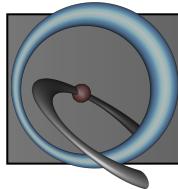


Protein-Protein Interactions Tools and Approaches

John “Scooter” Morris

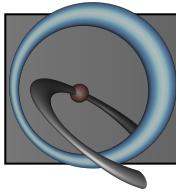
June 6, 2016
Eötvös Loránd University
Budapest, Hungary



Today's Agenda



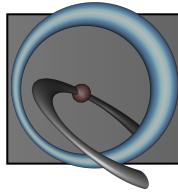
Time	Topic
10.00 – 11.00 10.00 – 11.00	Introductions – Why are protein-protein interactions important? Introduction – Why are protein-protein interactions important?
11.00 – 12.30	Protein-protein interaction networks – I Getting the data Experimental techniques for determining protein-protein interactions
12.30 – 13.30	High throughput techniques Lunch Low throughput techniques
13.30 – 15.00	Computational approaches Protein-protein interaction networks – II Literature mining • Network approaches to visualizing and analyzing protein-protein interactions Orthology Public repositories of protein-protein interaction data Lunch Introduction to Cytoscape - I
13.30 – 15.00 15.00 – 15.30	Break Protein-protein interactions networks – II • Network approaches to visualizing and analyzing protein-protein interactions
15.30 – 17.00 15.00 – 15.30 15.30 – 17.00	• Introduction to Cytoscape – II Break Hands-on Introduction to Cytoscape
17.00 – 18.00	Workflows



Introductions



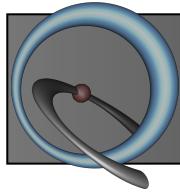
- John “Scooter” Morris
 - 2010-Current
 - Adjunct Assistant Professor, Pharmaceutical Chemistry
 - 2004-Current
 - Director, NCRR Resource for Biocomputing, Visualization, and Informatics (RBVI) @ UCSF
 - 1985-2004
 - Principal Systems Architect: Genentech, Inc.
 - Cytoscape core team since 2006
 - Author or co-author of several Cytoscape apps
 - *structureViz2, clusterMaker2, chemViz2, stringApp, CyAnimator*



What is this course about?



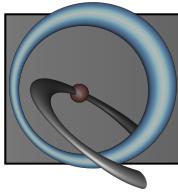
- Protein-protein interactions
 - Some background
 - Some techniques
- Two tools:
 - Cytoscape – network visualization and analysis
 - Chimera – structural visualization and analysis
- Why those tools?
 - It's what our lab works on ☺



Why?



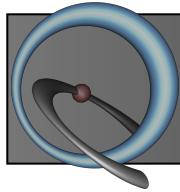
- Why are protein-protein interactions important?
 - You tell me...



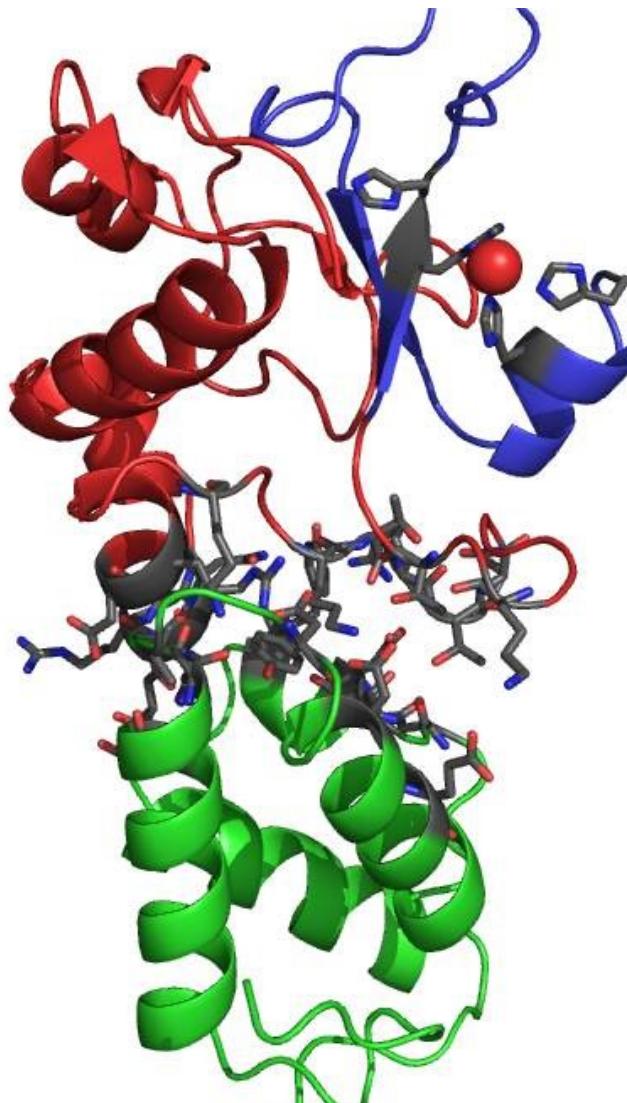
Why?

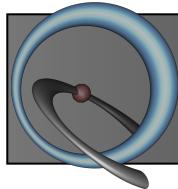


- Each participant
 - Name
 - Introduce your image
 - Say why protein-protein interactions are important

