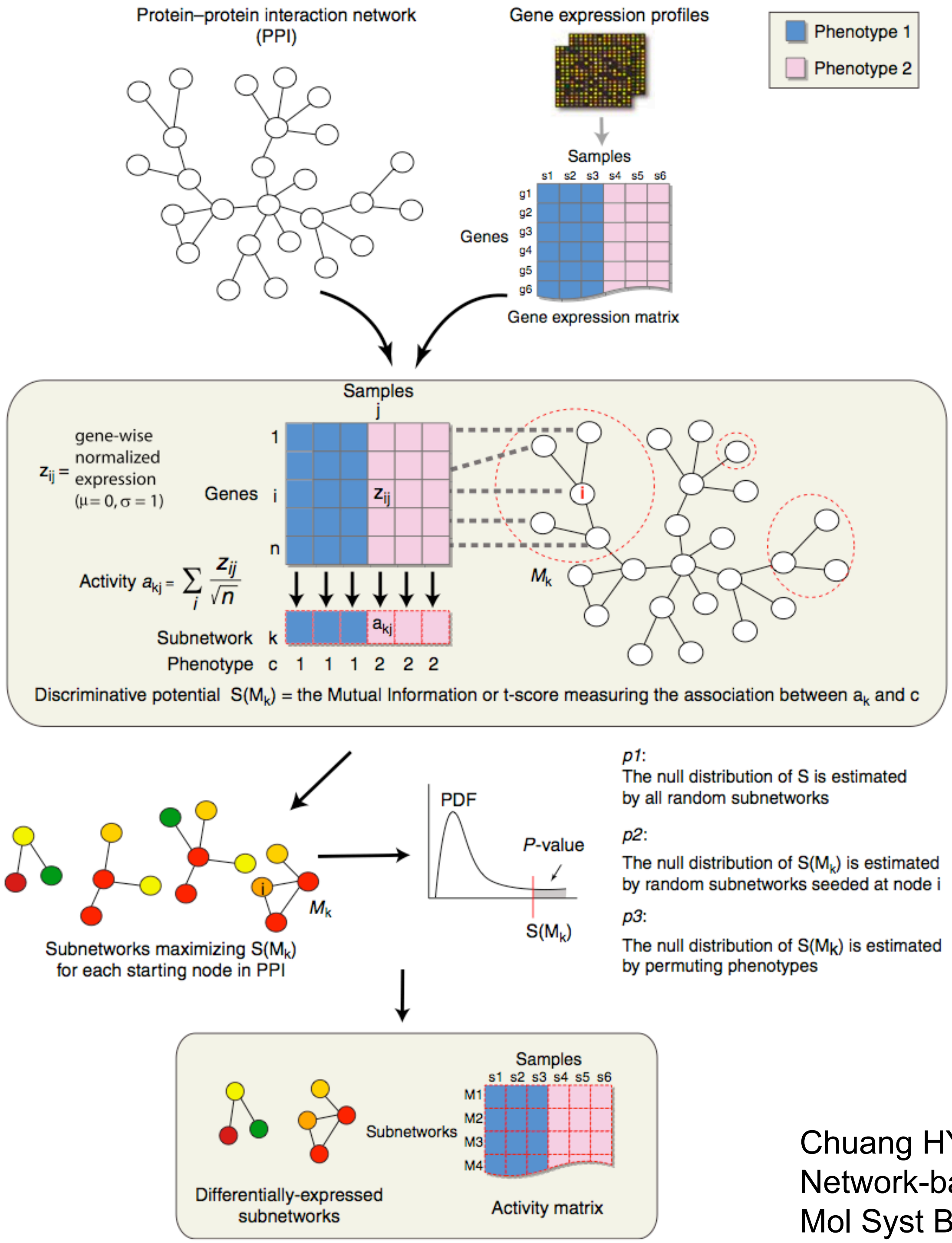
Network classification of disease

- Traditional: Gene association
- Limitations: Too many genes reduces statistical power
- New: Active cell map based approaches combining network and molecular profiles
- http://www.connotea.org/user/ACMT

Chuang HY, Lee E, Liu YT, Lee D, Ideker T
Network-based classification of breast cancer metastasis
Mol Syst Biol. 2007;3:140. Epub 2007 Oct 16

Liu M, Liberzon A, Kong SW, Lai WR, Park PJ, Kohane IS, Kasif S
Network-based analysis of affected biological processes in type 2 diabetes models
PLoS Genet. 2007 Jun;3(6):e96

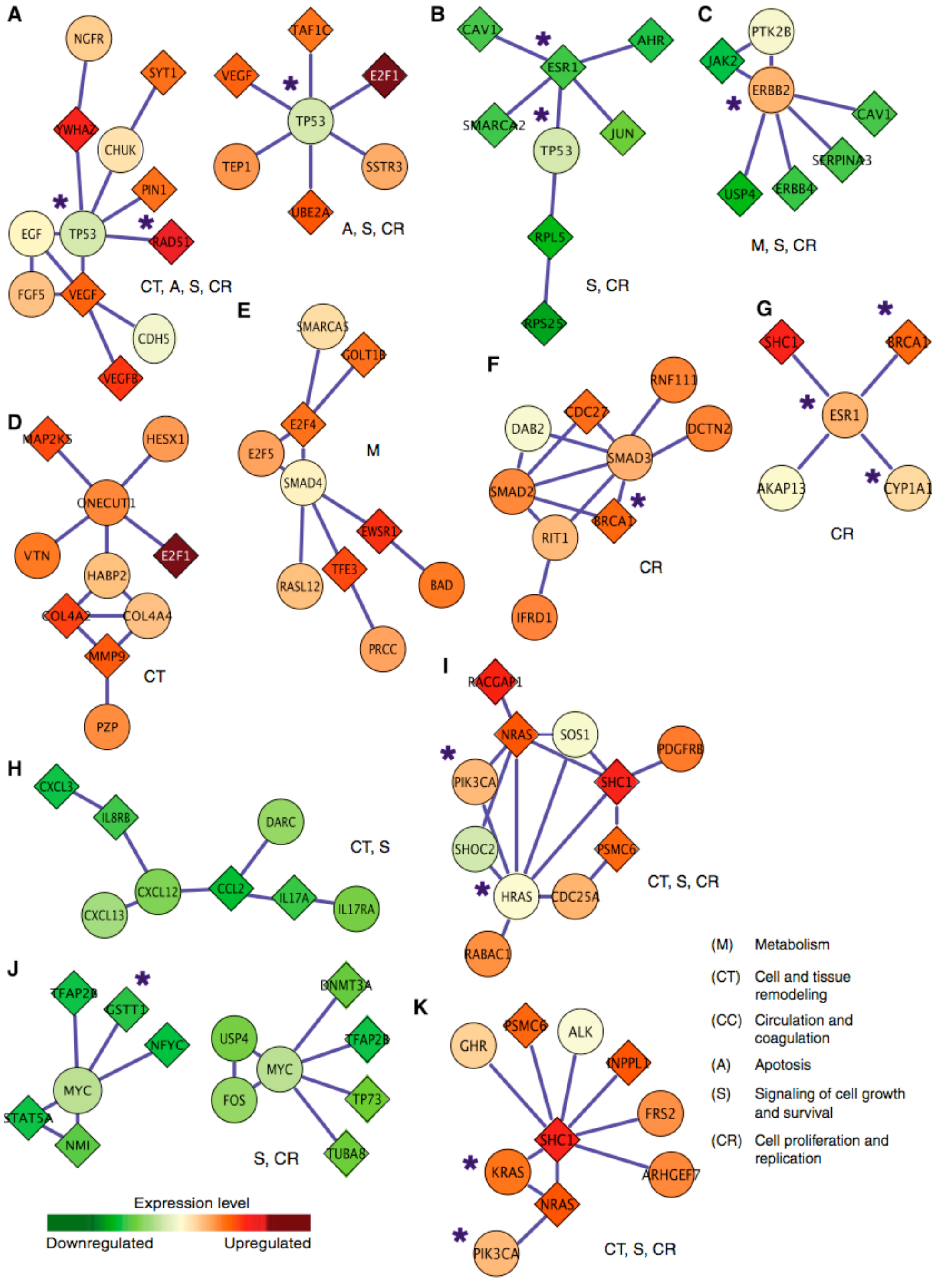
Efroni S, Schaefer CF, Buetow KH
Identification of key processes underlying cancer phenotypes using biologic pathway analysis
PLoS ONE. 2007 May 9;2(5):e425



Network-based breast cancer classification

- 57k intx from Y2H, orthology, co-citation, HPRD, BIND, Reactome
- 2 breast cancer cohorts, different expression platforms
- Metastasis: 78 vs. 217 (van de Vijver) & 106 vs. 180 (Wang)

Chuang HY, Lee E, Liu YT, Lee D, Ideker T
Network-based classification of breast cancer metastasis
Mol Syst Biol. 2007;3:140. Epub 2007 Oct 16



- Similar network markers across 2 data sets (better than original overlap)
- Increased classification accuracy
- Better coverage of known cancer risk genes (*)

Integration of Heterogeneous Data

Gary Bader
Terrence Donnelly Center for Cellular and Biomolecular Research (CCBR)
BBDMR, SLRI
University of Toronto

MoGen Topic Course - Mar.30.2009

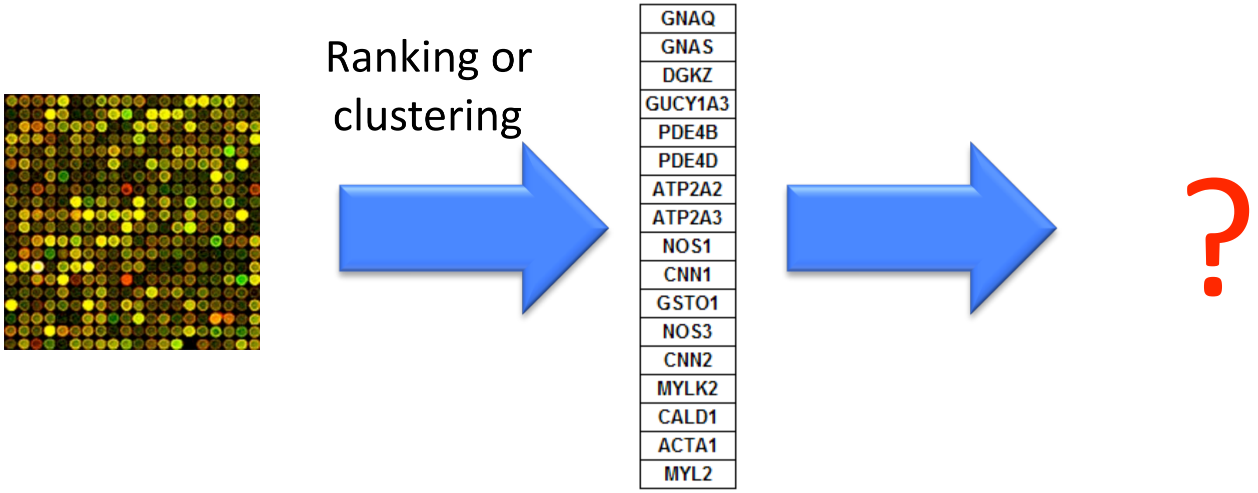
http://baderlab.org

Outline

- Gene lists from Omics data
- Networks as scaffold for data integration
- Statistical integration
  - Gene function prediction
- Database and semantic integration
  - Biological pathways
- Visualizing integrated data with Cytoscape

Interpreting Gene Lists

- My cool new screen worked and produced 1000 hits! …Now what?
- Genome-Scale Analysis (Omics)
  - Genomics, Proteomics



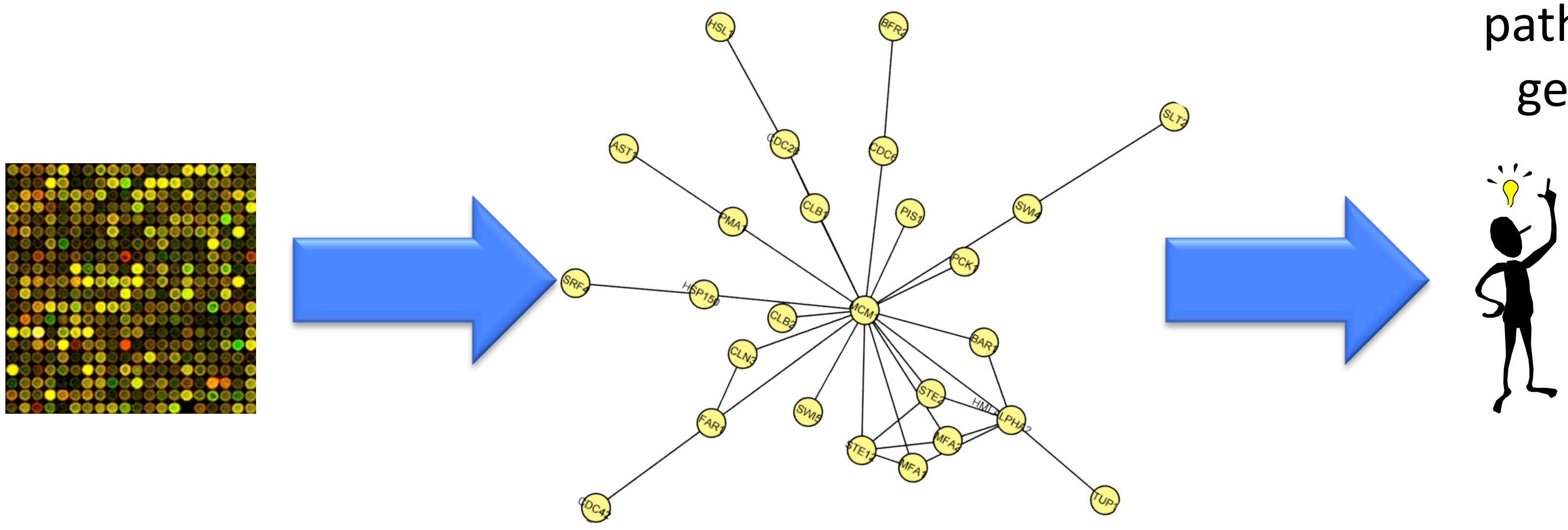
Ranking or clustering

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

?

Interpreting Gene Lists using Networks

- My cool new screen worked and produced 1000 hits! …Now what?
- Genome-Scale Analysis (Omics)
  - Genomics, Proteomics



Eureka! New pathway gene!

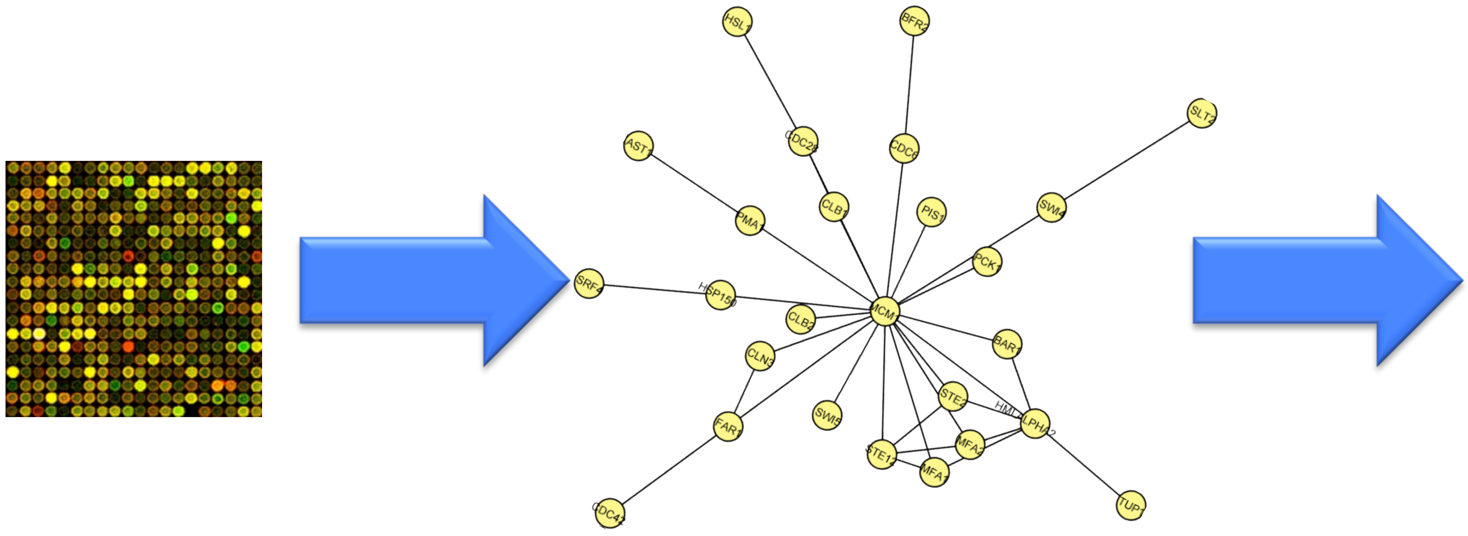
Why Network Analysis?

Intuitive to Biologists

- Provide a biological context for results
- More efficient than searching databases gene-by-gene
- Intuitive display for sharing data

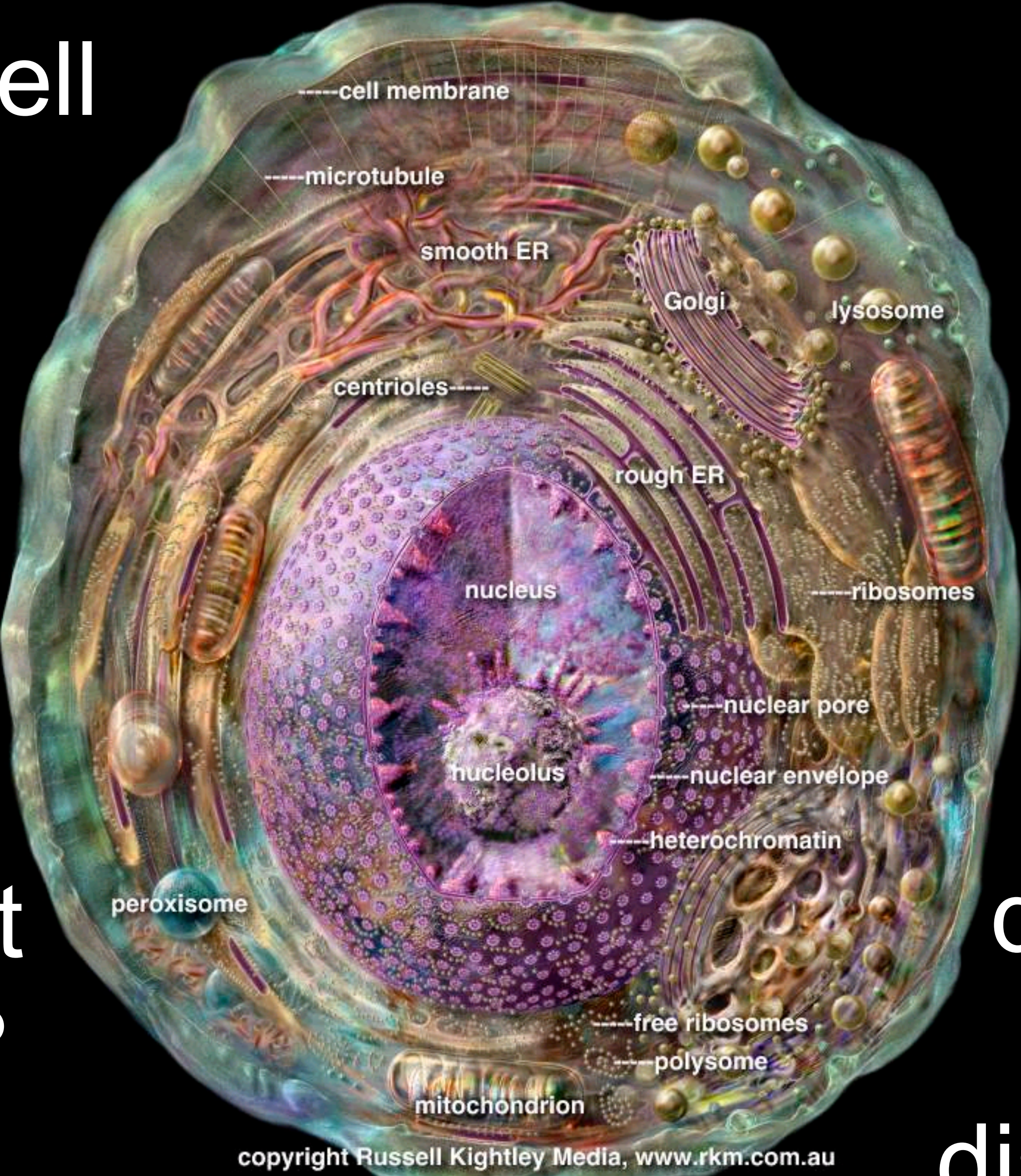
Computationally Query to Answer Specific Questions

- Visualize multiple data types on a network
- Cluster, Find active pathways, Compare, Search



Eureka! New
pathway
gene!

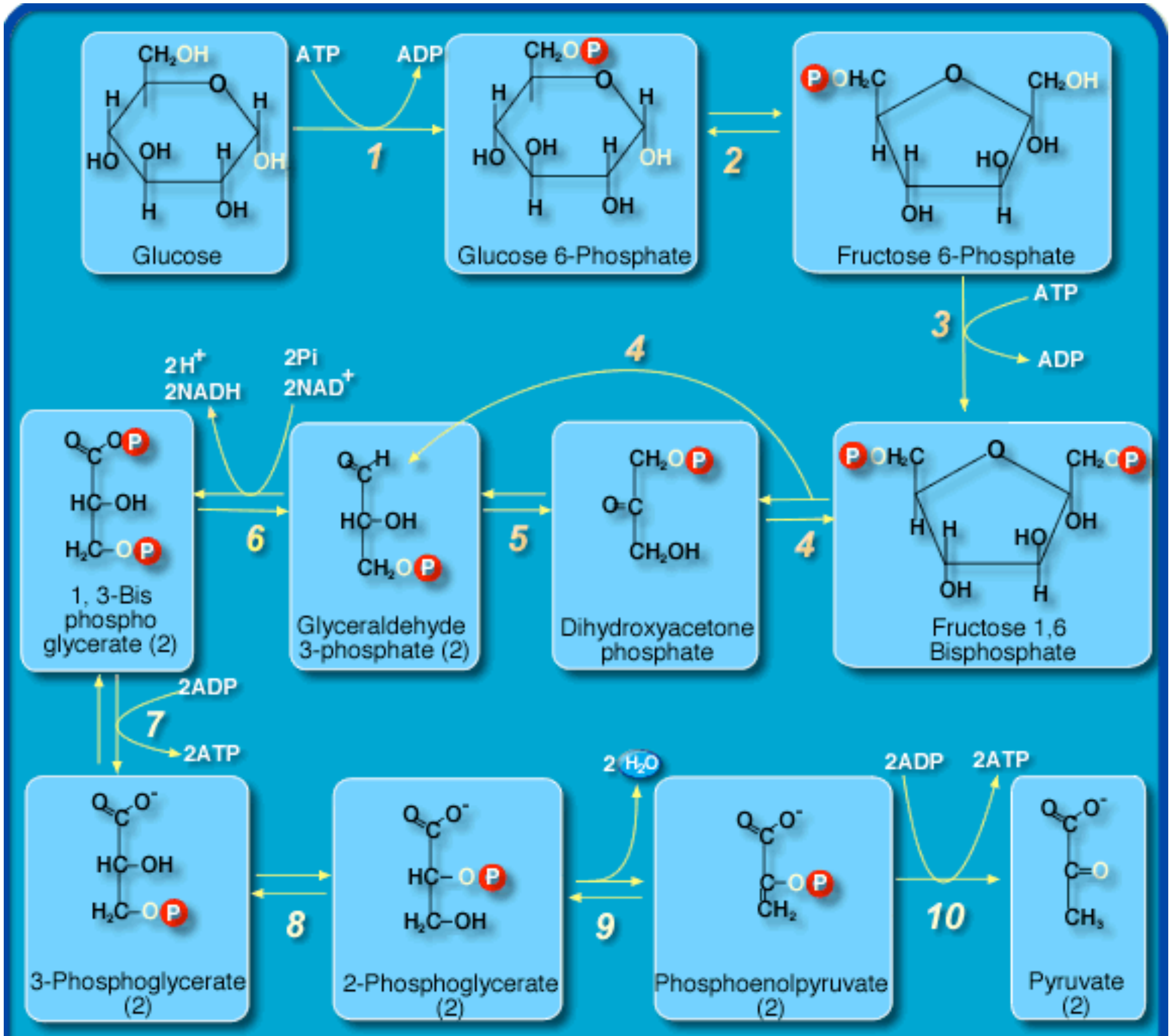
The Cell



copyright Russell Kightley Media, www.rkm.com.au

How
does it
work?

How
does it
fail in
disease?



ENZYMES ● Preparatory phase ● Payoff phase

1 Hexokinase
2 Glucose Phosphate Isomerase
3 Phosphofructokinase
4 Fructose diphosphate aldolase
5 Triose phosphate Isomerase
6 Glyceraldehyde Phosphate Dehydrogenase
7 Phosphoglycerate Kinase
8 Phosphoglyceromutase
9 Enolase
10 Pyruvate Kinase

Signaling Pathway

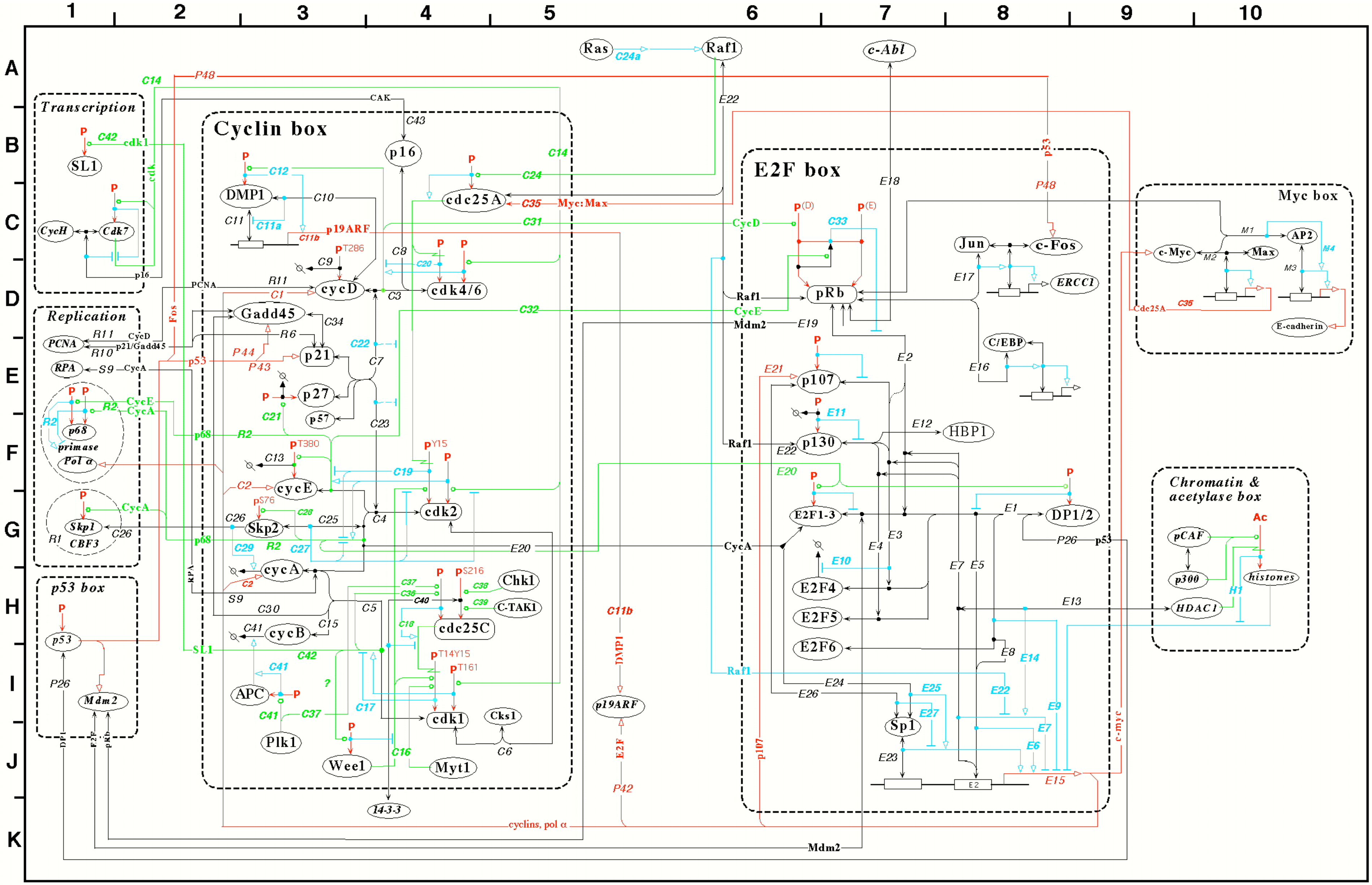
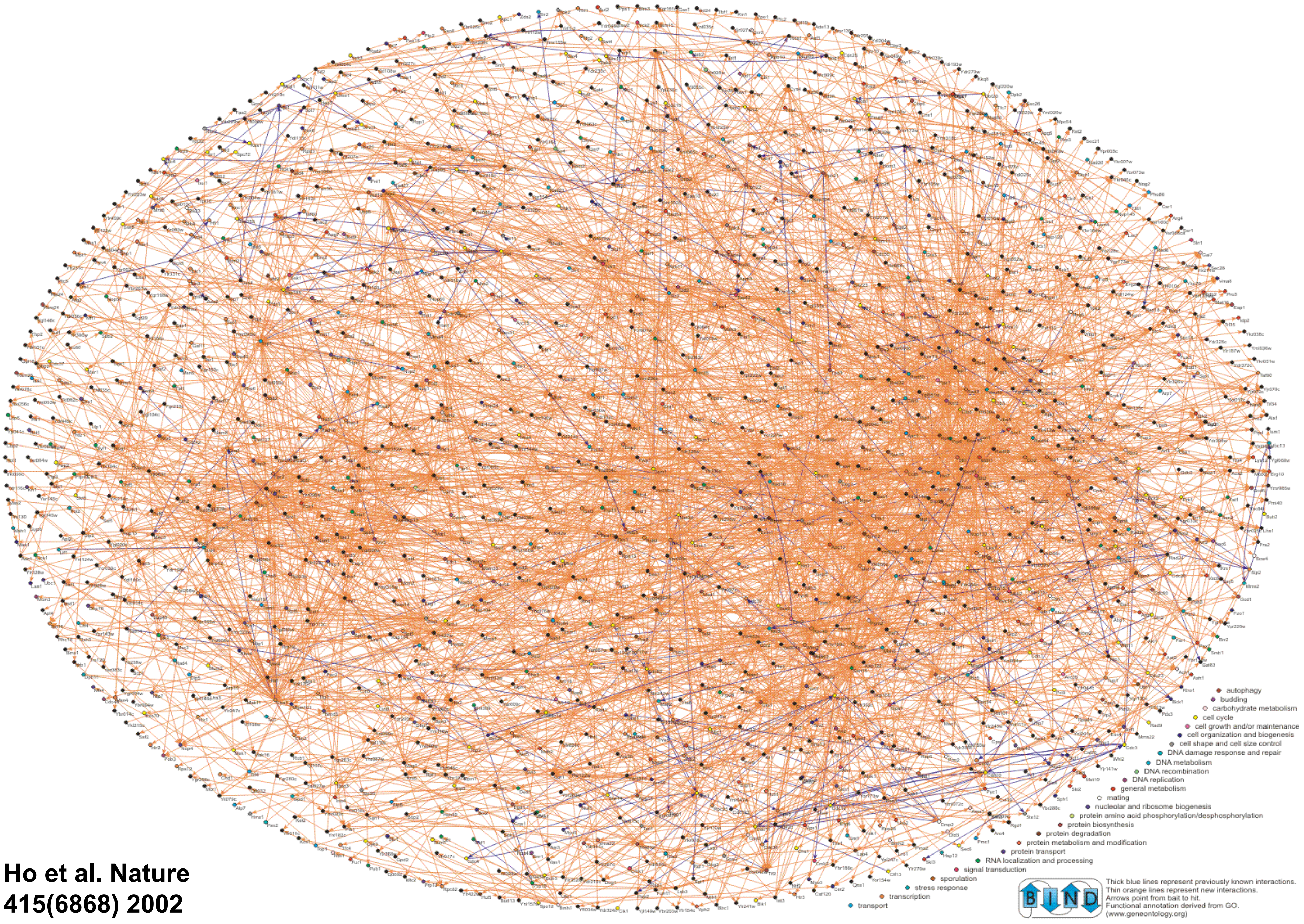


