

Network Visualization and Analysis Outline

- Network introduction
- Network visualization
- Cytoscape software tool for network visualization and analysis
- Network analysis

Module 3: Network Visualization with Cytoscape

bioinformatics.ca

_computational
BIOLOGY

PRIMER

How to visually interpret biological data using networks

Daniele Merico, David Gfeller & Gary D Bader

Networks in biology can appear complex and difficult to decipher. We illustrate how to interpret biological networks with the help of frequently used visualization and analysis patterns.

Module 3: Network Visualization with Cytoscape

bioinformatics.ca

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
- Network analysis

Module 3: Network Visualization with Cytoscape bioinformatics.ca

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?

<http://www.time.com/time/techtime/200406/community.html>

Module 3: Network Visualization with Cytoscape bioinformatics.ca

Applications of Network Biology

[a] JActiveModules, UCSD

JActiveModules, UCSD

[b] PathBlast, UCSD

PathBlast, UCSD

[c] MCODE, University of Toronto

MCODE, University of Toronto

[d] DomainGraph, Max Planck Institute

DomainGraph, Max Planck Institute

- **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

humangenetics-amc.nl

Applications of Network Informatics in Disease

[a] Agilent Literature Search

Agilent Literature Search

[b] Mondrian, MSKCC

Mondrian, MSKCC

[c] PinnacleZ, UCSD

PinnacleZ, UCSD

- **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

humangenetics-amc.nl

What's Missing?

- Dynamics
 - Pathways/networks represented as static processes
 - Difficult to represent a calcium wave or a feedback loop
 - More detailed mathematical representations exist that handle these e.g. Stoichiometric modeling, Kinetic modeling (VirtualCell, E-cell, ...)
 - Need to accumulate or estimate comprehensive kinetic information
- Detail – atomic structures
- Context – cell type, developmental stage

What Have We Learned?

- 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
 - Good to determine your question and search for a solution
 - Or get to know many methods and see how they can be applied to your data

Network Visualization and Analysis using Cytoscape

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

Module 3: Network Visualization with Cytoscape

bioinformatics.ca

<http://cytoscape.org> UCSD, ISB, Agilent, MSKCC,
Pasteur, UCSF

Network visualization and analysis

Pathway comparison

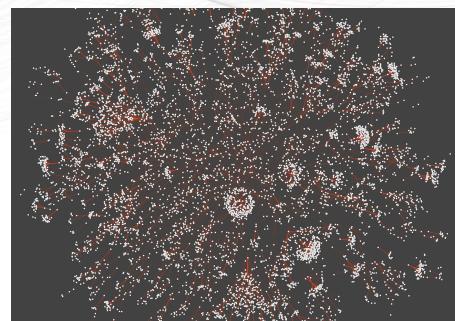
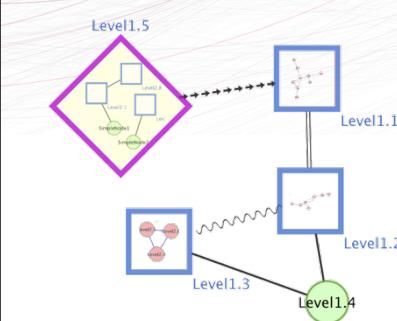
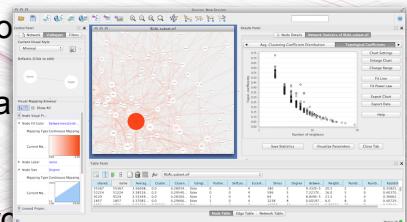
Literature mining

Gene Ontology ana

Active modules

Complex detection

Network motif search..



Manipulate Networks

Automatic Layout

Filter/Query

Interaction Database Search

The Cytoscape App Store

http://apps.cyto.org

- network generation**: DiGeNET, GENE MANIA, LitSearch, BioGRID, xDrill, UniProt, UniProtKB, CABIN, R
- online data import**: BioGRID, xDrill, UniProt, UniProtKB, CABIN, R, Pathway analysis, Gene expression analysis, Complex detection, Literature mining, Network motif search, Pathway comparison
- graph analysis**: CluePedia, CPDB, Super, JG, Wall of Apps, StarNet, RIN, CTL, MetScape, CML, nbIC, PC, G3D

Active Community

<http://www.cytoscape.org>

- 10,000s users, >8,000 downloads/month
- Help
 - Documentation, data sets
 - Mailing lists
 - <http://tutorials.cytoscape.org>
- Cline MS et al. Integration of biological networks and gene expression data using Cytoscape Nat Protoc. 2007;2(10):2366-82
- Annual Conference
- >200 Apps Extend Functionality
 - Build your own, requires programming

Module 3: Network Visualization with Cytoscape

bioinformatics.ca



What Have We Learned?

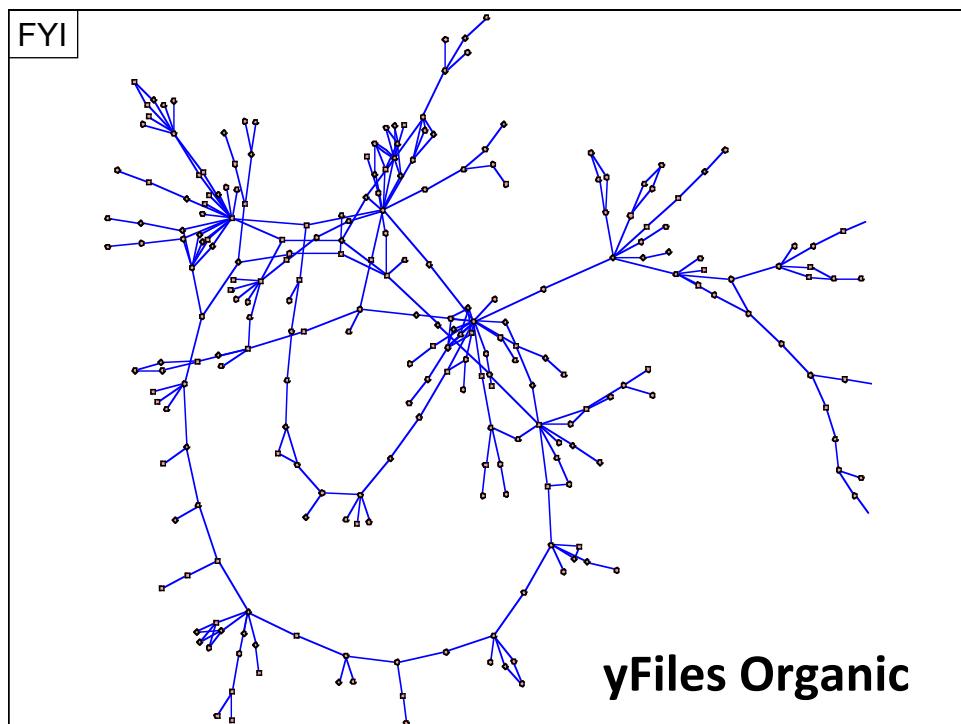
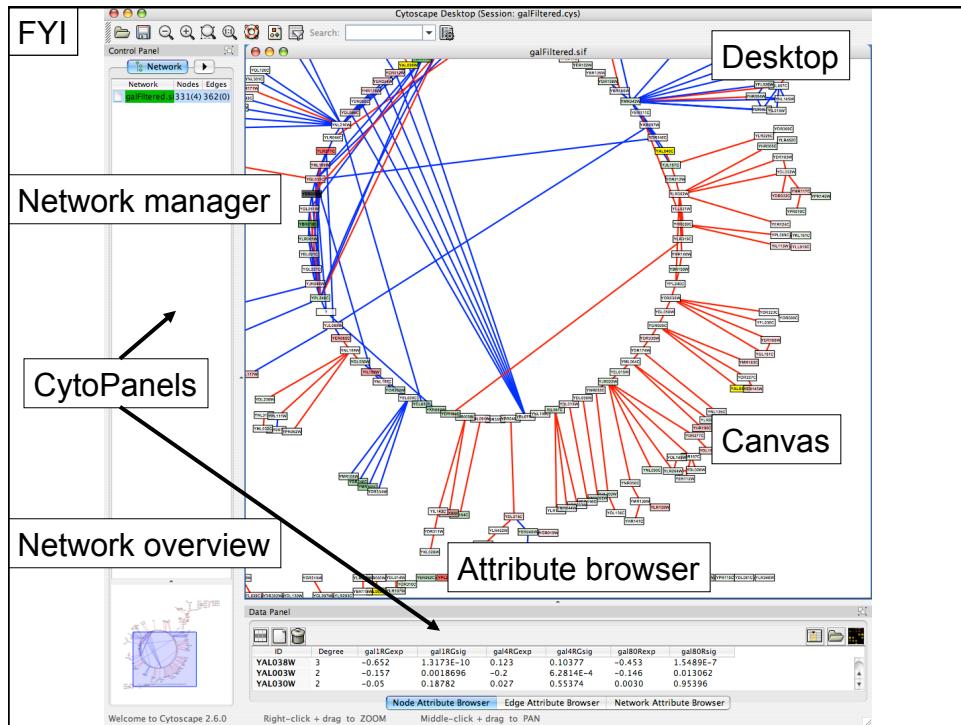
- Cytoscape is a useful, free software tool for network visualization and analysis
- Provides basic network manipulation features
- Apps are available to extend the functionality

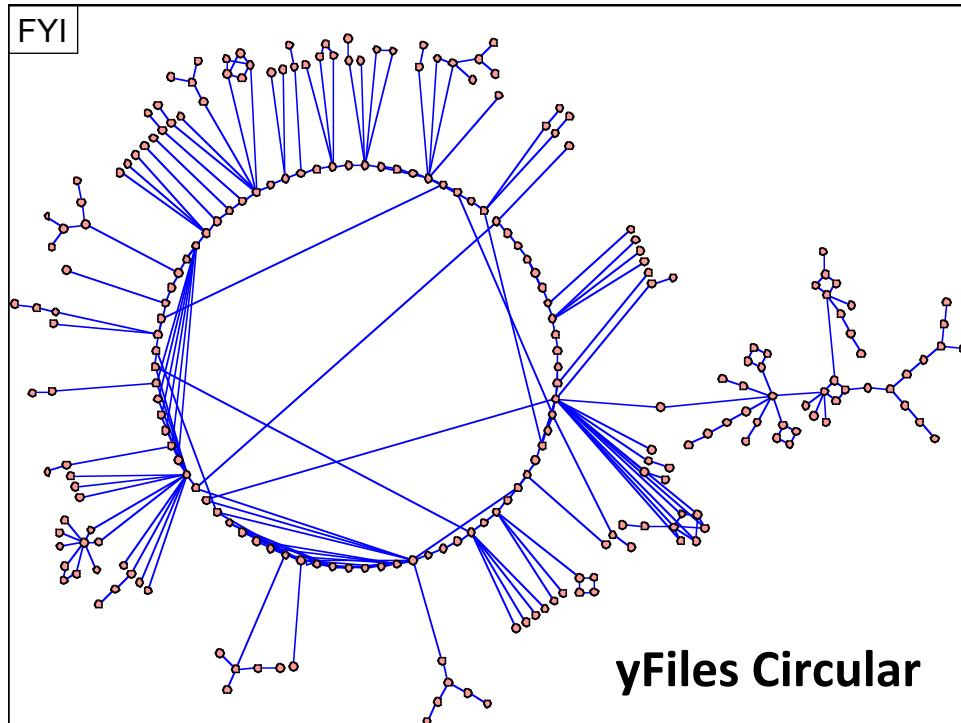
Module 3: Network Visualization with Cytoscape

bioinformatics.ca

Cytoscape Demo

Version 3.2.1
www.cytoscape.org





FYI

Network Layout

- Many algorithms available through apps
- Demo: Move, zoom/pan, rotate, scale, align