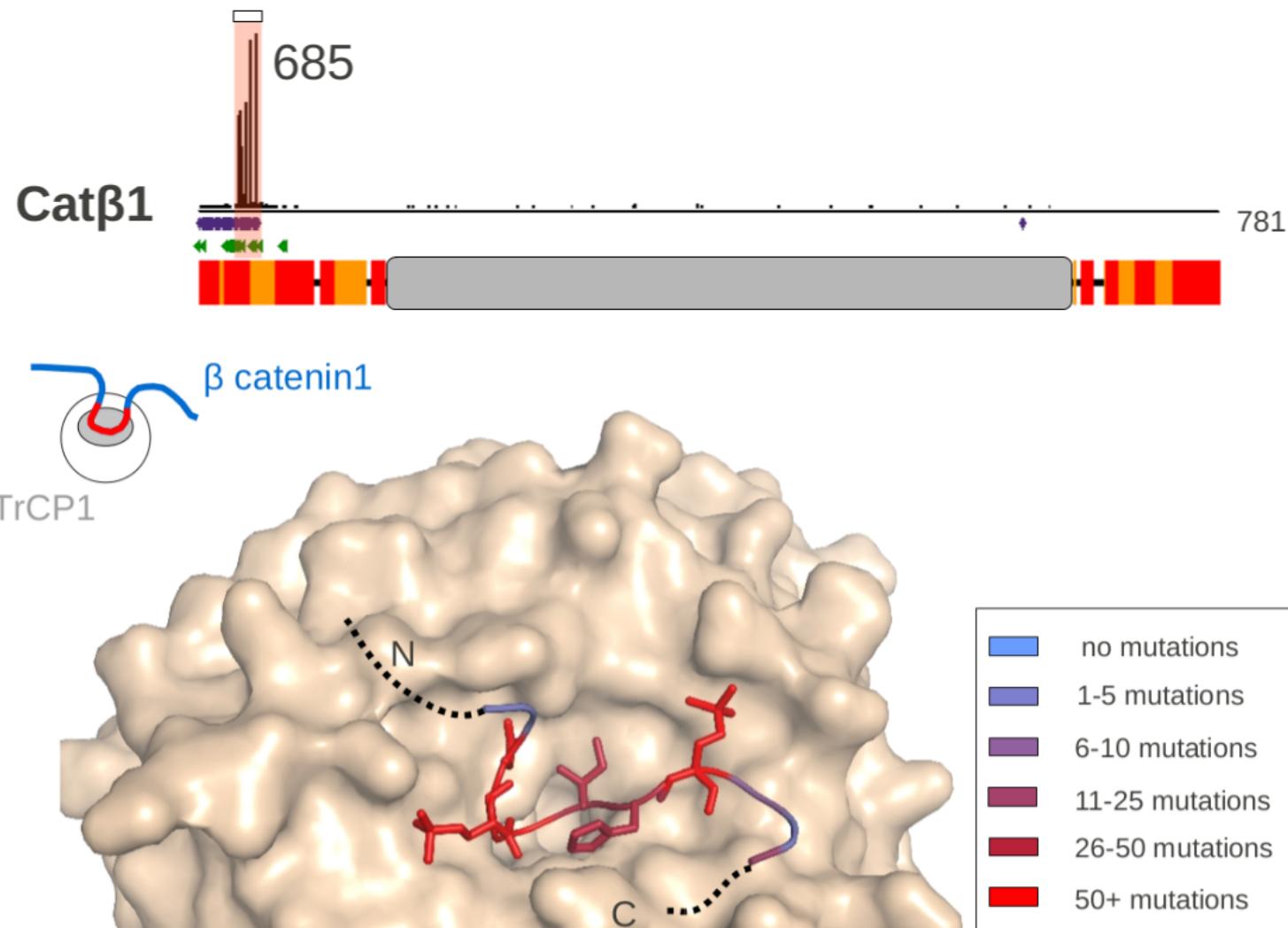


Katalin Borsos

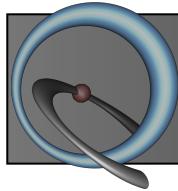




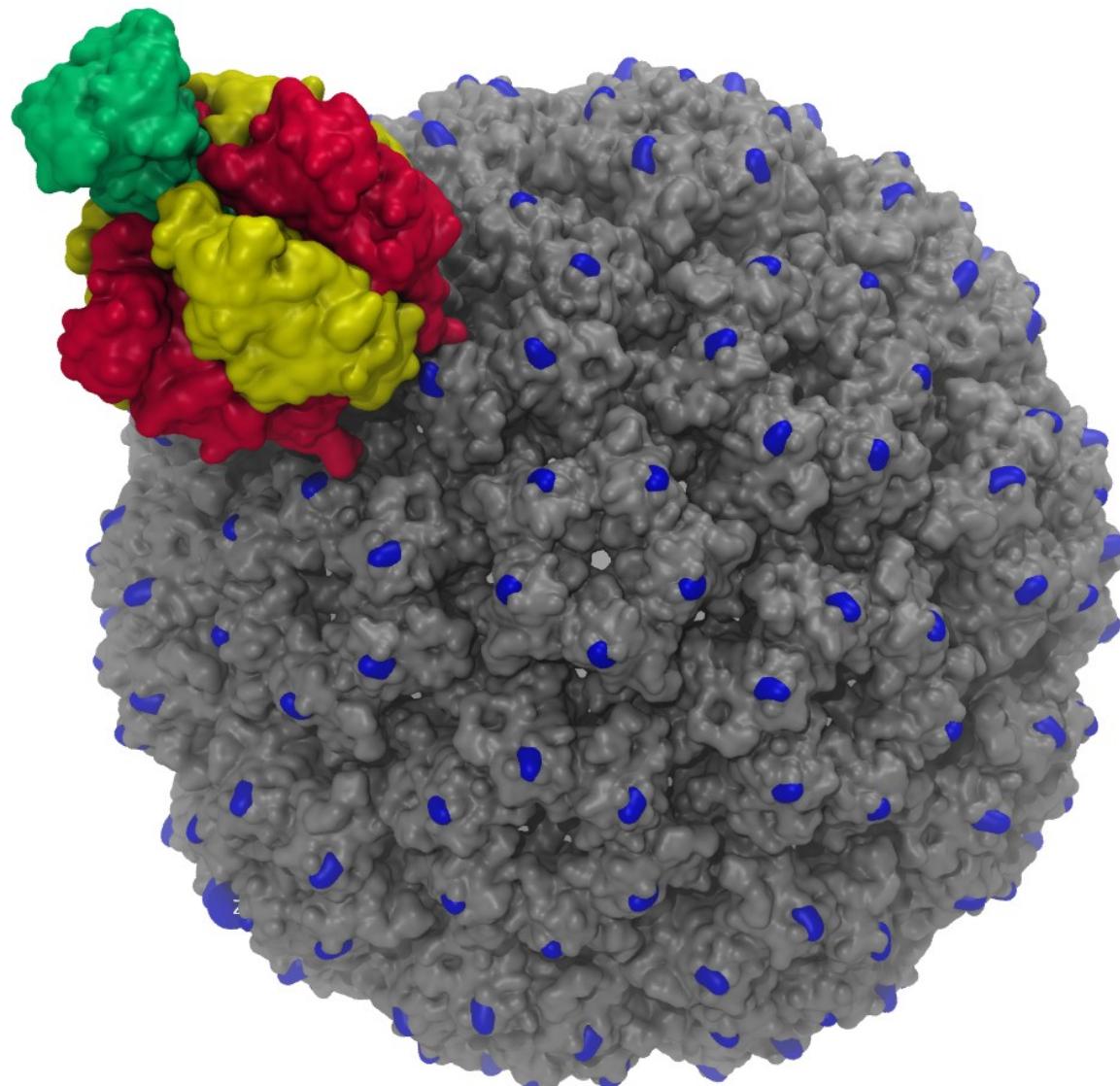
Borbála Soltész

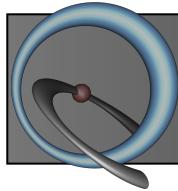


30-YLDSGIHSGAT-40

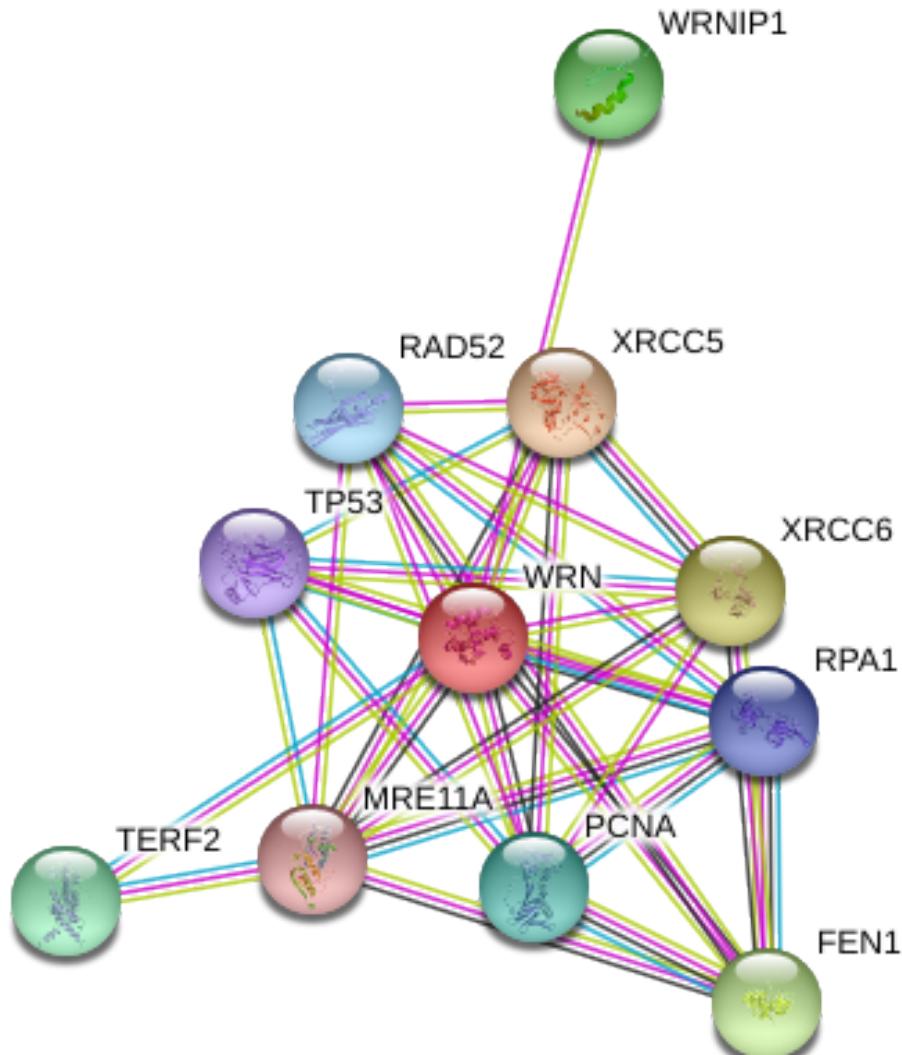


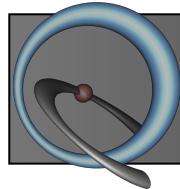
Ákos Gellért



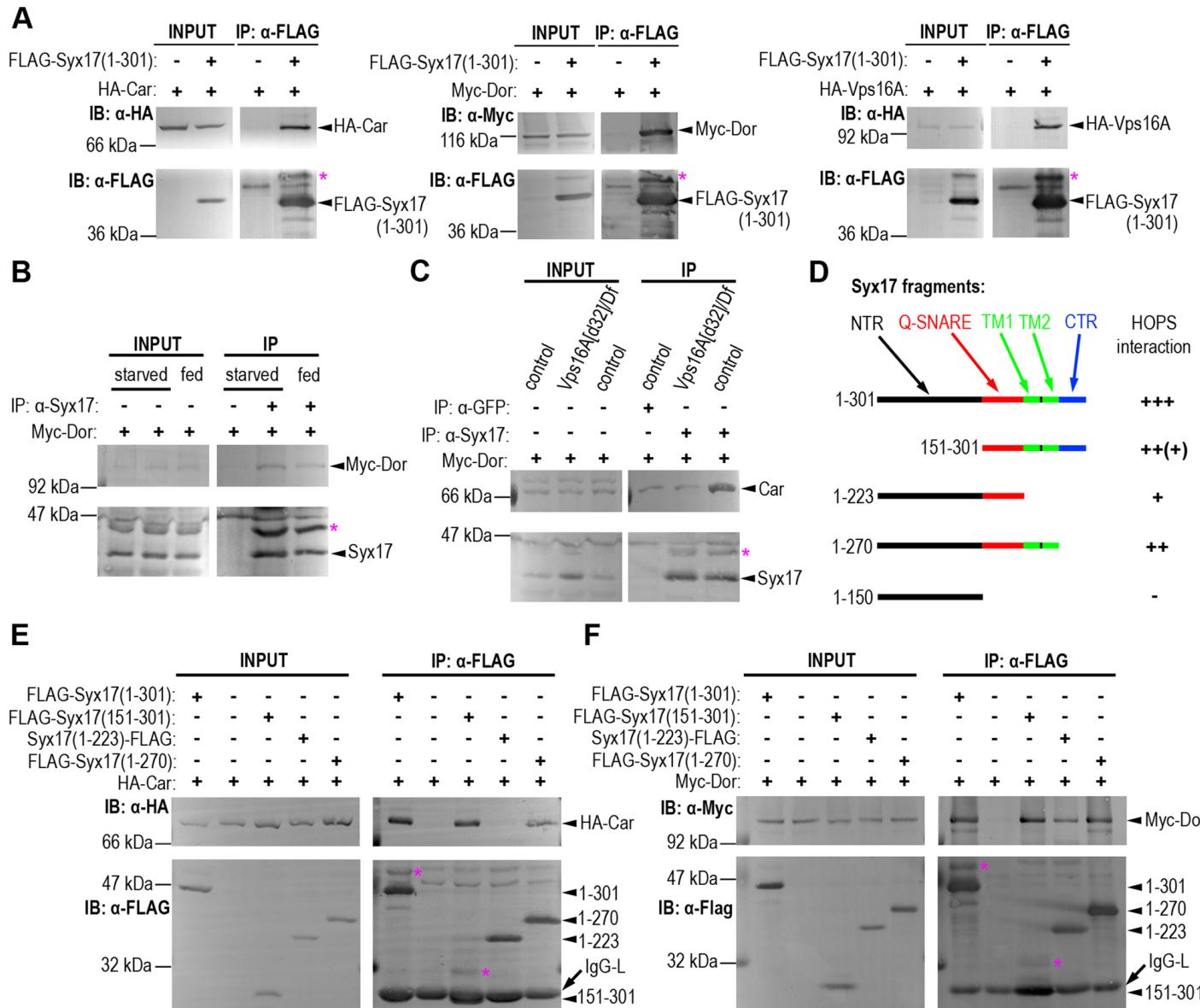


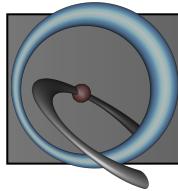
Kerepesi Csaba



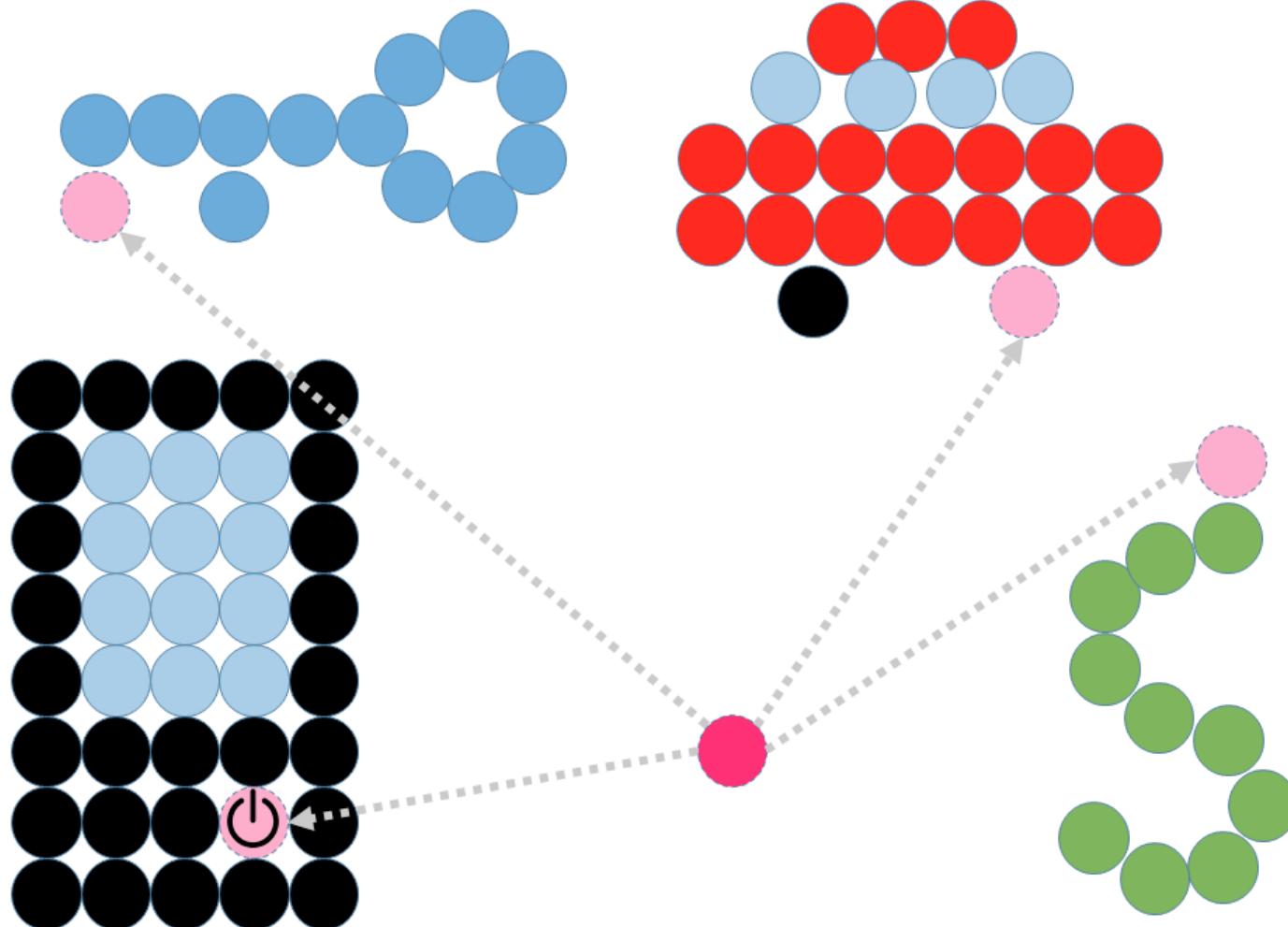


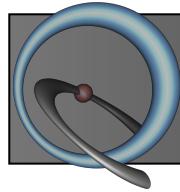
Takáts Szabolcs



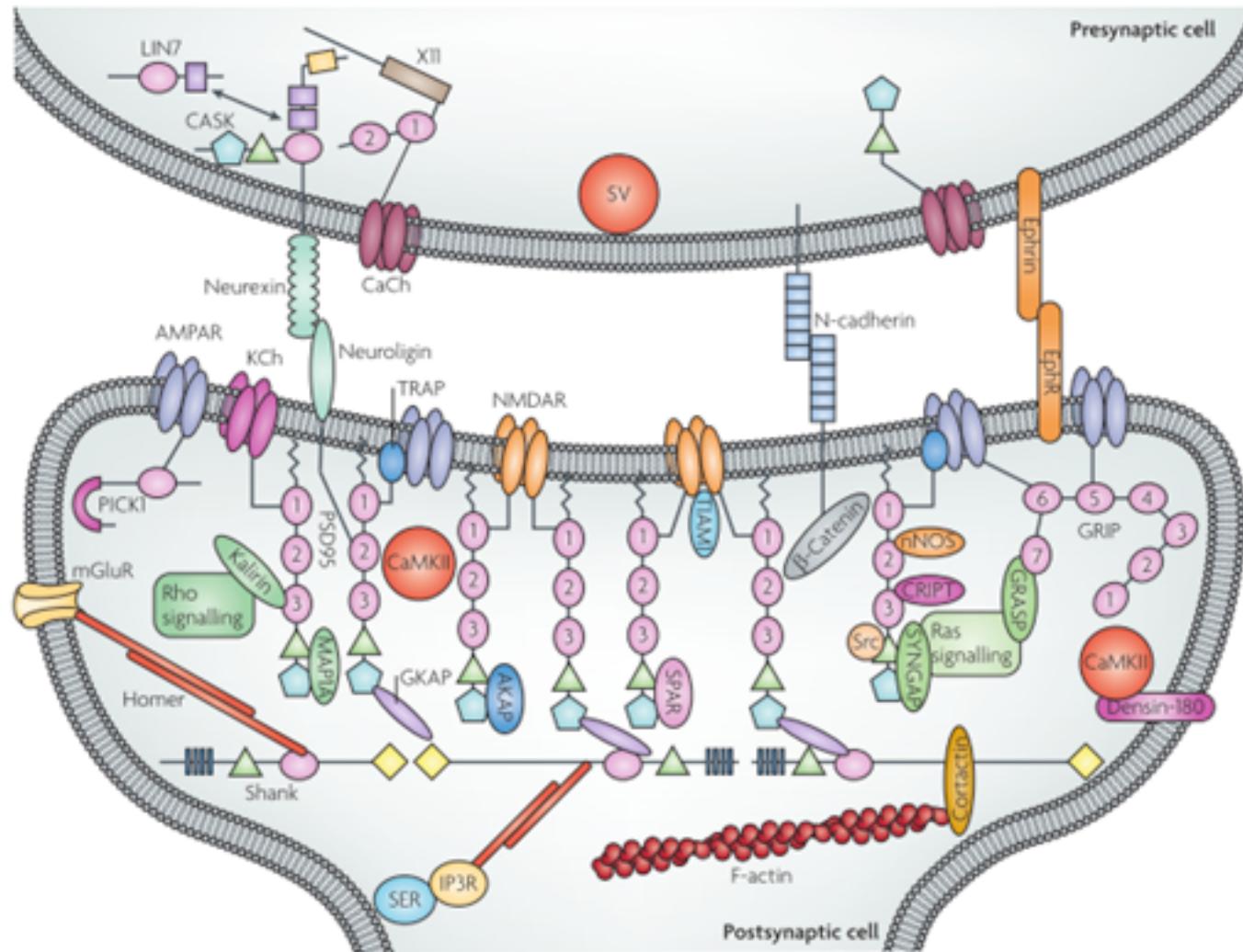


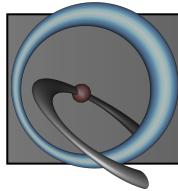
Lajos Pintér



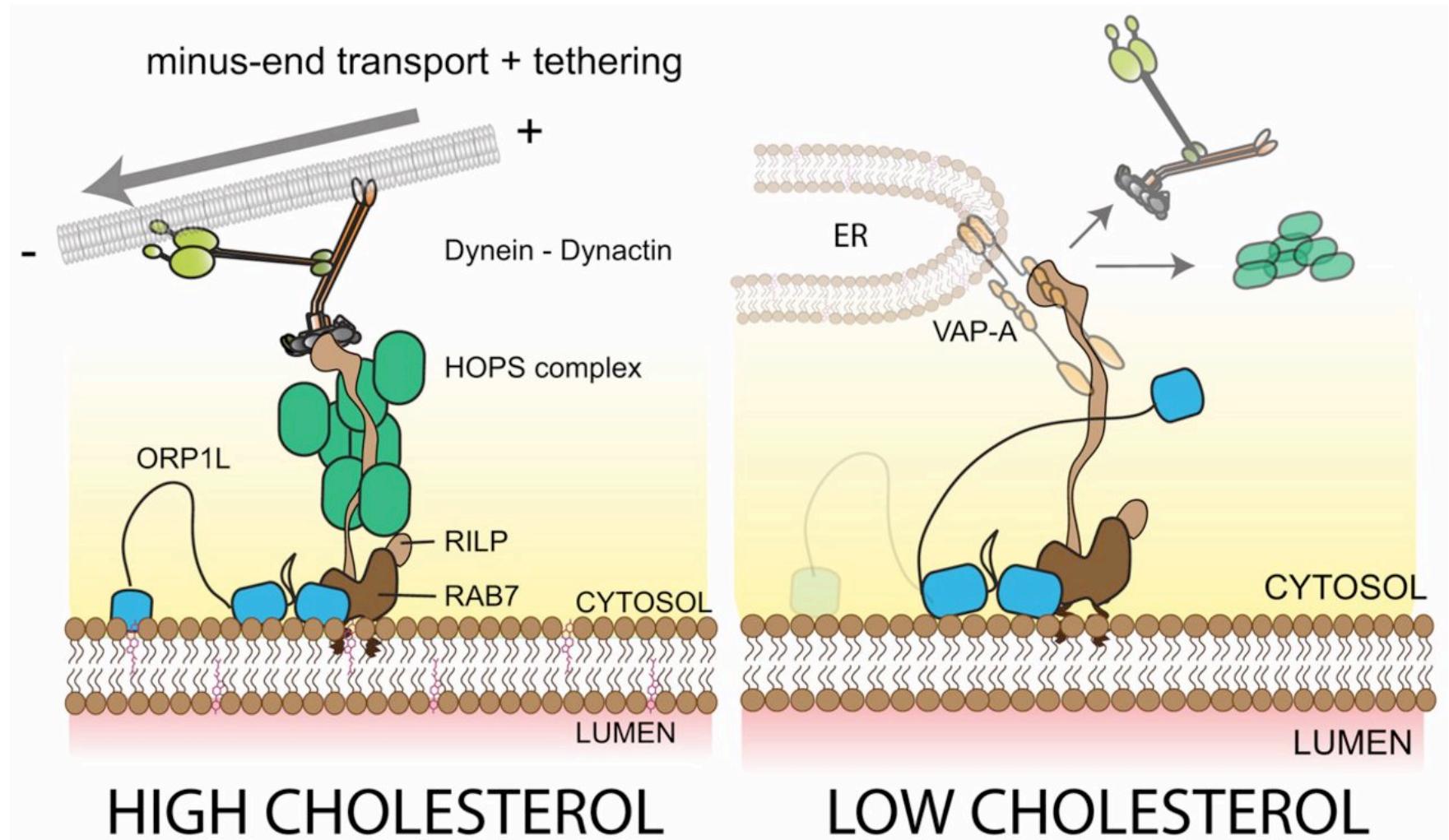


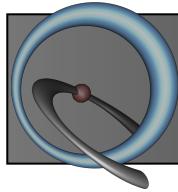
Bertalan Kovács



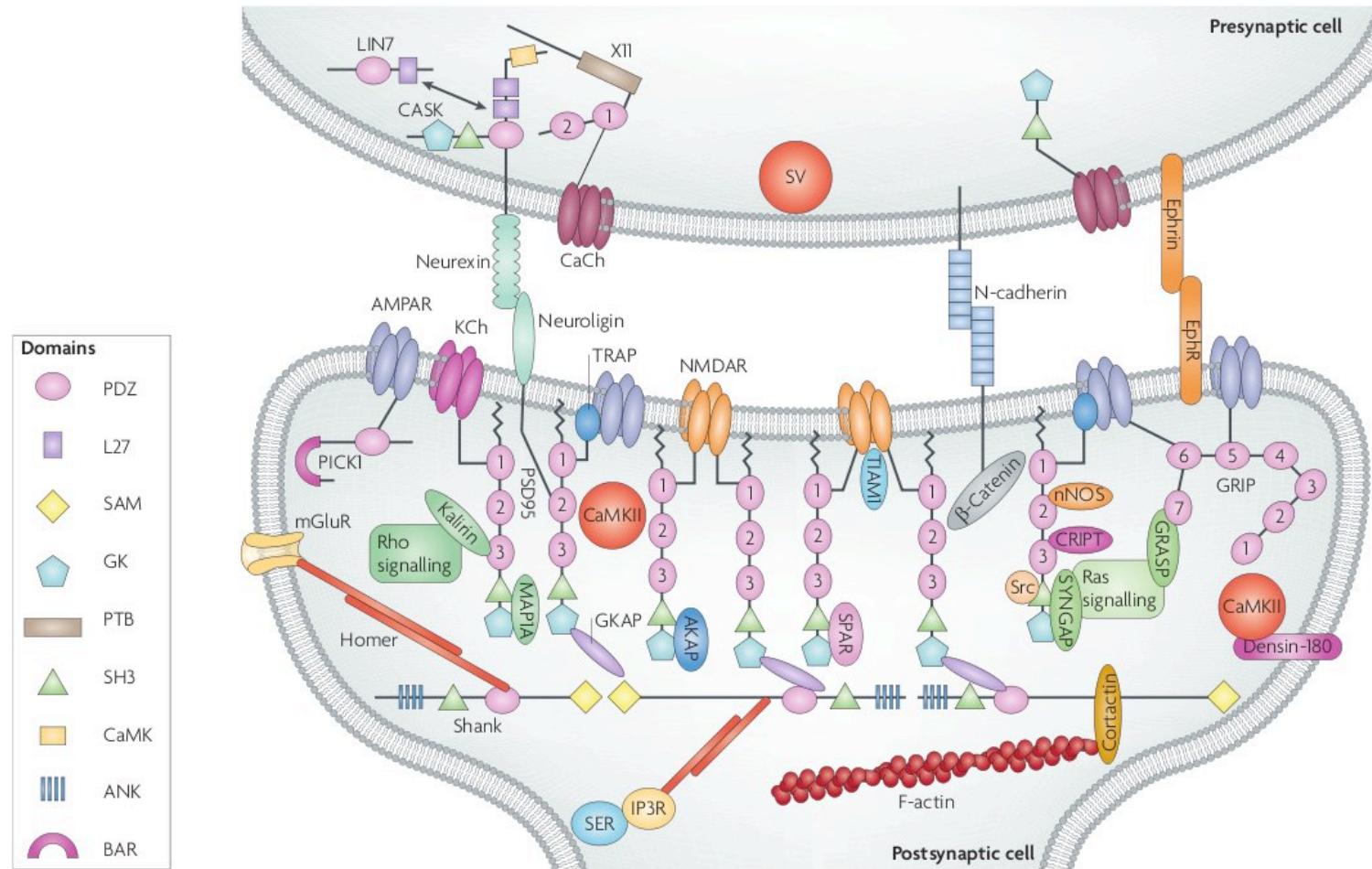


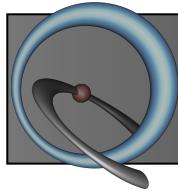
Tóth Sarolta



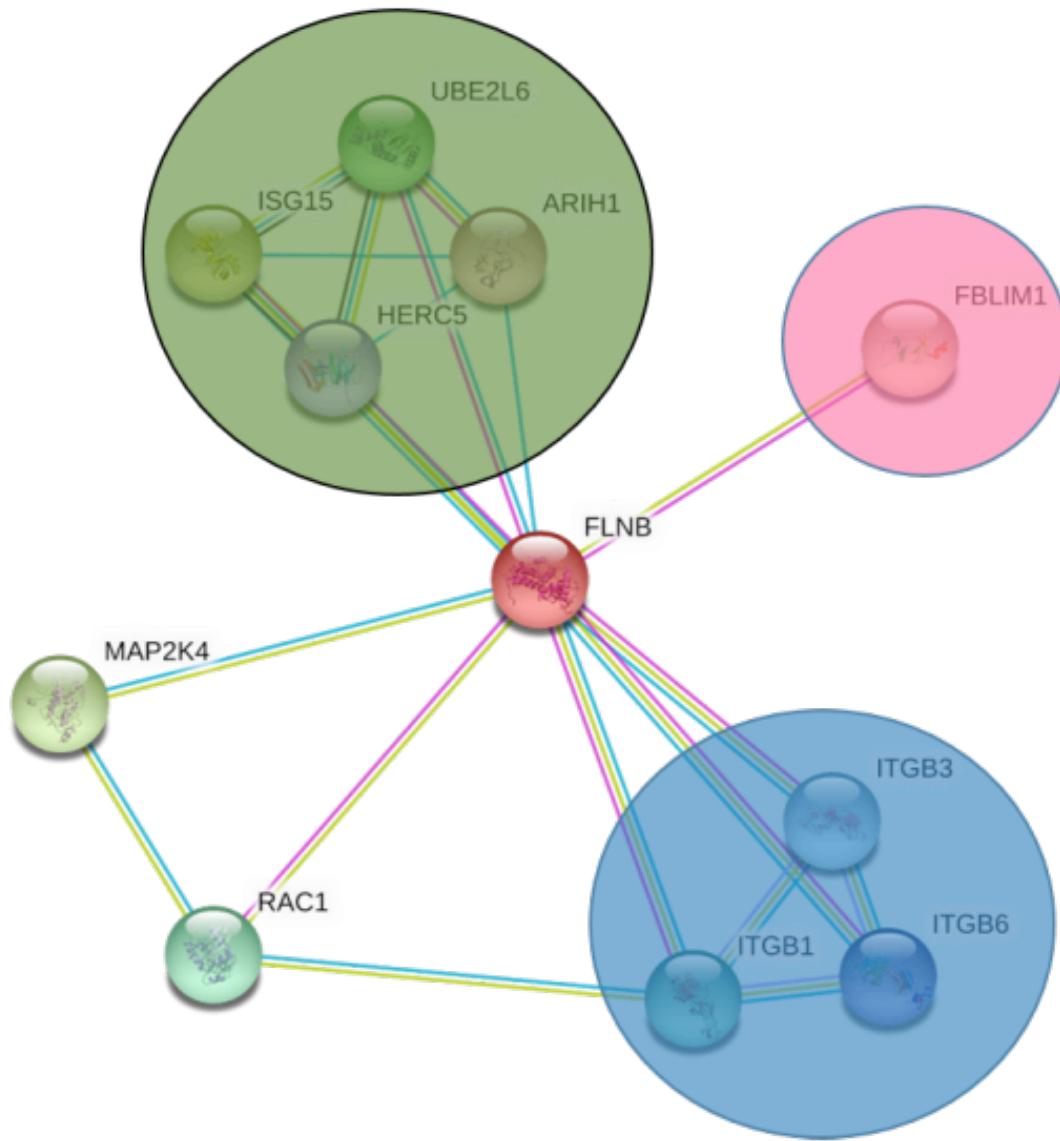


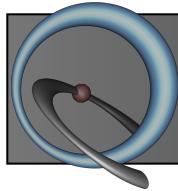
Korai Anett Márta



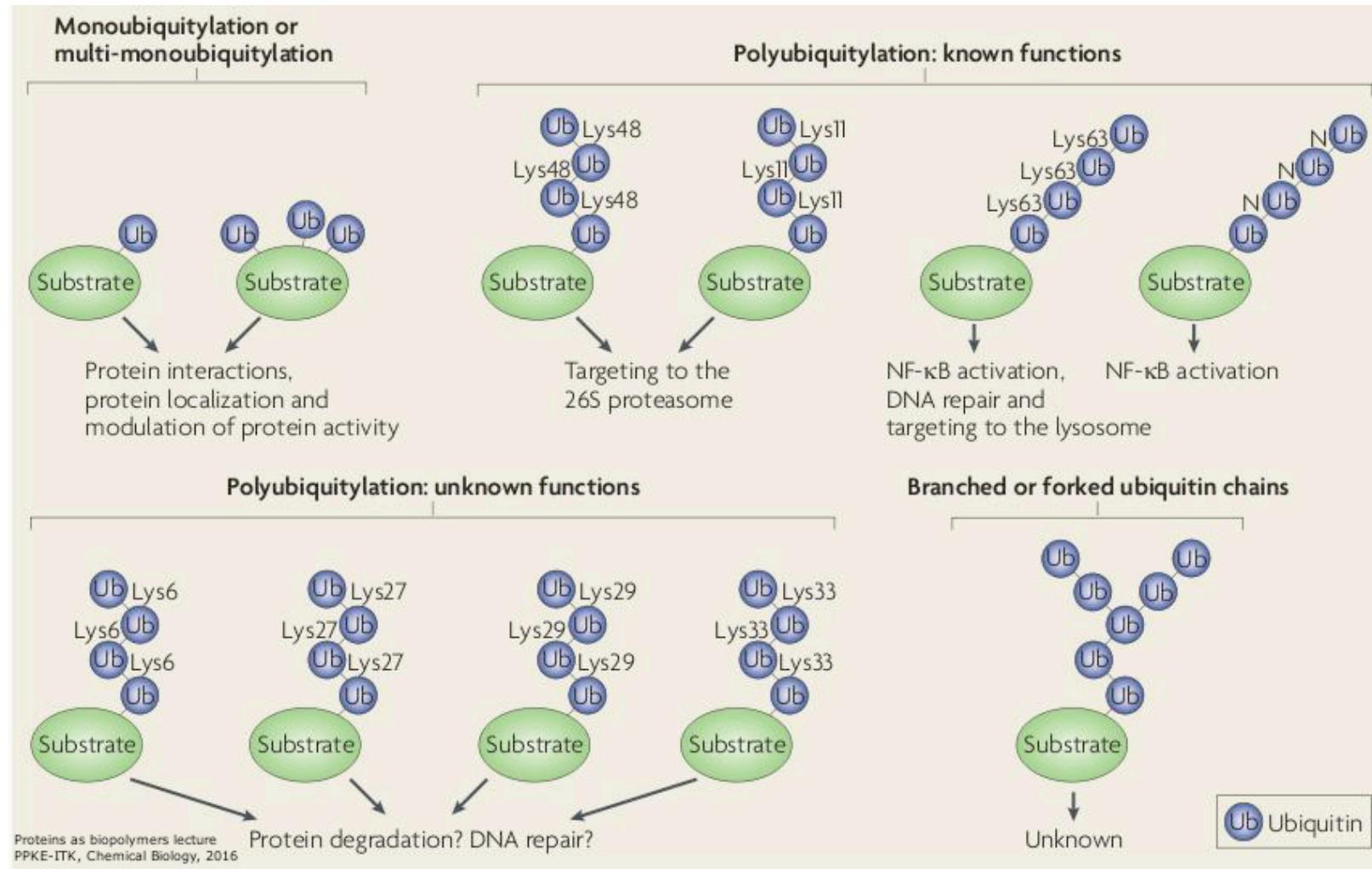


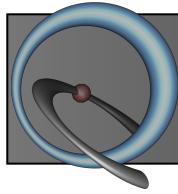
Sándor Babik





Tóth Ágota

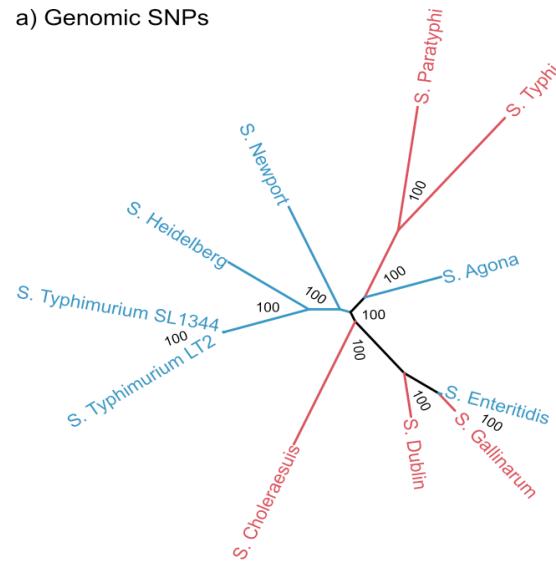




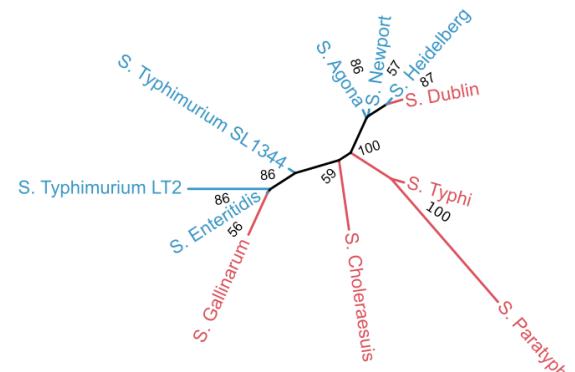
Ari Eszter



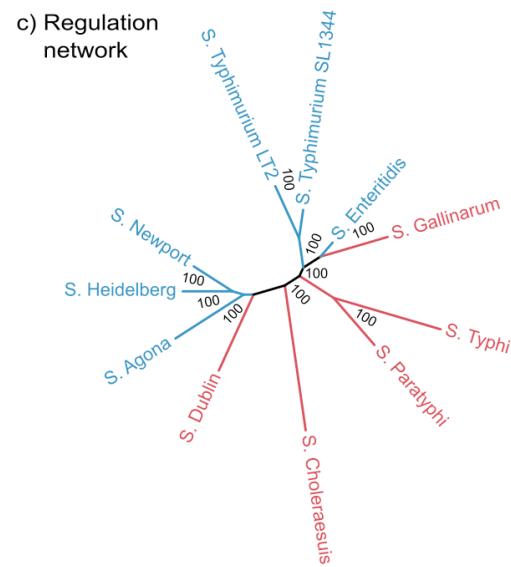
a) Genomic SNPs



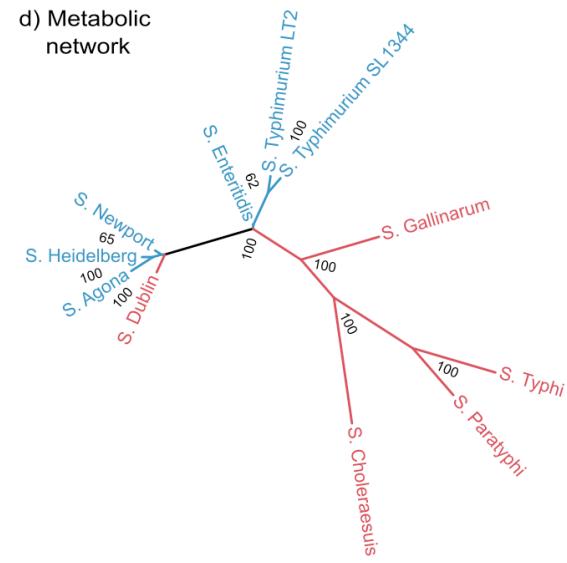
b) PPI network

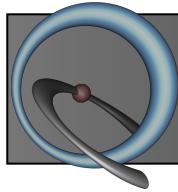


c) Regulation network



d) Metabolic network





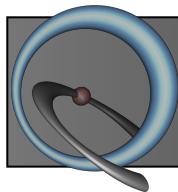
Balázs Vedelek



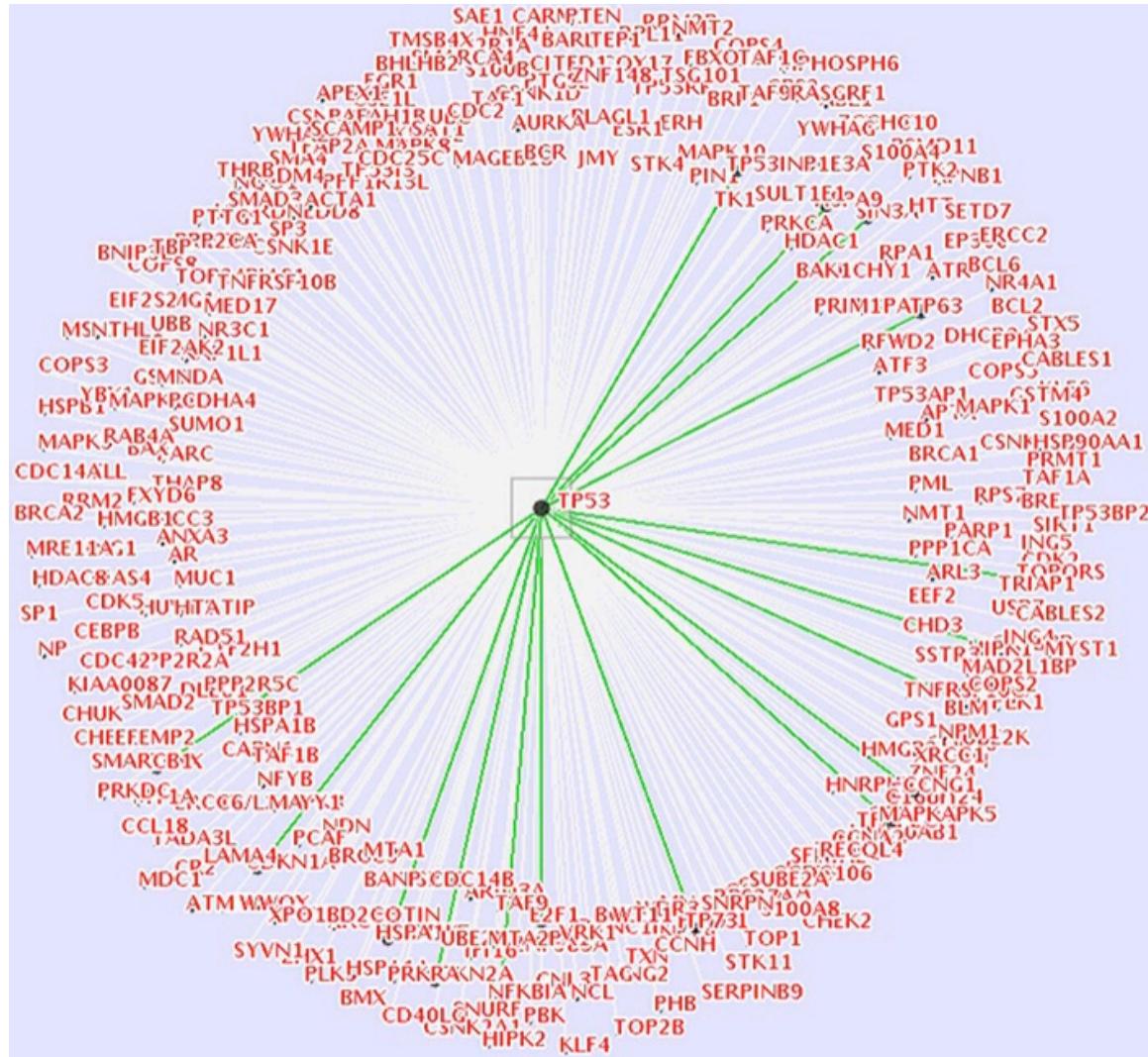
Protein-protein interactions in my project

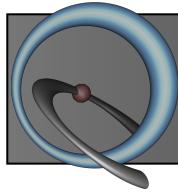
Protein-protein interactions in my mind (brain)

The image shows a man from behind, wearing a red t-shirt, looking at a computer screen. On the screen, there is a Western blot image on the left and a diagram of protein-protein interactions on the right. Below the screen, there is a collage of four images: a fluorescence micrograph of neurons, a detailed diagram of a synaptic vesicle, a diagram of a neuron with various labeled parts, and a network graph representing a brain. The overall theme is the study of protein-protein interactions in the context of neuroscience.

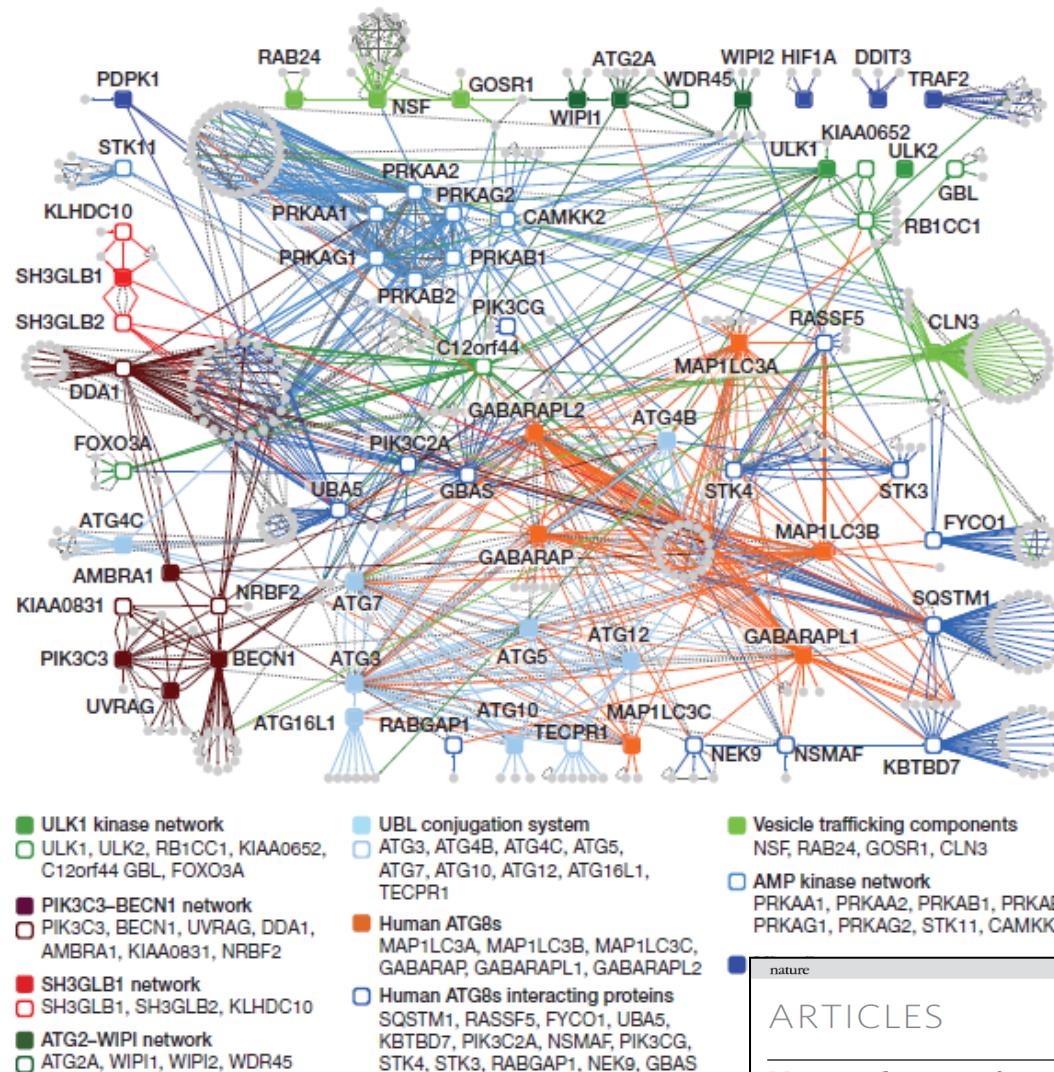


Fichó Erzsébet





Hegedűs Krisztina



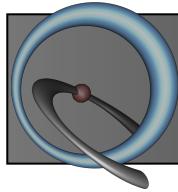
nature

Vol 466 | 1 July 2010 | doi:10.1038/nature09204

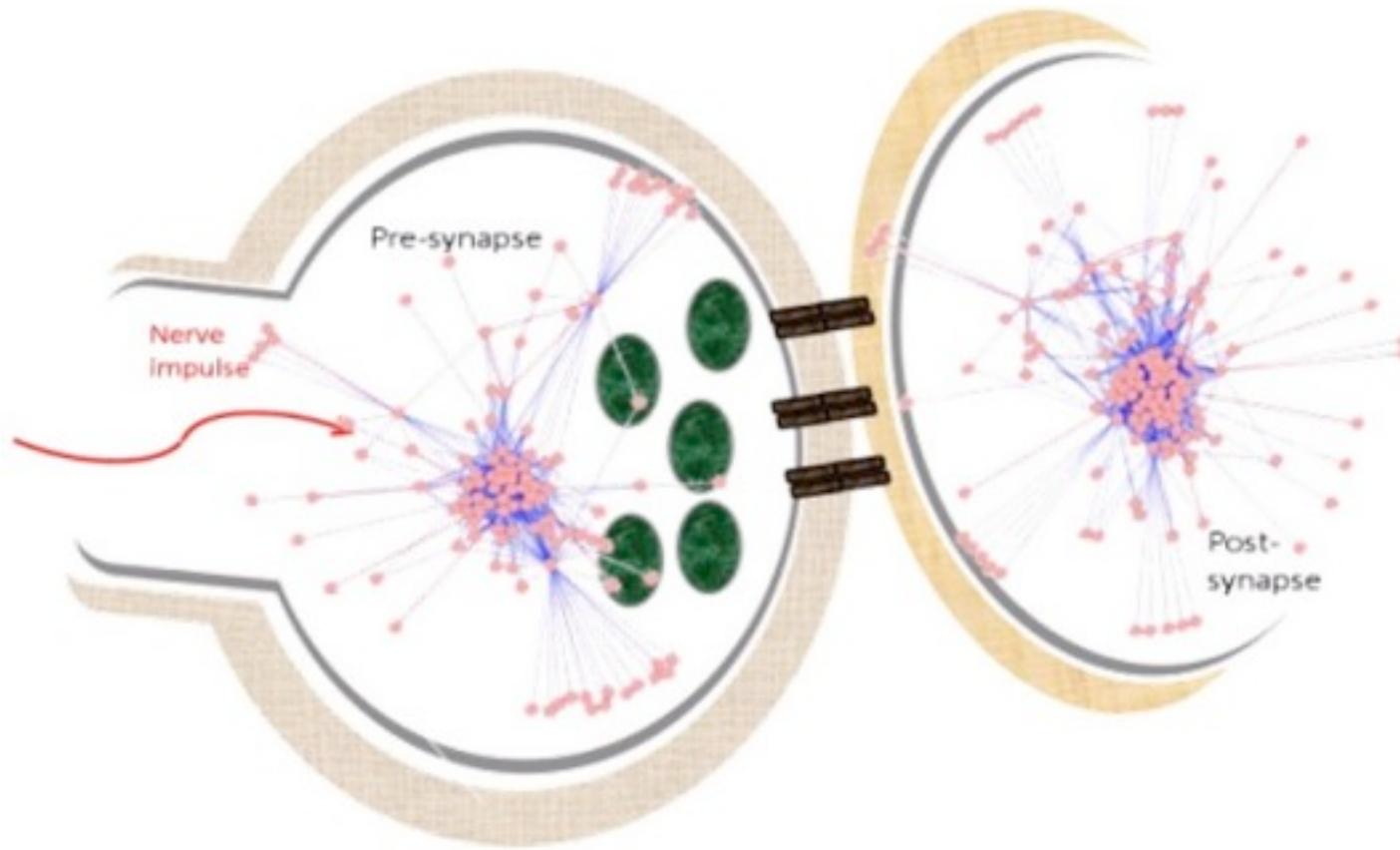
ARTICLES

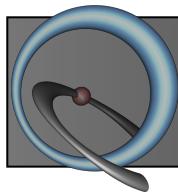
Network organization of the human autophagy system