Figure 6A: The Cyclin - E2F cell cycle control system (version 3a - June 8, 1999)

http://discover.nci.nih.gov/kohnk/interaction_maps.html

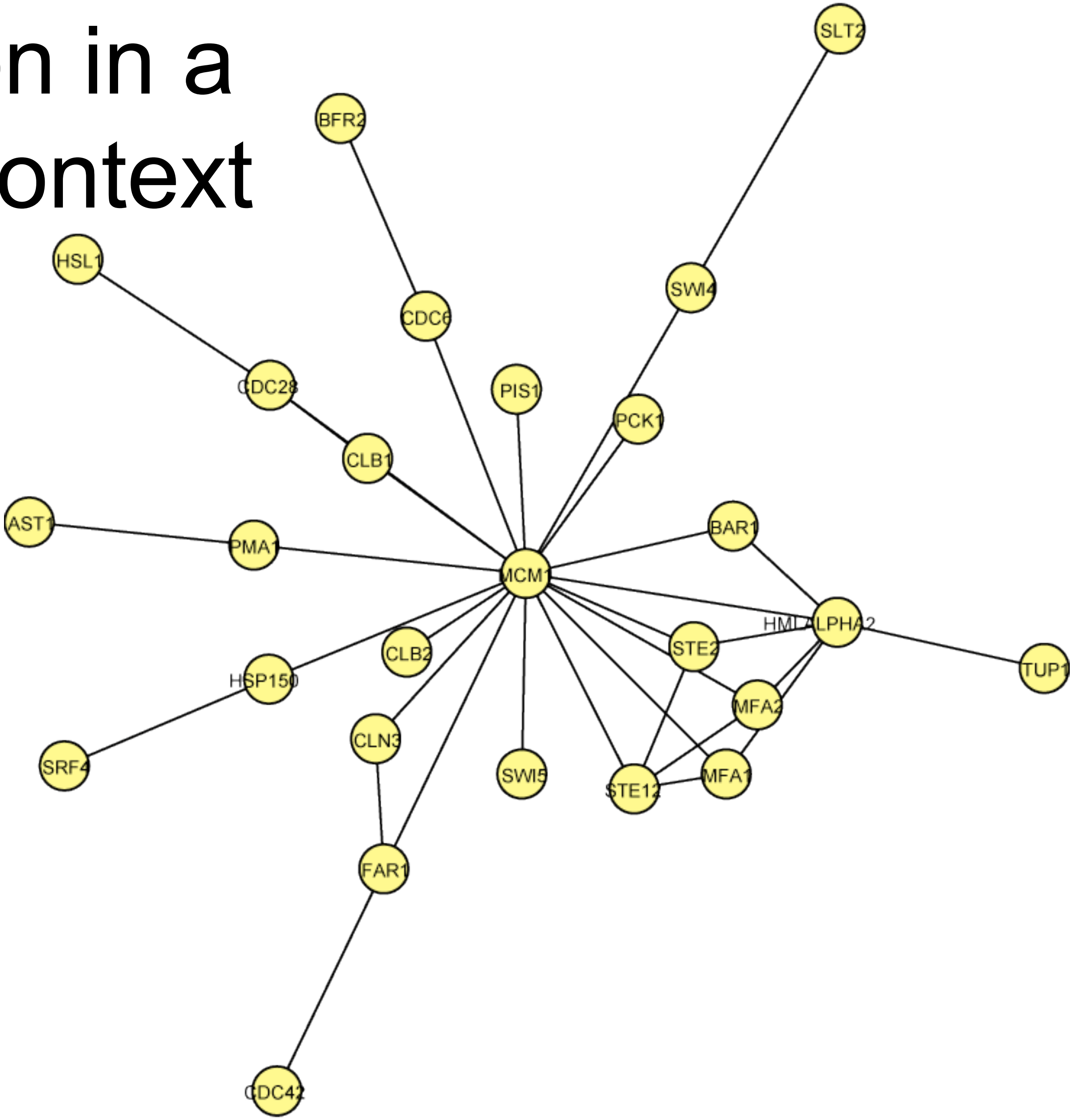
MIM©KE 2k

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

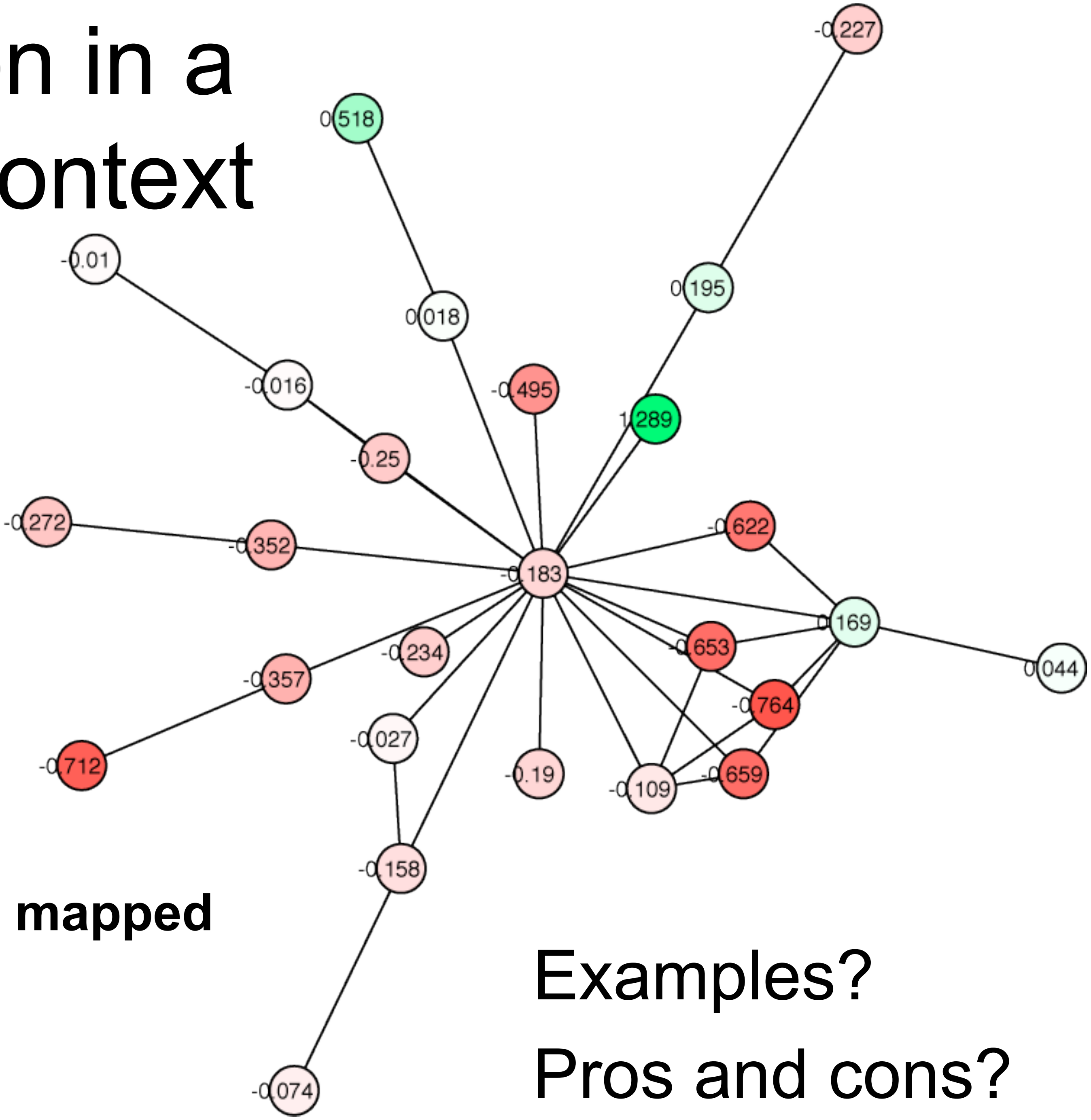


Ho et al. Nature
415(6868) 2002

Integration in a network context



Integration in a network context



Expression data mapped to node colours

Examples?

Pros and cons?

Integration in a network context

Advantages:

-Interpretable

-Broader coverage

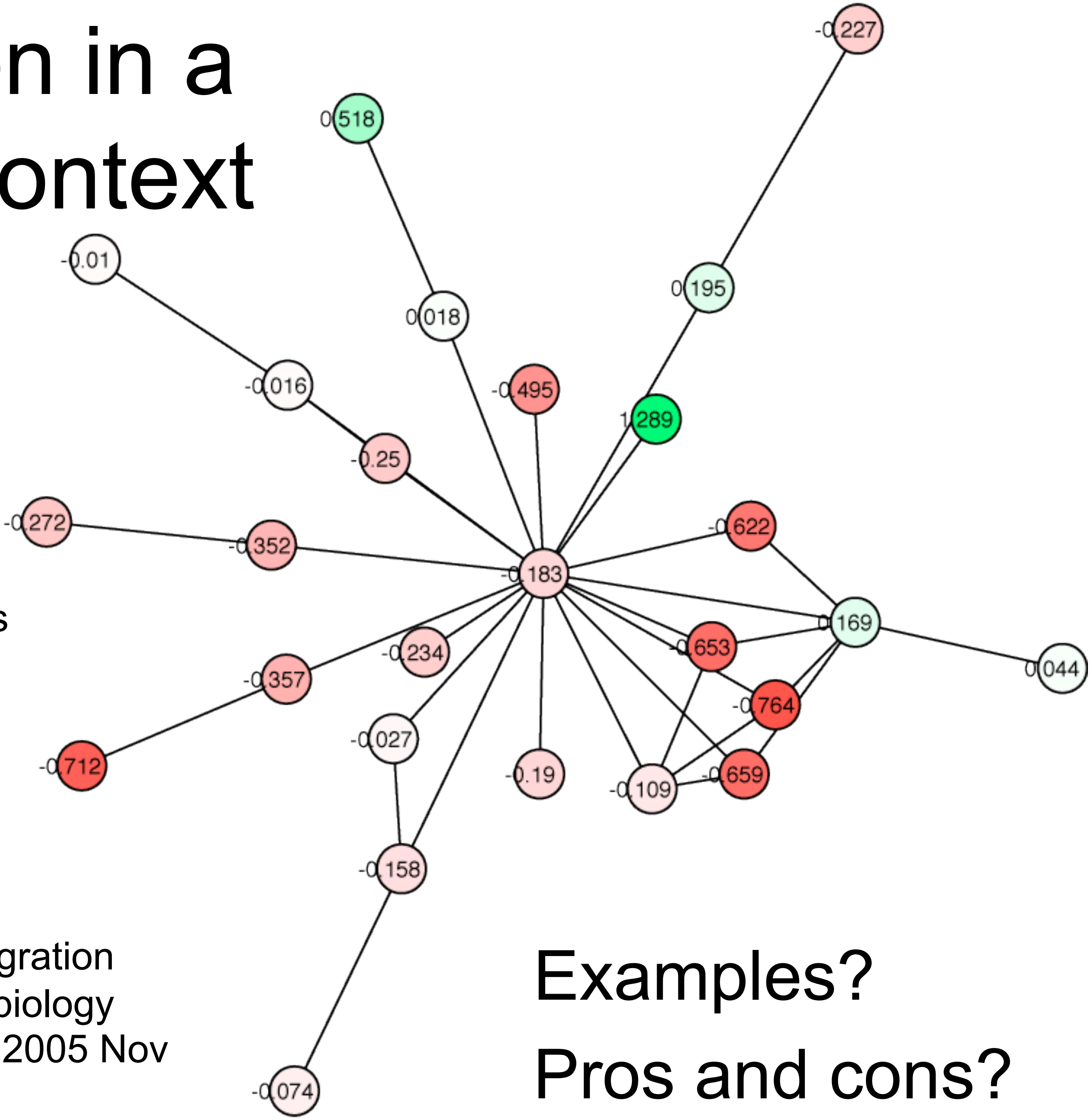
-Error reduction

Challenges:

-Must carefully match data sets to avoid errors e.g. different interaction experiments

-Consider data set bias

-Consider binary vs. discrete vs. continuous

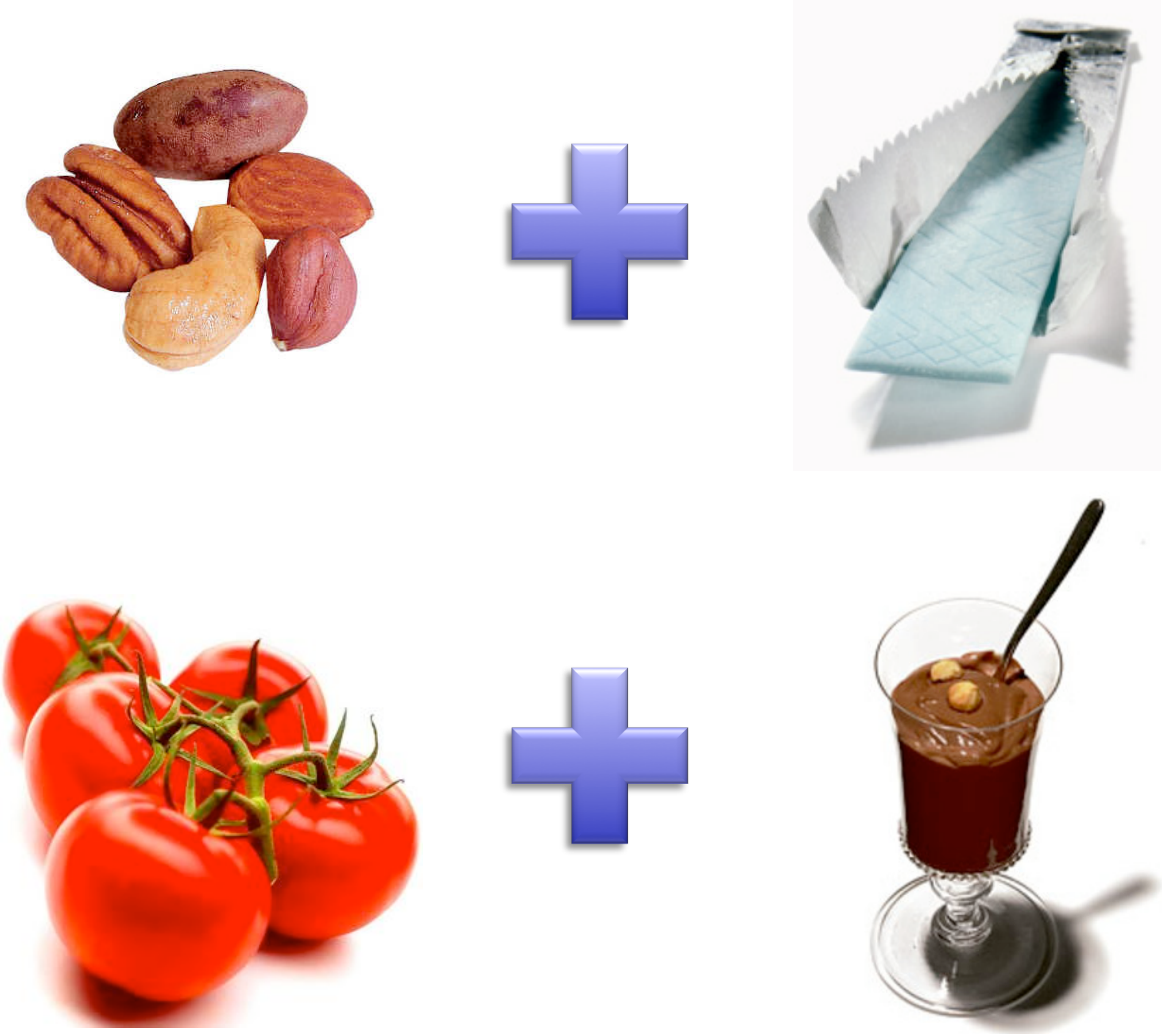


Hwang D et al. A data integration methodology for systems biology Proc Natl Acad Sci U S A. 2005 Nov 29;102(48):17296-301

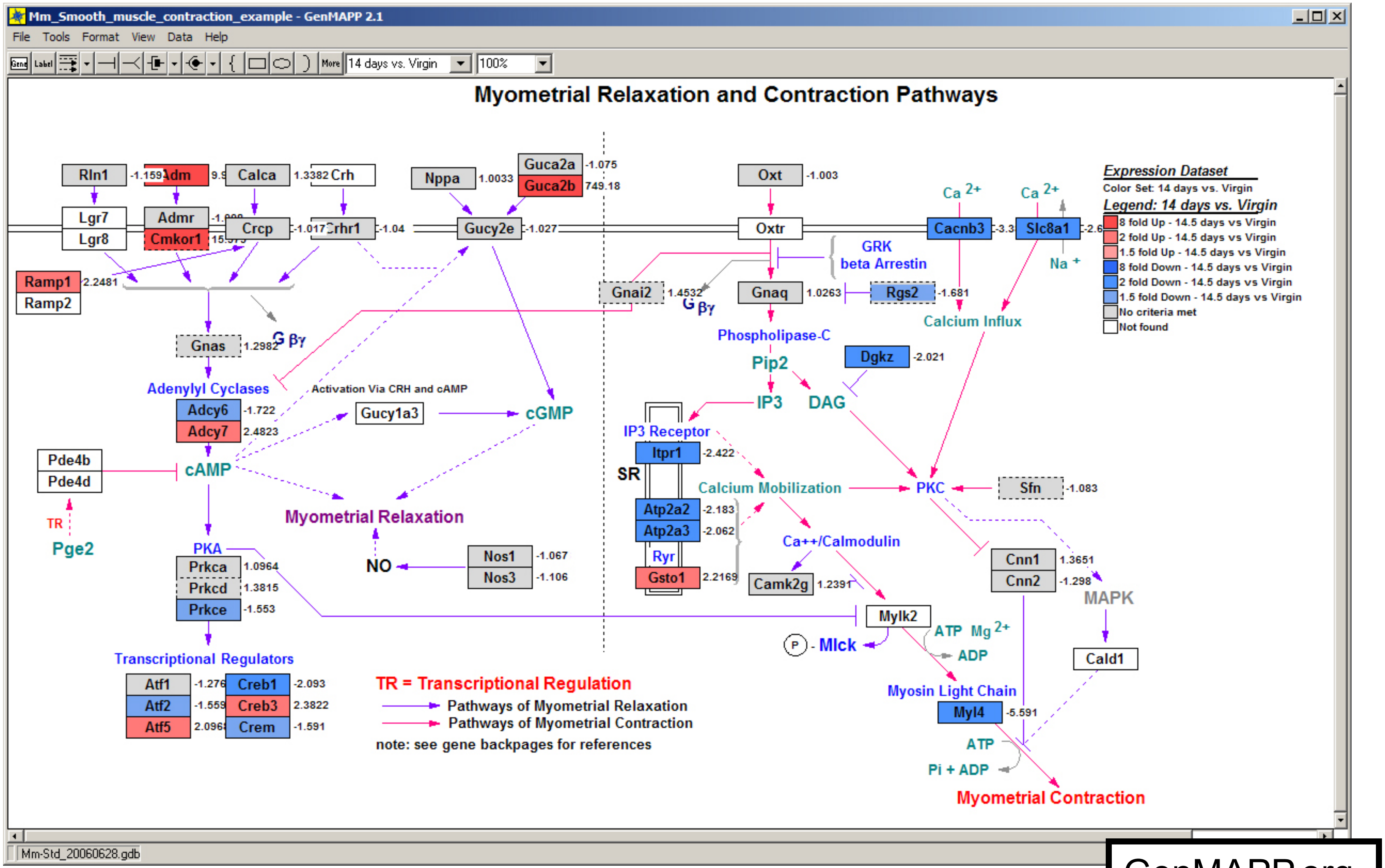
Examples?

Pros and cons?

Data Integration



Visualizing Time Course Data on Pathways: Single Comparison View



TR = Transcriptional Regulation

Pathways of Myometrial Relaxation

Pathways of Myometrial Contraction

note: see gene backpages for references

Expression Dataset

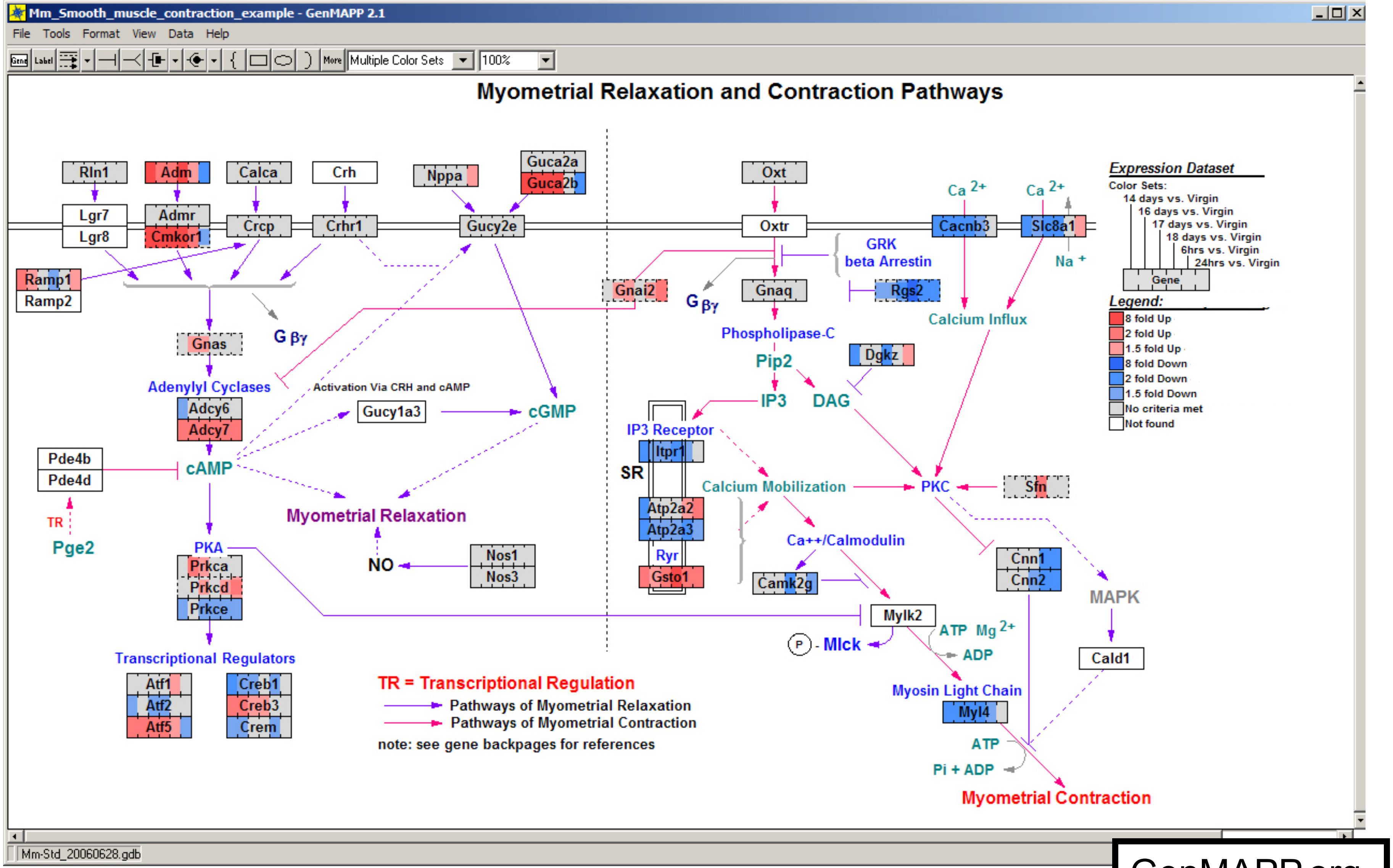
Color Set: 14 days vs. Virgin

Legend: 14 days vs. Virgin

- 🟥 8 fold Up - 14.5 days vs Virgin
- 🟥 2 fold Up - 14.5 days vs Virgin
- 🟥 1.5 fold Up - 14.5 days vs Virgin
- 🟦 8 fold Down - 14.5 days vs Virgin
- 🟦 2 fold Down - 14.5 days vs Virgin
- 🟦 1.5 fold Down - 14.5 days vs Virgin
- ☐ No criteria met
- ☐ Not found

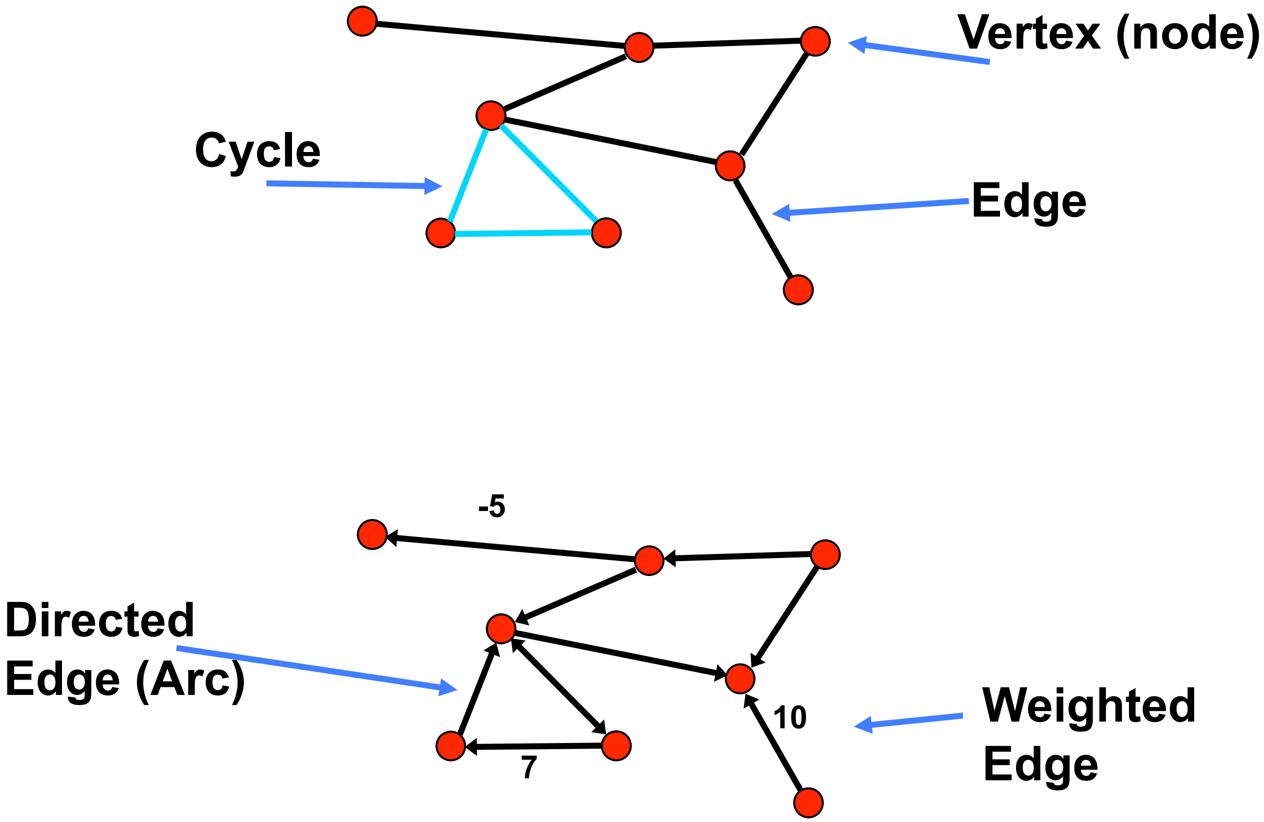
GenMAPP.org

Visualizing Time Course Data on Pathways: Multiple Comparison View



GenMAPP.org

Graph Theory



Vertex (node)

Cycle

Edge

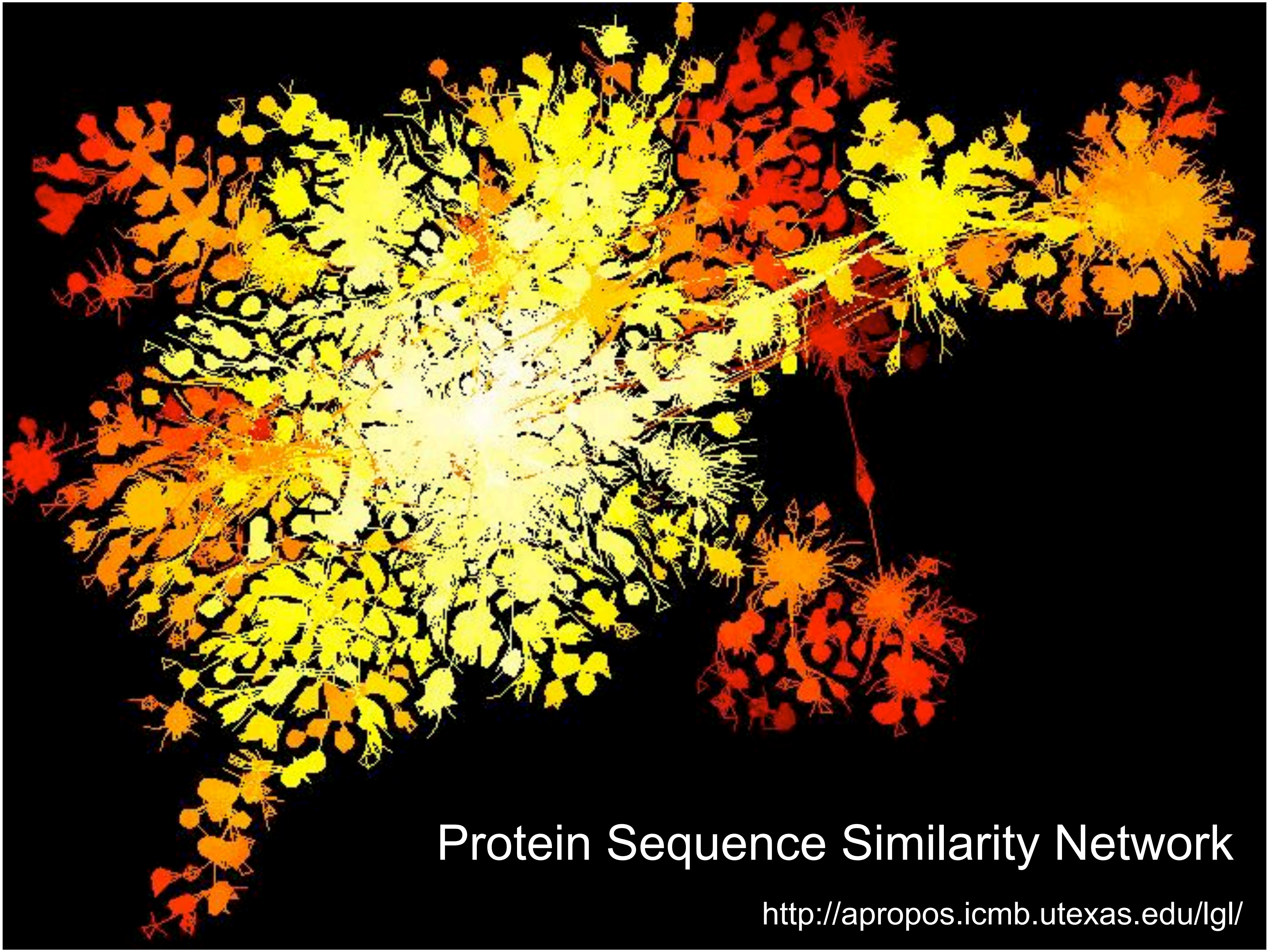
Directed Edge (Arc)

Weighted Edge

We map molecular interaction networks to graphs

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?