Rotate
Scale
Align and Distribute

Rotate
Scale
Align and Distribute

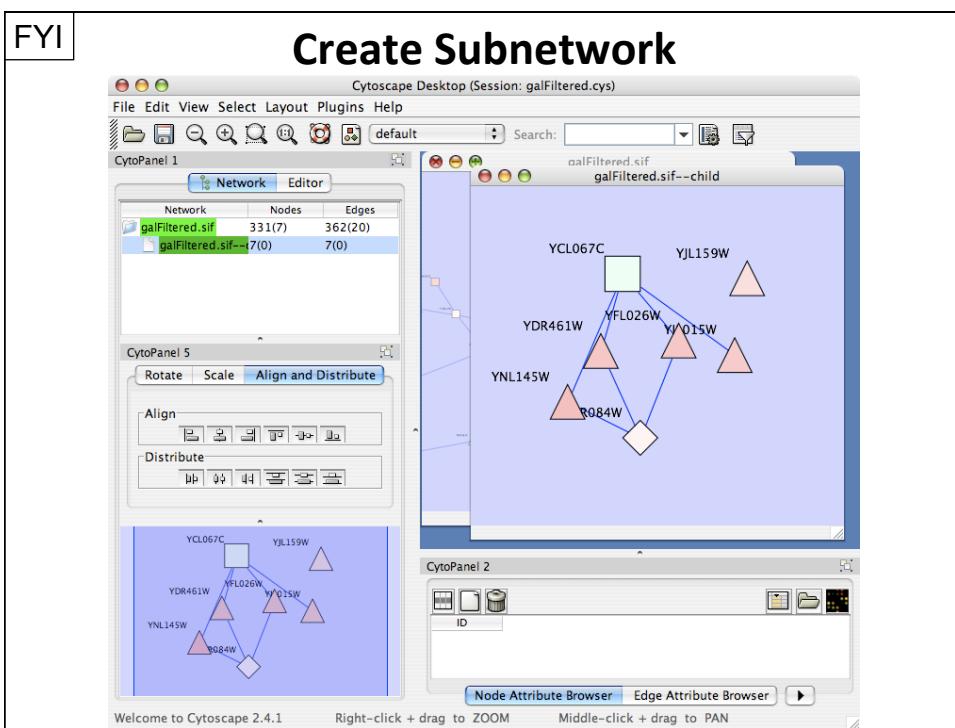
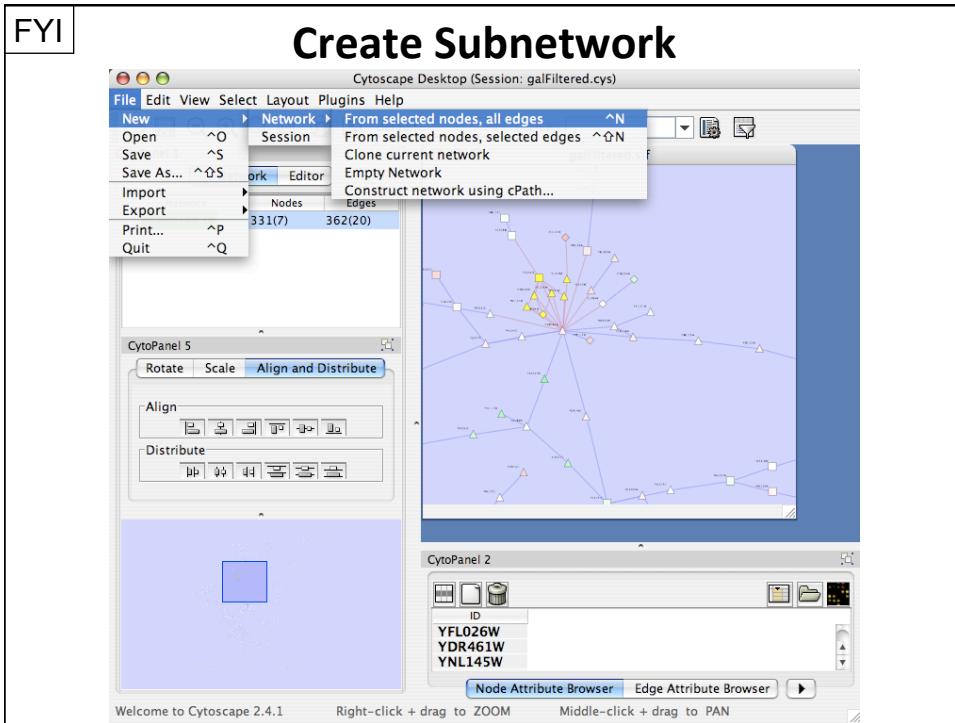
Rotate in Degrees:

Rotate Selected Nodes Only

Align

Distribute

Module 3: Network Visualization with Cytoscape bioinformatics.ca



FYI

Visual Style

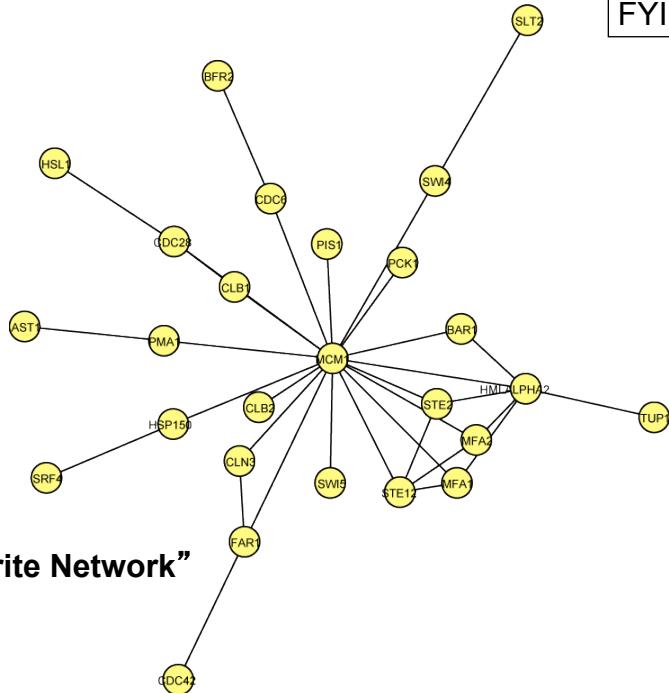
- Customized views of experimental data in a network context
- Network has node and edge attributes
 - E.g. expression data, GO function, interaction type
- Mapped to visual attributes
 - E.g. node/edge size, shape, colour...
- E.g. Visualize gene expression data as node colour gradient on the network

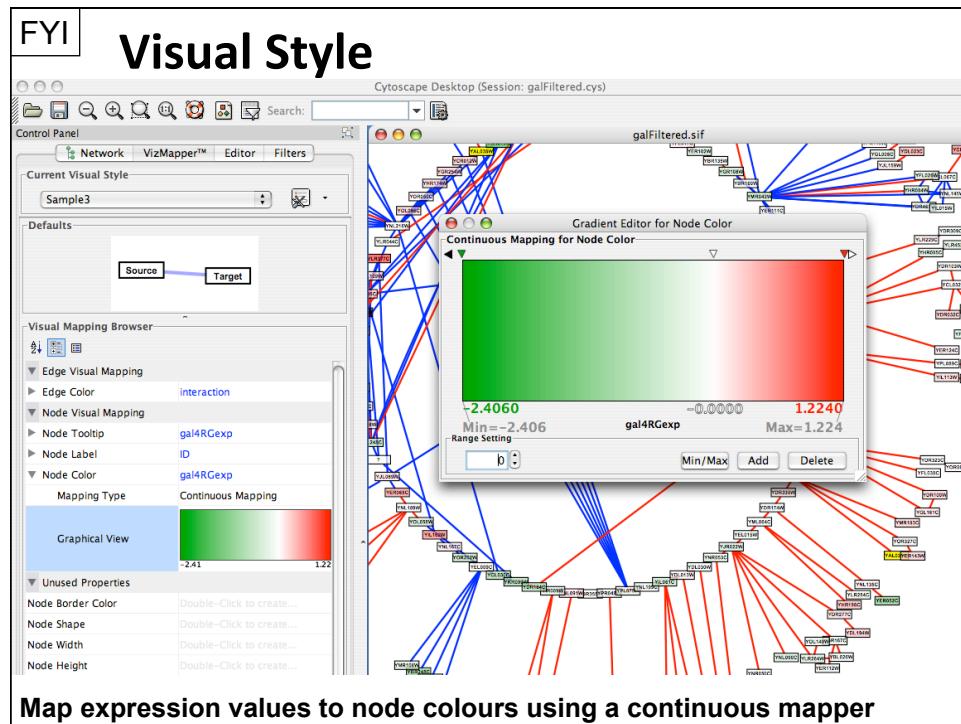
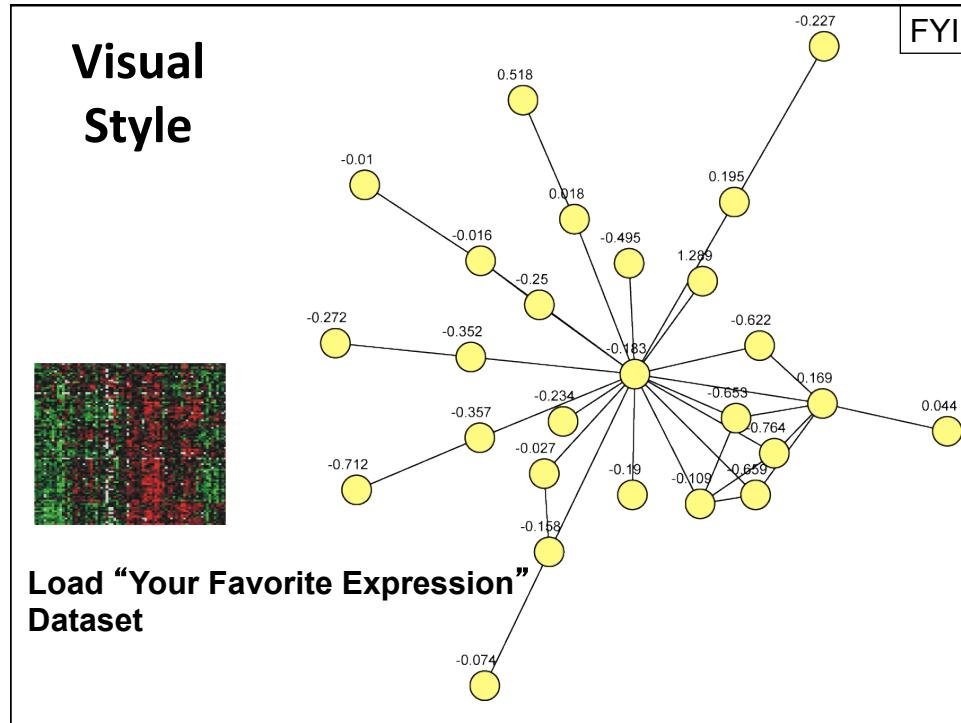
Module 3: Network Visualization with Cytoscape

