	0.102112676	61.05645008	22.65935206
GO:0043086	negative regulation of enzyme activity	0.008017022	9	0.040723982	53.28031076	17.48904224
GO:0006909	phagocytosis	0.008106069	10	0.080645161	55.66270253	12.47536747
GO:0002573	myeloid leukocyte differentiation	0.008174948	10	0.092592593	62.86577216	9.401887596
GO:0006959	humoral immune response	0.008396095	16	0.044568245	55.05654091	18.94209565
GO:0046649	lymphocyte activation	0.009044401	29	0.059917355	61.92213317	21.03553355
GO:0030595	leukocyte chemotaxis	0.009707319	7	0.101449275	56.33116709	6.945510559
GO:0006469	negative regulation of protein kinase activity	0.010782155	7	0.046357616	52.22863516	12.58524145
GO:0051348	negative regulation of transferase activity	0.010782155	7	0.04516129	52.22863516	12.58524145
GO:0007179	transforming growth factor beta receptor signaling pathw	0.012630825	13	0.071038251	83.49440788	12.63256309
GO:0005520	insulin-like growth factor binding	0.012950071	9	0.097826087	81.41963394	7.528247832
GO:0042110	T cell activation	0.013410548	20	0.064516129	59.77891783	26.06174863
GO:0002455	humoral immune response mediated by circulating immunogl	0.016780163	10	0.125	54.70766244	14.2572143
GO:0005830	cytosolic ribosome (sensu Eukaryota)	0.016907351	8	0.01843318	61.68933284	7.814673781

# Enrichment Map

## GENE SETS

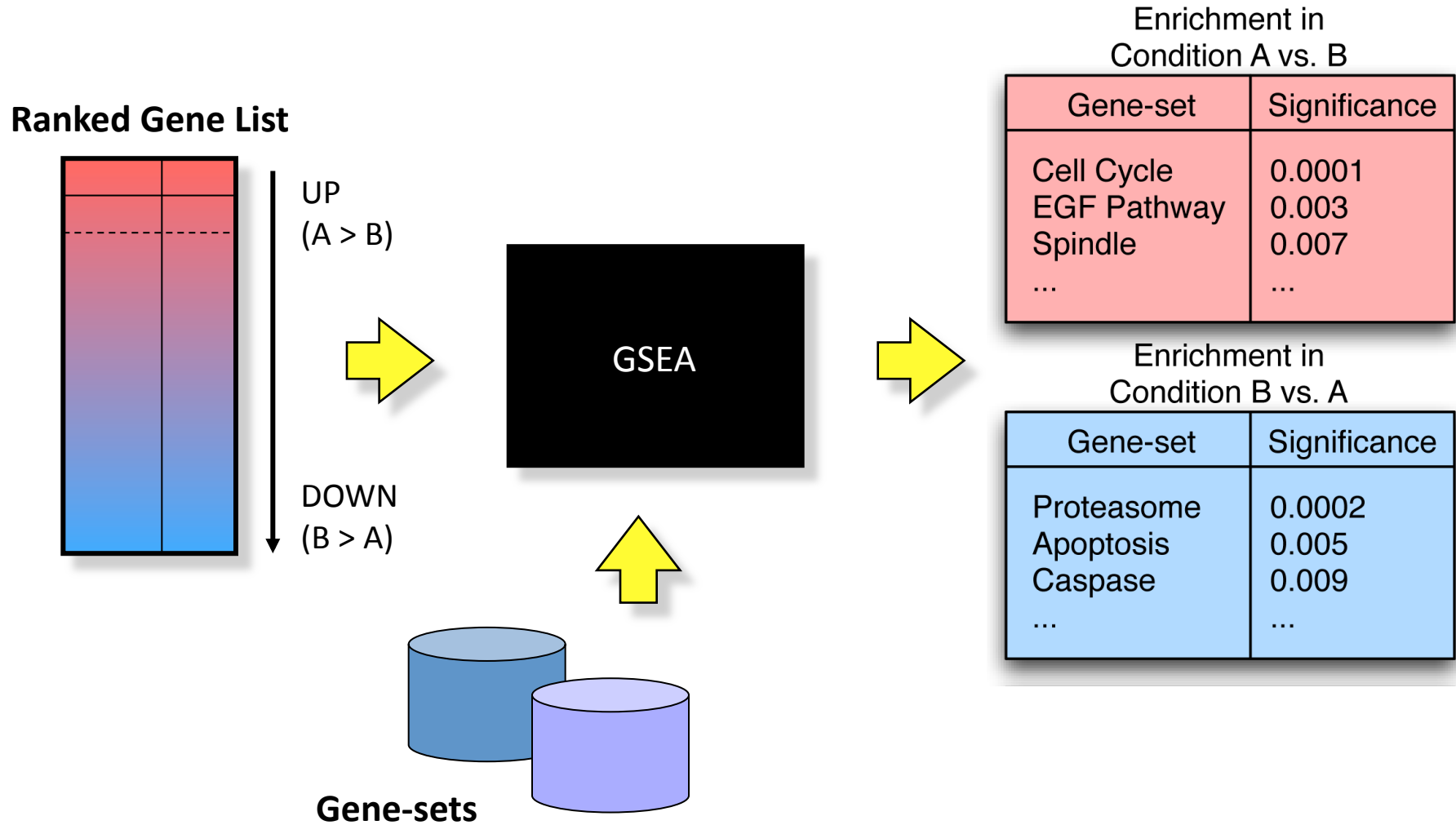


## ENRICHMENT MAP



- Use available gene-set scoring models
  - threshold dependent (e.g. Fisher's) or threshold free (e.g. GSEA, GOrilla)
- Use the network framework to organize gene-sets exploiting their inter-dependencies

# Gene Set Enrichment Analysis (GSEA)



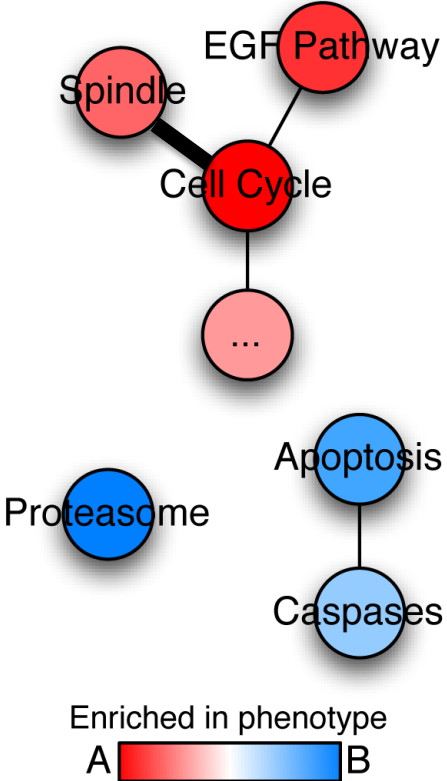
# Enrichment Map

Enrichment in Condition A vs. B

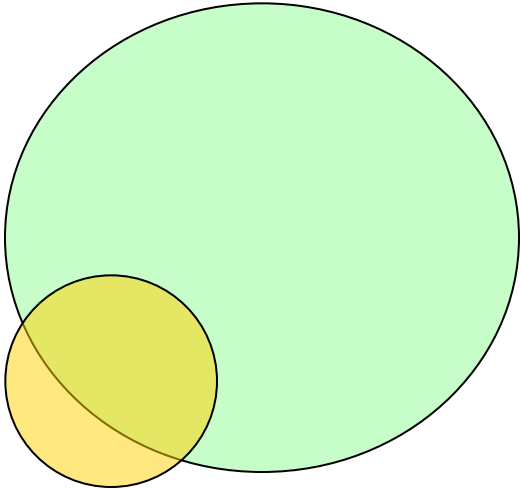
Gene-set	Significance
Cell Cycle	0.0001
EGF Pathway	0.003
Spindle	0.007
...	...

