CBIO Computational Biology @ UCT

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Practical Course on "Gene/Protein Functional Networks & Interactomes"

23 and 24 November 2015 - UCT, Cape Town, South Africa

(thanks to Prof. Nicola Mulder)

Dr. Javier De Las Rivas
Cancer Research Center (CiC-IBMCC, CSIC/USAL), Salamanca, Spain

CBIO Computational Biology @ UCT

CiC

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

DAY 1

Session 1 (9:30 - 12:30, 3h)

Bioinformatic tools for Functional Enrichment Analysis (FEA)

Session 2 (13:30 - 16:30, 3h)

Construction of gene functional networks

DAY2

Session 3 (9:30 - 12:30, 3h)

Protein interaction networks

Session 4 (13:30 - 16:30, 3h)

Construction and analysis of gene/protein networks

Dr. Javier De Las Rivas

Cancer Research Center (CiC-IBMCC, CSIC/USAL), Salamanca, Spain

CBIO Computational Biology @ UCT

CiC

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Session 3 (9:30 - 12:30, 3h)
Protein interaction networks
Session 4 (13:30 - 16:30, 3h)
Construction and analysis of gene/protein networks

- From gene expression signatures to gene coexpression networks
- Definition and properties of protein interaction networks
- Visualize and analyse biomolecular networks in Cytoscape
- Using on-line tools to build gene/protein networks: APID, STRING, GeneMANIA, PSICQUIC
- Network medicine: proteins and drugs interactions (STITCH)

Networks & Pathways

Comparison and combination of these type of complex data

genes/proteins in networks

and

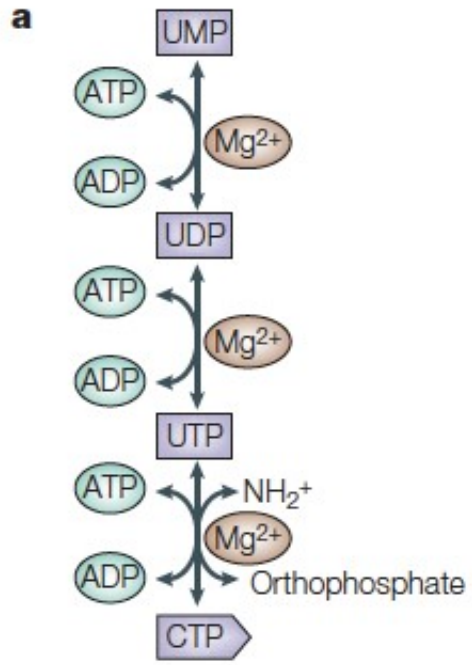
genes/proteins in pathways

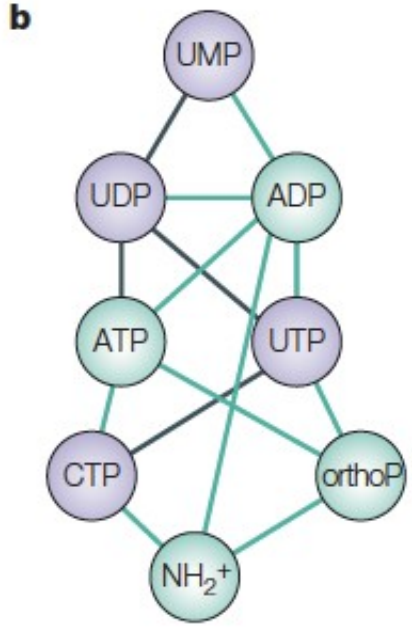
NETWORK BIOLOGY: UNDERSTANDING THE CELL'S FUNCTIONAL ORGANIZATION

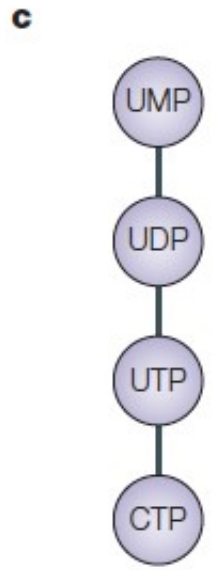
Albert-László Barabási & Zoltán N. Oltvai‡*

A key aim of postgenomic biomedical research is to systematically catalogue all molecules and their interactions within a living cell. There is a clear need to understand how these molecules and the interactions between them determine the function of this enormously complex machinery, both in isolation and when surrounded by other cells. Rapid advances in network biology indicate that cellular networks are governed by universal laws and offer a new conceptual framework that could potentially revolutionize our view of biology and disease pathologies in the twenty-first century.

From ***Barabasi et al. (2004) Nature Reviews Genetics 5, 101-113.***





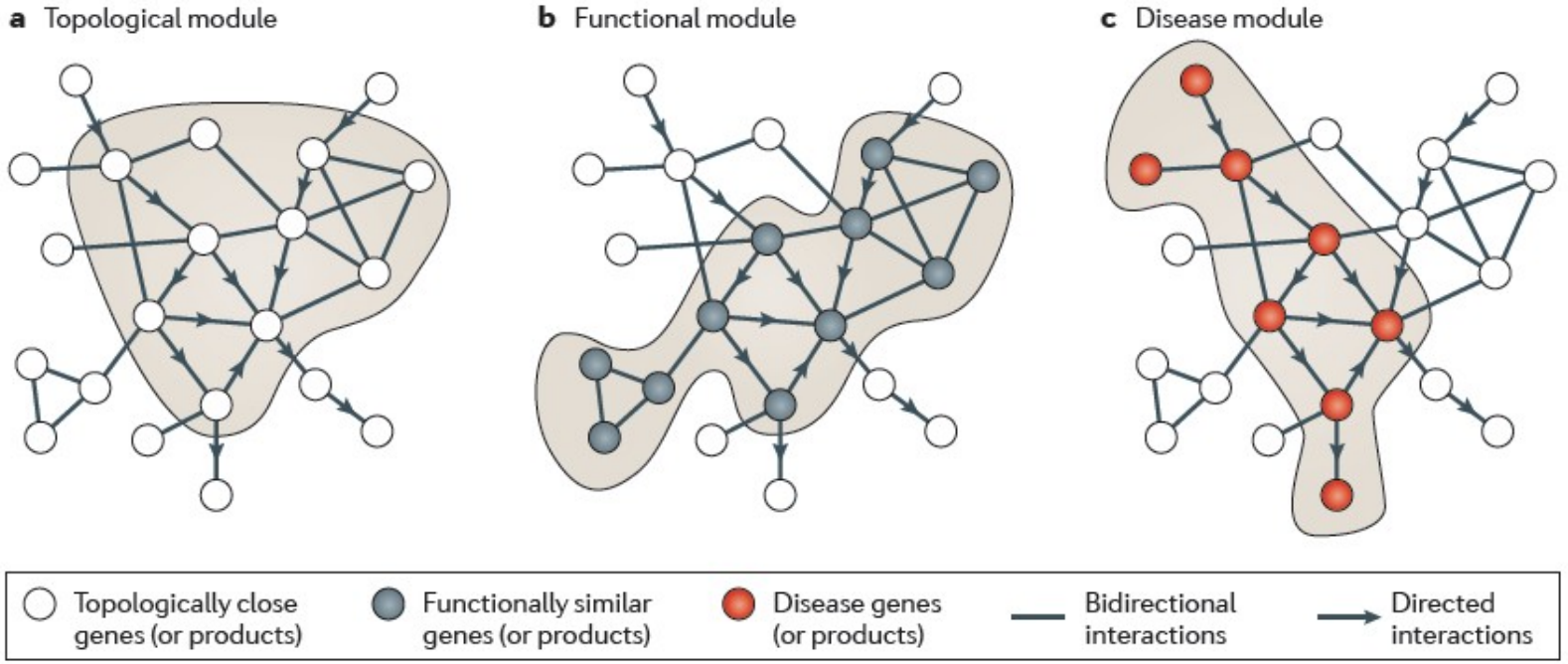


Network medicine: a network-based approach to human disease

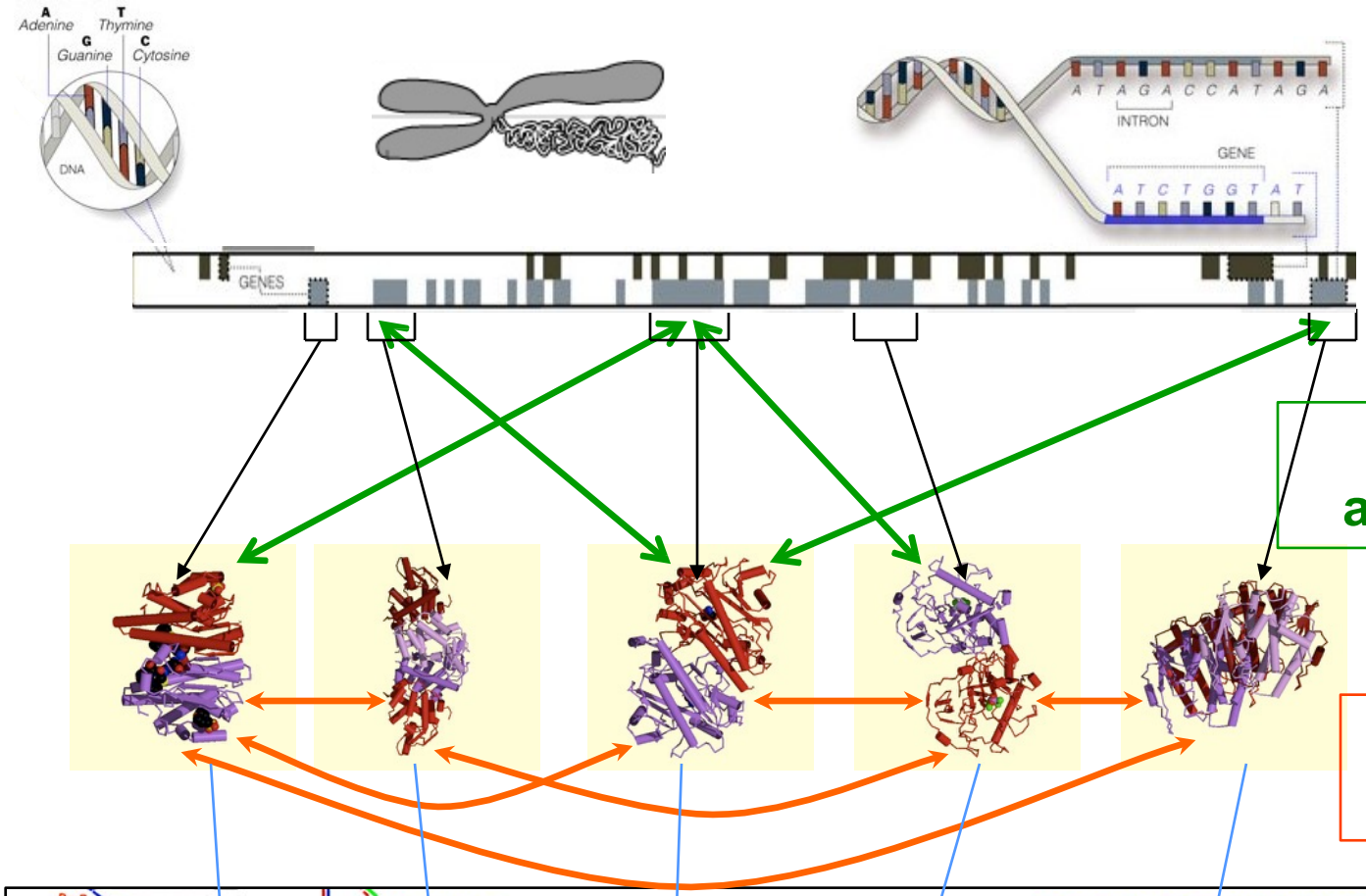
Albert-László Barabási [*‡§], *Natali Gulbahce* [*‡||] *and Joseph Loscalzo*[§]

Abstract | Given the functional interdependencies between the molecular components in a human cell, a disease is rarely a consequence of an abnormality in a single gene, but reflects the perturbations of the complex intracellular and intercellular network that links tissue and organ systems. The emerging tools of network medicine offer a platform to explore systematically not only the molecular complexity of a particular disease, leading to the identification of disease modules and pathways, but also the molecular relationships among apparently distinct (patho)phenotypes. Advances in this direction are essential for identifying new disease genes, for uncovering the biological significance of disease-associated mutations identified by genome-wide association studies and full-genome sequencing, and for identifying drug targets and biomarkers for complex diseases.

From ***Barabasi et al. (2011) Nature Reviews Genetics 12, 56-68.***



Biomolecular complexity of living systems

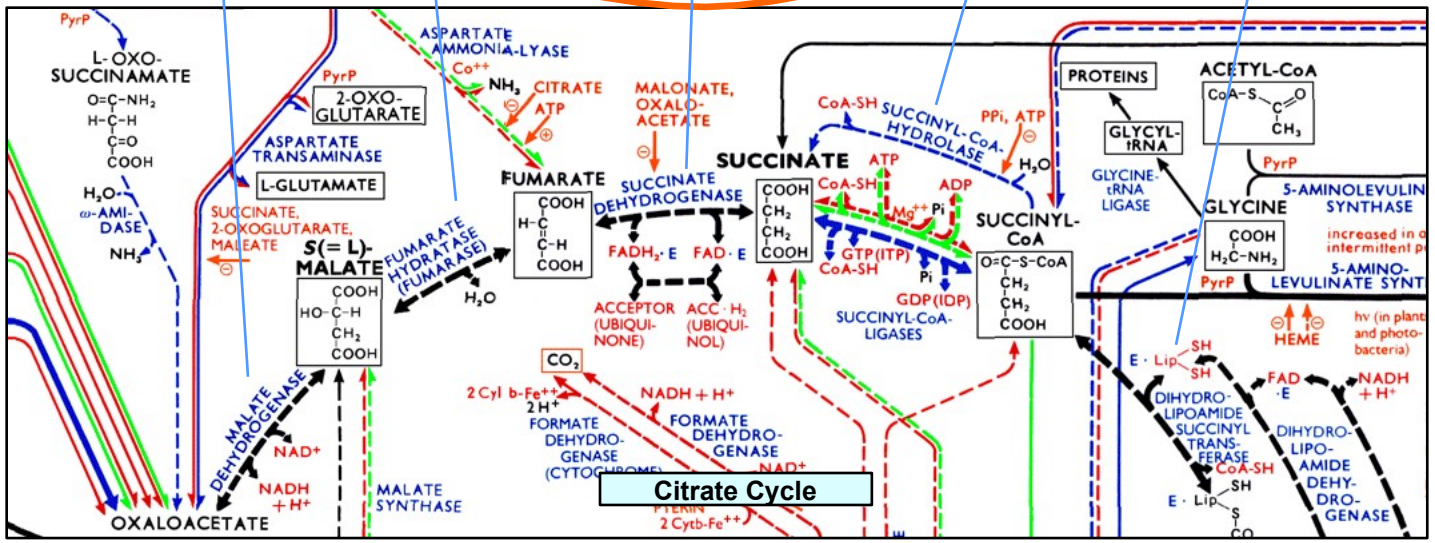


GENOME
TRANSCRIPTOME

genes/RNAs
actions & relations

PROTEOME

proteins
actions & relations

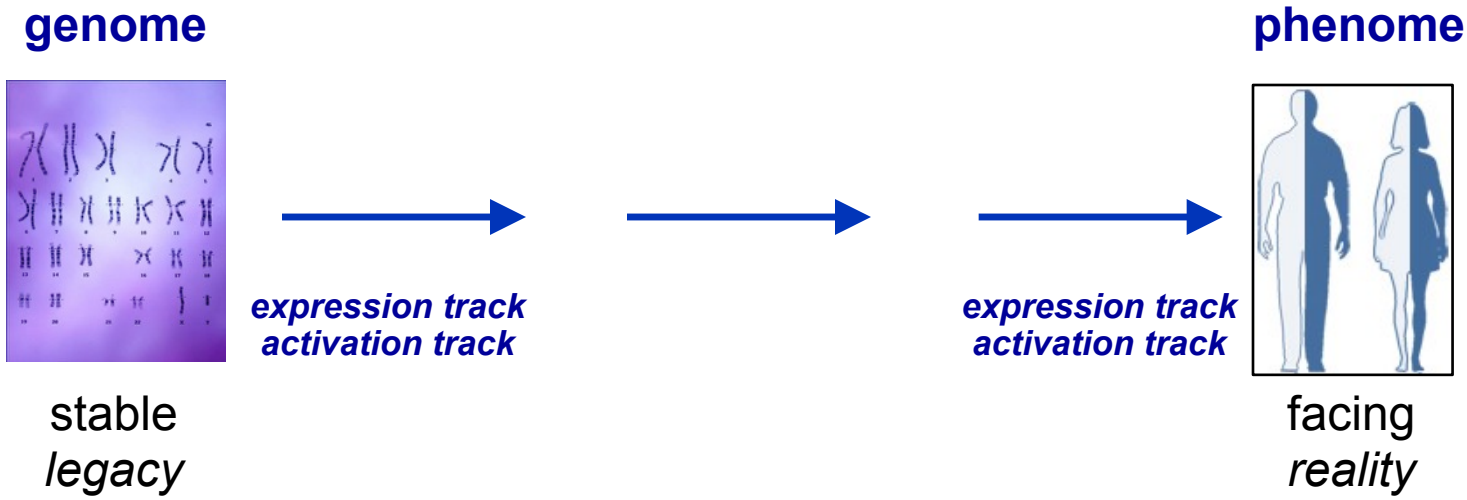


METABOLOME
INTERACTOME

structural, metabolic or signaling (molecular interactions)
protein-ligand actions & relations

Omics era: unraveling biological complexity

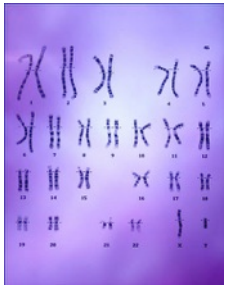
the paradox of the "genome alone"



Omics era: unraveling biological complexity

the paradox of the "genome alone"

genome



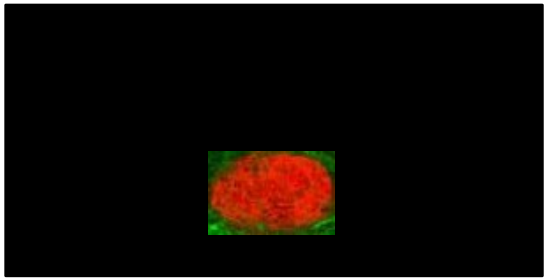
stable
legacy

expression track
activation track

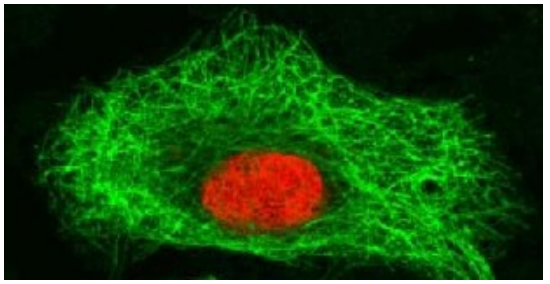
expression track
activation track

phenome

facing
reality



genome



living system

Genomes …

Is there a simple "genome factor"?