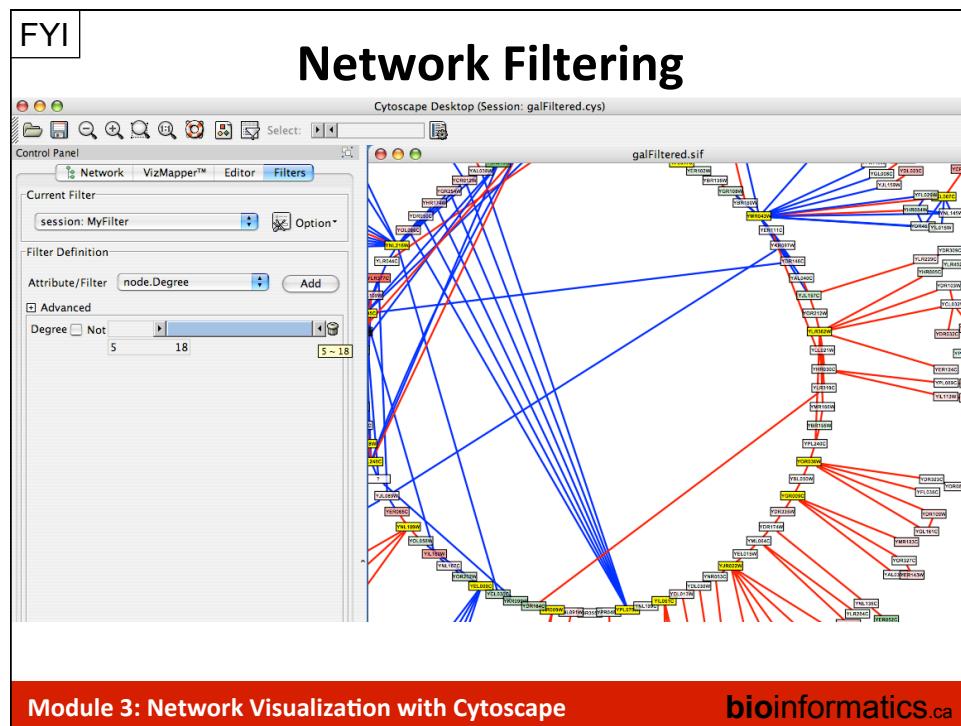
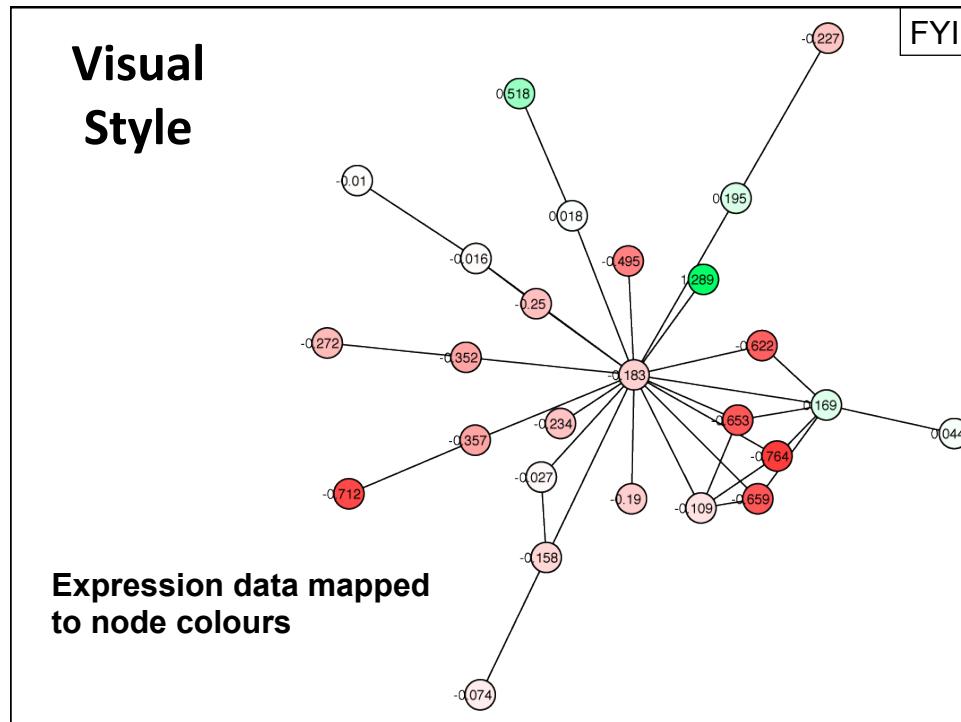
bioinformatics.ca

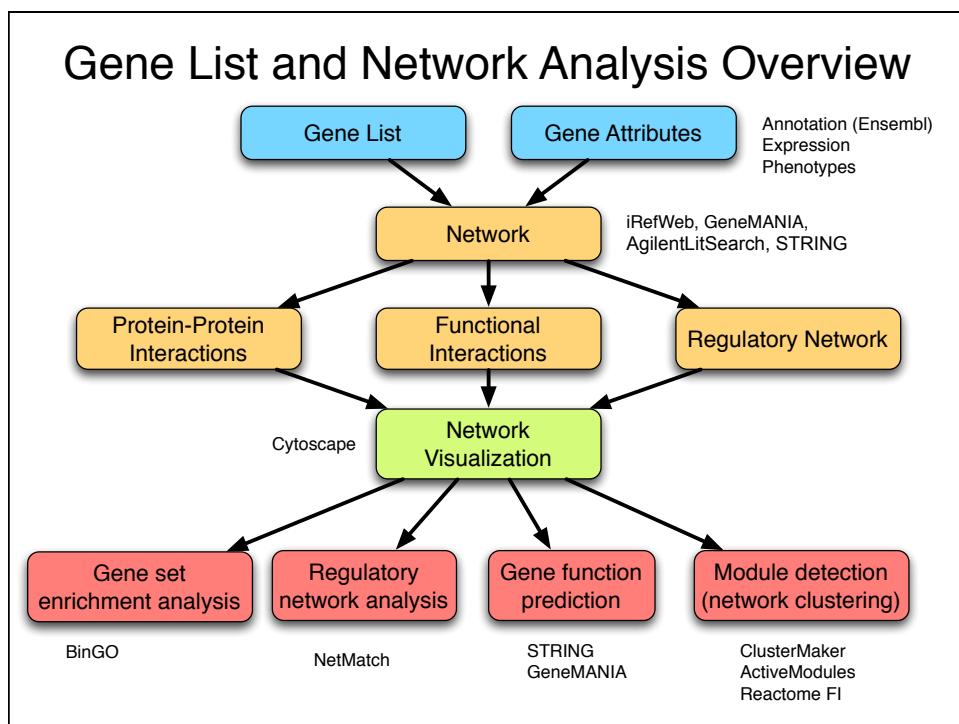
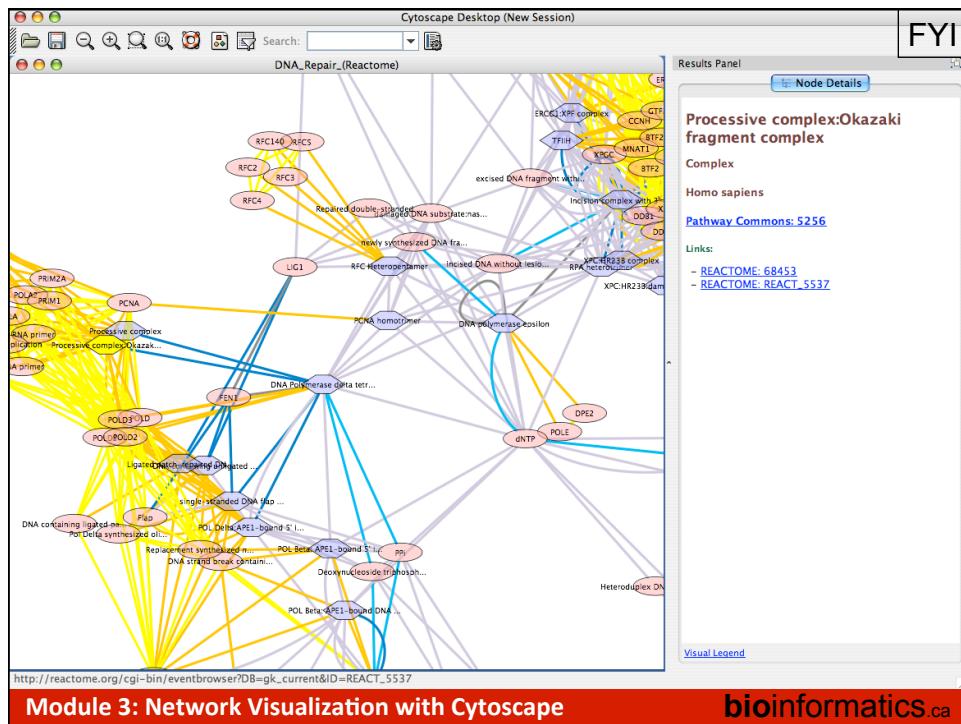
Visual Style

FYI







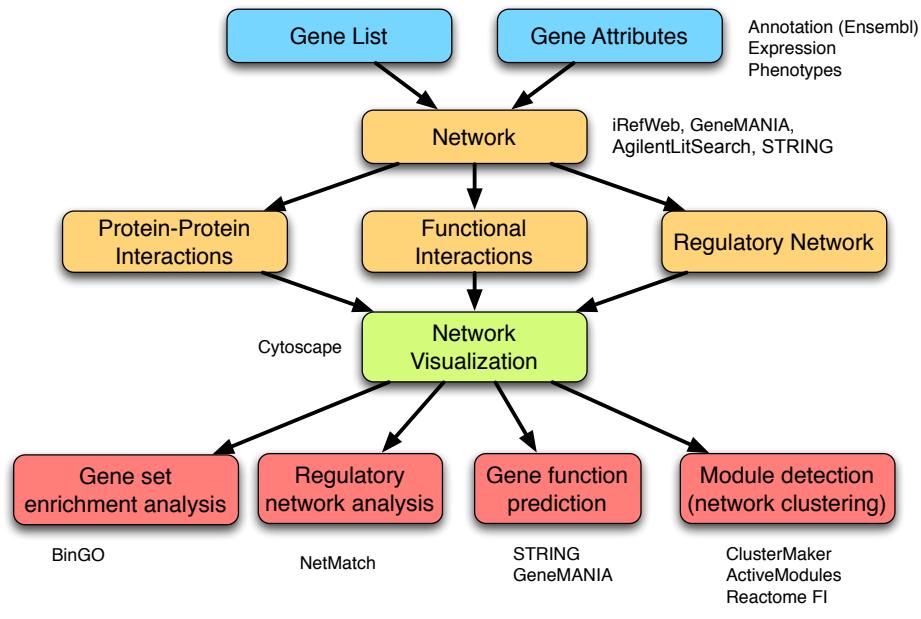


We are on a Coffee Break & Networking Session

Module 3: Network Visualization with Cytoscape

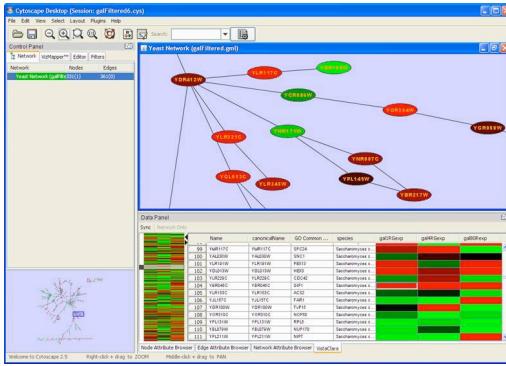
bioinformatics.ca

Gene List and Network Analysis Overview



VistaClara

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



The screenshot shows a Cytoscape window displaying a yeast network graph with nodes representing genes and edges representing interactions. Below the graph is a Data Panel containing a table with columns for Gene, Name, canonical_name, GO_Conserv..., species, gal1RGexp, gal4RGexp, and gal80Rexp. To the right of the Cytoscape window are four heatmaps labeled gal1RGexp, gal4RGexp, gal80Rexp, gal1RGexp, gal4RGexp, and gal80Rexp, showing expression levels across different conditions.