Enrichment in Condition B vs. A

Gene-set	Significance
Proteasome	0.0002
Apoptosis	0.005
Caspase	0.009
...	...



Overlap



$$\frac{|A \cap B|}{\min(|A|, |B|)}$$

GENE-SET LIST



ENRICHMENT MAP

# Enrichment Map: use case I

## *Single enrichment*

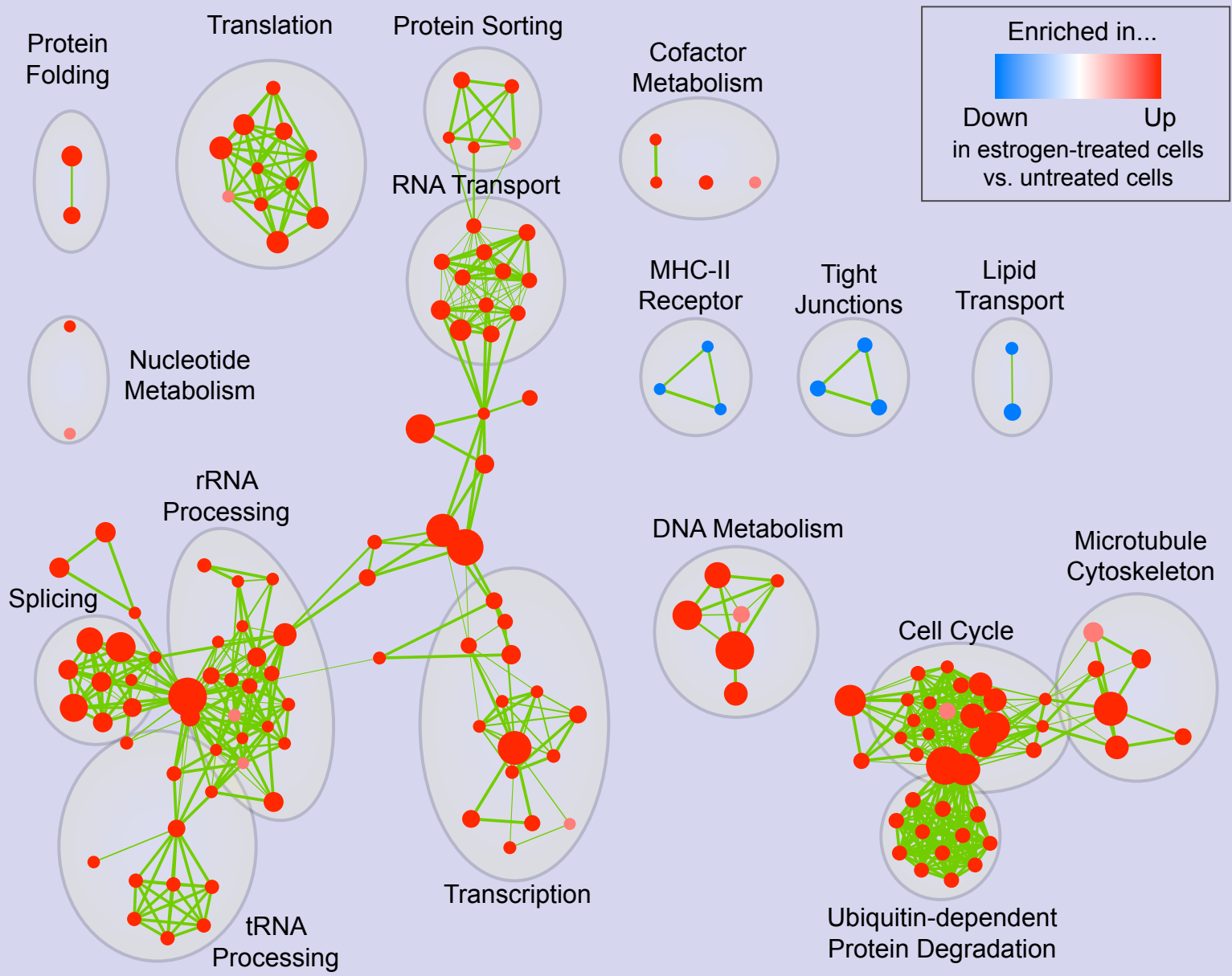
Estrogen treatment of breast cancer cells

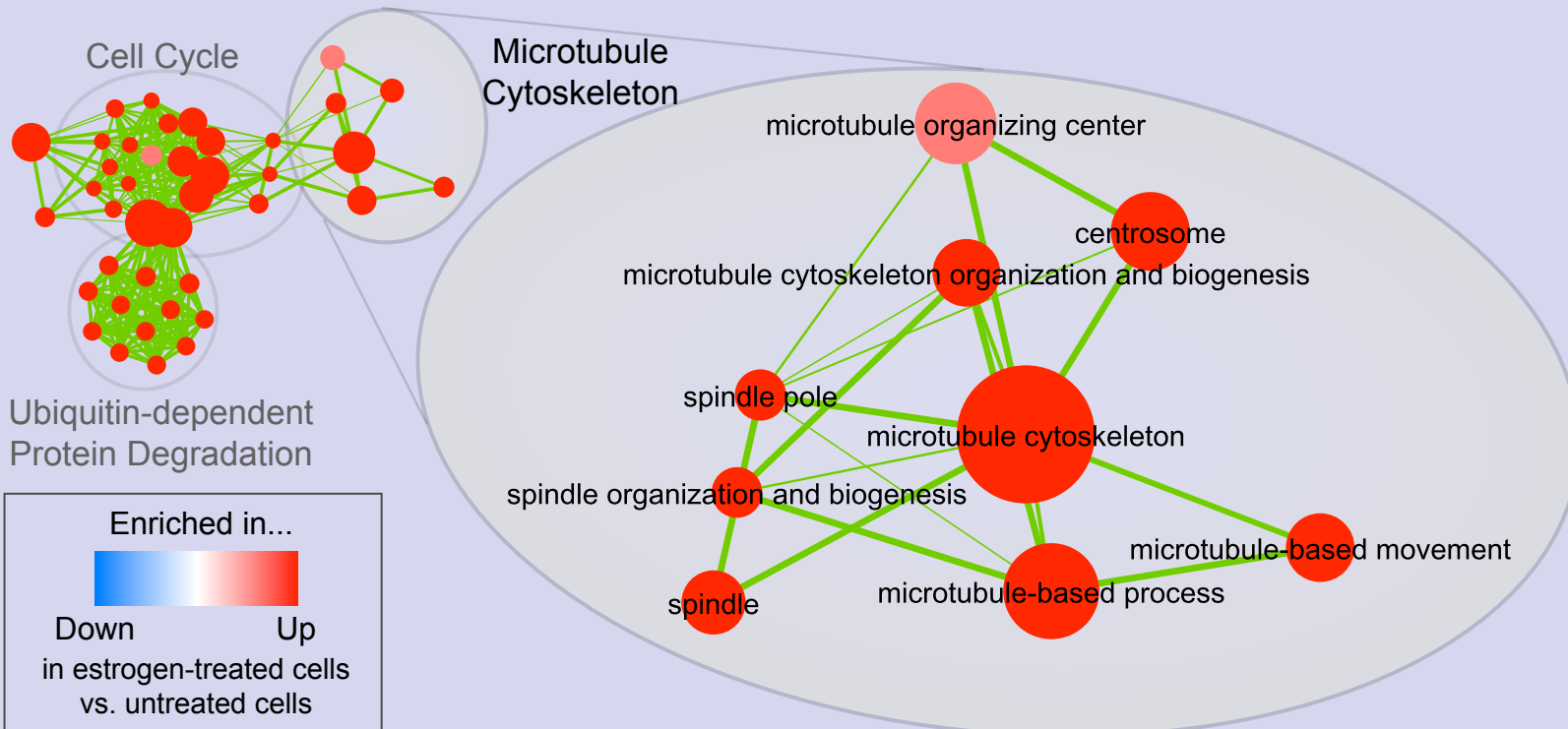
- Design:  
**2-time points, two-class**