Organism	Genome Genes	Genome Factor
Bacteria	3.000	
		x2
Yeast	6.000	
		x3
Worm	18.000	
		x2
Human	36.000	x12

From the mere genome numbers **HUMAN** is only about 12 times **BACTERIA**

BIOLOGY includes two other key factors:

➔ **Cellular Factor**

1 bacteria is 1 cell

1 human is 10^9 cells (and more than 300 cell-types)

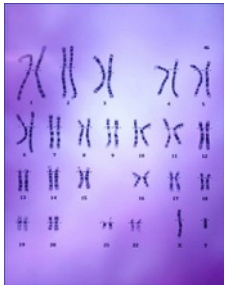
➔ **Relational Factor**

By interaction and relations the number of possible outputs grows exponentially

Omics era: unraveling biological complexity

proteins constitute the *keystones* of the cellular machinery

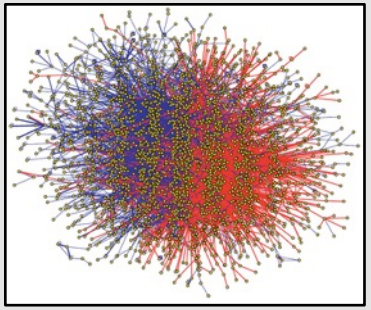
genome



stable
legacy

expression track
active track

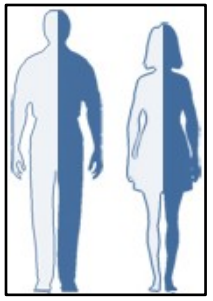
interactome



working-moving
machinery

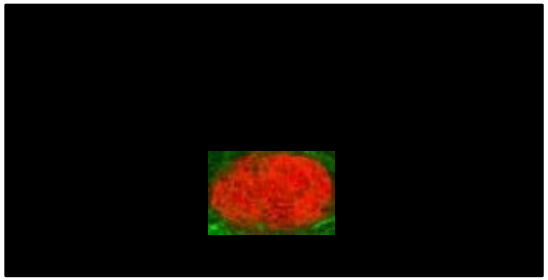
expression track
active track

phenome



facing
reality

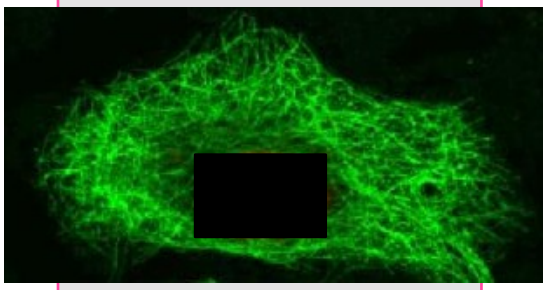
genes



genome

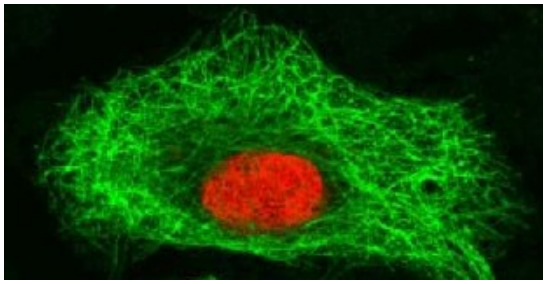
+

proteins



cellular machinery

=

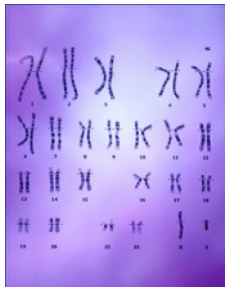


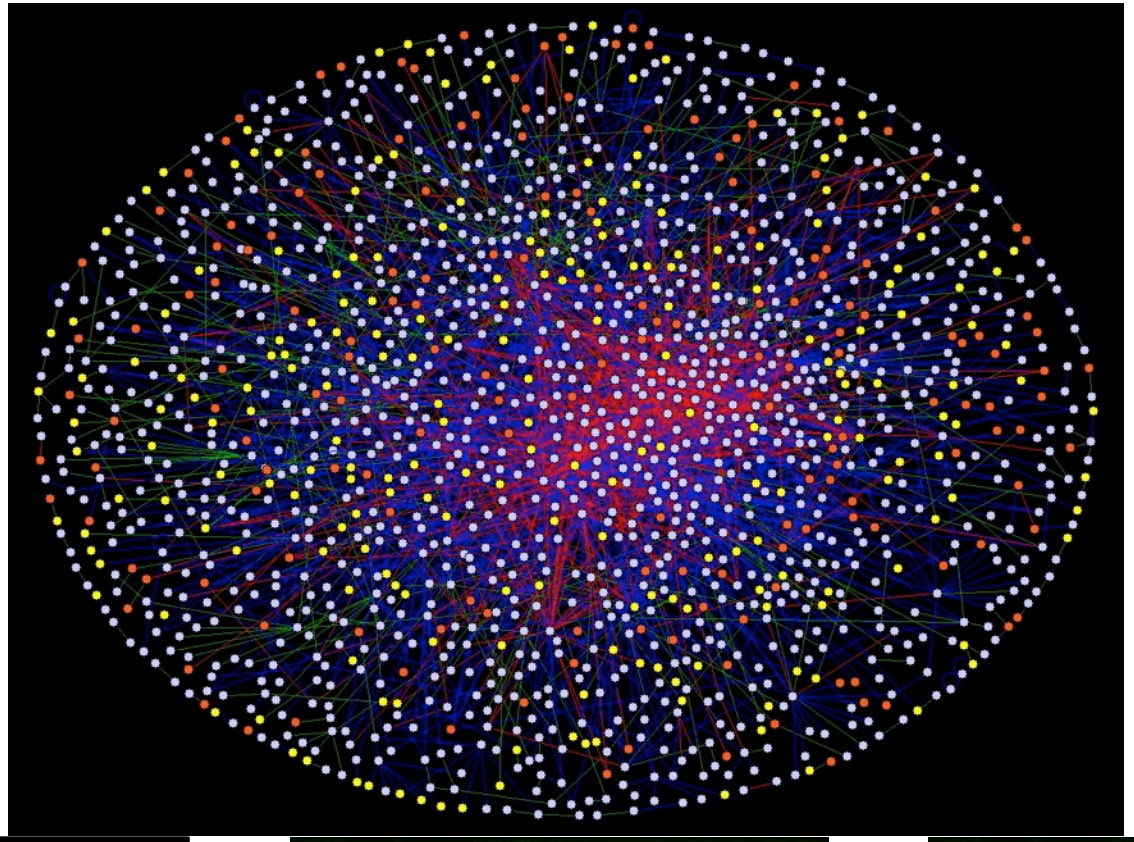
living system

Omics era: unraveling biological complexity

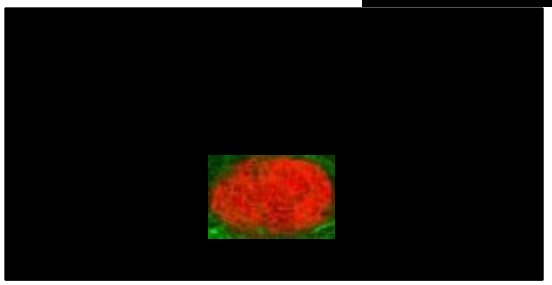
interactions (gene2gene, prot2prot) ... cellular machinery dynamics

genome

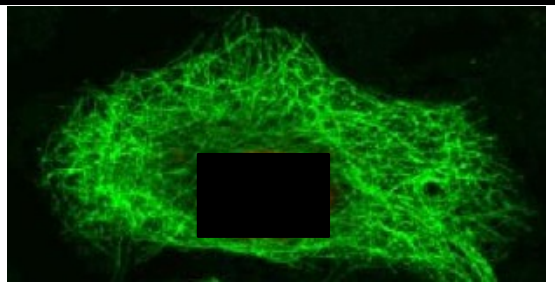




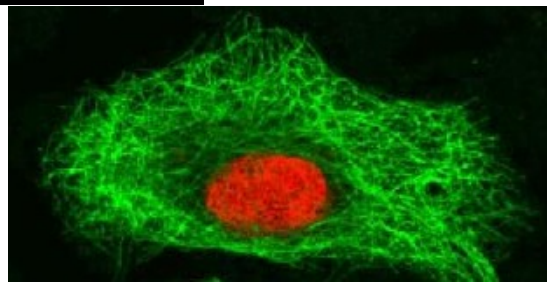
phenome



+



=

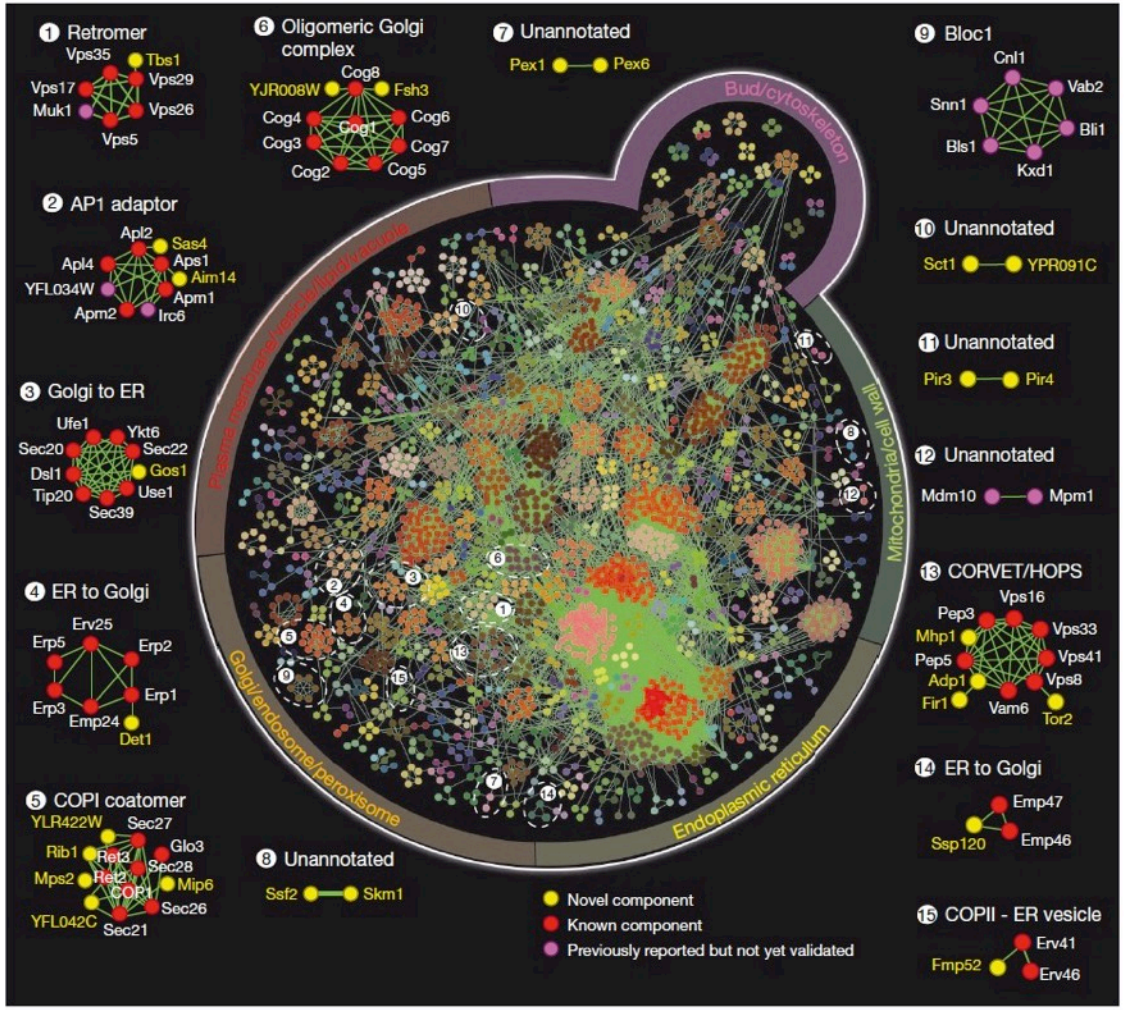


genome **cellular machinery** **living system**

Omics era: unraveling biological complexity

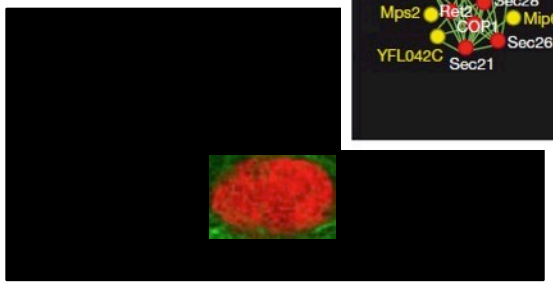
interactome ... network of interacting proteins

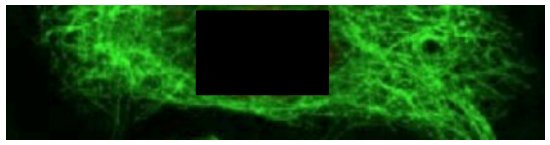
genome

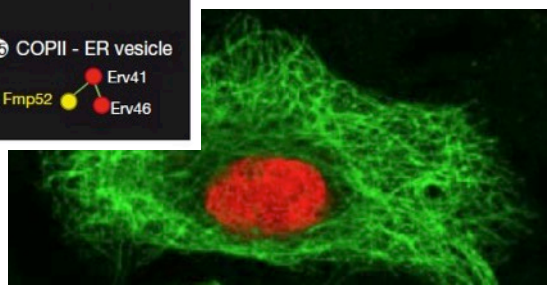


phenome

Babu et al. (2012) *Nature*







genome

molecular interactome

living system

CBIO Computational Biology @ UCT

CiC

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Session 3 (9:30 - 12:30, 3h)

Protein interaction networks

Session 4 (13:30 - 16:30, 3h)

Construction and analysis of gene/protein networks = biomolecular networks

- From gene expression signatures to gene coexpression networks
- Definition and properties of protein interaction networks
- Visualize and analyse biomolecular networks in Cytoscape
- Using on-line tools to build gene/protein networks: APID, STRING, GeneMANIA, PSICQUIC
- Network medicine: proteins and drugs interactions (STITCH)

Omics era: unraveling biological complexity

mapping **biological networks**

REVIEW

Zhu et al. (2007) Genes Dev.

Getting connected: analysis and principles of biological networks

Xiaowei Zhu,[1,2] Mark Gerstein,[3] and Michael Snyder[1,2,4]

How can we characterize **biomolecular networks** and measure parameters that allow to understand the role of different nodes & edges in a given network ? (graph & network theory)

Type of network	Species	Number of nodes	Number of interactions	Reference
Transcription factor-binding network	*S. cerevisiae*	3528	7419	Yu et al. 2003[a]
		3207	11231	Harbison et al. 2004[b]
Protein–protein interaction	*C. elegans*	2788	4441	Stark et al. 2006
	D. melanogaster	7546	25403	
	Homo sapiens	7509	20979	
	Mus musculus	209	393	
	S. cerevisiae	5325	51773	
Phosphorylation network	*S. cerevisiae*	1325	4200	Ptacek et al. 2005
Metabolic network	*E. coli*	473	574	Guimera and Nunes Amaral 2005
	S. cerevisiae	646	1149	Tong et al. 2004
Genetic network	*S. cerevisiae*	3258	13963	Reguly et al. 2006[c]

[a]Transcriptional factor-binding data collected at rich-media condition.
[b]Transcriptional factor-binding data collected at a variety of growth conditions.
[c]Synthetic lethal interactions among nonessential genes.

From ***Zhu et al. (2007) Genes Dev.***

Networks

Two major types of networks derived from experimental data

Two major types of networks derived from large-scale *omic* data

1.– **Gene Coexpression Networks**: ***ggcoe***
derived from gene expression profiling and transcriptomic studies

2.– **Protein-Protein Interaction Networks**: ***ppi***
derived from proteomic studies

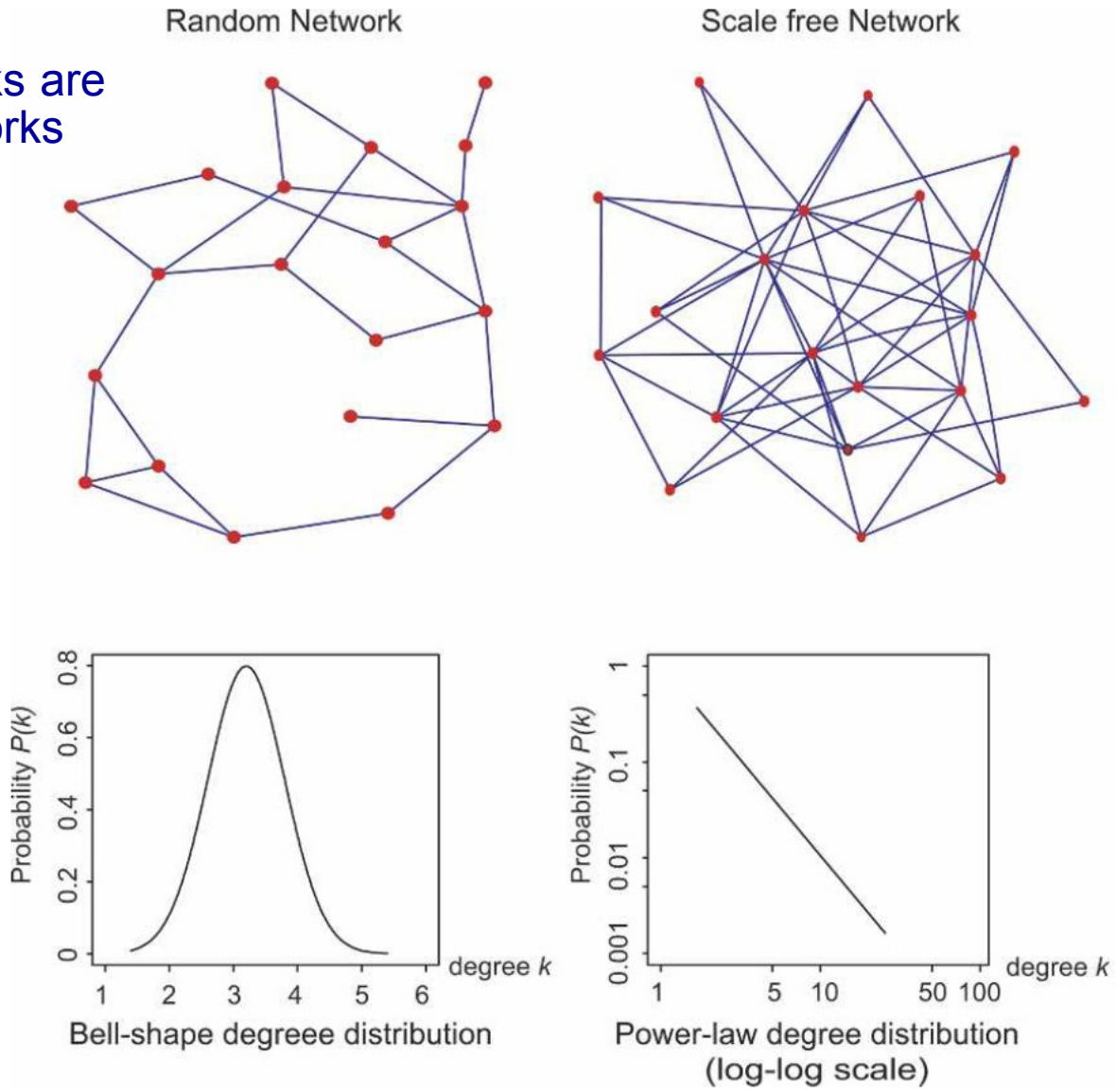
Omics era: unraveling biological complexity

mapping **biological networks**

The ***ggcoe*** and ***ppi*** networks are complex biomolecular networks

Biomolecular networks are **scale-free**

A scale-free network has more high-degree nodes and a power-law degree distribution, which leads to a straight line when plotting the total number of nodes with a particular degree versus that degree in log-log scales

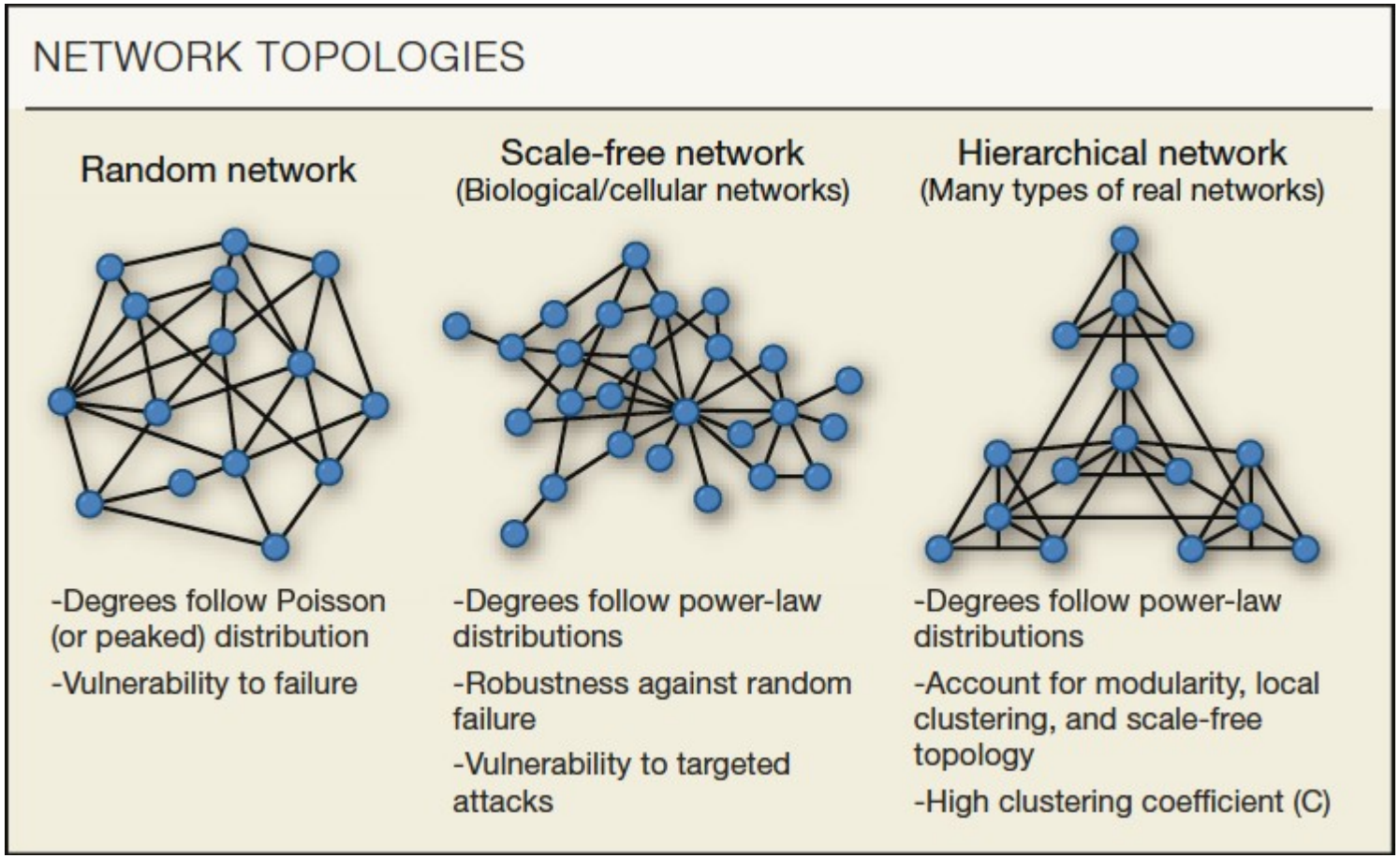


From ***Zhu et al. (2007) Genes Dev.***

Omics era: unraveling biological complexity

mapping biological networks

Biomolecular networks are **scale-free**: A scale-free network has more high-degree nodes and a power-law degree distribution, which leads to a straight line when plotting the total number of nodes with a particular degree versus that degree in log-log scales.