Module 3: Network Visualization with Cytoscape bioinformatics.ca

Cytoscape Lab

- Cytoscape – expression data visualization
 - Load the sample network file: galFiltered.sif
 - Lay it out – try different layouts
 - Load expression data - galExpData.pvals
 - Use File->Import->Attribute from Table
 - Examine node attributes
 - Visualize gene expression data using the Visual Mapper
 - Install the VistaClara plugin from the plugin manager
 - Play the expression data as a movie

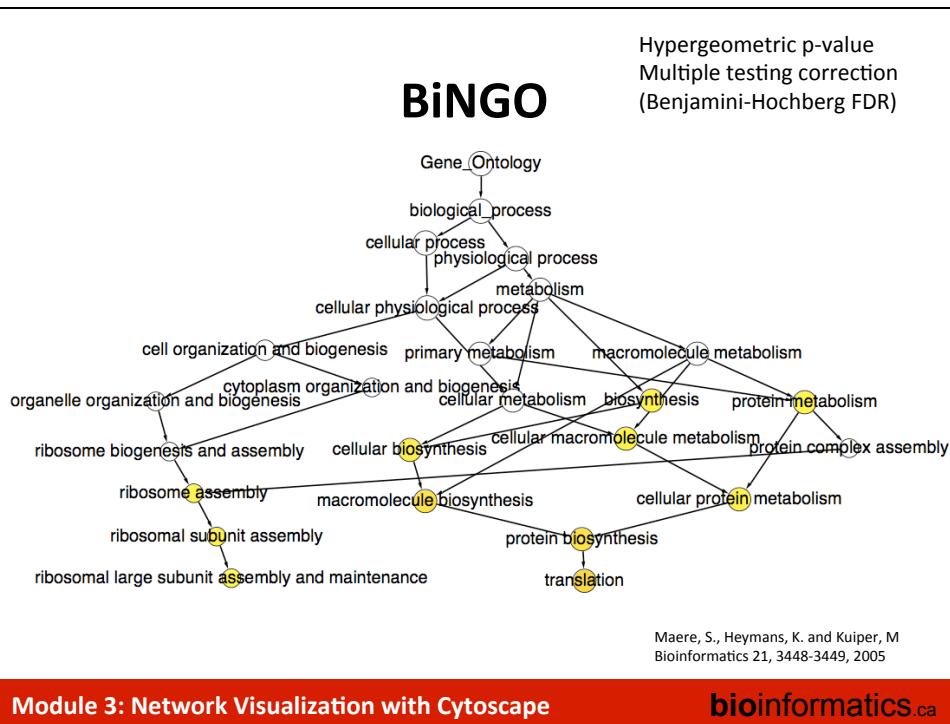
Module 3: Network Visualization with Cytoscape bioinformatics.ca

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

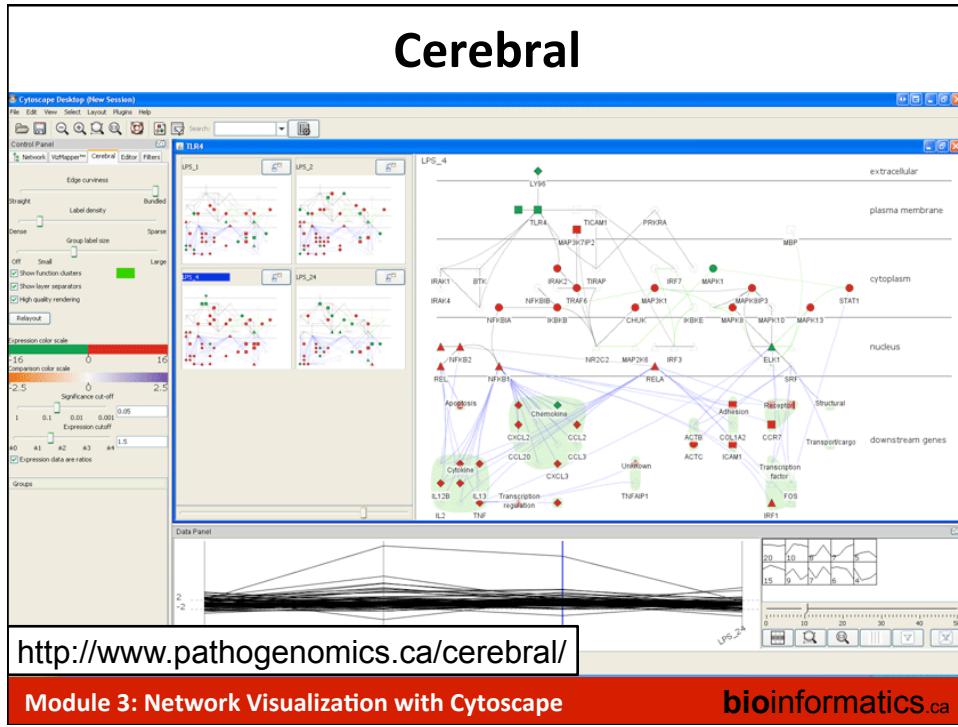
Module 3: Network Visualization with Cytoscape

bioinformatics.ca



Module 3: Network Visualization with Cytoscape

bioinformatics.ca



Find Active Subnetworks

- Active modules
 - Input: network + p-values for gene expression values e.g. from GCRMA
 - Output: significantly differentially expressed subgraphs
- Method
 - Calculate z-score/node, Z_A score/subgraph, correct vs. random expression data sampling
 - Score over multiple experimental conditions
 - Simulated annealing used to find high scoring networks

Ideker T, Ozier O, Schwikowski B, Siegel AF. Bioinformatics. 2002;18 Suppl 1:S233-40

Module 3: Network Visualization with Cytoscape bioinformatics.ca

Active Module Results

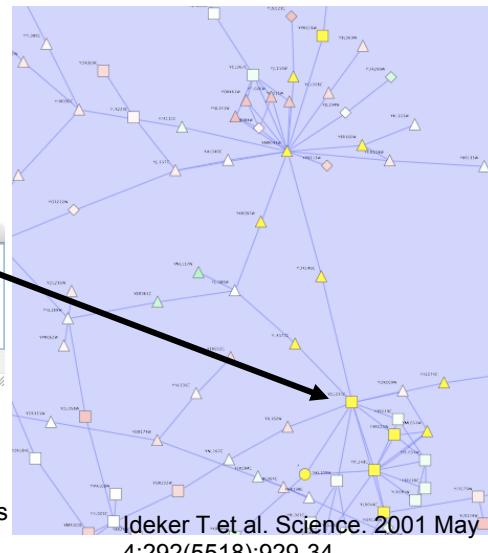
Network: yeast protein-protein and protein-DNA network

Expression data: 3 gene knock out conditions (enzyme, TF activator, TF repressor)

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

Note: non-deterministic, multiple runs required for confidence of result robustness



Module 3: Network Visualization with Cytoscape

bioinformatics.ca

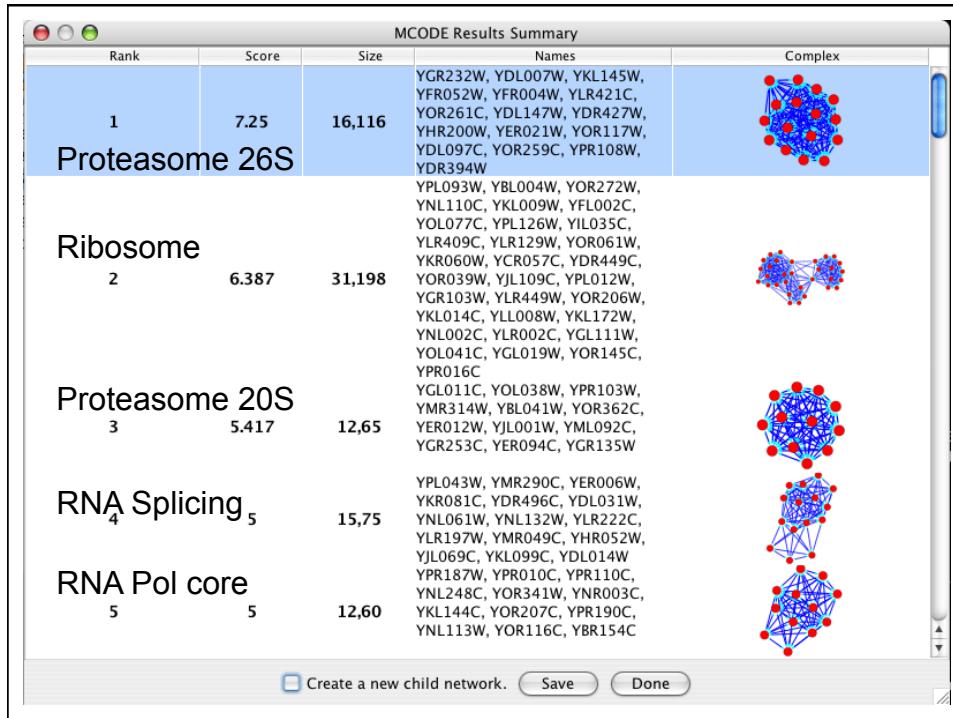
Network Clustering

- Clusters in a protein-protein interaction network have been shown to represent protein complexes and parts of pathways
- Clusters in a protein similarity network represent protein families
- Network clustering is available through the ClusterMaker Cytoscape plugin

Bader & Hogue, BMC Bioinformatics 2003 4(1):2

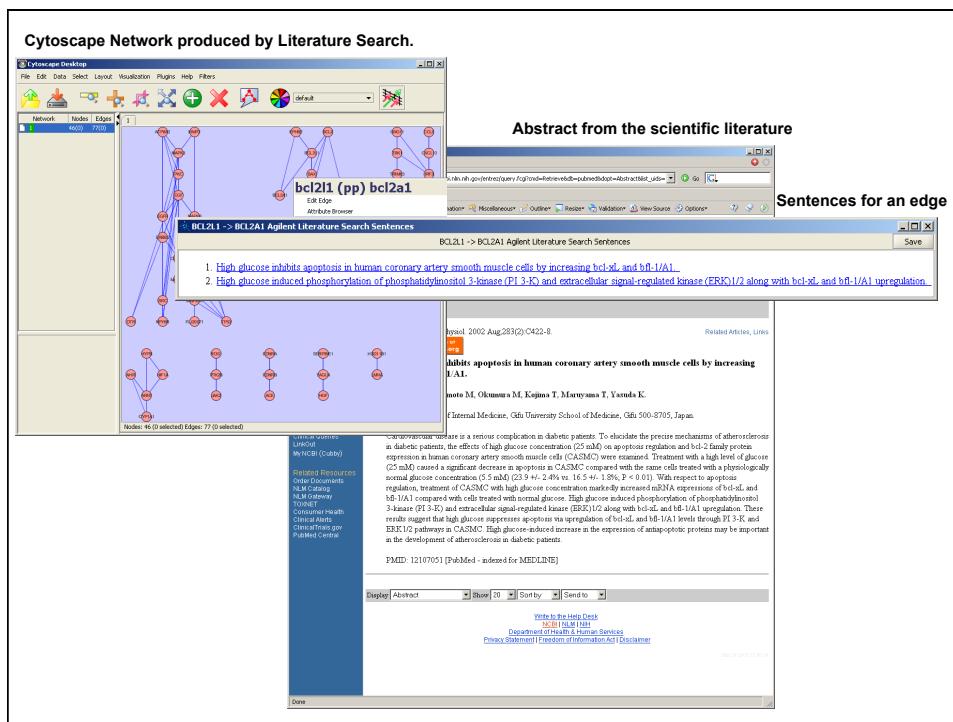
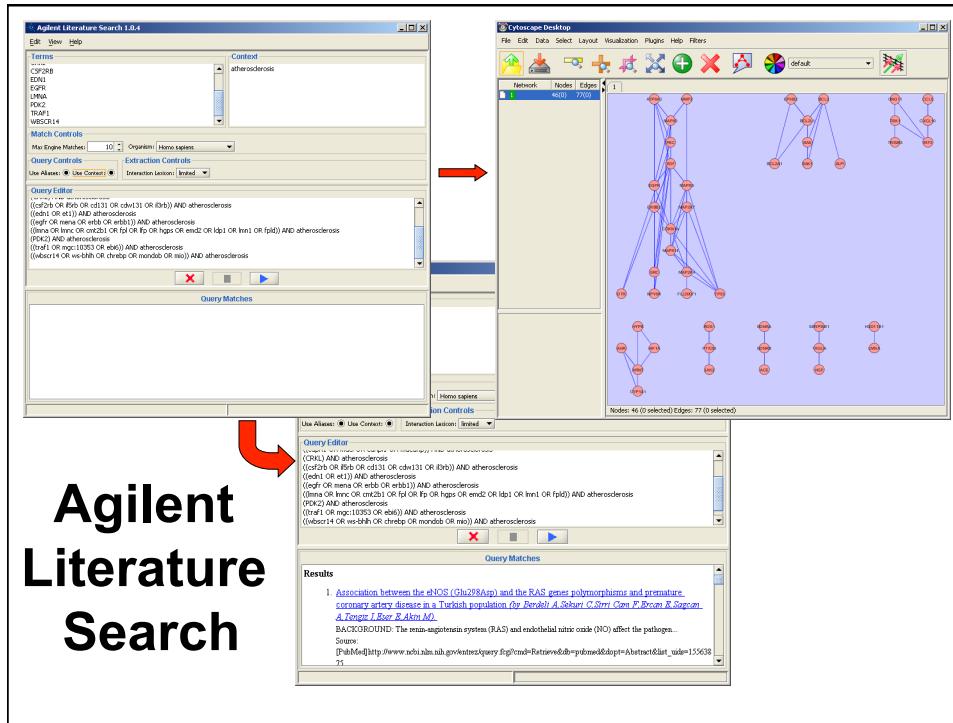
Module 3: Network Visualization with Cytoscape

bioinformatics.ca



Text Mining

- Computationally extract gene relationships from text, usually PubMed abstracts
- Useful if network is not in a database
 - Literature search tool
- BUT not perfect
 - Problems recognizing gene names
 - Natural language processing is difficult
- Agilent Literature Search Cytoscape plugin
- iHOP (www.ihop-net.org/UniPub/iHOP/)