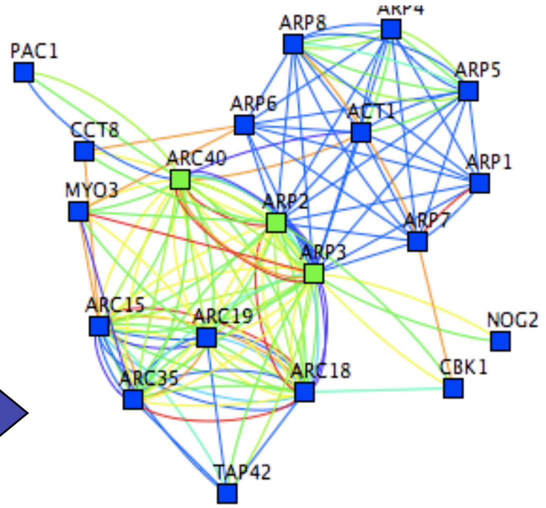
http://www.time.com/time/techtime/200406/community.html

Outline

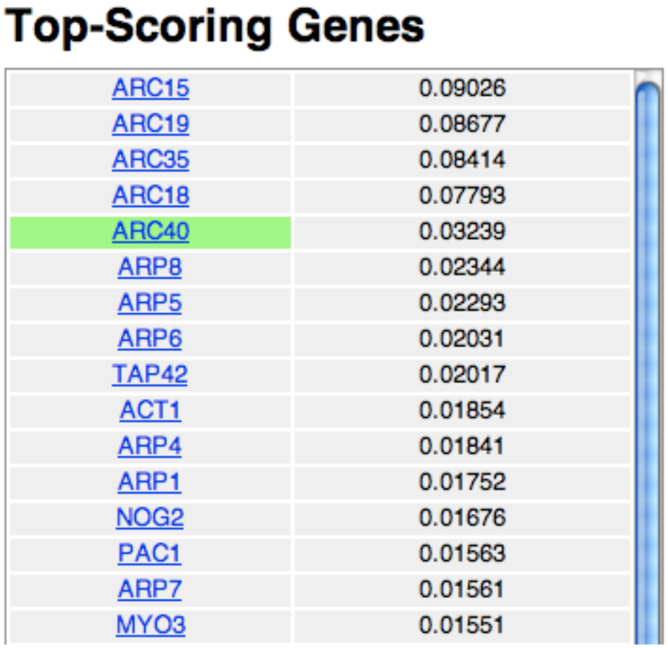
- Gene lists from Omics data
- Networks as scaffold for data integration
- Statistical integration
  - Gene function prediction
- Database and semantic integration
  - Biological pathways
- Visualizing integrated data with Cytoscape

Predicting Gene Function

arp2
arp3
arc40



- STRING
  - http://string.embl.de/
- bioPIXIE
  - http://pixie.princeton.edu/pixie/
- GeneMania
  - http://genemania.org

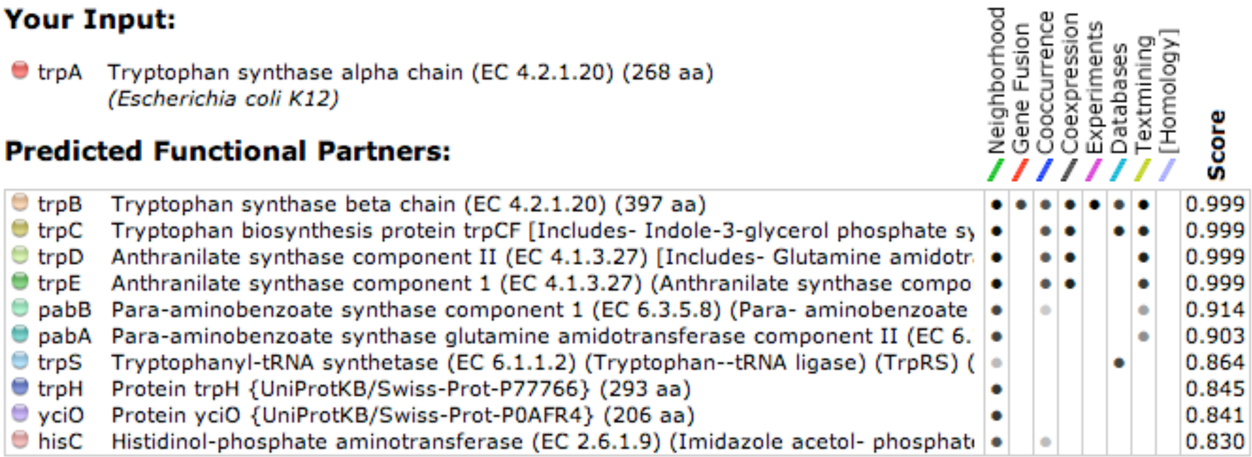


Top-Scoring Genes

ARC15	0.09026
ARC19	0.08677
ARC35	0.08414
ARC18	0.07793
ARC40	0.03239
ARP8	0.02344
ARP5	0.02293
ARP6	0.02031
TAP42	0.02017
ACT1	0.01854
ARP4	0.01841
ARP1	0.01752
NOG2	0.01676
PAC1	0.01563
ARP7	0.01561
MYO3	0.01551

Fraser AG, Marcotte EM - A probabilistic view of gene function - Nat Genet. 2004 Jun;36(6):559-64

STRING

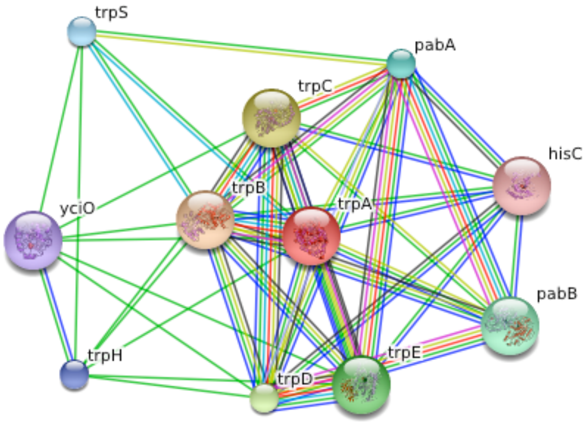


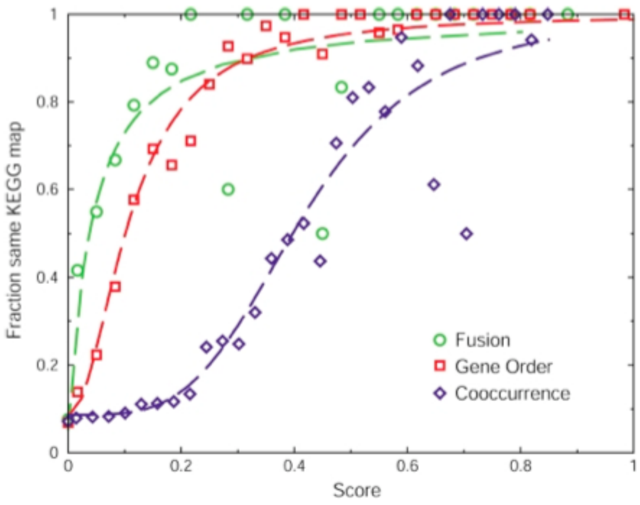
Your Input:

● trpA Tryptophan synthase alpha chain (EC 4.2.1.20) (268 aa)
(Escherichia coli K12)

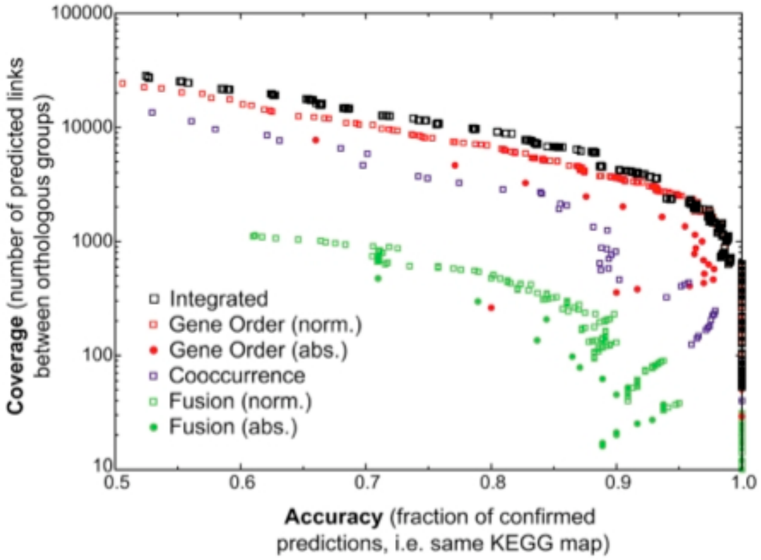
Predicted Functional Partners:

			Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	[Homology]	**Score**
●	trpB	Tryptophan synthase beta chain (EC 4.2.1.20) (397 aa)	●	●	●	●	●	●	●		0.999
●	trpC	Tryptophan biosynthesis protein trpCF [Includes- Indole-3-glycerol phosphate sy	●		●	●		●	●		0.999
●	trpD	Anthranilate synthase component II (EC 4.1.3.27) [Includes- Glutamine amidotr	●		●	●			●		0.999
●	trpE	Anthranilate synthase component 1 (EC 4.1.3.27) (Anthranilate synthase compo	●		●	●			●		0.999
●	pabB	Para-aminobenzoate synthase component 1 (EC 6.3.5.8) (Para- aminobenzoate	●		●				●		0.914
●	pabA	Para-aminobenzoate synthase glutamine amidotransferase component II (EC 6.	●						●		0.903
●	trpS	Tryptophanyl-tRNA synthetase (EC 6.1.1.2) (Tryptophan--tRNA ligase) (TrpRS) (	●					●			0.864
●	trpH	Protein trpH {UniProtKB/Swiss-Prot-P77766} (293 aa)	●								0.845
●	yciO	Protein yciO {UniProtKB/Swiss-Prot-P0AFR4} (206 aa)	●								0.841
●	hisC	Histidinol-phosphate aminotransferase (EC 2.6.1.9) (Imidazole acetol- phosphat	●		●						0.830



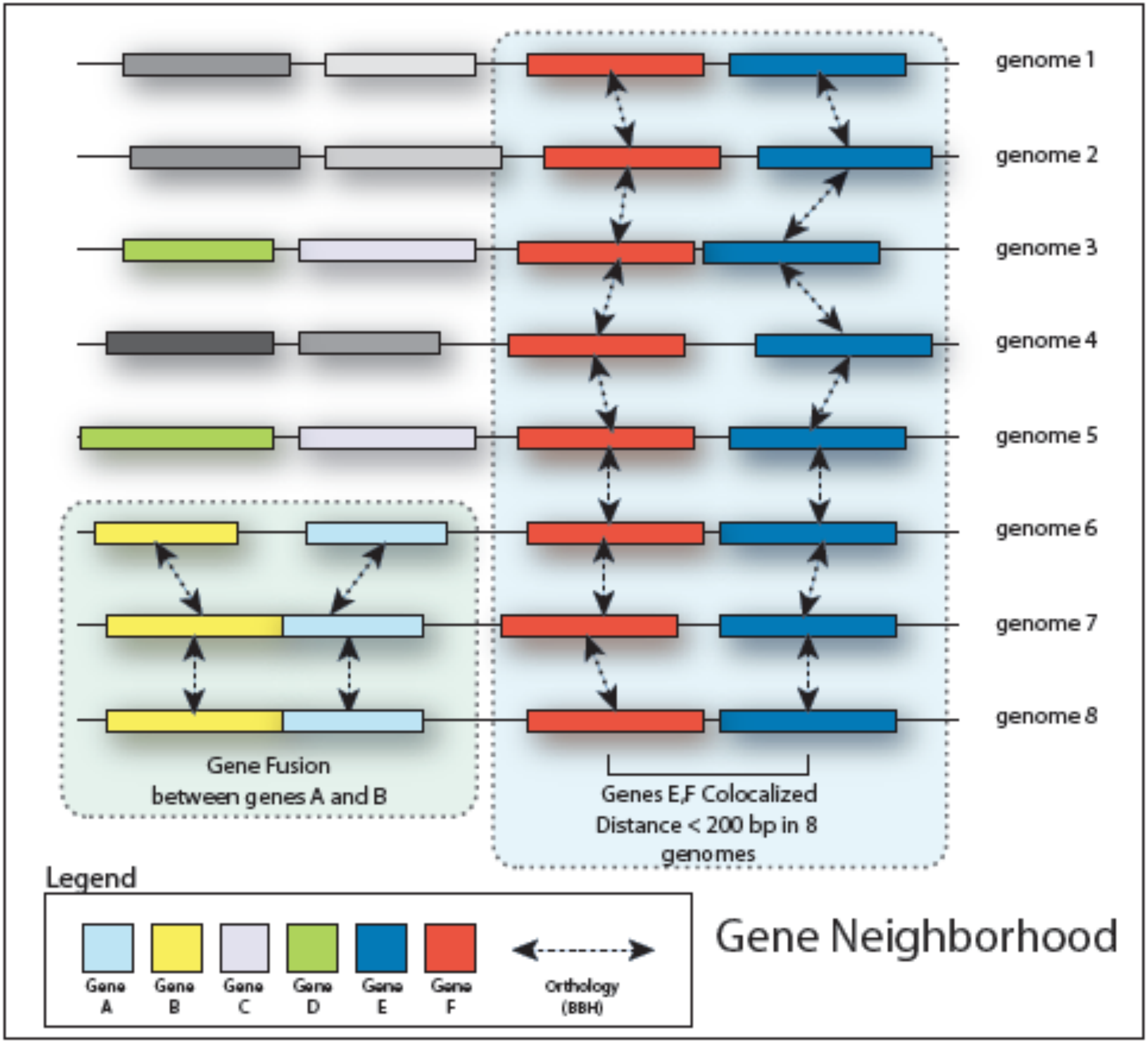


x=score
a=intercept
b=cooperativity
c=half max for x



$$f(x) = a + [(1 - a)x^b/(c^b + x^b)]$$

von Mering C et al. Nucleic Acids Res. 2003 Jan 1;31(1):258-61

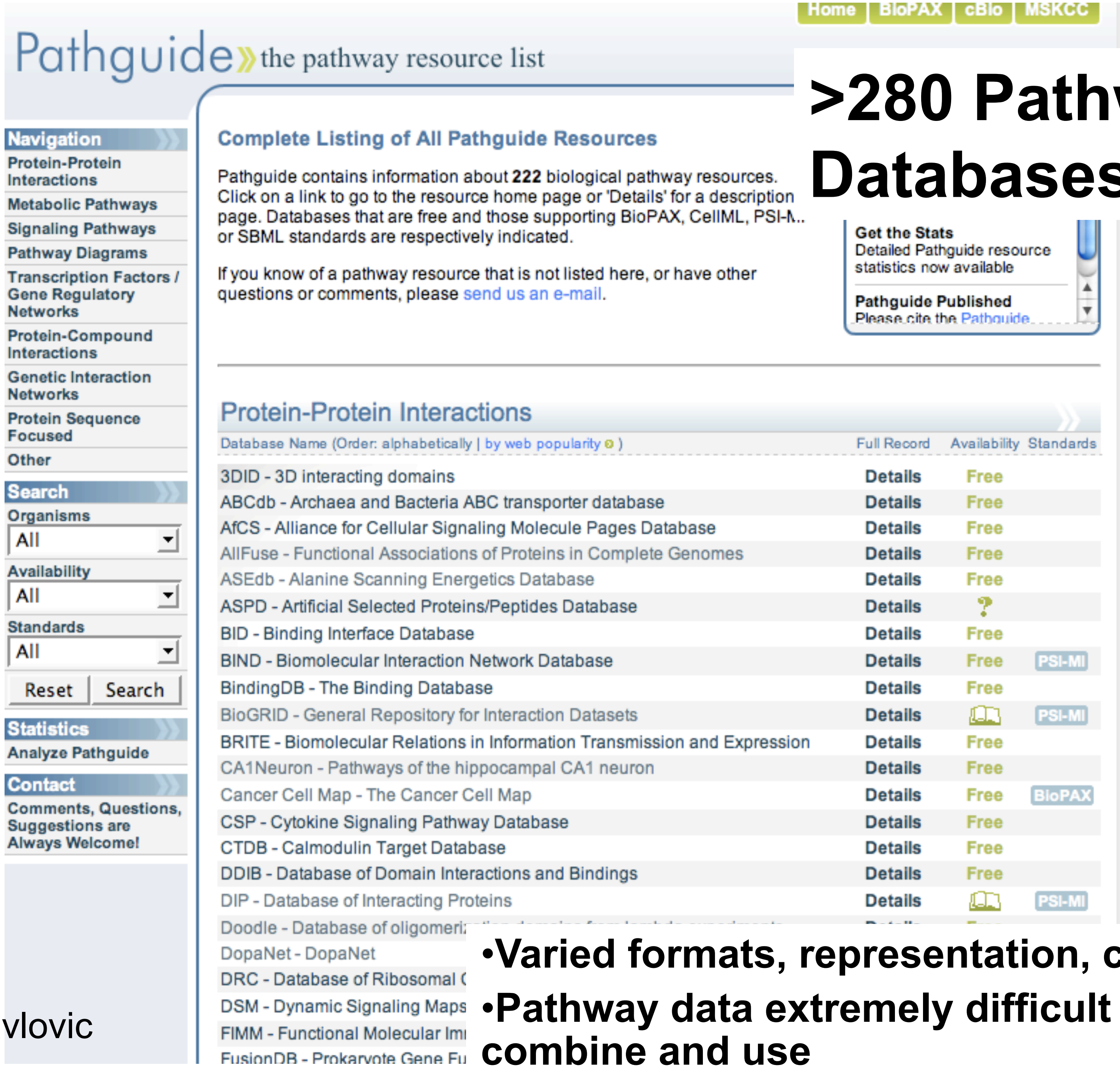

genome 1
genome 2
genome 3
genome 4
genome 5
genome 6
genome 7
genome 8
Gene Fusion
between genes A and B
Genes E,F Colocalized
Distance < 200 bp in 8
genomes
Legend
Gene A
Gene B
Gene C
Gene D
Gene E
Gene F
Orthology (BBH)
Gene Neighborhood

Outline

- Gene lists from Omics data
- Networks as scaffold for data integration
- Statistical integration
  - Gene function prediction
- Database and semantic integration
  - Biological pathways
- Visualizing integrated data with Cytoscape

http://pathguide.org

>280 Pathway Databases!



Pathguide» the pathway resource list

Navigation
- Protein-Protein Interactions
- Metabolic Pathways
- Signaling Pathways
- Pathway Diagrams
- Transcription Factors / Gene Regulatory Networks
- Protein-Compound Interactions
- Genetic Interaction Networks
- Protein Sequence Focused
- Other

Search

Organisms: All

Availability: All

Standards: All

Reset | Search

Statistics

Analyze Pathguide

Contact

Comments, Questions, Suggestions are Always Welcome!

Complete Listing of All Pathguide Resources

Pathguide contains information about **222** biological pathway resources. Click on a link to go to the resource home page or 'Details' for a description page. Databases that are free and those supporting BioPAX, CellML, PSI-M.. or SBML standards are respectively indicated.

If you know of a pathway resource that is not listed here, or have other questions or comments, please send us an e-mail.

Get the Stats
Detailed Pathguide resource statistics now available

Pathguide Published
Please cite the Pathguide

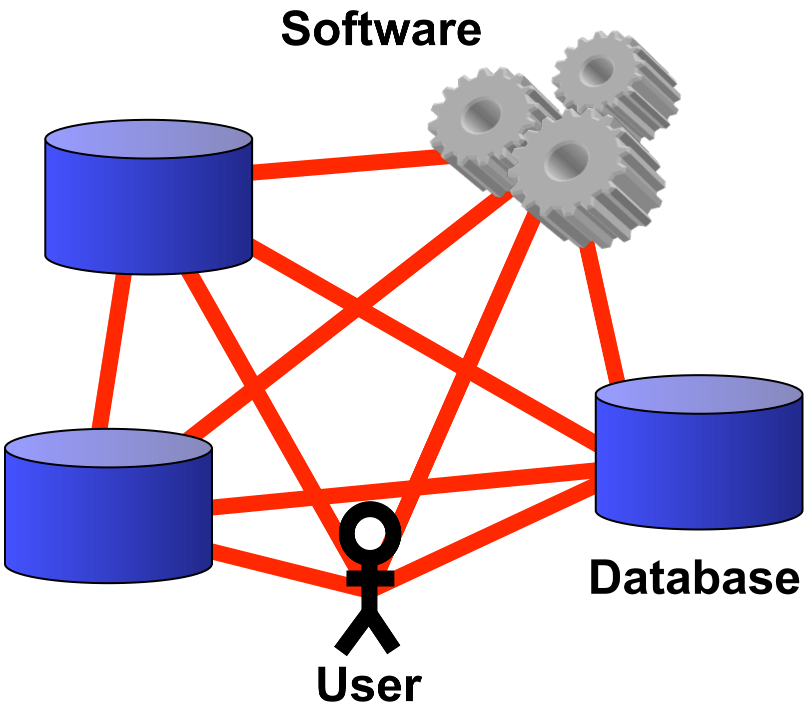
Protein-Protein Interactions

Database Name (Order: alphabetically | by web popularity ⊙)	Full Record	Availability	Standards
3DID - 3D interacting domains	Details	Free	
ABCdb - Archaea and Bacteria ABC transporter database	Details	Free	
AfCS - Alliance for Cellular Signaling Molecule Pages Database	Details	Free	
AllFuse - Functional Associations of Proteins in Complete Genomes	Details	Free	
ASEdb - Alanine Scanning Energetics Database	Details	Free	
ASPD - Artificial Selected Proteins/Peptides Database	Details	?	
BID - Binding Interface Database	Details	Free	
BIND - Biomolecular Interaction Network Database	Details	Free	PSI-MI
BindingDB - The Binding Database	Details	Free	
BioGRID - General Repository for Interaction Datasets	Details	📖	PSI-MI
BRITE - Biomolecular Relations in Information Transmission and Expression	Details	Free	
CA1Neuron - Pathways of the hippocampal CA1 neuron	Details	Free	
Cancer Cell Map - The Cancer Cell Map	Details	Free	BioPAX
CSP - Cytokine Signaling Pathway Database	Details	Free	
CTDB - Calmodulin Target Database	Details	Free	
DDIB - Database of Domain Interactions and Bindings	Details	Free	
DIP - Database of Interacting Proteins	Details	📖	PSI-MI
Doodle - Database of oligomeri			
DopaNet - DopaNet			
DRC - Database of Ribosomal (			
DSM - Dynamic Signaling Maps			
FIMM - Functional Molecular Im			
FusionDB - Prokaryote Gene Fu			

- Varied formats, representation, coverage
- Pathway data extremely difficult to combine and use

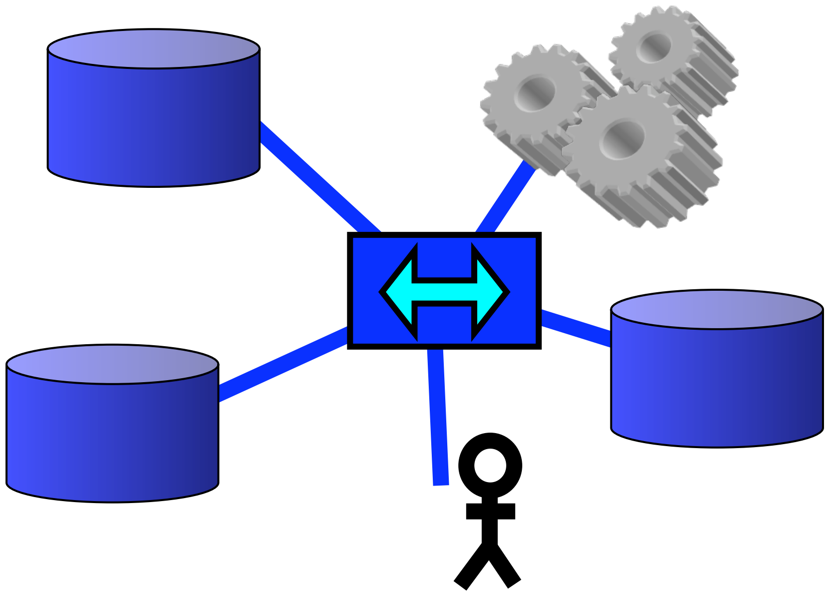
Vuk Pavlovic

Biological Pathway Exchange (BioPAX)



Before BioPAX

>100 DBs and tools
Tower of Babel



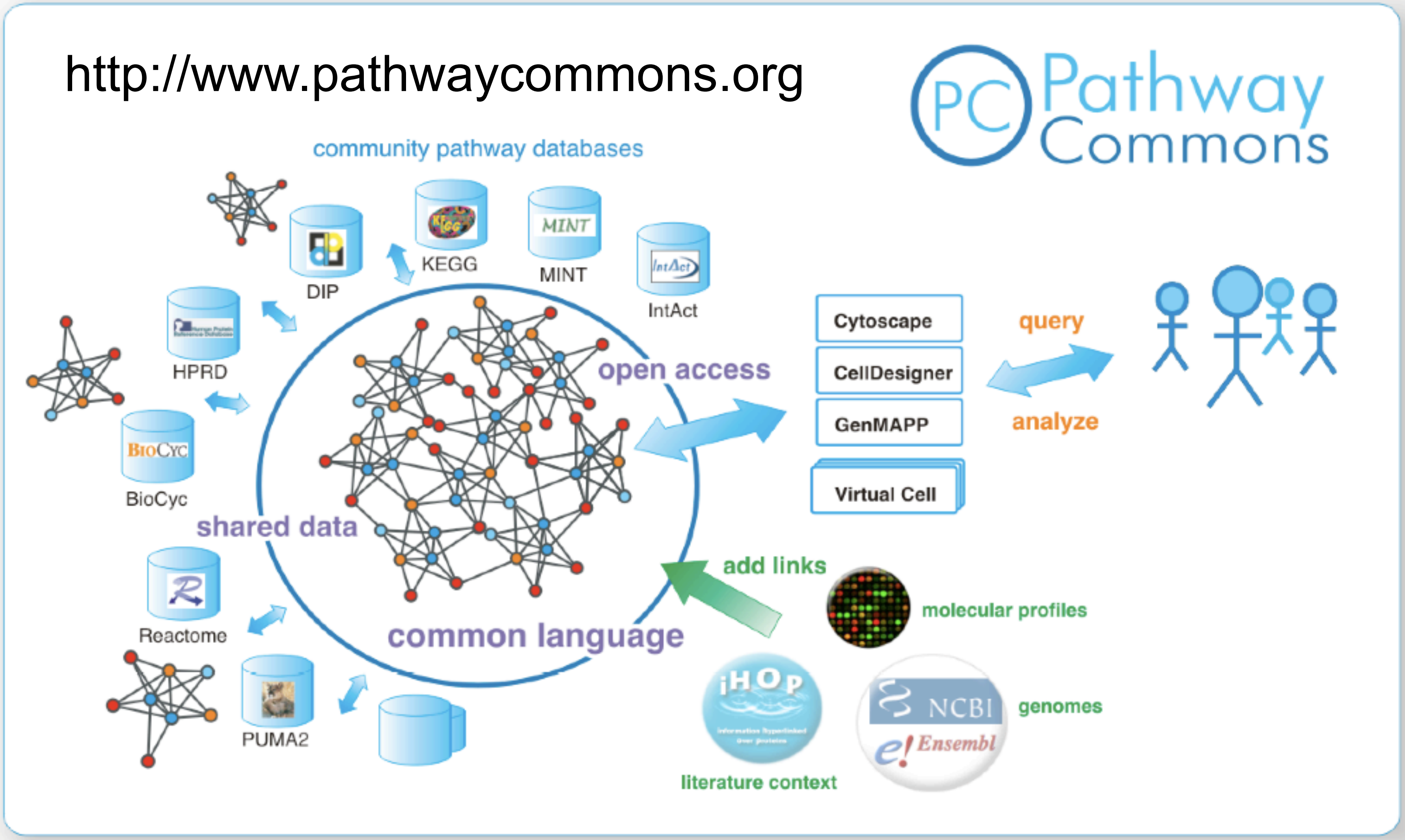
After BioPAX
Unifying language

Reduces work, promotes collaboration, increases accessibility

BioPAX Pathway Language

- Represent:
  - Metabolic pathways
  - Signaling pathways
  - Protein-protein, molecular interactions
  - Gene regulatory pathways
  - Genetic interactions
- Community effort: pathway databases distribute pathway information in standard format
  - Over 100 people, database groups, standard efforts

Aim: Convenient Access to Pathway Information



Facilitate creation and communication of pathway data
Aggregate pathway data in the public domain
Provide easy access for pathway analysis

Long term: Converge to integrated cell map http://pathwaycommons.org

Pathway Commons is a convenient point of access to biological pathway information collected from public pathway databases, which you can browse or search. Pathways include biochemical reactions, complex assembly, transport and catalysis events, and physical interactions involving proteins, DNA, RNA, small molecules and complexes. more...

Search Pathway Commons:

Search

To get started, enter a gene name, gene identifier or pathway name in the text box above. For example: p53, P38398 or mTOR.

To restrict your search to specific data sources or specific organisms, update your global filter settings.

Pathway Commons currently contains the following data sources:

Cancer Cell Map, Release: 1.0 [19-May-06]
Browse

HumanCyc, Release: 10.5 [18-Sep-06]
Browse

NCI / Nature Pathway Interaction Database [01-Jan-07]
Browse

Reactome, Release: 19 [16-Nov-06]
Browse

Pathway Commons Quick Stats:

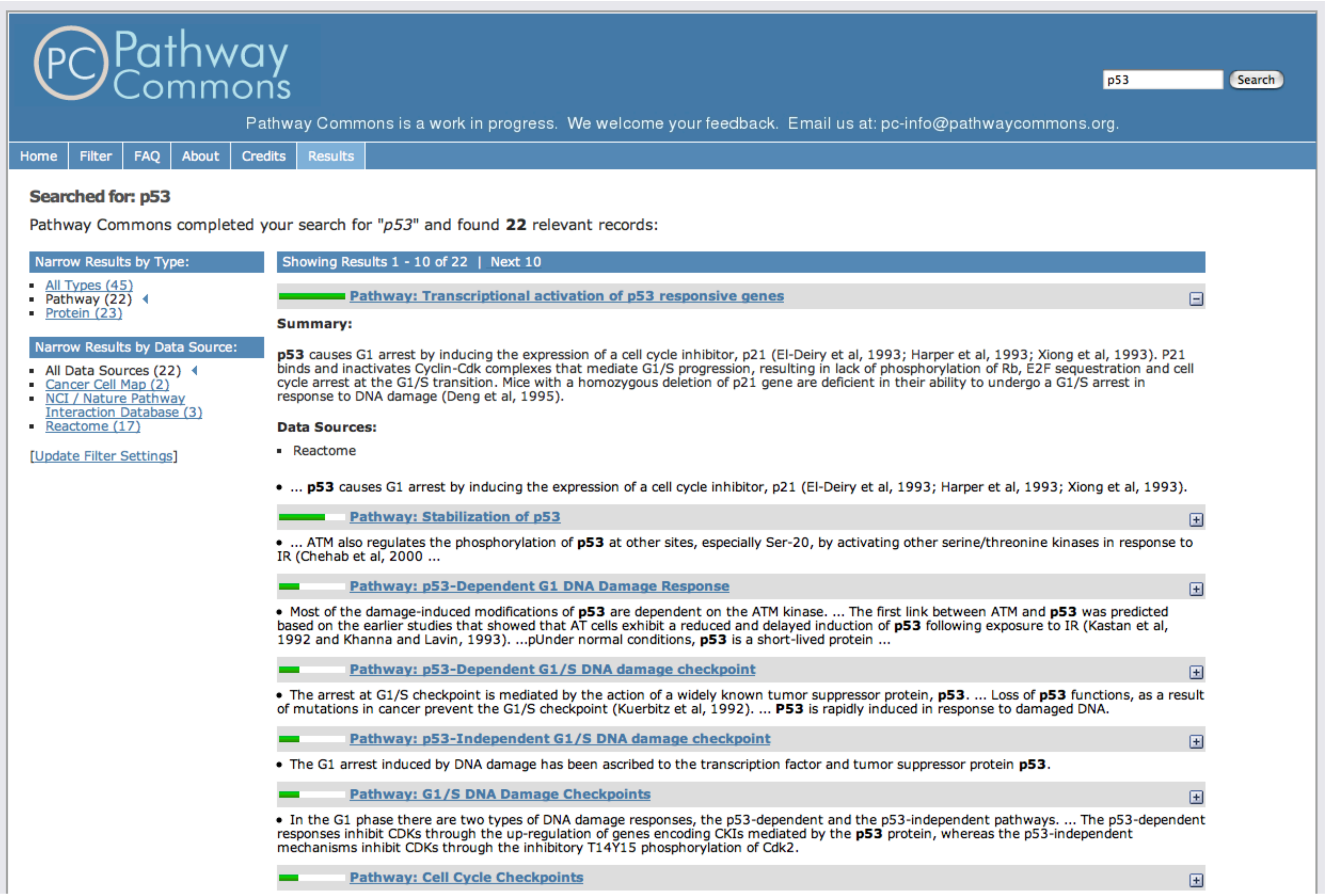
Number of Pathways:	921
Number of Interactions:	9,924
Number of Physical Entities:	15,515
Number of Organisms:	10

Biologists: Browse and search pathways across multiple valuable public pathway databases.

Computational biologists: Download an integrated set of pathways in BioPAX format for global analysis.

Software developers: Build software on top of Pathway Commons using our soon-to-be released web service API. Download and install the cPath software to create a local mirror.

Pathway Commons is hosted by the **Computational Biology Center** at **Memorial Sloan-Kettering Cancer Center** and the **University of Toronto**.



Pathway Commons is a work in progress. We welcome your feedback. Email us at: pc-info@pathwaycommons.org.

Searched for: p53

Pathway Commons completed your search for "*p53*" and found **22** relevant records:

Narrow Results by Type:

- All Types (45)
- Pathway (22)
- Protein (23)

Narrow Results by Data Source:

- All Data Sources (22)
- Cancer Cell Map (2)
- NCI / Nature Pathway Interaction Database (3)
- Reactome (17)

[Update Filter Settings]

Showing Results 1 - 10 of 22 | Next 10

Pathway: Transcriptional activation of p53 responsive genes

Summary:

p53 causes G1 arrest by inducing the expression of a cell cycle inhibitor, p21 (El-Deiry et al, 1993; Harper et al, 1993; Xiong et al, 1993). P21 binds and inactivates Cyclin-Cdk complexes that mediate G1/S progression, resulting in lack of phosphorylation of Rb, E2F sequestration and cell cycle arrest at the G1/S transition. Mice with a homozygous deletion of p21 gene are deficient in their ability to undergo a G1/S arrest in response to DNA damage (Deng et al, 1995).

Data Sources:

- Reactome

- ... **p53** causes G1 arrest by inducing the expression of a cell cycle inhibitor, p21 (El-Deiry et al, 1993; Harper et al, 1993; Xiong et al, 1993).

Pathway: Stabilization of p53

- ... ATM also regulates the phosphorylation of **p53** at other sites, especially Ser-20, by activating other serine/threonine kinases in response to IR (Chehab et al, 2000 ...

Pathway: p53-Dependent G1 DNA Damage Response

- Most of the damage-induced modifications of **p53** are dependent on the ATM kinase. ... The first link between ATM and **p53** was predicted based on the earlier studies that showed that AT cells exhibit a reduced and delayed induction of **p53** following exposure to IR (Kastan et al, 1992 and Khanna and Lavin, 1993). ...pUnder normal conditions, **p53** is a short-lived protein ...

Pathway: p53-Dependent G1/S DNA damage checkpoint

- The arrest at G1/S checkpoint is mediated by the action of a widely known tumor suppressor protein, **p53**. ... Loss of **p53** functions, as a result of mutations in cancer prevent the G1/S checkpoint (Kuerbitz et al, 1992). ... **P53** is rapidly induced in response to damaged DNA.

Pathway: p53-Independent G1/S DNA damage checkpoint

- The G1 arrest induced by DNA damage has been ascribed to the transcription factor and tumor suppressor protein **p53**.

Pathway: G1/S DNA Damage Checkpoints

- In the G1 phase there are two types of DNA damage responses, the p53-dependent and the p53-independent pathways. ... The p53-dependent responses inhibit CDKs through the up-regulation of genes encoding CKIs mediated by the **p53** protein, whereas the p53-independent mechanisms inhibit CDKs through the inhibitory T14Y15 phosphorylation of Cdk2.