Christian Behrends¹, Mathew E. Sowa¹, Steven P. Gygi² & J. Wade Harper¹

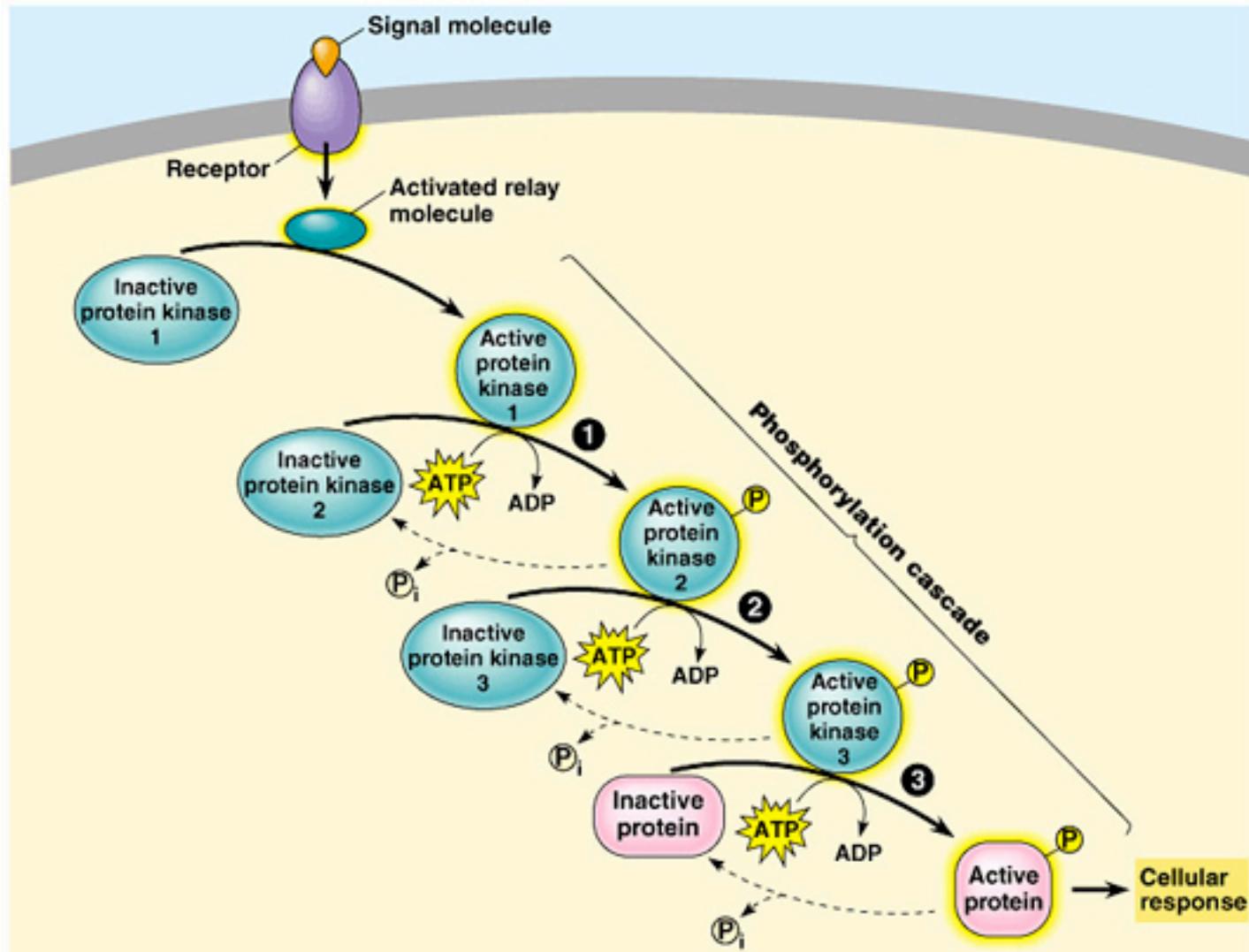


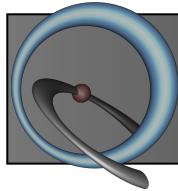
Annamária Kiss-Tóth



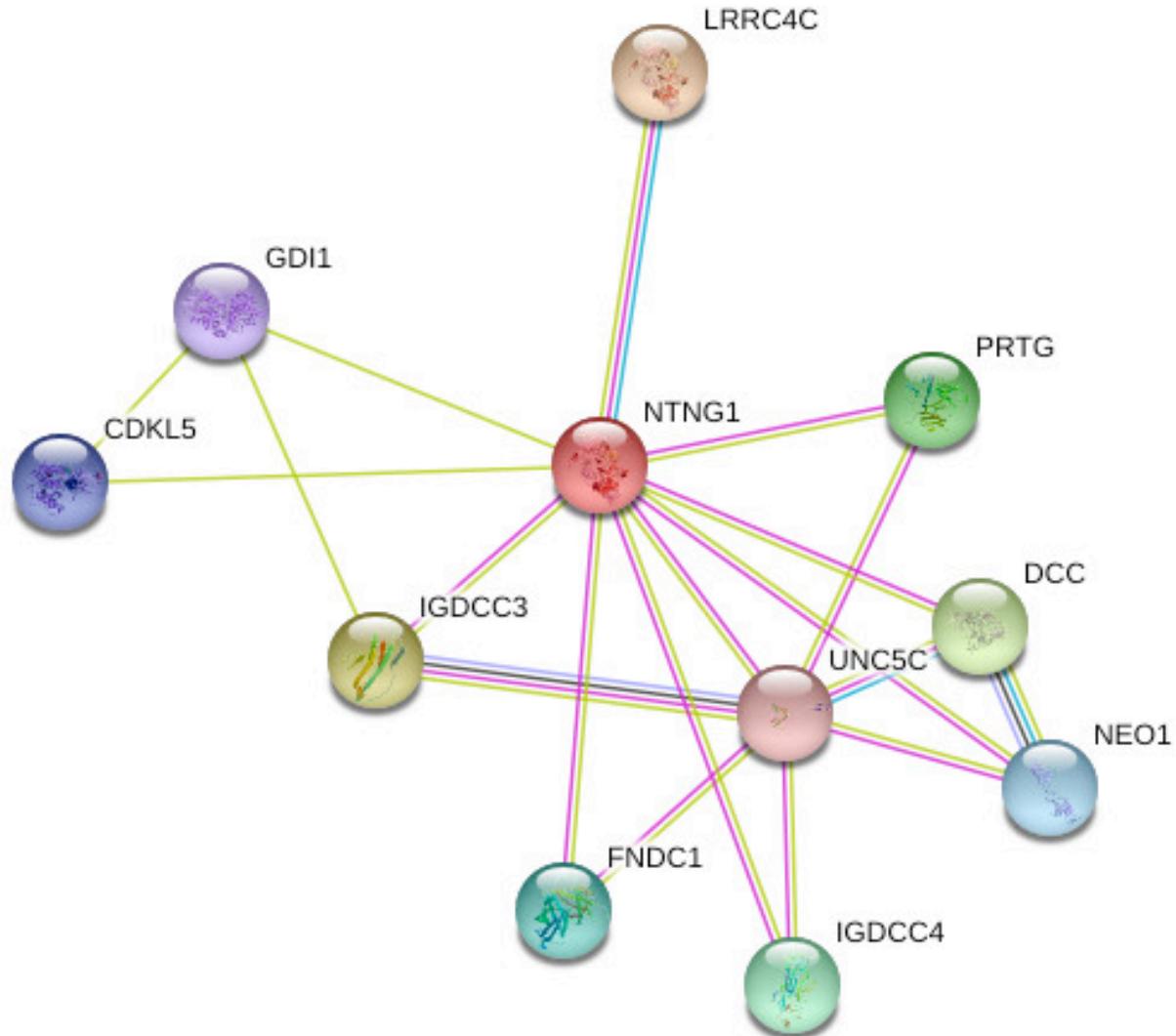


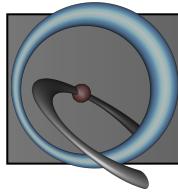
Brigitta Kállai



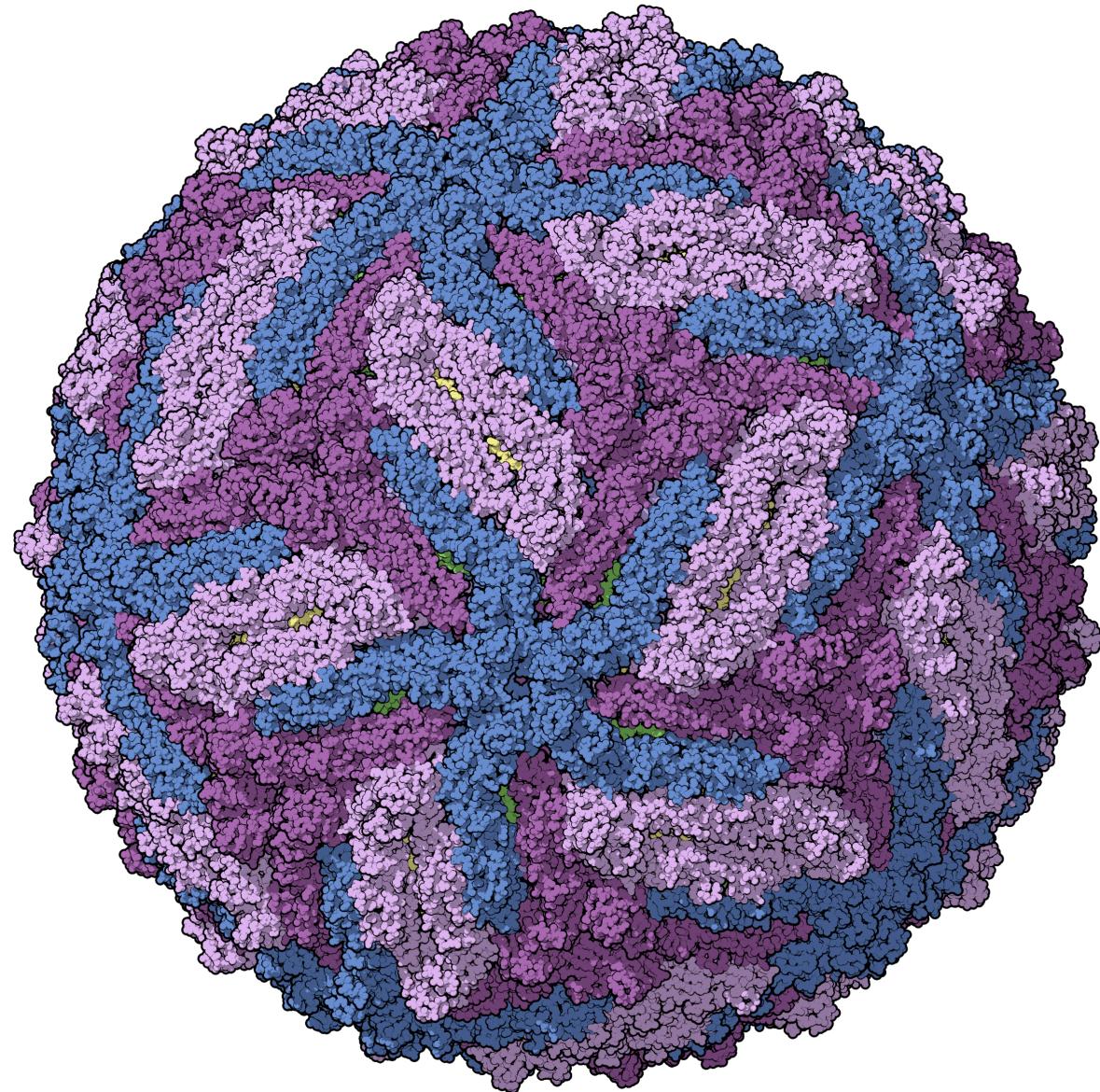


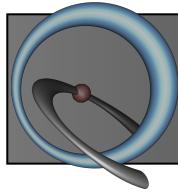
Laura Laszlo



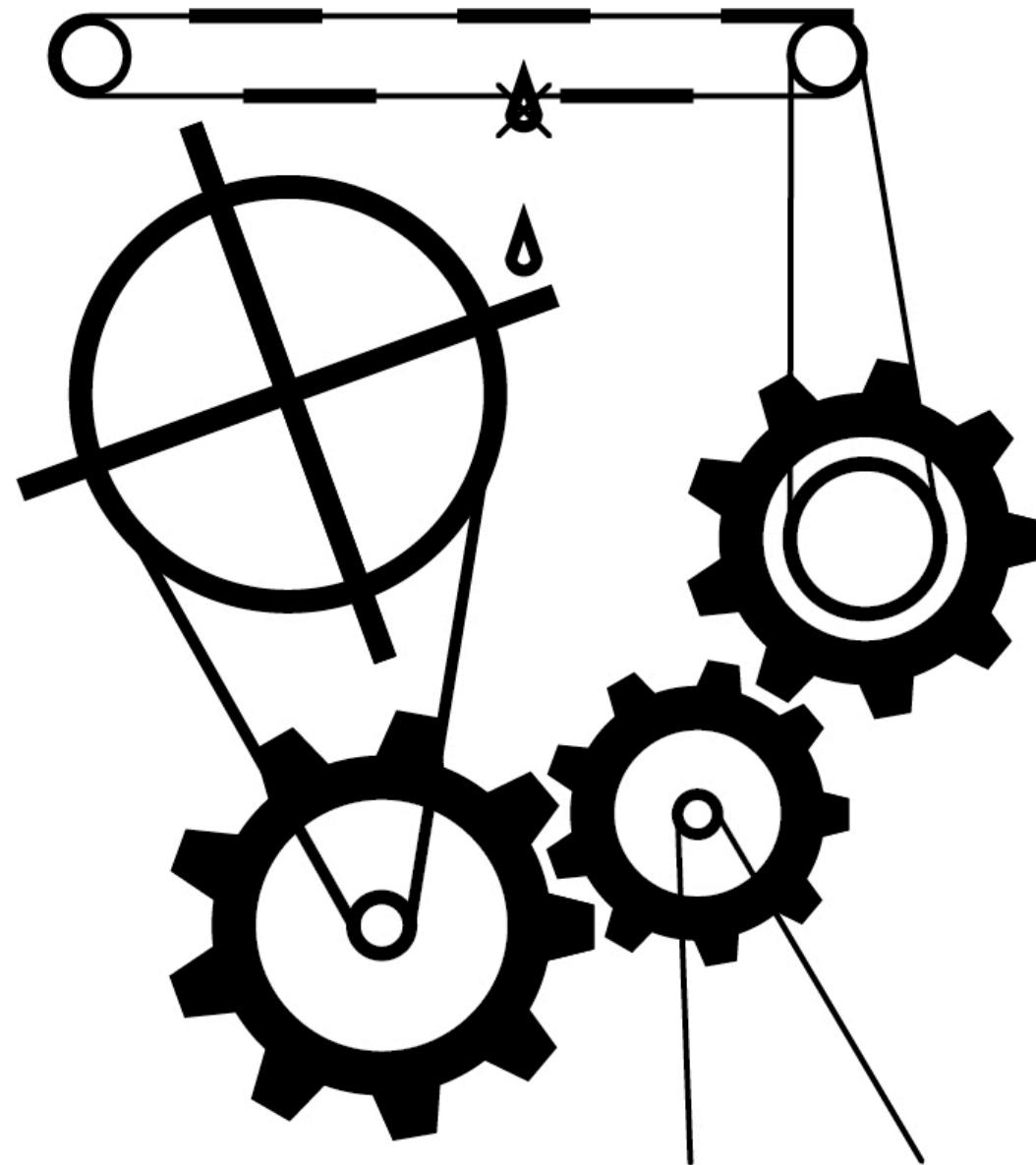


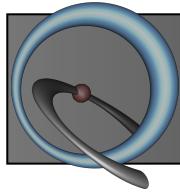
Balázs Szalkai





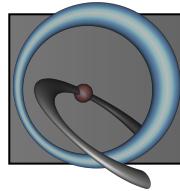
Bence Széplaki





Others?

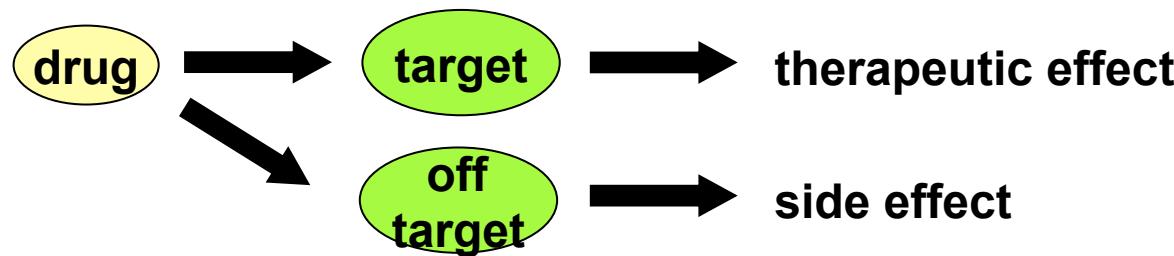




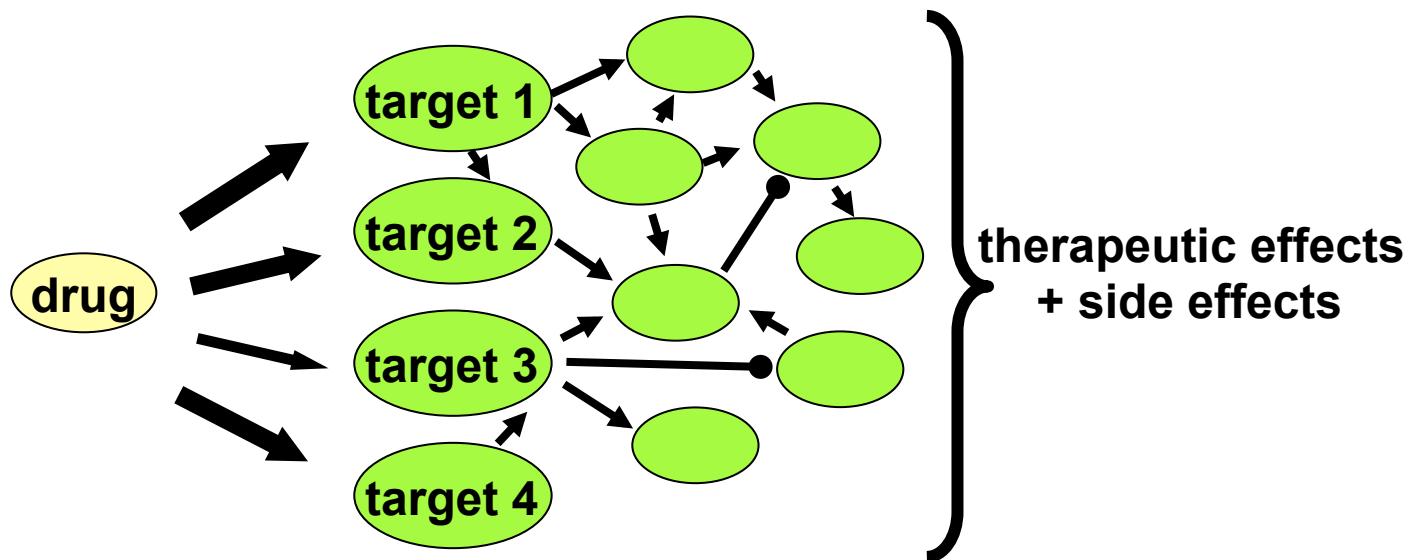
Why

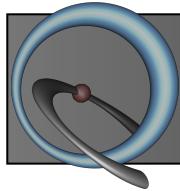


classic view of drug action



network view of drug action

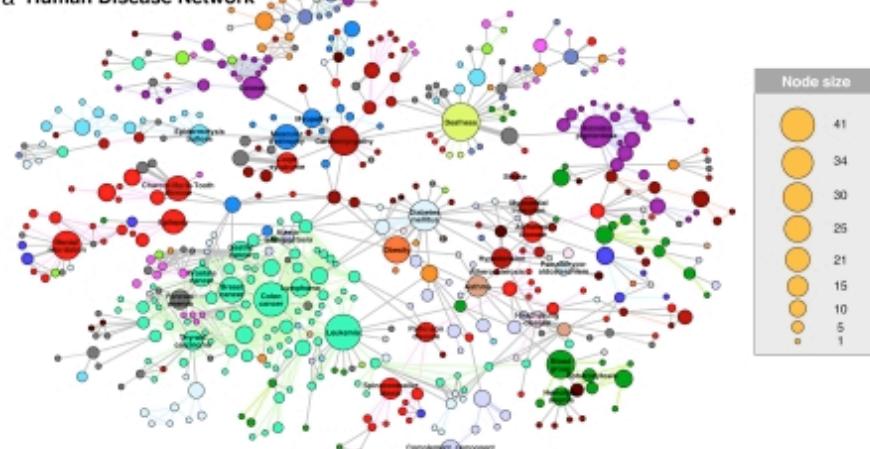




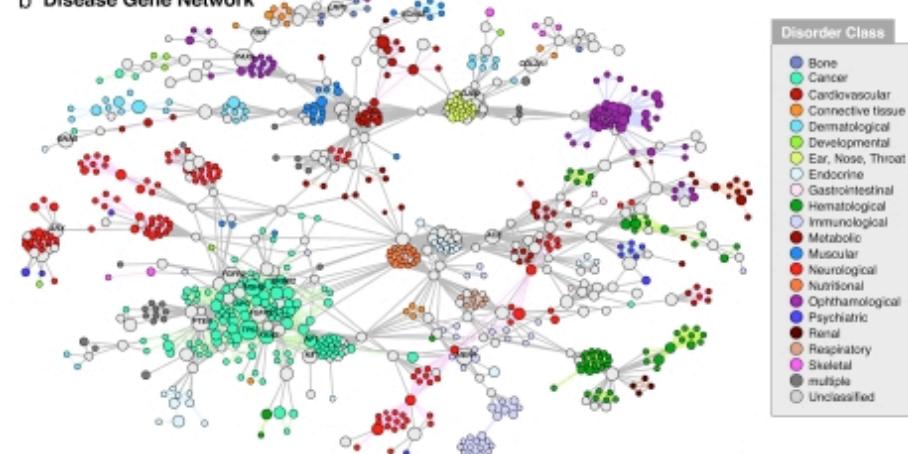
Why



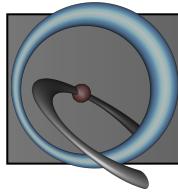
a Human Disease Network



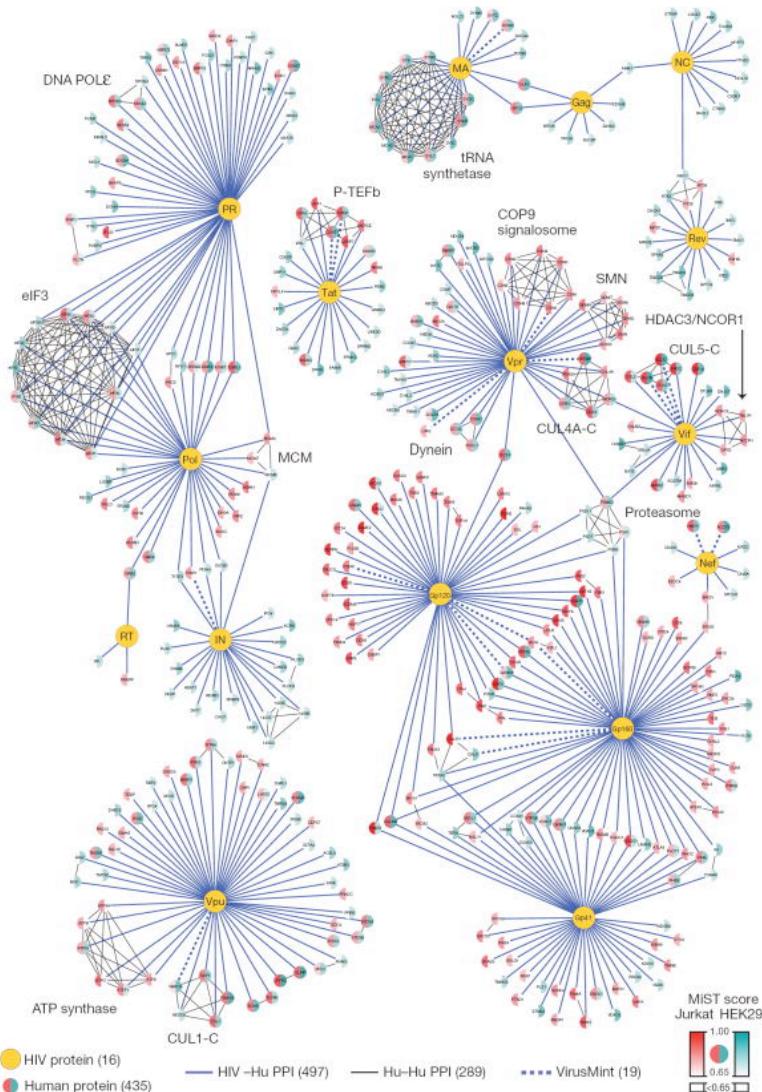
b Disease Gene Network



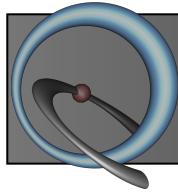
Proc Natl Acad Sci U S A. 2007 May 22;104(21):8685-90. Epub 2007 May 14.
The human disease network.
Goh KI¹, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL.



Why



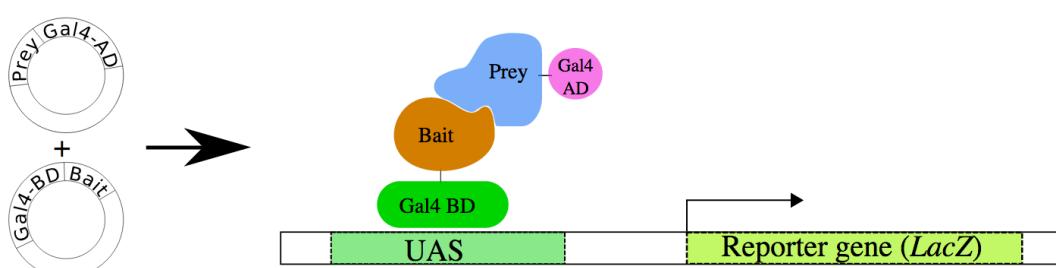
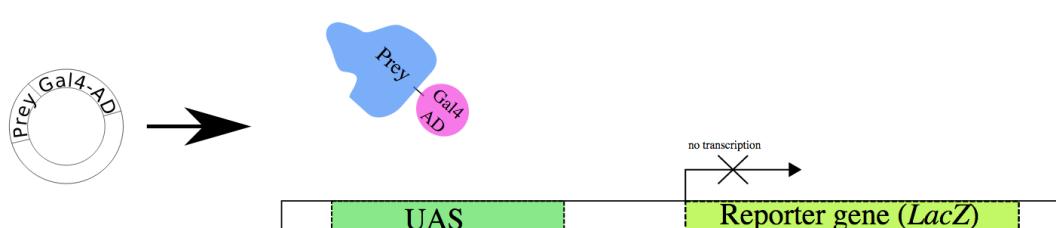
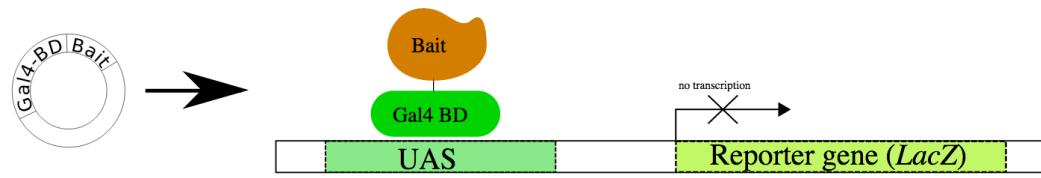
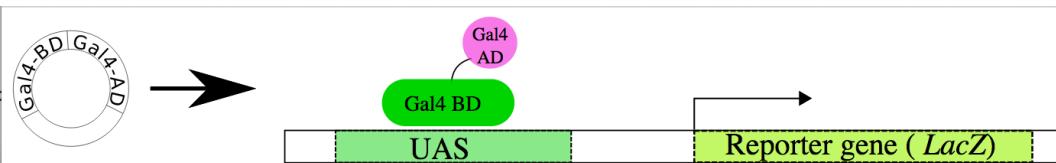
Nature. 2011 Dec 21;481(7381):365-70. doi: 10.1038/nature10719.
Global landscape of HIV-human protein complexes.
Jäger S et. al

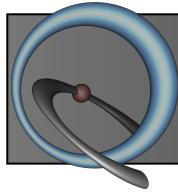


Experimental Techniques



- High - Y



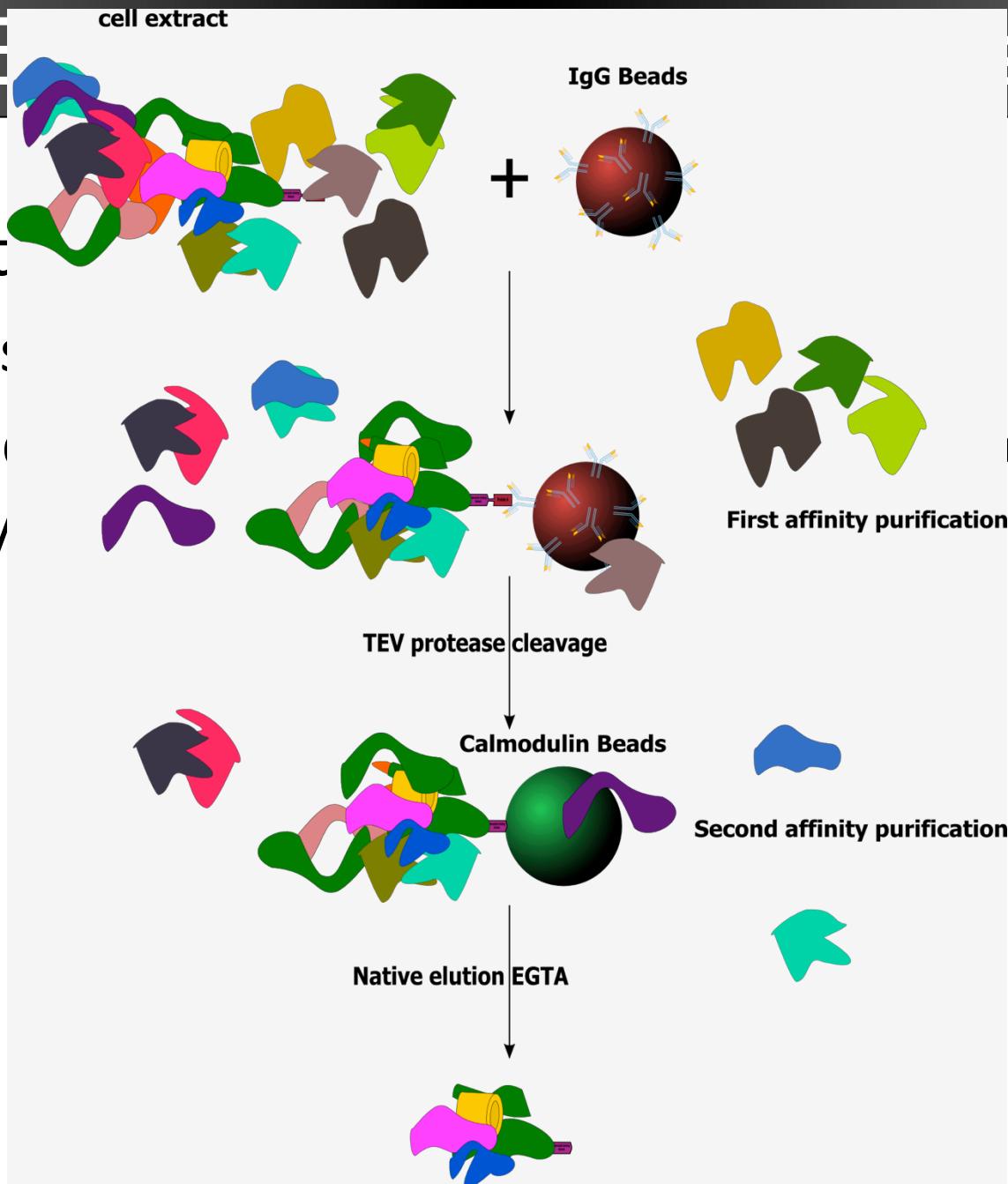


Experimental Techniques

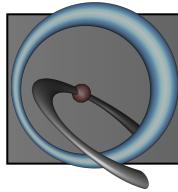


- High throughput
 - Yeast 2 hybrid
 - Pros:
 - Relatively low-tech
 - Scalable
 - Cons:
 - High false positive and false negative rate

- High titer
- Yeast
- Tandem
(AP)



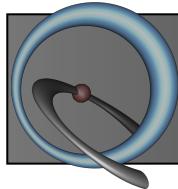
ometry



Experimental Techniques



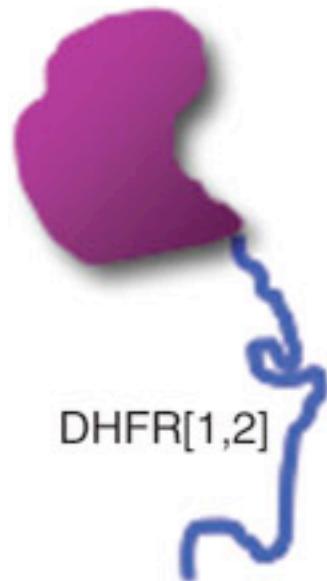
- High throughput
 - Yeast 2 hybrid
 - Tandem Affinity Purification/Mass Spectrometry (AP/MS)
 - Pros:
 - Quantitative results *in vivo*
 - Cons:
 - Tag can obscure binding
 - Protease can cleave protein
 - Not good for transient interactions



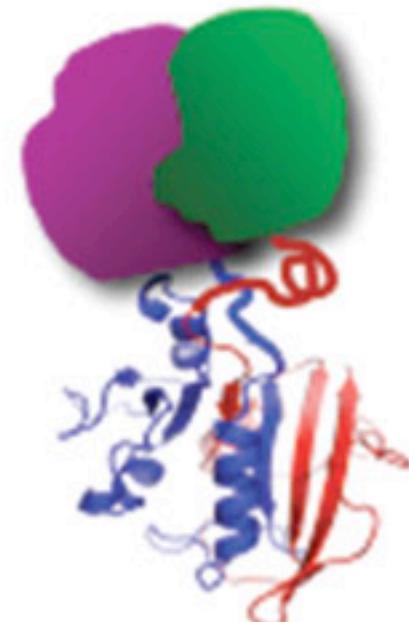
Experimental Techniques



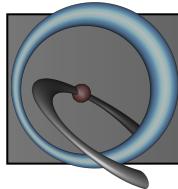
Protein X



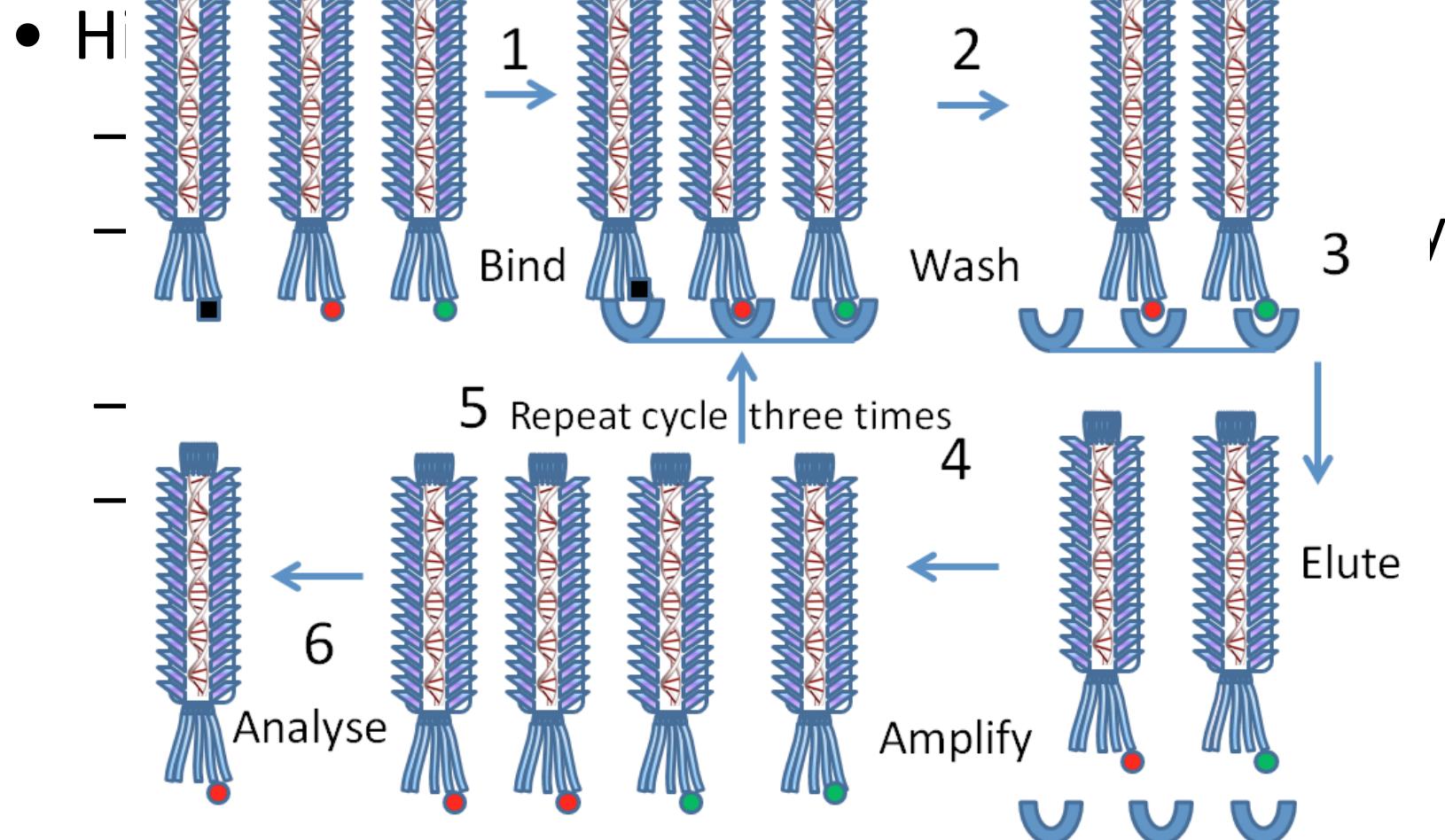
Protein Y

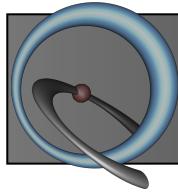


Reconstituted
enzyme
activity



Experimental Techniques

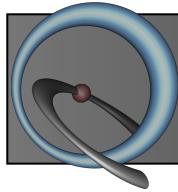




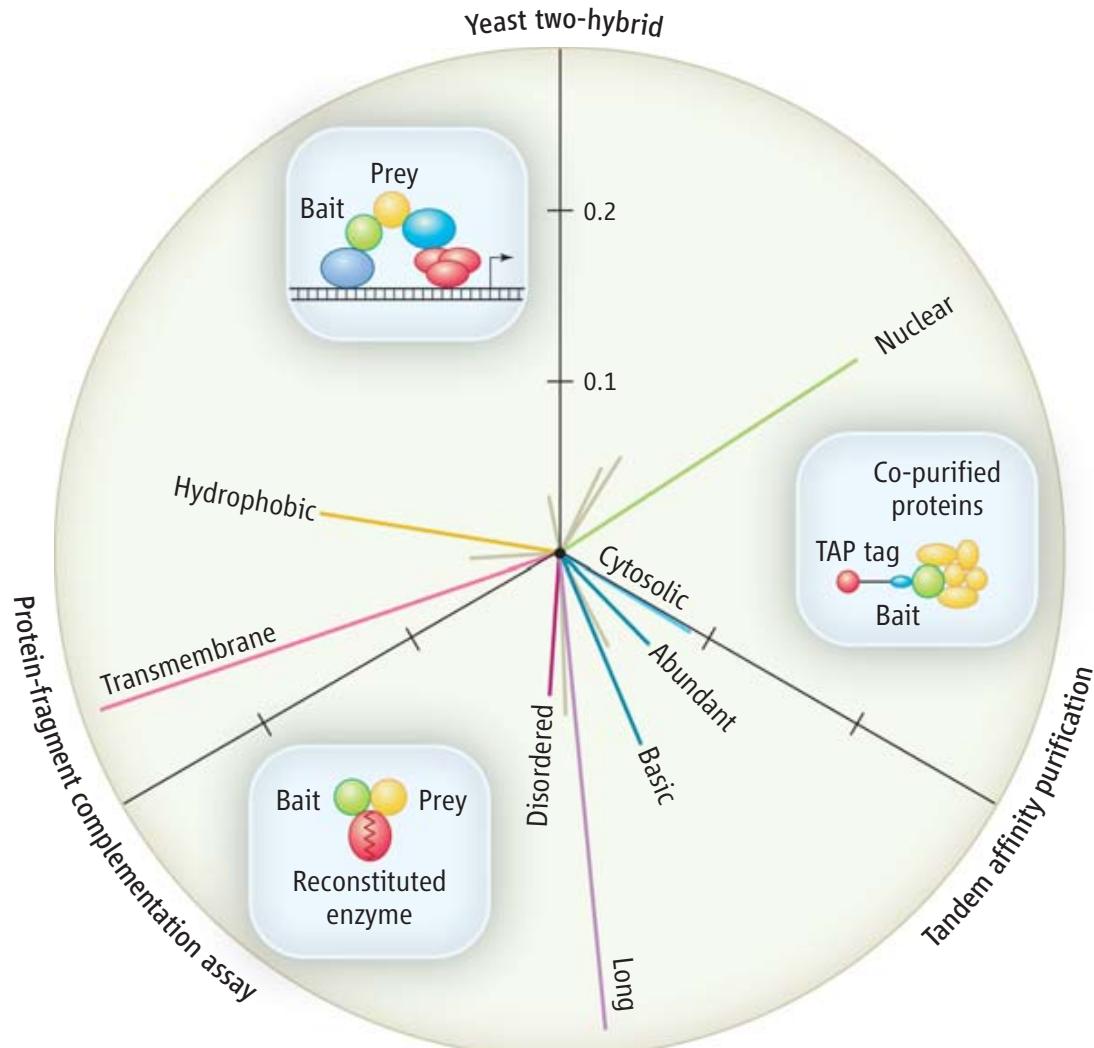
Experimental Techniques



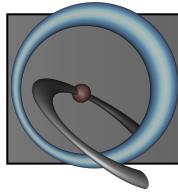
- High throughput
 - Yeast 2 hybrid
 - Tandem Affinity Purification/Mass Spectrometry (AP/MS)
 - Protein-fragment complementation assays
 - Phage display
 - Protein microarrays
 - Chemical crosslinking