From ***Seebacher & Gavin (2011) Cell***

Omics era: unraveling biological complexity

mapping biological networks

Biological networks derived from PPIs are not randomly organized but rather have a scale-free format, containing a small number of nodes (hubs) with many connections (*Barabasi and Oltvai 2004*).

This organization was originally discovered in World Wide Web (www) interactions and later found to exist in biological networks (*Barabasi and Albert 1999; Jeong et al. 2000, 2001; Guelzim et al. 2002; Tong et al. 2004*).

Compared with a bell-shaped degree distribution in random networks, scale-free networks have a typical power law distribution: a fat-tailed distribution in which there are vertices with high degrees termed hubs. The advantage of this type of organization is that the system is more robust: random loss of individual non-hub vertices is less disruptive in scale-free networks than random networks.

Network topology plays a vital role in understanding network architecture and performance. It is important to know the most important and commonly used topological parameters that can be calculated in a network.

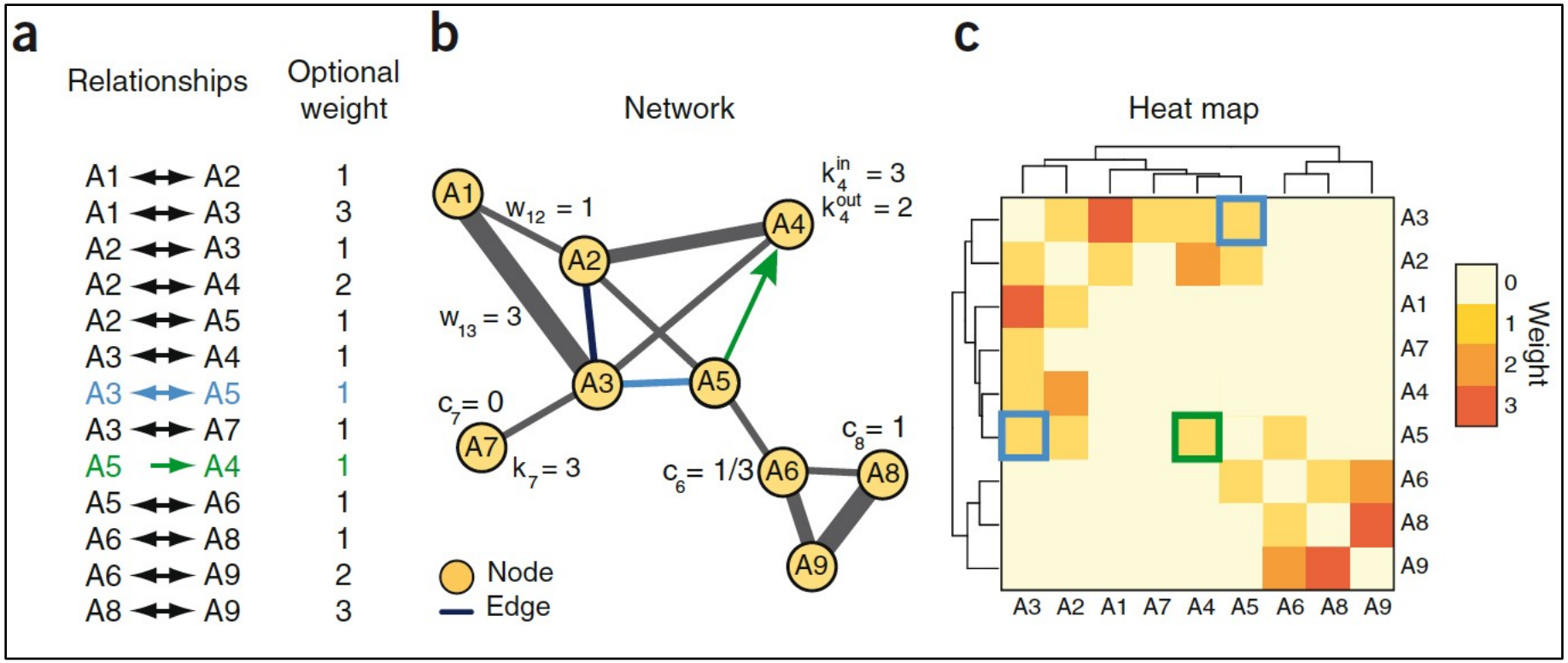
From ***Zhu et al. (2007) Genes Dev.***

Omics era: unraveling biological complexity

mapping biological networks

Networks with **undirected links**: ***ggcoe*** and ***ppi*** biological networks

Representation and visualization of networks

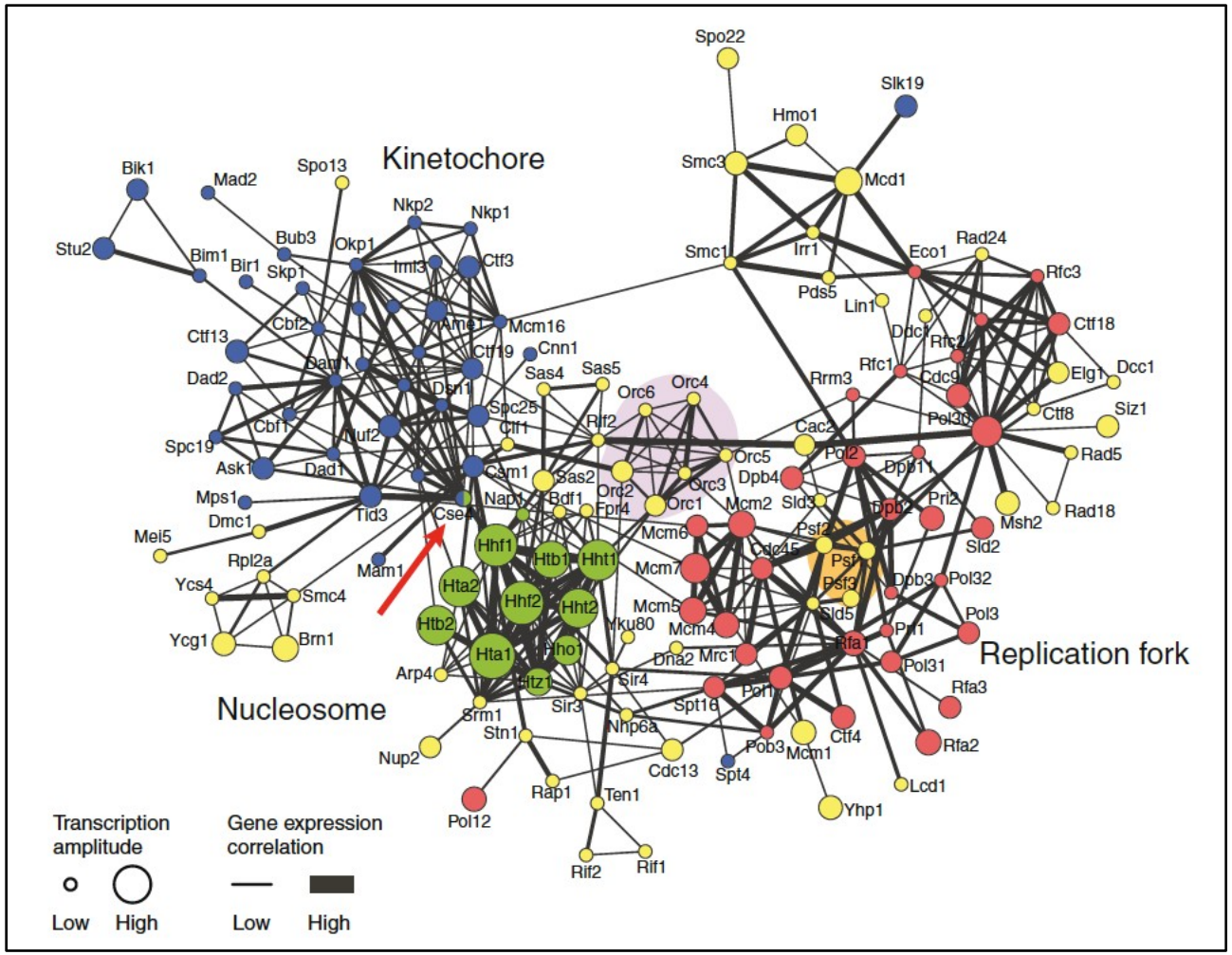


From ***Merico, Gfeller & Bader (2010) Nature Biotechnology***

Omics era: unraveling biological complexity

mapping **biological networks**

Example of *gene to gene coexpression network*: ***ggcoe***

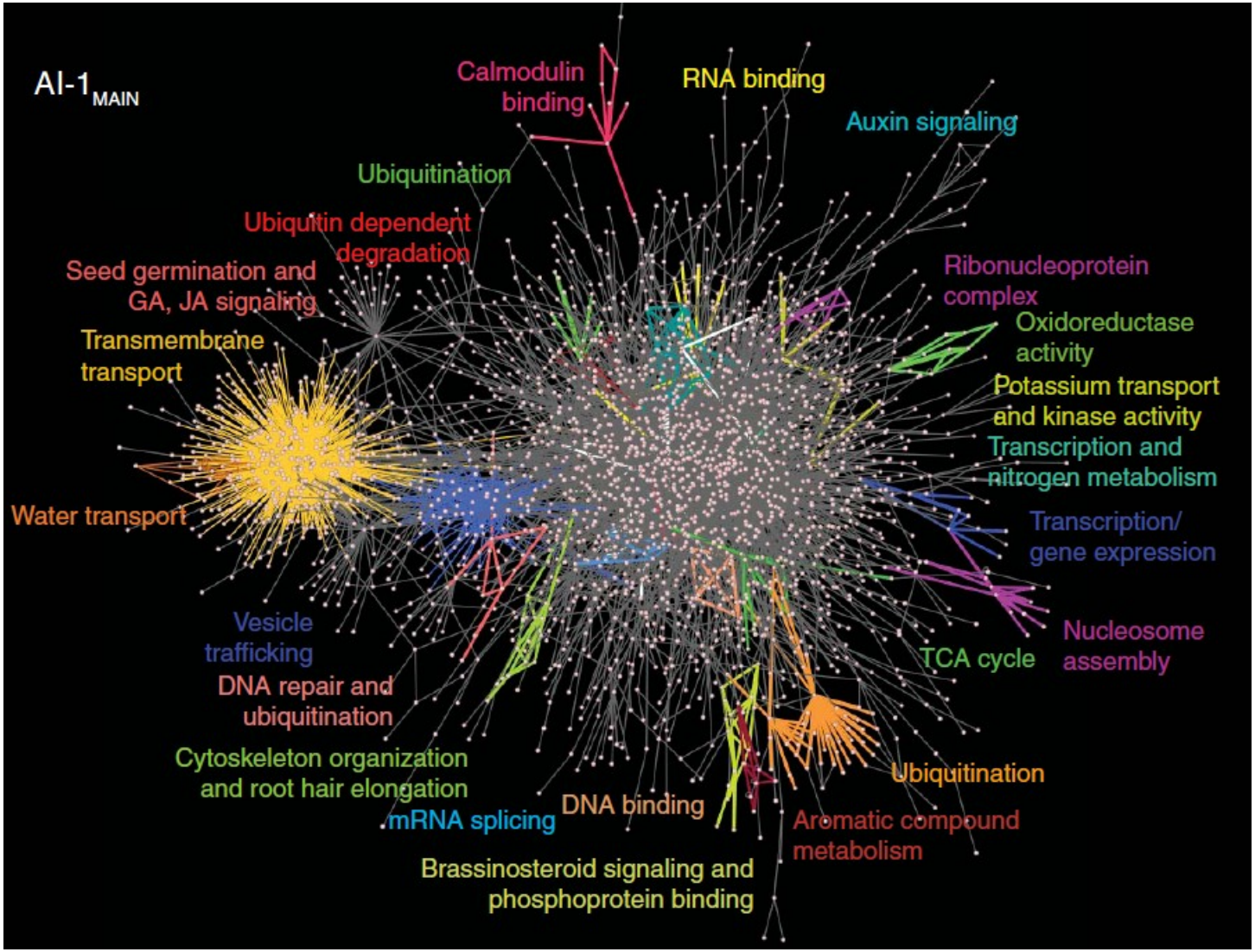


From ***Merico, Gfeller & Bader (2010) Nature Biotechnology***

Omics era: unraveling biological complexity

mapping **biological networks**

Example of *protein to protein interaction network*: ***ppi***

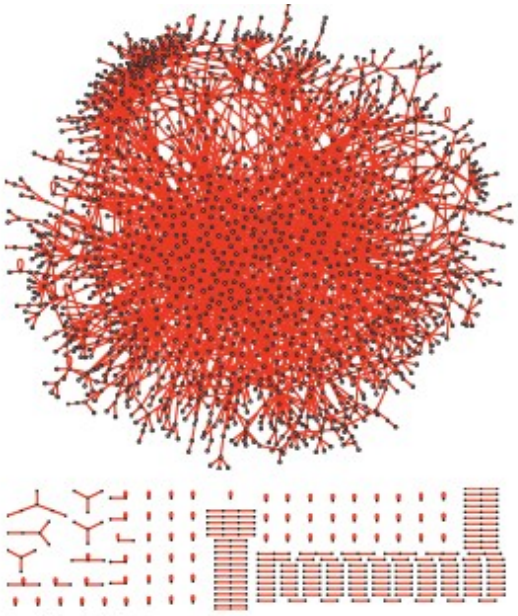


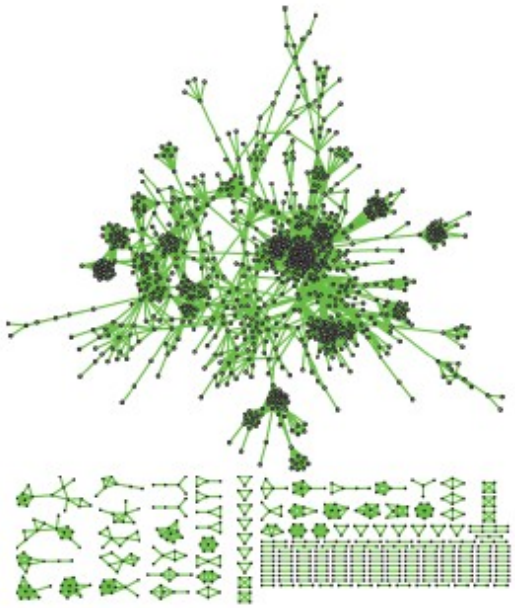
From ***Braun et al. (2011) Science***

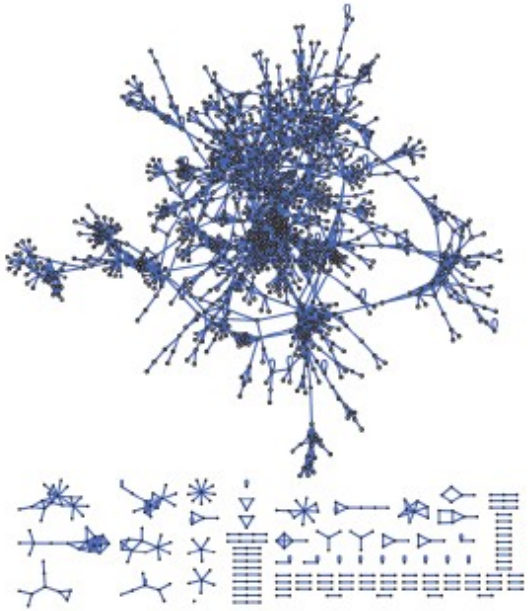
Omics era: unraveling biological complexity

mapping **biological networks**

Example of *protein to protein interaction network*: ***ppi***





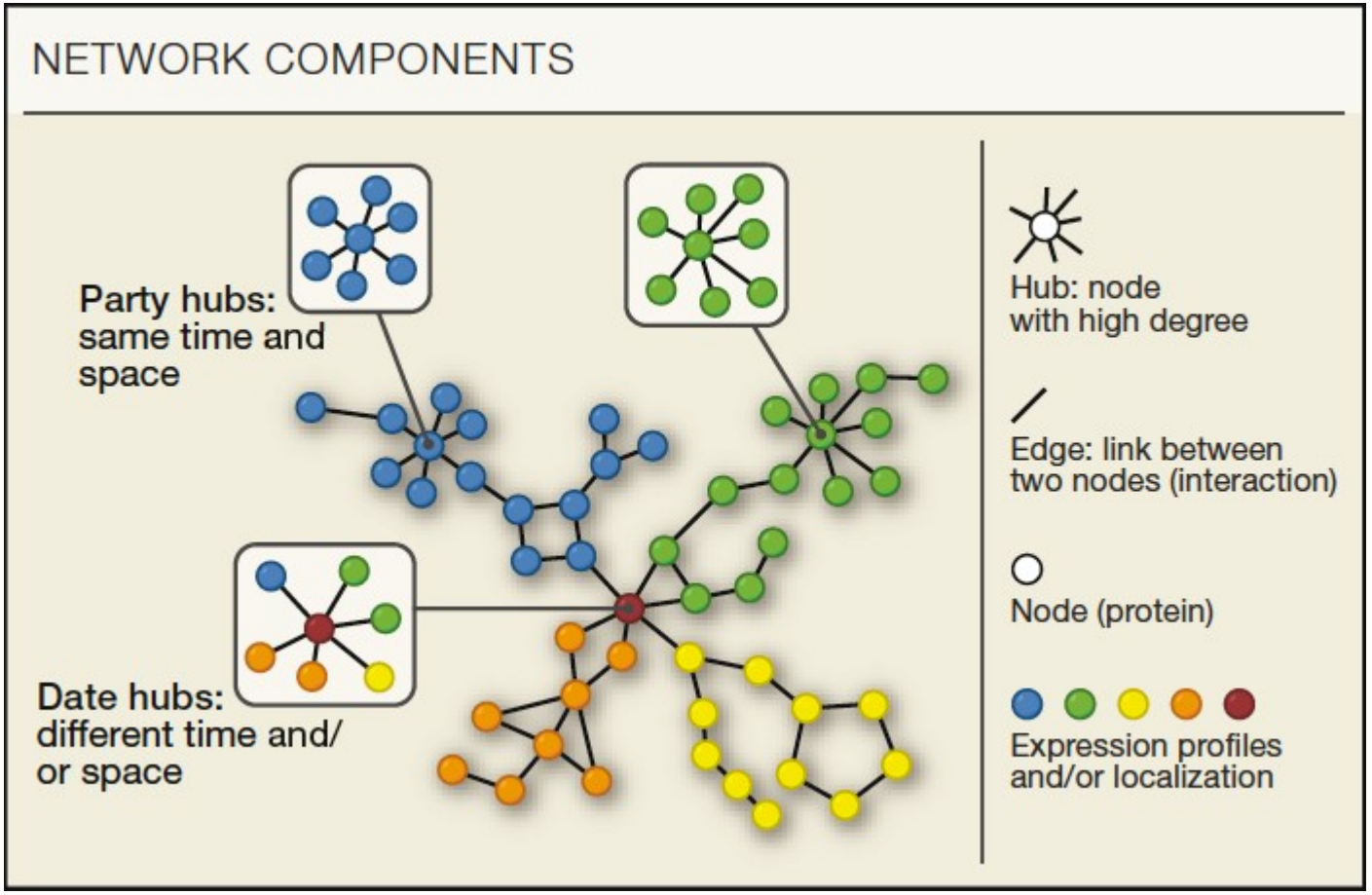


Omics era: unraveling biological complexity

mapping biological networks

Networks with **undirected links**: ***ggcoe*** and ***ppi*** biological networks

Representation and visualization of networks: **hubs**



From ***Seebacher & Gavin (2011) Cell***

Omics era: unraveling biological complexity

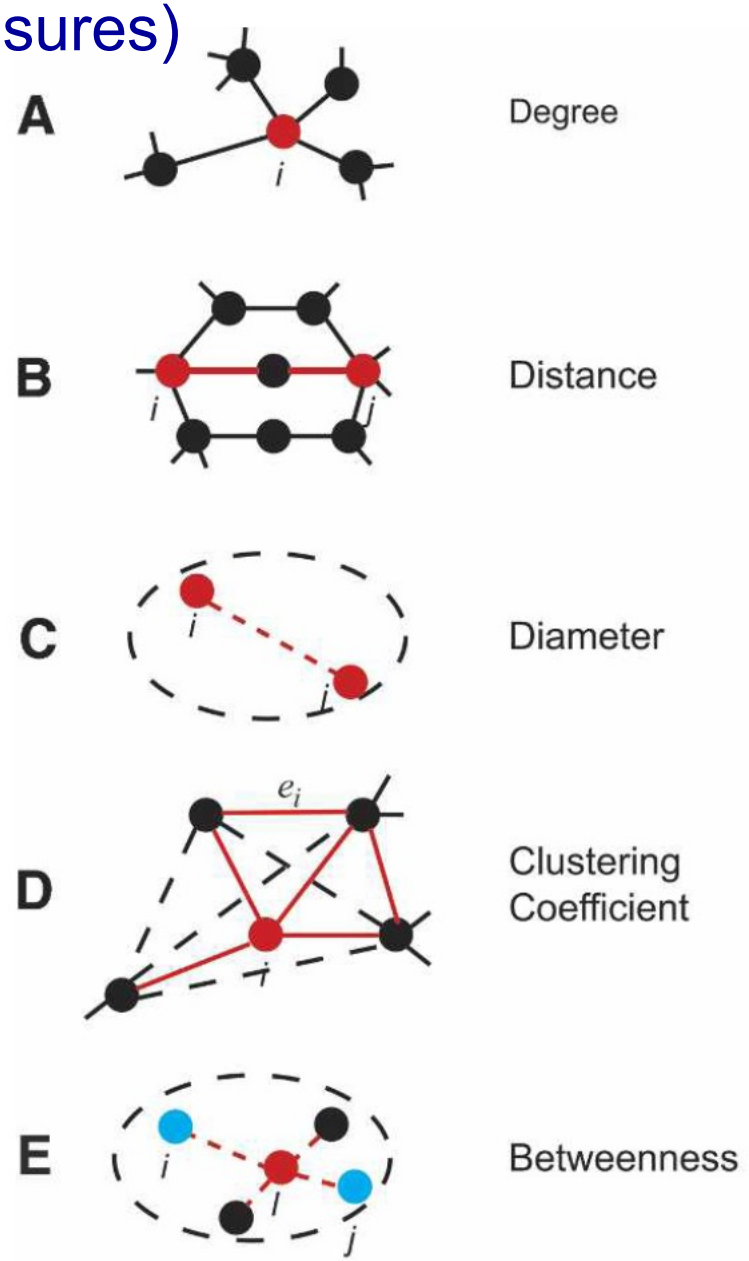
topological parameters (network measures)

The ***ggcoe*** and ***ppi*** networks are complex biomolecular networks

Network topology plays a vital role in understanding network architecture and performance.