Pathway: Cell Cycle Checkpoints

http://pathwaycommons.org

Pathway Commons Status

Cancer Cell Map, Release: 1.0 [19-May-06]
Browse

HPRD [01-Sep-07]
Browse

HumanCyc, Release: 10.5 [18-Sep-06]
Browse

IntAct [14-Dec-07]
Browse

MINT [21-Dec-07]
Browse

NCI / Nature Pathway Interaction Database [28-Jan-08]
Browse

Reactome, Release: 24 [12-Mar-08]
Browse

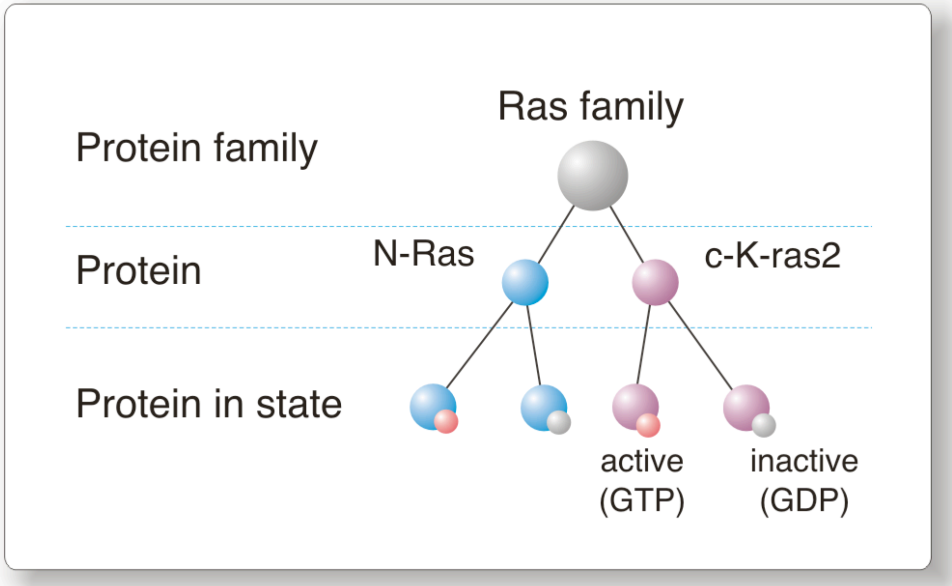
- More databases
  - INOH, Panther
- Neighborhood visualization
- Download service

Pathway Commons Quick Stats:

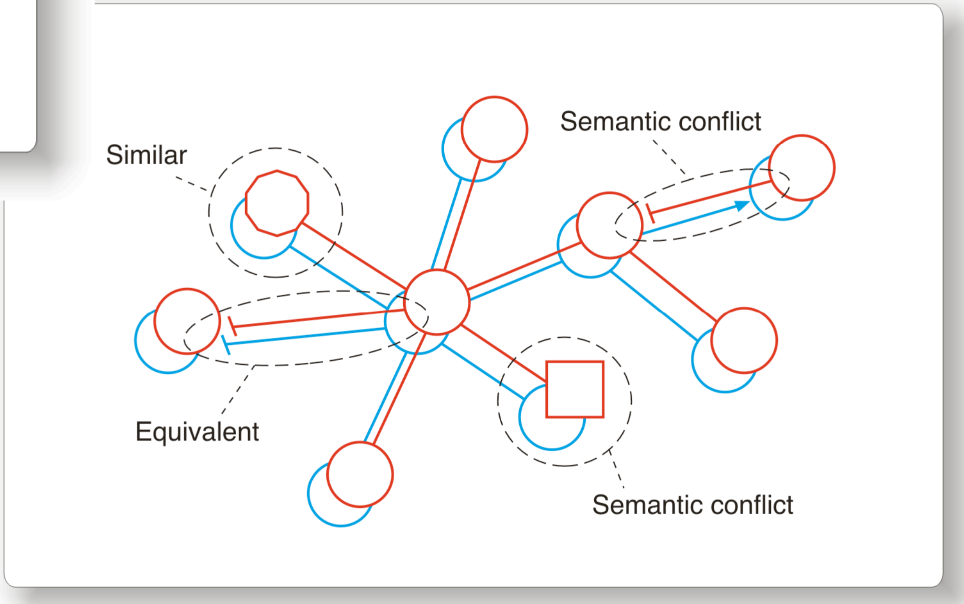
Number of Pathways:	1,391
Number of Interactions:	418,544
Number of Physical Entities:	85,548
Number of Organisms:	1,005

Towards an Integrated Cell Map

- Semantic pathway integration is difficult



Relationships



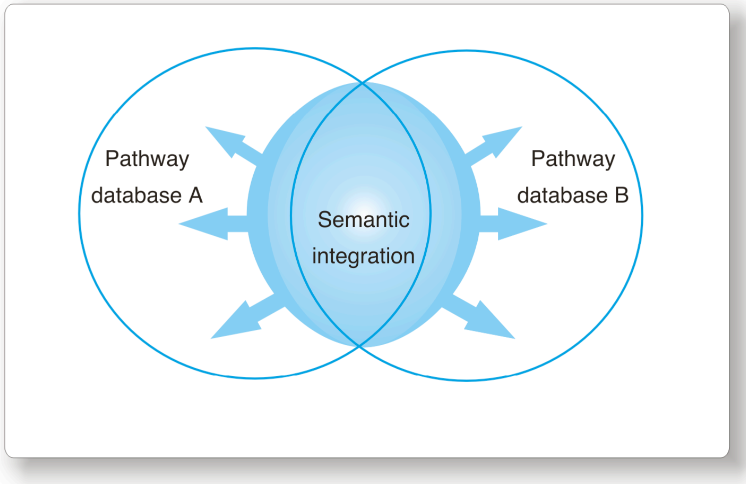
Physical entities

Determining equivalent entities is critical

Practical Semantic Integration

- Minimize errors
  - Integrate only where possible with high accuracy
  - Detect and flag conflicts, errors for users, no revision
  - Promote best-practices to minimize future errors
  - Interaction confidence algorithms
  - Validation software
  - Allow users to filter and select trusted sources
- Converge to standard representation
  - Community process

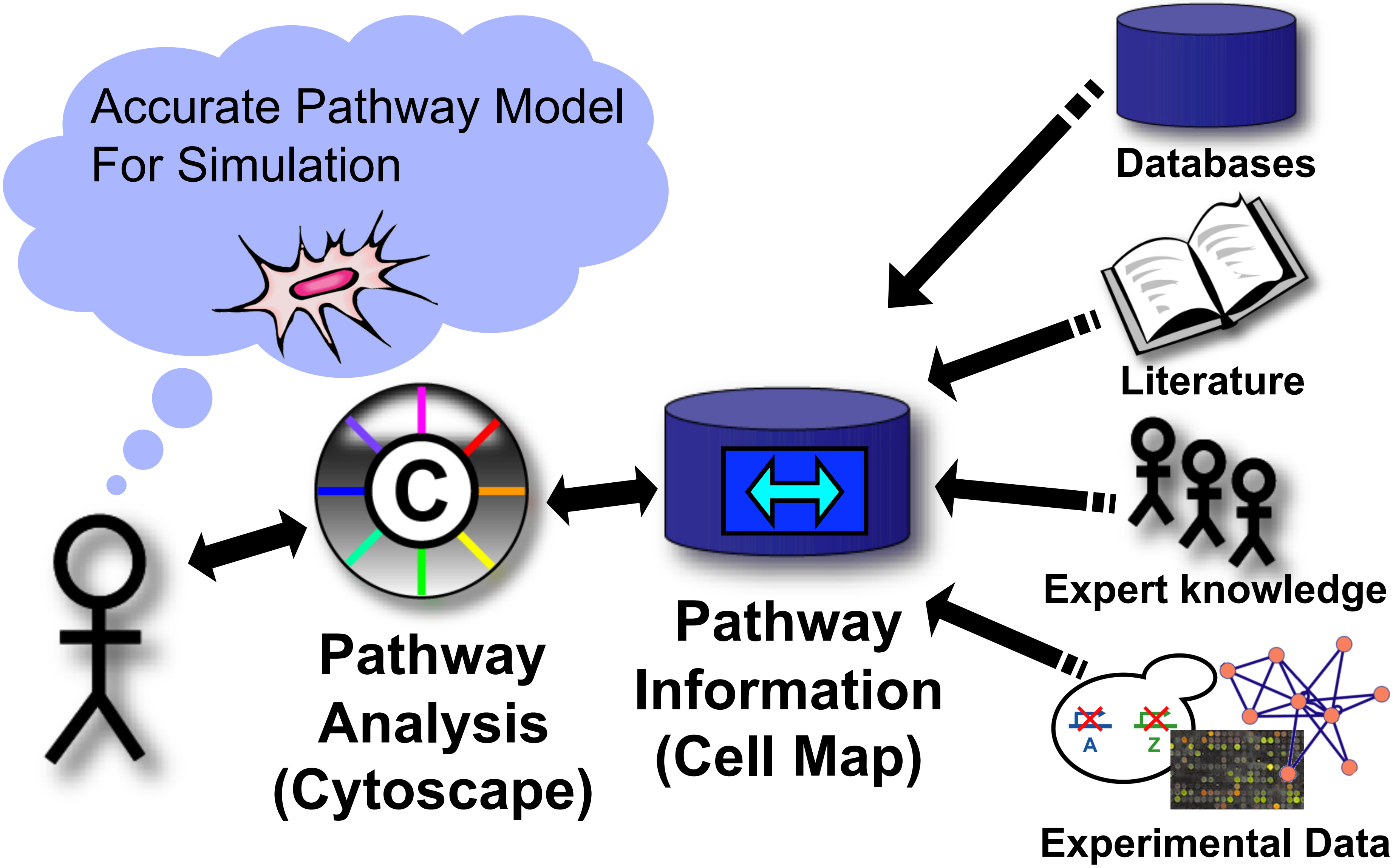
Doable: hundreds of curators globally in >200 databases (GDP) - make it more efficient



Outline

- Gene lists from Omics data
- Networks as scaffold for data integration
- Statistical integration
  - Gene function prediction
- Database and semantic integration
  - Biological pathways
- Visualizing integrated data with Cytoscape

Using Pathway Information



Accurate Pathway Model
For Simulation

Pathway
Analysis
(Cytoscape)

Pathway
Information
(Cell Map)

Databases

Literature

Expert knowledge

Experimental Data

Cytoscape

UNIVERSITY OF TORONTO

UCSF

NCIBI
National Center for Integrative Biomedical Informatics

Unilever

MEMORIAL SLOAN-KETTERING CANCER CENTER · ESTABLISHED 1884 ·

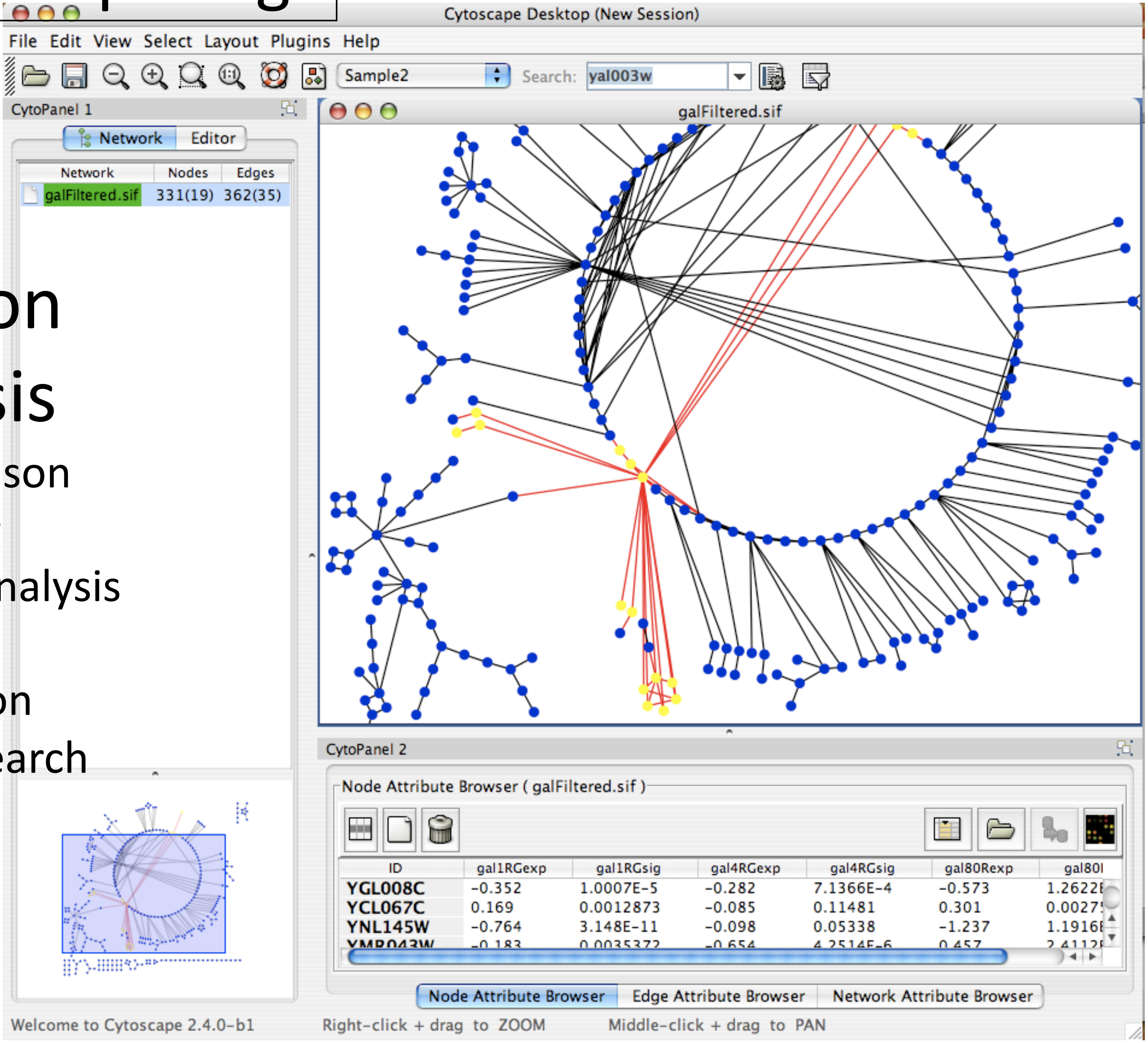
Agilent Technologies

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 Michigan

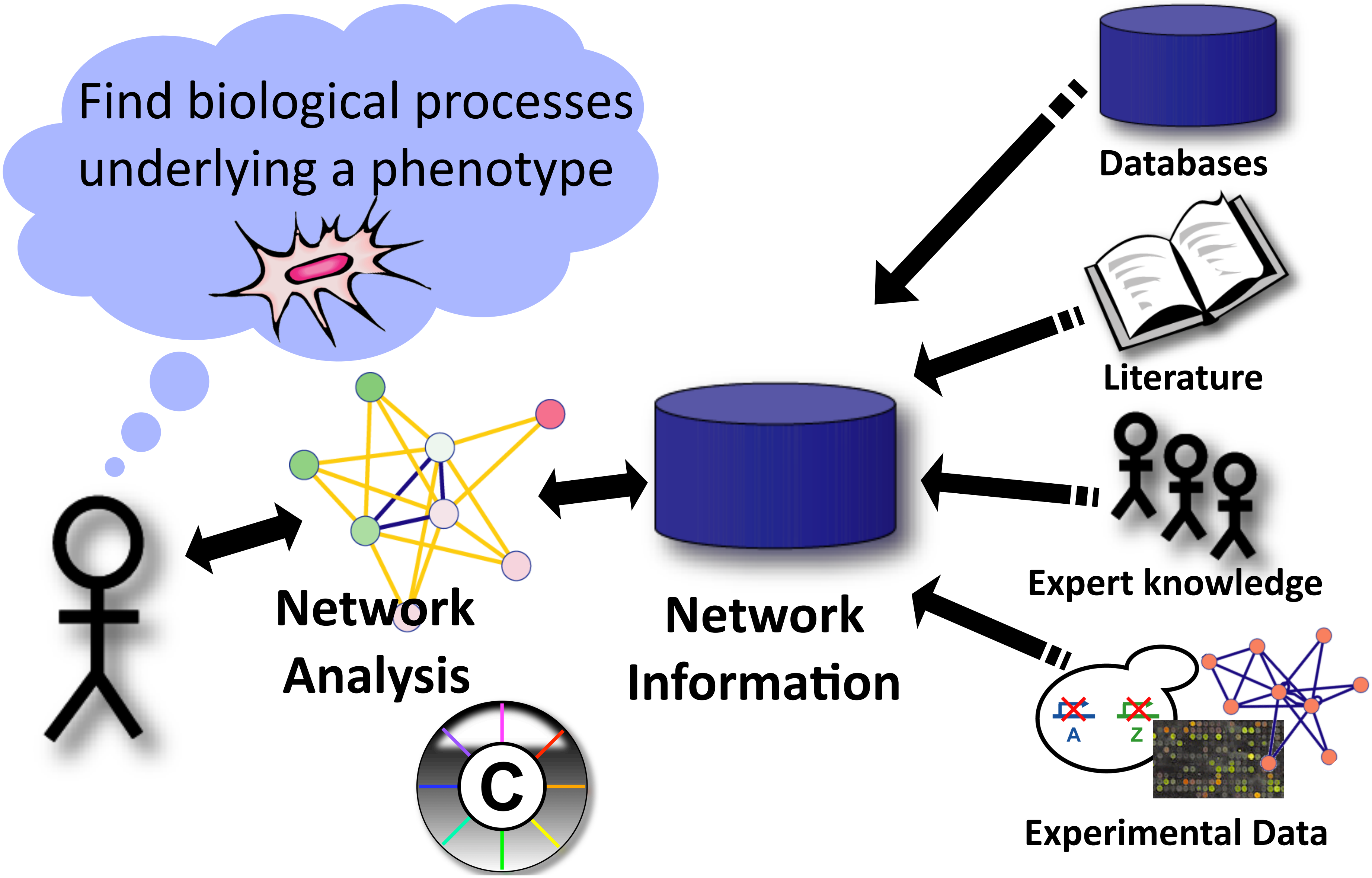


Network	Nodes	Edges
galFiltered.sif	331(19)	362(35)

ID	gal1RGexp	gal1RGsig	gal4RGexp	gal4RGsig	gal80Rexp	gal80I
YGL008C	-0.352	1.0007E-5	-0.282	7.1366E-4	-0.573	1.2622E
YCL067C	0.169	0.0012873	-0.085	0.11481	0.301	0.0027
YNL145W	-0.764	3.148E-11	-0.098	0.05338	-1.237	1.1916E
YMR043W	-0.183	0.0035372	-0.654	4.2514E-6	0.457	2.4112E

Welcome to Cytoscape 2.4.0-b1
Right-click + drag to ZOOM
Middle-click + drag to PAN

Network Analysis using Cytoscape



Find biological processes underlying a phenotype

Network Analysis

Network Information

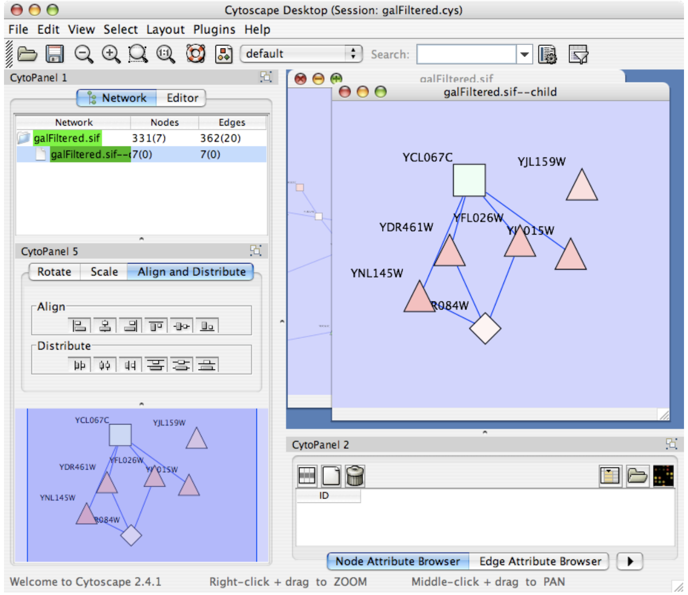
Databases

Literature

Expert knowledge

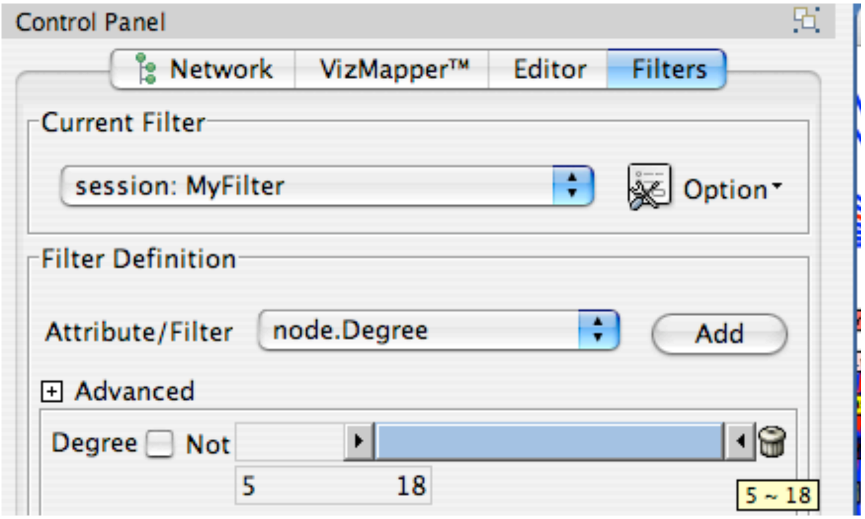
Experimental Data

Manipulate Networks



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

Filter/Query



Current Filter: session: MyFilter

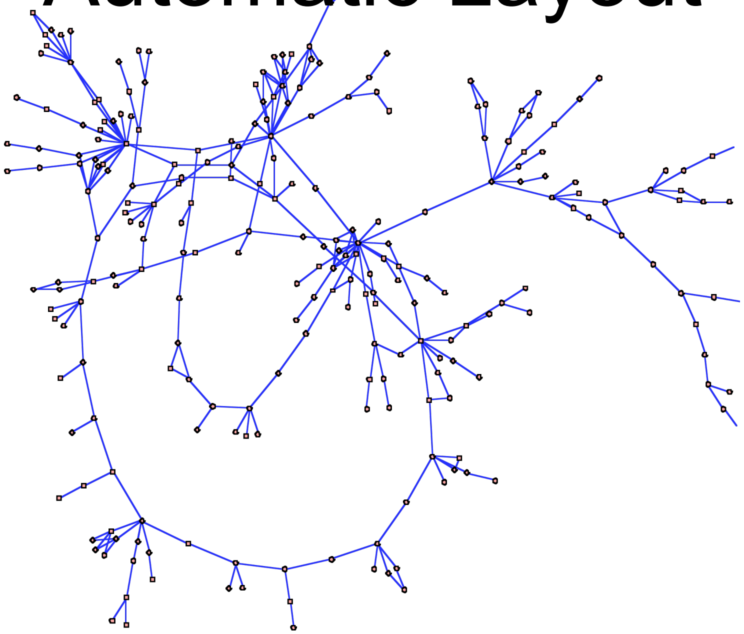
Attribute/Filter: node.Degree

Degree ☐ Not

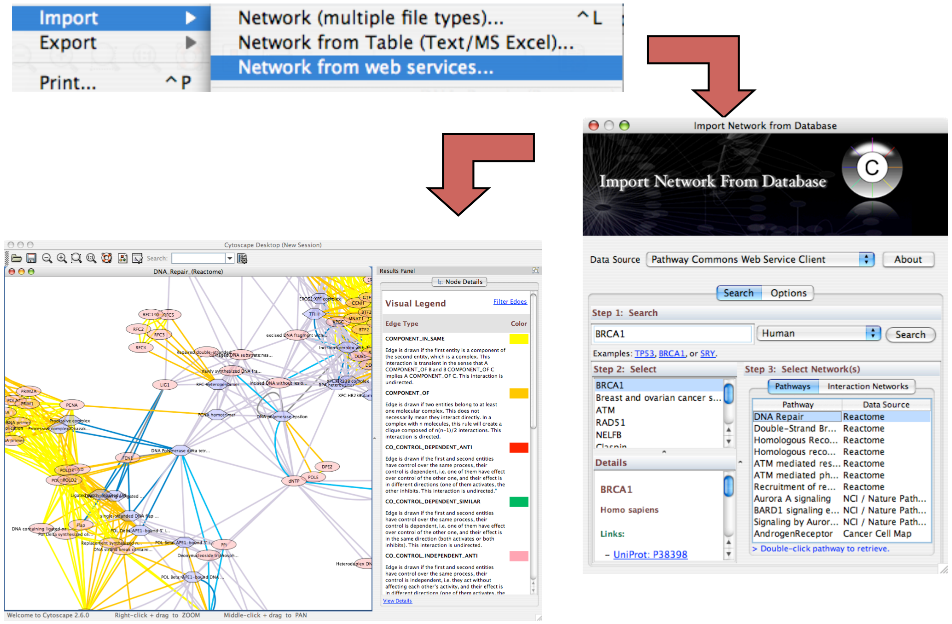
5 18

5 ~ 18

Automatic Layout



Interaction Database Search



Data Source: Pathway Commons Web Service Client

Step 1: Search

BRCA1 Human

Examples: TP53, BRCA1, or SRY.

Step 2: Select

BRCA1
Breast and ovarian cancer s...
ATM
RAD51
NELFB

Details

BRCA1

Homo sapiens

Links:

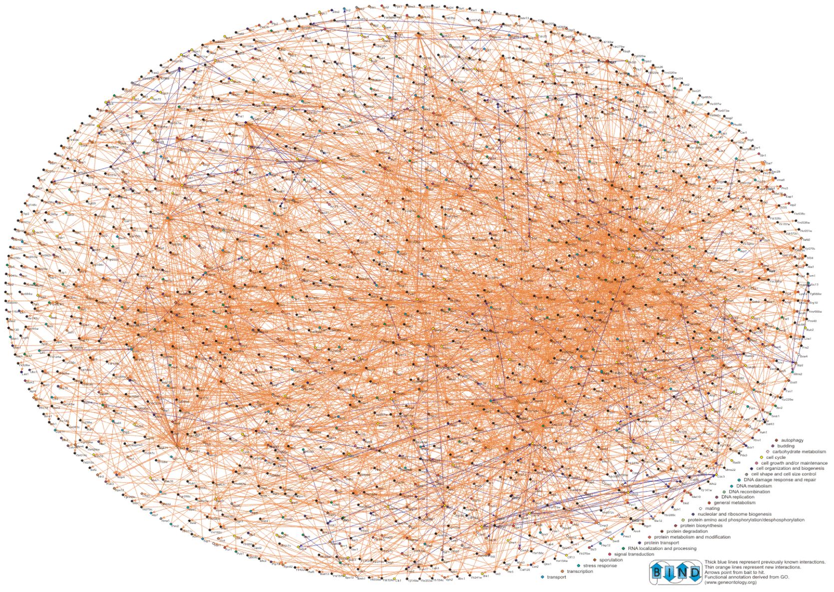
– UniProt: P38398

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

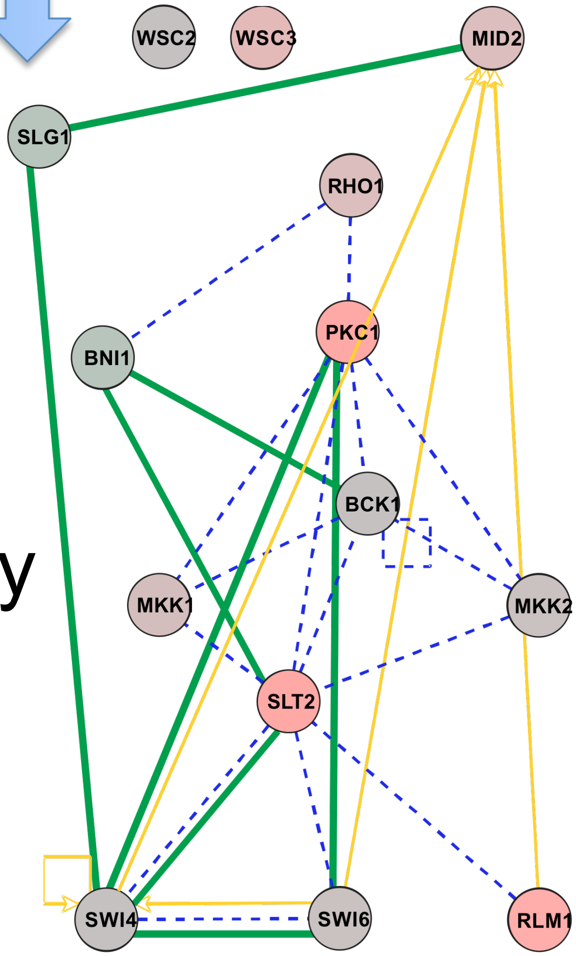
Overview

Systematic identification of protein complexes in
Saccharomyces cerevisiae by mass spectrometry



Zoom

Focus



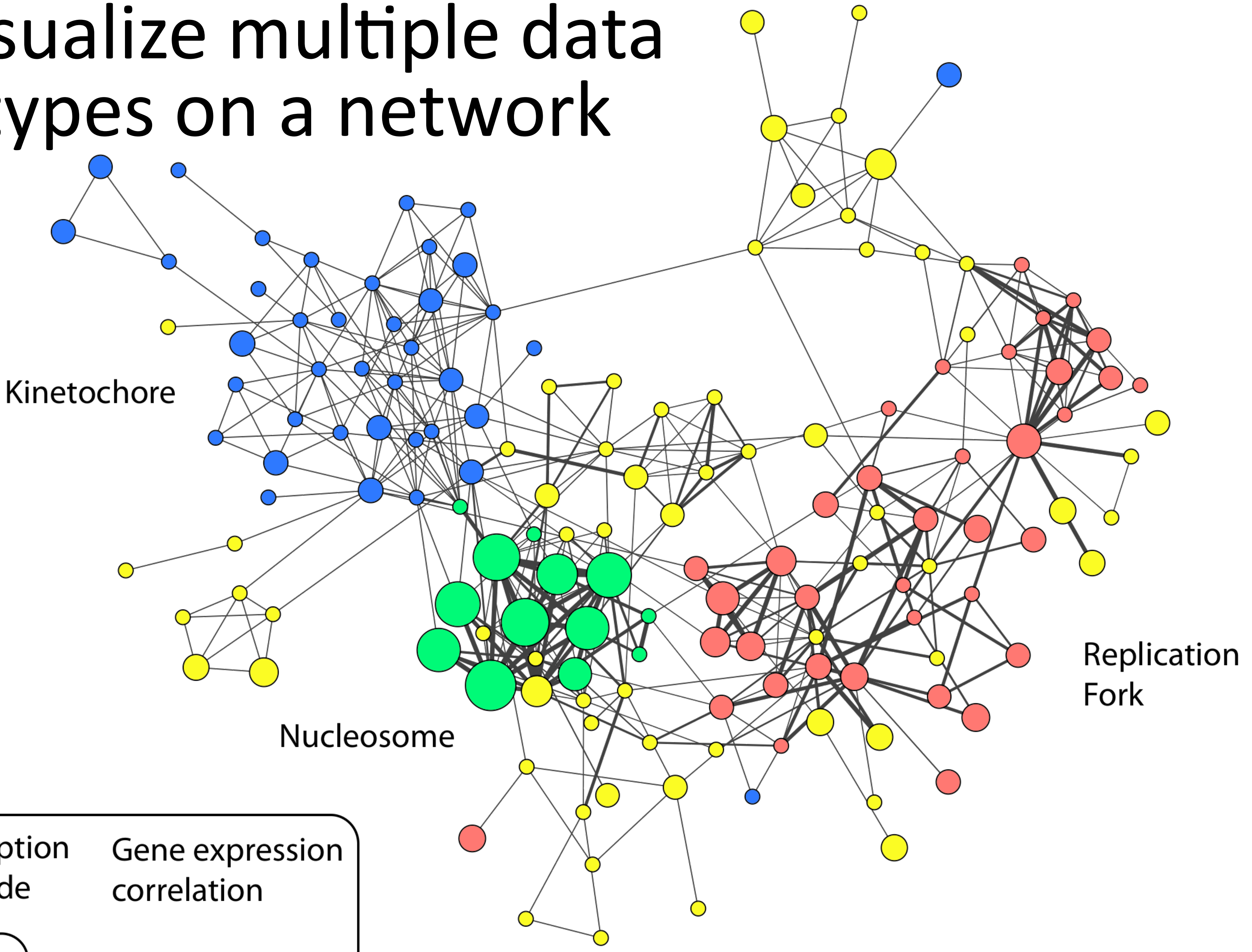
PKC
Cell
Wall
Integrity

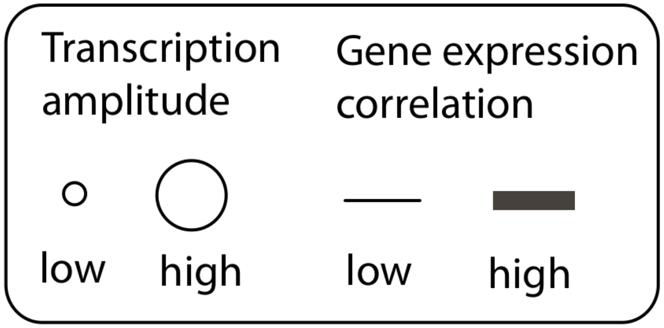
━━━ **Synthetic Lethal**
━━━ **Transcription Factor Regulation**
▬ ▬ ▬ **Protein-Protein Interaction**

● **Up Regulated Gene Expression**

● **Down Regulated Gene Expression**

Visualize multiple data types on a network





Control: node/edge size, shape, color...