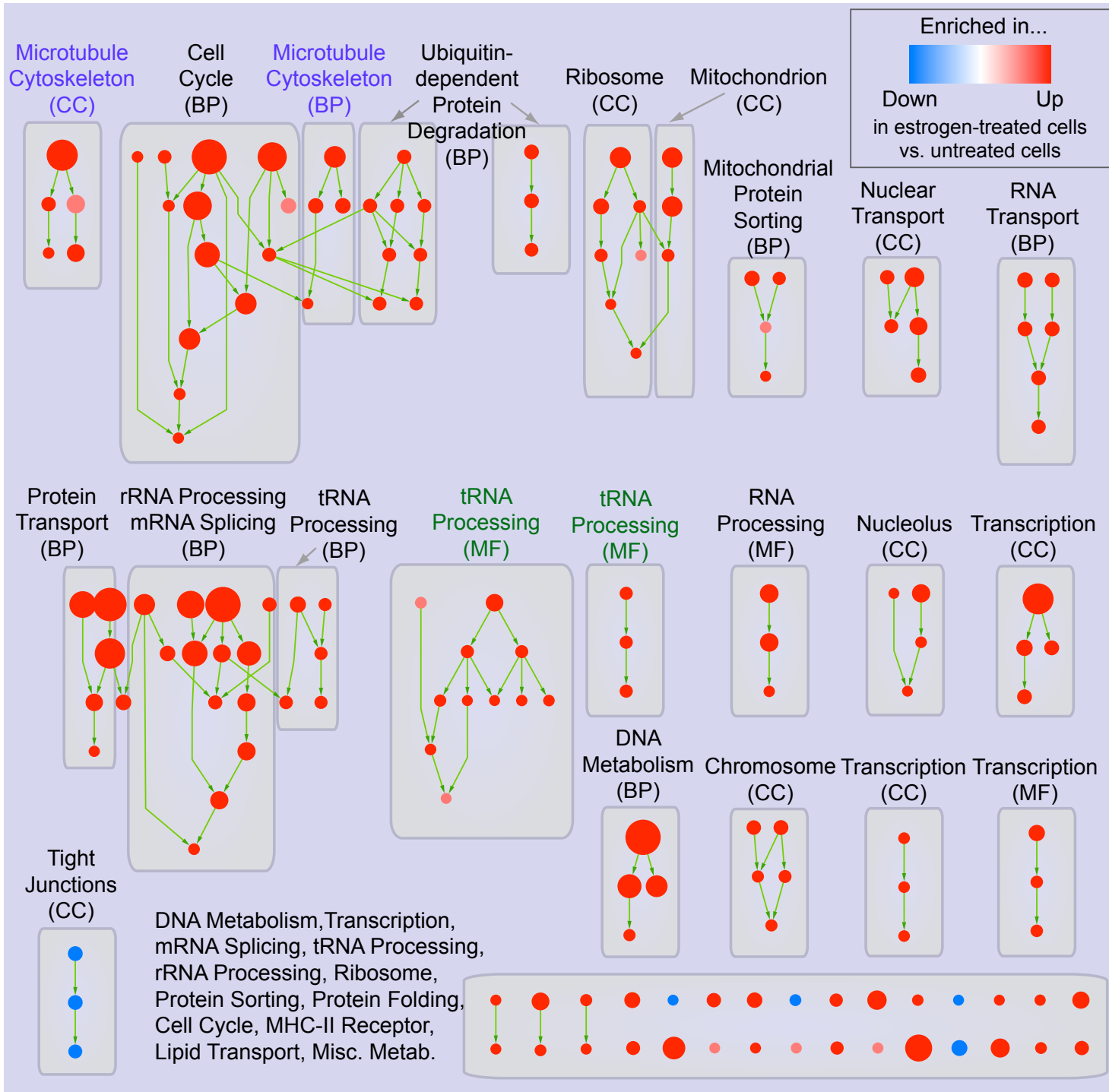
	12 hrs	24 hrs
Estrogen-treated	3	3
Untreated	3	3

- Gene set Database:  
**Gene Ontology**









# Enrichment Map: use case II

## *Comparison of two enrichments*

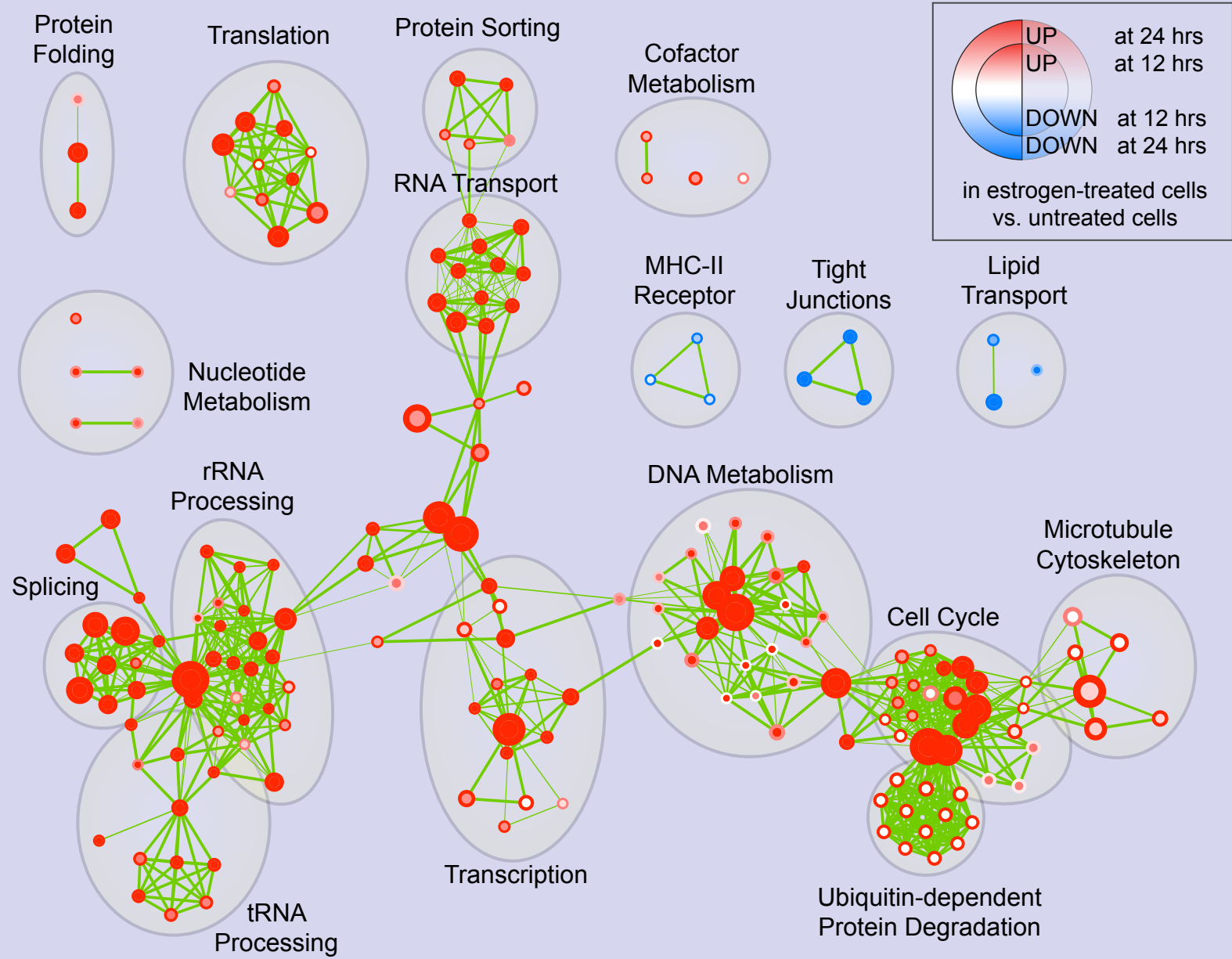
Estrogen treatment of breast cancer cells