Several of the most important and commonly used topological parameters include:

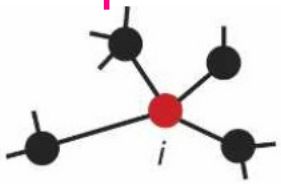
- **degree** number of links connected to 1 vertex
- **distance** shortest path length
- **diameter** maximum distance between any two nodes
- **clustering coefficient** number of links between the vertices within its neighborhood divided by the number of possible links between them
- **betweenness** fraction of the shortest paths between all pairs of vertices that pass through one vertex or link



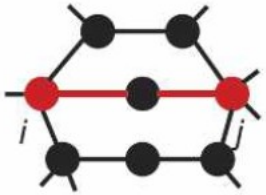
From ***Zhu et al. (2007) Genes Dev.***

Omics era: unraveling biological complexity

topological parameters (network measures)

A 

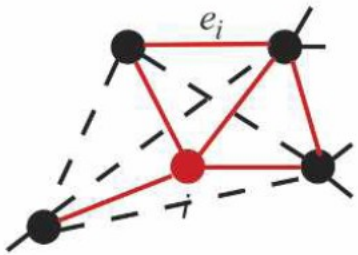
Degree k_i = number of links connected to node i

B 

Distance d_{ij} = shortest path length between node i and j

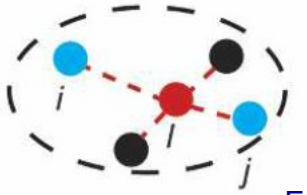
C 

Diameter $D = \max\{d_{ij} \mid i, j \in N\}$ N : all nodes in the network

D 

Clustering Coefficient $c_i = \frac{2e_i}{k_i(k_i-1)}$

e_i : number of existing links (labeled in red) among the k_i nodes that connect to node i

E 

Betweenness $b_l = \sum_{ij} p_{ij}(l)/p_{ij}$

p_{ij} : number of shortest paths between i and j

$p_{ij}(l)$: number of shortest paths between i and j going through node l

From ***Zhu et al. (2007) Genes Dev.***

Omics era: unraveling biological complexity

topological parameters (network measures)

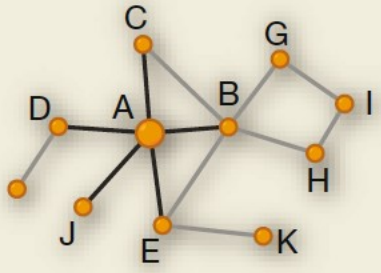
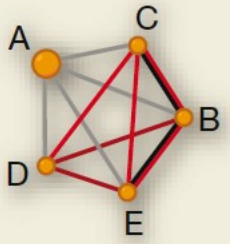
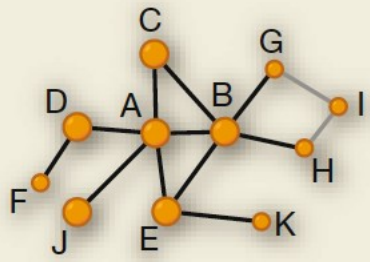
Network measures related to "number of friends" (connectivity):

– **degree** = connectivity

– **clustering coefficient** = inter-connectivity

– **assortativity** = average nearest neigbor's connectivity

NETWORK MEASURES

Degree/ connectivity (k)	Clustering coefficient/ interconnectivity (C)	Assortativity/average nearest neighbor's connectivity (NC)
 k_A=Nb of edges through A=5	 $C_A = \frac{\text{Actual links between A's neighbors (black)}}{\text{Possible links between A's neighbors (orange)}}$ $C_A=n_A/[k_A(k_A-1)/2]$ $=2/[4x(4-1)/2]=0.333$	 $NC_A=(k_B+k_C+k_D+k_E+k_J)/5$ $=(5+2+2+3+1)/5=2.6$

From ***Seebacher & Gavin (2011) Cell***

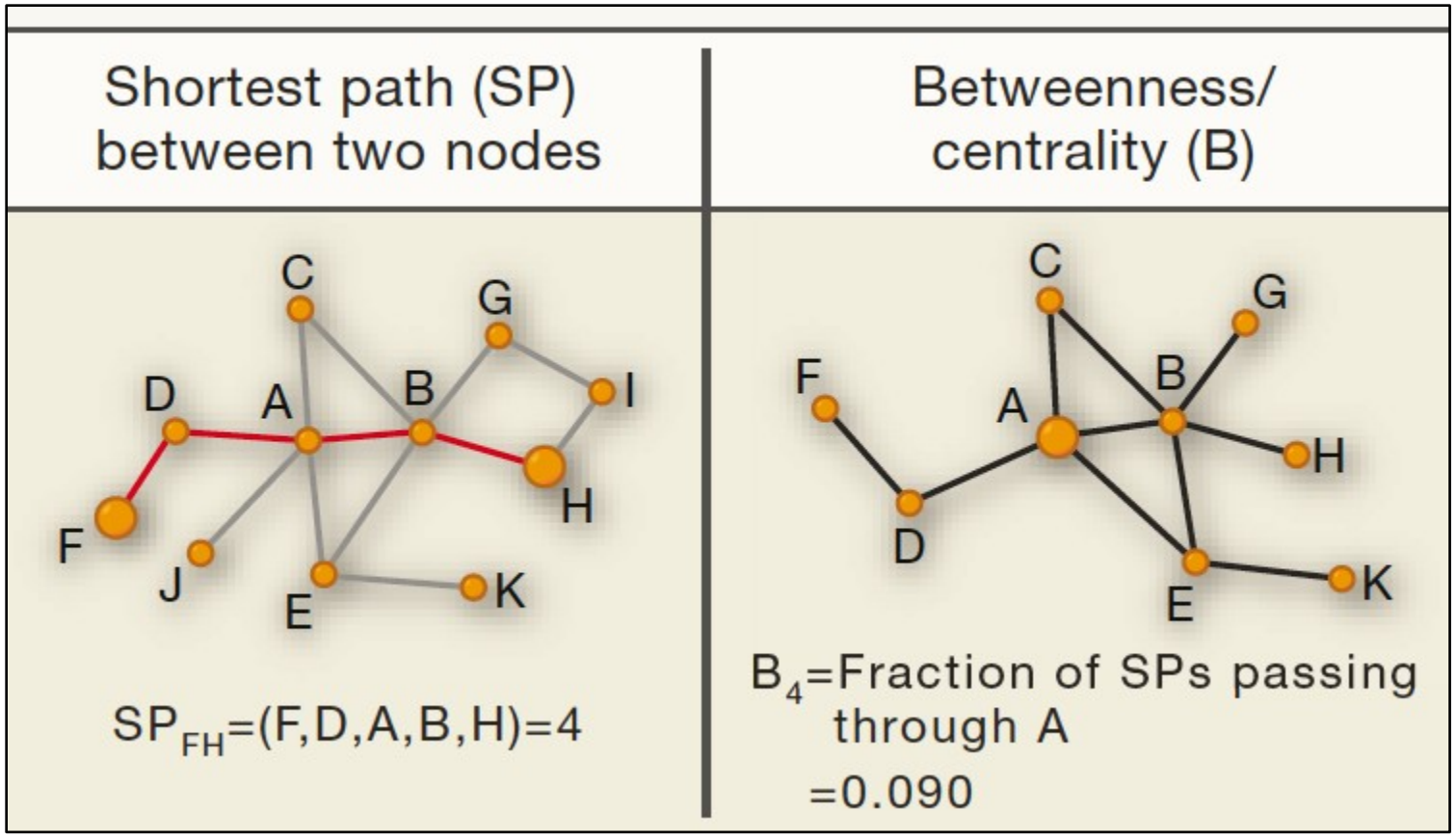
Omics era: unraveling biological complexity

topological parameters (network measures)

Network measures related to "number of ways" (path-ways):

– **shortest path**

– **betweenness** = centrality



From ***Seebacher & Gavin (2011) Cell***

Session 3 (9:30 - 12:30, 3h)

Protein interaction networks

Session 4 (13:30 - 16:30, 3h)

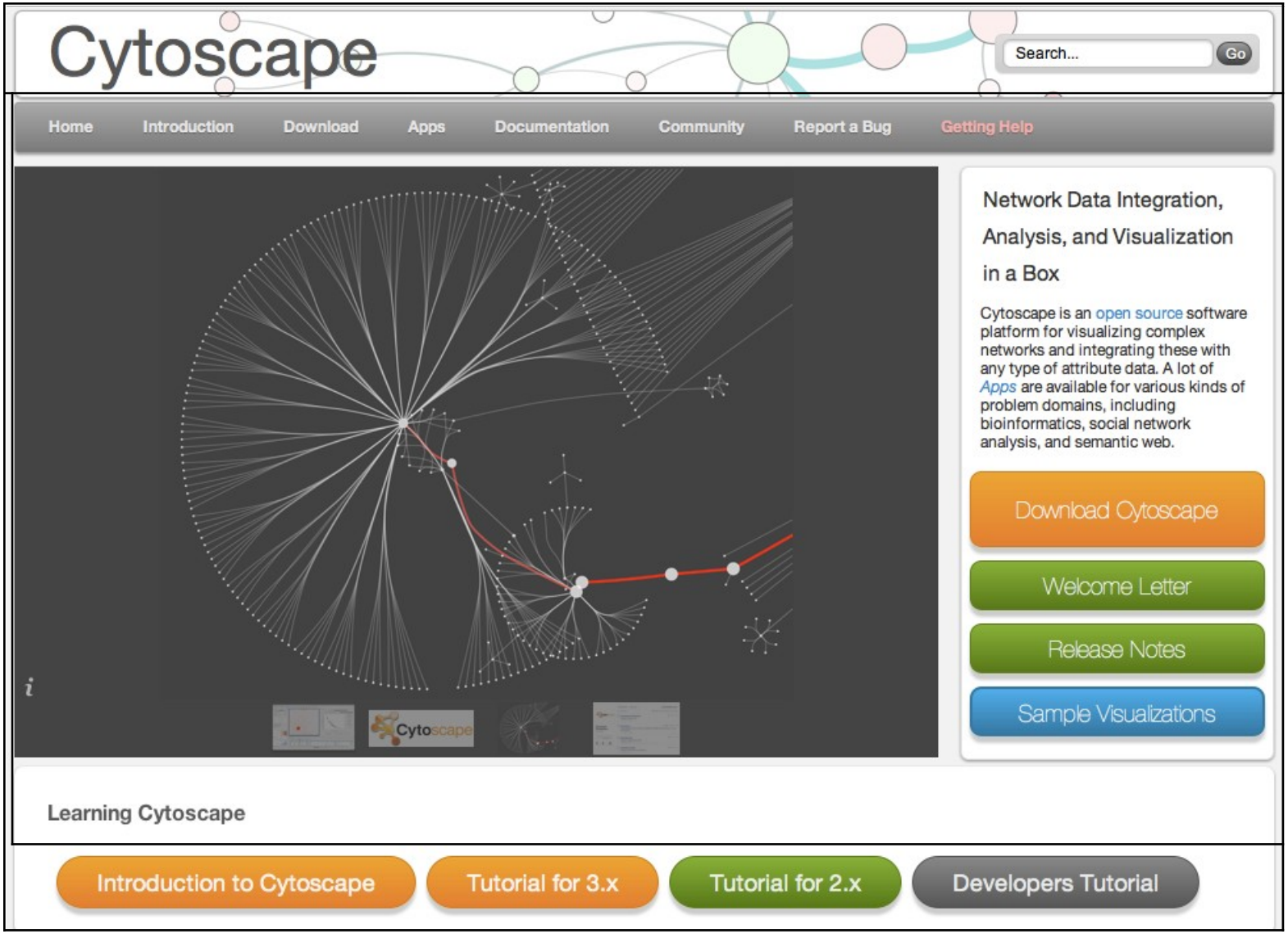
Construction and analysis of gene/protein networks

- From gene expression signatures to gene coexpression networks
- Definition and properties of protein interaction networks
- Visualize and analyse biomolecular networks in Cytoscape
- Using on-line tools to build gene/protein networks: APID, STRING, GeneMANIA, PSICQUIC
- Network medicine: proteins and drugs interactions (STITCH)

Networks tool = Cytoscape

The most powerful tool to build, visualize and analyse networks

Cytoscape: open source bioinformatics tool for biological network visualization & data integration (desktop Java application released under GNU License, LGPL)



Protein-Protein Interactions (PPIs)

build networks from experimental data: Cytoscape

Challenge: improve data integration and analytic methods to understand **networks**

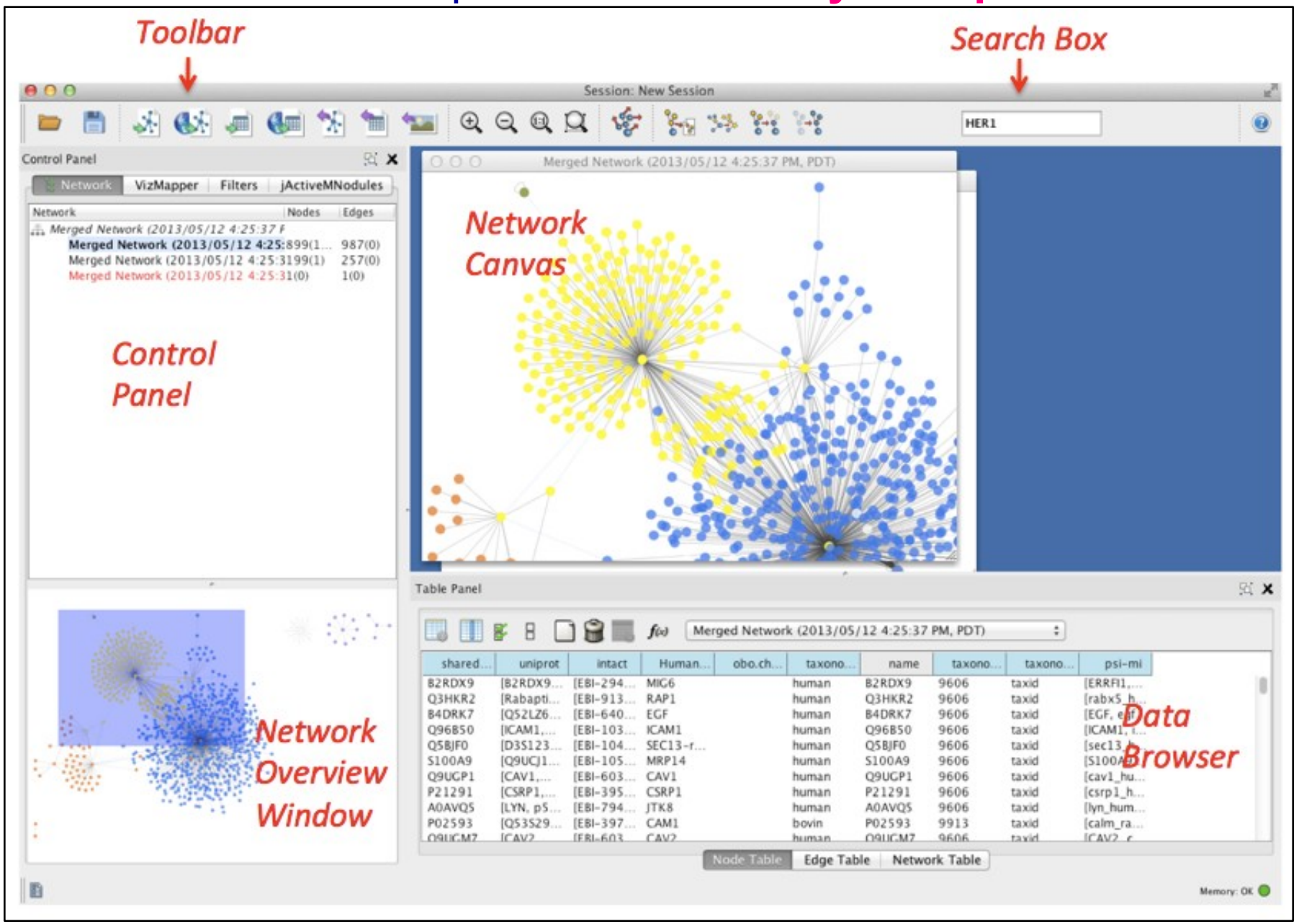
http://www.cytoscape.org/

What is Cytoscape?

Cytoscape is an open source software platform for ***visualizing*** molecular interaction networks and biological pathways and ***integrating*** these networks with annotations, gene expression profiles and other state data. Although Cytoscape was originally designed for biological research, now it is a general platform for complex network analysis and visualization. Cytoscape *core* distribution provides a basic set of features for data integration, analysis, and visualization. Additional features are available as ***Apps*** (formerly called *Plugins*). Apps are available for network and molecular profiling analyses, new layouts, additional file format support, scripting, and connection with databases. They may be developed by anyone using the Cytoscape open API based on Java™ technology and App community development is encouraged. Most of the Apps are freely available from Cytoscape App Store.

Protein-Protein Interactions (PPIs)

build networks from experimental data: Cytoscape



Networks tool = Cytoscape

The most powerful tool to build, visualize and analyse networks

Cytoscape is a open source bioinformatics package for
biological network visualization and data integration
(desktop Java application released under GNU License, LGPL)

Main page

http://www.cytoscape.org/

Web

http://cytoscapeweb.cytoscape.org/

Wiki

http://wiki.cytoscape.org/

http://opentutorials.cgl.ucsf.edu/index.php/Tutorial:Introduction_to_Cytoscape
http://opentutorials.cgl.ucsf.edu/index.php/Tutorial:Introduction_to_Cytoscape-part2

http://wiki.cytoscape.org/Cytoscape_3/UserManual/Network_Formats

Networks tool = Cytoscape

The most powerful tool to build, visualize and analyse networks

Cytoscape is a open source bioinformatics package for
biological network visualization and data integration
(desktop Java application released under GNU License, LGPL)

BIOINFORMATICS **APPLICATIONS NOTE**

Vol. 27 no. 3 2011, pages 431–432
doi:10.1093/bioinformatics/btq675

Systems biology

Advance Access publication December 12, 2010

Cytoscape 2.8: new features for data integration and network visualization

Michael E. Smoot[1,2,*], Keiichiro Ono[1,2], Johannes Ruscheinski and Trey Ideker[1,2]

[1]Department of Medicine and [2]Department of Bioengineering, University of Califor
La Jolla, CA 92093, USA

Associate Editor: Joaquin Dopazo

Important publications:
Nature Protocols (2007)
Bioinformatics (2011)

Protocol

Nature Protocols **2**, 2366 - 2382 (2007)
Published online: 27 September 2007 | doi:10.1038/nprot.2007.324

Subject Category: Computational and theoretical biology

Integration of biological networks and gene expression data using Cytoscape

Melissa S Cline[1,2], Michael Smoot[3], Ethan Cerami[4], Allan Kuchinsky[5], Nerius Landys[3], Chris Workman[6], Rowan Christmas[7], Iliana Avila-Campilo[7,8], Michael Creech[9], Benjamin Gross[4], Kristina Hanspers[10], Ruth Isserlin[11,12], Ryan Kelley[3], Sarah Killcoyne[7], Samad Lotia[3], Steven Maere[13,14], John Morris[15], Keiichiro Ono[3], Vuk Pavlovic[11,12], Alexander R Pico[10], Aditya Vailaya[5,16], Peng-Liang Wang[3], Annette Adler[5], Bruce R Conklin[10], Leroy Hood[7], Martin Kuiper[13,14], Chris Sander[4], Ilya Schmulevich[7], Benno Schwikowski[1], Guy J Warner[17], Trey Ideker[3] & Gary D Bader[11,12]

Cytoscape is a free software package for visualizing, modeling and analyzing molecular and genetic interaction networks. This protocol explains how to use Cytoscape to analyze the results of mRNA expression profiling, and other functional genomics and proteomics experiments, in the context of an interaction network obtained for genes of interest. Five major steps are described: (i) obtaining a gene or protein network, (ii) displaying the network using layout algorithms, (iii) integrating with gene expression and other functional attributes, (iv) identifying putative complexes and functional modules and (v) identifying enriched Gene Ontology annotations in the network. These steps provide a broad sample of the types of analyses performed by Cytoscape.