Active Community

http://www.cytoscape.org

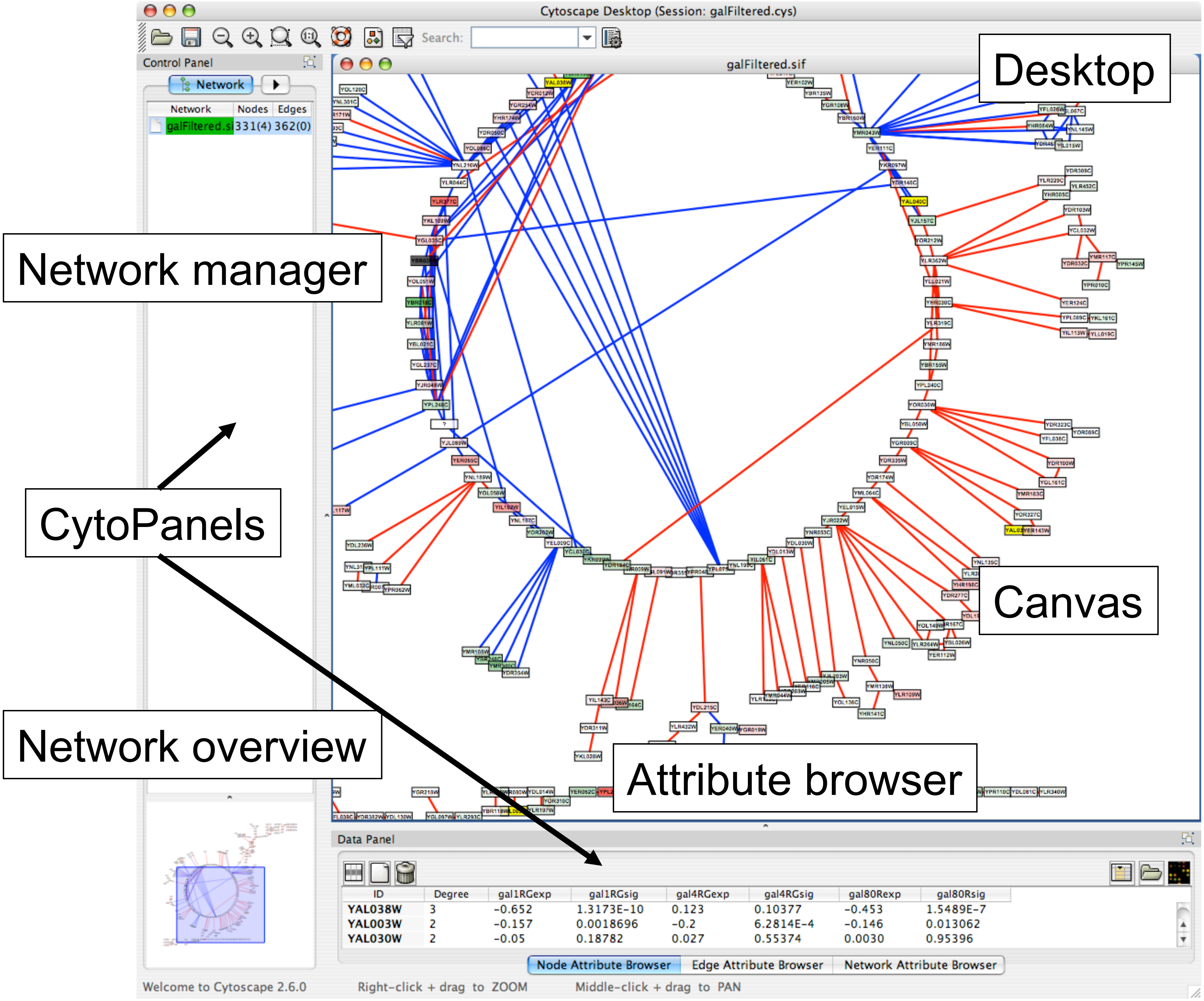
- Help
  - 8 tutorials, >10 case studies
  - Mailing lists for discussion
  - Documentation, data sets

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

- Annual Conference: Houston Nov 6-9, 2009
- 10,000s users, 2500 downloads/month
- >40 Plugins Extend Functionality
  - Build your own, requires programming
    - e.g. Retina Workbench

Demo

- Desktop
- Layout
- Subnetworks
- Visual styles
- Filtering/Searching
- Getting data
- Plugins: Lit search, Active Modules, Bingo, MCL



Desktop

Network manager

CytoPanels

Canvas

Network overview

Attribute browser

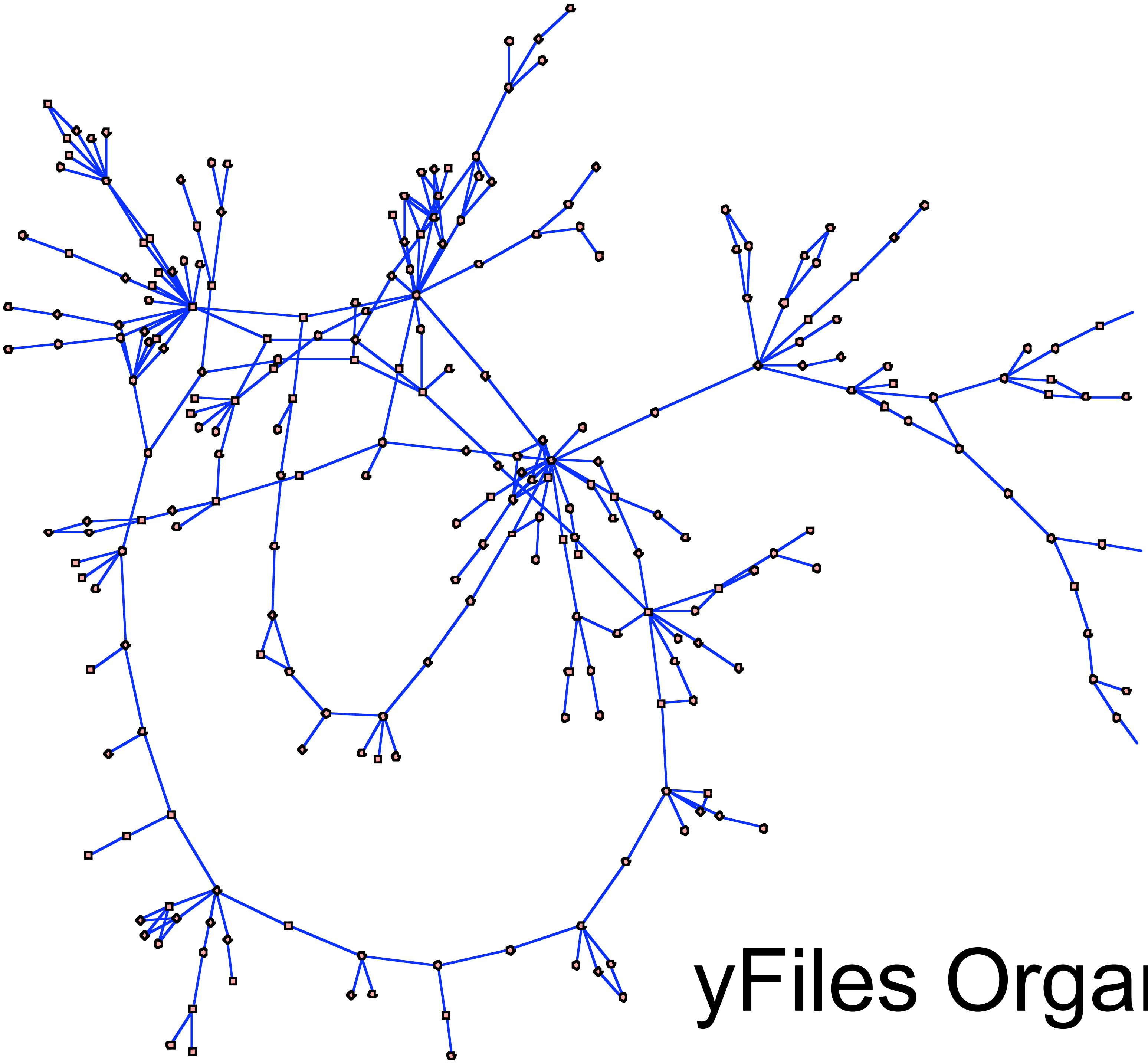
Network	Nodes	Edges
galFiltered.si	331(4)	362(0)

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

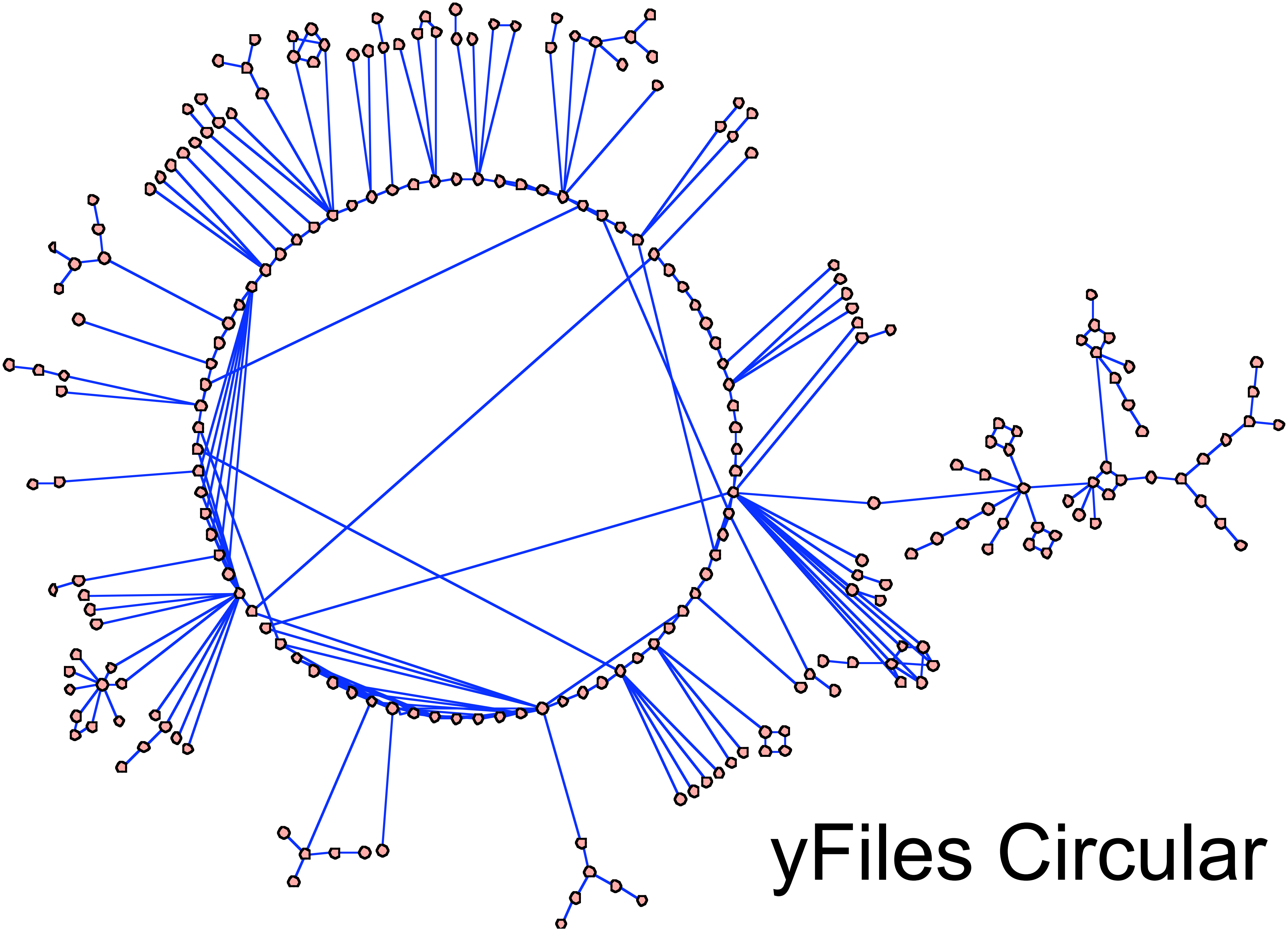
Welcome to Cytoscape 2.6.0

Right-click + drag to ZOOM

Middle-click + drag to PAN



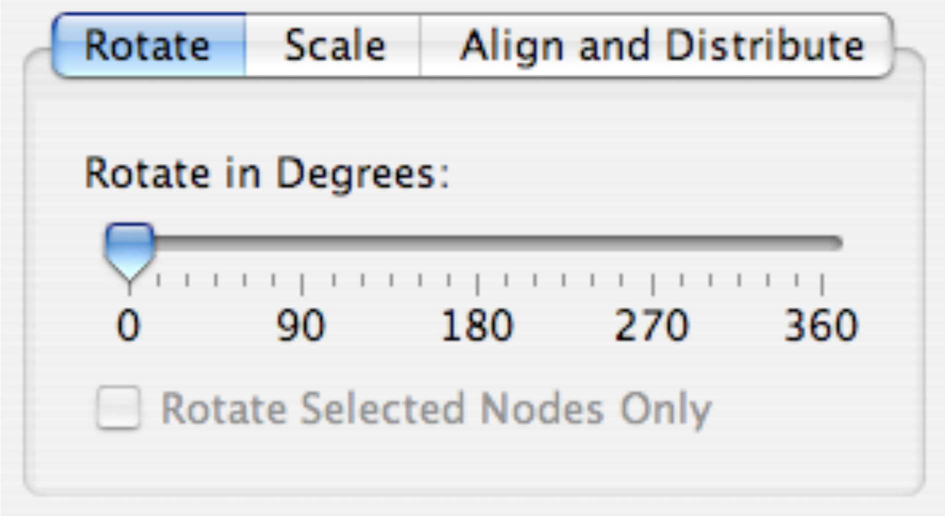
yFiles Organic

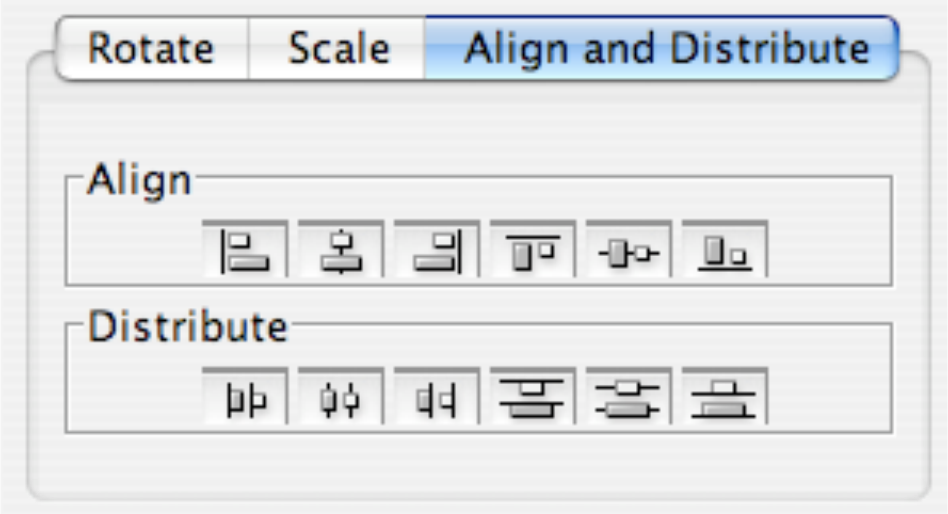


yFiles Circular

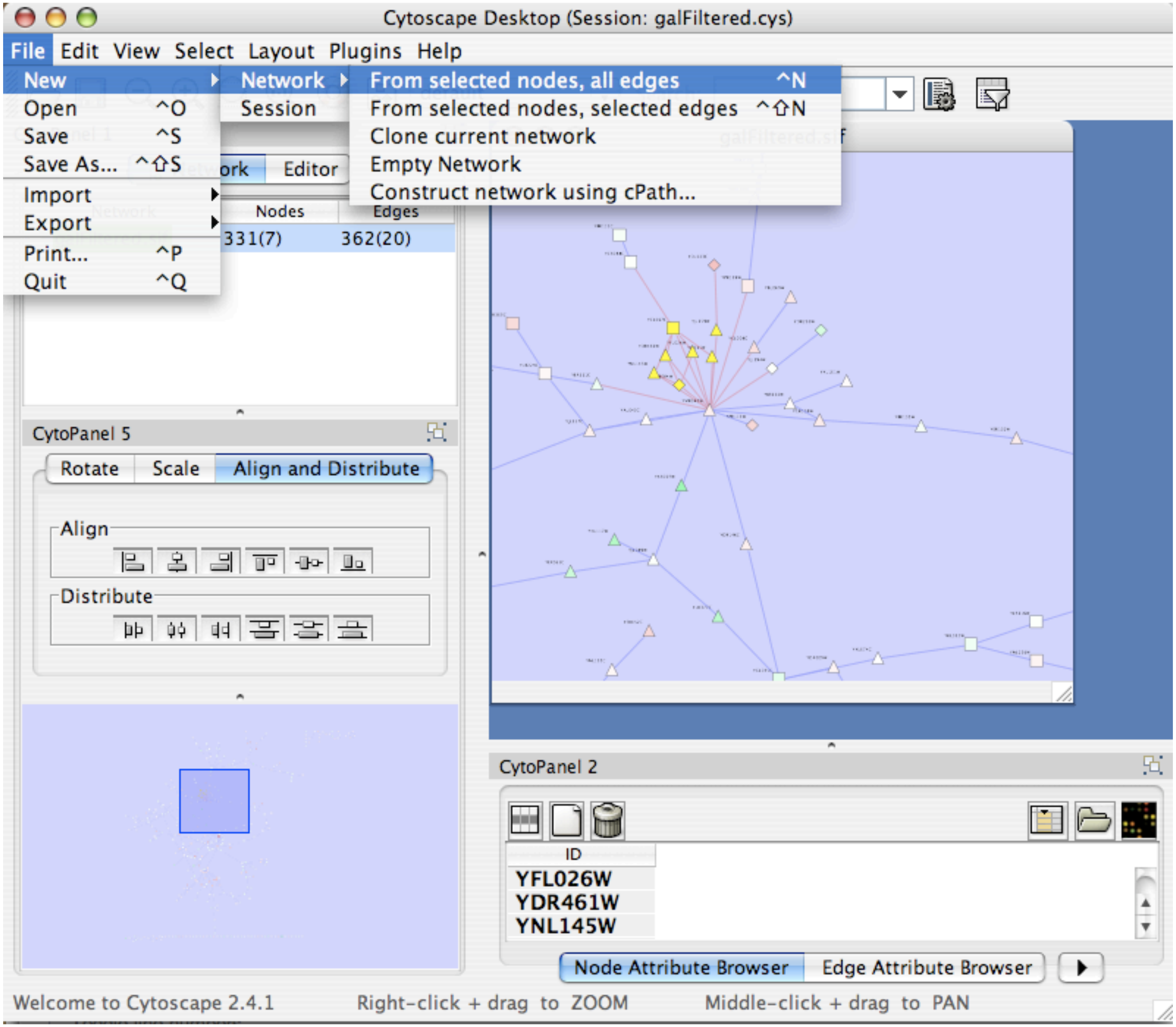
Network Layout

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



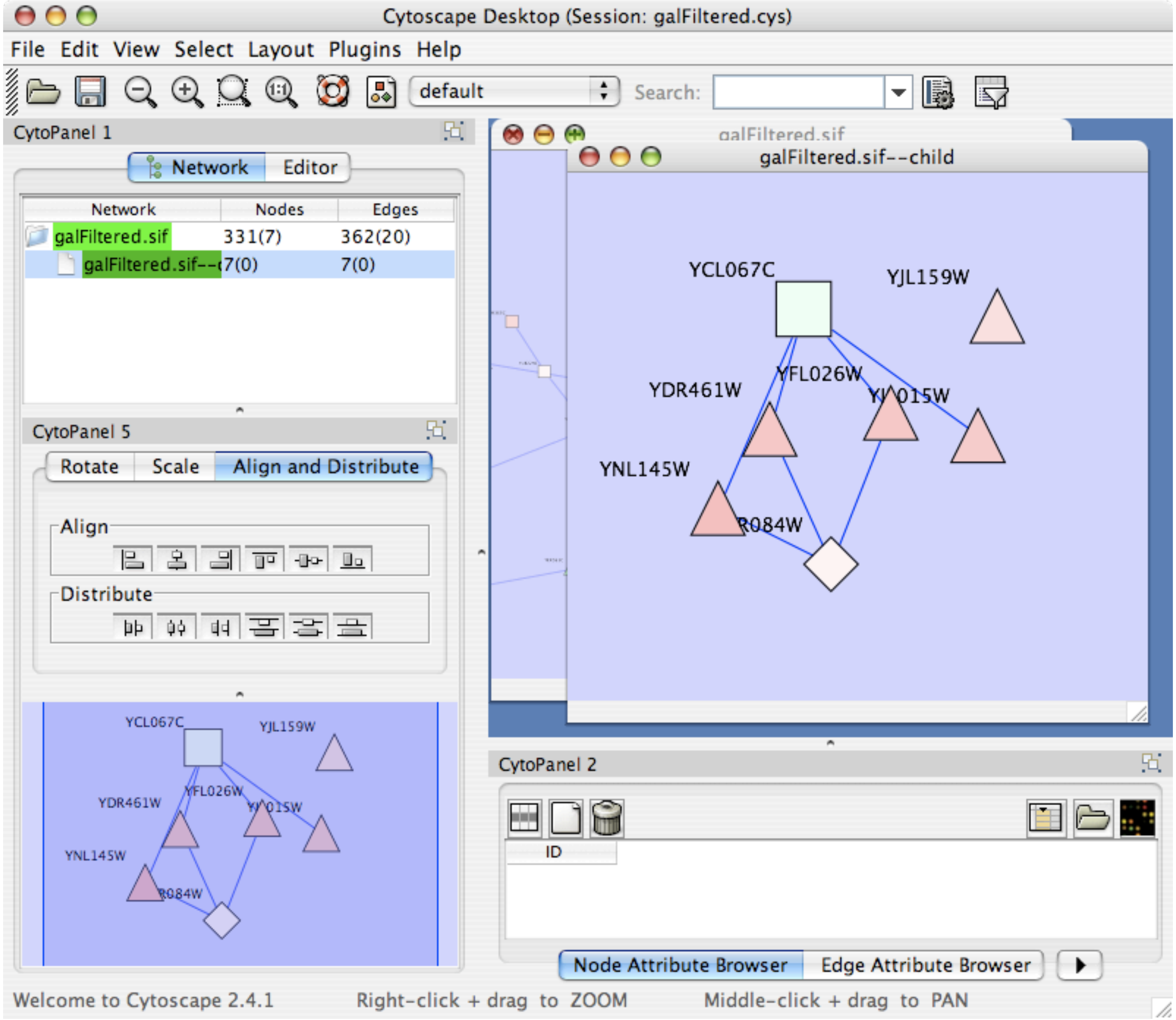


Create subnetwork



ID
YFL026W
YDR461W
YNL145W

Create subnetwork



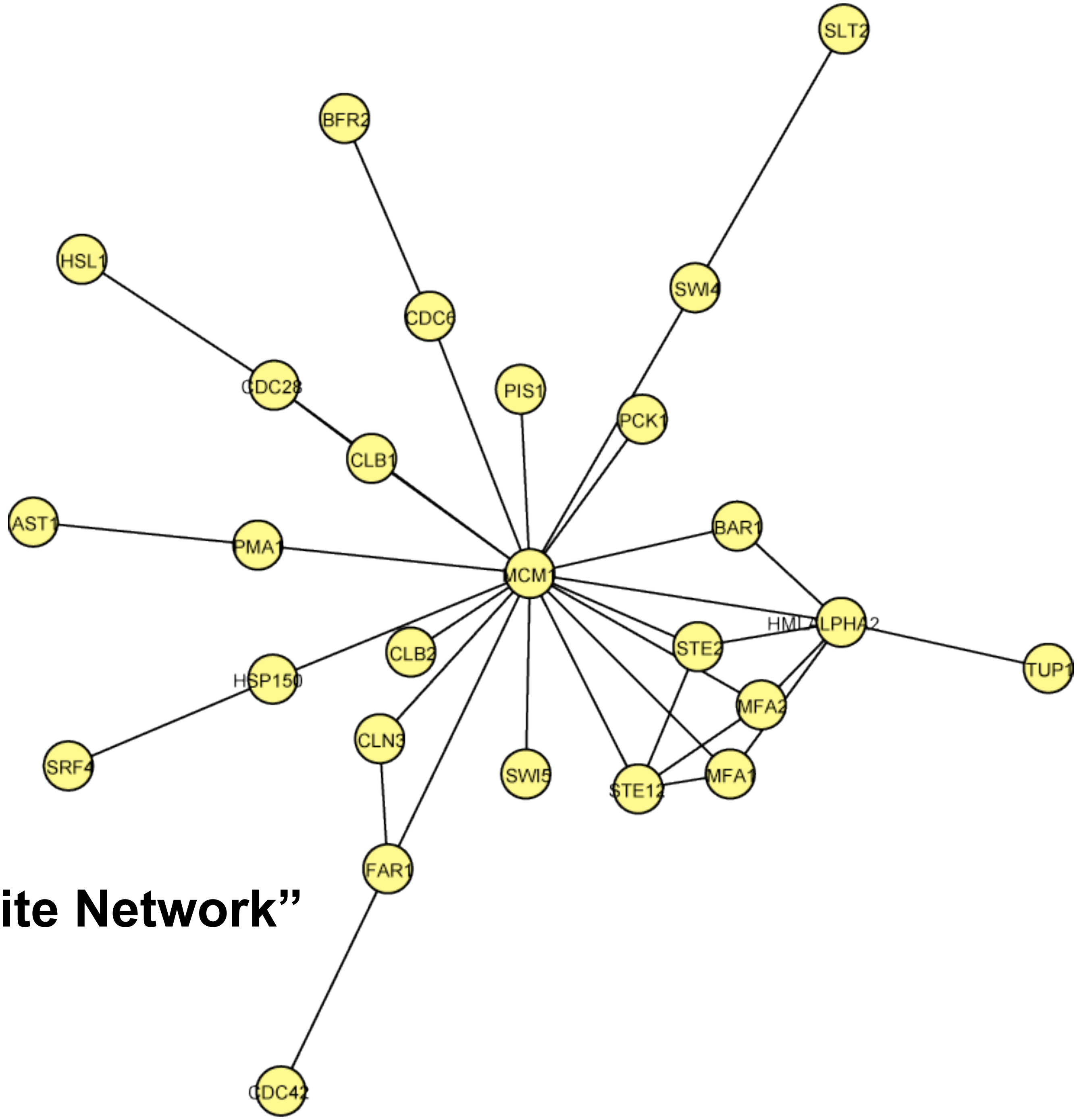
Network	Nodes	Edges
galFiltered.sif	331(7)	362(20)
galFiltered.sif--(	7(0)	7(0)

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

Visual Style

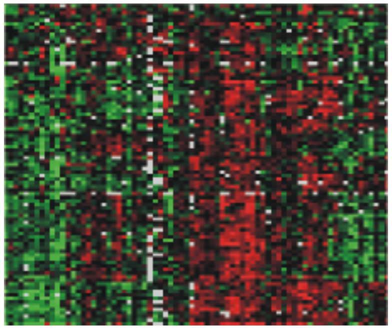
- Customized views of experimental data in a network context
- Network has node and edge attributes
    - E.g. expression data, GO function, interaction type
- Mapped to visual attributes
    - E.g. node/edge size, shape, colour…
- E.g. Visualize gene expression data as node colour gradient on the network

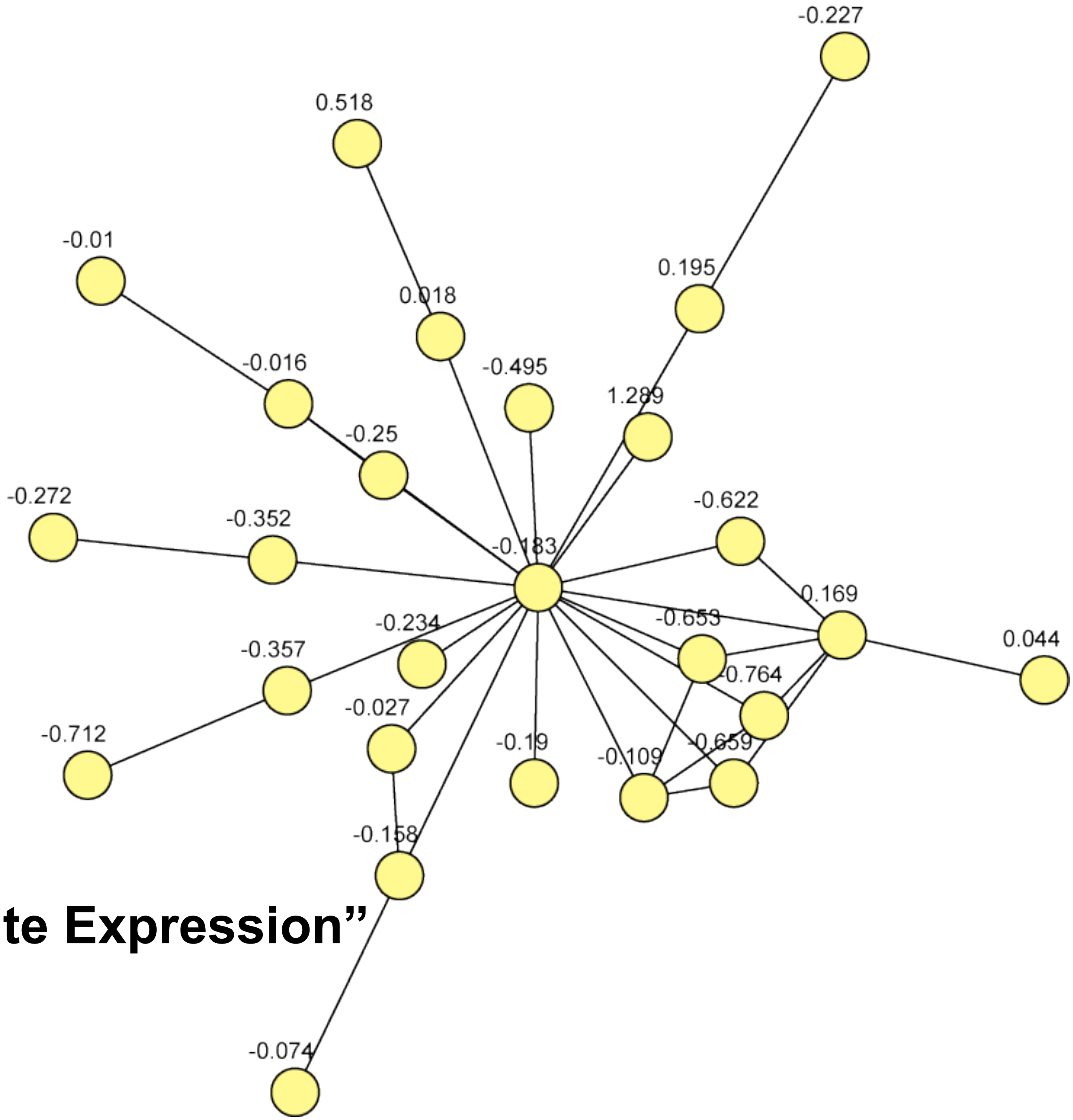
Visual Style



Load “Your Favorite Network”

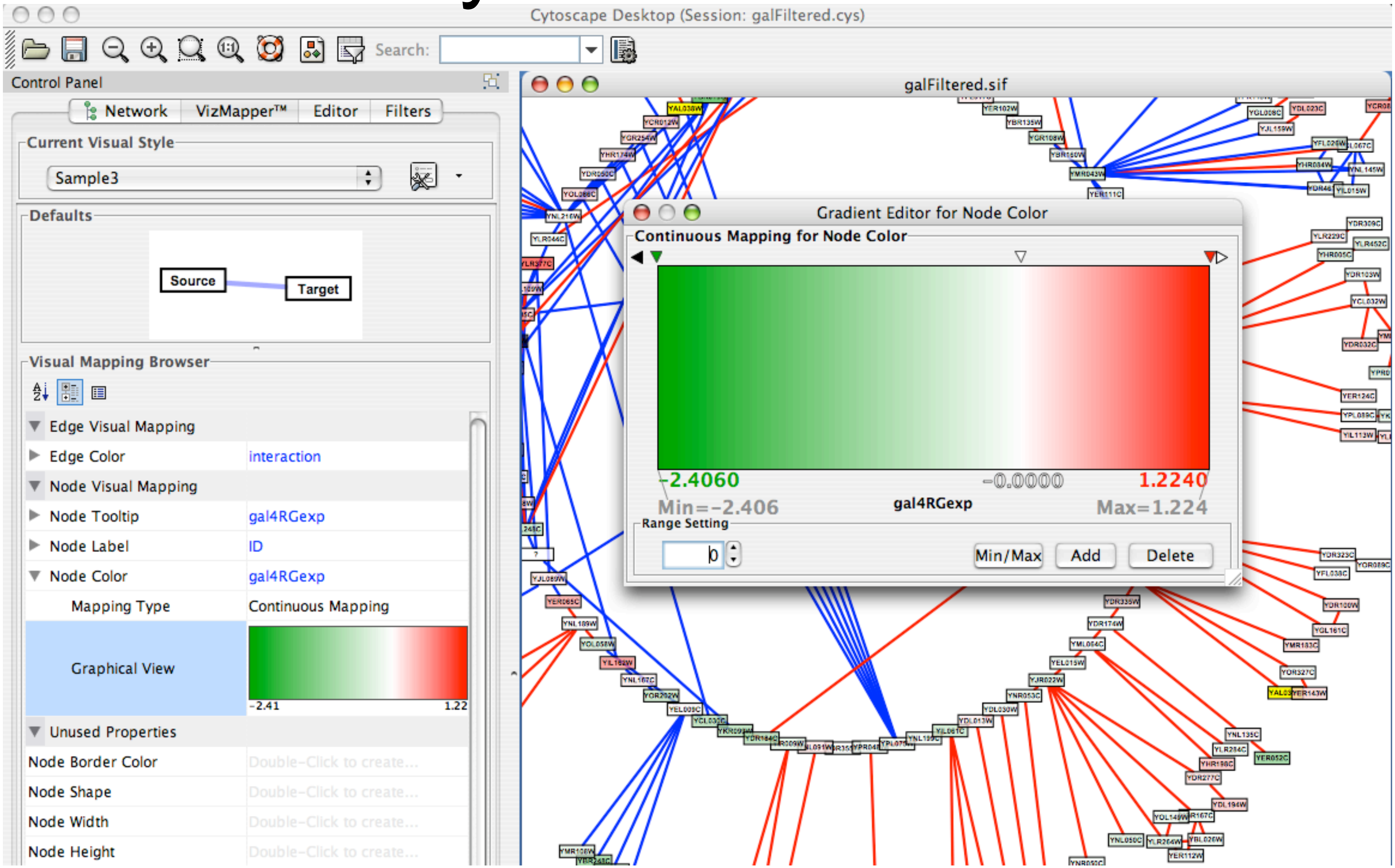
Visual Style





Load "Your Favorite Expression" Dataset

Visual Style



Cytoscape Desktop (Session: galFiltered.cys)