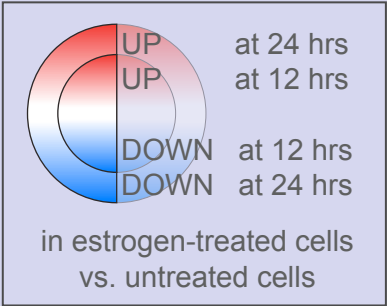
- Design:

**2-time points, two-class**

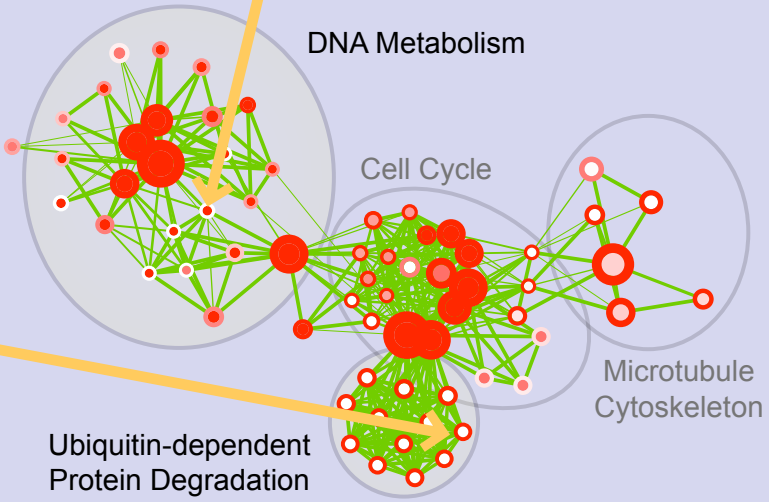
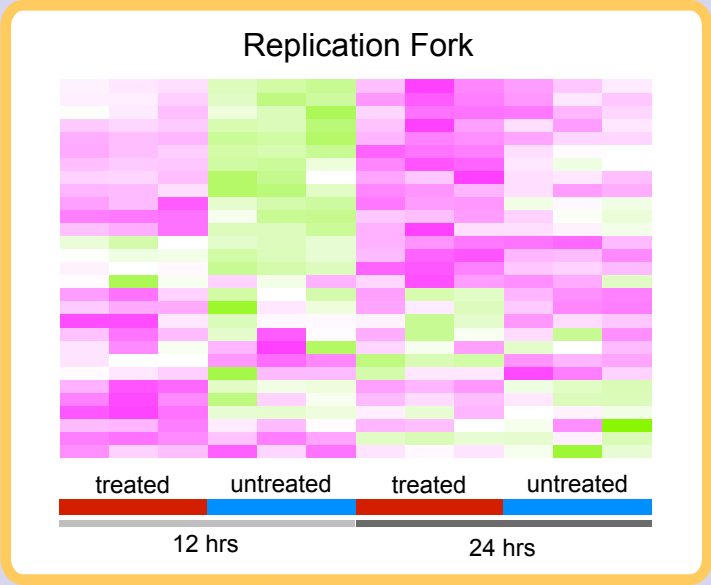
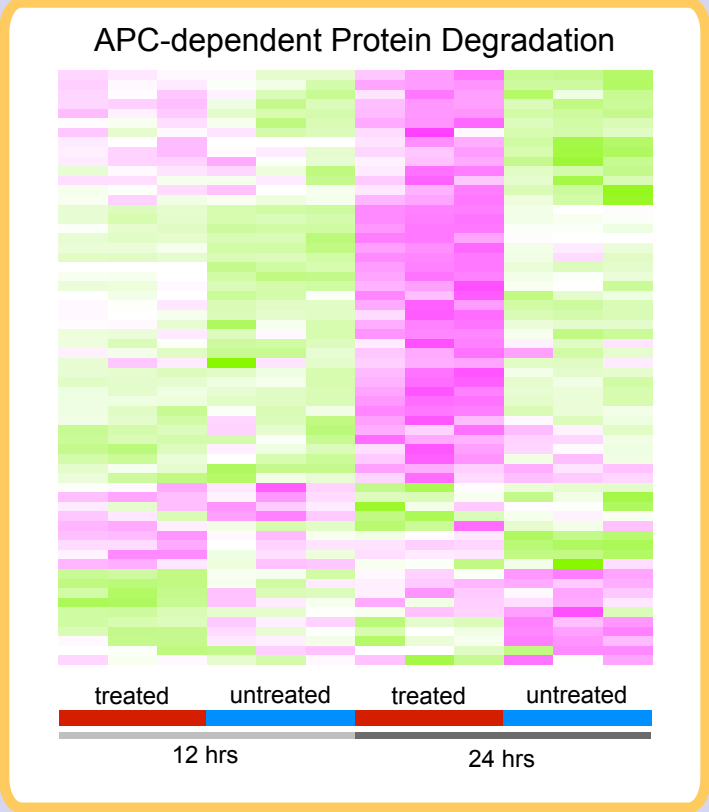
	12 hrs	24 hrs
Estrogen-treated	3	3
Untreated	3	3