Networks tool = Cytoscape

The most powerful tool to build, visualize and analyse networks

Cytoscape 2.8: new features for data integration and network visualization

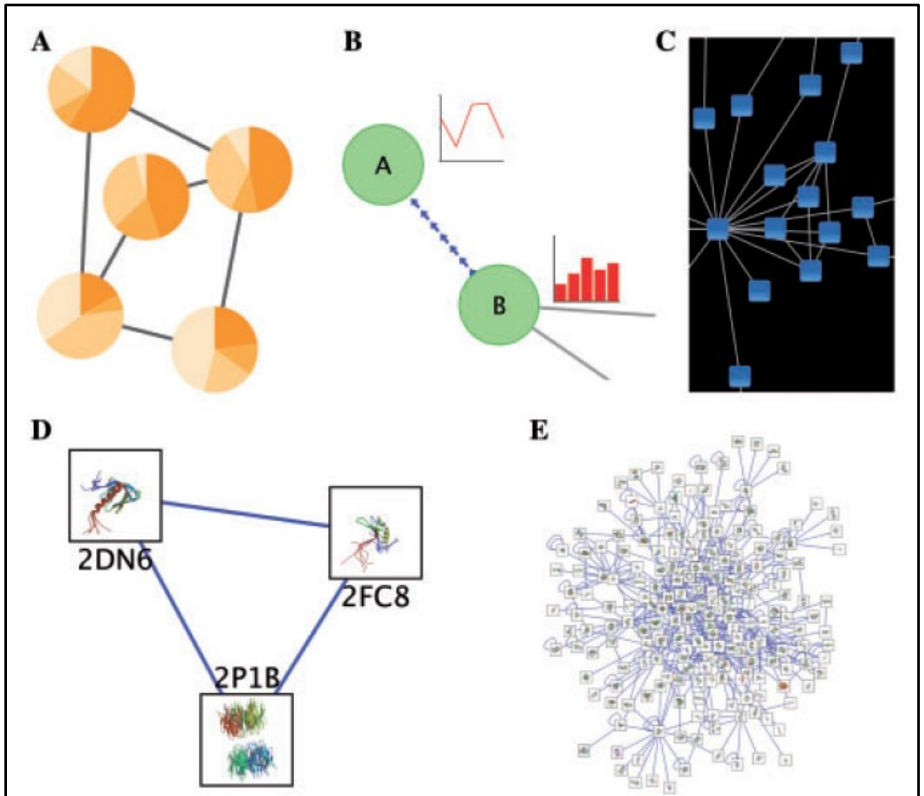


Fig. 1. Rich network visualizations enabled by the new Cytoscape features. Simple networks are shown with custom node images based on (**A**) pie chart displays or (**B**) line plots and bar charts generated using Google's Chart API. (**C**) Nodes have a transparent custom graphic to give the appearance of shading. (**D** and **E**) Protein–protein interaction networks in which each node contains a 3D image of the protein structure of the protein represented by the node.

Protein-Protein Interactions (PPIs)

build networks from experimental data: Cytoscape

http://www.cytoscape.org/

Comparison of network analyses platforms

TABLE 1 | Comparison of network analysis platforms.

Feature	CY	GM	VA	OS	CD	AR	IN	GG	PI	PR	BL	PA
Free for academic use	√	√	√	√	√				√	√	√	√
Free for commercial use	√	√	√		√				√	√	√	
Open source	√	√							√	√	√	
Curated pathway/network content		√		√		√	√	√				
Standard file format support	√		√		√				√	√		√
User-defined networks/pathways	√	√	√	√	√	√	√	√	√	√	√	√
Functionality to infer new pathways	√		√			√		√	√			
GO/pathway enrichment analysis	√	√	√				√	√				
Automated graph layout	√		√	√	√	√	√	√		√	√	√
Complex criteria for visual properties		√				√	√	√		√	√	√
Multiple visual styles	√		√	√		√	√			√		
Advanced node selection	√		√	√		√	√	√	√	√	√	√
Customizable gene/protein database		√	√			√		√	√			
Rich graphical annotation		√	√				√	√				√
Statistical network analysis	√		√				√	√	√		√	
Extensible functionality: plugins or API	√		√		√	√	√	√	√			
Quantitative pathway simulation					√	√						

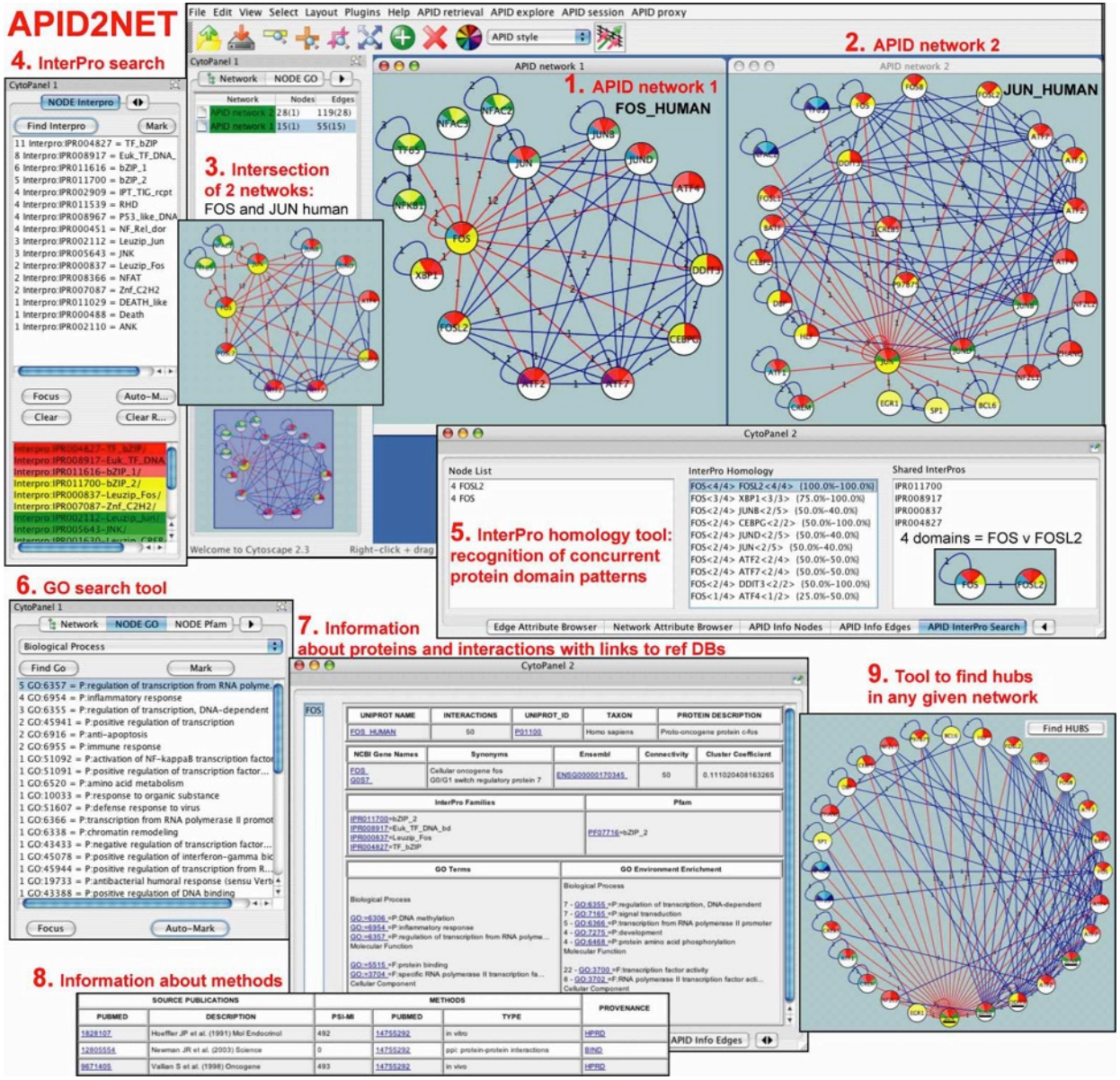
CY, Cytoscape[31]; GM, GenMAPP[26]; VA, VisANT[24]; OS, Osprey[23] (http://biodata.mshri.on.ca/osprey/); CD, CellDesigner[25]; AR, Ariadne Genomics Pathway Studio; IN, Ingenuity Pathways Analysis; GG, GeneGo; PI, PIANA (http://sbi.imim.es/piana/); PR, ProViz (http://cbi.labri.fr/eng/proviz.htm); BL, BioLayout; PA, PATIKA.

From ***Cline et al. (2007) Nature Protocols***

Protein-Protein Interactions (PPIs)

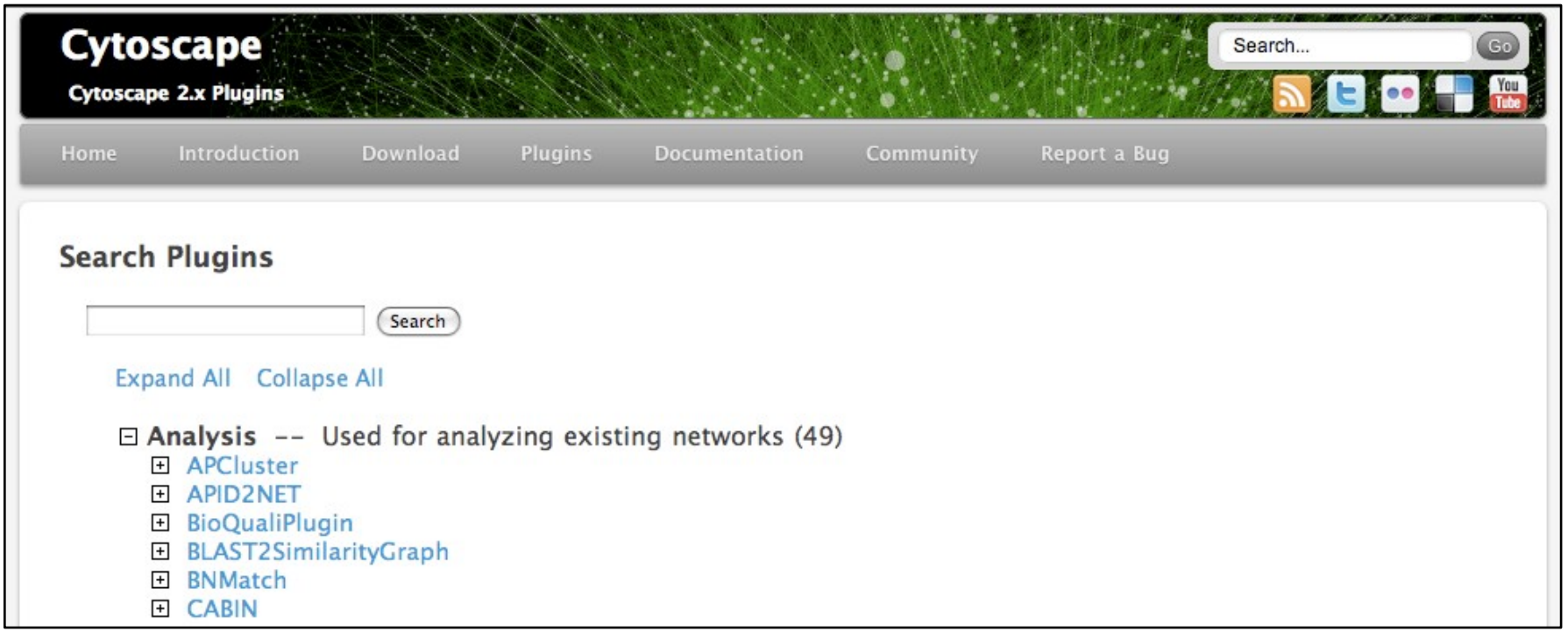
build networks from experimental data: APID2NET

Using *Cytoscape* and the plugin **APID2NET** you can build a **PPI network** by direct query and retrieval from **APID**



Protein-Protein Interactions (PPIs)

build networks from experimental data: APID2NET



APID2NET	1.52	5929
	1.51	1667
	1.5	2030

APID2NET

9626 downloads

(October 2015)

Protein-Protein Interactions (PPIs)

two publications

W298–W302 *Nucleic Acids Research, 2006, Vol. 34, Web Server issue*
doi:10.1093/nar/gkl128

APID: Agile Protein Interaction DataAnalyzer

Carlos Prieto and Javier De Las Rivas*

Bioinformatics and Functional Genomics Research Group, Cancer Research Center (CIC, CSIC/USAL), 37007 Salamanca, Spain

BIOINFORMATICS **APPLICATIONS NOTE** *Vol. 23 no. 18 2007, pages 2495–2497*
doi:10.1093/bioinformatics/btm373

Systems biology

APID2NET: unified interactome graphic analyzer

Juan Hernandez-Toro, Carlos Prieto and Javier De Las Rivas*

Bioinformatics and Functional Genomics Research Group, Cancer Research Center (IBMCC-CIC, CSIC-USAL), Salamanca, Spain

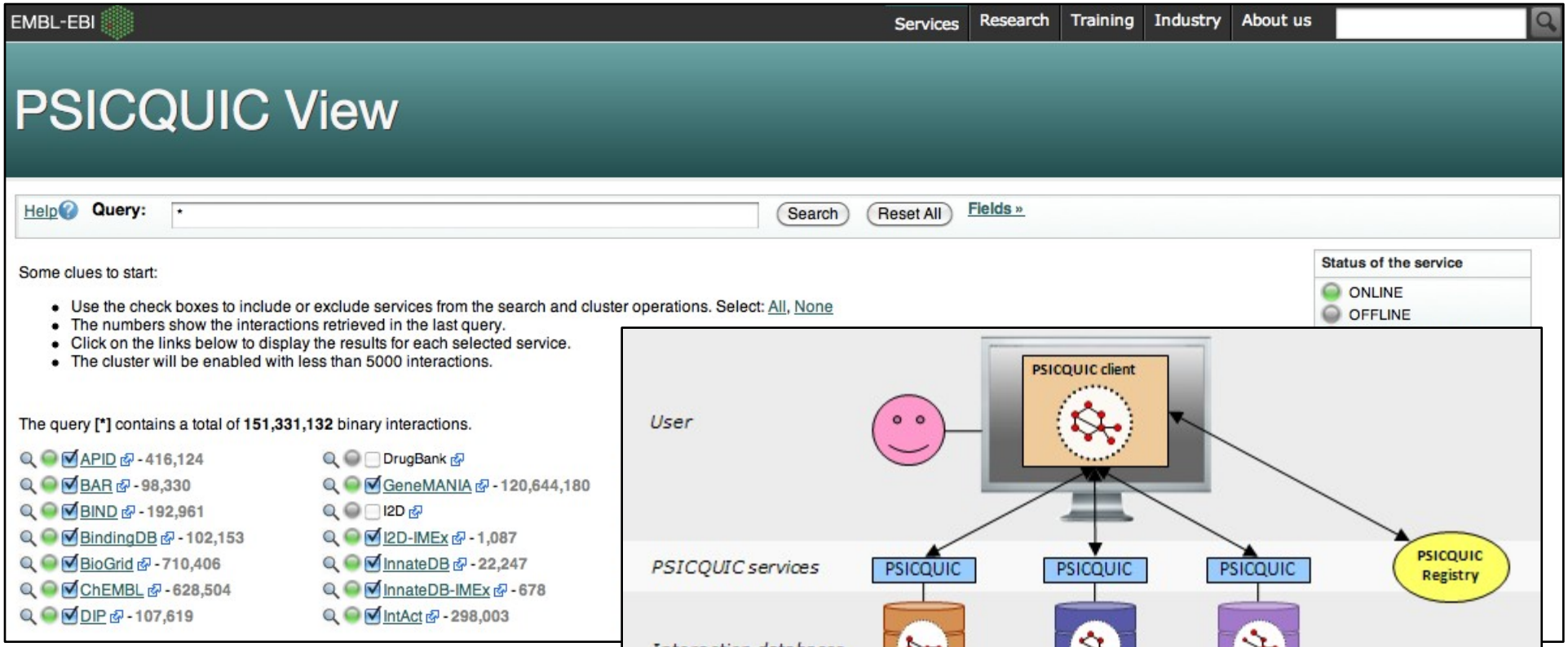
Received on April 13, 2007; revised on June 7, 2007; accepted on July 11, 2007

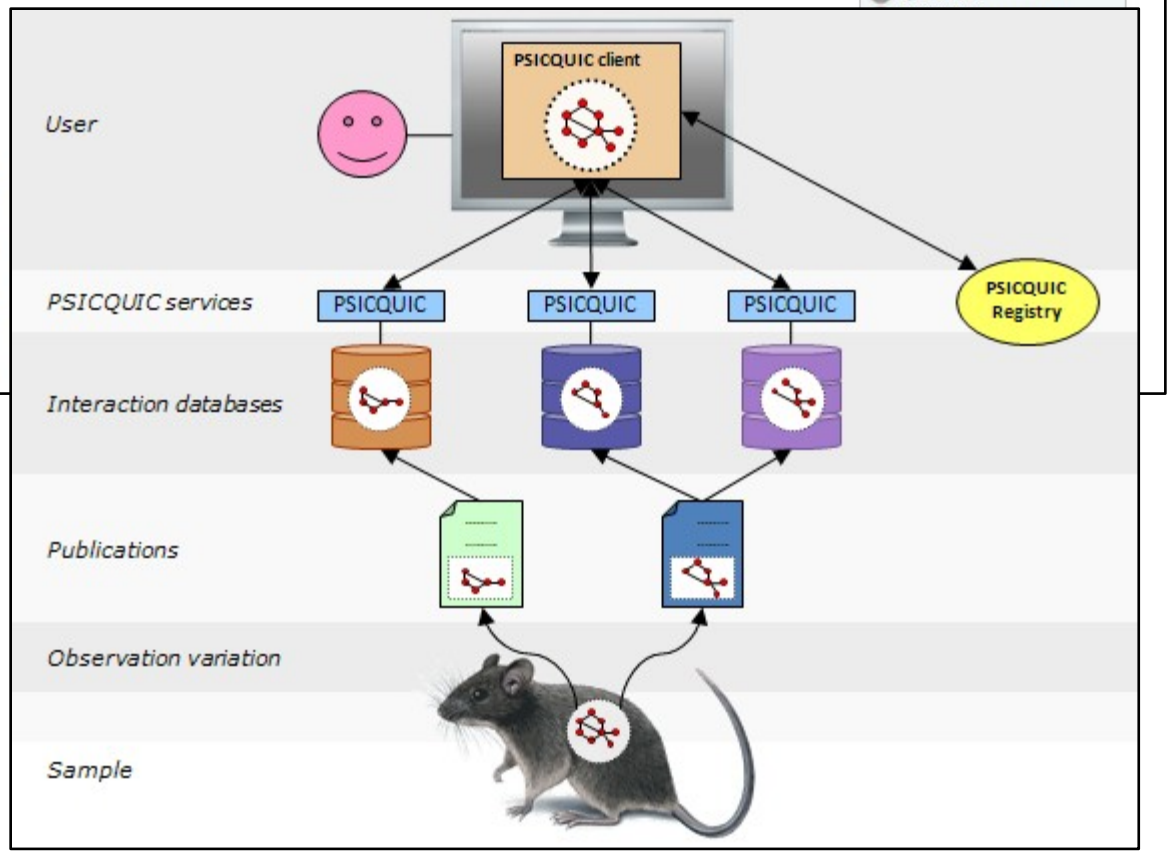
Advance Access publication July 21, 2007

Associate Editor: Trey Ideker

Protein-Protein Interactions (PPIs)