Current Visual Style: Sample3

Edge Color: interaction

Node Tooltip: gal4RGexp

Node Label: ID

Node Color: gal4RGexp

Mapping Type: Continuous Mapping

Gradient Editor for Node Color

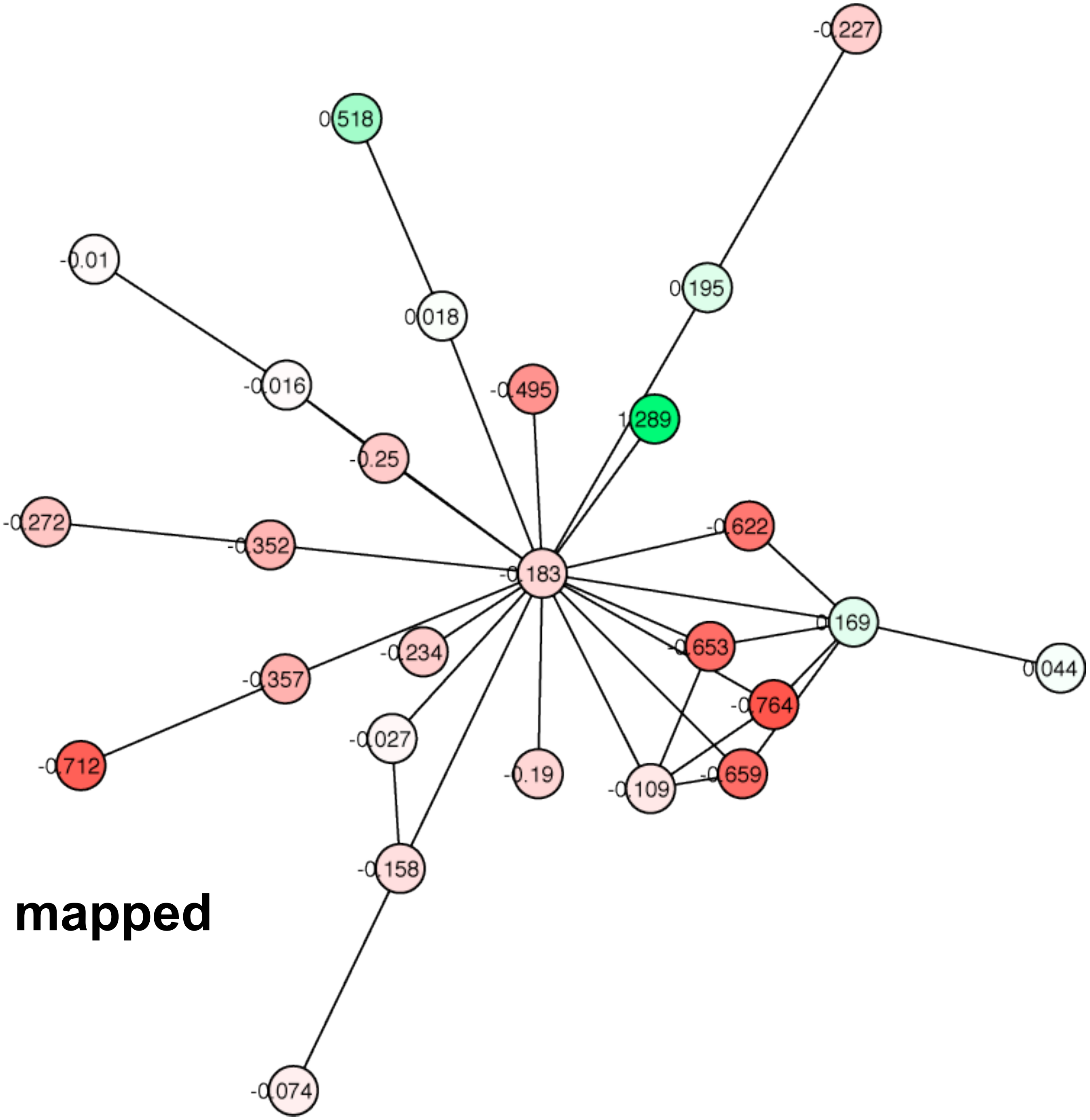
Continuous Mapping for Node Color

-2.4060 -0.0000 1.2240

Min=-2.406 gal4RGexp Max=1.224

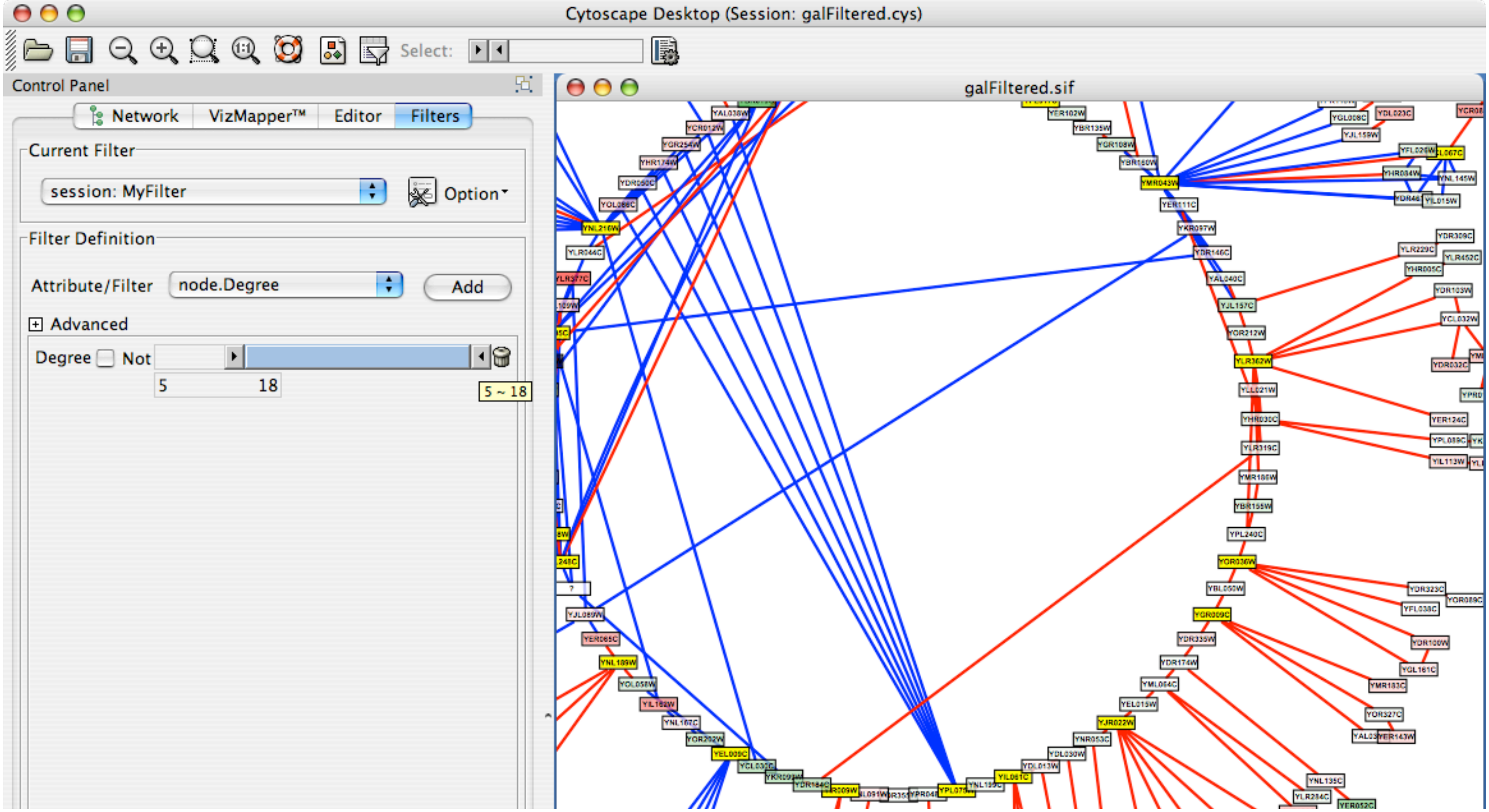
Map expression values to node colours using a continuous mapper

Visual Style



Expression data mapped to node colours

Network Filtering



session: MyFilter

node.Degree

Degree ☐ Not

5 18

5 ~ 18

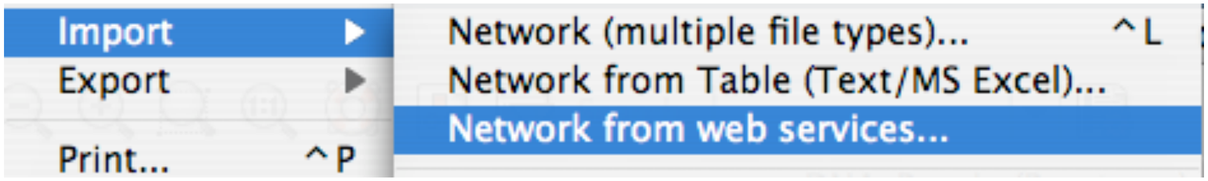
Demo

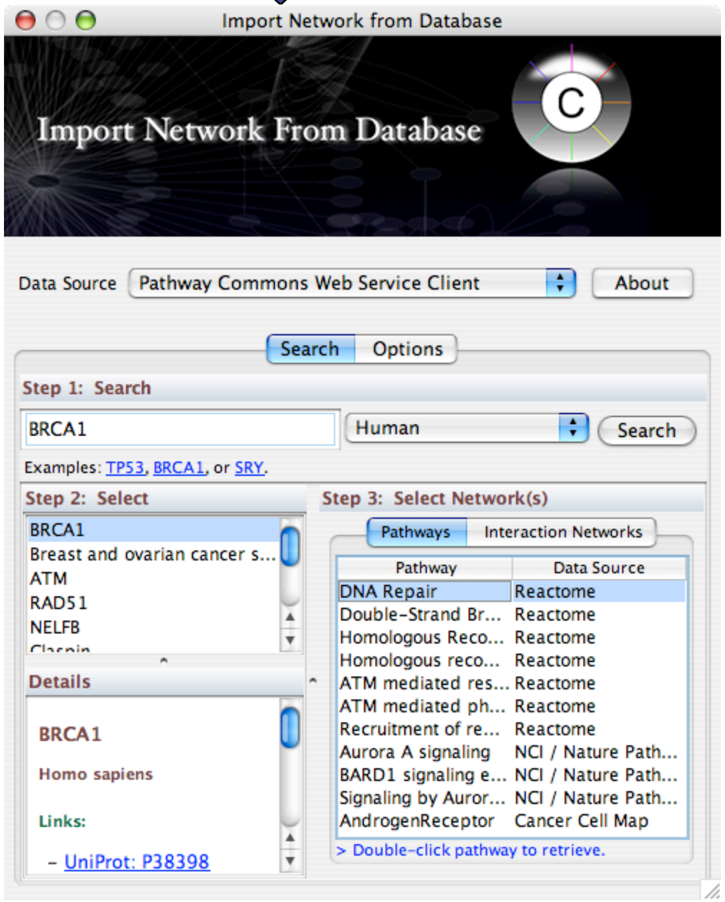
- Desktop
- Layout
- Subnetworks
- Visual styles
- Filtering/Searching
- Getting data
- Plugins: Lit search, Active Modules, Bingo, MCL

Analyzing gene expression data in a network context

- Input
  - Gene expression data
  - Network data
- Output
  - Visual diagram of expression data on network
  - Active network regions
- Outline
  - Where to find network data?
    - Interaction database (cPath)
    - Literature associations via text mining
  - Load expression data
  - Identify active pathways

Interaction Database Search





Data Source: Pathway Commons Web Service Client

Step 1: Search

BRCA1 — Human

Examples: TP53, BRCA1, or SRY.

Step 2: Select

BRCA1
Breast and ovarian cancer s...
ATM
RAD51
NELFB
Claspin

Details

BRCA1

Homo sapiens

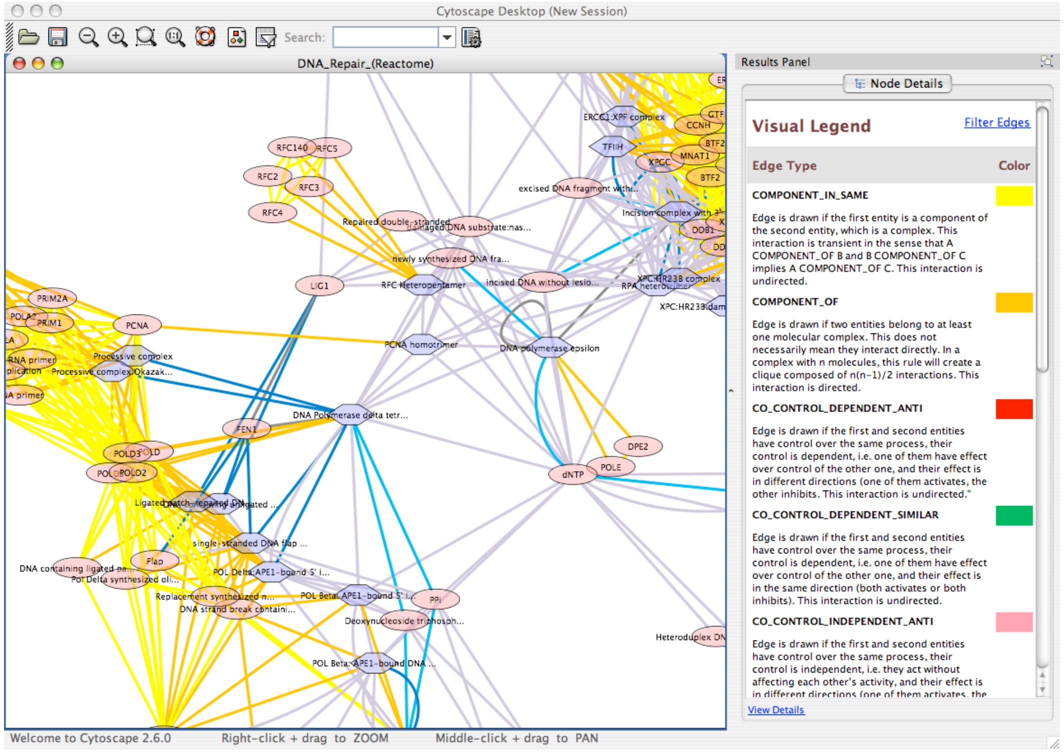
Links:

- UniProt: P38398

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

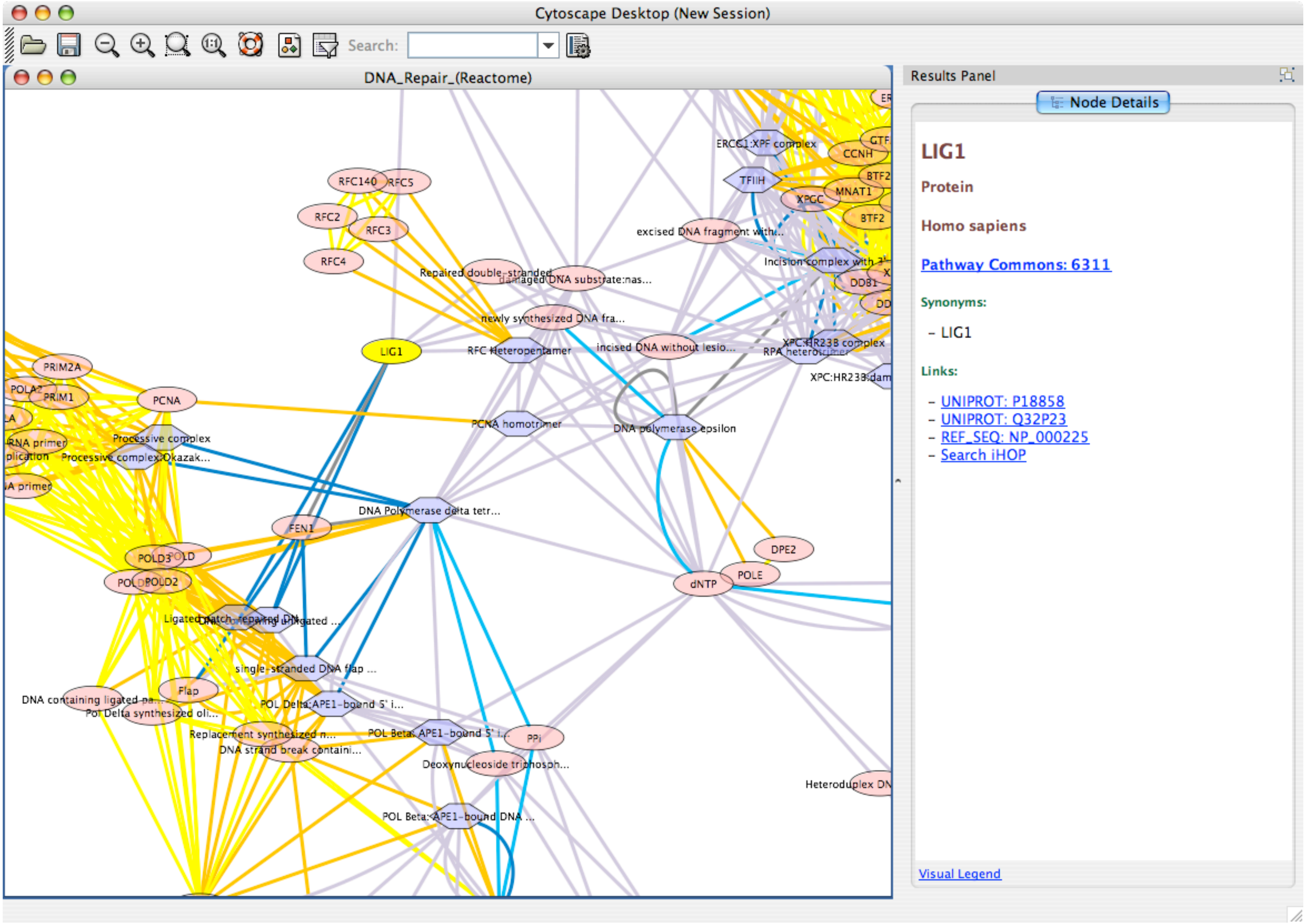
> Double-click pathway to retrieve.



Visual Legend — Filter Edges

Edge Type	Color
COMPONENT_IN_SAME Edge is drawn if the first entity is a component of the second entity, which is a complex. This interaction is transient in the sense that A COMPONENT_OF B and B COMPONENT_OF C implies A COMPONENT_OF C. This interaction is undirected.	🟨
COMPONENT_OF Edge is drawn if two entities belong to at least one molecular complex. This does not necessarily mean they interact directly. In a complex with n molecules, this rule will create a clique composed of n(n-1)/2 interactions. This interaction is directed.	🟧
CO_CONTROL_DEPENDENT_ANTI Edge is drawn if the first and second entities have control over the same process, their control is dependent, i.e. one of them have effect over control of the other one, and their effect is in different directions (one of them activates, the other inhibits. This interaction is undirected."	🟥
CO_CONTROL_DEPENDENT_SIMILAR Edge is drawn if the first and second entities have control over the same process, their control is dependent, i.e. one of them have effect over control of the other one, and their effect is in the same direction (both activates or both inhibits). This interaction is undirected.	🟩
CO_CONTROL_INDEPENDENT_ANTI Edge is drawn if the first and second entities have control over the same process, their control is independent, i.e. they act without affecting each other's activity, and their effect is in different directions (one of them activates, the	▨

View Details



LIG1

Protein

Homo sapiens

Pathway Commons: 6311

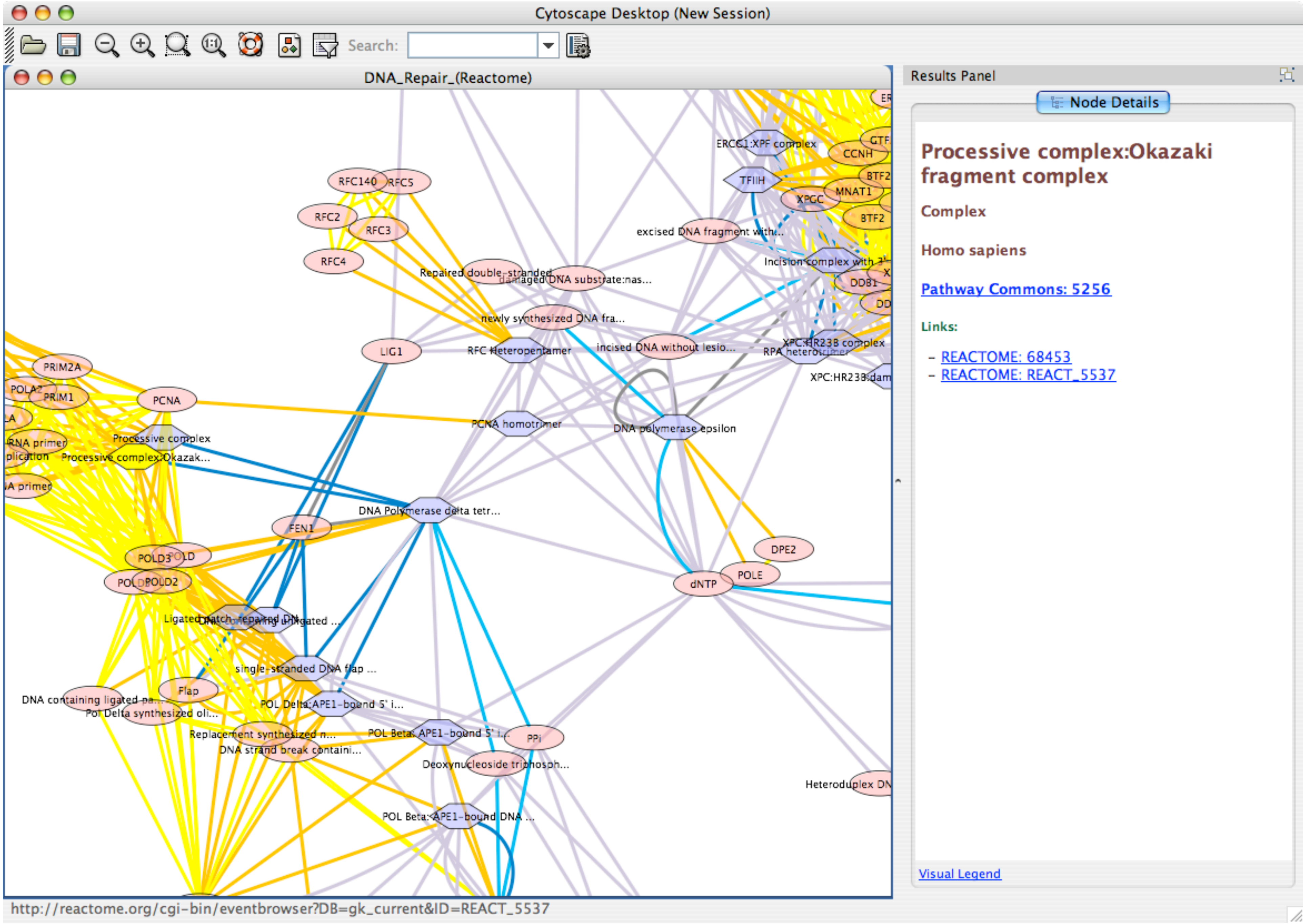
Synonyms:

- 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

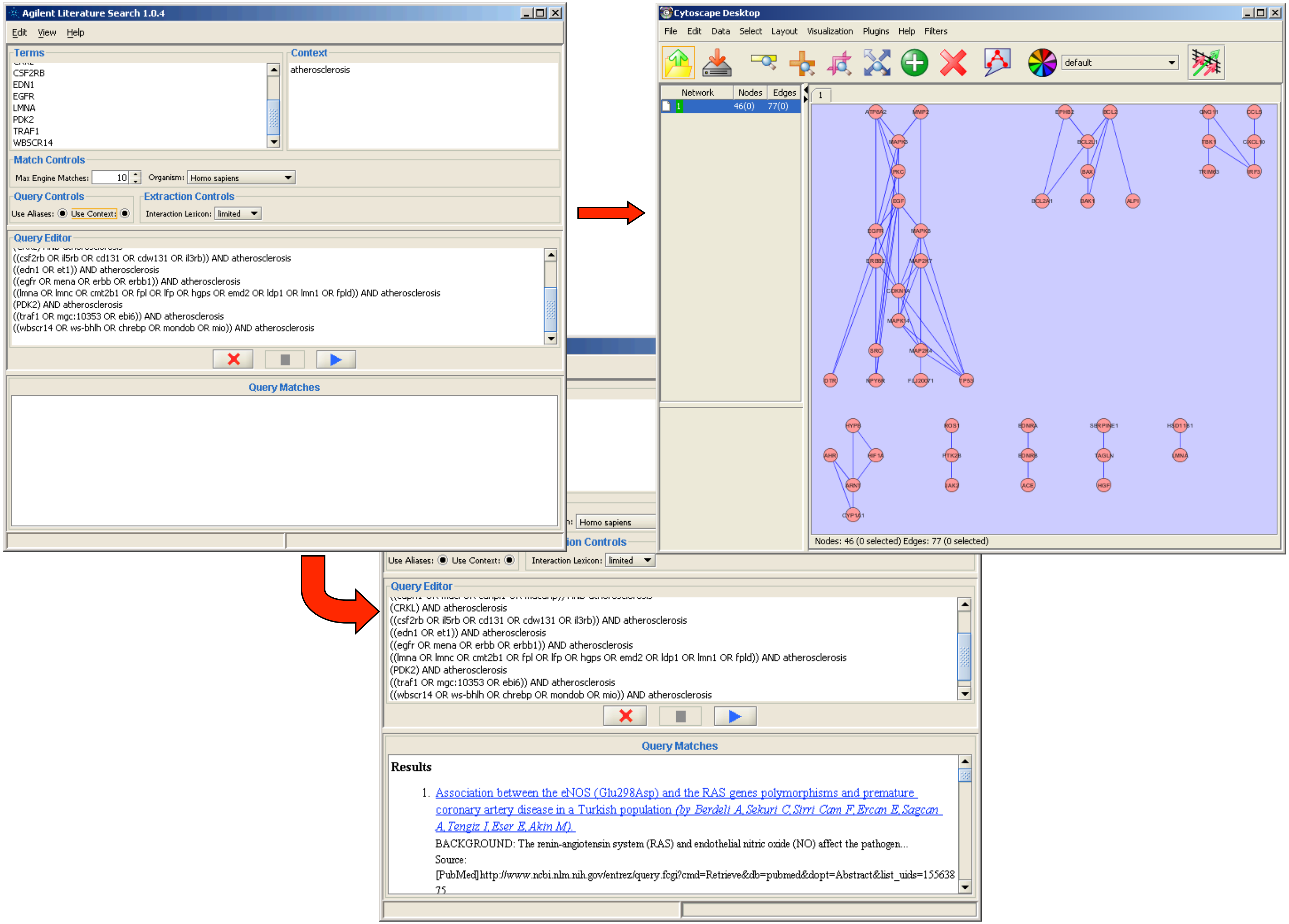
Links:

- REACTOME: 68453
- REACTOME: REACT_5537

http://reactome.org/cgi-bin/eventbrowser?DB=gk_current&ID=REACT_5537

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/



Terms:
CSF2RB
EDN1
EGFR
LMNA
PDK2
TRAF1
WBSCR14

Context: atherosclerosis

Max Engine Matches: 10

Organism: Homo sapiens

Use Aliases: ● Use Context: ●

Interaction Lexicon: limited

```
((csf2rb OR il5rb OR cd131 OR cdw131 OR il3rb)) AND atherosclerosis
((edn1 OR et1)) AND atherosclerosis
((egfr OR mena OR erbb OR erbb1)) AND atherosclerosis
((lmna OR lmnc OR cmt2b1 OR fpl OR lfp 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
```

Network 1 — Nodes 46(0) — Edges 77(0)

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

Organism: Homo sapiens

Use Aliases: ● Use Context: ●

Interaction Lexicon: limited

```
(CRKL) AND atherosclerosis
((csf2rb OR il5rb OR cd131 OR cdw131 OR il3rb)) AND atherosclerosis
((edn1 OR et1)) AND atherosclerosis
((egfr OR mena OR erbb OR erbb1)) AND atherosclerosis
((lmna OR lmnc OR cmt2b1 OR fpl OR lfp 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
```

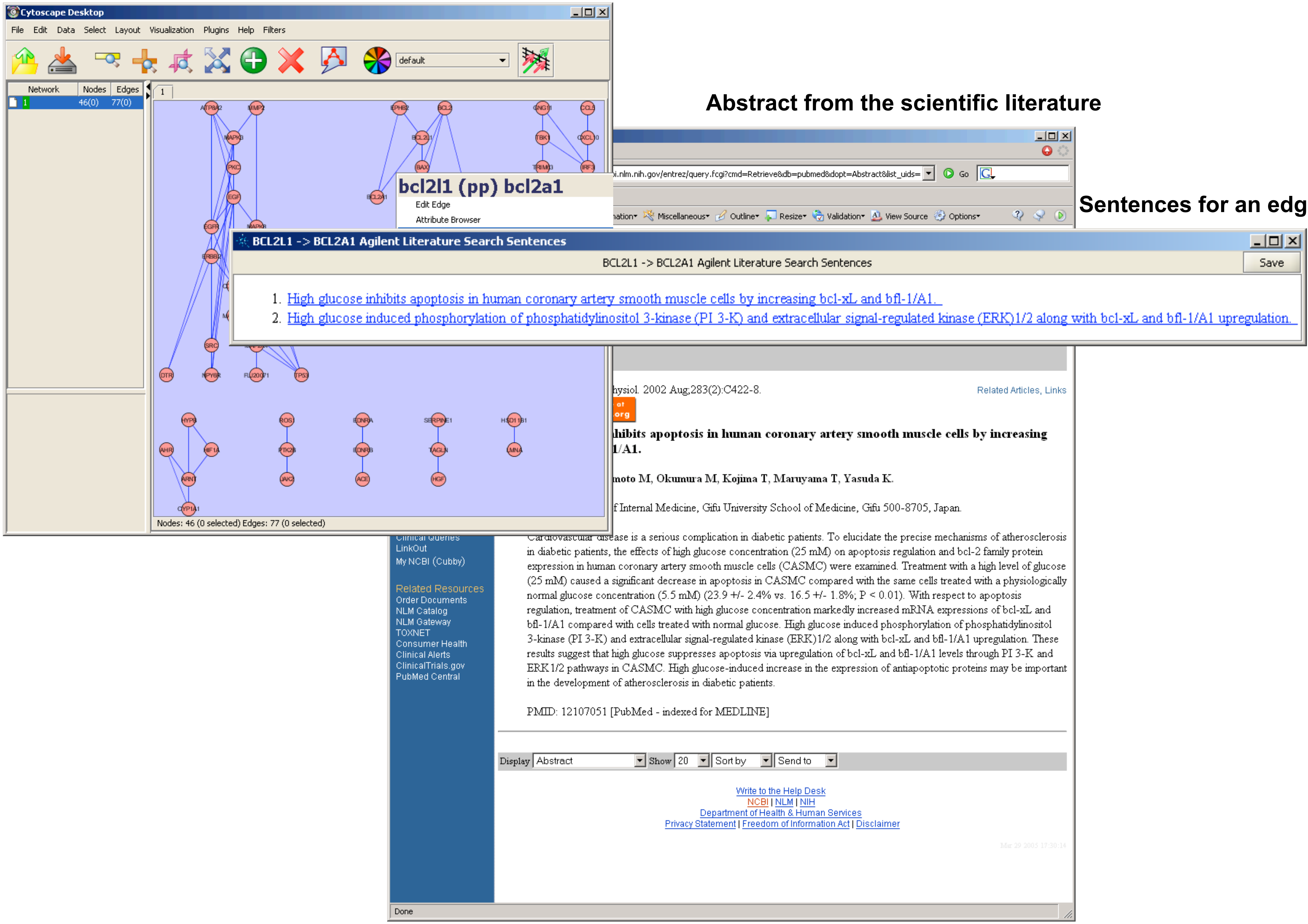
Results

1. 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 Cam F, Ercan E, Sagcan A, Tengiz I, Eser E, Akin 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

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.

hysiol. 2002 Aug;283(2):C422-8.

hibits apoptosis in human coronary artery smooth muscle cells by increasing 1/A1.