- Gene set Database:  
**Gene Ontology**



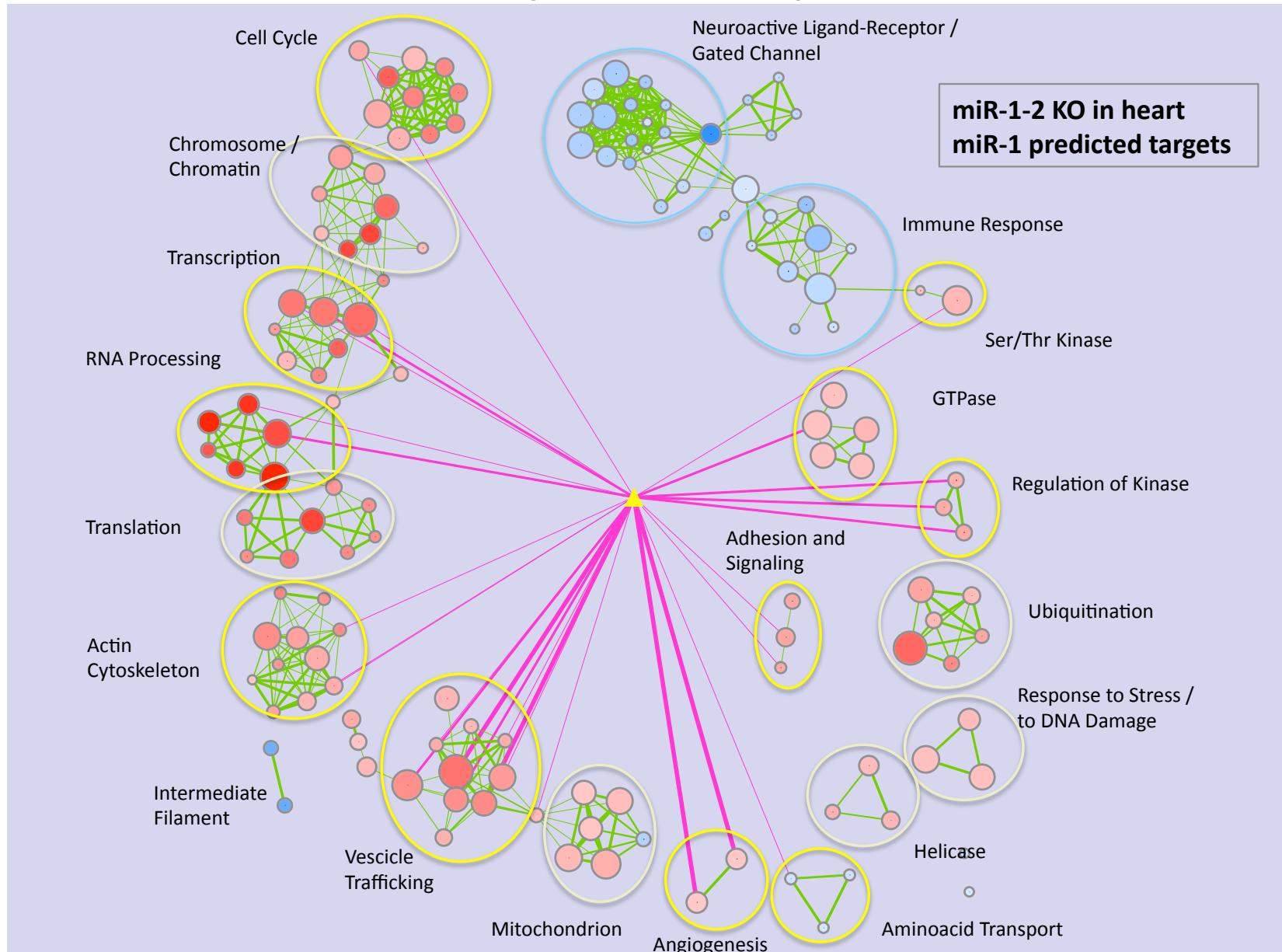


Estrogen-treated vs untreated:  
 ■ Up  
 ■ Down



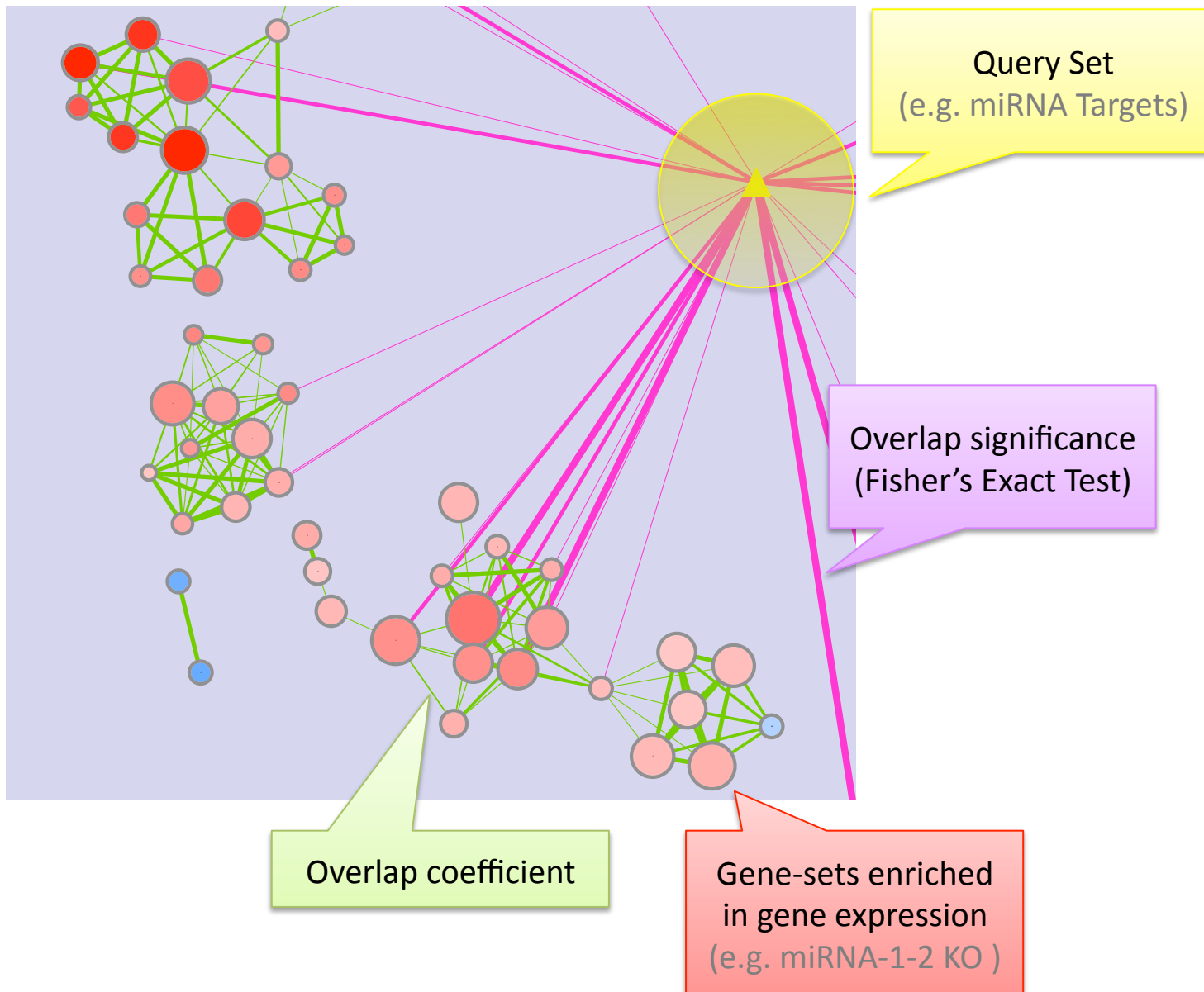
# Enrichment Map: use case III

## Query Set Analysis



# Enrichment Map: use case III

## *Query Set Analysis*





# Autism

- Autism Spectrum Disorder (ASD) definition:
- Genetics
  - highly heritable
    - monozygotic twin concordance 90-60%  
(depending on the stringency of diagnosis)
  - known genetics:
    - 5-15% rare single-gene disorders and chromosomal re-arrangements
    - de-novo CNV previously reported in 5-10% of ASD cases
    - GWA (Genome-wide Association Studies) have been able to explain only a small amount of heritability