PSICQUIC new tool and service

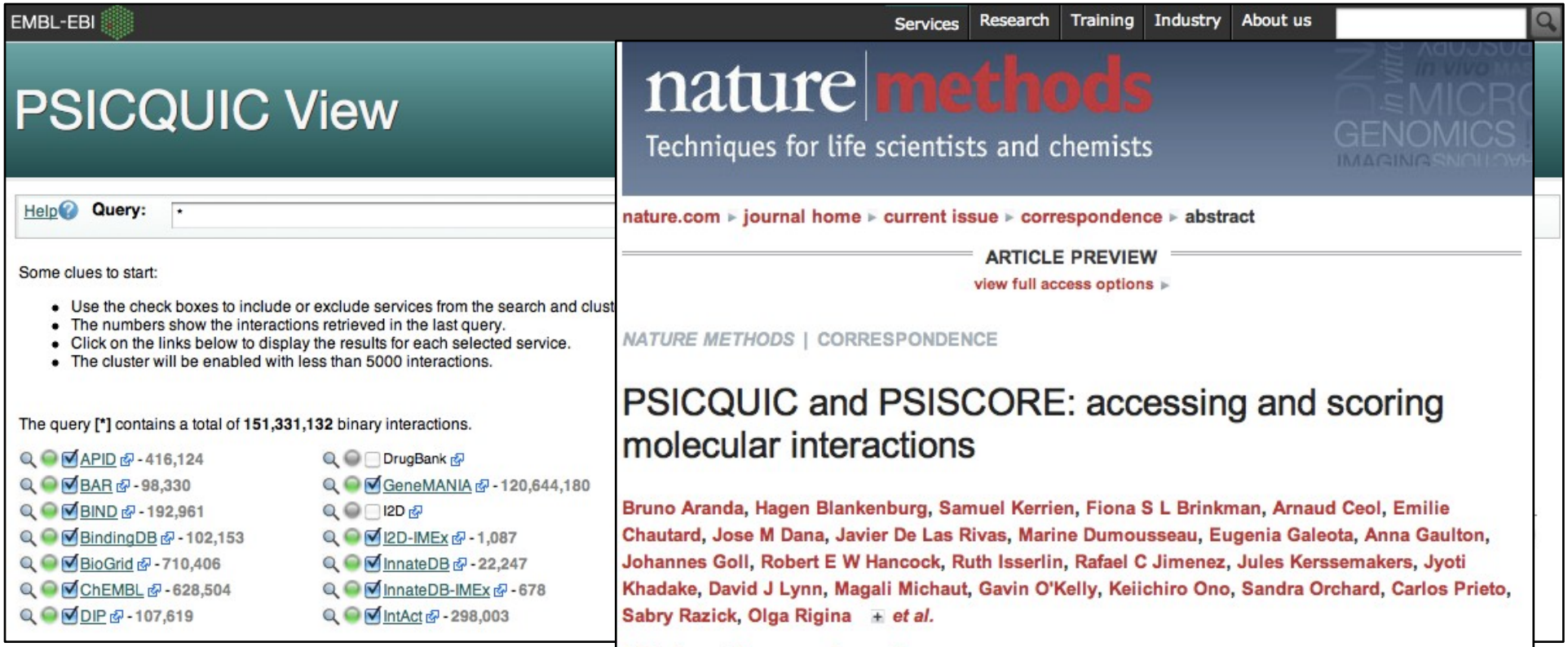




Aranda et al. (2011) Nature Methods

Protein-Protein Interactions (PPIs)

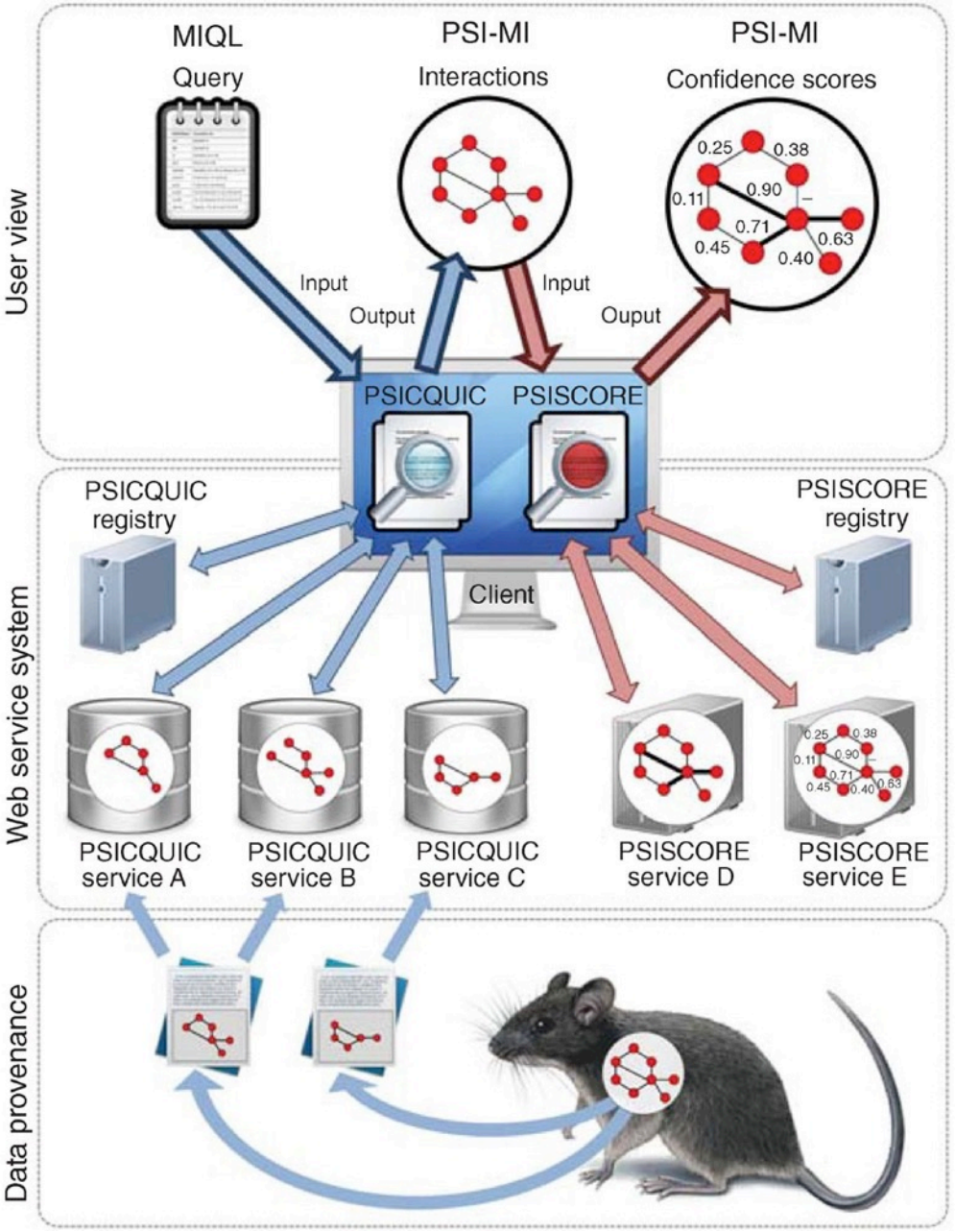
PSICQUIC new tool and service



Aranda et al. (2011) Nature Methods

Protein-Protein Interactions (PPIs)

PSICQUIC & PSISCORE



Aranda et al. (2011) Nature Methods

Protein-Protein Interactions (PPIs)

Javier De Las Rivas

References

- Aranda *et al.* (2011) **PSICQUIC and PSISCORE: accessing and scoring molecular interactions**. *Nature Methods* 8, 528–529.
- De Las Rivas J, Fontanillo C. (2010) **Protein-Protein Interactions essentials: key concepts to building and analyzing Interactome Networks**. *PLoS Computational Biology* 6(6): e1000807.
- Prieto C, De Las Rivas J. (2010) **Structural domain-domain interactions: assessment and comparison with protein-protein interaction data to improve the interactome**. *Proteins* 78:109-117.
- Hernandez-Toro J., Prieto C, De Las Rivas J. (2007). **APID2NET: unified interactome graphic analyzer**. *Bioinformatics* 23: 2495-2497.
- Prieto C, De Las Rivas J. (2006). **APID: Agile Protein Interaction DataAnalyzer**. *Nucleic Acids Research* 34: W298-302.

WEB References

http://bioinfow.dep.usal.es/apid

http://ubioinfo.cicancer.org/en/index-en.html

Networks & Pathways

Comparison and combination of these type of complex data

Networks & Pathways

¿The data?: databases, data sources

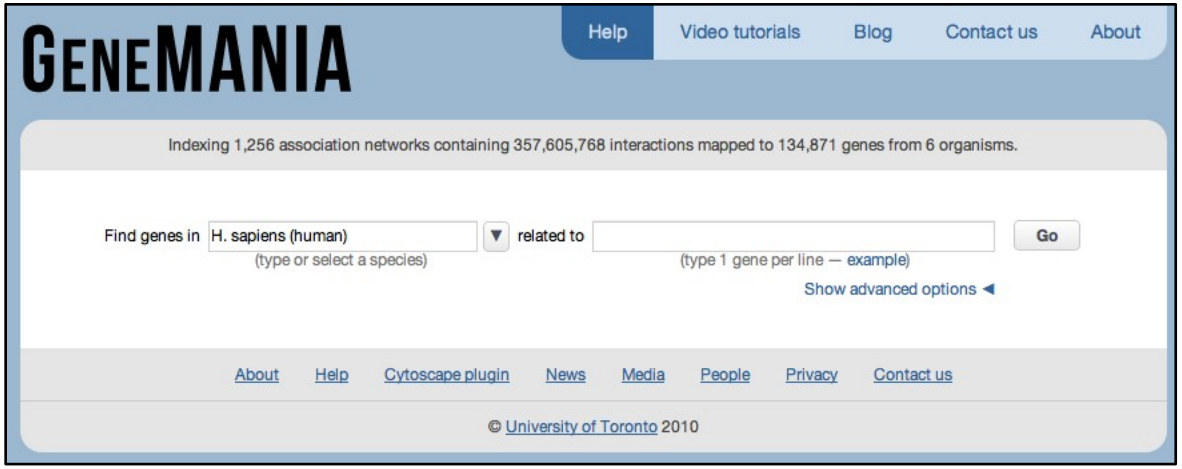
genes/proteins in networks

and

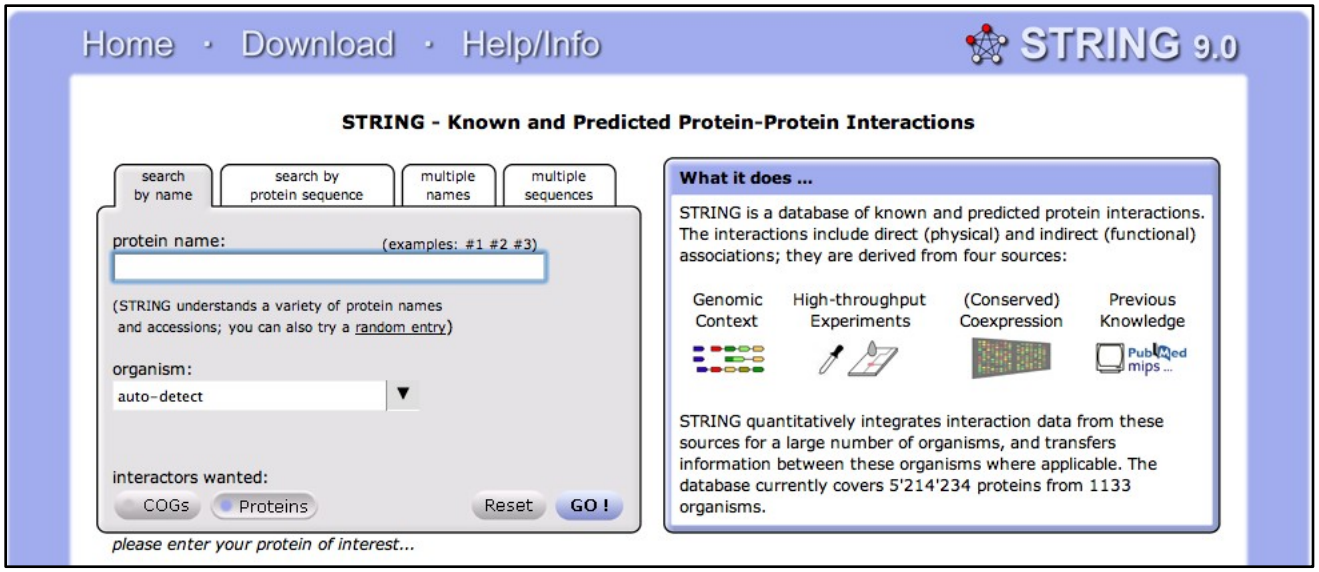
genes/proteins in pathways

Network databases
GeneMANIA and STRING

http://www.genemania.org/



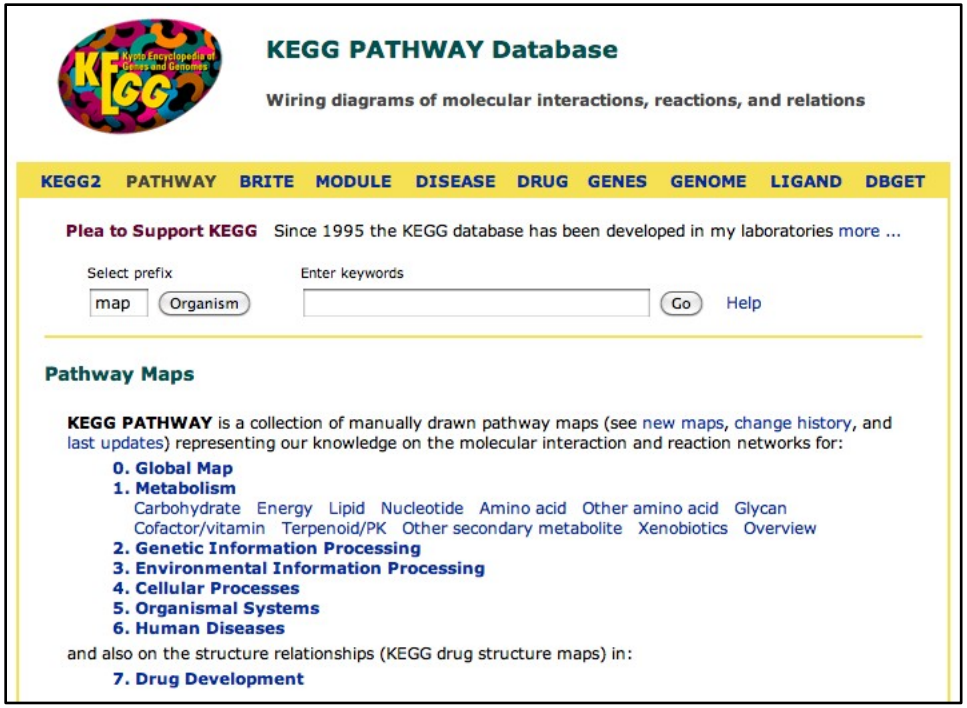
http://string-db.org/



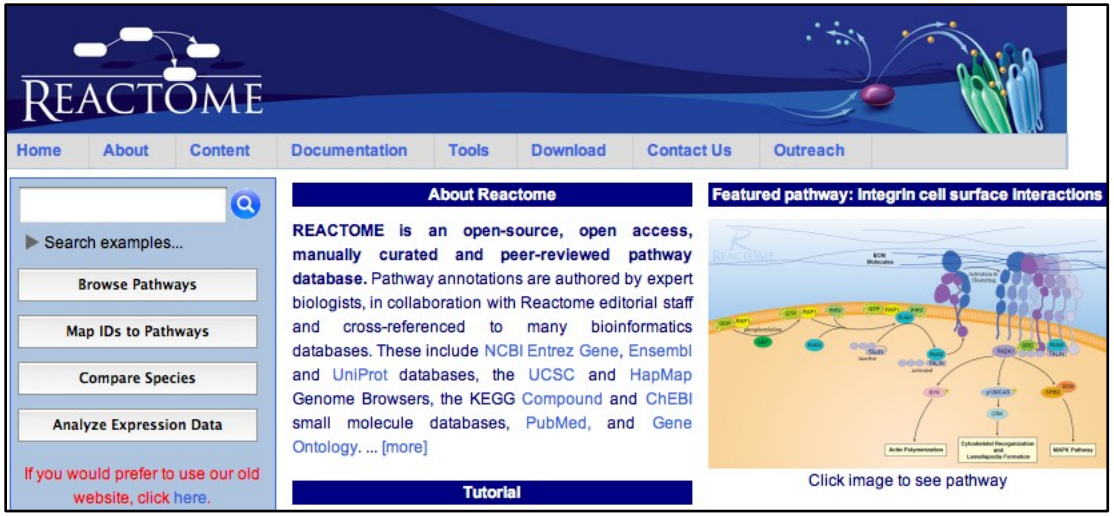
Pathways databases

KEGG and Reactome

http://www.genome.jp/kegg/



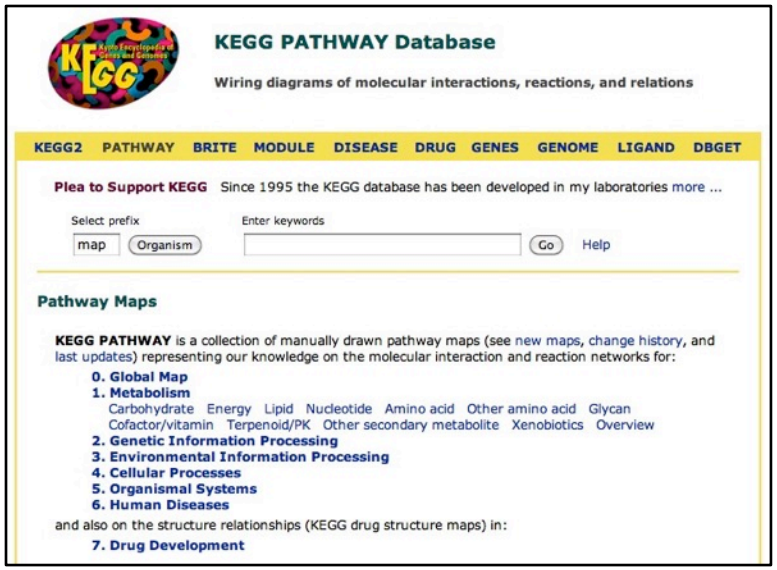
http://www.reactome.org/



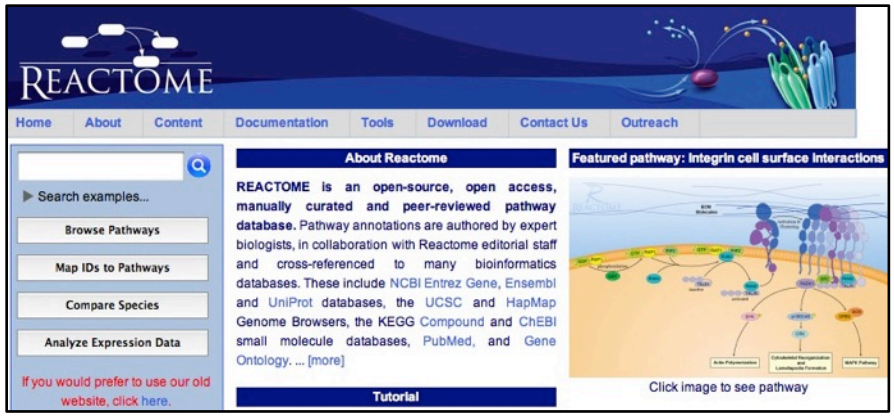
Networks & Pathways

Comparison and combination of these type of complex data

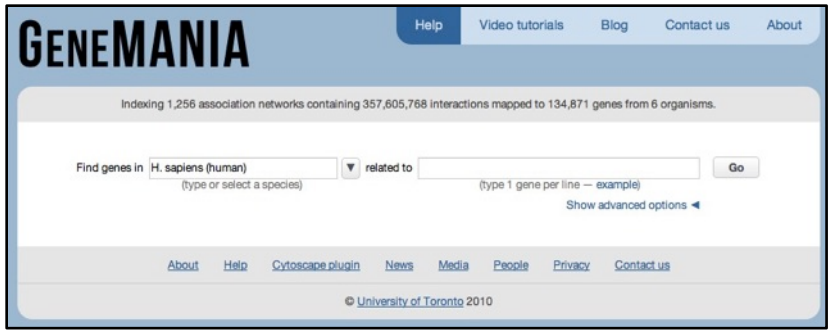
http://www.genome.jp/kegg/



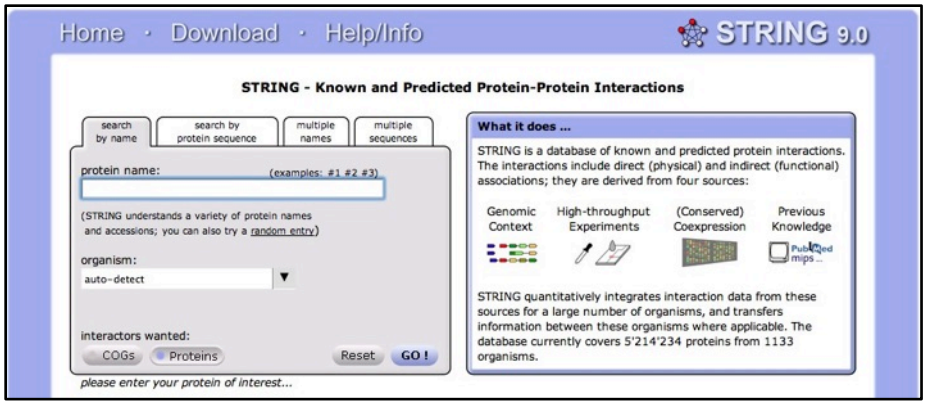
http://www.reactome.org/



http://www.genemania.org/



http://string-db.org/



CBO Computational Biology @ UCT

CiC

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Hands-on: Practical Examples

Explore web resources & tools: GeneMANIA, STRING

Protein_SETs_2014.xls
(175g human, 59g5pc yeast)

CBO Computational Biology @ UCT

CiC

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Hands-on: Practical Examples

Start using: Cytoscape

Cytoscape sampleData (yeastHighQuality.sif file)