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>

Module 3: Network Visualization with Cytoscape

bioinformatics.ca

Finding specific biological relevant TF-PPI sub-networks

The screenshot shows the NetMatch V1.0.1 software interface. On the left, the "NetMatch Query Editor - new query*" window displays a "Motifs" palette with a "Feed Forward Loop" motif selected. Below it is a "Query" panel containing a small graph diagram with three nodes and two edges. At the bottom of the editor are buttons for "Pass Query to NetMatch" and status information: "Nodes: 6 Edges: 6 Paths: 0 Loops: 0". On the right, the "NetMatch V1.0.1" window shows the "Results" panel. This panel includes a table of matches:

Match Number	Nodes	Image
1	YMR309C, YOR361C, YPR043W	
2	YOR310C, YDL014W, YLR197W	
3	YDR100W, YGL161C, YOR036W	
4	YIL015W, YMR043W, YCL067C	

Below the table, there is a list of search results:

```

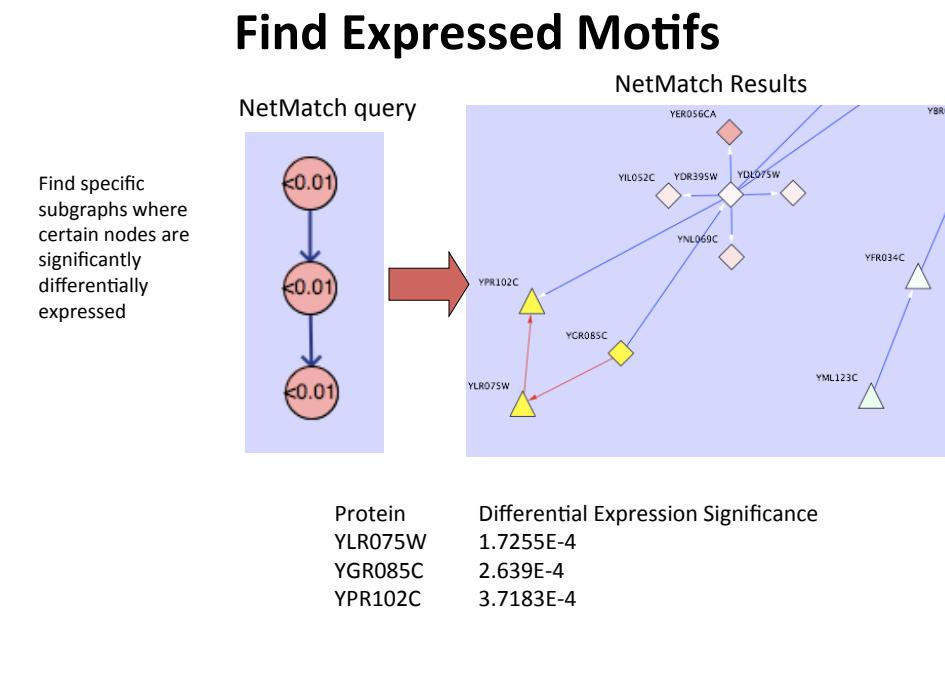
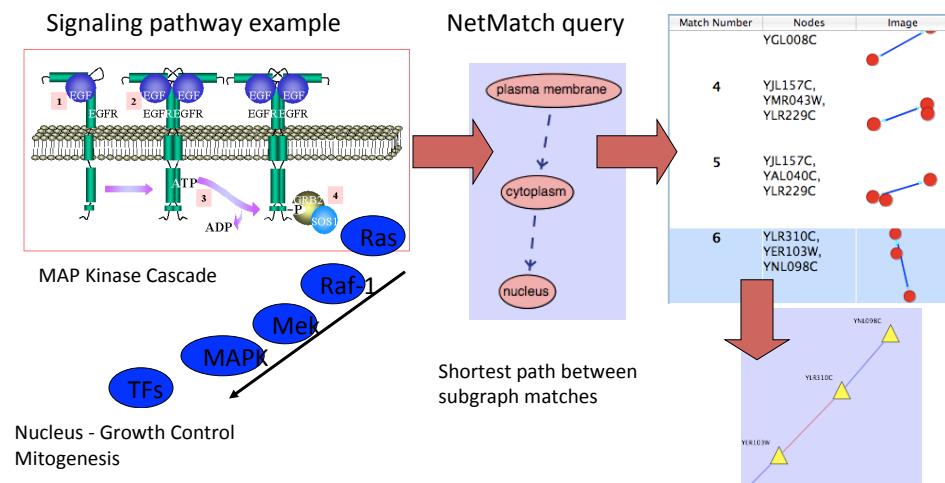
1 matches YBR020W
2 matches YGL035C
*****
0 matches YPL248C
1 matches YML051W
2 matches YML052W
*****
0 matches YML053W
1 matches YML054W
2 matches YML055W
*****

```

At the bottom right of the results panel, it says "Ferro et al. Bioinformatics 2007".

Find Signaling Pathways

- Potential signaling pathways from plasma membrane to nucleus via cytoplasm



Cytoscape 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

Cytoscape Tips & Tricks

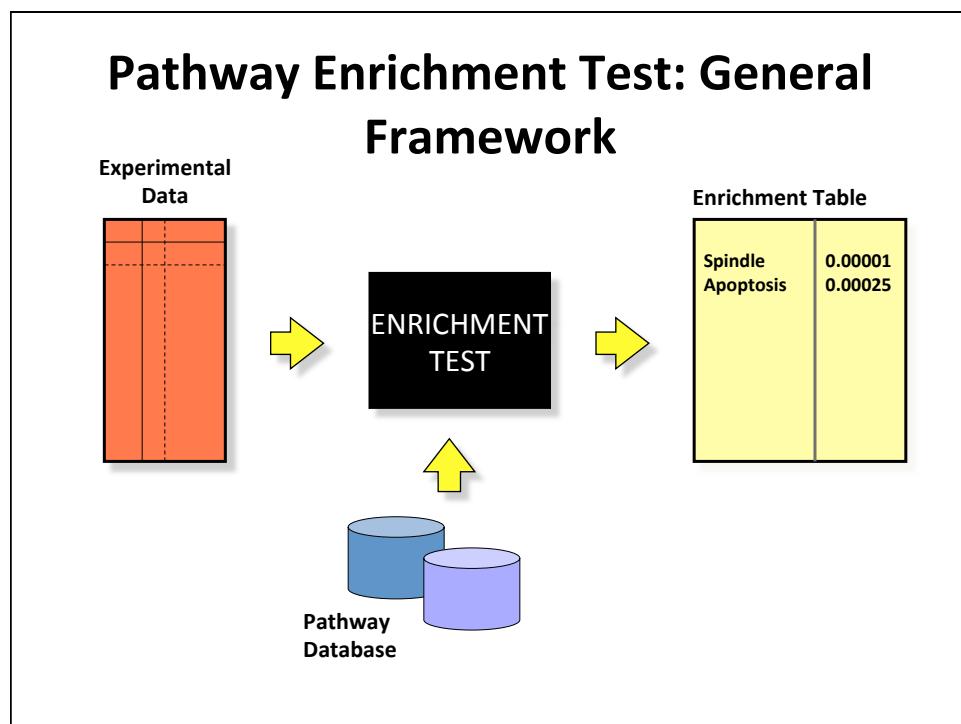
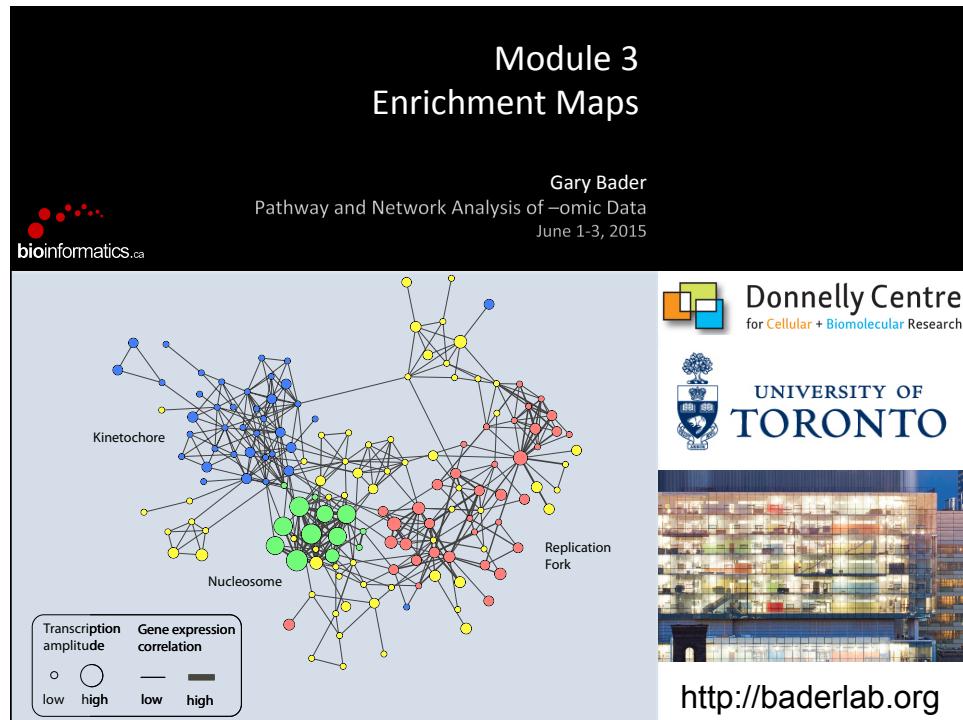
- Sessions
 - Sessions save pretty much everything:
 - Networks
 - Properties
 - Visual styles
 - Screen sizes
 - Saving a session on a large screen may require some resizing when opened on your laptop

Cytoscape 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
 - Cytoscape tries to “guess” good default memory settings based on how much memory you have, but it can fail