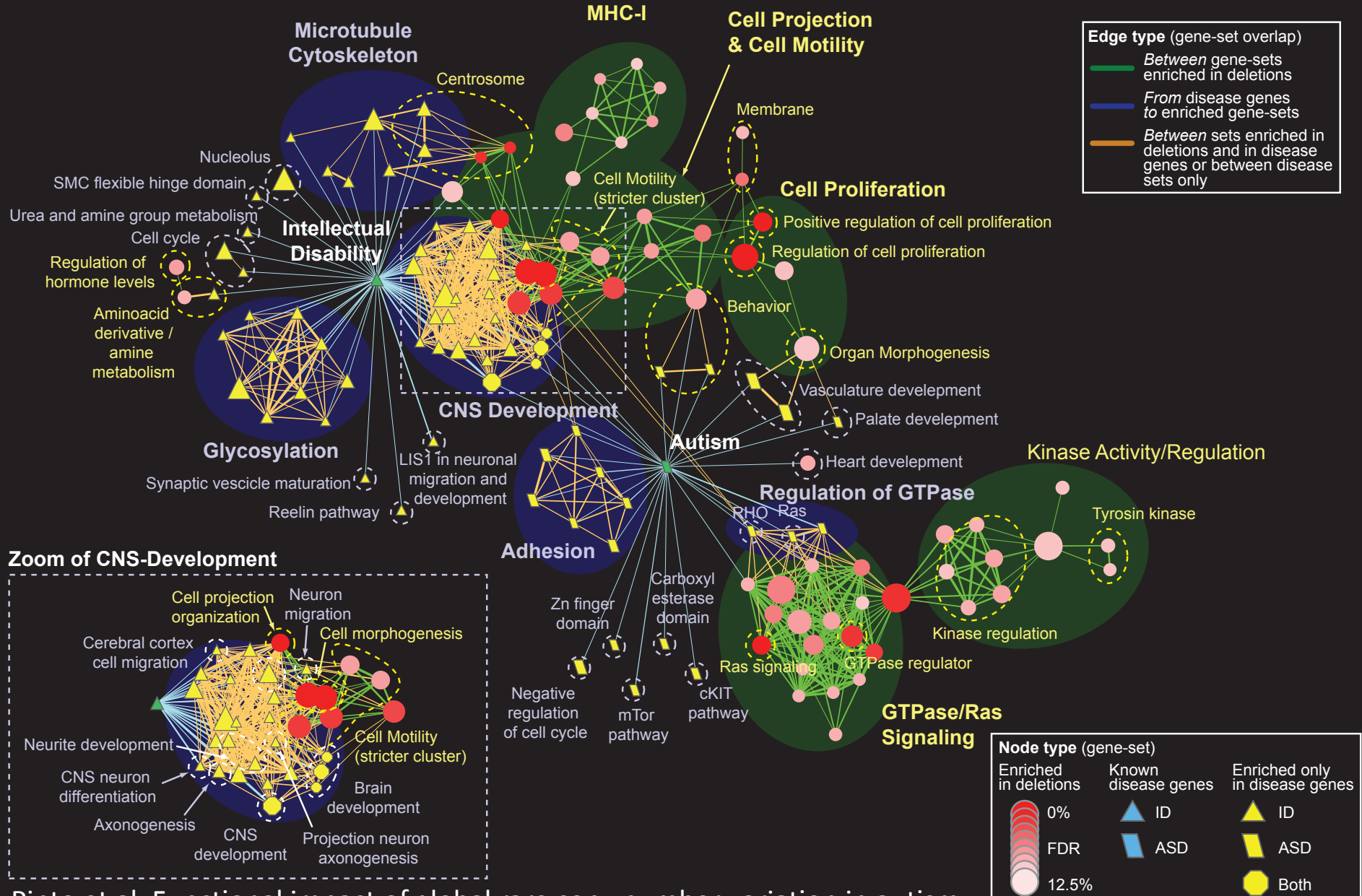
# CNV screening of Autism

- Rare Copy Number Variation screening (DEL, DUP)
  - 889 Case and 1146 Ctrl (European Ancestry)
  - Illumina Infinium 1M-single SNP
  - high quality rare CNV (90% PCR validation)
    - identification by the three main algorithms required for detection (QuantiSNP, iPattern, PennCNV)
    - frequency < 1%
    - length > 30 kb
- Statistics
  - average CNV size: 182.7 kb
  - median CNV number per individual: 2
  - difference in CNV distribution: only genic DEL higher
  - > 5.7% ASD individuals carry at least one de-novo CNV

# Gene-set sources

- **Gene Ontology**
  - Biological Process
  - Cellular Component
  - Molecular Function
- **Pathways**
  - KEGG
  - NCI
  - Reactome
- **PFAM domains**
- Number of gene-sets:
  - Unfiltered (all): 14,433
  - Filtered (5 << 700 genes): 6,129
  - Tested (counts > 0): 3,493

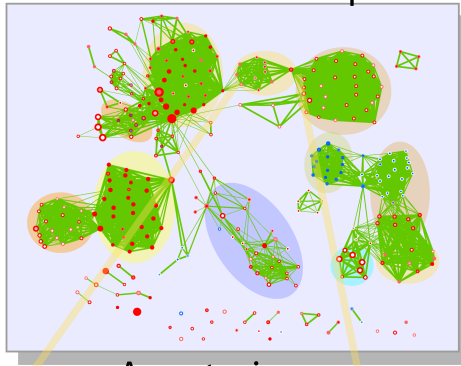
Pinto et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 2010 Jun 9.



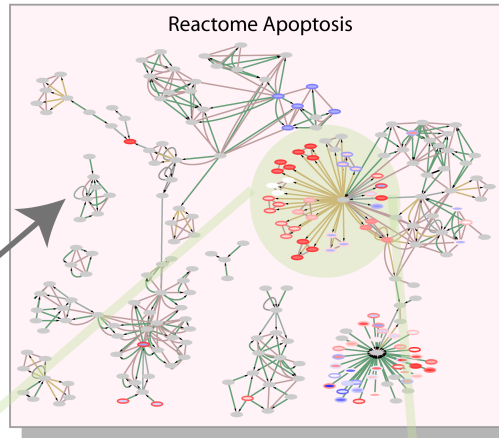
Pinto et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 2010 Jun 9.