Experimental Techniques



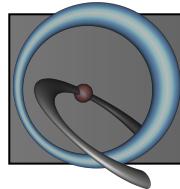
Jensen & Bork, *Science*, 2008



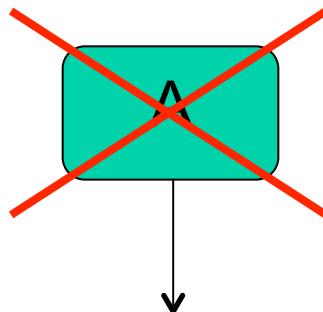
Experimental Techniques



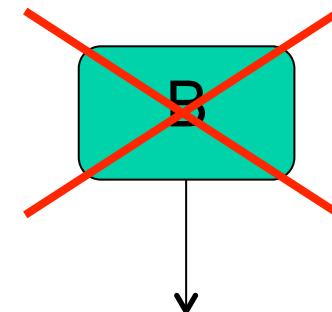
- Genetic interactions
 - Measures functional linkage between genes
 - Orthogonal to physical interaction techniques
 - Two approaches
 - SGA (Synthetic genetic array)
 - E-MAP (Epistatic MiniArray Profile)



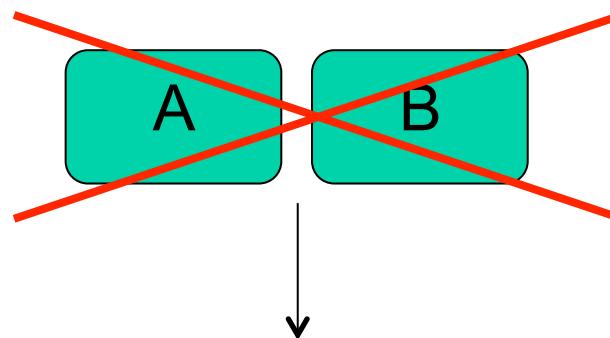
Experimental Techniques



Growth_A

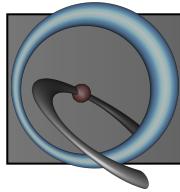


Growth_B



Growth_{AB}

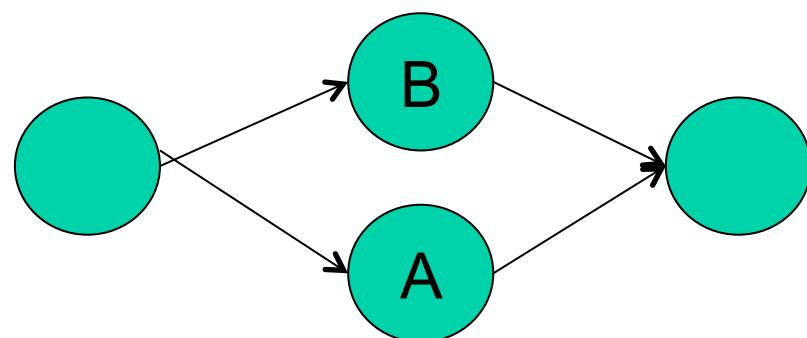
$$\text{Interaction}_{AB} = \text{Growth}_{AB} - \text{Growth}_A * \text{Growth}_B$$



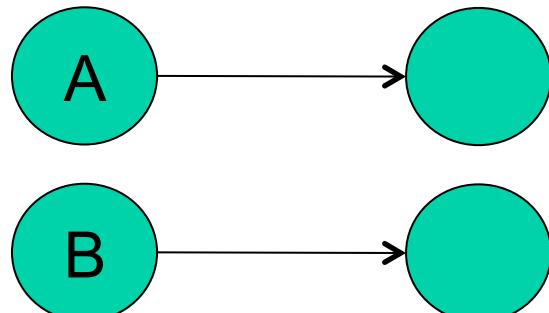
Experimental Techniques



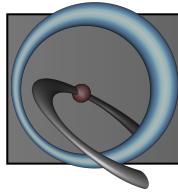
+



-



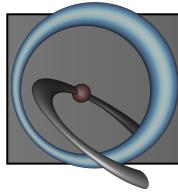
0



Experimental Techniques



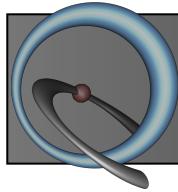
- Genetic Interactions (E-MAP)
 - High throughput
 - Quantitative
 - Can provide corroboration to PPI data



Experimental Techniques



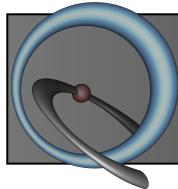
- Low throughput
 - Xray crystallography
 - NMR
 - SAXS
 - ELISA Binding Assays
 - Co-IP



Computational Techniques



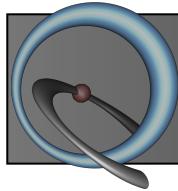
- Text mining
 - Search literature for references to protein-protein interactions
- Orthology
 - Predicts interaction based on orthologous pairs in another species
- Domain pairs
 - Predicts interaction based on domains interacting in other proteins



Public Repositories



- Pathway data
 - [WikiPathways](#)
 - [Reactome](#)
 - [BioCyc](#)
 - [KEGG](#)
 - [Pathway Commons](#)



Public Repositories



- PPI Data
 - [Pathway Commons](#)
 - [BioGRID](#)
 - [IntAct](#)
 - Model organisms databases (e.g. [SGD](#))
 - [STRING](#)



Scooter Morris

June 6, 2016
Eötvös Loránd University
Budapest, Hungary



Outline

- Biological Networks
 - Why Networks?
 - Biological Network Taxonomy
 - Analytical Approaches
 - Visualization
- Introduction to Cytoscape
- Hands on Tutorial
 - Data import
 - Layout and apps



Introductions

- John “Scooter” Morris
 - 2010-Current
 - Adjunct Assistant Professor, Pharmaceutical Chemistry
 - 2004-Current
 - Director, NCRR Resource for Biocomputing, Visualization, and Informatics (RBVI) @ UCSF
 - 1985-2004
 - Principal Systems Architect: Genentech, Inc.
 - Cytoscape core team since 2006
 - Author or co-author of several Cytoscape apps
 - *structureViz2, clusterMaker2, chemViz2, stringApp, CyAnimator*



Introductions

- Nadezhda T. Doncheva
 - 2016-Current
 - Postdoc researcher at the Novo Nordisk Foundation Center for Protein Research & Center for non-coding RNA in Technology and Health in Copenhagen
 - 2010-2016:
 - Research scientist and PhD student at the Max Planck Institute for Informatics in Saarbrücken
 - Author or co-developer of several Cytoscape plugins:
 - RINalyzer, structureViz2, setsApp, NetworkPrioritizer, NetworkAnalyzer,



Introductions

- Barry Demchak
 - 2012-Current
 - Chief Software Architect, Project Manager for National Resource for Network Biology (NRNB, Ideker Lab)
 - 2005-2012
 - PhD Computer Science and Engineering, UC San Diego
 - 1987-current
 - President, Torrey Pines Software, Inc
 - Cytoscape core team since 2012
 - Architect of Cytoscape Cyberinfrastructure



Introductions

- Christian Zmasek
 - 2015-Current
 - Cytoscape core team (Ideker Lab)
 - 2006-2015
 - Postdoc (Sanford-Burnham) Comparative functional genomics
- Rintaro Saito, PhD
 - 2014-Current
 - Associate Project Scientist (Kumar Lab)
 - 2011-2014
 - Visiting Assistant Professor (Ideker Lab)
 - 2002-2011
 - Assistant Professor (Keio University)



Introductions

- You?
 - Clinician
 - Biologist
 - Bioinformatician
 - Computer Scientist
 - Chemist
 - Other?



Installation

- How many have installed:
 - Cytoscape 3.4.0
 - Apps:
 - Omics Analysis Collection



Why Networks?

- **Networks are...**
 - Commonly understood
 - Structured to reduce complexity
 - More efficient than tables

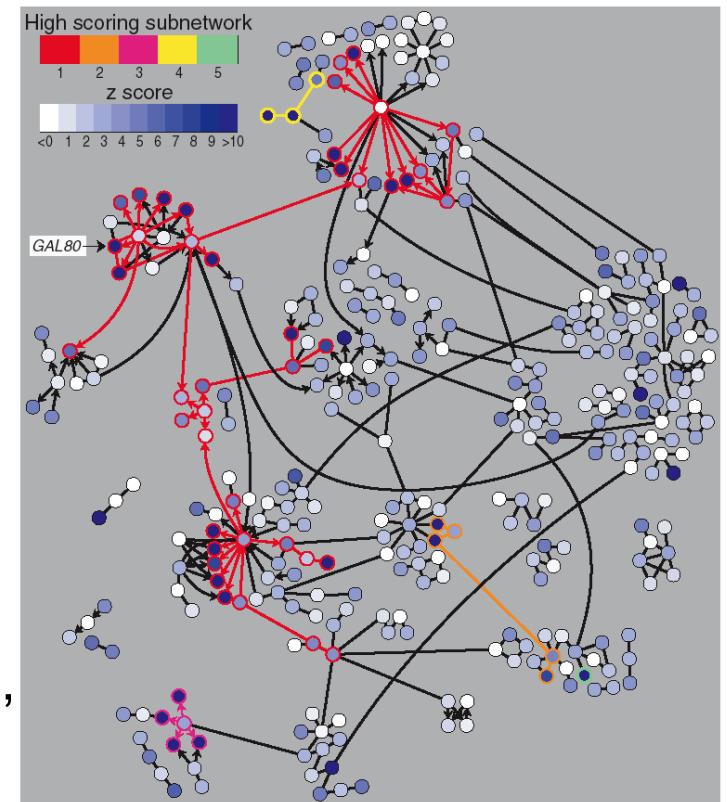
- **Network tools allow...**

Analysis

- Characterize network properties
- Identify hubs and subnets
- Classify, quantify and correlate, e.g., cluster nodes by associated data

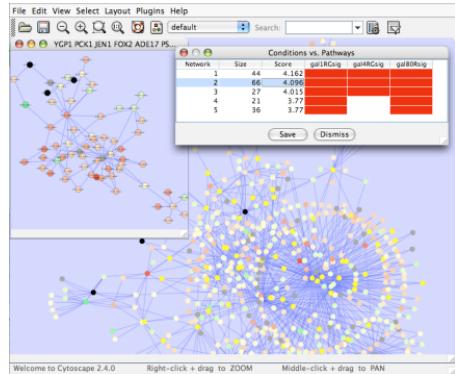
Visualization

- Explore data overlays
- Interpret mechanisms, e.g., how a process is modulated or attenuated by a stimulus





Applications of Network Biology



jActiveModules, UCSD

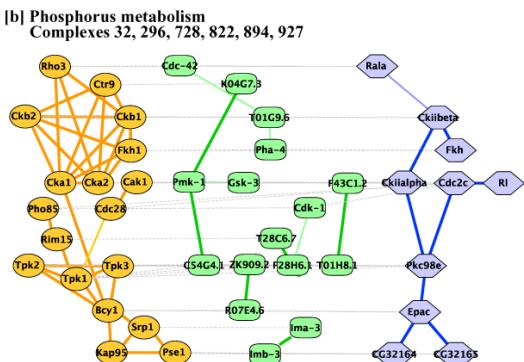
- **Gene Function Prediction** shows connections to sets of genes/proteins involved in same biological process

- **Detection of protein complexes/subnetworks** discover modularity & higher order organization (motifs, feedback loops)



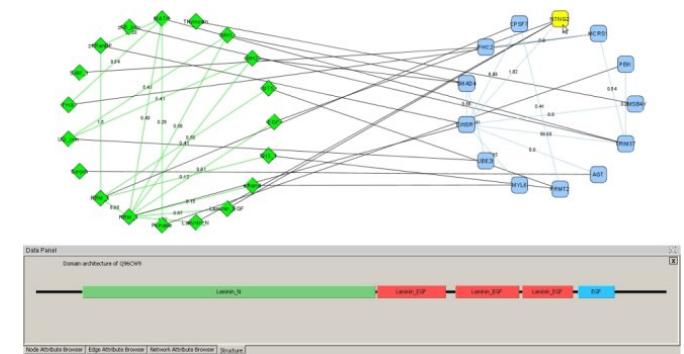
mCode, University of Toronto

- **Network evolution** biological process(s) conservation across species



PathBlast, UCSD

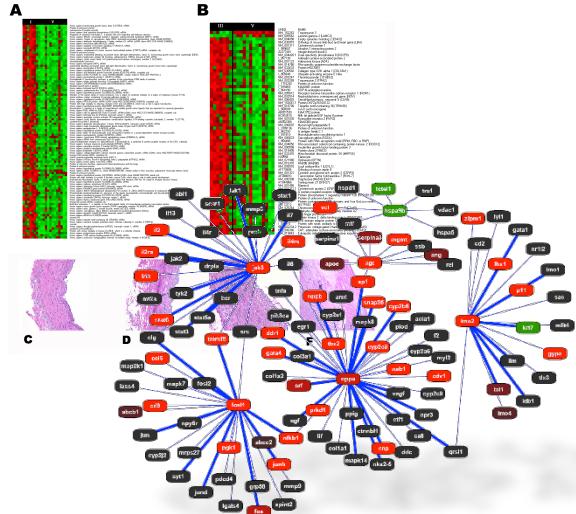
- **Prediction of interactions & functional associations** statistically significant domain-domain correlations in protein interaction network to predict protein-protein or genetic interaction



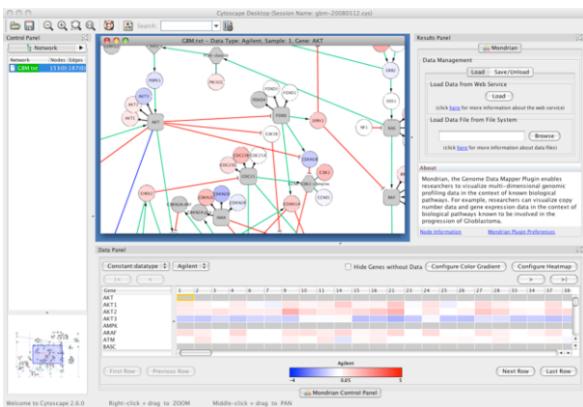
DomainGraph, Max Planck Institute



Applications in Disease

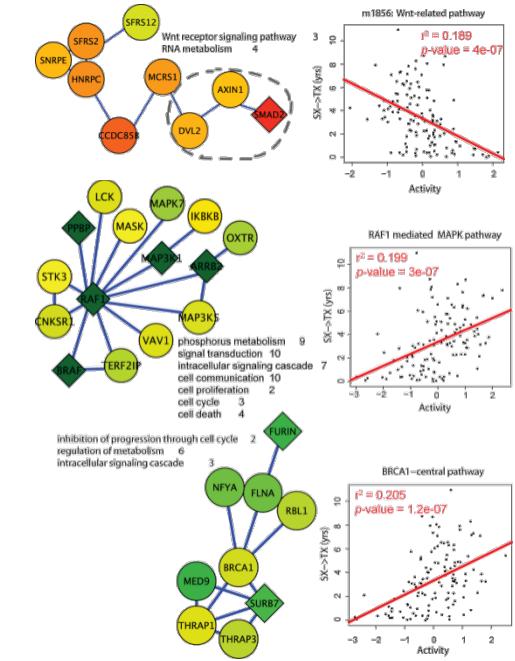


Agilent Literature Search



Mondrian, MSKCC

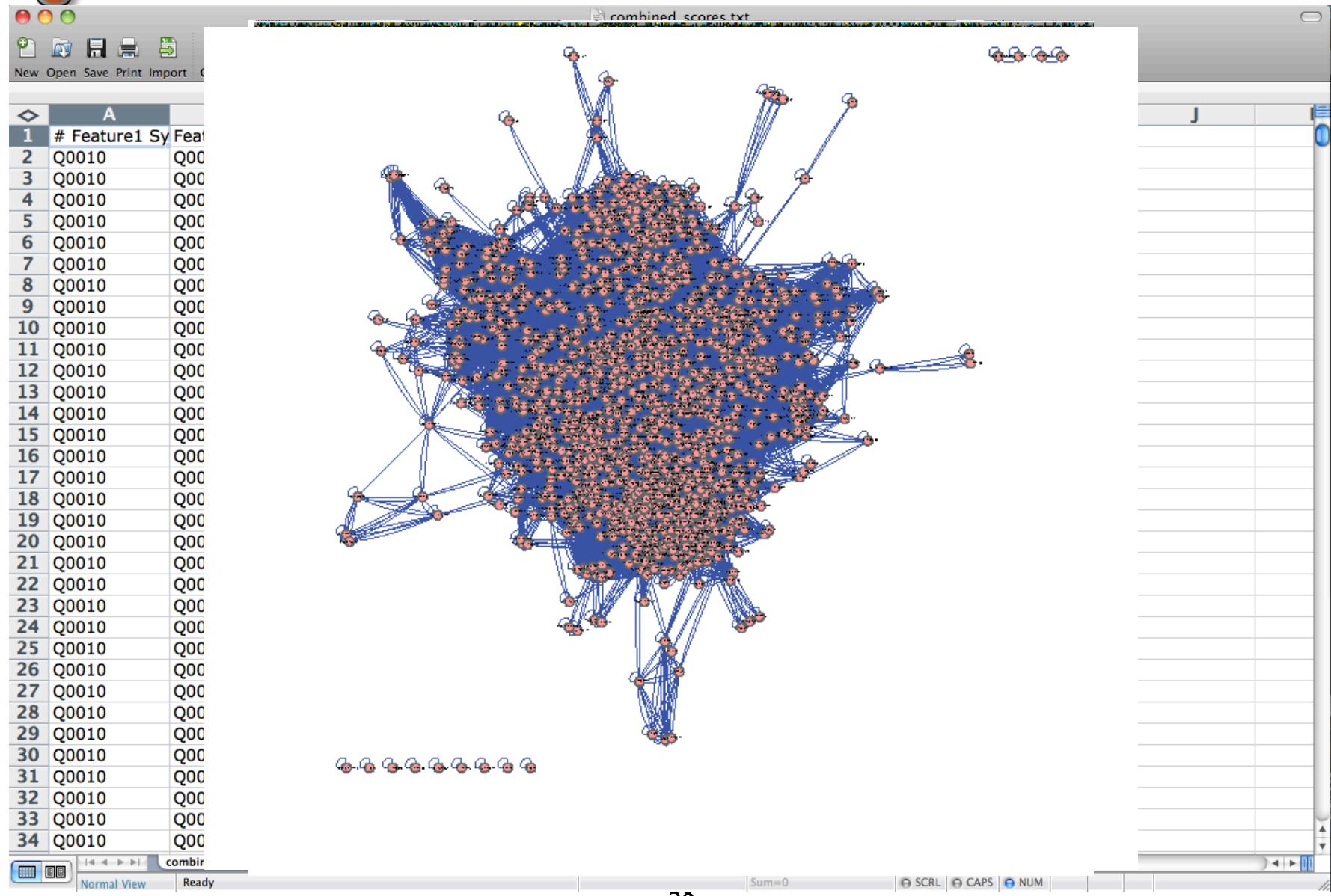
- **Identification of disease subnetworks** – identification of disease network subnetworks that are transcriptionally active in disease.
- **Subnetwork-based diagnosis** – source of biomarkers for disease classification, identify interconnected genes whose aggregate expression levels are predictive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



PinnacleZ, UCSD



The Challenge



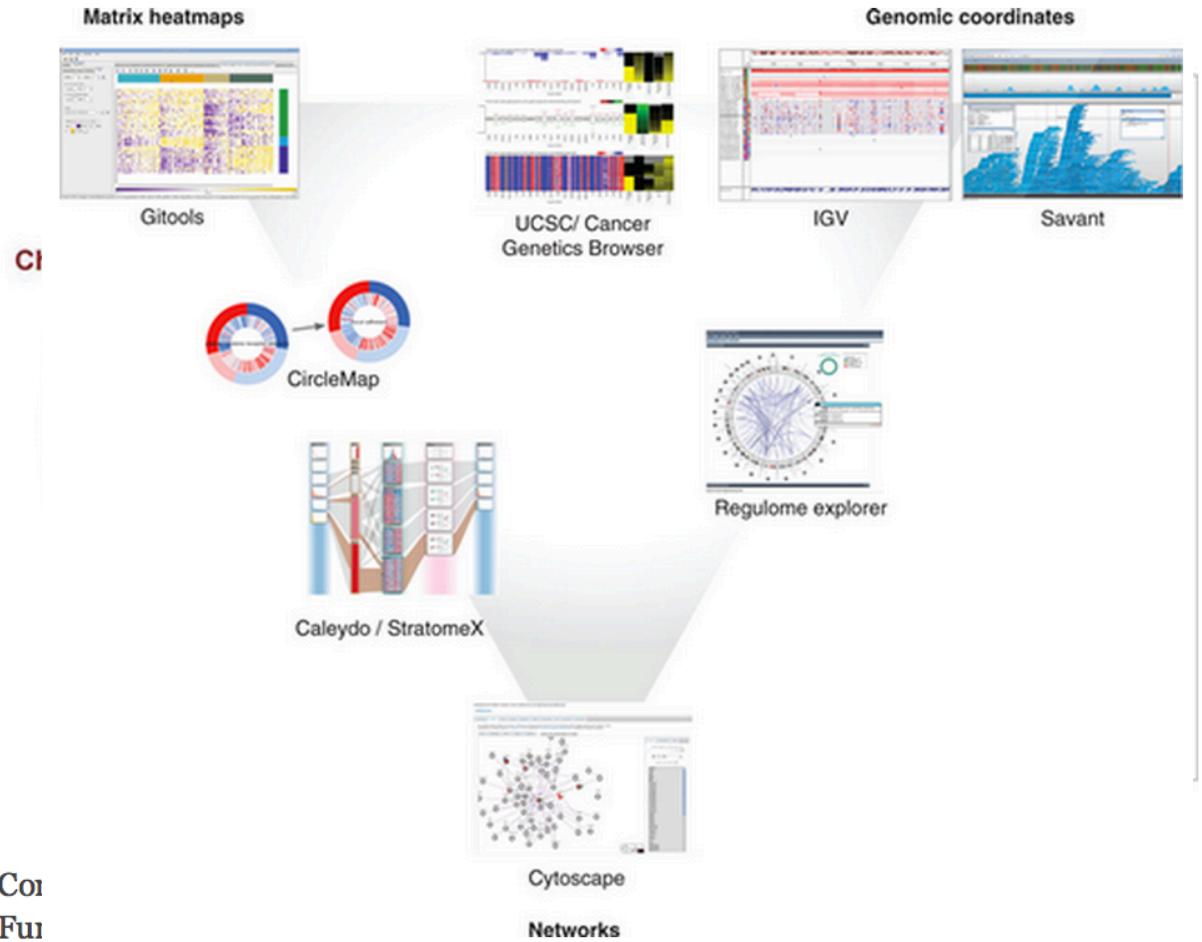


The Challenge

- Biological networks
 - Seldom tell us anything by themselves
 - **Analysis** involves:
 - Understanding the characteristics of the network
 - Modularity
 - Comparison with other networks (i.e., random networks)
 - **Visualization** involves:
 - Placing nodes in a meaningful way (layouts)
 - Mapping biologically relevant data to the network
 - Node size, node color, edge weights, etc
 - *...which then allowing for more analysis!*



The Challenge

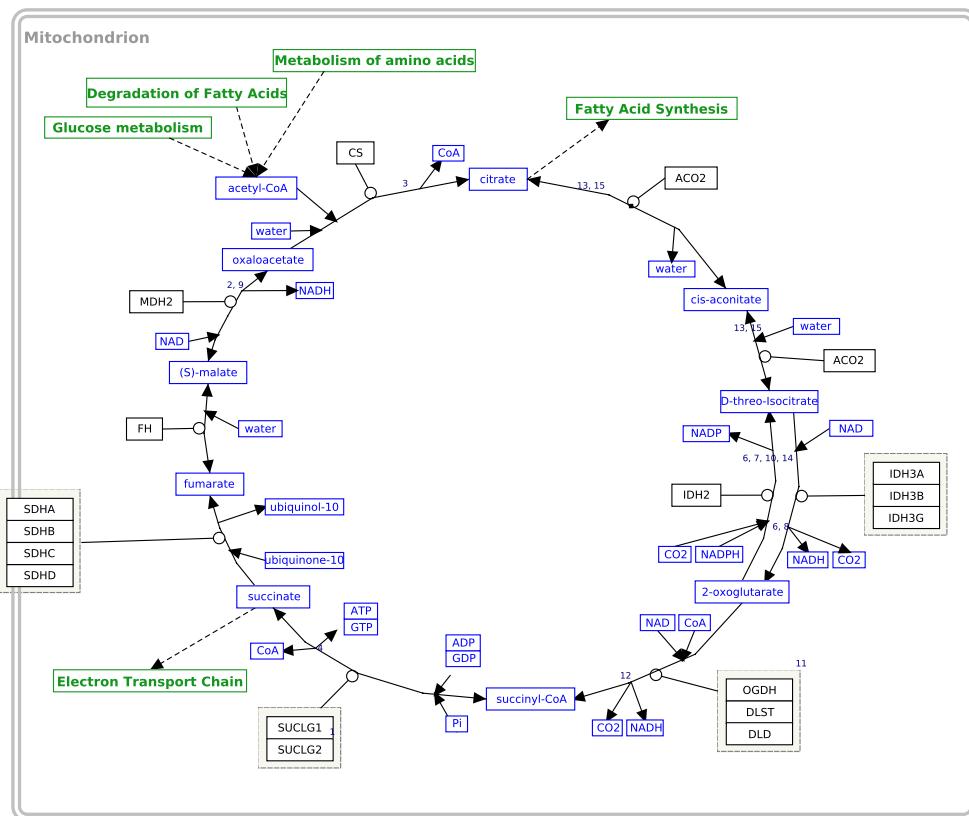


<http://cytoscape-publications.tumblr.com/archive>

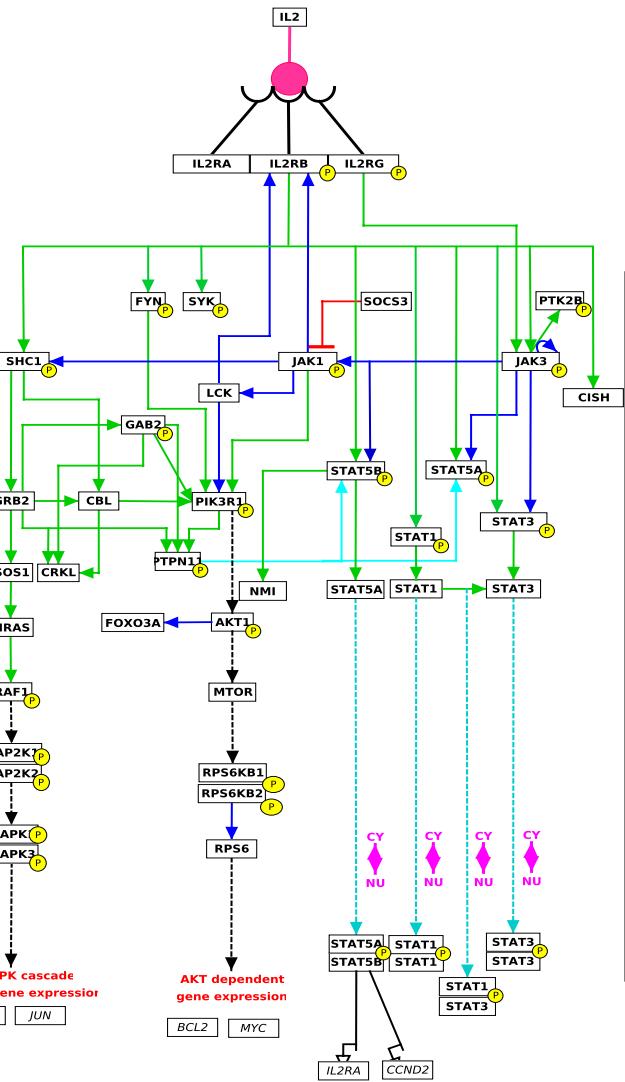


Biological Network Taxonomy

- Pathways
 - Signaling, Metabolic, Regulatory



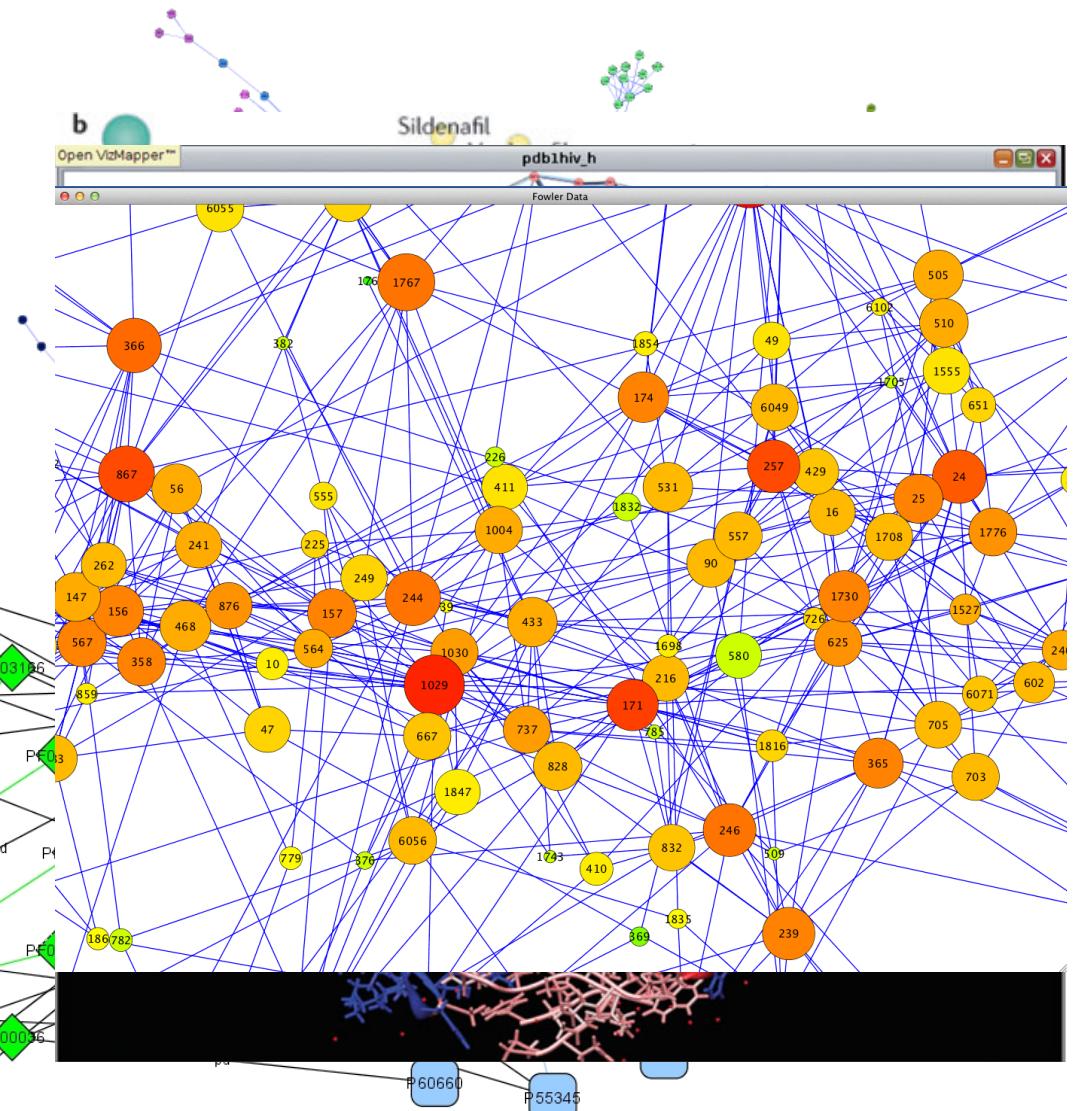
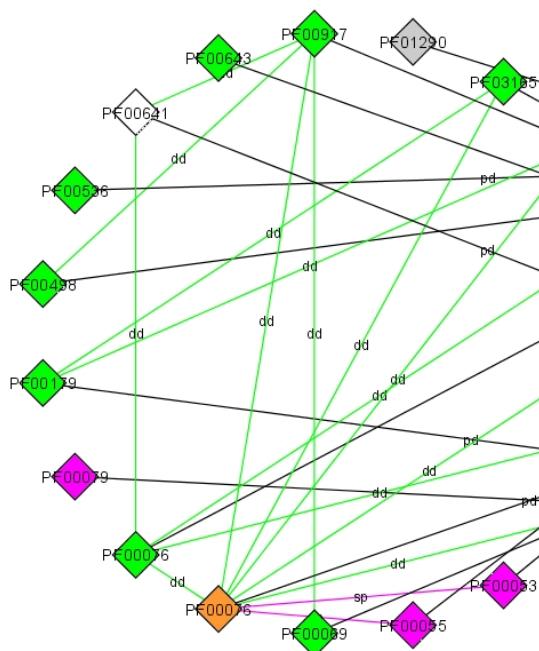
IL-2 Signaling Pathway