Session 3 (9:30 - 12:30, 3h)

Protein interaction networks

Session 4 (13:30 - 16:30, 3h)

Construction and analysis of gene/protein networks

- From gene expression signatures to gene coexpression networks
- Definition and properties of protein interaction networks
- Visualize and analyse biomolecular networks in Cytoscape
- Using on-line tools to build gene/protein networks: APID, STRING, GeneMANIA, PSICQUIC
- Network medicine: proteins and drugs interactions (STITCH)

Networks

Two major types of networks derived from experimental data

Two major types of networks derived from large-scale *omic* data

1.– **Gene Coexpression Networks**: ***ggcoe***
derived from gene expression profiling and transcriptomic studies

2.– **Protein-Protein Interaction Networks**: ***ppi***
derived from proteomic studies

Human coexpression studies

Stuart et al. (2003) Science

A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules

Joshua M. Stuart,[1*†] Eran Segal,[2*] Daphne Koller,[2‡] Stuart K. Kim[3‡]

Lee et al. (2004) Genome Research

Coexpression Analysis of Human Genes Across Many Microarray Data Sets

Homin K. Lee,[1] Amy K. Hsu,[1,2] Jon Sajdak,[1] Jie Qin,[1] and Paul Pavlidis[1,3,4]

[1]Columbia Genome Center, [2]College of Physicians and Surgeons, and [3]Department of Biomedical Informatics, Columbia University, New York, New York 10032, USA

Griffith et al. (2005) Genomics

Assessment and integration of publicly available SAGE, cDNA microarray, and oligonucleotide microarray expression data for global coexpression analyses

Obi L. Griffith[a], Erin D. Pleasance[a], Debra L. Fulton[b], Mehrdad Oveisi[a], Martin Ester[c], Asim S. Siddiqui[a], Steven J.M. Jones[a,*]

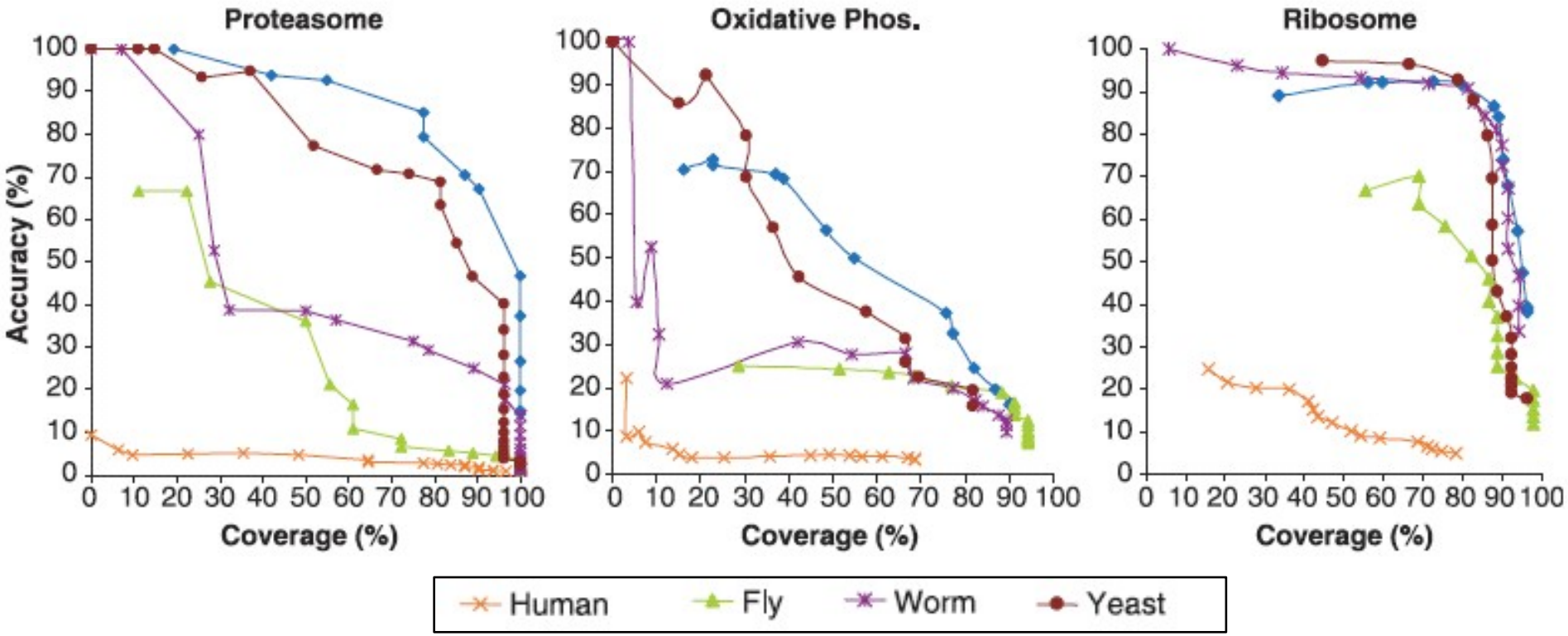
Human coexpression

low signal & high noise

Stuart et al. (2003) Science

A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules

Joshua M. Stuart,[1*†] Eran Segal,[2*] Daphne Koller,[2‡] Stuart K. Kim[3‡]



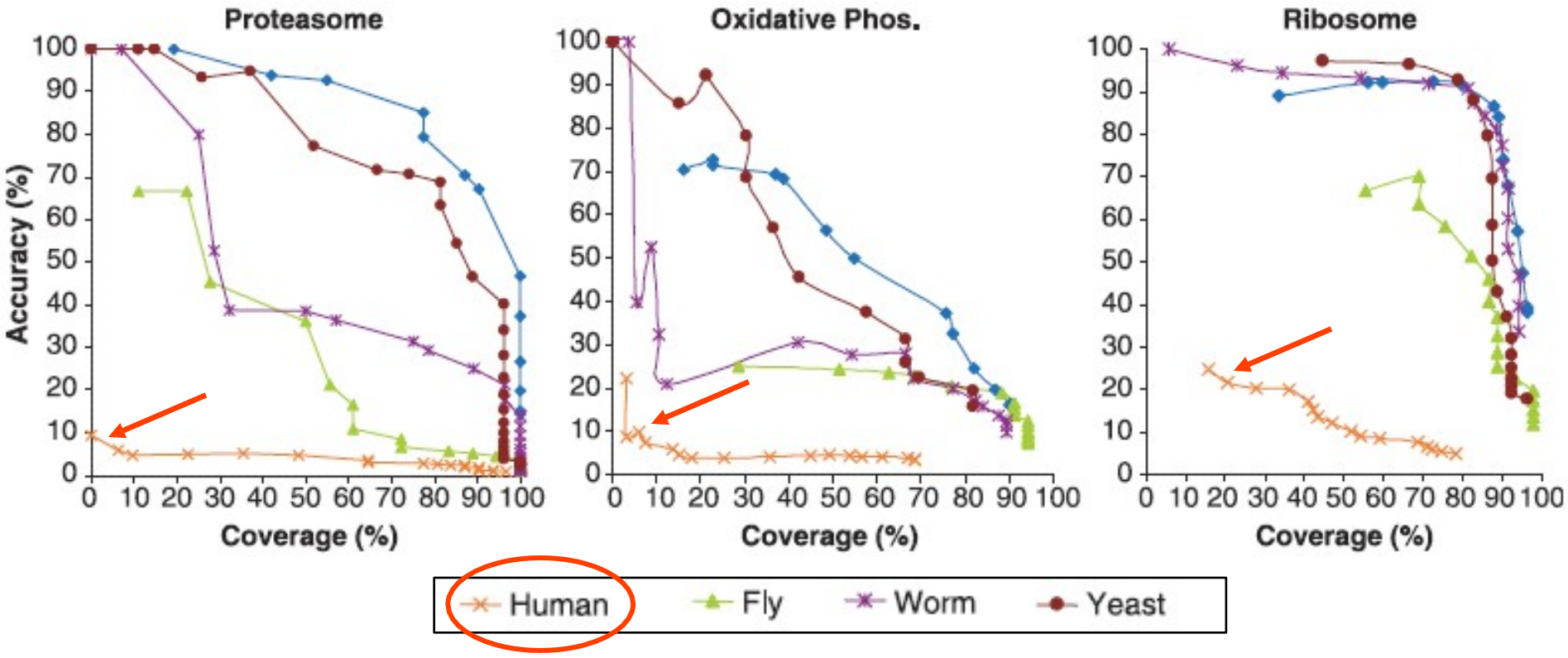
Human coexpression

low signal & high noise

Stuart et al. (2003) Science

A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules

Joshua M. Stuart,[1*†] Eran Segal,[2*] Daphne Koller,[2‡] Stuart K. Kim[3‡]



Human coexpression studies

Stuart et al. (2003) Science

A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules

Joshua M. Stuart,[1*†] Eran Segal,[2*] Daphne Koller,[2‡] Stuart K. Kim[3‡]

Lee et al. (2004) Genome Research

Coexpression Analysis of Human Genes Across Many Microarray Data Sets

Homin K. Lee,[1] Amy K. Hsu,[1,2] Jon Sajdak,[1] Jie Qin,[1] and Paul Pavlidis[1,3,4]

[1]Columbia Genome Center, [2]College of Physicians and Surgeons, and [3]Department of Biomedical Informatics, Columbia University, New York, New York 10032, USA

Griffith et al. (2005) Genomics

Assessment and integration of publicly available SAGE, cDNA microarray, and oligonucleotide microarray expression data for global coexpression analyses

Obi L. Griffith[a], Erin D. Pleasance[a], Debra L. Fulton[b], Mehrdad Oveisi[a], Martin Ester[c], Asim S. Siddiqui[a], Steven J.M. Jones[a,*]

Experimental dataset selection

Sample bias

Lee et al. (2004) Genome Research

Coexpression Analysis of Human Genes Across Many Microarray Data Sets

Homin K. Lee,[1] Amy K. Hsu,[1,2] Jon Sajdak,[1] Jie Qin,[1] and Paul Pavlidis[1,3,4]

[1]Columbia Genome Center, [2]College of Physicians and Surgeons, and [3]Department of Biomedical Informatics, Columbia University, New York, New York 10032, USA

Sample bias

"malignant data"

Lee et al. (2004) Genome Research

Table 1. Summary of the Microarray Data Sets Used[a]

Reference[b]	Samples[c]	Genes[d]	Raw links[e]	Reference	Samples	Genes	Raw links
(Alizadeh et al. 2000)	96	1759	25748	(Nielsen et al. 2002)	46	3359	25175
(Allander et al. 2001)	19	1205	1251	(Perou et al. 1999)	26	3027	42105
(Armstrong et al. 2002)	72	8242	213456	(Perou et al. 2000)	84	5701	167826
(Bhattacharjee et al. 2001)	203	8243	243303	(Pomeroy et al. 2002)	90	5418	85909
(Bittner et al. 2000)	38	4382	16141	(Ramaswamy et al. 2001)	255	9528	372500
(Butte et al. 2000)	68	4906	81755	(Rickman et al. 2001)	51	5418	60169
(Chang et a							129814
(Chaussabe							3409
(Chen et al.							45468
(Cheok et a							87486
(Dhanaseka							19895
(Diehn et a							248952
(Dyrskjot et							1578
(Dyrskjot et							14351
(Erraji-BenC							4760
(Garber et a							128303
(Golub et a							12944
(Greenberg							51286
(Gruvberge							241088
(Hedenfalk							7941
(Hedenfalk							60268
(Huang et a							752390
(Huang et a							5923
(Jazaeri et a							2670
(Khan et al. 2001)	88	1952	19868	(Welsh et al. 2001)	49	5418	52459
(Khatua et al. 2003)	13	8257	10072	(Welsh et al. 2001)	55	8258	260155
(Leung et al. 2002)	126	12657	993195	(West et al. 2001)	49	5418	84842
(Luo et al. 2001)	25	4354	14873	(Whitfield et al. 2002)	114	12801	1547199
(Ma et al. 2003)	61	1569	10086	(Yeoh et al. 2002)	248	8257	257979
(MacDonald et al. 2001)	31	1309	3179	(Yoon et al. 2002)	12	5418	53305

≈ 80% of these datasets correspond to "cancer" samples

¿how "normal" is this?

¿ do we consider that "tumor cells" usually have totally aberrant genome with many altered chromosomes?

Sample bias

"malignant data"

microarray datasets in

Lee et al. (2004) Genome Research

lymphoma
GIST **sarcoma**
leukemia
lung **cancer**
melanoma
leprosy
NCI-60 **tumor** cell lines
fibroblasts
parasite response
liver **cancer**
leukemia
breast **cancer**
prostate **cancer**
T-cells
bladder **tumors**
bladder **tumors**
lung **cancer**
leukemia
inflammatory myopathy
breast **cancer**

breast **cancer**
breast **cancer**
obesity
breast **cancer**
thyroid papillary **tumors**
breast **cancer**
blue cell **tumors**
astrocytoma
gastric **cancer**
prostate **cancer**
breast **cancer**
medulloblastoma
sarcoma
breast **cancer**
breast **cancer**
brain **tumors**
tumor and normal
glioma
leukemia
breast **cancer**

NCI-60 **tumor** cell lines
lymphoma
prefrontal cortex
prostate **cancer**
breast **cancer**
breast **cancer**
asthma
NCI-60 **tumor** cell lines
diverse tissues
dermatomyositis
viral infection
breast **cancer**
leukemia
muscle
ovarian **cancer**
prostate **cancer**
breast **cancer**
cell cycle, **tumors**
leukemia
colorectal **cancer**

Human transcriptomic network of normal tissues: a global map without malignant data

Key questions

•Can we use global human gene expression data (i.e. transcriptomic genome-wide microarray data) to derive **gene coexpression networks** ?

•Is it a **reliable** way to find coexpression (knowing the large noise and background in microarrays and the bad effect of outliyers on correlation) ?

• How **reliable** are the data coming from **microarrays** ?
Can we calculate and improve the reliability of microarray data ?

• Which **algorithm** is good enough to provide a sensible reliable **expression signal**: MAS5, RMA, dCHIP, PLIER, FARMS ...?

Experimental dataset selection

Normal Samples

Prieto et al. (2008) PLoS ONE

OPEN ACCESS Freely available online

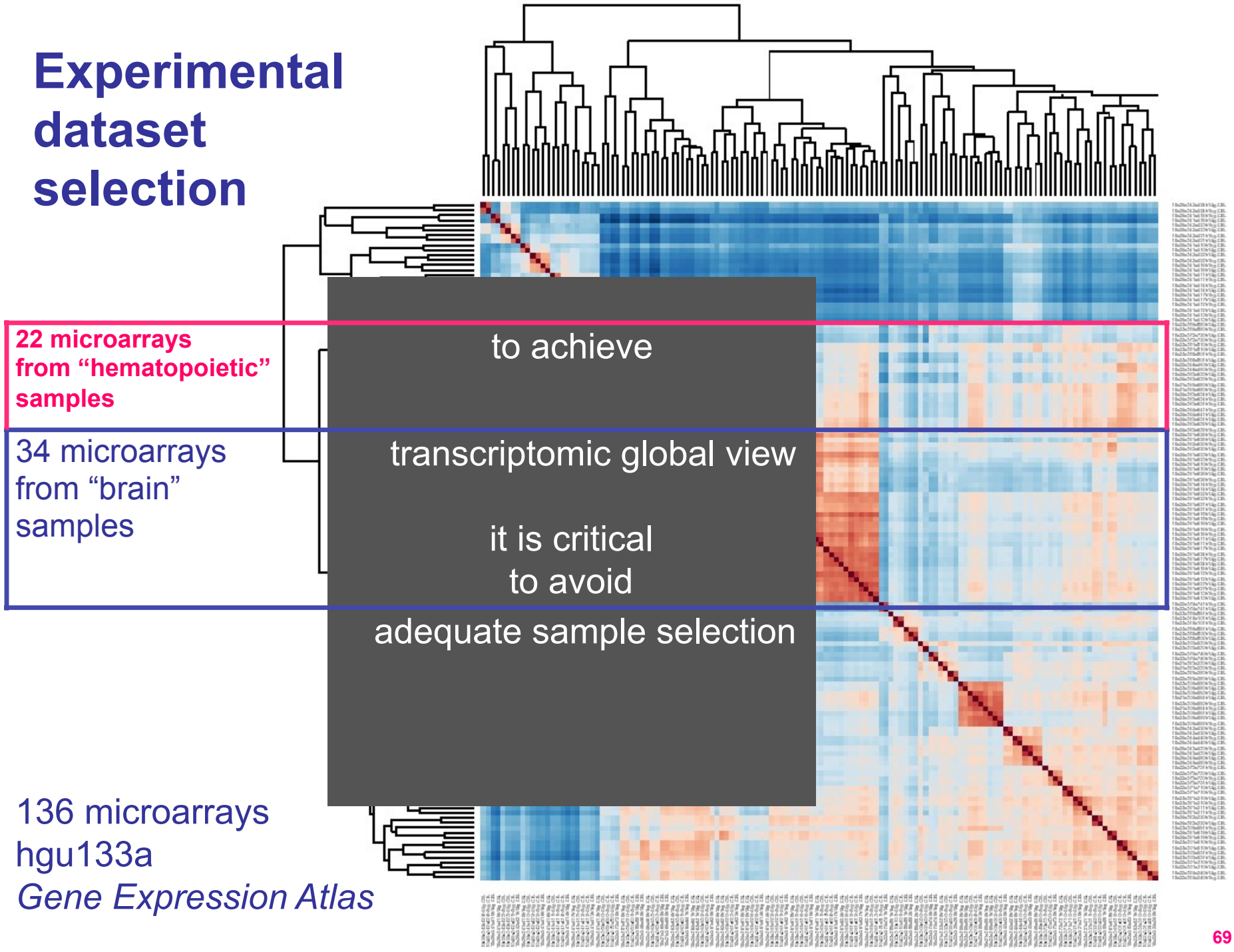
PLoS one

Human Gene Coexpression Landscape: Confident Network Derived from Tissue Transcriptomic Profiles

Carlos Prieto, Alberto Risueño, Celia Fontanillo, Javier De Las Rivas*

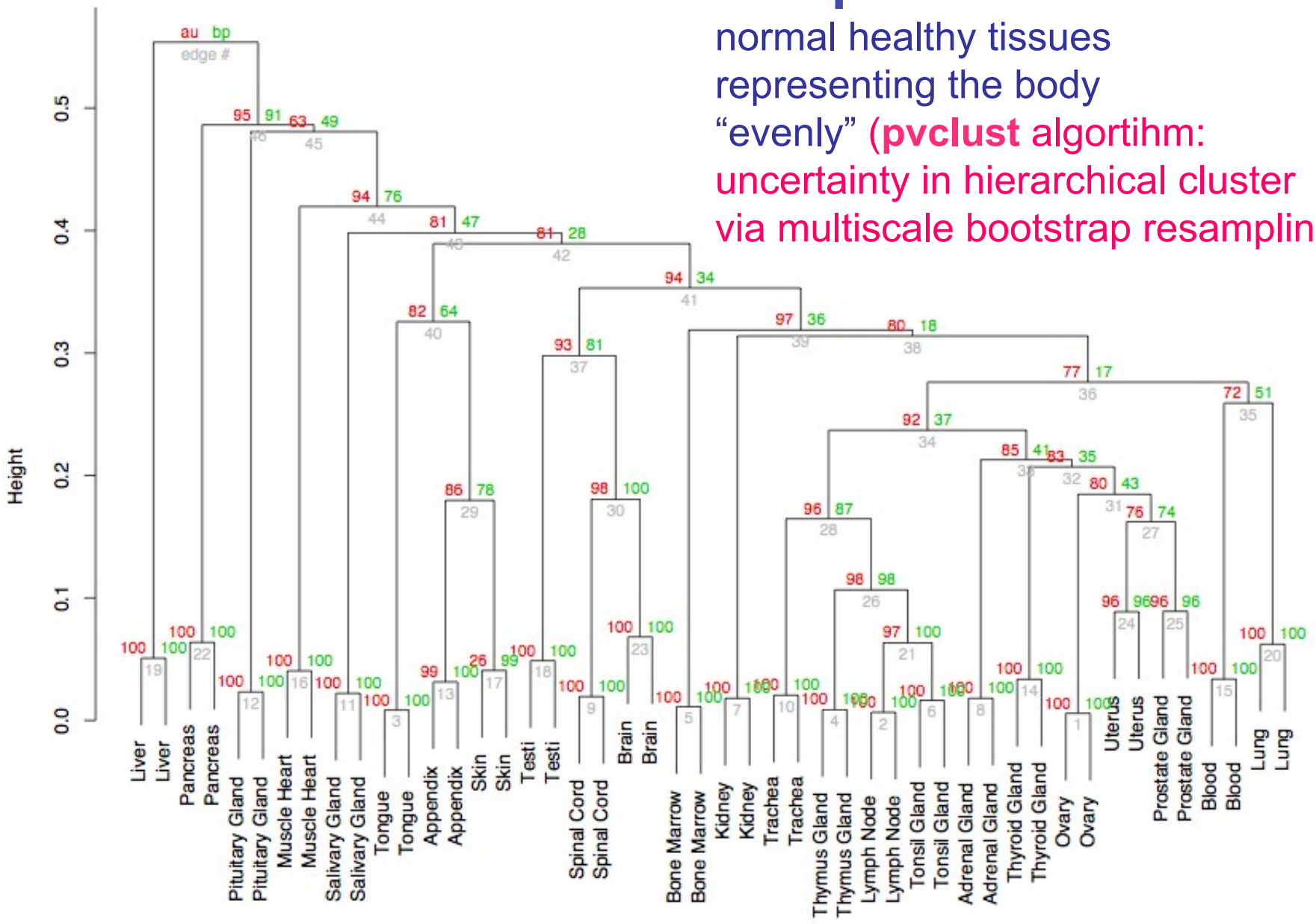
Bioinformatics and Functional Genomics Research Group, Cancer Research Center (CIC-IBMCC, CSIC/USAL), Salamanca, Spain