## Enrichment Map



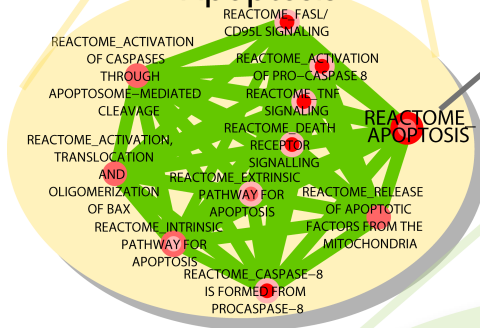
## Pathway Network



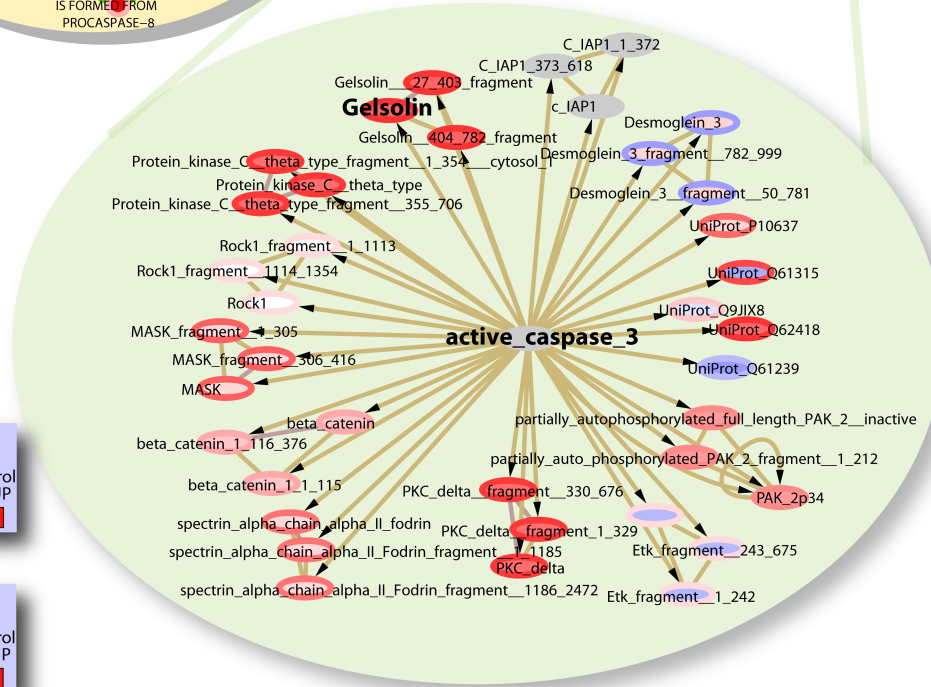
## Future Work

- Add Network visualization support
- Pathway visualization and analysis tools

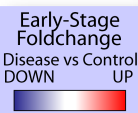
## Apoptosis



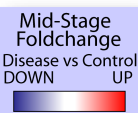
## Caspase Neighbourhood

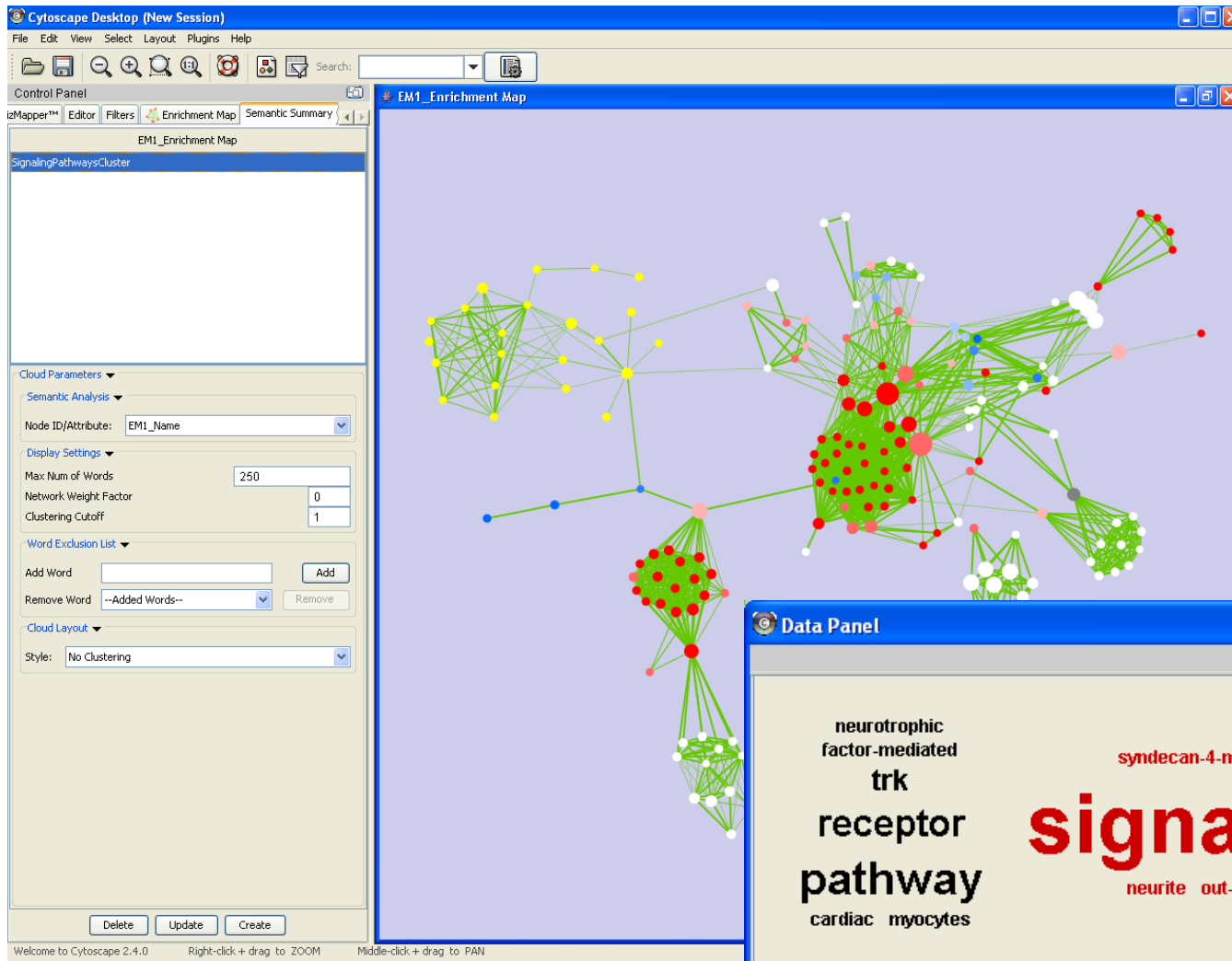


### Inner Circle

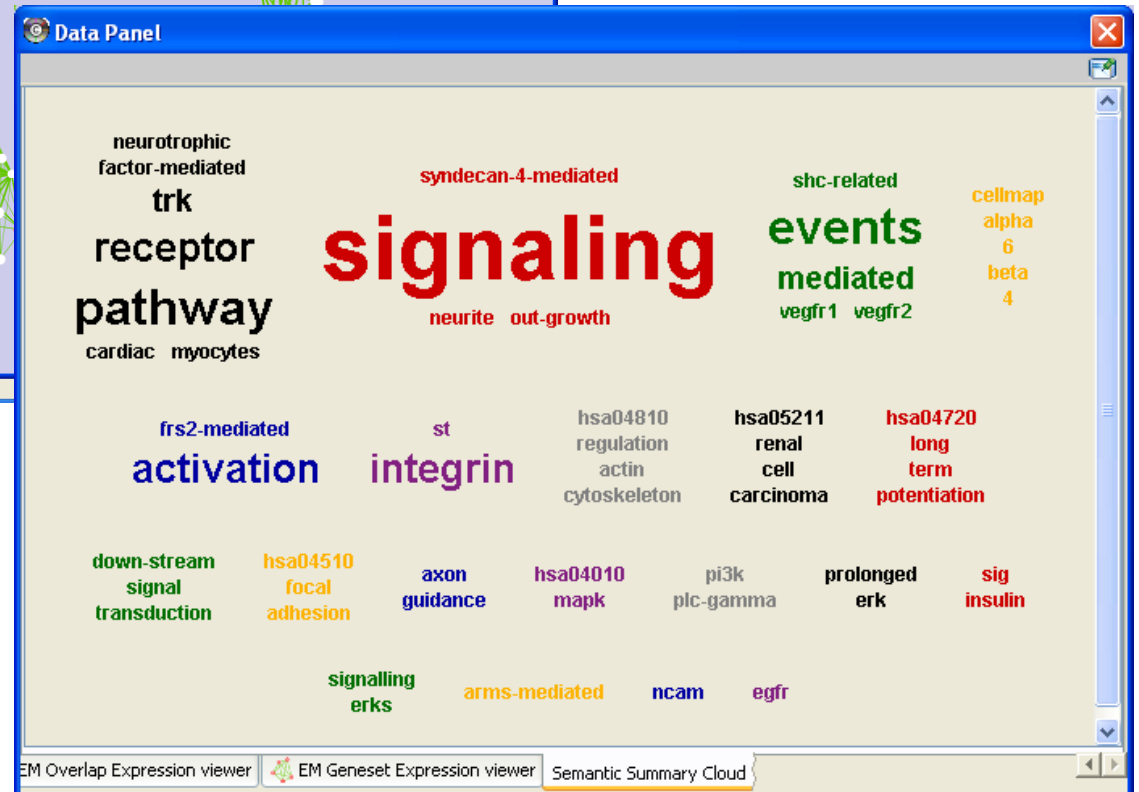


### Outer Circle

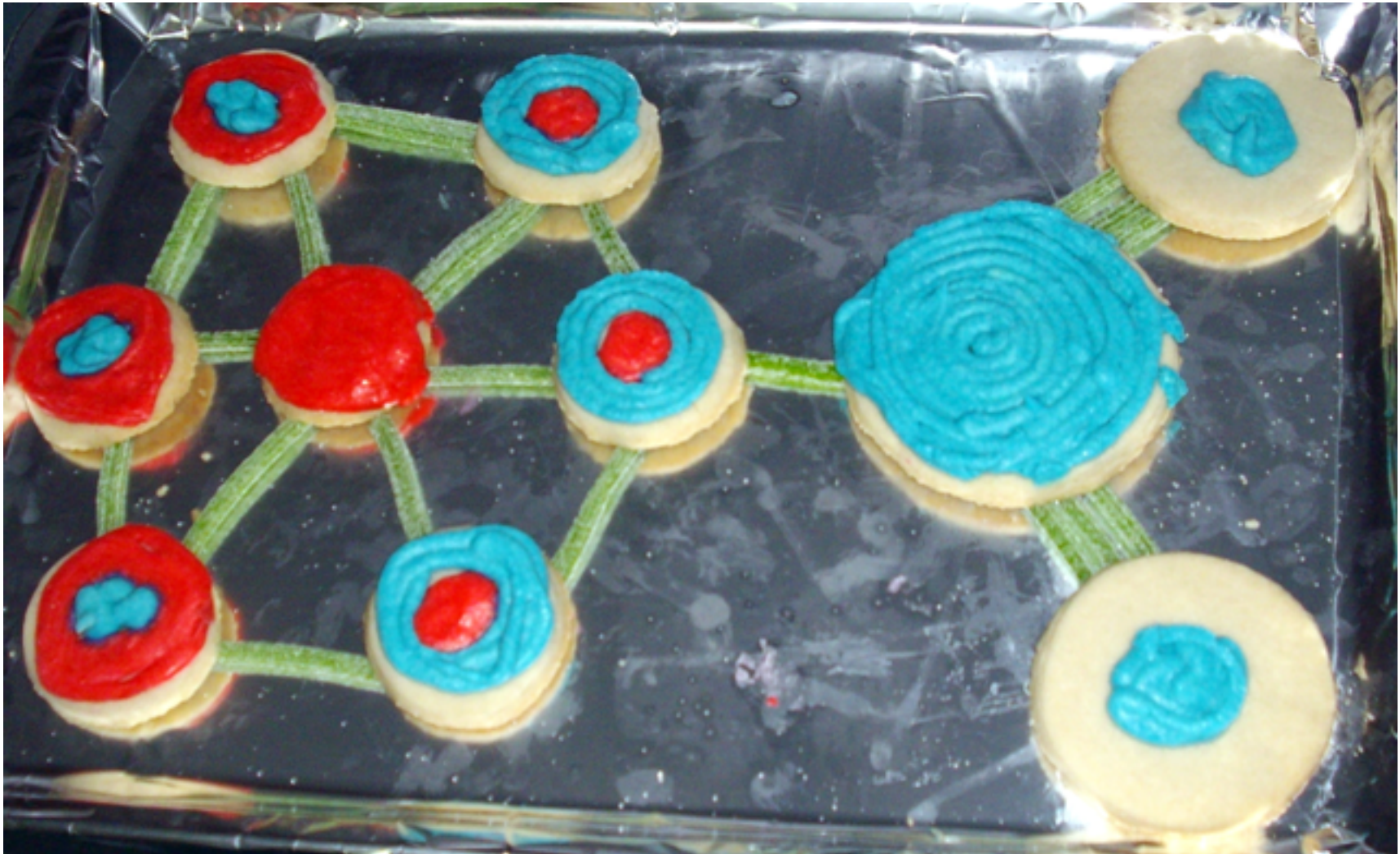