Biological Network Taxonomy

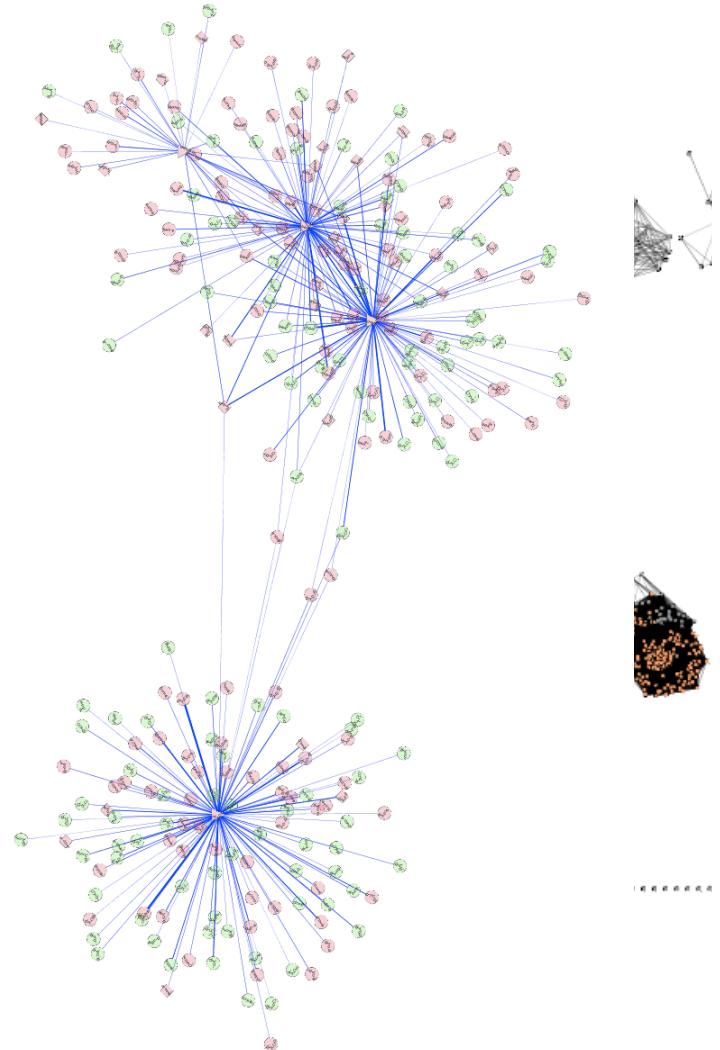
- Interactions
 - Protein-Protein
 - Protein-Ligand
 - Domain-Domain





Biological Network Taxonomy

- Similarity
 - Protein-Protein
 - Chemical similarity
 - Ligand similarity (SEA)
 - Others
 - Tag clouds
 - Topic maps





Biological Network Taxonomy

Where do I get *the* network?

There is no such thing!

550 different interaction databases!

... in 2013





Biological Network Taxonomy

Where do I get *the* network?

There is no such thing!

... in 2016



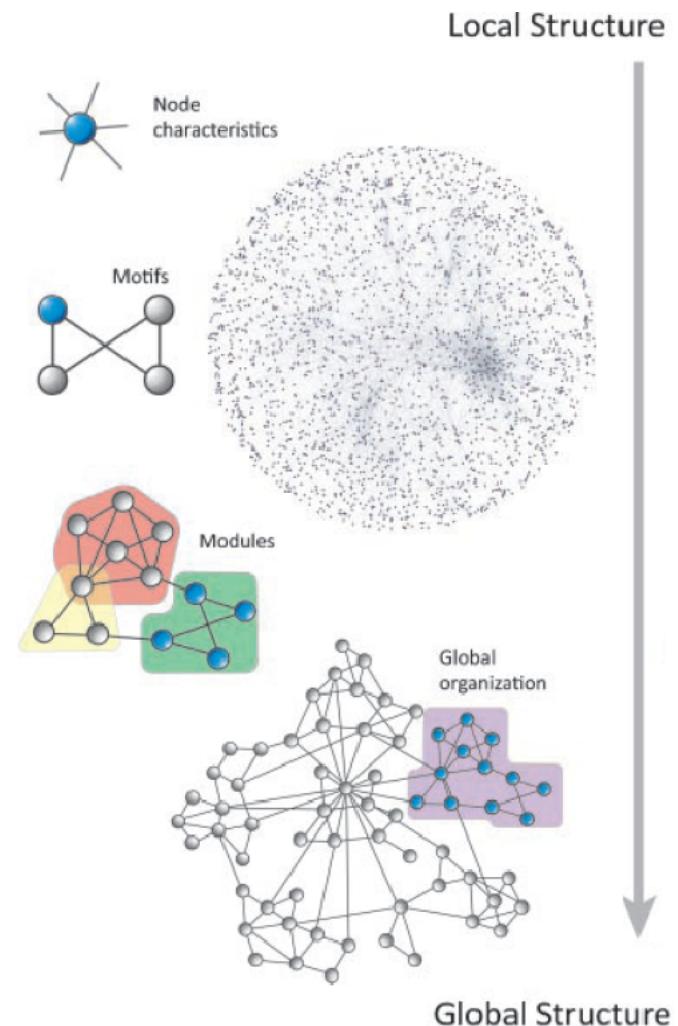
<http://ndexbio.org>



Analytical Approaches

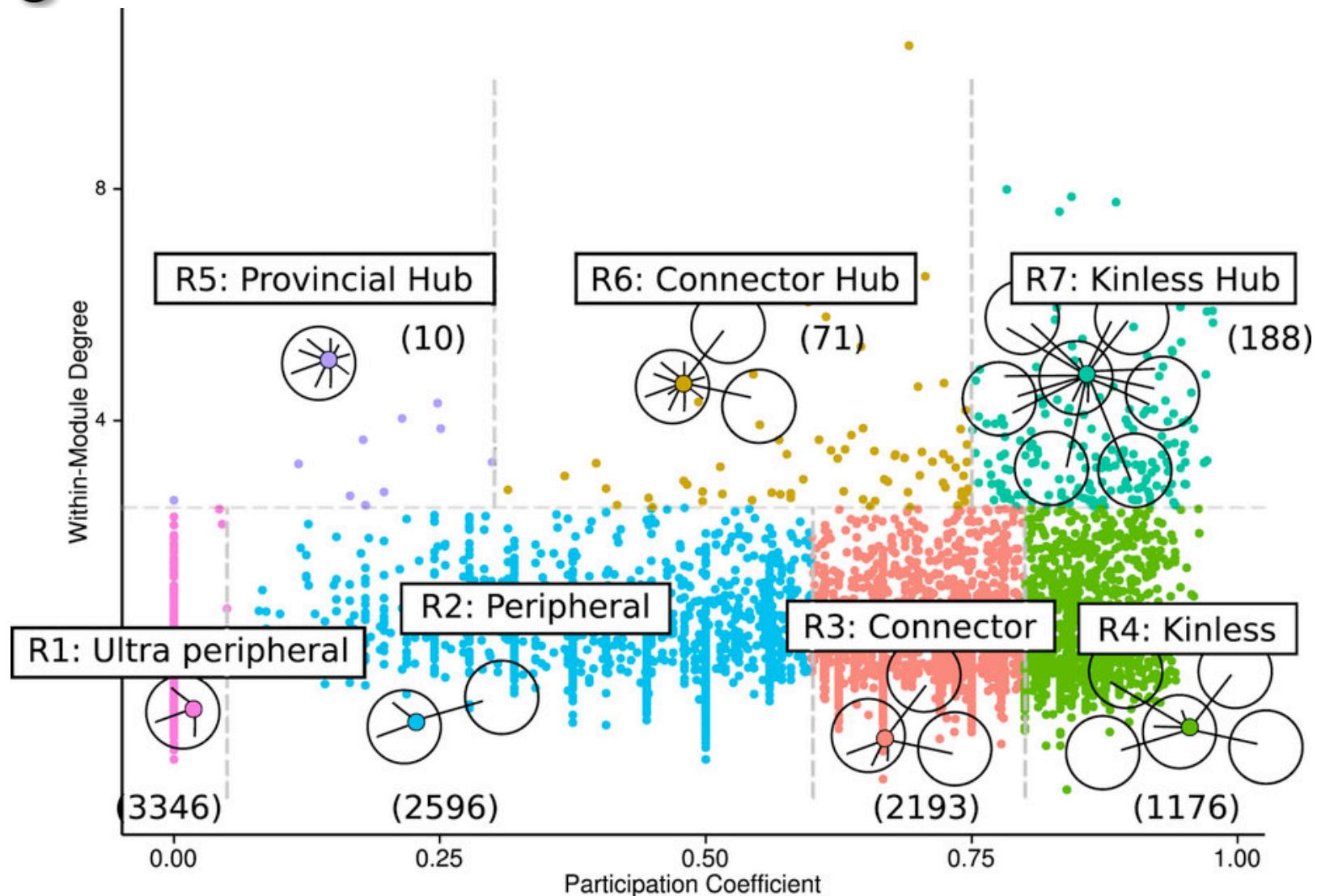
The levels of organization of complex networks:

- ▶ Node **degree** provides information about single nodes
- ▶ Three or more nodes represent a **motif**
- ▶ Larger groups of nodes are called **modules** or **communities**
- ▶ **Hierarchy** describes how the various structural elements are combined





Analytical Approaches





Analytical Approaches

Network topology statistics such as node degree, degree distribution, centralitiy, clustering coefficient, shortest paths, and robustness of the network to the random removal of single nodes are important network characteristics.

Modularity refers to the identification of sub-networks of interconnected nodes that might represent molecules physically or functionally linked that work coordinately to achieve a specific function.

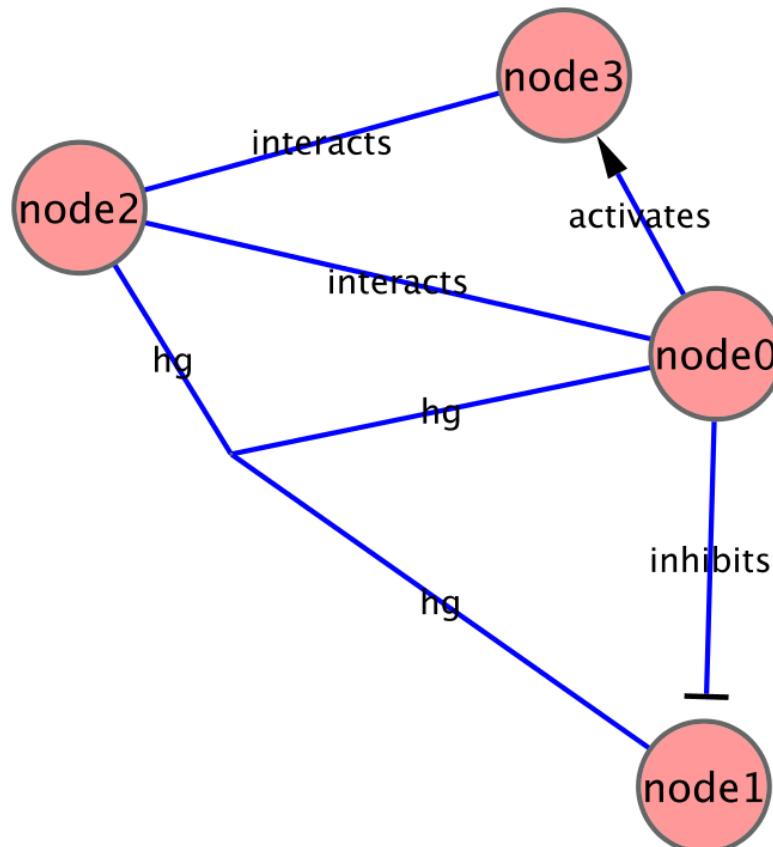
Motif analysis is the identification of small network patterns that are over-represented when compared with a randomized version of the same network. Regulatory elements are often composed of such motifs.

Network alignment and comparison tools can identify similarities between networks and have been used to study evolutionary relationships between protein networks of organisms.



Analytical Approaches

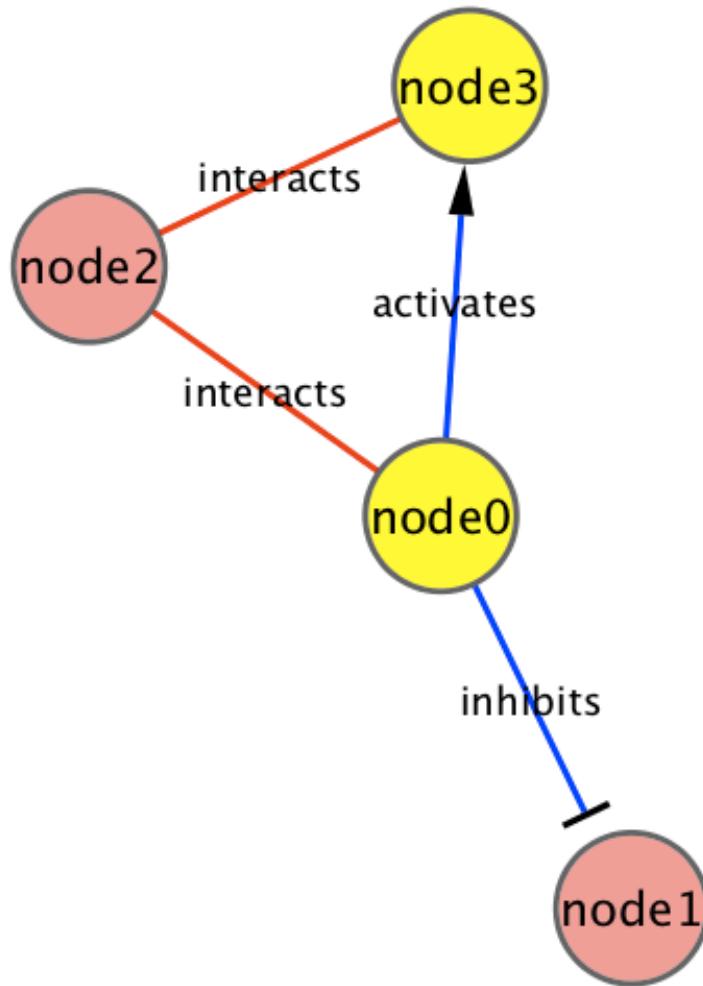
-





Analytical Approaches

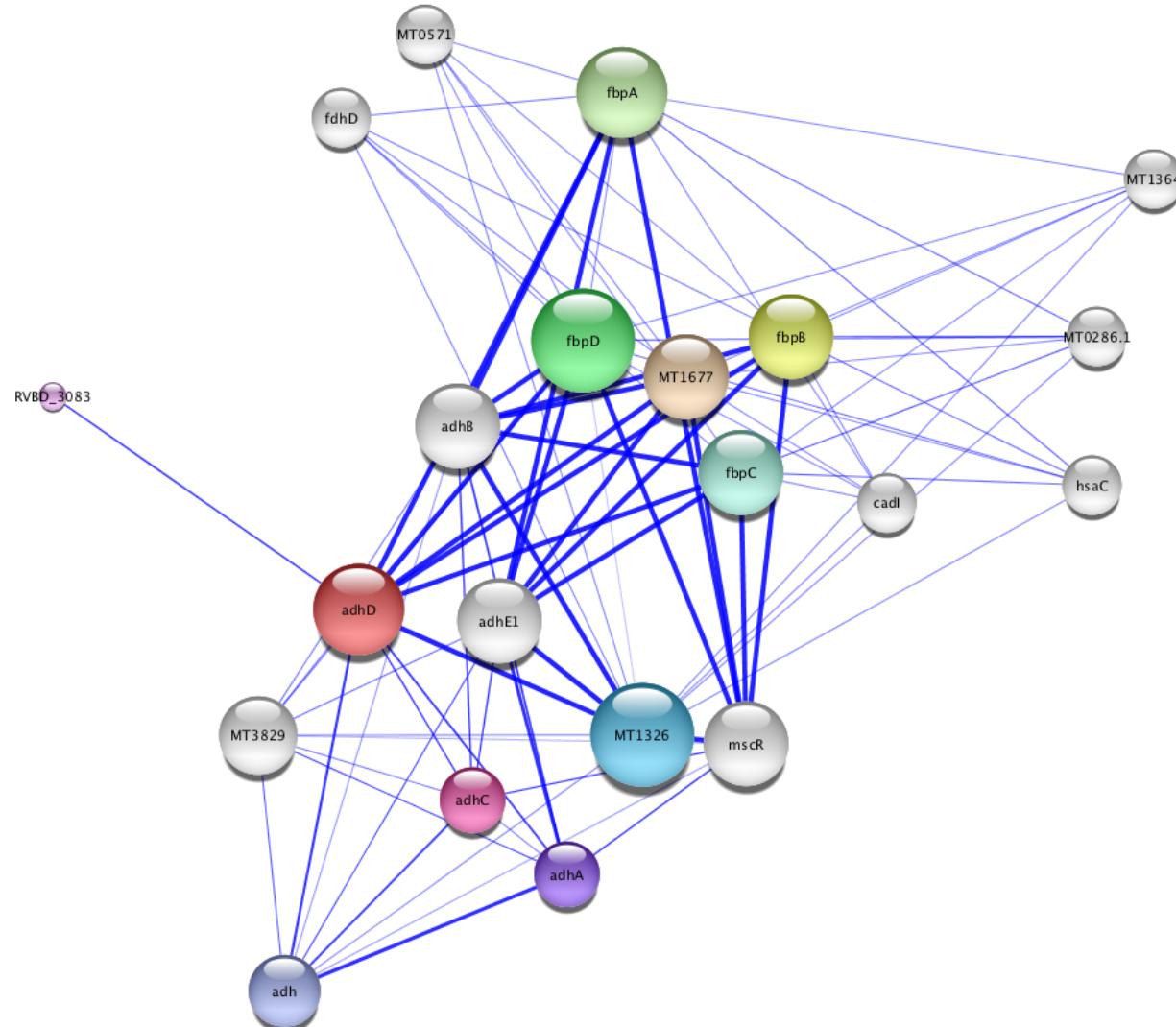
- N



not

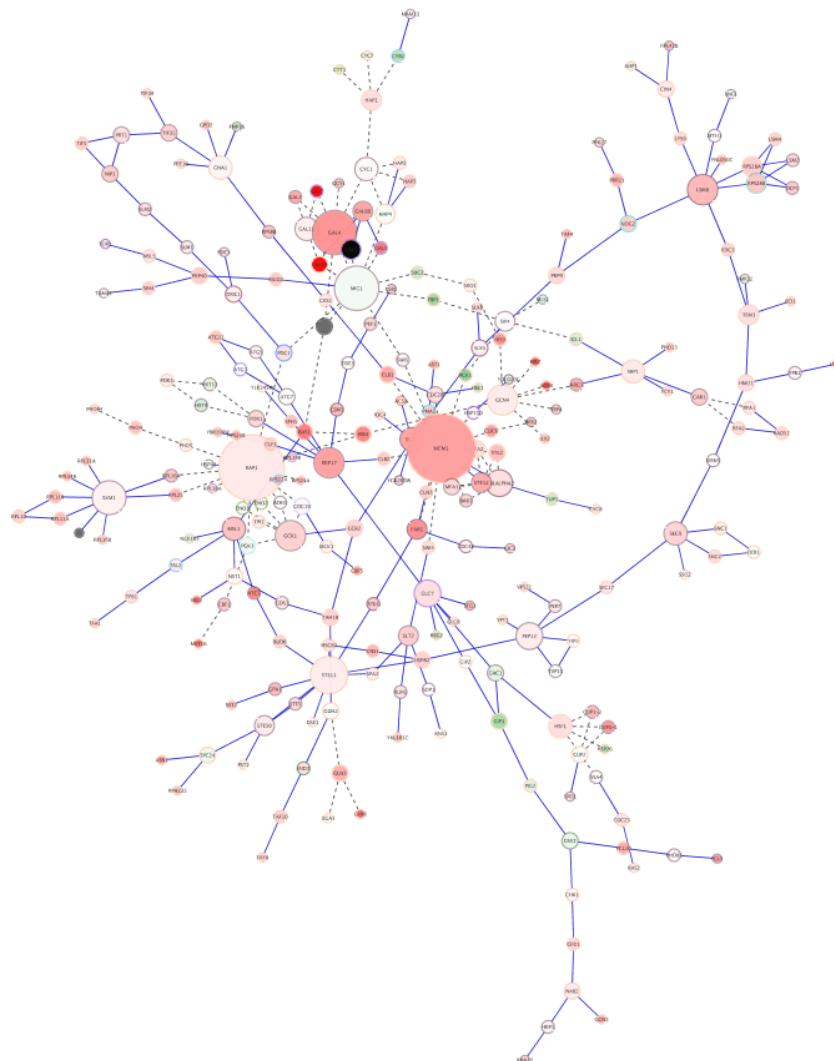
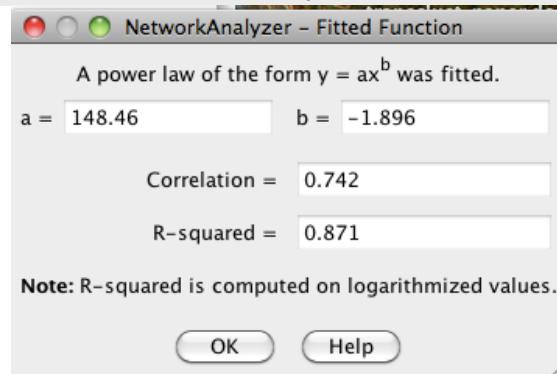
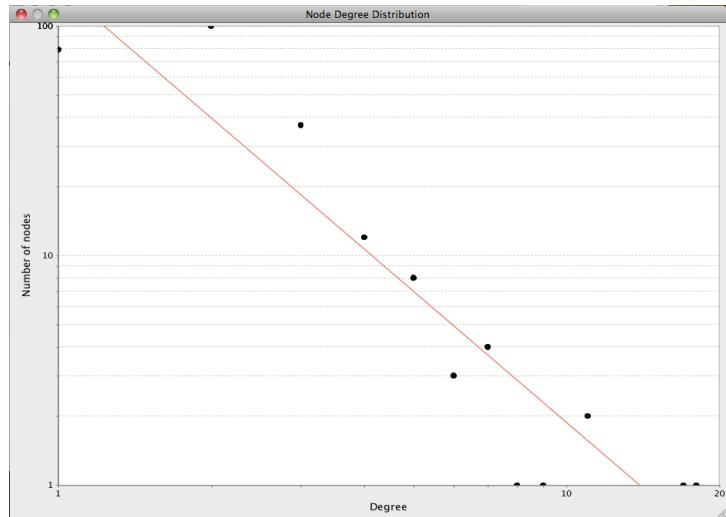


Analytical Approaches





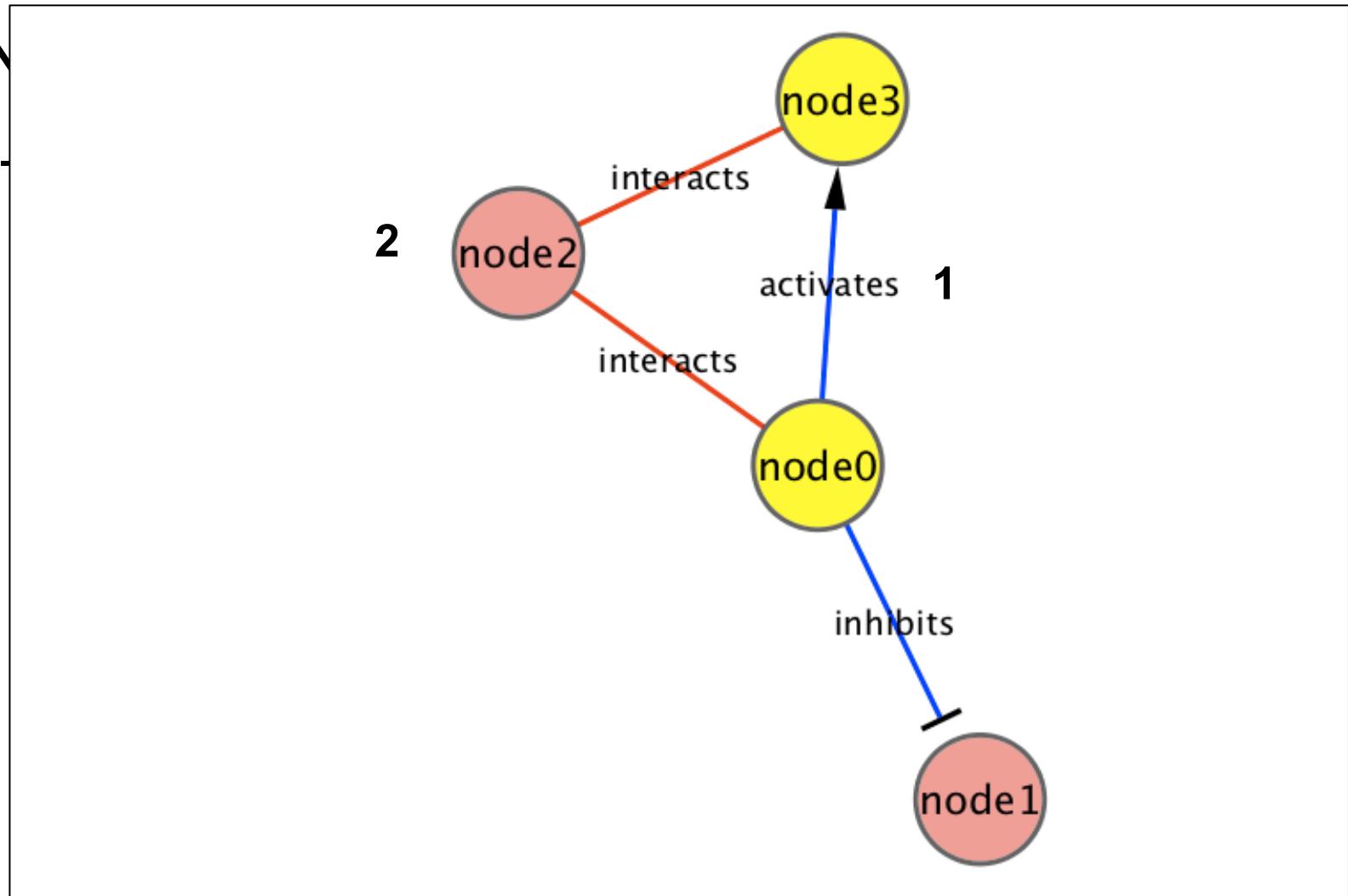
Analytical Approaches





Analytical Approaches

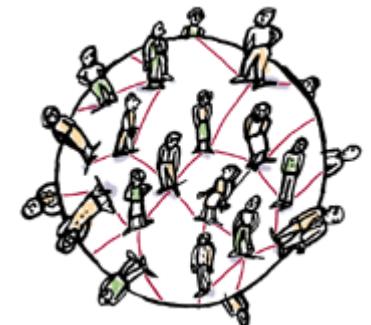
-





Analytical Approaches

- Small-world networks
 - any two arbitrary nodes are connected by a small number of intermediate edges
 - the network has an average shortest path length much smaller than the number of nodes in the network (Watts, Nature, 1998).
 - Interaction networks have been shown to be small-world networks (Barabási, Nature Reviews in Genetics, 2004)



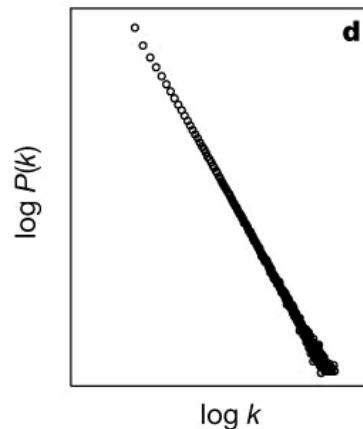
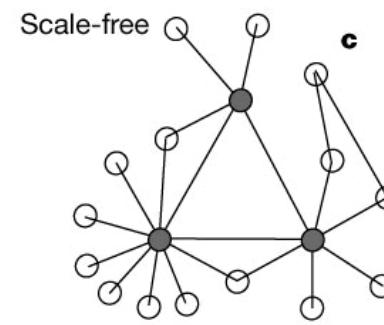
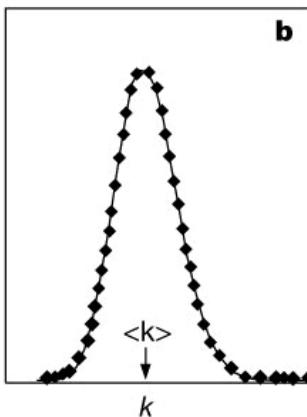
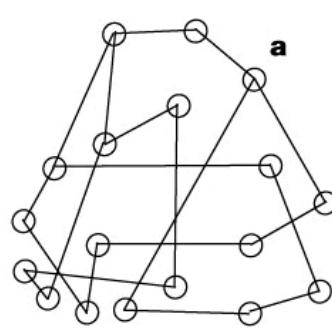


Analytical Approaches

- Random networks
 - Algorithms exist to create random networks

- Flat random networks
- Scale-free
- Small-world

- Useful to understand network

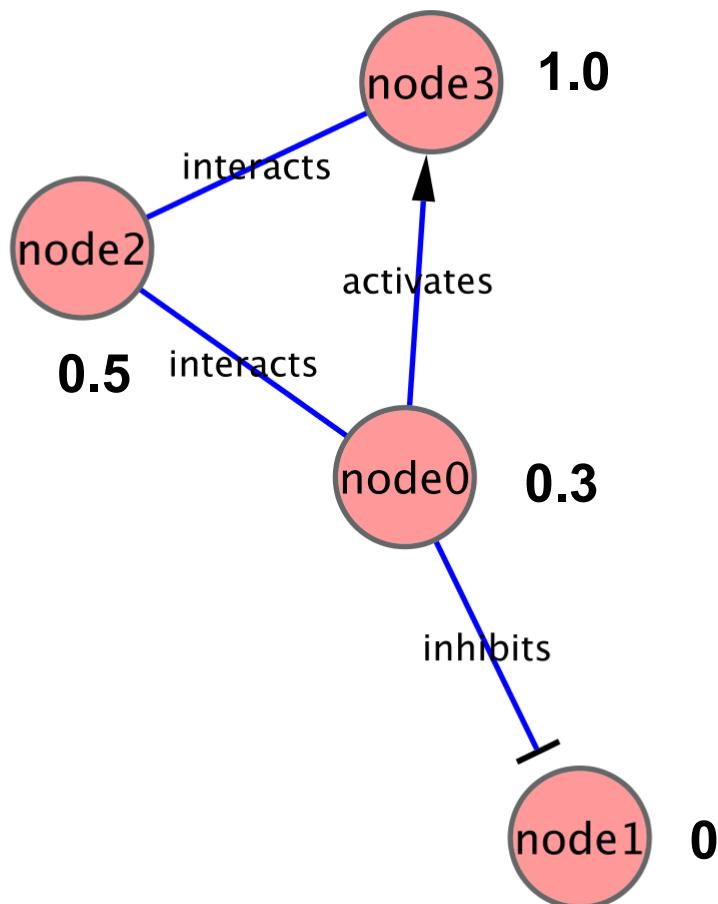


homogeneous, burst to
ogatz
a random



Analytical Approaches

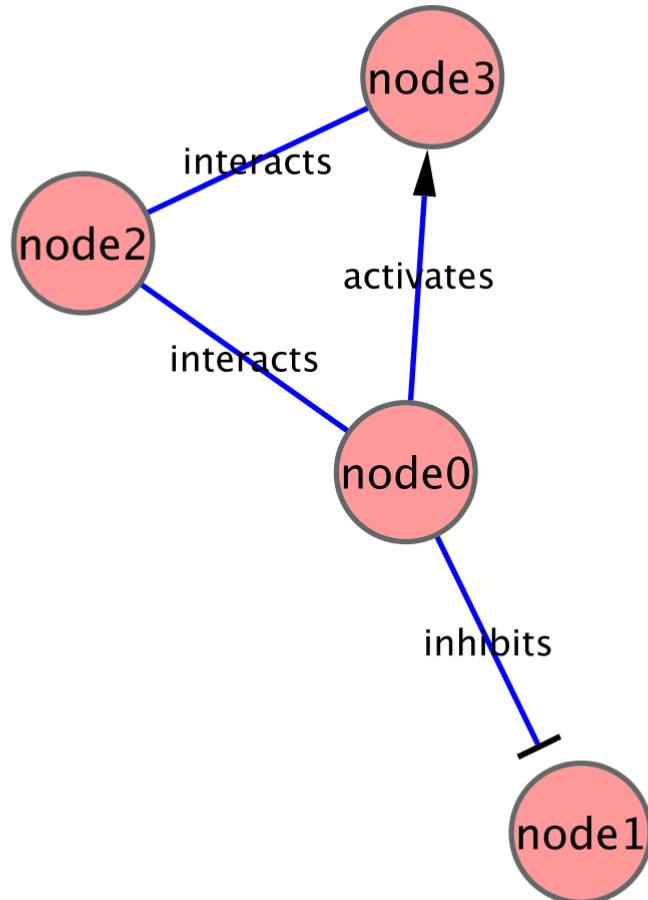
- N





Analytical Approaches

-



ID ▼	AverageShortestPathLength	BetweennessCentrality	ClosenessCentrality	ClusteringCoefficient	Degree
node0	1.0	0.66666667	1.0	0.33333333	3
node1	0.0	0.0	0.0	0.0	1
node2	1.33333333	0.66666667	0.75	0.5	2
node3	2.0	0.0	0.5	1.0	2



Analytical Approaches

- Network Analyzer Demo...