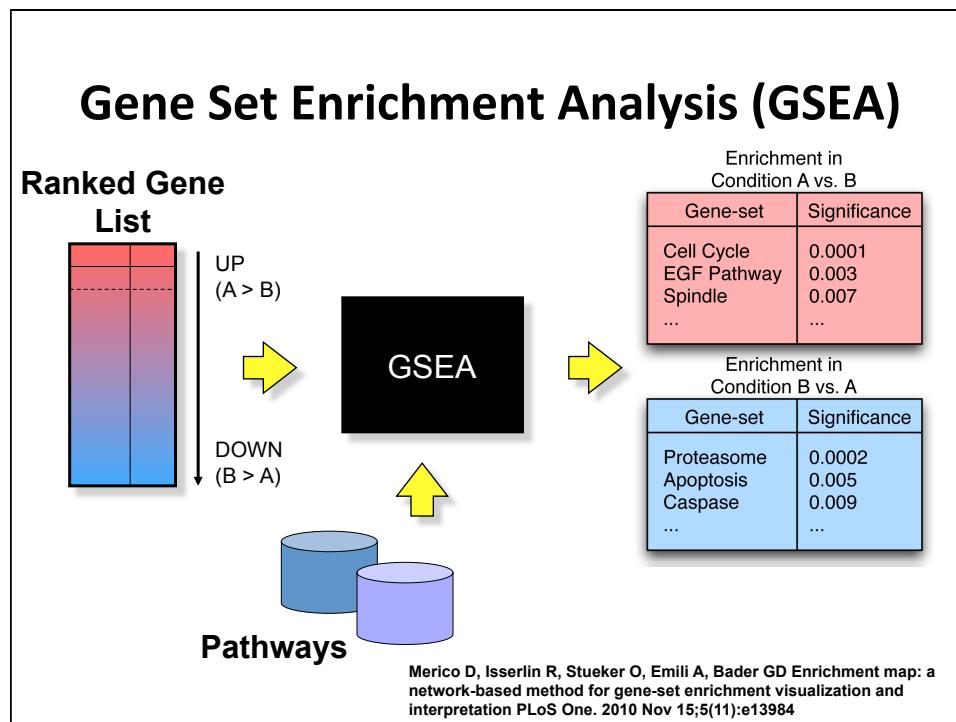
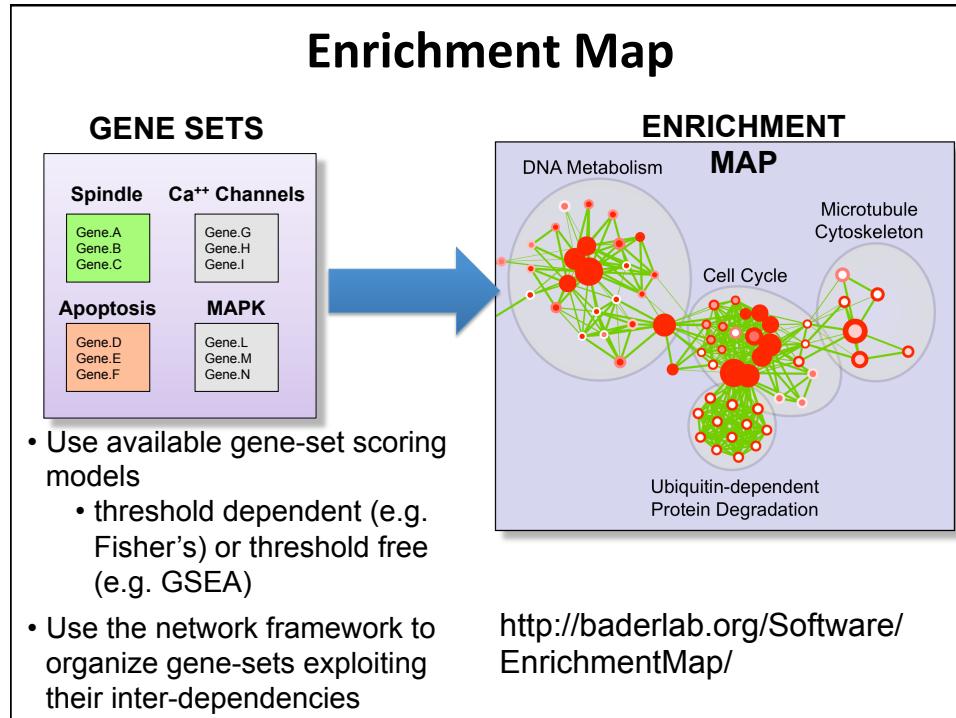
Cytoscape Tips & Tricks

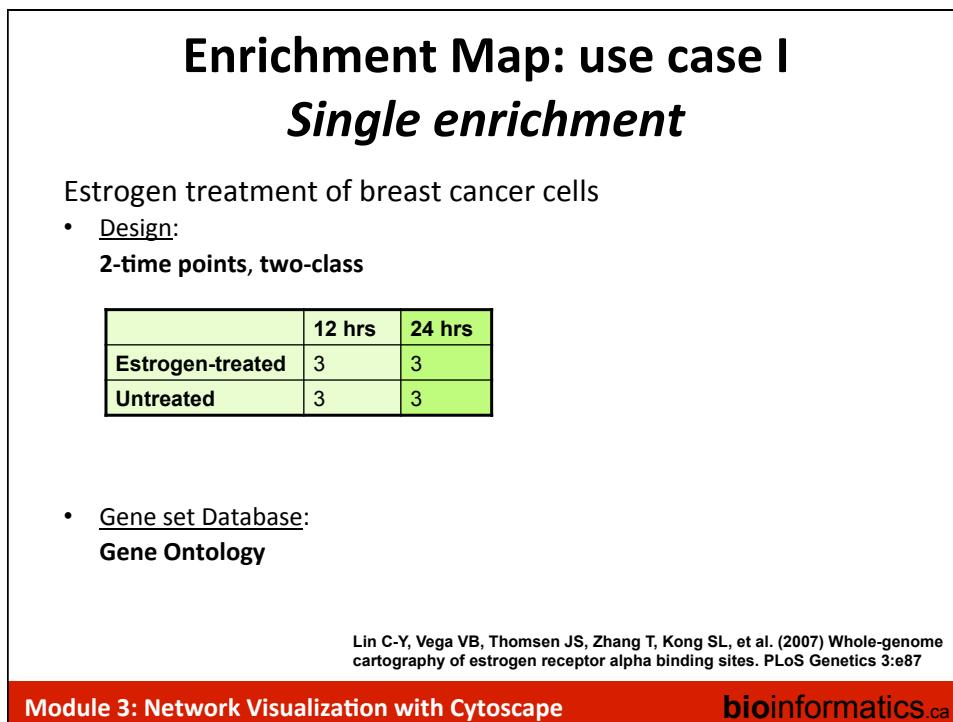
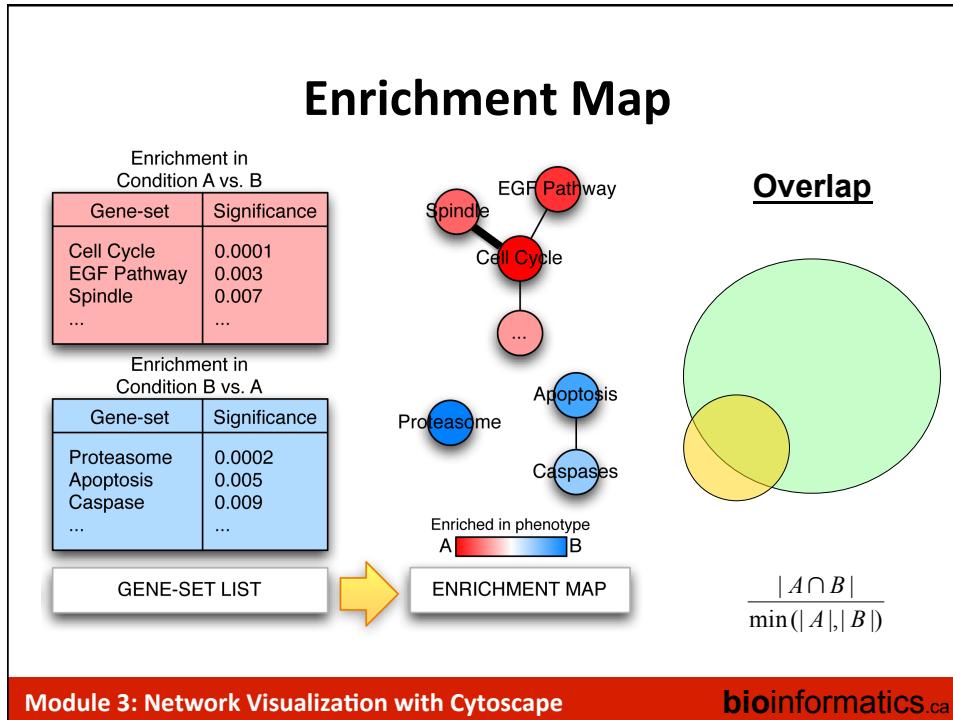
- CytoscapeConfiguration directory
 - In your user/home directory
 - Your defaults and any apps downloaded from the app store will go here
 - Sometimes, if things get really messed up, deleting (or renaming) this directory can give you a “clean slate”

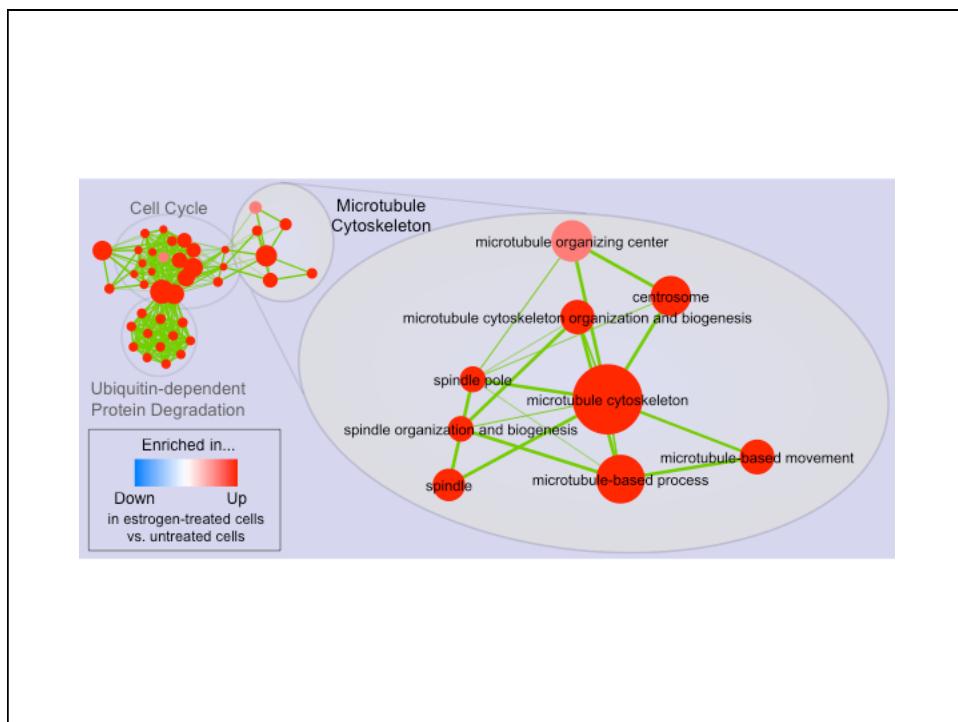
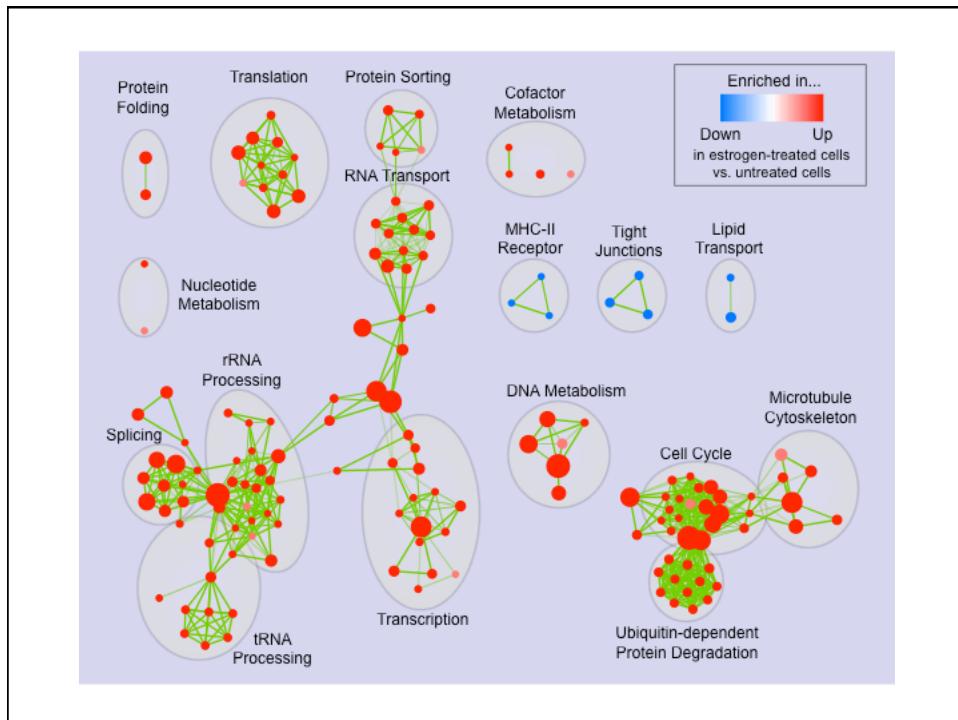