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

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

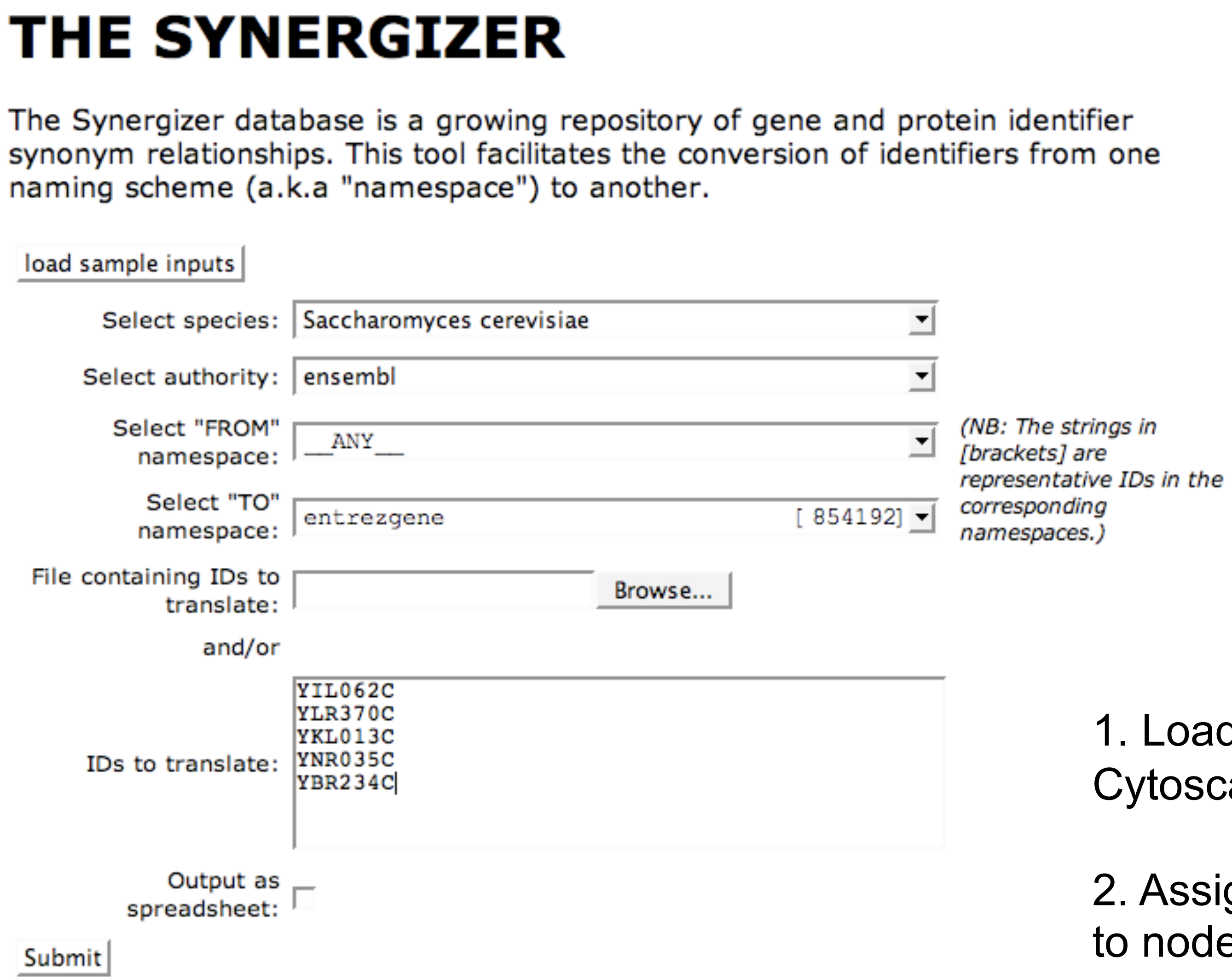
Cardiovascular disease is a serious complication in diabetic patients. To elucidate the precise mechanisms of atherosclerosis in diabetic patients, the effects of high glucose concentration (25 mM) on apoptosis regulation and bcl-2 family protein expression in human coronary artery smooth muscle cells (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 ERK1/2 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]

Gene Expression/Network Integration

- Identifier (ID) mapping
  - Translation from network IDs to gene expression IDs e.g. Affymetrix probe IDs
  - Also: Unification, link out, query
  - Entrez gene IDs (genes), UniProt (proteins)
- Synergizer
  - llama.med.harvard.edu/cgi/synergizer/translate
- More ID mapping services available
  - http://baderlab.org/IdentifierMapping

Gene Expression/Network Integration



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.

Select species: Saccharomyces cerevisiae

Select authority: ensembl

Select "FROM" namespace: __ANY__

Select "TO" namespace: entrezgene [854192]

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

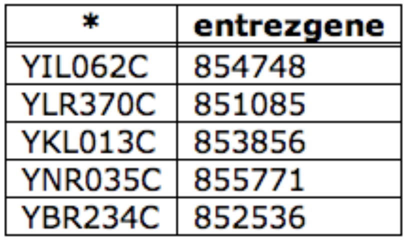
File containing IDs to translate:

and/or

IDs to translate:

```
YIL062C
YLR370C
YKL013C
YNR035C
YBR234C
```

Output as spreadsheet: ☐

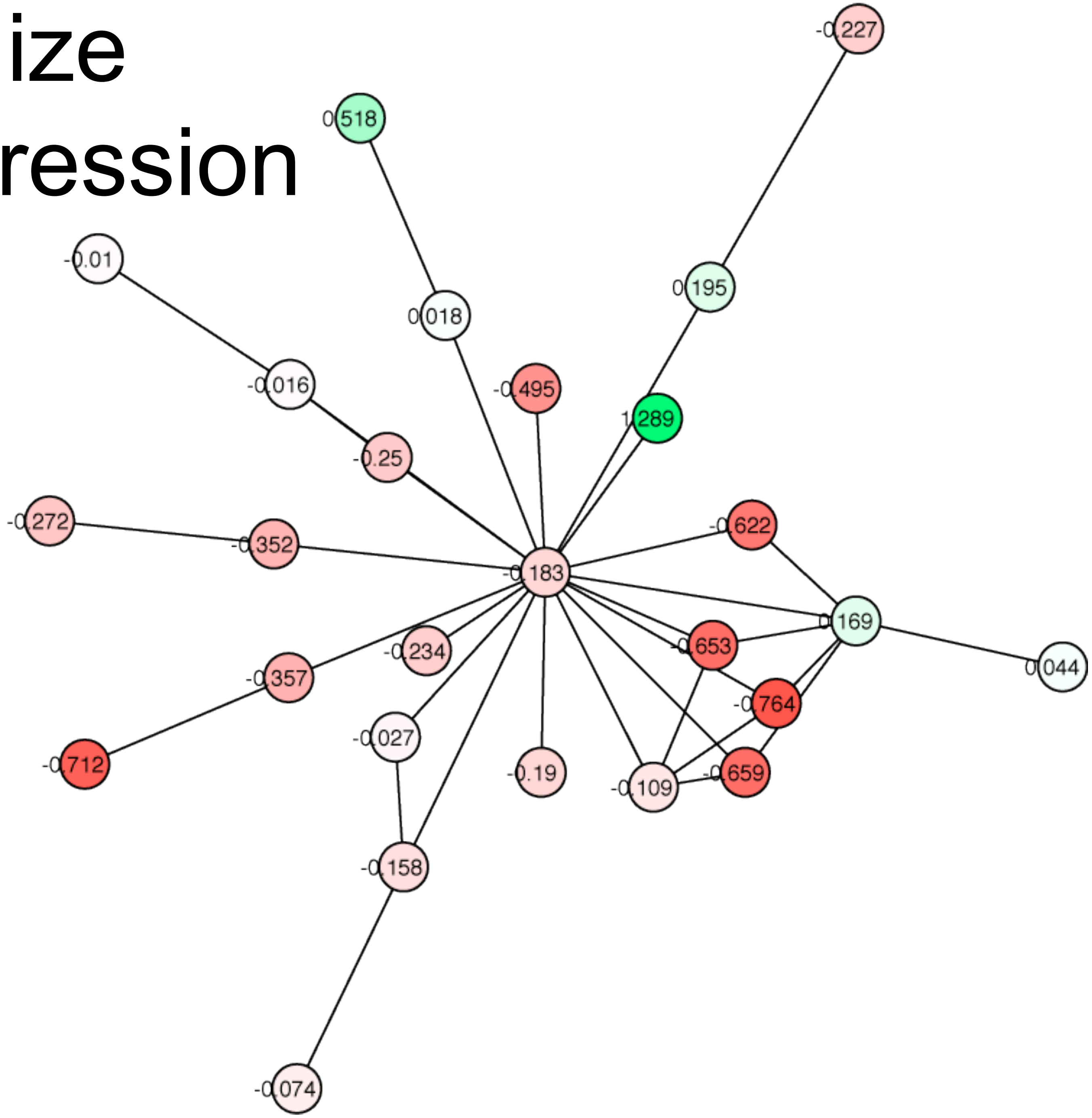


*	entrezgene
YIL062C	854748
YLR370C	851085
YKL013C	853856
YNR035C	855771
YBR234C	852536

1. Load as attributes in Cytoscape

2. Assign expression values to nodes using this attribute set

Visualize Gene Expression



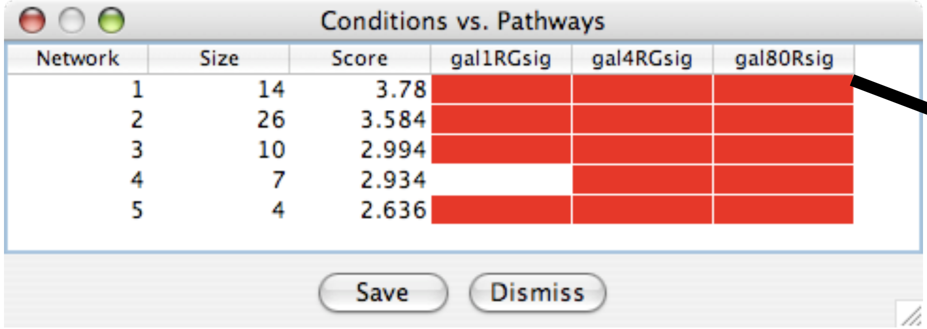
Find Active Subnetworks

- 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

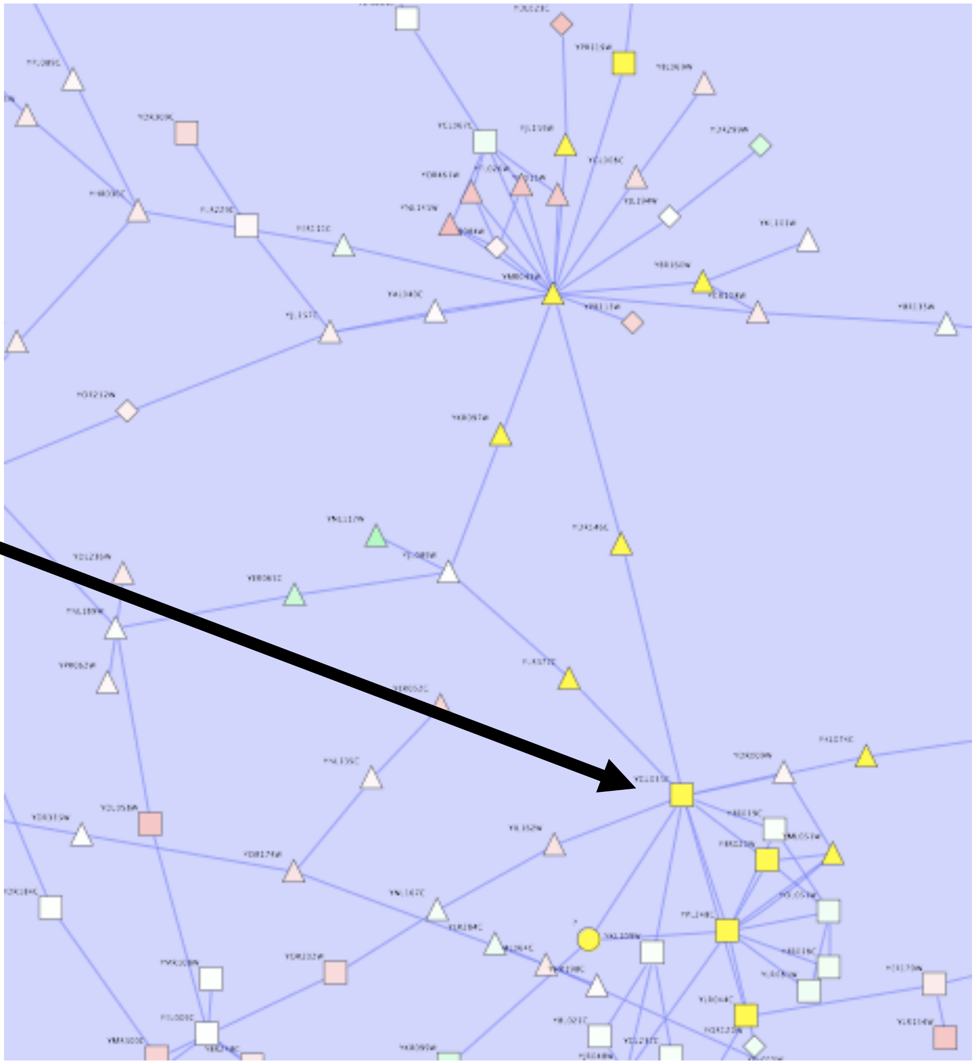
Active Module Results

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	🟥	🟥	🟥

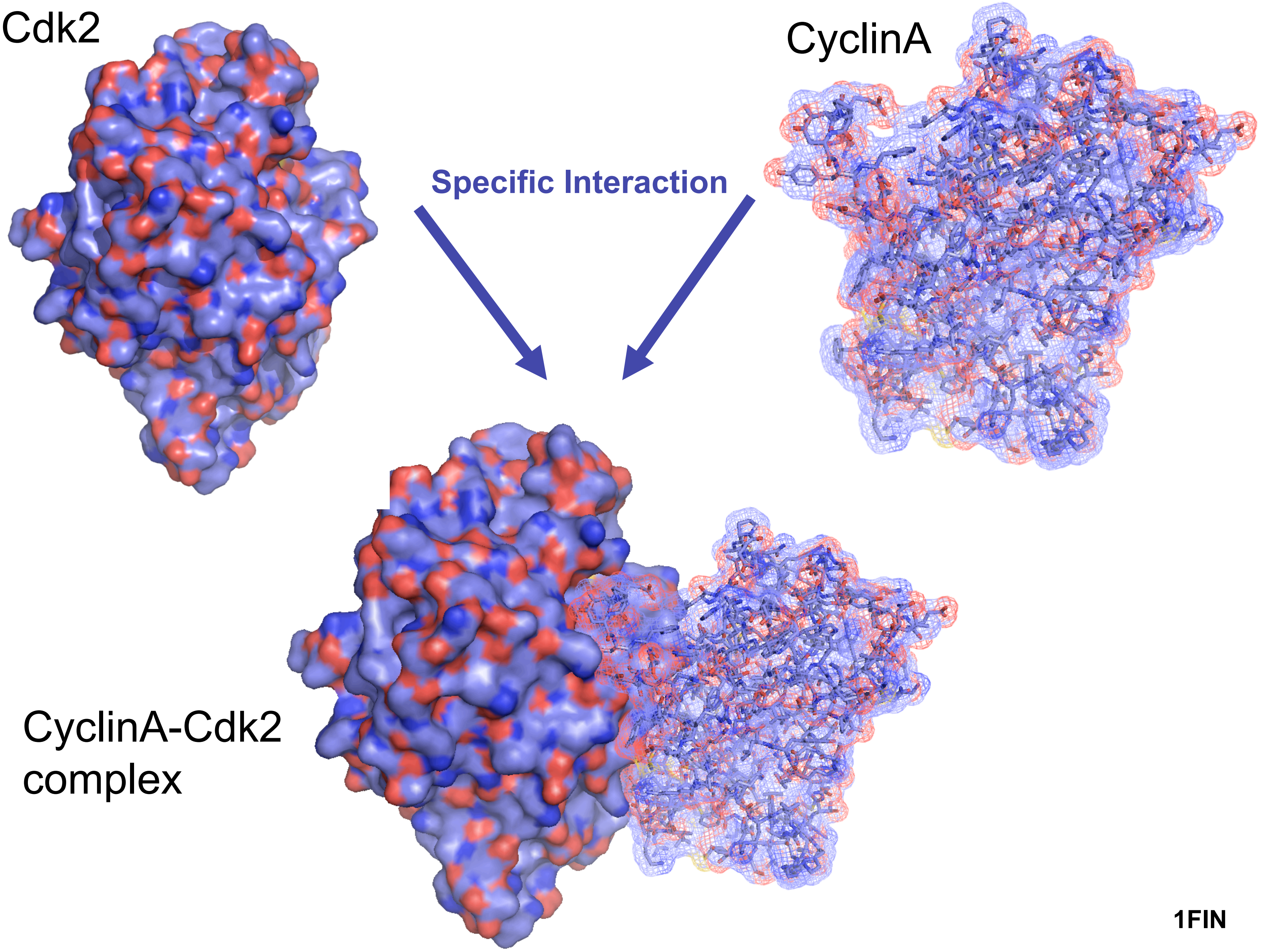


Note: non-deterministic, multiple runs required for confidence of result robustness

Ideker T et al. Science. 2001 May 4;292(5518):929-34.

Additional Plugins

- Bingo: over-representation analysis
- ClusterMaker: clusters networks, includes MCL
- NetworkAnalyzer: calculates statistics about a network
- (You may have to use an earlier version of Cytoscape to get some plugins to run)



Cdk2

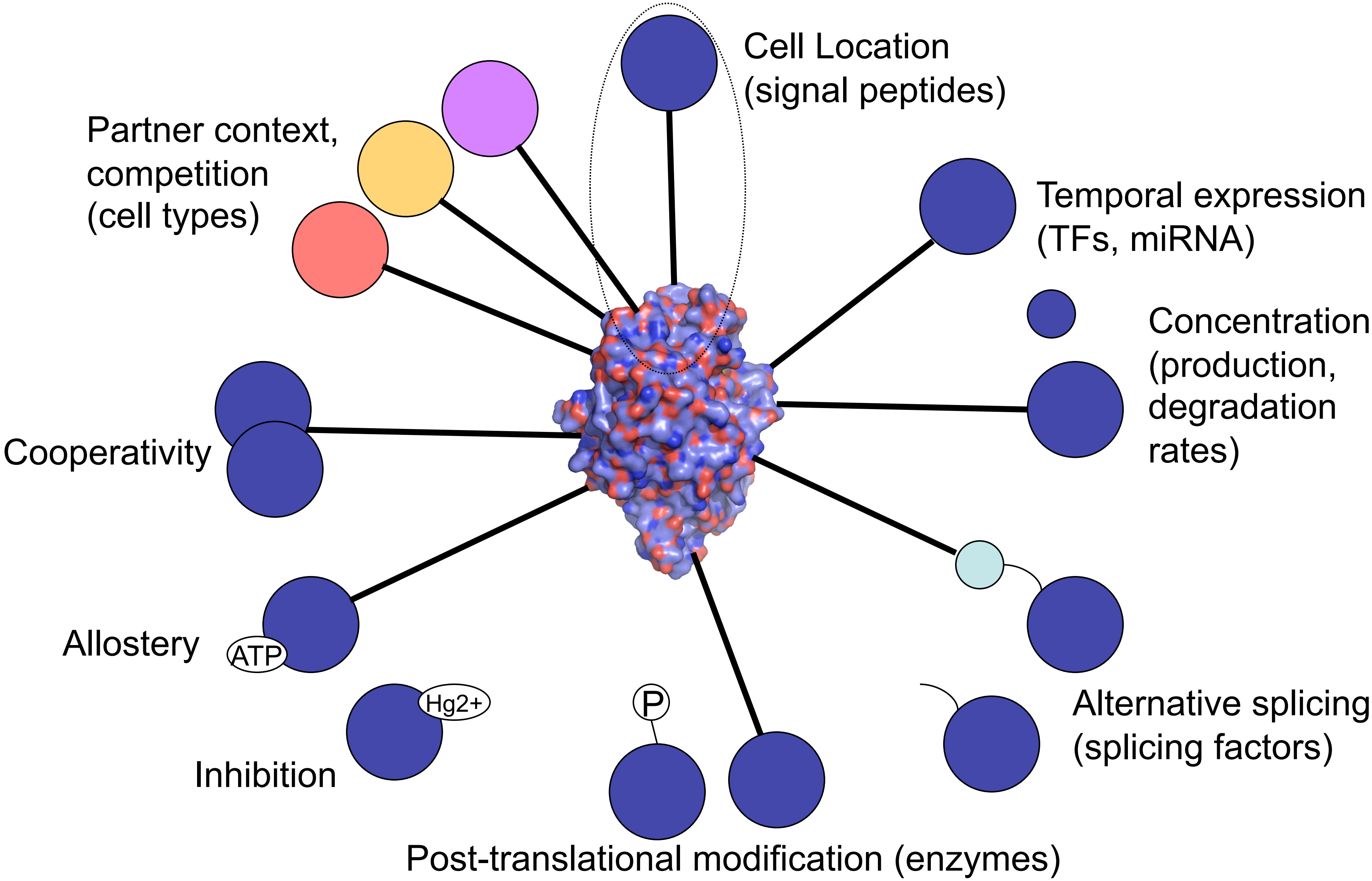
CyclinA

Specific Interaction

CyclinA-Cdk2 complex

1FIN

Towards More Biologically Relevant Networks



Cell Location
(signal peptides)

Partner context,
competition
(cell types)

Temporal expression
(TFs, miRNA)

Concentration
(production,
degradation
rates)

Cooperativity

Allostery

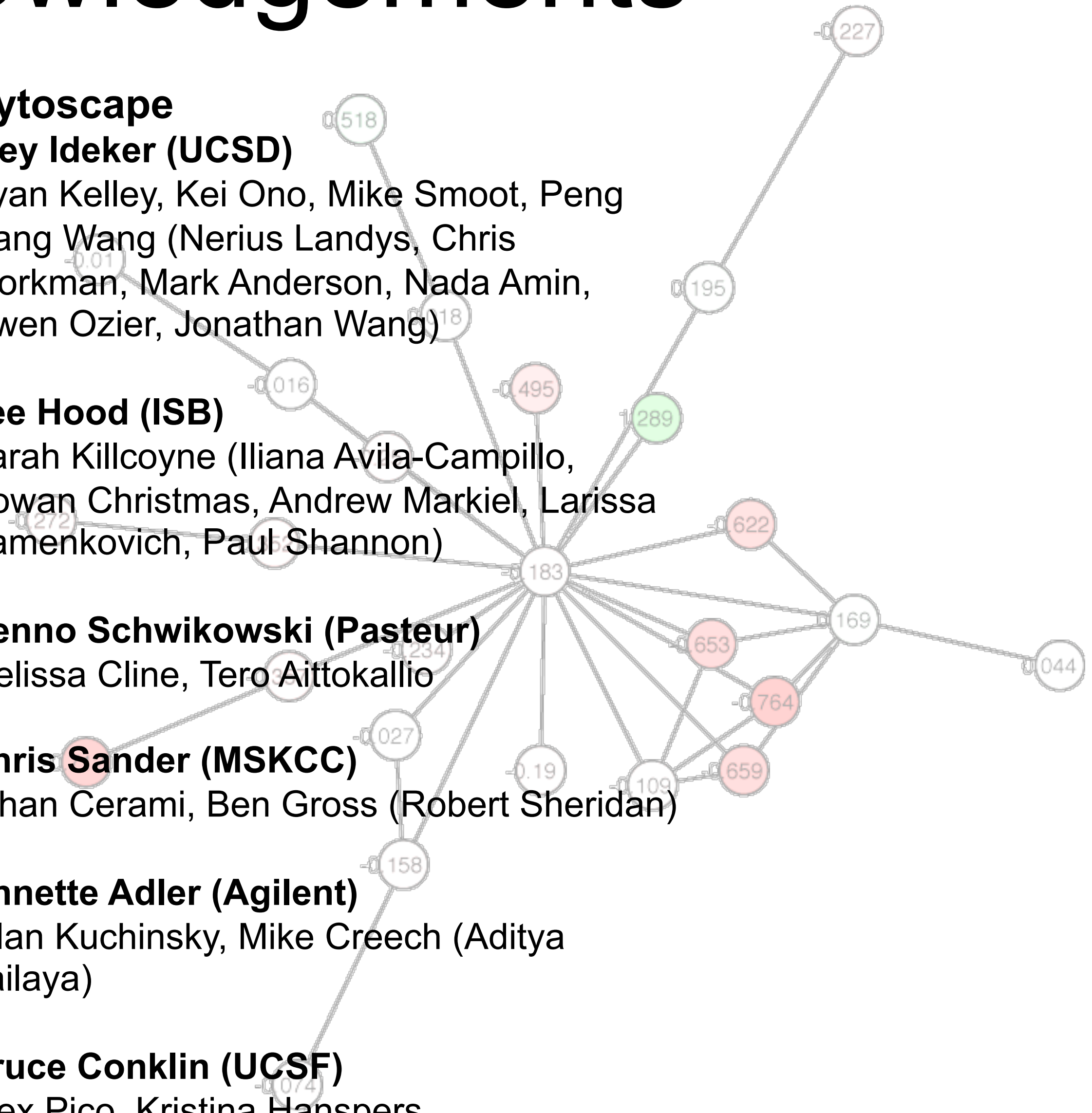
Alternative splicing
(splicing factors)

Inhibition

Post-translational modification (enzymes)

Acknowledgements

Pathway Commons
Chris Sander
Ethan Cerami
Ben Gross
Emek Demir
Robert Hoffmann



Cytoscape
Trey Ideker (UCSD)
Ryan Kelley, Kei Ono, Mike Smoot, Peng Liang Wang (Nerius Landys, Chris Workman, Mark Anderson, Nada Amin, Owen Ozier, Jonathan Wang)

Lee Hood (ISB)
Sarah Killcoyne (Iliana Avila-Campillo, Rowan Christmas, Andrew Markiel, Larissa Kamenkovich, Paul Shannon)

Benno Schwikowski (Pasteur)
Melissa Cline, Tero Aittokallio

Chris Sander (MSKCC)
Ethan Cerami, Ben Gross (Robert Sheridan)

Annette Adler (Agilent)
Allan Kuchinsky, Mike Creech (Aditya Vailaya)

Bruce Conklin (UCSF)
Alex Pico, Kristina Hanspers

Funding
CIHR
Genome Canada
Canada Foundation for Innovation

http://baderlab.org

Gene and Protein Identifiers

- Identifiers (IDs) are 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 use the correct record type
  - E.g. Entrez Gene records don’t store sequence. They link to DNA regions, RNA transcripts and proteins.

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

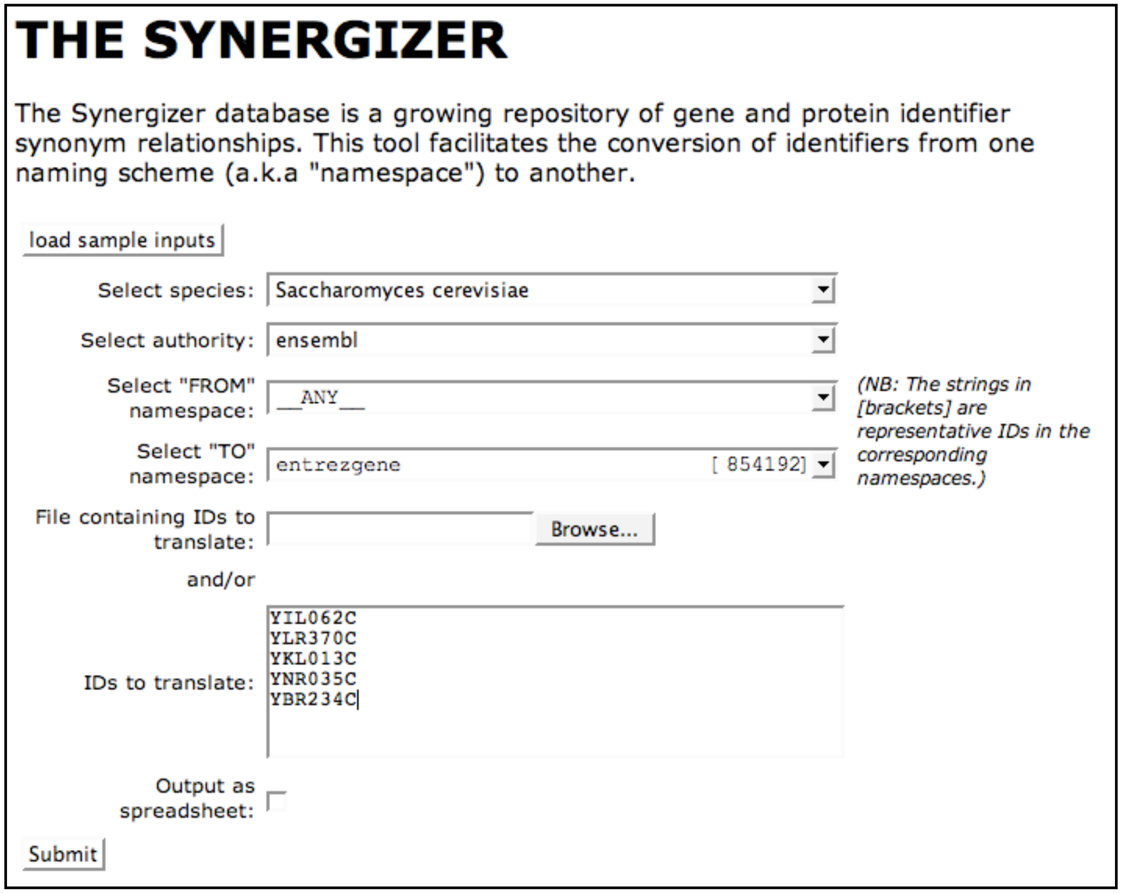
InterPro IPR015252
OMIM 600185
Pfam PF09104
Gene Ontology GO:0000724
SNPs rs28897757

Experimental Platform

Affymetrix 208368_3p_s_at
Agilent A_23_P99452
CodeLink GE60169
Illumina GI_4502450-S

Red = Recommended

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.

Select species: Saccharomyces cerevisiae

Select authority: ensembl

Select "FROM" namespace: __ANY__

Select "TO" namespace: entrezgene [854192]

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

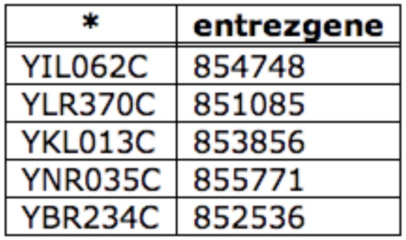
File containing IDs to translate:

and/or

IDs to translate:

```
YIL062C
YLR370C
YKL013C
YNR035C
YBR234C
```

Output as spreadsheet: ☐



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

- Synergizer
  - http://llama.med.harvard.edu/cgi/synergizer/translate
- Ensembl BioMart
  - http://www.ensembl.org
- PIR
  - http://pir.georgetown.edu/pirwww/search/idmapping.shtml

ID Mapping Challenges

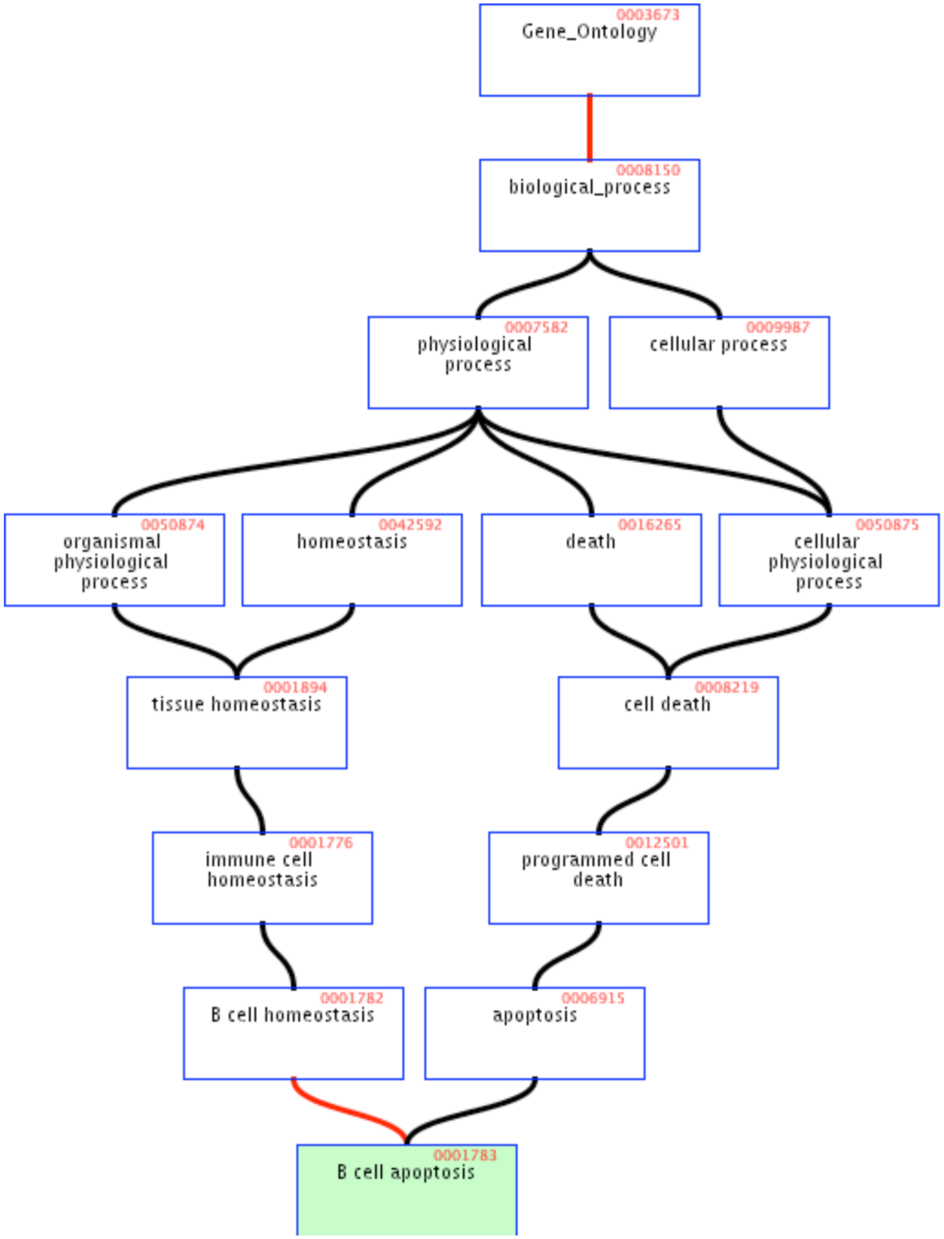
- Gene name ambiguity
  - Not a good ID, but official gene symbol is ok e.g. HGNC/HUGO gene symbol
- Excel error-introduction
  - OCT4 is changed to October-4
- Problems reaching 100% coverage
  - E.g. due to version issues
  - Use multiple sources to increase coverage

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

The Gene Ontology (GO)