Analytical Approaches

- Guilt by association
 - Combine weak signals to get a stronger one.
 - Two main applications:
 - Suggest network or functional relationship based on related data (e.g. co-expression).
 - Infer function or role (i.e. in disease) based on related data
 - Some algorithmic approaches
 - Random walk with restarts
 - Semi-supervised learning
 - Belief Propagation
 - Example: GeneMANIA



Analytical Approaches

- GeneMANIA Demo...

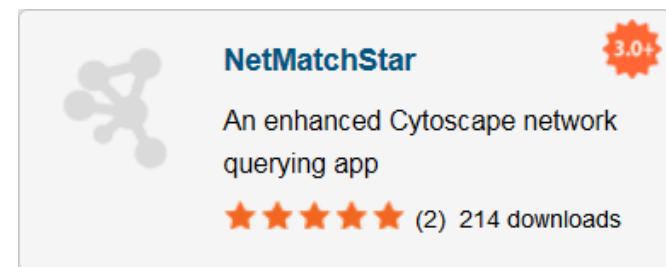


Analytical Approaches

- Motif finding
 - Search directed networks for network motifs (feed-forward loops, feedback loops, etc.)

The screenshot shows the NetMatch application window. On the left is a control panel with various network models (Shuffling, Erdos-Renyi, Watts-Strogatz, Barabasi-Albert, Duplication, Geometric, Forest-fire) and parameters like seed, shuffle, and edge probability. In the center, a tree view shows a network named 'BINDhuman' with a child node 'parFiltered'. A 'NetMatch?' dialog box is open, displaying network statistics: Average degree (1.099634440876132), Average clustering coefficient (0.79708301570997), Assortativity (-0.2429914561566887), and a list of network models with their respective average degree, clustering coefficient, and assortativity values. On the right is a results panel showing a table of motif occurrences and their corresponding network diagrams.

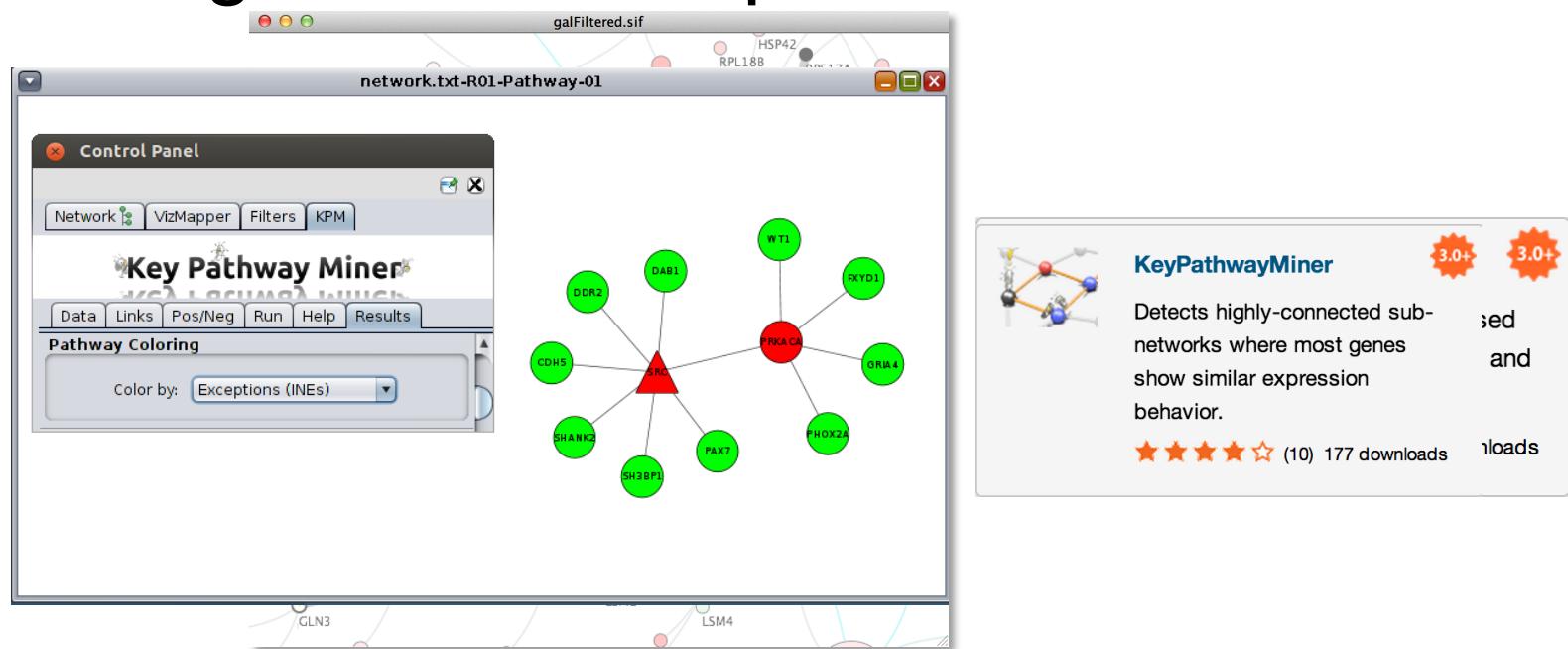
Match number	Occurrences	Nodes	Image
1	2	YPR102C, YLR075W, YDR395W, YGR085C	
2	2	YPR102C, YDR395W, YLR075W, YGR085C	
2	2	YGL153W, YLR191W, YDR142C, YDR244W	
2	2	YGL153W, YLR191W, YML214W, YDR244W	





Analytical Approaches

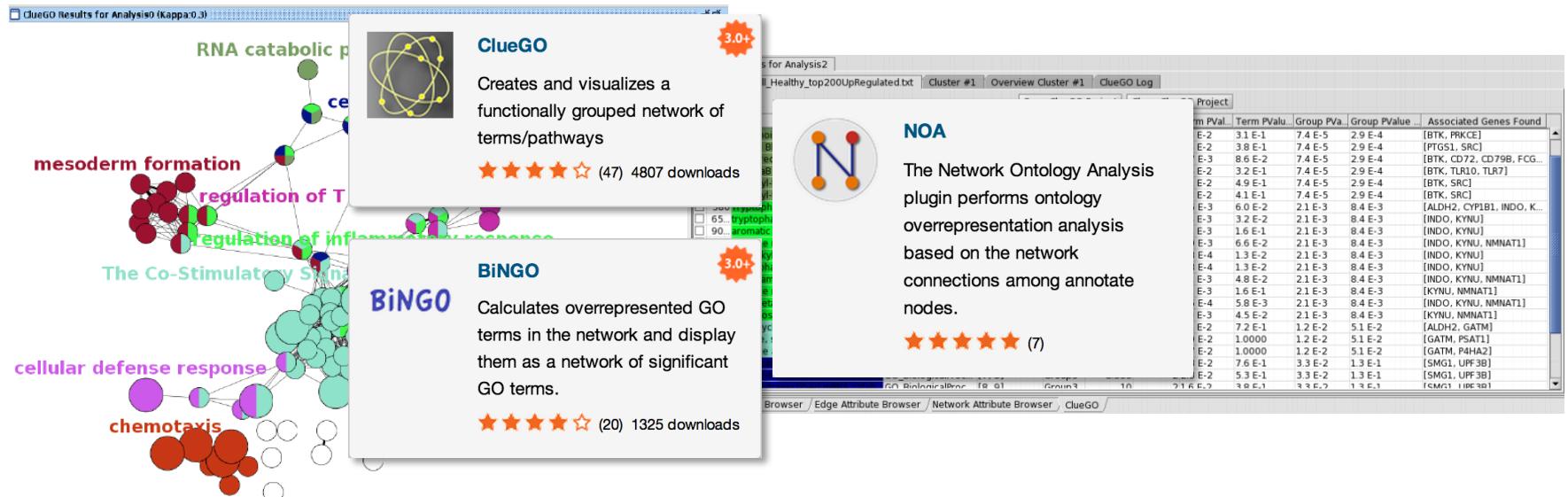
- Finding subnetworks of similarly expressed genes
- Finding the shortest path between nodes





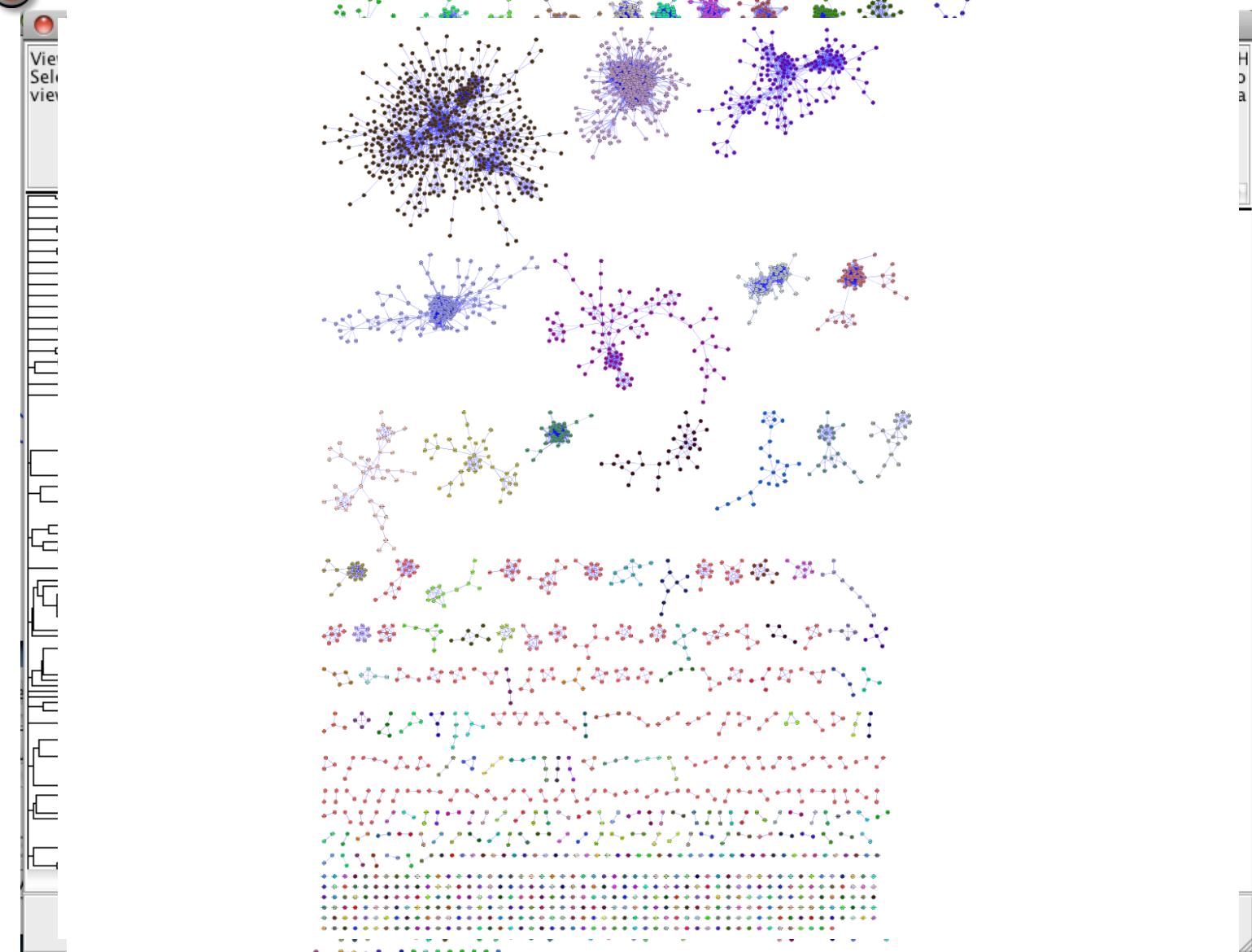
Analytical Approaches

- Overrepresentation analysis
 - Find terms (GO) that are statistically overrepresented in a network
 - Not really a network analysis technique
 - Very useful for visualization





Analytical Approaches



edges

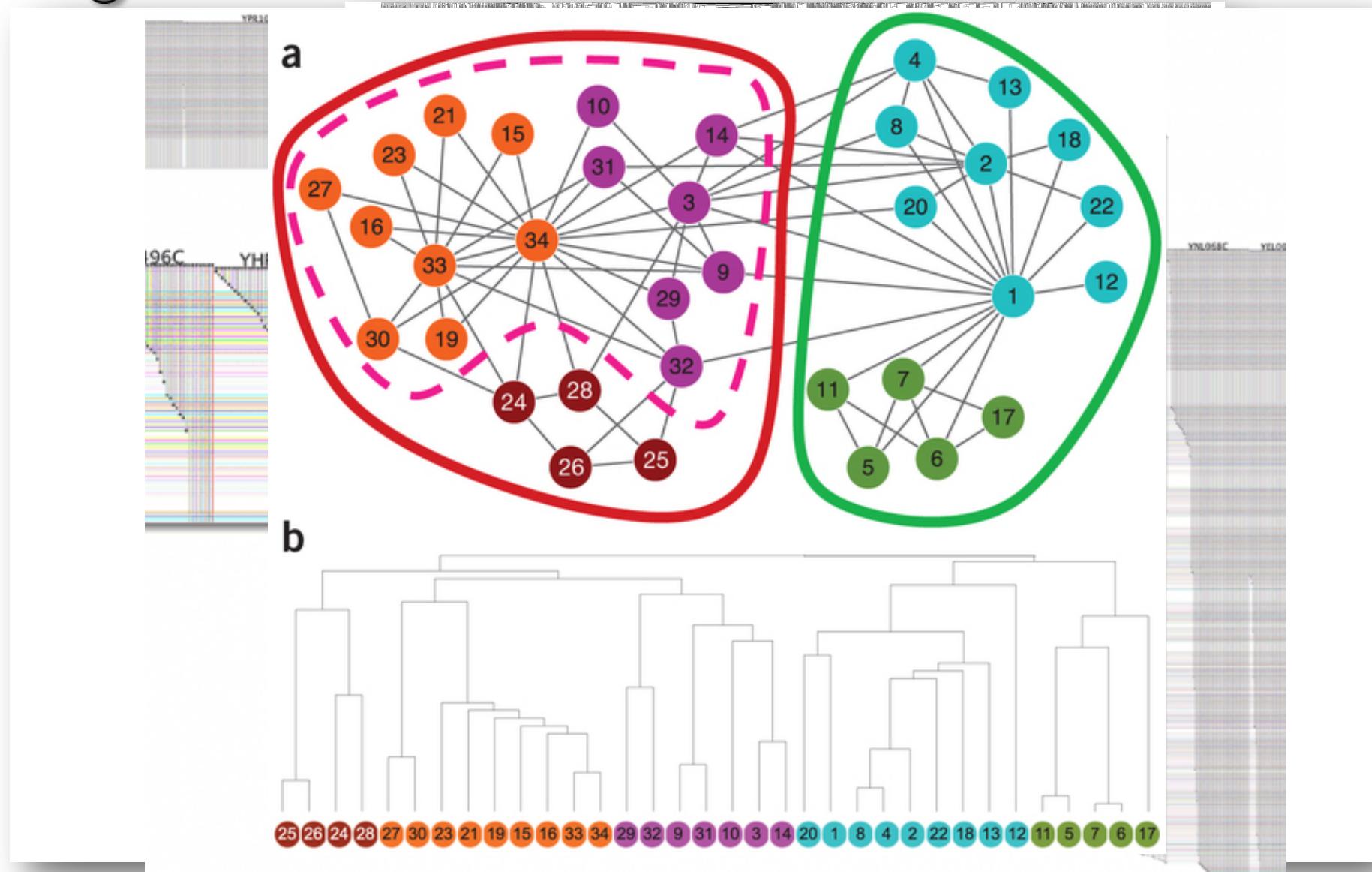


Visualization of Biological Networks

- Data Mapping
- Layouts
- Animation

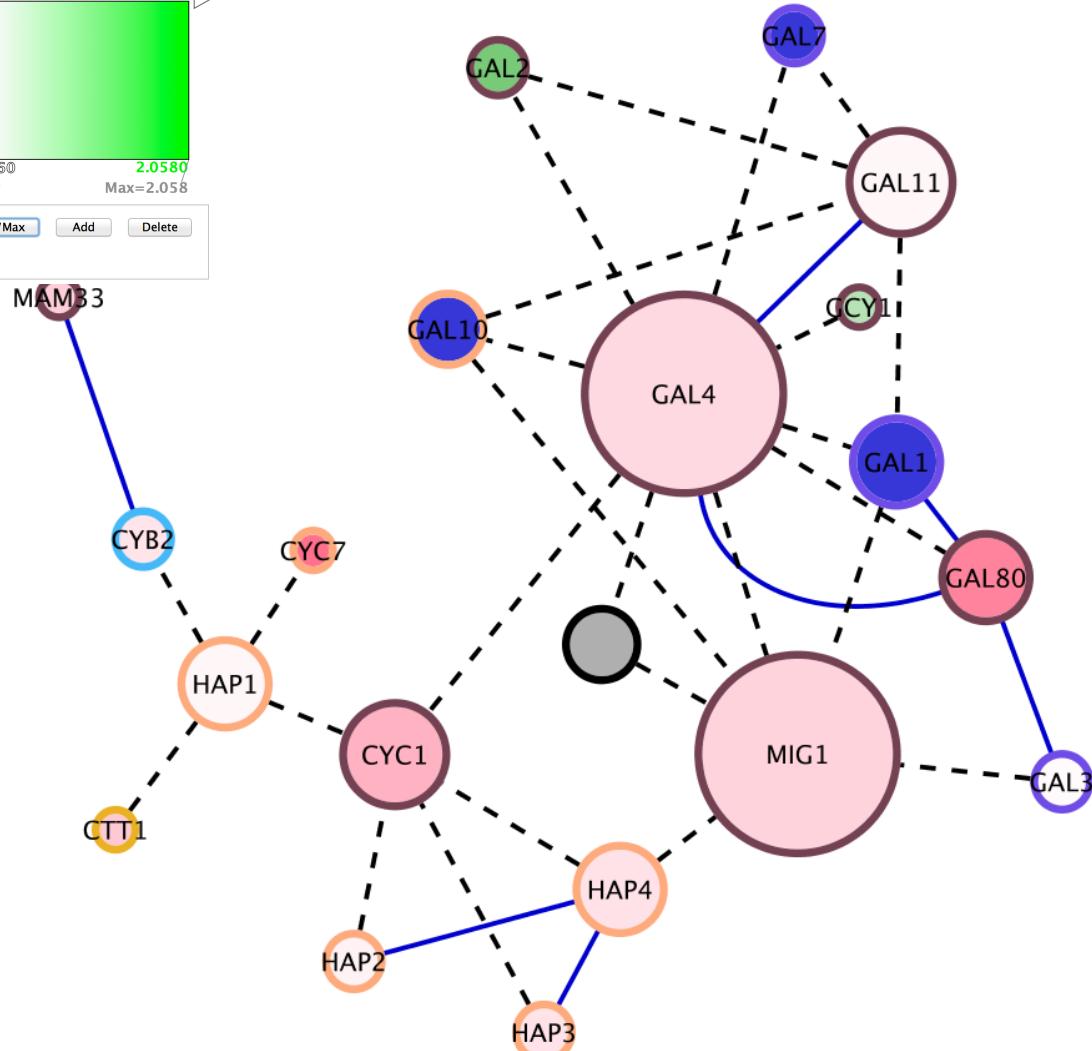
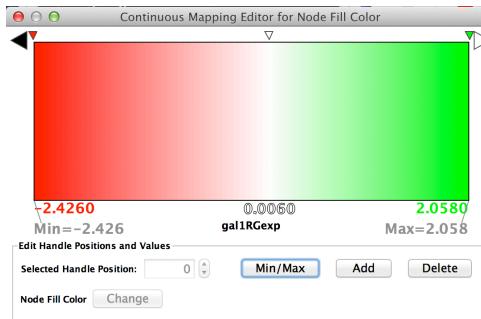


Depiction





Data Mapping



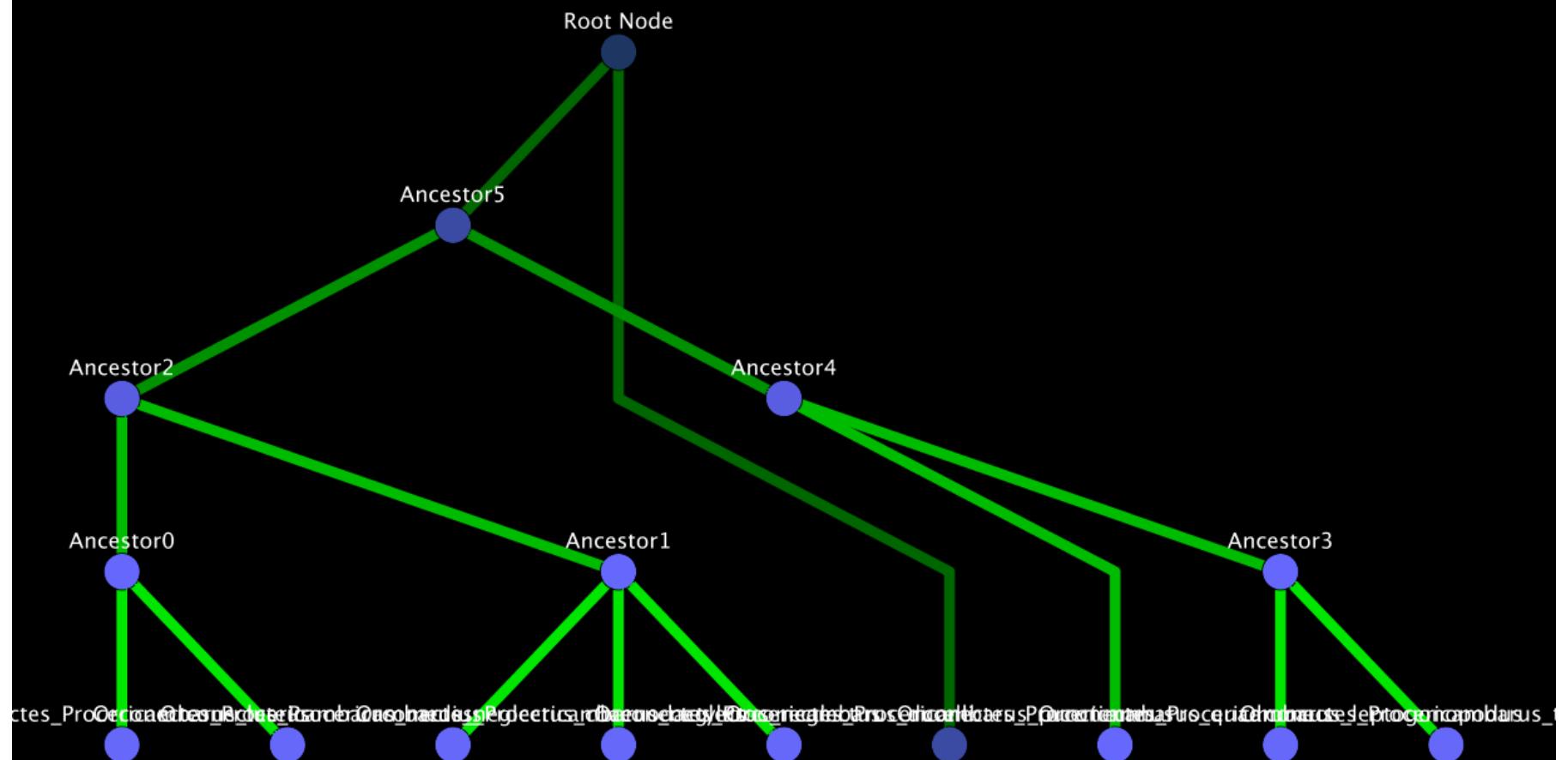


Data Mapping

- Avoid cluttering your visualization with too much data
 - Highlight meaningful differences
 - Avoid confusing the viewer
 - Consider creating multiple network images

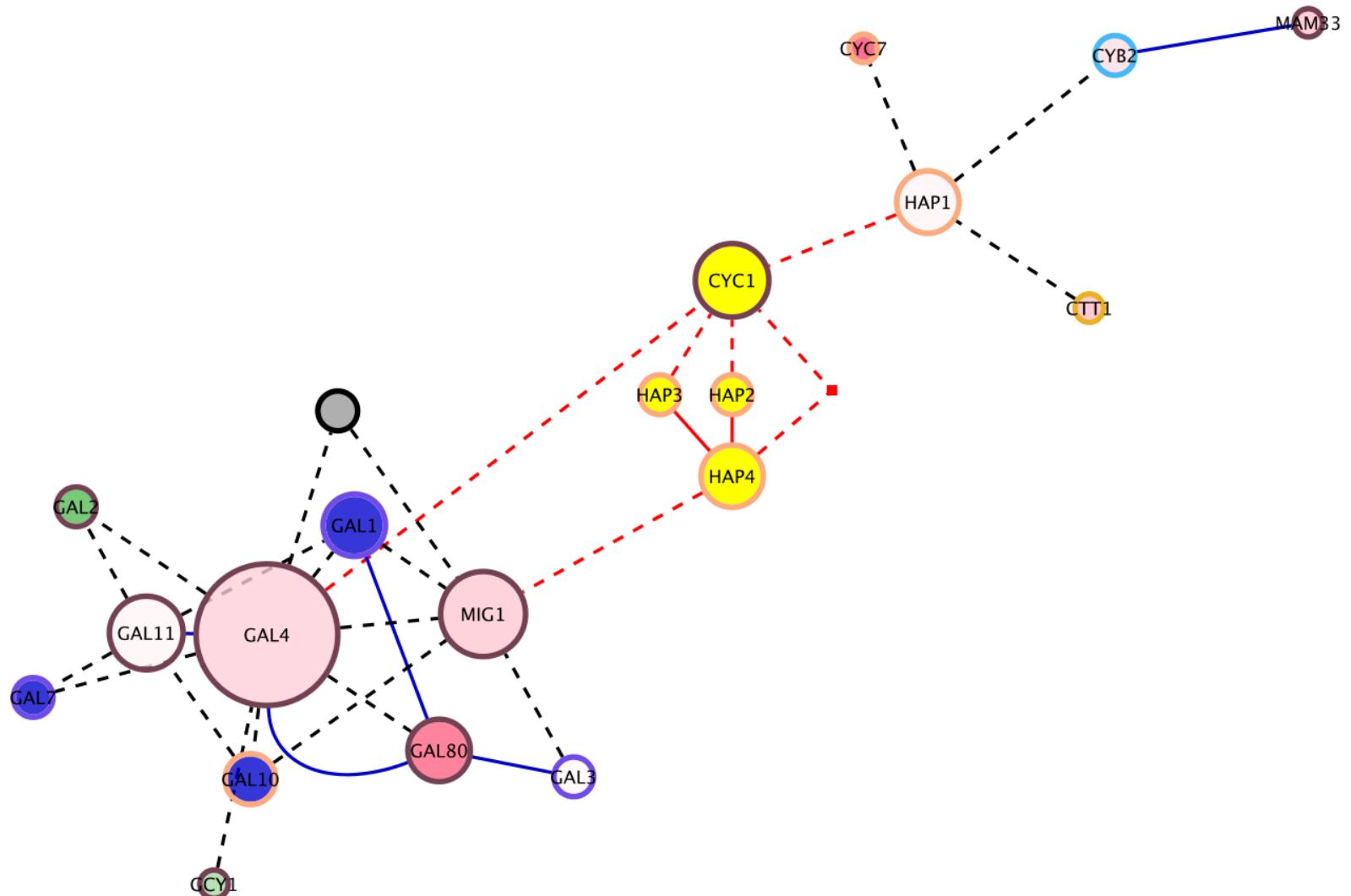


Layouts





Layouts



— Many, many others

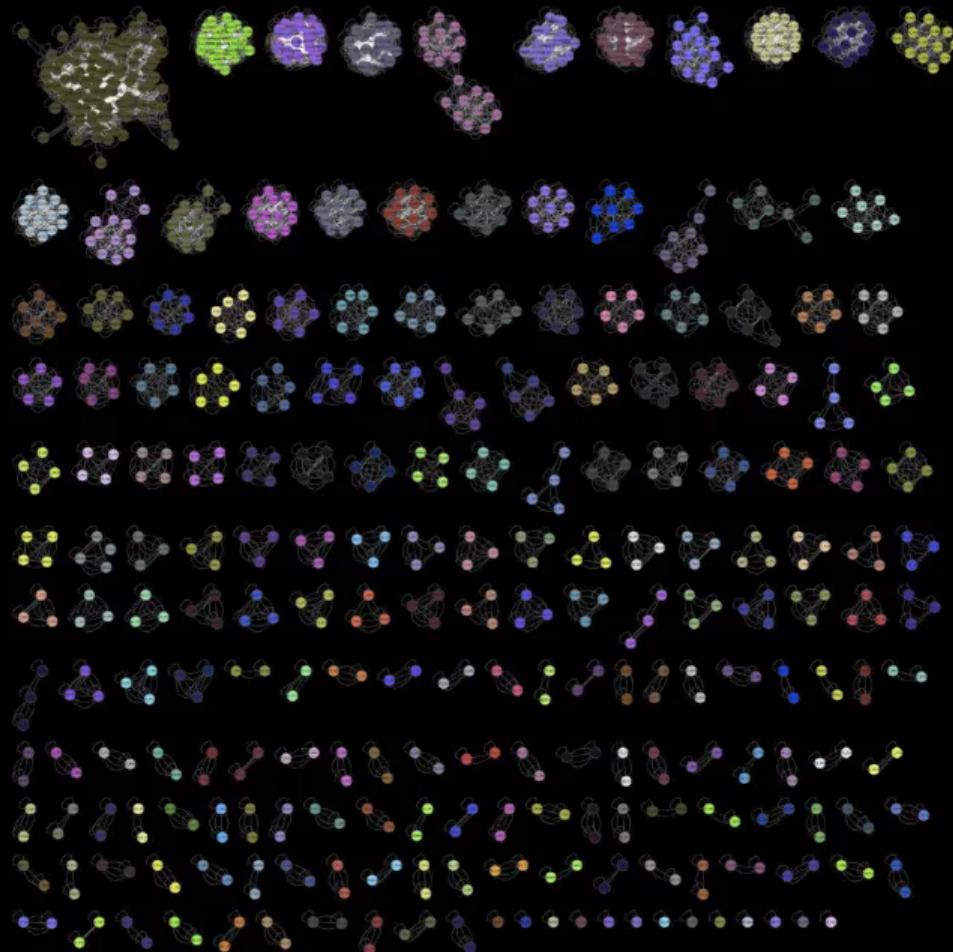


Layouts

- Use layouts to convey the relationships between the nodes.
- Layout algorithms may need to be “tuned” to fit your network.
- There is not one *correct* layout. Try different things.