GO.id	GO.name	p.value	cover	cover.rat	Deg.mdn	Deg.iqr
GO:0042330	taxis	2.18E-06	23	0.056930693	54.94499375	9.139238998
GO:0006935	chemotaxis	2.18E-06	23	0.060209424	54.94499375	9.139238998
GO:0002460	adaptive immune response based on somatic recombination	7.10E-05	25	0.111111111	57.32306955	16.97054864
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	62.2271598	10.35284833
GO:0002522	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	Tymphocyte mediated immunity	0.00087216	22	0.101851852	57.84161096	16.13171132
GO:0019838	growth factor binding	0.000913285	15	0.068181818	83.04050808	10.58734852
GO:0051258	protein polymerization	0.00108876	17	0.080952381	57.97543252	17.3163968
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	18	0.027173913	52.75249873	23.23104637
GO:0030100	regulation of endocytosis	0.002658555	11	0.090909099	56.38041132	16.02486889
GO:0002526	acute inflammatory response	0.002660742	24	0.103004292	57.80087676	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	60.29679988	18.41103376
GO:002684	positive regulation of immune system process	0.006581245	27	0.111570248	60.29679988	22.05051447
GO:0050778	positive regulation of immune response	0.006581245	27	0.113924051	60.29679988	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:0050776	negative regulation of protein kinase activity	0.008017022	9	0.040723982	53.28031076	17.48904224
GO:0043086	phagocytosis	0.008106069	10	0.080645161	55.66270253	12.47536747
GO:0006909	myeloid leukocyte differentiation	0.008174948	10	0.092592593	62.86577216	9.401887596
GO:0002573	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	56.33116708	6.945510559
GO:0006469	negative regulation of protein kinase activity	0.010782155	7	0.046357616	52.28635156	12.58524145
GO:0051348	negative regulation of transferase activity	0.010782155	7	0.04516129	52.28635156	12.58524145
GO:0007179	transforming growth factor beta receptor signaling pathway	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

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	62.2271598	10.35284833
GO:0002522	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	Tymphocyte mediated immunity	0.00087216	22	0.101851852	57.84161096	16.13171132
GO:0019838	growth factor binding	0.000913285	15	0.068181818	83.04050808	10.58734852
GO:0051258	protein polymerization	0.00108876	17	0.080952381	57.97543252	17.3163968
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	18	0.027173913	52.75249873	23.23104637
GO:0030100	regulation of endocytosis	0.002658555	11	0.090909099	56.38041132	16.02486889
GO:0002526	acute inflammatory response	0.002660742	24	0.103004292	57.80087676	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	60.29679988	18.41103376
GO:002684	positive regulation of immune system process	0.006581245	27	0.111570248	60.29679988	22.05051447
GO:0050778	positive regulation of immune response	0.006581245	27	0.113924051	60.29679988	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:0066909	phagocytosis	0.00808106069	10	0.080645161	55.66270253	12.47536747
GO:002573	myeloid leukocyte differentiation	0.008174948	10	0.029529593	62.86577216	9.401887596
GO:0006959	humoral immune response	0.008396095	18	0.044568245	55.05654091	18.94209565
GO:0046649	Tymphocyte activation	0.009044401	29	0.059917355	61.92213317	21.03553355
GO:0030595	Teukocyte chemotaxis	0.0090707319	7	0.101449275	56.33116708	6.945510559
GO:0066909	negative regulation of protein kinase activity	0.010782155	7	0.046357616	52.28635156	12.58524145
GO:0051348	negative regulation of transferase activity	0.010782155	7	0.04516129	52.28635156	12.58524145
GO:0007179	transforming growth factor beta receptor signaling pathway	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: use case II

Comparison of two enrichments

Estrogen treatment of breast cancer cells

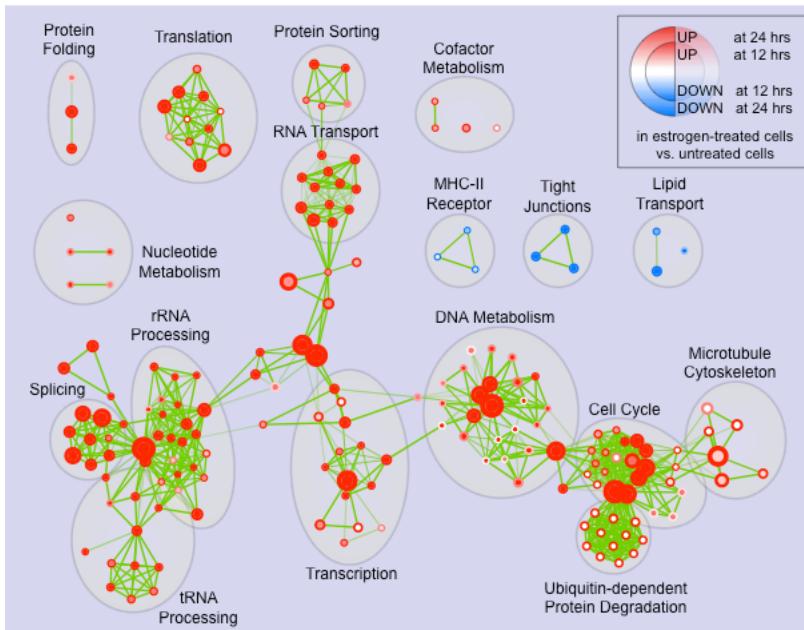
- Design:
2-time points, two-class

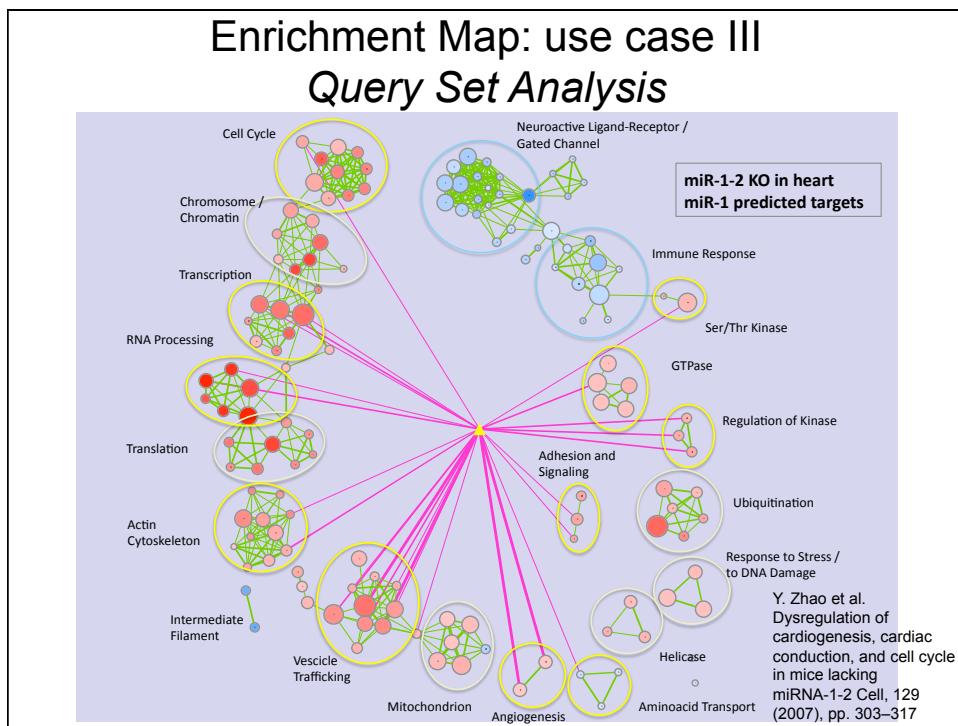
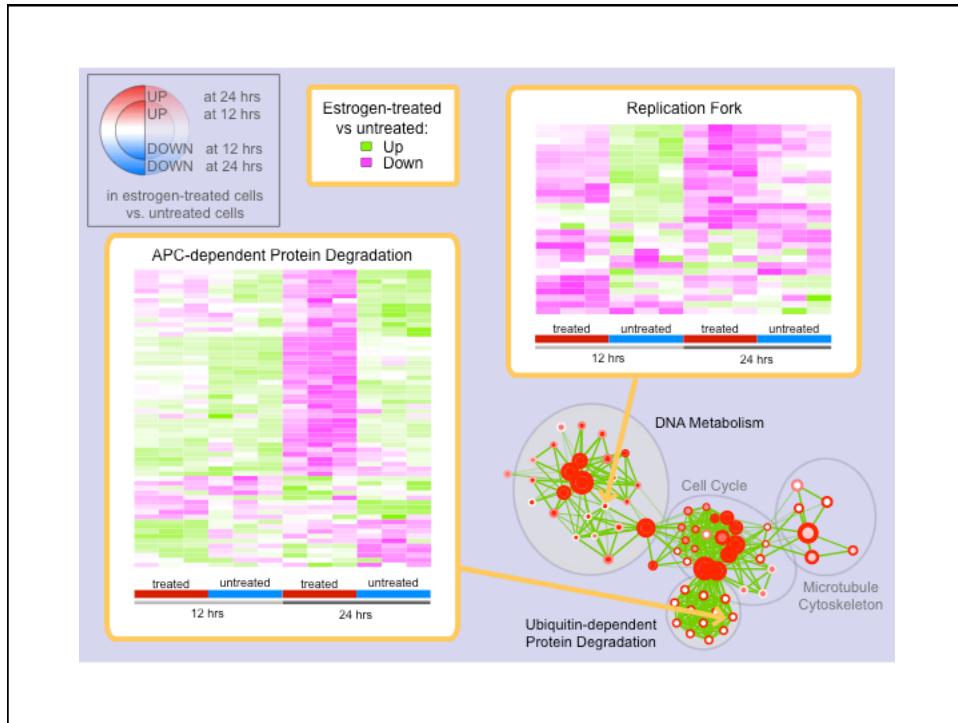
	12 hrs	24 hrs
Estrogen-treated	3	3
Untreated	3	3