Cluster words according to their co-occurrence in labels to preserve semantic meaning.



Layla Oesper  
Google Summer of Code

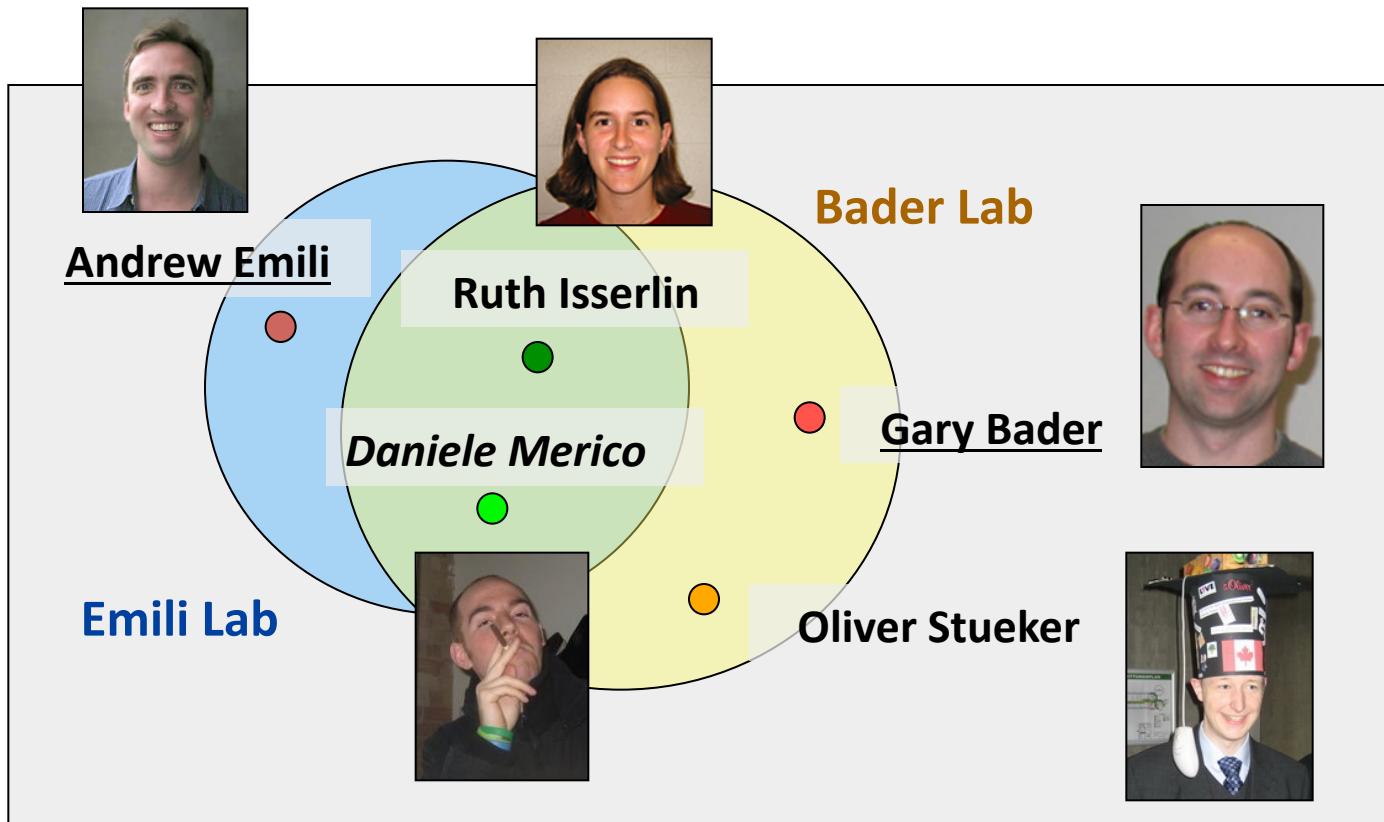


**Thanks for your attention!**



# Enrichment Map

## *Acknowledgments*



**Terrence Donnelly CCBR  
(University of Toronto)**

# Lab Time

- Try out Cytoscape
- [http://opentutorials.rbvi.ucsf.edu/index.php/Tutorial:Introduction\\_to\\_Cytoscape](http://opentutorials.rbvi.ucsf.edu/index.php/Tutorial:Introduction_to_Cytoscape)