Animation



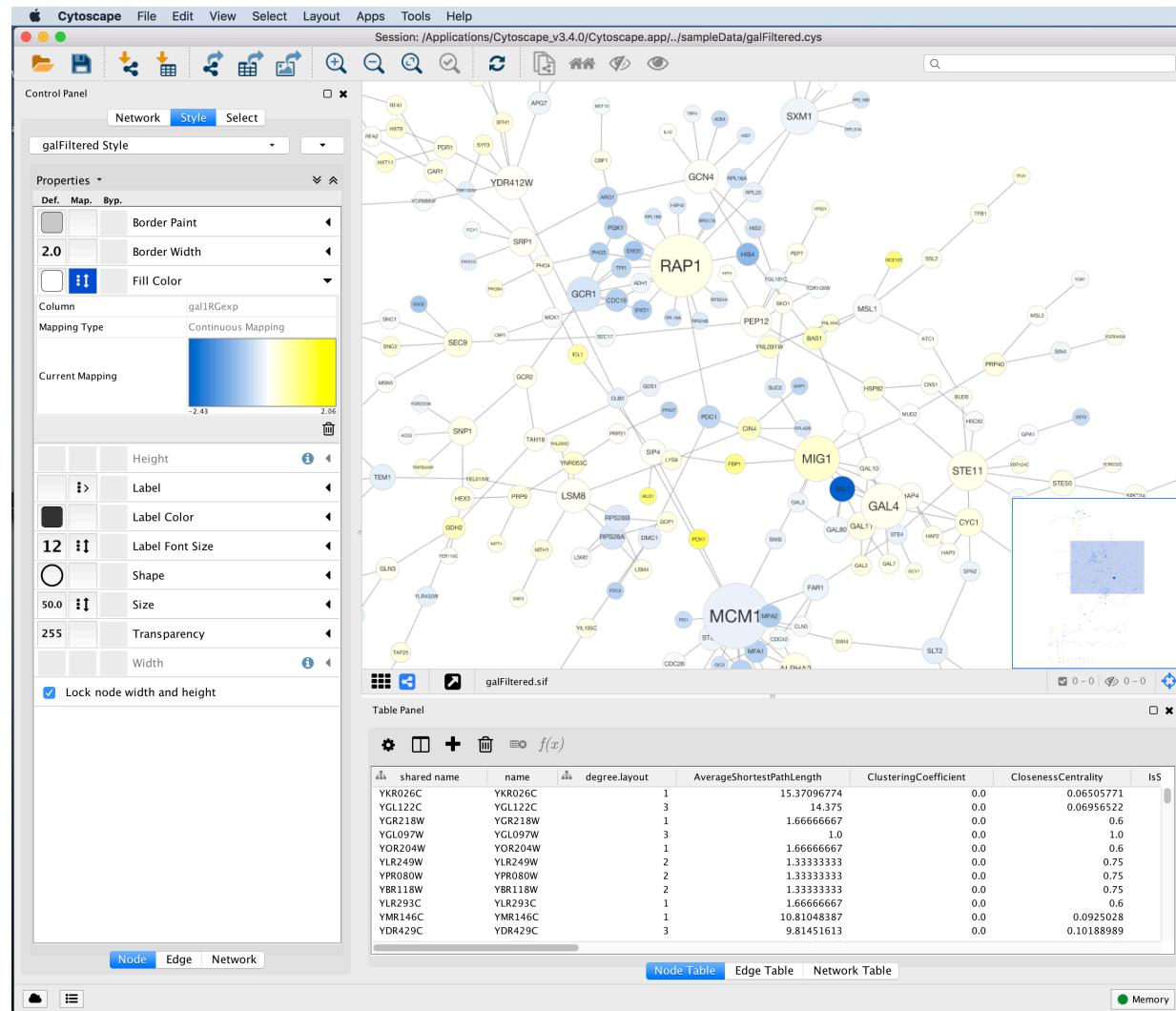


Introduction to Cytoscape

- Overview
- Core Concepts
 - Networks and Tables
 - Visual Properties
 - Cytoscape Apps
- Working with Data
 - Loading networks from files and online databases
 - Loading data tables from CSV or Excel files
 - The Table Panel



Cytoscape



- Open source
- Cross platform
- Consortium



Institute for Systems Biology



University of California at San Diego



Memorial Sloan-Kettering Cancer Center



Institut Pasteur



Agilent Technologies



University of California at San Francisco



University of Toronto



GLADSTONE INSTITUTES

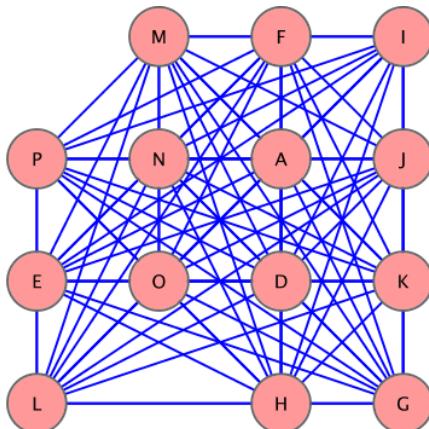


NRNB



Core Concepts

- Networks and Tables



	A	B	C	D	E	F	G	H	I	J
1	GI_Number	Gene	HPRD_ID	REFSEQ	Entrez_ID	OMIM_ID	SWISSPROT	Gene_Description	Architecture	Arch
2	gi10923959	SDSL	11542	NP_012441.1	113675		SDSL		SP	Motif
3	gi14591926	SEC23B	11543	NP_116781.1	10483	Q15437	Protein transport protein SEC23B	GEL	Dom	
4	gi14602492	SCMH1	11540	AAH0752.1	22955	Q96GD3	SCMH1	SAM	Dom	
5	gi21361625	EXOC2	11544	NP_060773.3	55770	Q96KP1	SEC5 like 1	IPT	Dom	
6	gi17998551	SERPINB12	11548	NP_536722.1	89777	Q96P63	Serpin B12	SERPIN	Dom	
7	gi65287717	EIF2AK4	18447	NP_00101372	440275	609280	Similar to GCN2 eIF2alpha kinase	S_T_Y_Kinase	Dom	
8	gi4507755	TYROBP	4996	NP_003323.1	7305	604142	043914	DAP12	ITAM	Dom
9	gi5803086	IL24	4995	NP_006841.1	11009	604136	Q13007	Interleukin 24	IL10	Dom
10	gi21265034	ADAMTS13	4994	NP_620594.1	11093	604134	ADAMTS 13	TSP1	Dom	
11	gi1743873	PIP5K1A	4470	AAC50911.1	8394	603275	Q99755	Phosphatidylinositol-4-phosphate 5 PIPKc	Dom	
12	gi1459382	SVIL	4992	NP_00101006.1	6840	604139	095425	Supeverillin	VHP	Dom
13	gi1633955	SLC17B1	4991	NP_014440.1	6002	604129	001232	ATP-binding transporter family 2B, member 1	ABC	
14	gi21040399	RBBP8	4990	AAH30590.1	5932	604124	099708	CTIP	LZ	Motif
15	gi4758378	FIGF	2102	NP_004460.1	2277	300091	043915	VEGFD	PDGF	Dom
16	gi11321617	DPYSL4	7463	NP_006417.1	10570	608407	014531	Collapsin response mediator protein 3		
17	gi8923202	TASP1	7460	NP_060184.1	55617	608270	Q9H6P5	Threonine aminopeptidase 1		
18	gi5454090	SSR4	2101	NP_006271.1	6748	300090	P51571	Signal sequence receptor delta	SP	Motif
19	gi5730045	SLC16A2	2108	NP_006508.1	6567	300095	P36021	X linked PEST containing transport	TM	Dom
20	gi11968027	FTS	7467	NP_071921.1	64400	608483	Q9H8T0	FTS	UBC	Dom
21	gi48255885	PRKC1	2109	NP_002731.3	5584	600539	P41743	Protein kinase C, iota type	S_T_Kinase	Dom
22	gi9100000	RPA4	6593	NP_074779.1	29335		RPA4		TRNA	
23	gi3376312	PPP1R2P9	6593	NP_074780.1	80316			Type I protein phosphatase inhibitor		
24	gi15826862		6595	NP_296379.1	90000			JM11 protein	CC	Motif
25	gi7661844	CCDC22	6594	NP_054727.1	28952			JM1 protein	CC	Motif
26	gi6005794	PRAF2	6596	NP_009144.1	11230			JM4 protein	TM	Dom

Networks

e.g., PPIs or pathways

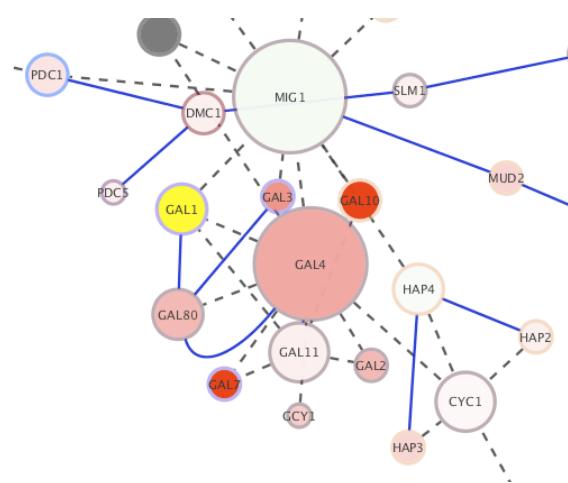
Tables

e.g., data or annotations



Core Concepts

- Networks and Tables



	A	B	C	D	E	F	G	H	I
	GI_Number	Gene	HPRD_ID	REFSEQ	Entrez_ID	OMIM_ID	SWISSPROT_Gene_Description	Architecture	Arch
1	gi10923959	SDSL	11542	NP_012441.1	113675		Protein transport protein SEC23B	SP	Motif
2	gi14591926	SEC23B	11543	NP_116781.1	10483	Q15437	Protein transport protein SEC23B	GEL	Dom
3	gi14620492	SCMH1	11540	AAH09752.1	22955	Q96GD3	SCMH1	SAM	Dom
4	gi21361625	EXOC2	11544	NP_060773.3	55770	Q96KP1	SEC5 like 1	IPT	Dom
5	gi17998551	SERPINB12	11548	NP_536722.1	89777	Q96P63	Serpin B12	SERPIN	Dom
6	gi65287717	EIF2AK4	18447	NP_00101372	440275	609280	Similar to GCN2 eIF2alpha kinase	S_T_Y_Kinase	Dom
7	gi4507755	TYROBP	4996	NP_003323.1	7305	604142	043914 DAP12	ITAM	Dom
8	gi5803086	IL24	4995	NP_006841.1	11009	604136	Q13007 Interleukin 24	IL10	Dom
9	gi21265034	ADAMTS13	4994	NP_620594.1	11093	604134	ADAMTS 13	TSP1	Dom
10	gi1743873	PIP5K1A	4470	AAC50911.1	8394	603275	Q99755 Phosphatidylinositol-4-phosphate 5 PIPKc	VHP	Dom
11	gi14591982	SVD	4992	NP_00101006.1	6840	604149	Q95425 Superfamily		
12	gi1633955	SLC12A2	4991	NP_014440.1	6002	600123	Na+/K+ ATPase family 2B, member 1		
13	gi21040399	RBBP8	4990	AAH30590.1	5932	604124	Q99708 CTIP	LZ	Motif
14	gi4758378	FIGF	2102	NP_004460.1	2277	300091	043915 VEGF D	PDGF	Dom
15	gi11321617	DPYSL4	7463	NP_006417.1	10570	608407	Q14531 Collapsin response mediator protein 3		
16	gi8923202	TASP1	7460	NP_060184.1	55617	608270	Q9H6P5 Threonine aminopeptidase 1		
17	gi5454090	SSR4	2101	NP_006271.1	6748	300090	P51571 Signal sequence receptor delta	SP	Motif
18	gi5730045	SLC16A2	2108	NP_006508.1	6567	300095	P36021 X linked PEST containing transport	TM	Dom
19	gi1968027	FTS	7467	NP_071921.1	64400	608483	Q9H8T0 FTS	UBC	Dom
20	gi48255885	PRKC1	2109	NP_002731.3	5584	600539	P41743 Protein kinase C, iota type	S_T_Y_kinase	Dom
21	gi9197459	RPA4	6592	NP_0747479.1	29335	600539	RPA4	TRNA	Dom
22	gi3376312	PPP1R2P9	6593	NP_0747479.1	80316		Type 1 protein phosphatase inhibitor		
23	gi15826862		6595	NP_296379.1	9000		JM11 protein	CC	Motif
24	gi7661844	CCDC22	6594	NP_054727.1	28952		JM1 protein	CC	Motif
25	gi6005794	PRAF2	6596	NP_009144.1	11230		JM4 protein	TM	Dom
26									

Networks

Tables

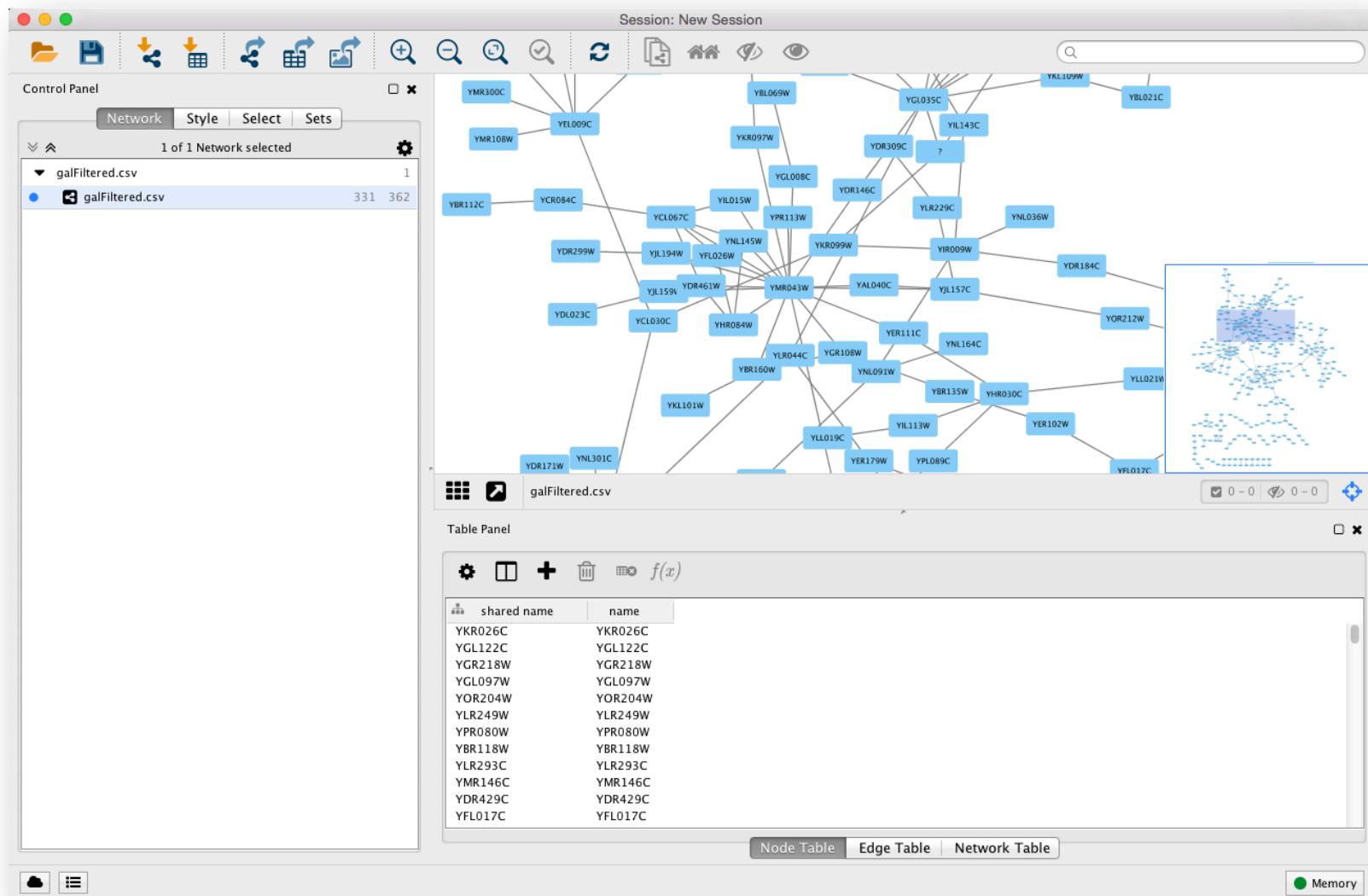
Visual Styles

Wall of Apps 173 total





Loading Networks





Loading Networks

Session: New Session

Control Panel Network Style Select Import Network from Public Databases

Data Source: Universal Interaction Database Client About

1. Enter Search Conditions

Search Mode: Search by ID (gene/protein/compound ID) Search

p53

2. Select Databases

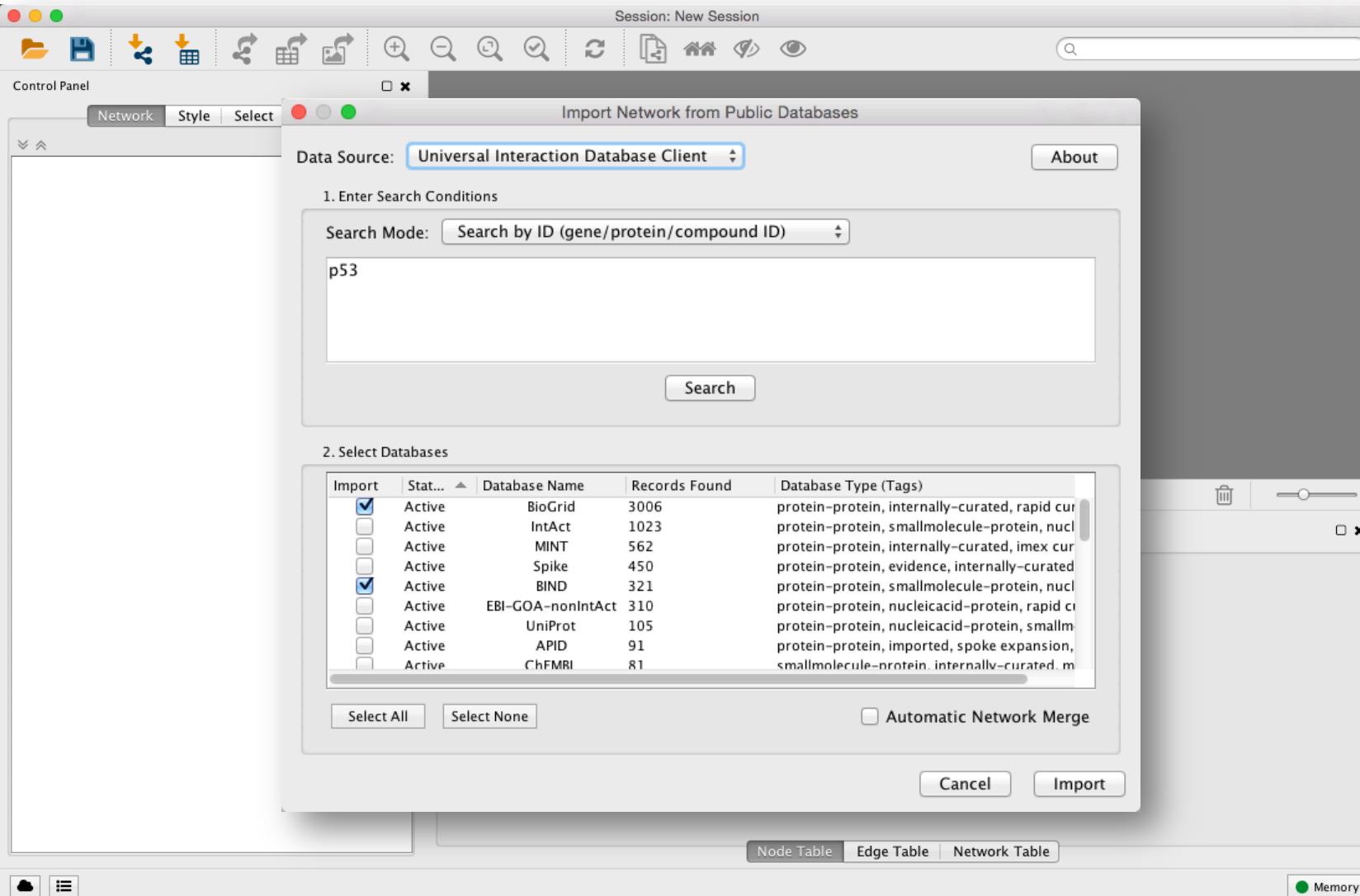
Import	Status	Database Name	Records Found	Database Type (Tags)
<input checked="" type="checkbox"/>	Active	BioGrid	3006	protein-protein, internally-curated, rapid cur
<input type="checkbox"/>	Active	IntAct	1023	protein-protein, smallmolecule-protein, nucl
<input type="checkbox"/>	Active	MIINT	562	protein-protein, internally-curated, imex cur
<input type="checkbox"/>	Active	Spike	450	protein-protein, evidence, internally-curated
<input checked="" type="checkbox"/>	Active	BIND	321	protein-protein, smallmolecule-protein, nucl
<input type="checkbox"/>	Active	EBI-GOA-nonIntAct	310	protein-protein, nucleicacid-protein, rapid cr
<input type="checkbox"/>	Active	UniProt	105	protein-protein, nucleicacid-protein, smallm
<input type="checkbox"/>	Active	APID	91	protein-protein, imported, spoke expansion,
<input type="checkbox"/>	Active	ChEMBL	81	smallmolecule-protein, internally-curated, m

Select All Select None Automatic Network Merge

Cancel Import

Node Table Edge Table Network Table

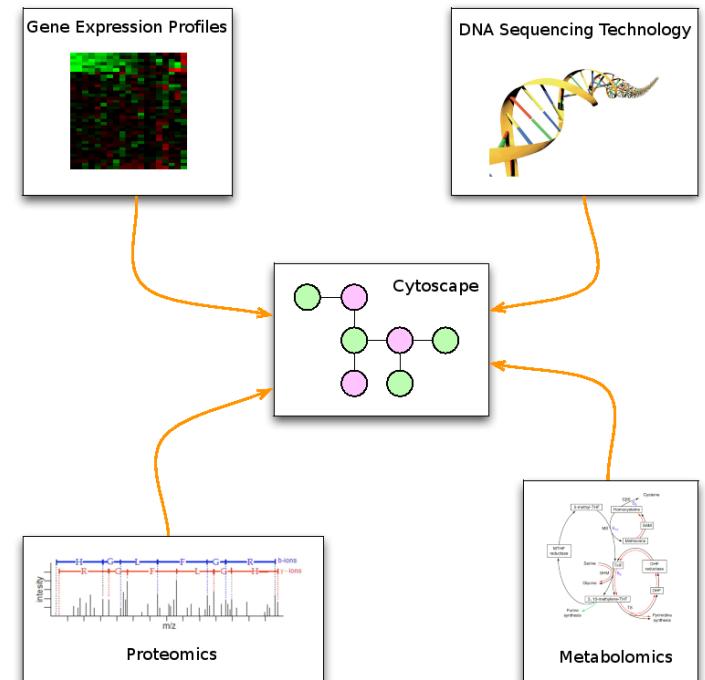
Memory





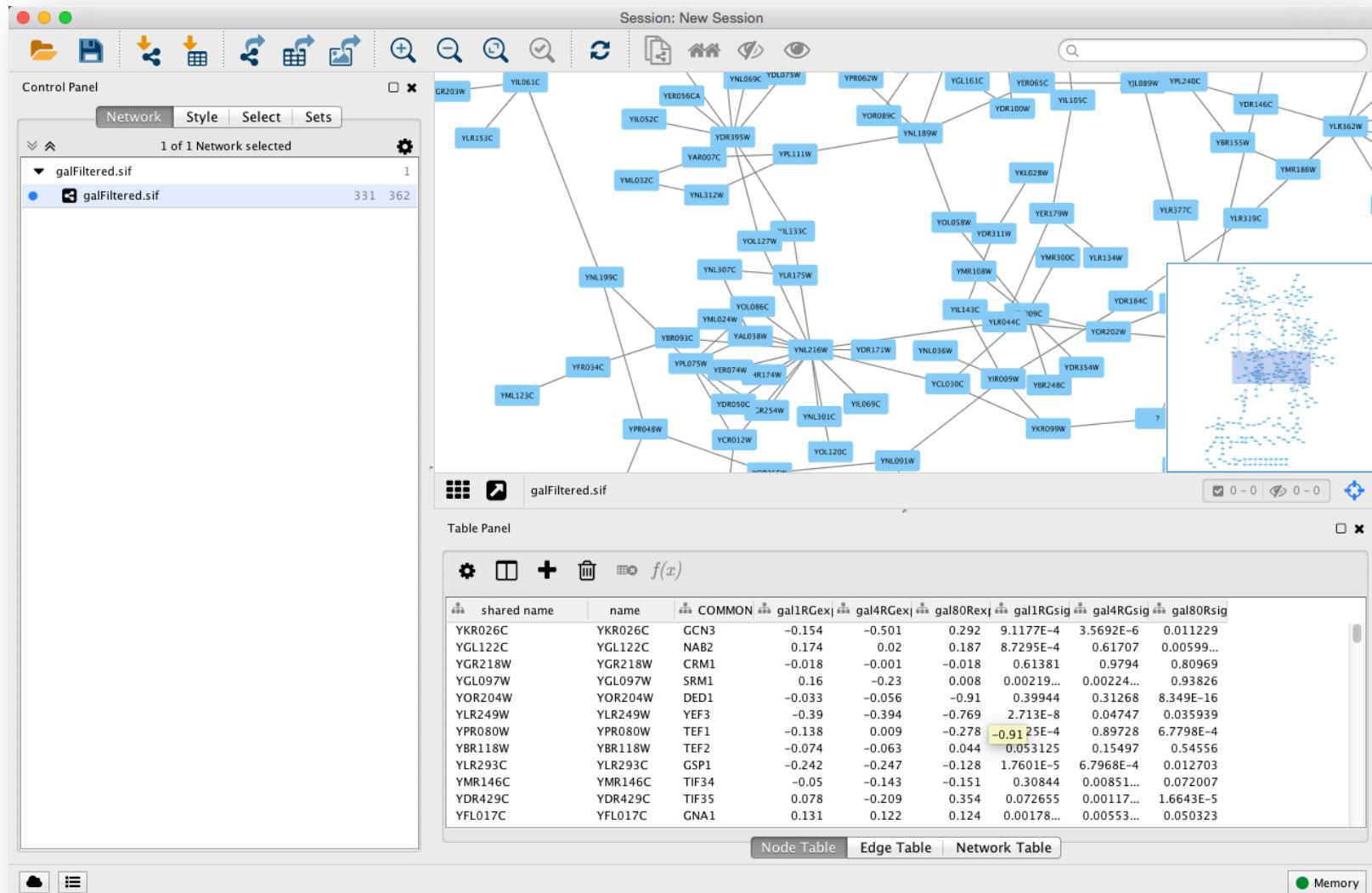
Loading Tables

- Nodes and edges can have data associated with them
 - Gene expression data
 - Mass spectrometry data
 - Protein structure information
 - Gene Ontology terms, etc.
- Cytoscape supports multiple data types: Numbers, Text, Boolean, Lists...





Loading Tables





Visual Style Manager

Session: /Applications/Cytoscape_v3.4.0-milestone2/sampleData/galFiltered.cys

Control Panel

Network Style Select Sets

Big Labels BioPAX BioPAX_SIF

Box Custom Graphics Style default

default black Directed galFiltered Style

Marquee Metallic Minimal

Nested Network Style Ripple Sample1

Solid Universe

Node Edge Network

Node Table Edge Table Network Table

Memory

Filtered.sif

e	name	degree.layout	AverageShortestPathLength	ClusteringCoefficient	ClosenessCentrality
YKR026C		1	15.37096774	0.0	0.065051
YGL122C		3	14.375	0.0	0.069565
YGR218W		1	1.66666667	0.0	0
YGL097W		3	1.0	0.0	0
YOR204W		1	1.66666667	0.0	0
YLR249W		2	1.33333333	0.0	0
YPR080W		2	1.33333333	0.0	0
YBR118W		2	1.33333333	0.0	0
YLR293C		1	1.66666667	0.0	0
YMR146C		1	10.81048387	0.0	0.09250
YDR429C		3	9.81451613	0.0	0.101885



Selection Filters

Session: /Applications/Cytoscape_v3.4.0-milestone2/sampleData/galFiltered.cys

Control Panel

Network | Style | Select | Sets

Default filter

Match any (OR)

~~x~~ Node: gal1RGexp is between 0.5 and 2.058 inclusive.

~~x~~ Node: gal1RGexp is between -2.426 and -.5 inclusive.

[+]

Selected 45 nodes and 0 edges in 127ms

Apply when filter changes

Apply

Filter | Chain

Table Panel

f(x)

shared name	name	degree.layout	AverageShortestPathLength	ClusteringCoefficient	ClosenessCentrality
YLR452C	YLR452C	1	8.84677419	0.0	0.113035
YDR299W	YDR299W	1	7.8266129	0.0	0.127769
YGR088W	YGR088W	1	9.52822581	0.0	0.104951
YER081W	YER081W	1	2.0	0.0	
YDR070C	YDR070C	1	10.4516129	0.0	0.095679
YNL036W	YNL036W	1	7.78225806	0.0	0.128491
YDL023C	YDL023C	1	7.8266129	0.0	0.127769
YBR045C	YBR045C	3	9.64516129	0.0	0.103678
YGL229C	YGL229C	2	1.75	0.0	0.571428
YAL038W	YAL038W	3	6.92741935	0.0	0.14435
YCR012W	YCR012W	3	6.92741935	0.0	0.14435

Node Table | Edge Table | Network Table

Memory



Saving and Exporting

- Sessions save everything as .cys files:
Networks, Tables, Styles, Screen sizes, etc
- Export networks in different formats:
SIF, GML, XGMML, BioPAX, PSI-MI 1 & 2.5
- Publication quality graphics in several formats:
PDF, EPS, SVG, PNG, JPEG, and BMP



Getting Help

Session: C:\Program Files\Cytoscape_v3.4.0\sampleData\galFiltered.cys

File Edit View Select Layout Apps Tools Help

User Manual Show Welcome Screen... Citations... Contact Help Desk... Report a Bug... About...

Control Panel Network Style Select jActiveModules 1 of 1 Network selected galFiltered.sif galFiltered.sif

Enter search term...