- Describes gene function

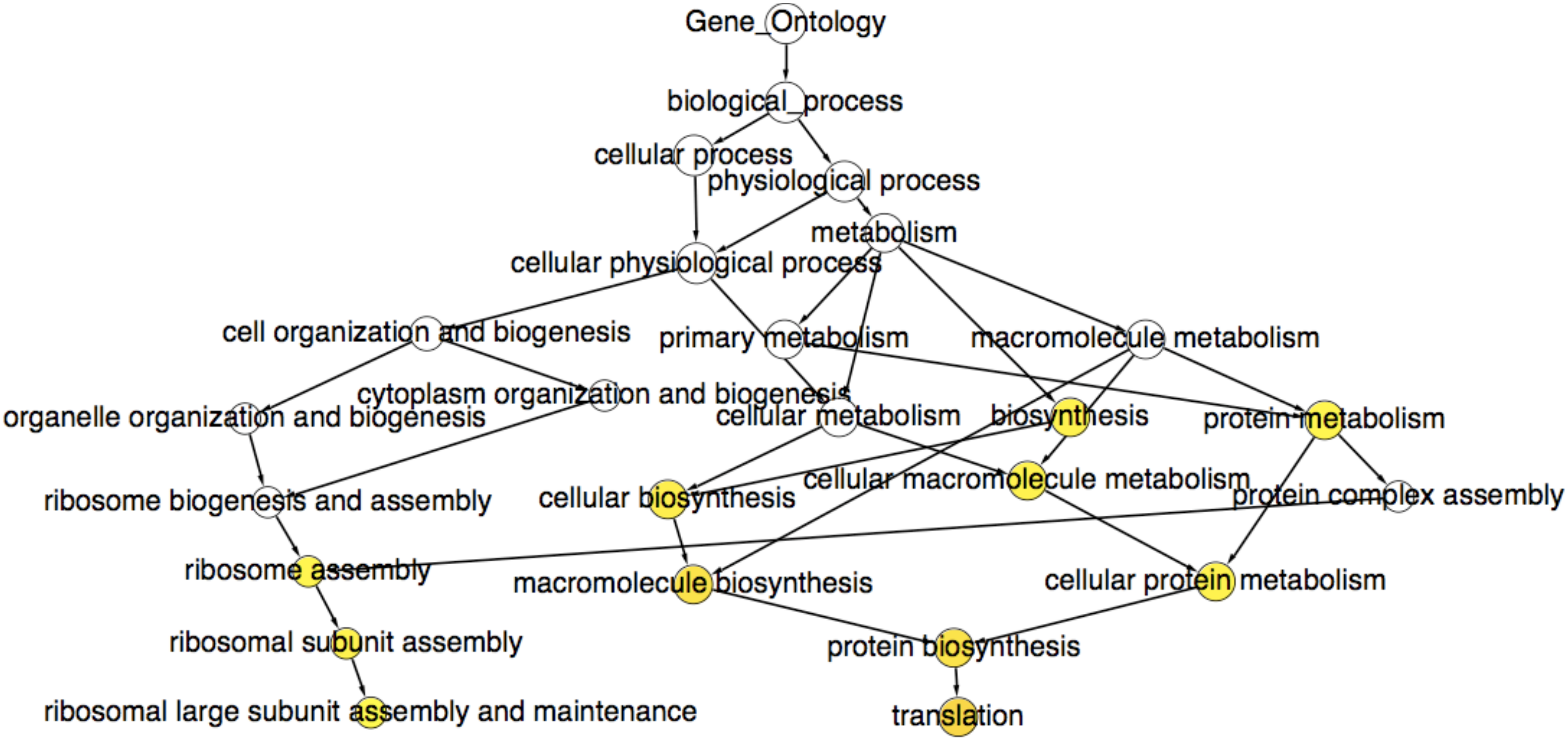
1. Agreed upon terms (controlled vocabulary)
    - Biological process
    - Cellular component
    - Molecular function
2. Genome annotation



www.geneontology.org

BiNGO

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



Caveats: Gene identifiers must match;
low GO term coverage, GO bias

Maere, S., Heymans, K. and Kuiper, M
Bioinformatics 21, 3448-3449, 2005

NetMatch

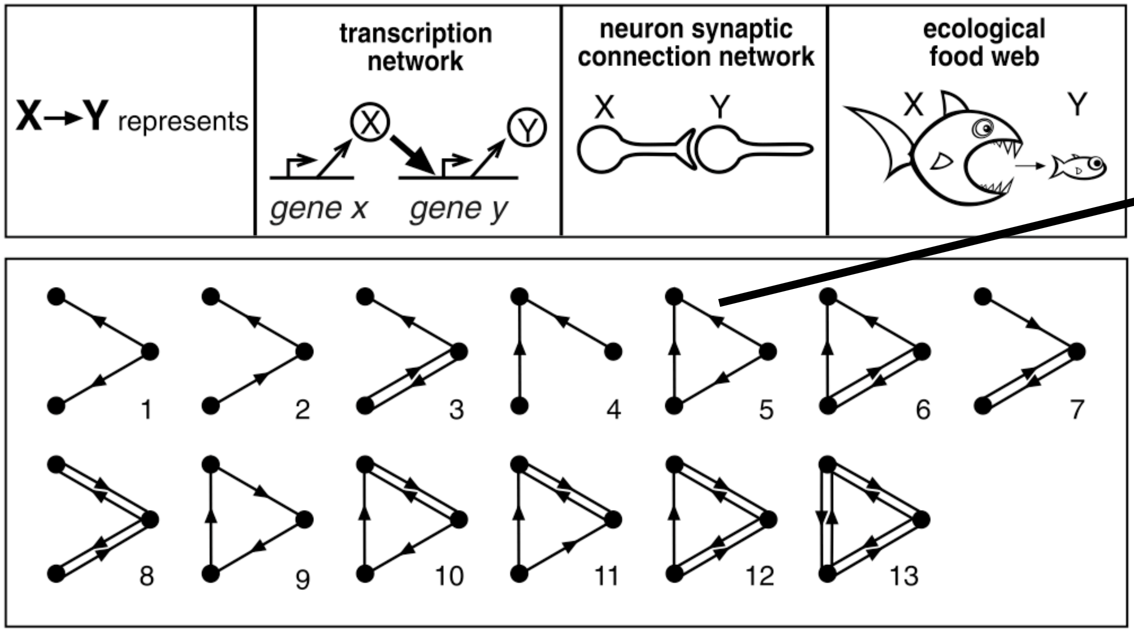
- Query a network for topological matches
- Input: query and target networks, optional node/edge labels
- Output: Topological query matches as subgraphs of target network
- Supports: subgraph matching, node/edge labels, label wildcards, approximate paths
- http://alpha.dmi.unict.it/~ctnyu/netmatch.html

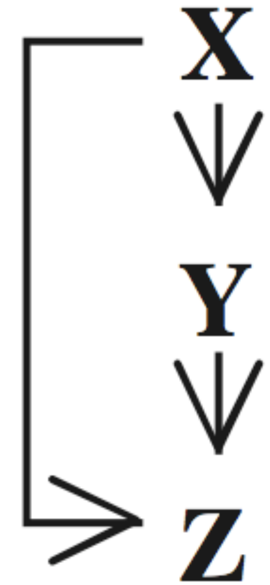
Ferro A, Giugno R, Pigola G, Pulvirenti A, Skripin D, Bader GD, Shasha D
Bioinformatics 2007 Feb 3

Extends state space representation based search from Cordella et al. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26, 10, 1367--1372

Find Feed-Forward Motifs

- Graph motifs over-represented in many network types



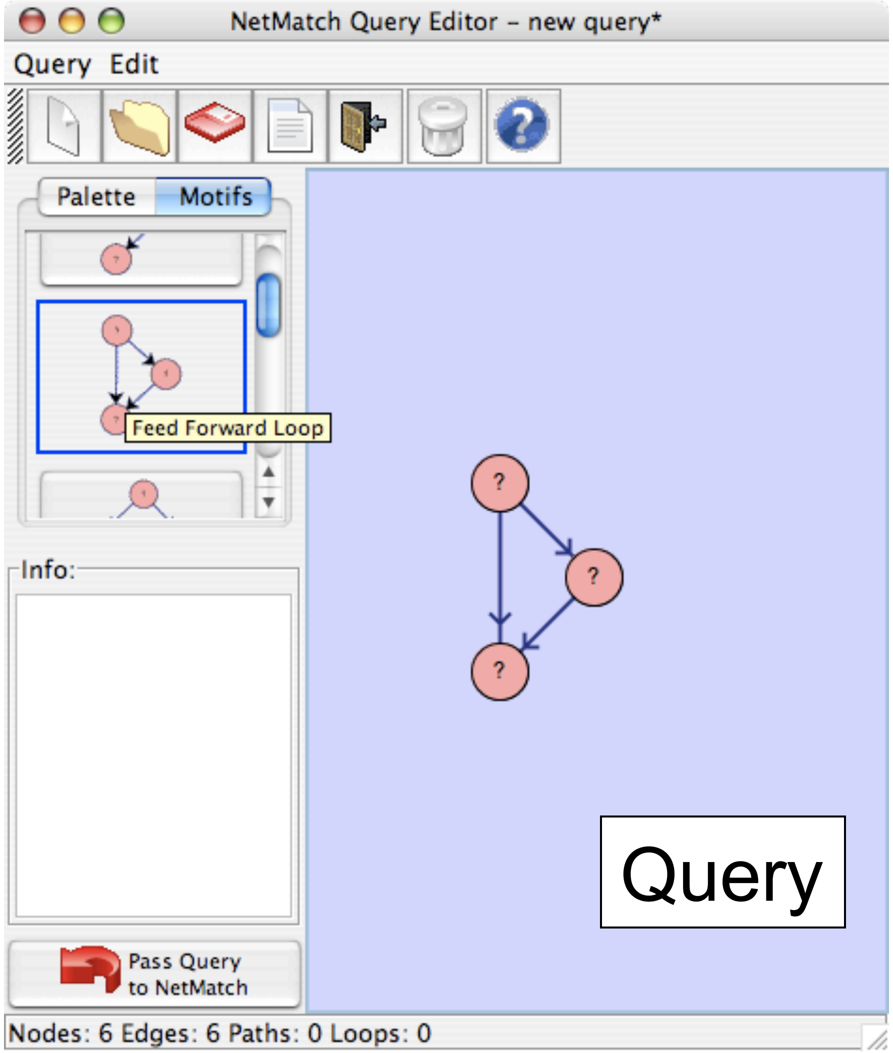


Feed-forward loop

Gene regulation
Neurons
Electronic circuits

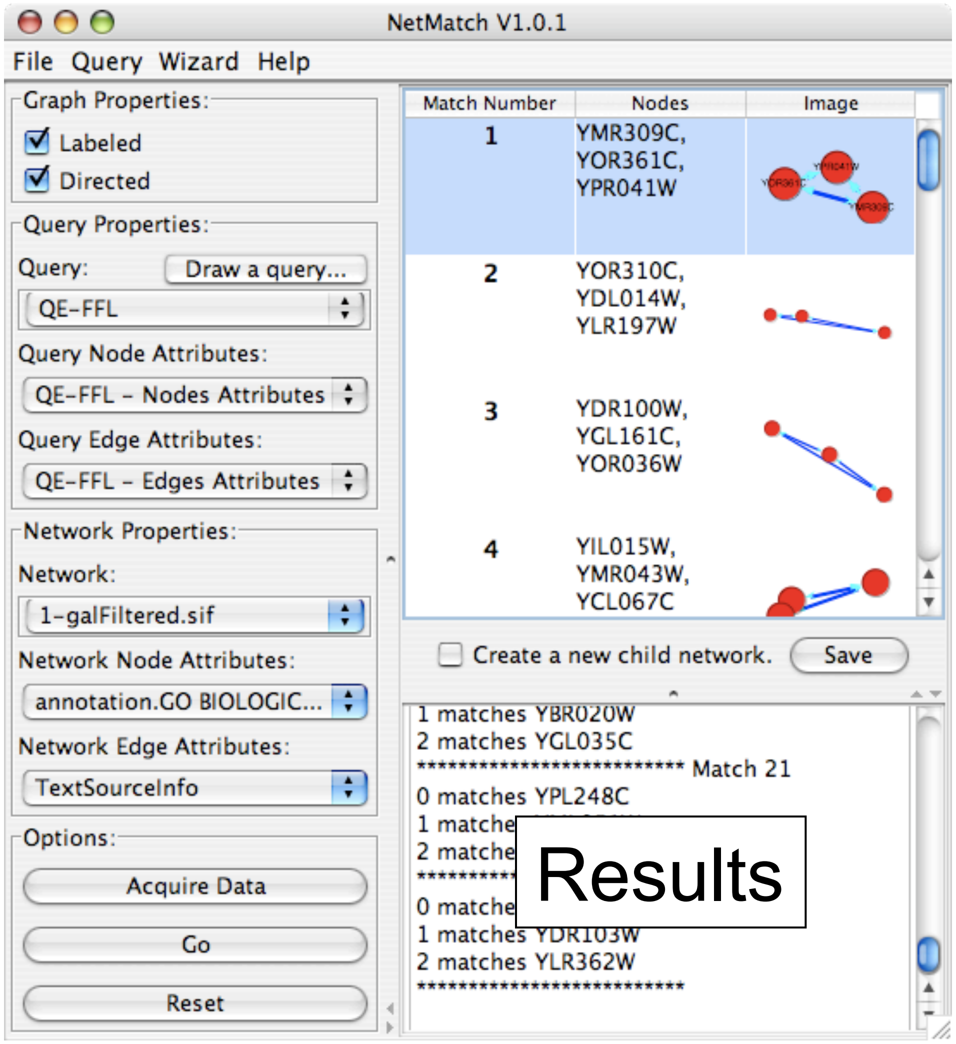
Milo et al. Science 298, 2002

Find Feed-Forward Motifs



Nodes: 6 Edges: 6 Paths: 0 Loops: 0

Query



Graph Properties:

☑ Labeled

☑ Directed

Query Properties:

Query: 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:

Match Number	Nodes	Image
1	YMR309C, YOR361C, YPR041W	
2	YOR310C, YDL014W, YLR197W	
3	YDR100W, YGL161C, YOR036W	
4	YIL015W, YMR043W, YCL067C	

☐ Create a new child network.

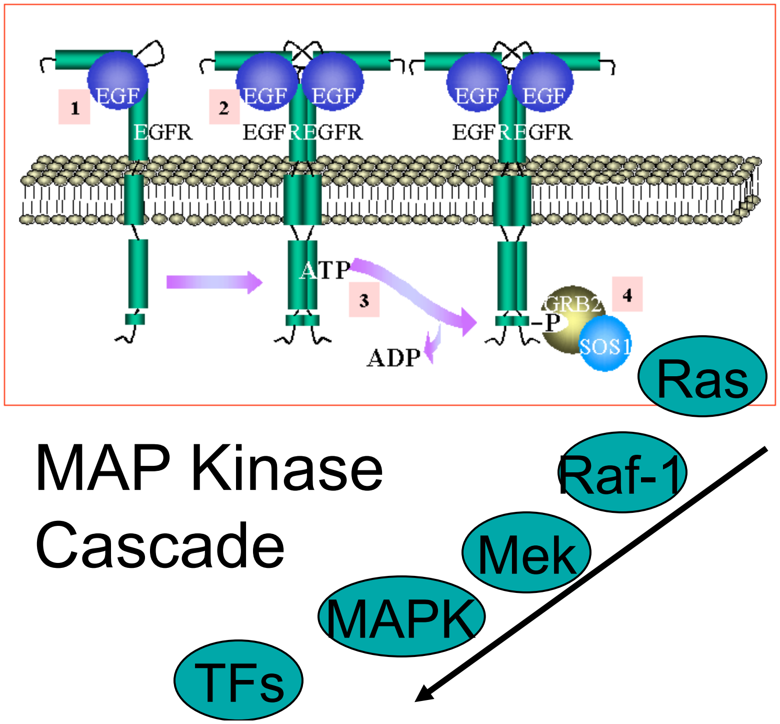
```
1 matches YBR020W
2 matches YGL035C
************************** Match 21
0 matches YPL248C
1 matche
2 matche
**********
0 matche
1 matches YDR103W
2 matches YLR362W
**************************
```

Results

Find Signaling Pathways

- Potential signaling pathways from plasma membrane to nucleus via cytoplasm

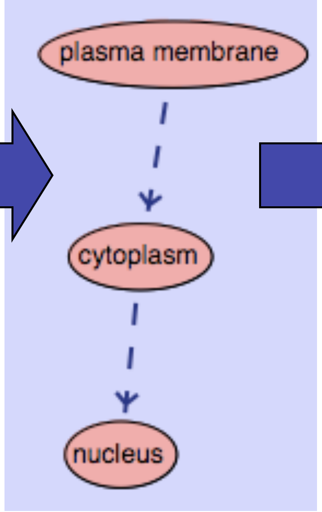
Signaling pathway example



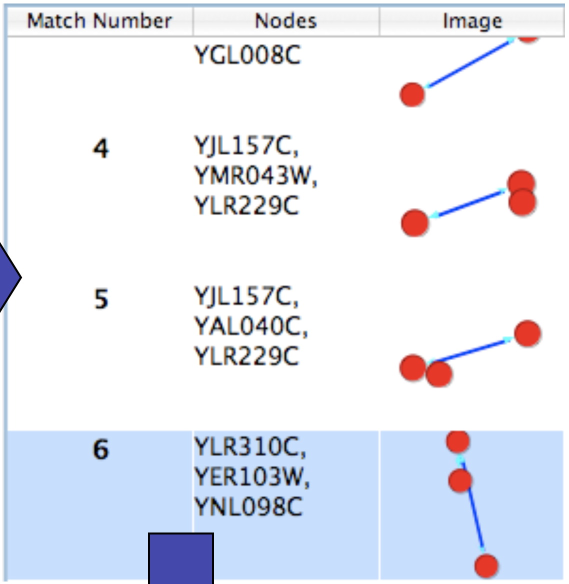
MAP Kinase Cascade

Nucleus - Growth Control Mitogenesis

NetMatch query

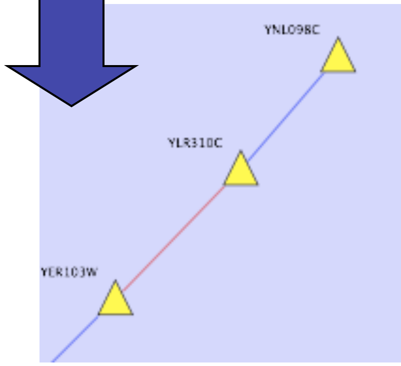


NetMatch Results



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

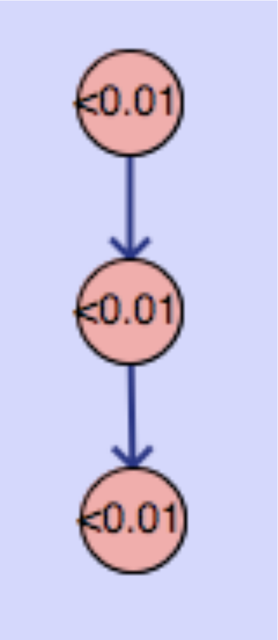
Shortest path between subgraph matches



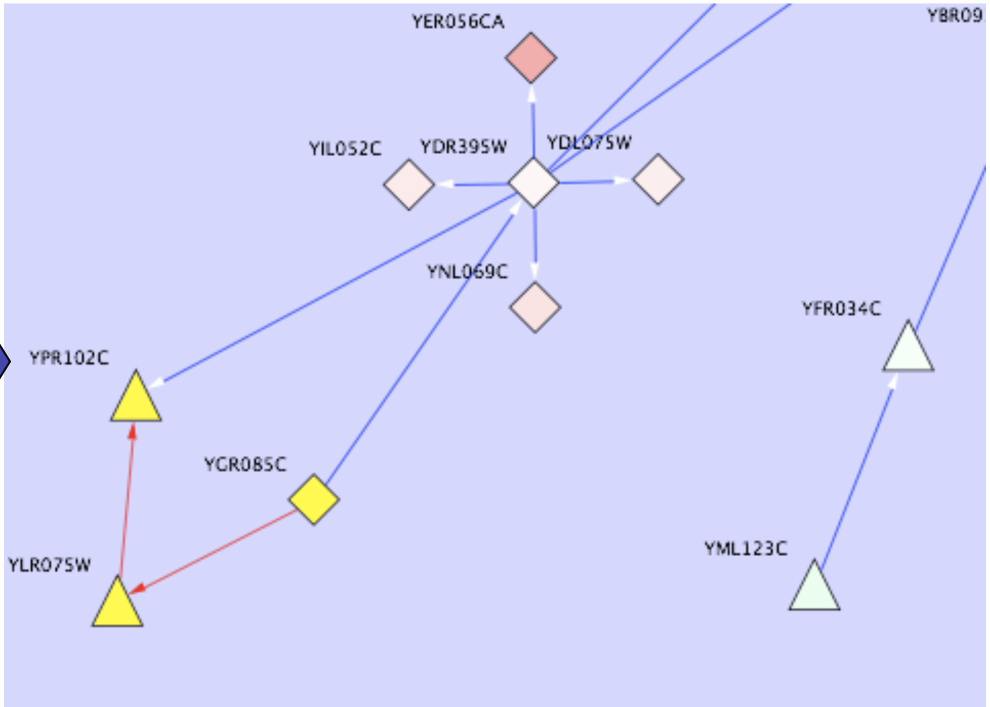
Find Expressed Motifs

Find specific subgraphs where certain nodes are significantly differentially expressed

NetMatch query



NetMatch Results



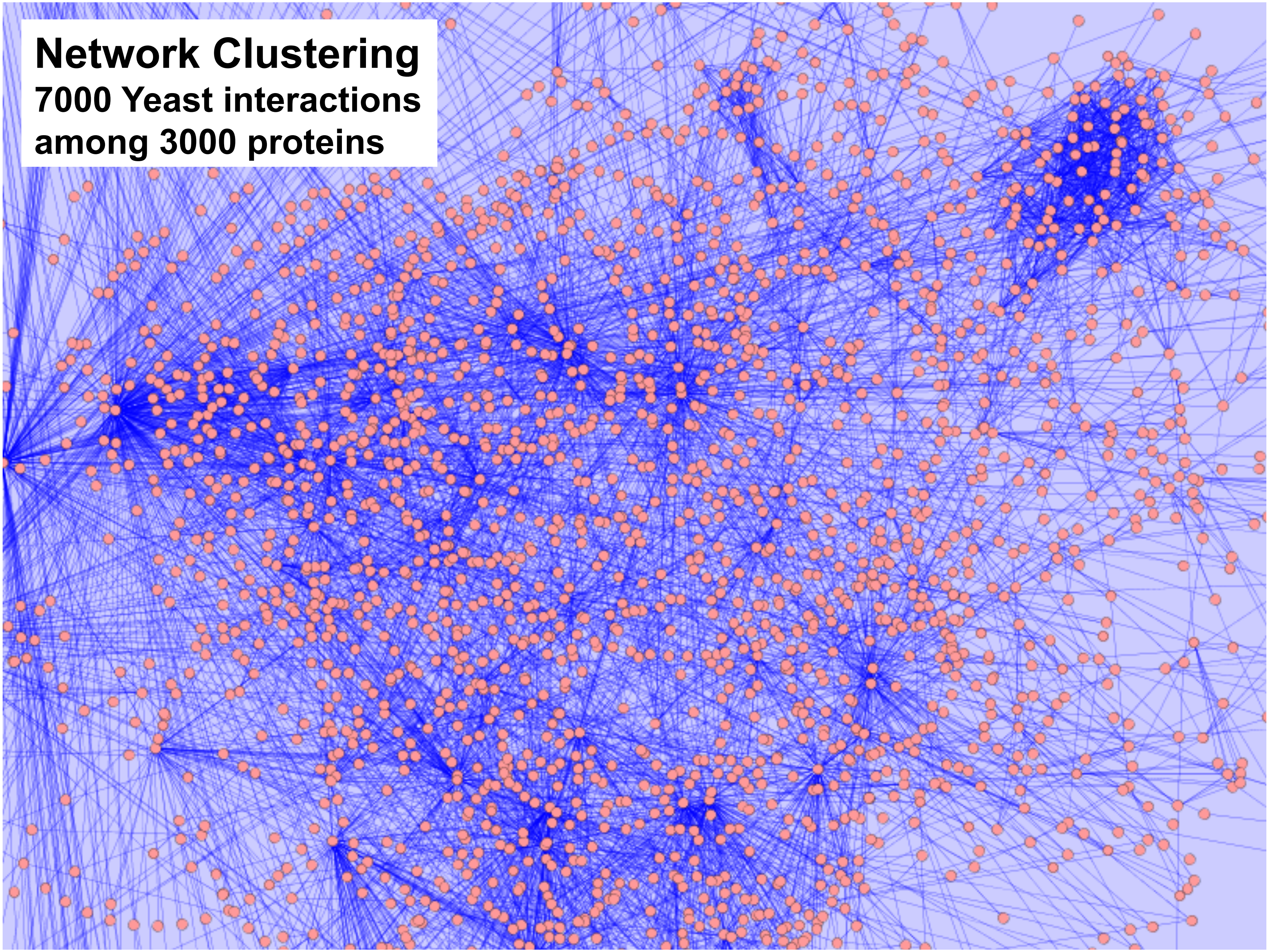
Protein	Differential Expression Significance
YLR075W	1.7255E-4
YGR085C	2.639E-4
YPR102C	3.7183E-4

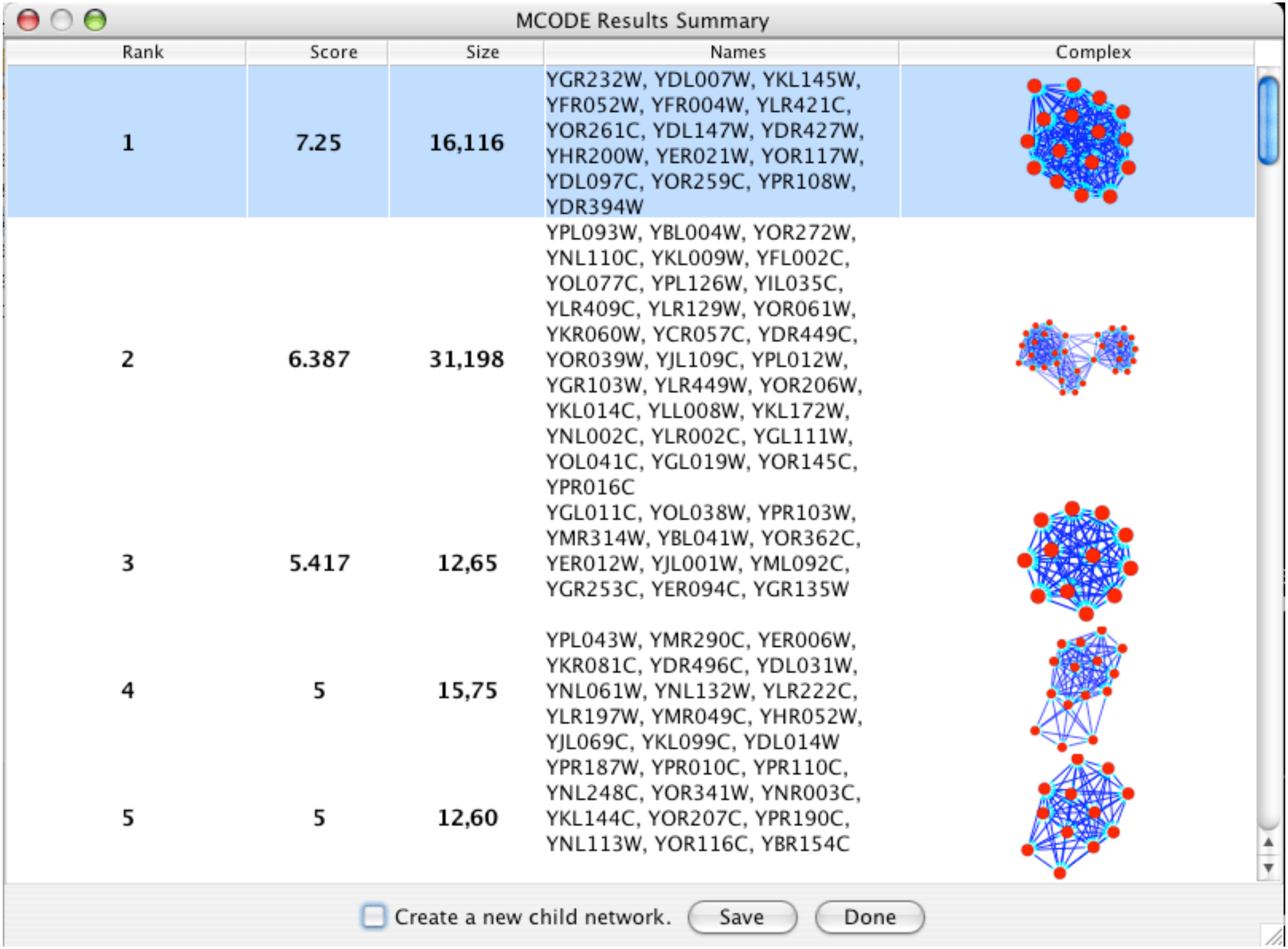
Graph Clustering - MCODE Plugin

- 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 MCODE Cytoscape plugin

Network Clustering

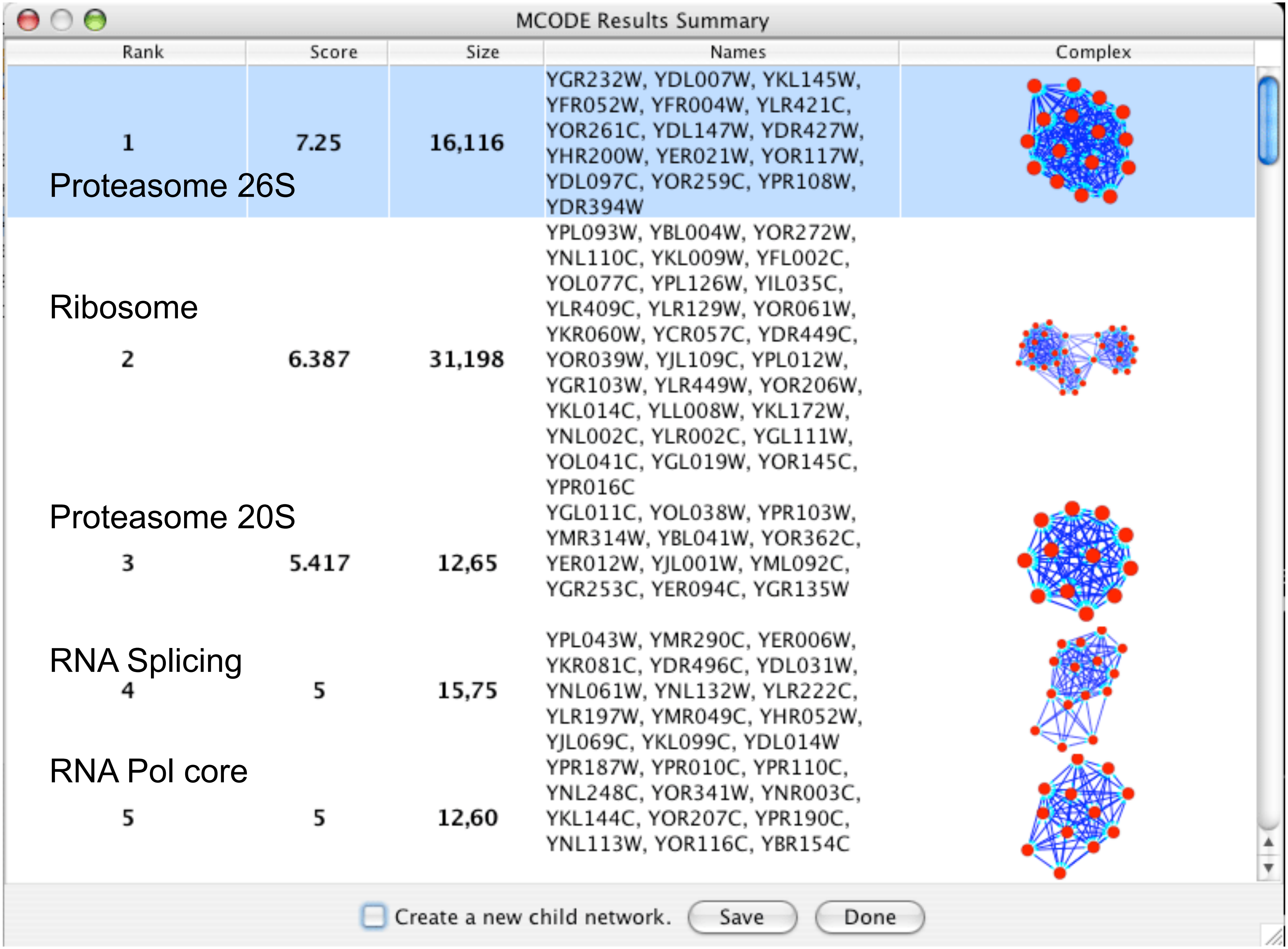
**7000 Yeast interactions
among 3000 proteins**





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	

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



Rank	Score	Size	Names	Complex
1 Proteasome 26S	7.25	16,116	YGR232W, YDL007W, YKL145W, YFR052W, YFR004W, YLR421C, YOR261C, YDL147W, YDR427W, YHR200W, YER021W, YOR117W, YDL097C, YOR259C, YPR108W, YDR394W	
Ribosome 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	
Proteasome 20S 3	5.417	12,65	YGL011C, YOL038W, YPR103W, YMR314W, YBL041W, YOR362C, YER012W, YJL001W, YML092C, YGR253C, YER094C, YGR135W	
RNA Splicing 4	5	15,75	YPL043W, YMR290C, YER006W, YKR081C, YDR496C, YDL031W, YNL061W, YNL132W, YLR222C, YLR197W, YMR049C, YHR052W, YJL069C, YKL099C, YDL014W	
RNA Pol core 5	5	12,60	YPR187W, YPR010C, YPR110C, YNL248C, YOR341W, YNR003C, YKL144C, YOR207C, YPR190C, YNL113W, YOR116C, YBR154C	

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

Systems Biology Graphical Notation

http://sbgn.org

In progress