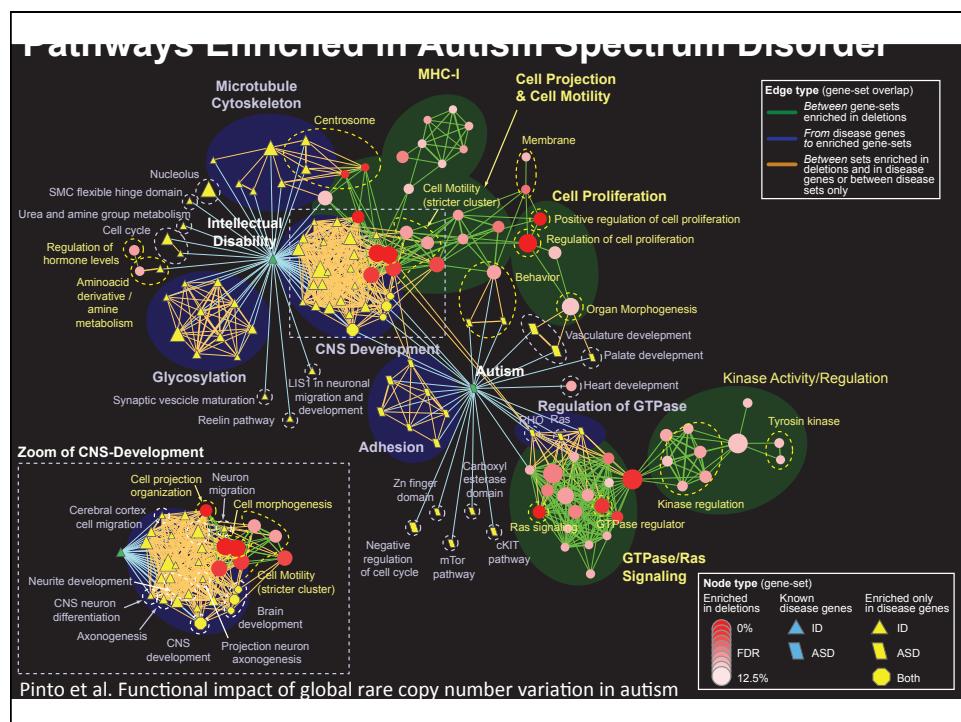
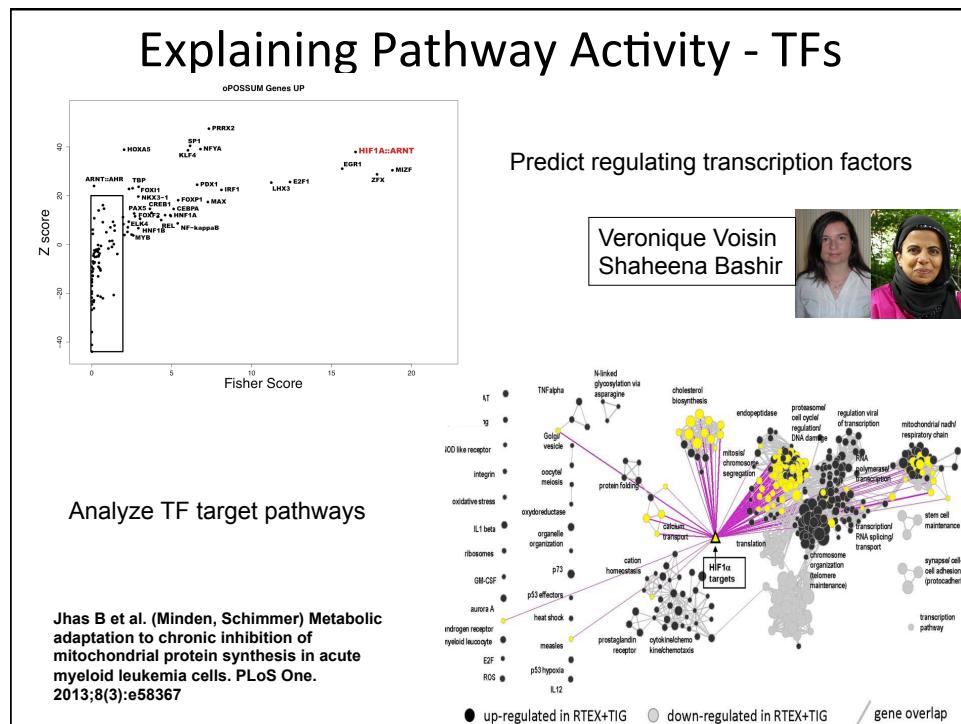
- Gene set Database:
Gene Ontology

Module 3: Network Visualization with Cytoscape

bioinformatics.ca



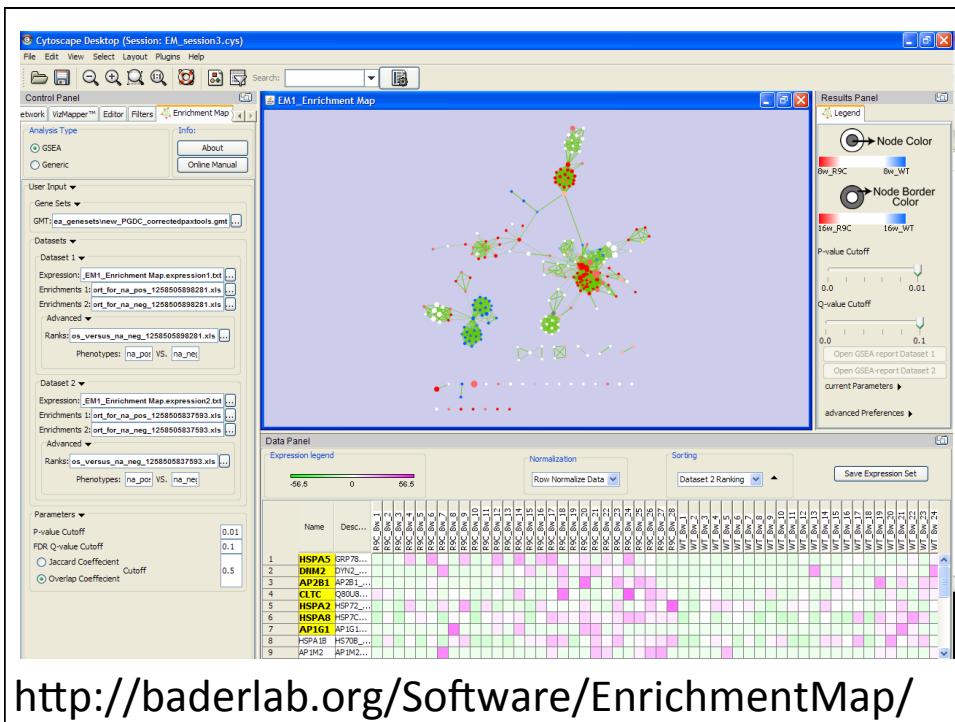


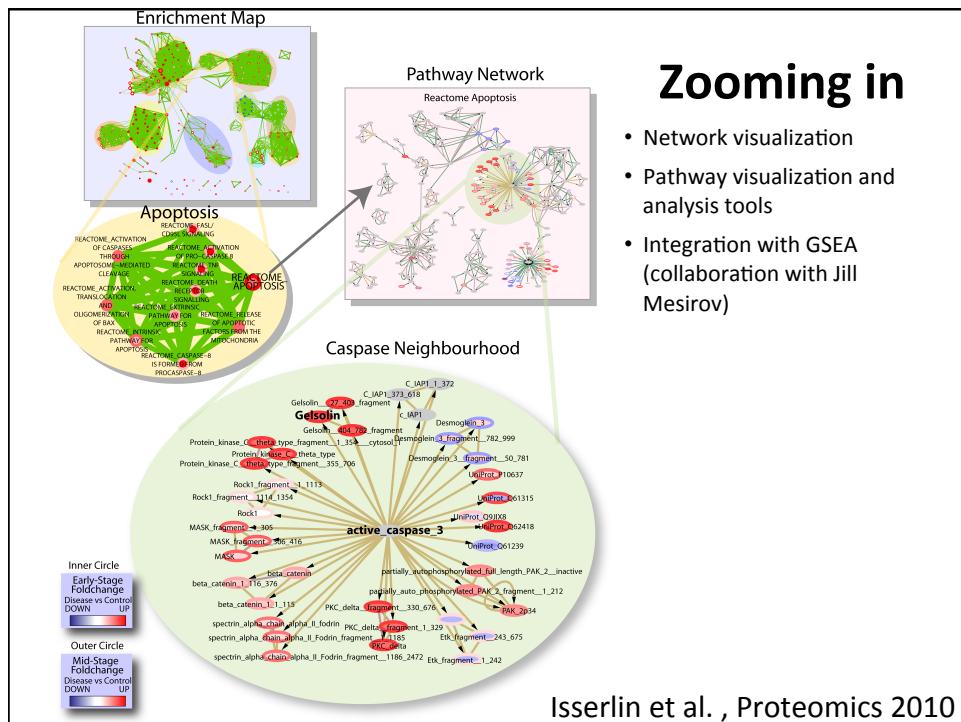


Pathway and gene set sources

- 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

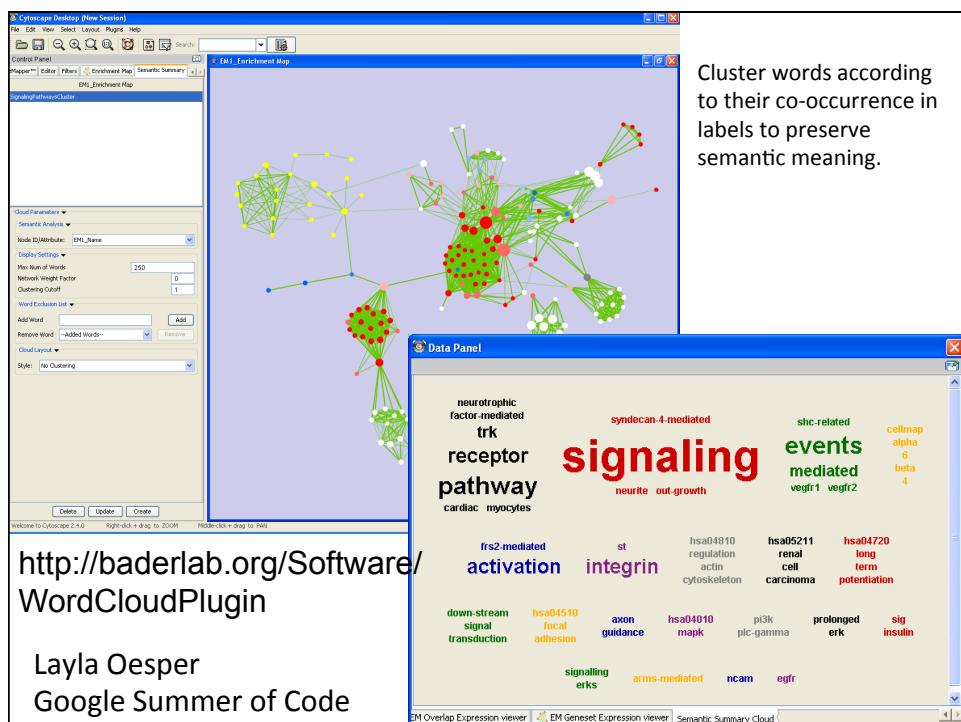
Pinto et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 2010 Jun 9.

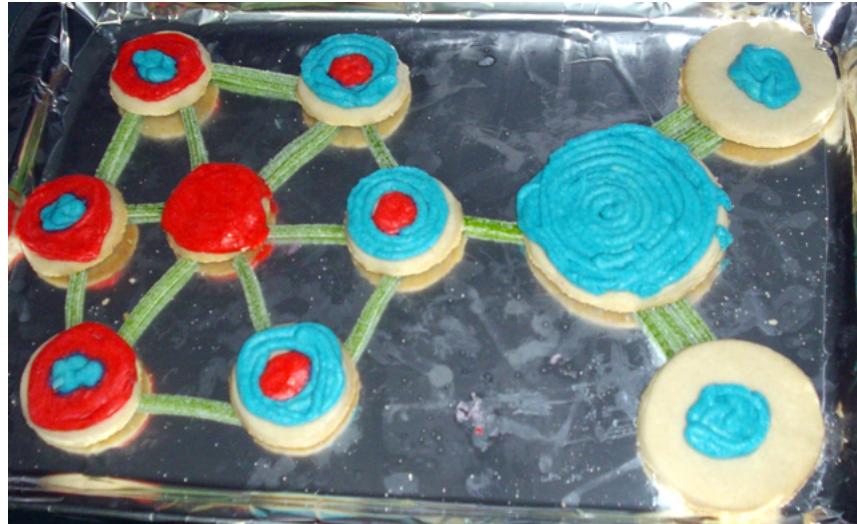




Zooming in

- Network visualization
- Pathway visualization and analysis tools
- Integration with GSEA (collaboration with Jill Mesirov)





Enrichment Map Lab

- Try out enrichment map – load the app from the app store
- Load your g:Profiler results – or - load the GSEA enrichment analysis file -
EM_EstrogenMCF7_TestData.zip (unzip) available at
 - <http://baderlab.org/Software/EnrichmentMap>
- Optional: integrated assignment data (Liver) enrichment map
- GSEA tutorial: <http://www.baderlab.org/Software/EnrichmentMap/Tutorial>