Network Table Edge Table Network Table

shared name	name	degree.layout	AverageShortestPathLength	ClusteringCoefficient	ClosenessCentrality
YKR026C	YKR026C	1	15.37096774	0.0	0.06505771
YGL122C	YGL122C	3	14.375	0.0	0.06956522
YGR218W	YGR218W	1	1.66666667	0.0	0.6
YGL097W	YGL097W	3	1.0	0.0	1.0
YOR204W	YOR204W	1	1.66666667	0.0	0.6
YLR249W	YLR249W	2	1.33333333	0.0	0.75
YPR080W	YPR080W	2	1.33333333	0.0	0.75
YBR118W	YBR118W	2	1.33333333	0.0	0.75
YLR293C	YLR293C	1	1.66666667	0.0	0.6
YMR145C	YMR145C	1	10.61040507	0.0	0.000000

Memory



Tips & Tricks

- Network Collections
 - Each collection has a “root” network
 - Changing the attribute for a node in one network *will* also change that attribute for a node with the same SUID in all other networks within the collection
 - You can clone a network into a new collection to “decouple” it and start a new root



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

- Task monitor
 - Current task displayed in status bar (lower left)
 - Icon opens complete task history
- Memory
 - Current status (lower right)
 - Toggle open for details and “Free Unused Memory” button



Tips & Tricks

- CytoscapeConfiguration directory
 - Your defaults and any apps downloaded from the App Store will go here
- App Manager
 - This is where you search/install/update/uninstall apps
 - You now have the option of disabling vs. uninstalling...
 - Can also install and update apps directly from the App Store website, if you have Cytoscape 3 up and running



Cytoscape: Platform

- Cytoscape as a platform
 - App architecture
 - <http://apps.cytoscape.org>
 - Use cases
 - Expression data analysis
 - Protein complexes
 - Agilent literature search



Hands-on Tutorial

Introduction to Cytoscape:
Networks, Data, Styles, Layouts and App Manager

tutorials.cytoscape.org



Hands-on Tutorial

App Tutorials:

<http://tutorials.cytoscape.org>

[http://rbvi.ucsf.edu/cytoscape/
structureViz2/tutorial.html](http://rbvi.ucsf.edu/cytoscape/structureViz2/tutorial.html)

<http://rinalyzer.de/tutorials.php>



Examples/Demos

- clusterMaker
 - Clustering and cluster visualizations
- Agilent LitSearch Tool
 - Extracting networks from abstracts
- WikiPathways
 - Search and load pathway diagrams



Expression Data Analysis

Control Panel

Network Style Filter Filter (New)

Network Nodes

galFiltered.sif galFiltered.sif 331(0)

Hierarchical cluster

Linkage pairwise average-linkage

Distance Metric Euclidean distance

Array sources

Node attributes for cluster

SelfLoops
TopologicalCoefficient
gal1RGexp
gal1RGsig
gal4RGexp
gal4RGsig
gal80Rexp
gal80Rsig

Edge column for cluster --None--

Clustering Parameters

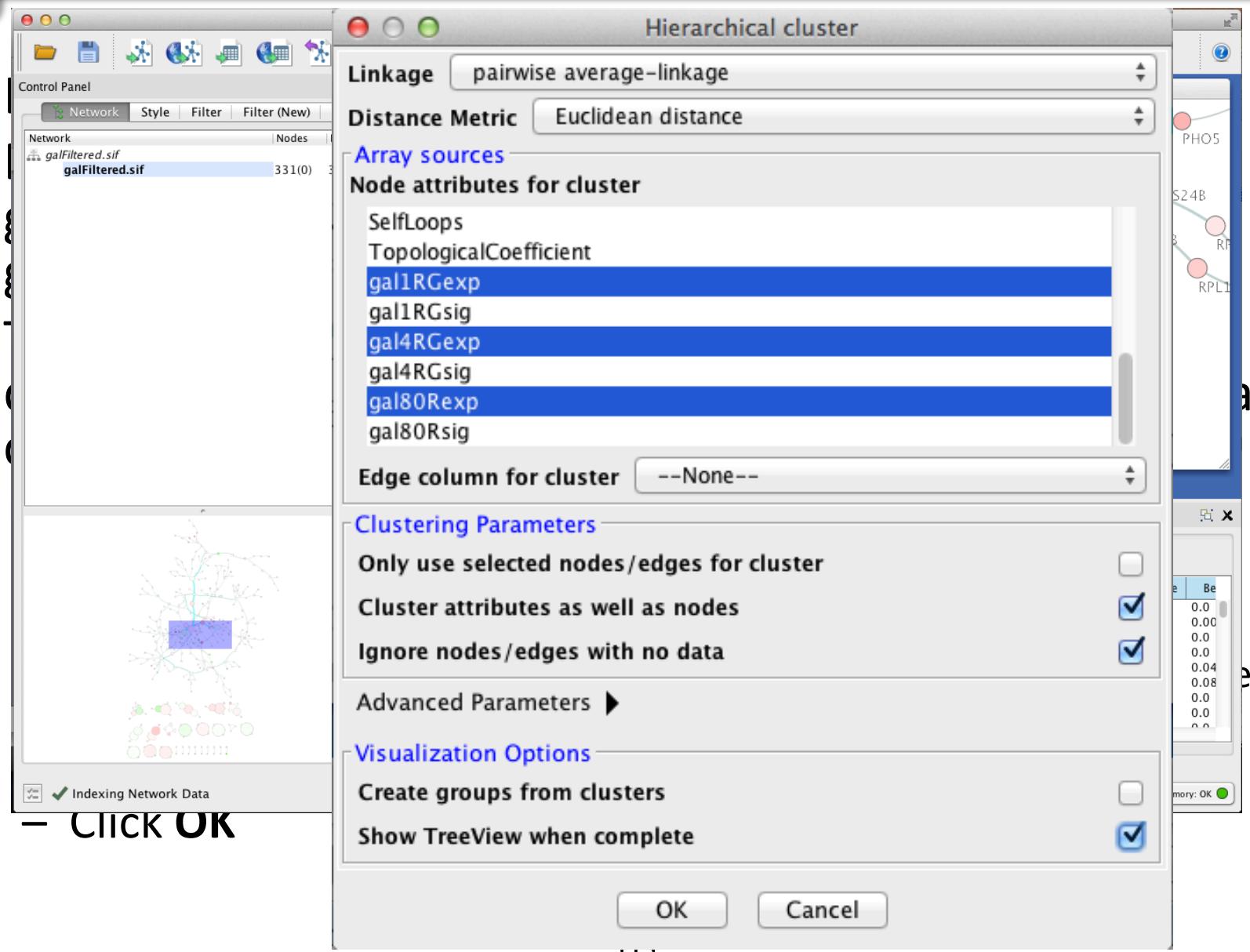
Only use selected nodes/edges for cluster
Cluster attributes as well as nodes
Ignore nodes/edges with no data

Advanced Parameters ▶

Visualization Options

Create groups from clusters
Show TreeView when complete

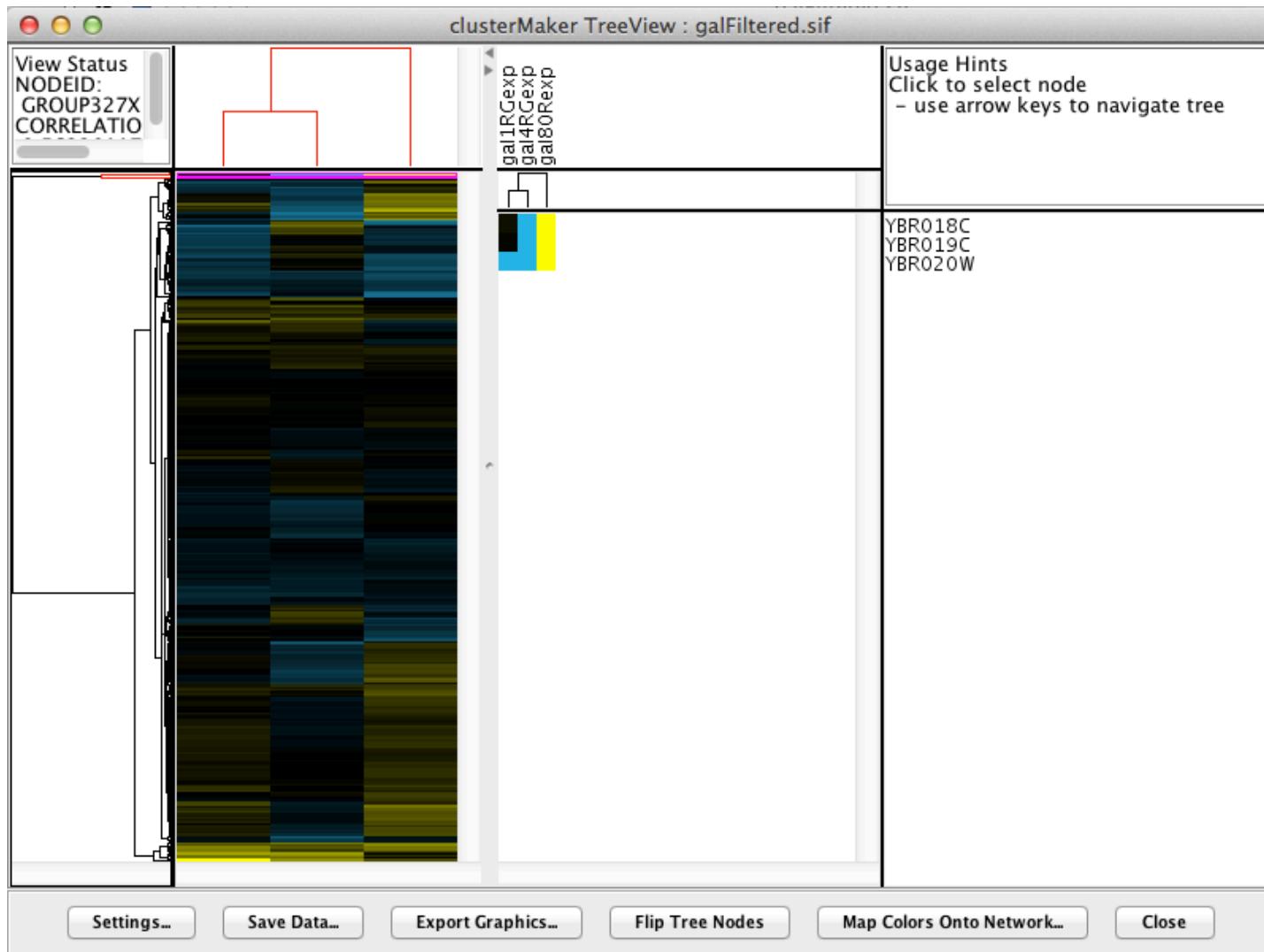
OK Cancel



– CLICK OK



Expression Data Analysis





Expression Data Analysis

BiNGO Settings
Session: /Users/scooter/Documents/galFiltered3.cys

Control Panel

Network Style Filter Filter (New) Sets

Network Nodes Edges

galFiltered.sif 331(3) 362(0)

cluster1 cluster1 72(0) 127(0)

BiNGO output

cluster1

GO_Saccharomyces cerevisiae,default,bingo,namespace close

Select All Unselect All Select nodes

Start BiNGO

The BiNGO software interface displays a metabolic pathway network and a detailed table of GO terms.

Network View: Shows a graph of biological processes. Nodes include "galactose catabolic process via UDP-galactose", "hexose catabolic process", "hexose metabolic process", "galactose metabolic process", and "monosaccharide catabolic process". Edges represent relationships between these processes.

Output Table: A detailed table of GO terms and their associated statistics.

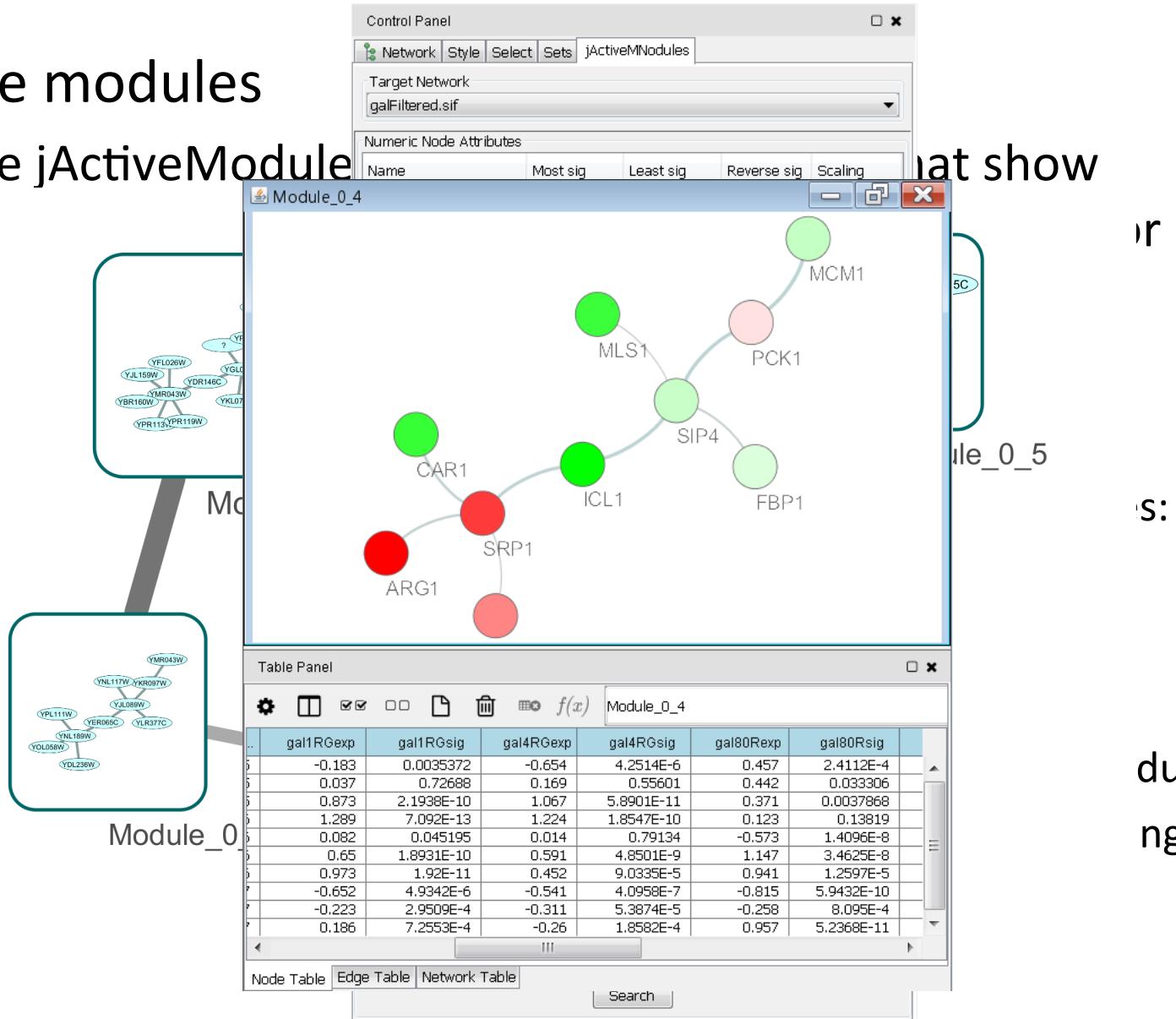
GO-ID Description	p-val	corr p-val	cluster freq	total freq	genes
33499 galactose catabolic process via UDP-galactose	2.5090...	1.9069...	3/3	100.0%	3/6208 ... YBR020W YBR018C YBR019C
19388 galactose catabolic process	5.0181...	1.9069...	3/3	100.0%	6/6208 ... YBR020W YBR018C YBR019C
6012 galactose metabolic process	4.1399...	1.0488...	3/3	100.0%	11/6208 ... YBR020W YBR018C YBR019C
19320 hexose catabolic process	7.3414...	1.3726...	3/3	100.0%	57/6208 ... YBR020W YBR018C YBR019C
46365 monosaccharide catabolic process	9.0300...	1.3726...	3/3	100.0%	61/6208 ... YBR020W YBR018C YBR019C
46164 alcohol catabolic process	1.1481...	1.4543...	3/3	100.0%	66/6208 ... YBR020W YBR018C YBR019C
44275 cellular carbohydrate catabolic process	1.9088...	2.0724...	3/3	100.0%	78/6208 ... YBR020W YBR018C YBR019C
16052 carbohydrate catabolic process	2.5677...	2.4394...	3/3	100.0%	86/6208 ... YBR020W YBR018C YBR019C
19318 hexose metabolic process	6.6970...	5.6552...	3/3	100.0%	118/6208 ... YBR020W YBR018C YBR019C
5996 monosaccharide metabolic process	9.4003...	7.1443...	3/3	100.0%	132/6208 ... YBR020W YBR018C YBR019C
11787 small molecule catabolic process	1.2782...	0.1771...	2/2	100.0%	148/6208 ... YBR020W YBR018C YBR019C



Expression Data Analysis

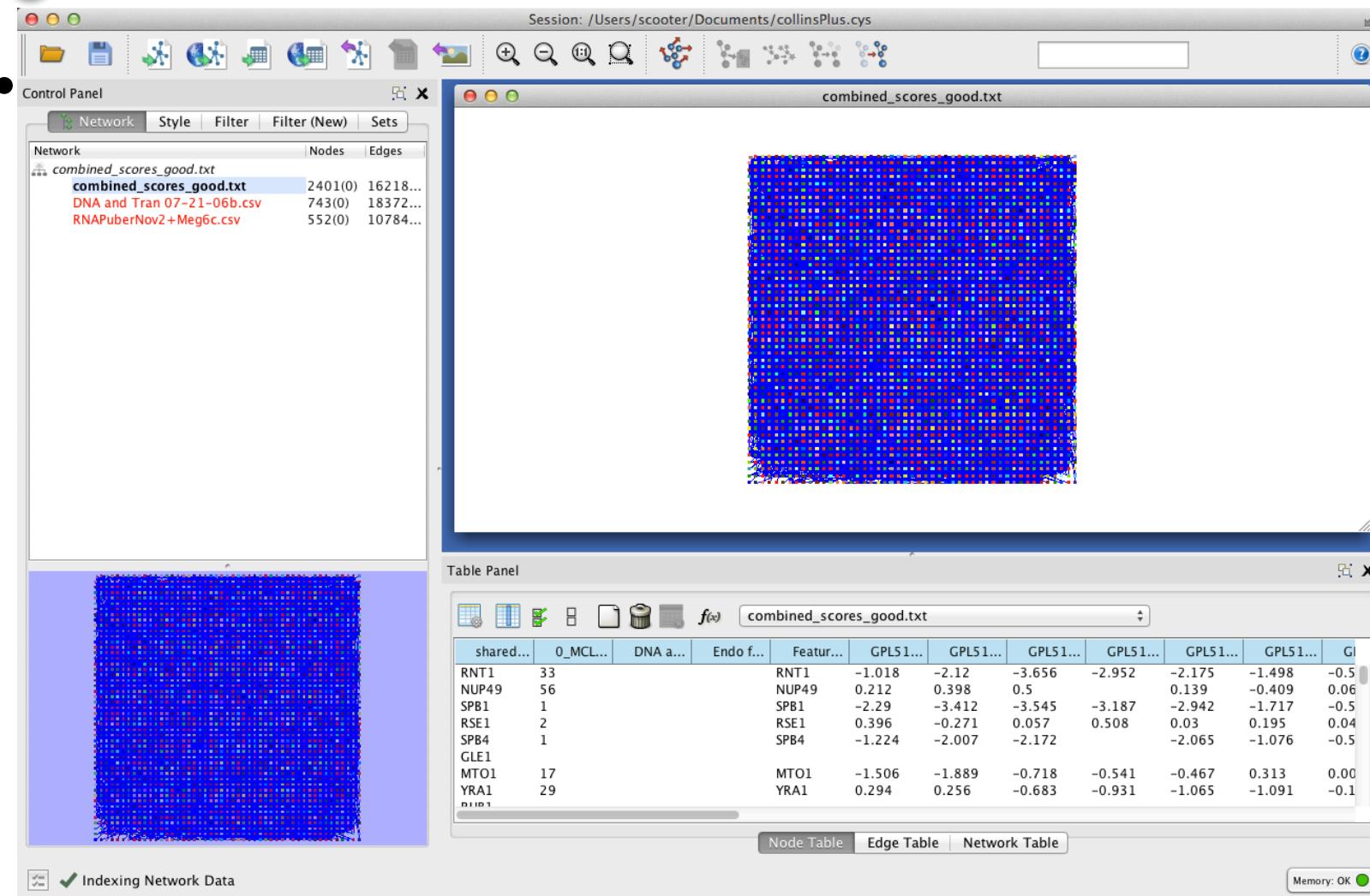
- Active modules

- The jActiveModule



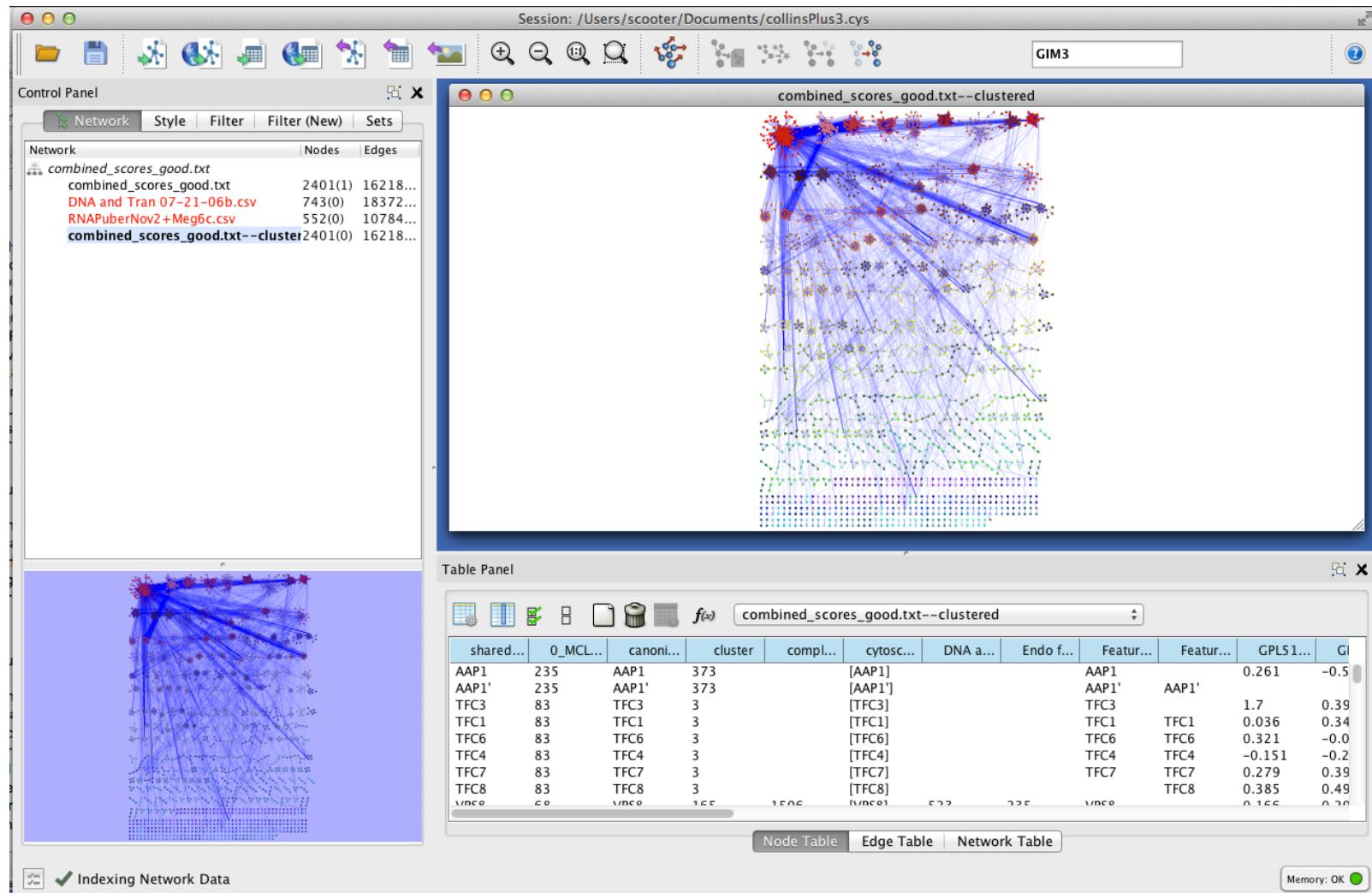


Protein Complexes





Protein Complexes



ges



Protein Complexes

e

The screenshot shows a software interface for protein complex analysis. On the left, a 'Control Panel' window displays a file tree under 'Network' with files like 'combined_scores_good.txt', 'combined_scores_good.csv', 'DNA and Tran 07-21-06b.csv', and 'RNAPuberNov2+Meg6c.csv'. Below the tree is a network visualization showing nodes and edges. A status bar at the bottom says 'Indexing Network Data' with a progress bar.

The main window is titled 'Hierarchical cluster' and contains the following settings:

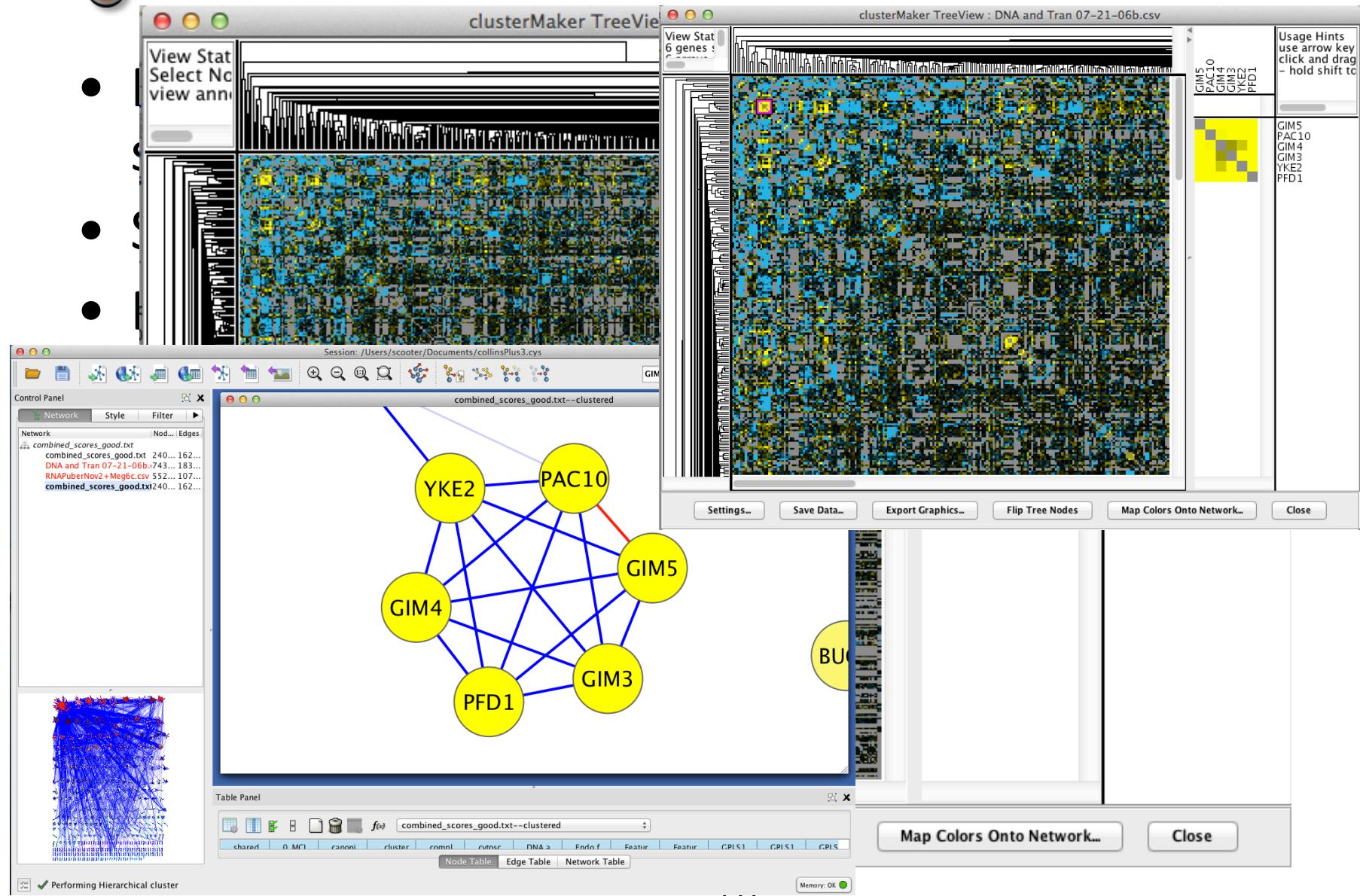
- Linkage:** pairwise average-linkage
- Distance Metric:** Uncentered correlation
- Array sources:** Node attributes for cluster (selected: --None--)
- Edge column for cluster:** DNA Strength
- Clustering Parameters:**
 - Only use selected nodes/edges for cluster:
 - Cluster attributes as well as nodes:
 - Ignore nodes/edges with no data:
- Advanced Parameters:** (with a right-pointing arrow)
- Visualization Options:**
 - Create groups from clusters:
 - Show TreeView when complete:

At the bottom are 'OK' and 'Cancel' buttons, and tabs for 'Node Table', 'Edge Table', and 'Network Table'. A status bar at the bottom right says 'Memory: OK'.

Feature	Feature	GPL51...	GI
AP1	AAP1'	0.261	-0.5
AP1'	AAP1'	1.7	0.39
FC3	TFC1	0.036	0.34
FC1	TFC6	0.321	-0.0
FC6	TFC4	-0.151	-0.2
FC4	TFC7	0.279	0.39
FC7	TFC8	0.385	0.49
FC8		0.166	0.20

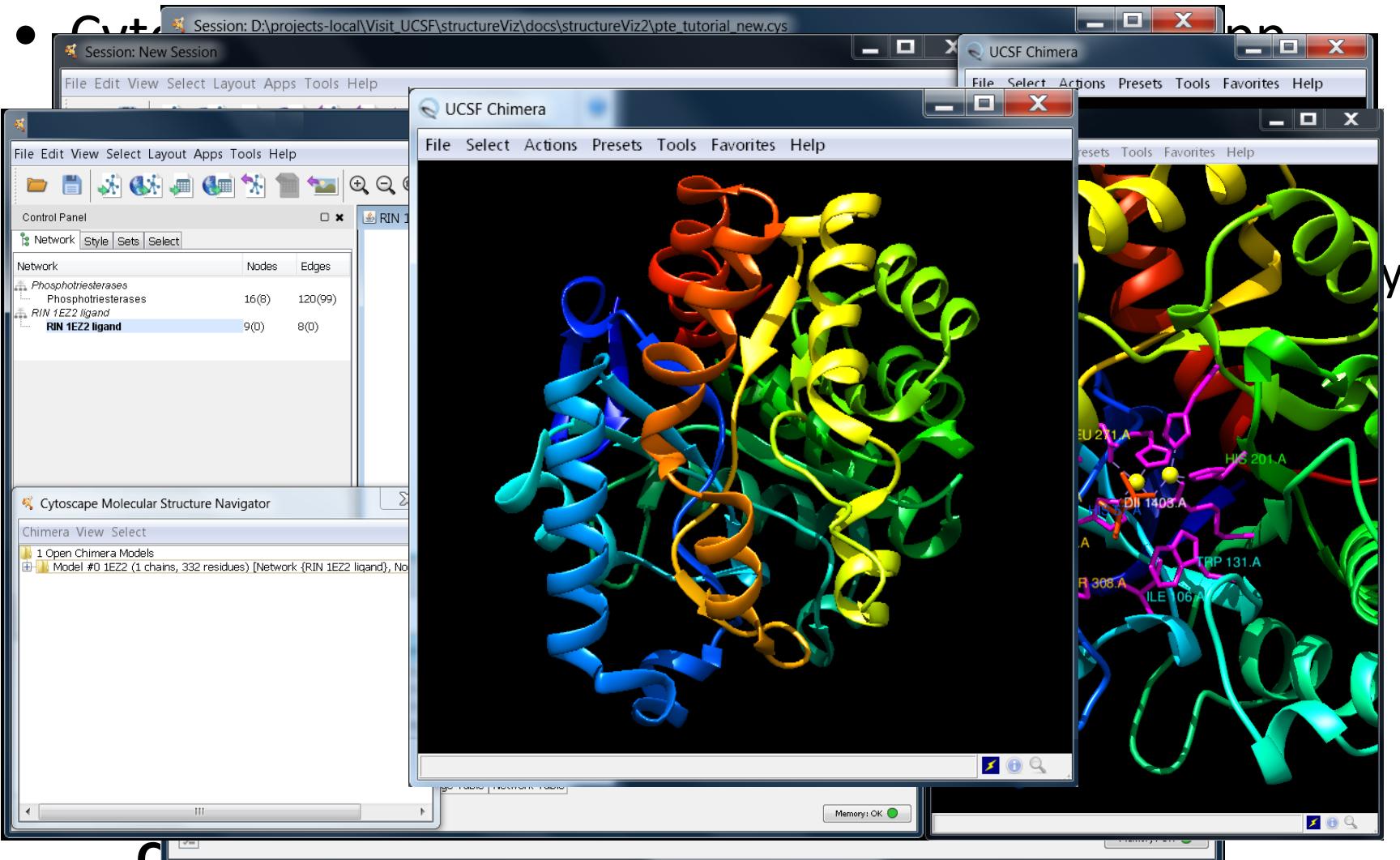


Protein Complexes





From Networks to Structures





Questions?

- scooter@cgl.ucsf.edu
- cytoscape-helpdesk@googlegroups.com