Experimental dataset selection



Sample selection

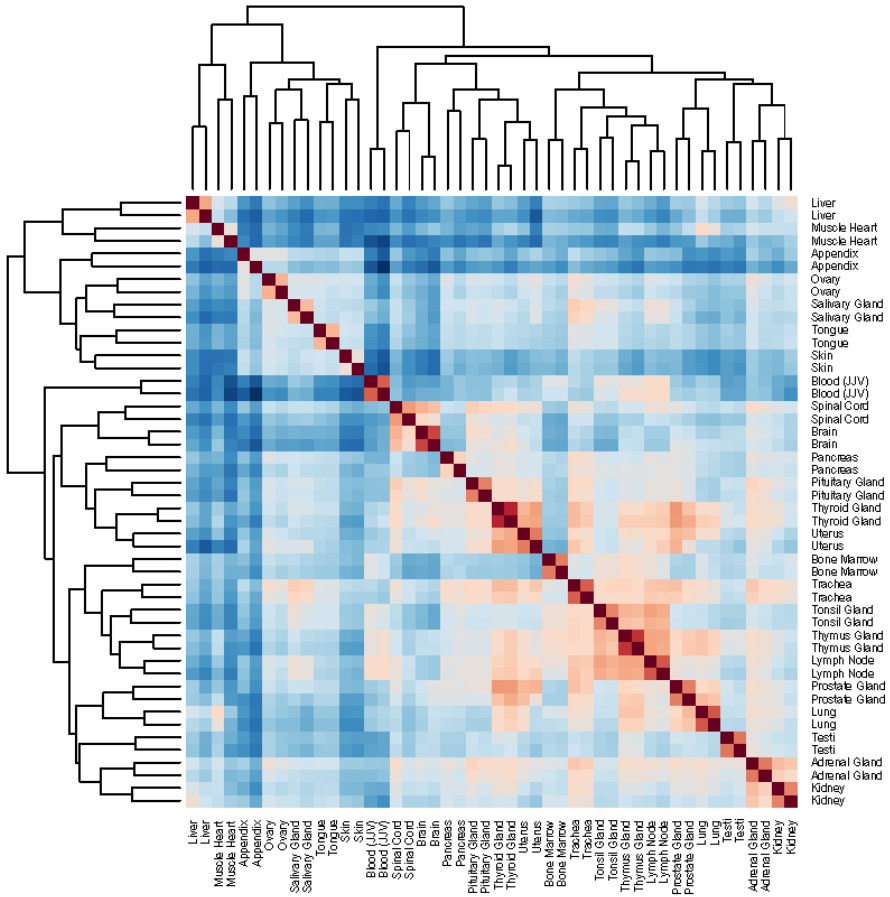
normal healthy tissues representing the body "evenly" (**pvclust** algortihm: uncertainty in hierarchical cluster via multiscale bootstrap resampling)



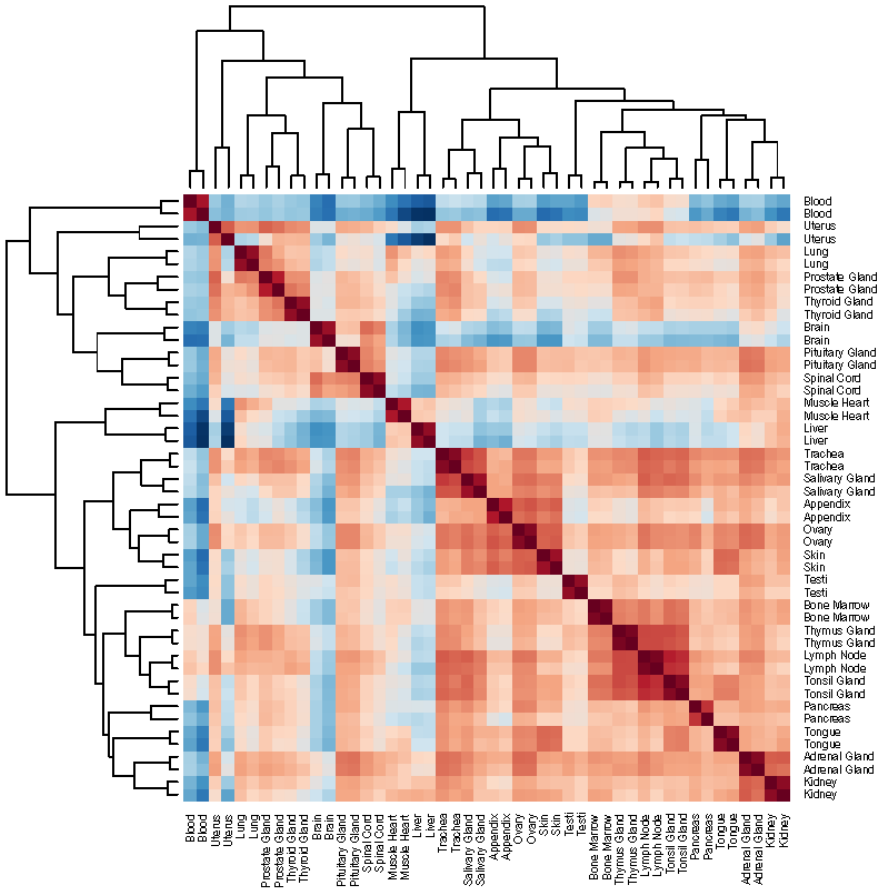
Experimental dataset selection

48 microarrays of whole tissues / organs
normal healthy samples (hgu133a)
Gene Expression Atlas

A **MAS5 - Spearman**



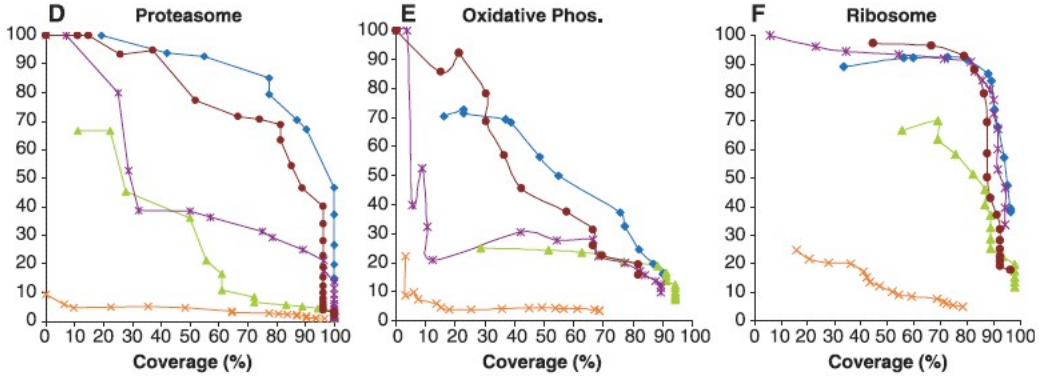
B **RMA - Pearson**



Human coexpression

comparative
study
using *Stuart et al.* approach

Stuart et al. (2003) Science



Lee et al. (2004) Genome Research

Coexpression Analysis of Human Genes Across Many Microarray Data Sets

Homin K. Lee,[1] Amy K. Hsu,[1,2] Jon Sajdak,[1] Jie Qin,[1] and Paul Pavlidis[1,3,4]

[1]Columbia Genome Center, [2]College of Physicians and Surgeons, and [3]Department of Biomedical Informatics, Columbia University, New York, New York 10032, USA

Griffith et al. (2005) Genomics

Assessment and integration of publicly available SAGE, cDNA microarray, and oligonucleotide microarray expression data for global coexpression analyses

Obi L. Griffith[a], Erin D. Pleasance[a], Debra L. Fulton[b], Mehrdad Oveisi[a], Martin Ester[c], Asim S. Siddiqui[a], Steven J.M. Jones[a,*]

Human coexpression studies

mapping coexpressing genes into ***KEGG*** **pathways** to check functional coherence

done as in:

Stuart et al. (2003) Science

i.e. detection of the number of genes within each pathway that coexpress

but still noisy data !!!

This work (2008)

pathway name (KEGG ID number)	nºgenes	genes coexp / genes	% gn coexp	mean r
Proteasome (3050)	31	**28 / 28**	**1.00**	0.69
Ribosome (3010)	120	**52 / 55**	**0.95**	0.75
Oxidative phosphorylation (190)	129	**88 / 95**	**0.93**	0.73
Focal adhesion (4510)	194	**154 / 168**	**0.92**	0.68
Antigen processing and presentation (4612)	86	**71 / 78**	**0.91**	0.75
Glycan structures - degradation (1032)	30	**20 / 22**	**0.91**	0.65
Neuroactive ligand-receptor interact. (4080)	299	**227 / 255**	**0.89**	0.68
Cell cycle (4110)	114	**90 / 102**	**0.88**	0.66
Regulation of actin cytoskeleton (4810)	208	**141 / 161**	**0.88**	0.66
Cytokine-cytokine receptor interact. (4060)	256	**196 / 223**	**0.88**	0.69

Lee et al. (2004)

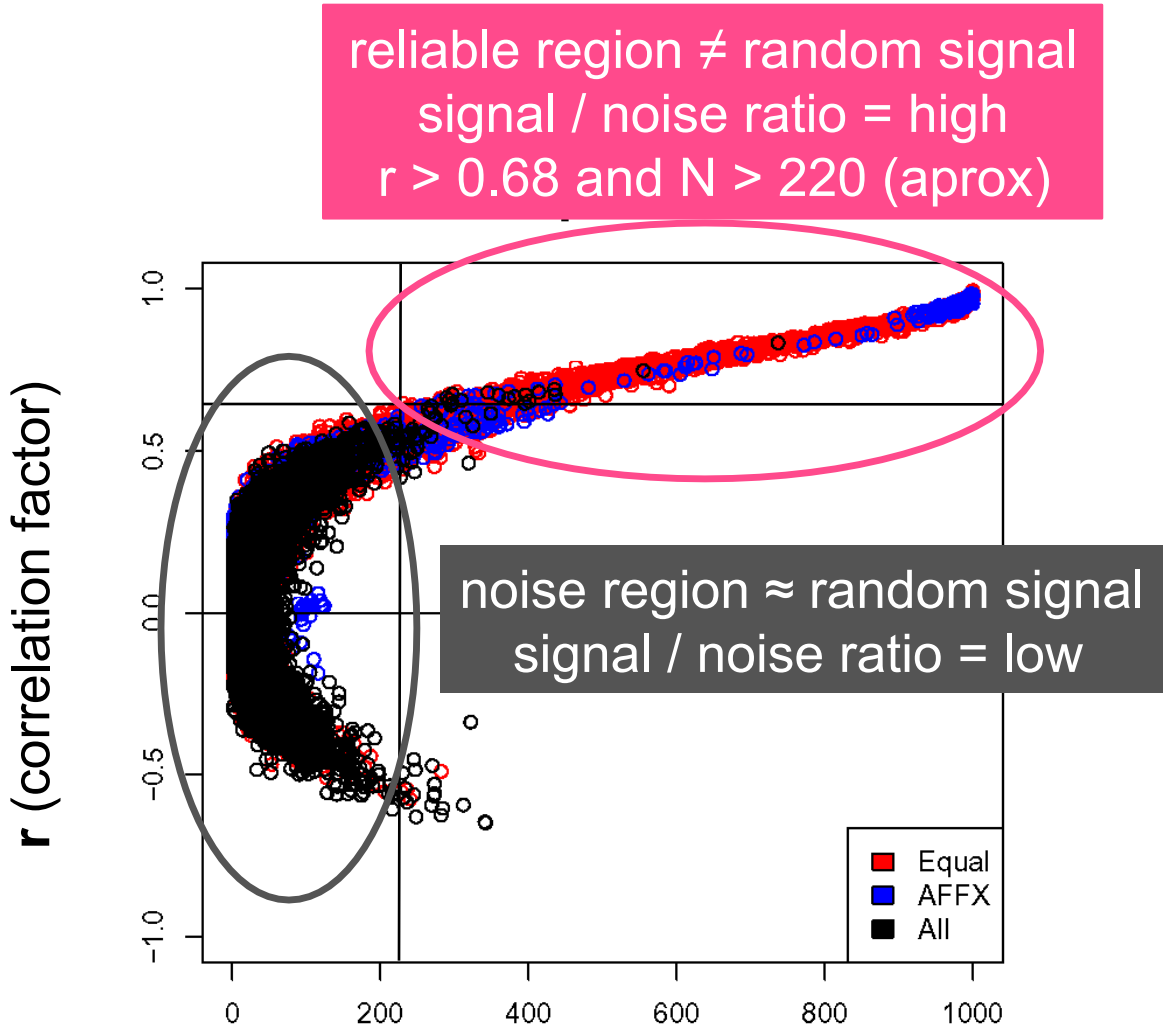
pathway name (KEGG ID number)	nºgenes	genes coexp / genes	% gn coexp
Ribosome (3010)	120	**43 / 44**	**0.98**
Proteasome (3050)	31	**19 / 22**	**0.86**
Oxidative phosphorylation (190)	129	**31 / 44**	**0.70**
Cell cycle (4110)	114	**33 / 47**	**0.70**
ECM-receptor interaction (4512)	87	**16 / 23**	**0.70**
Gap junction (4540)	92	**9 / 13**	**0.69**
Pathogenic Escherichia coli infection (5130)	49	**11 / 16**	**0.69**
Pathogenic Escherichia coli infection (5131)	49	**11 / 16**	**0.69**
T cell receptor signaling pathway (4660)	93	**15 / 22**	**0.68**
Metabolism of xenobiotics by cytP450 (980)	70	**7 / 11**	**0.64**

Griffith et al. (2005)

pathway name (KEGG ID number)	nºgenes	genes coexp / genes	% gn coexp
Ribosome (3010)	120	**36 / 38**	**0.95**
Proteasome (3050)	31	**20 / 24**	**0.83**
Oxidative phosphorylation (190)	129	**55 / 67**	**0.82**
Val, Leu and isoleucine degradation (280)	50	**15 / 19**	**0.79**
ECM-receptor interaction (4512)	87	**16 / 22**	**0.73**
Cell cycle (4110)	114	**36 / 51**	**0.71**
Propanoate metabolism (640)	34	**9 / 14**	**0.64**
Butanoate metabolism (650)	44	**9 / 14**	**0.64**
Hematopoietic cell lineage (4640)	88	**18 / 28**	**0.64**
beta-Alanine metabolism (410)	24	**7 / 11**	**0.64**

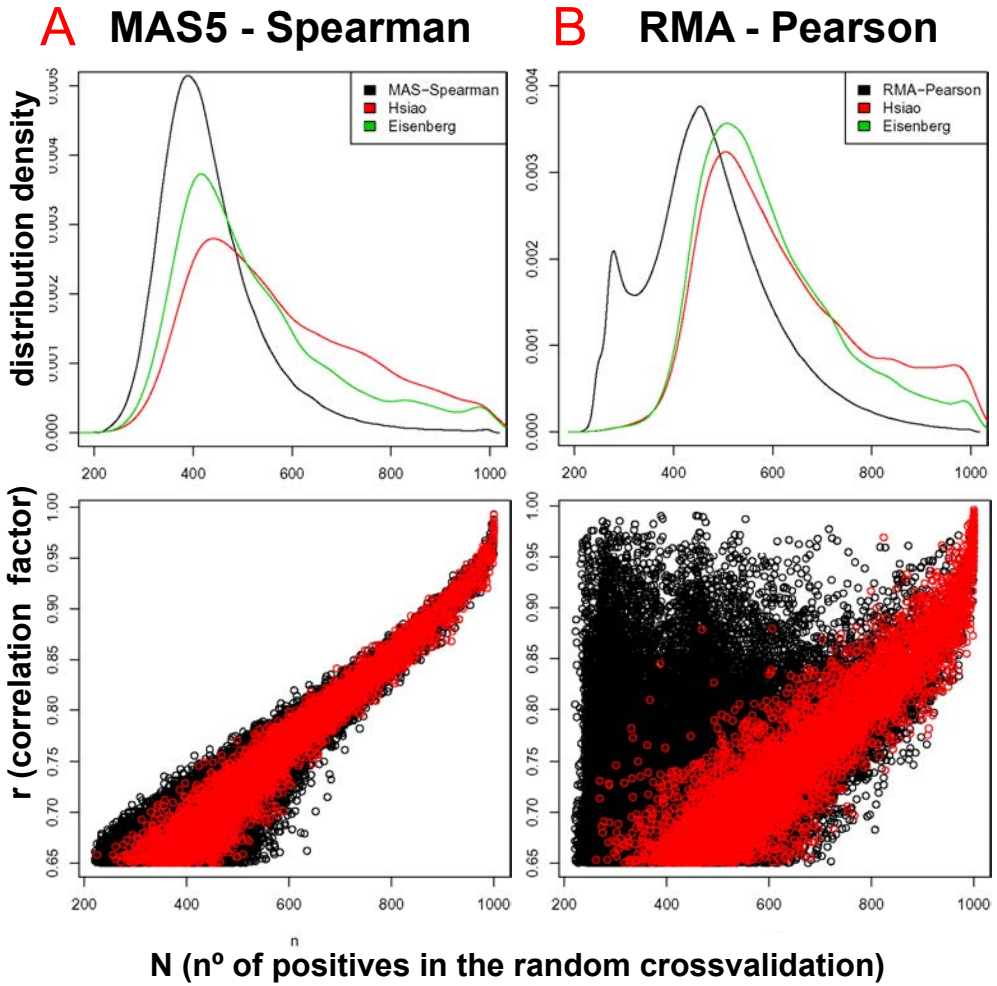
Gene2gene coexpression method

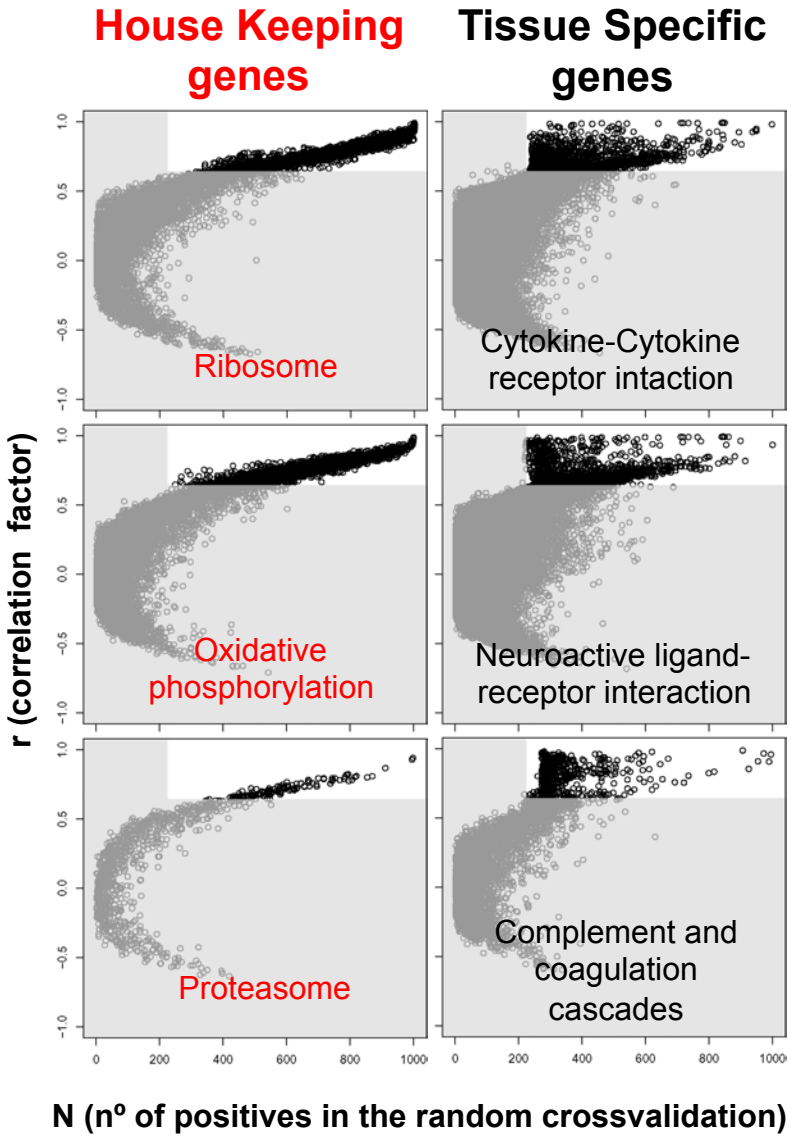
(based in combination of correlation **r** and crossvalidation **N**)



Gene2gene coexpression method

mapping coexpression of **house keeping genes** and **tissue specific genes** (based in KEGG pathways)

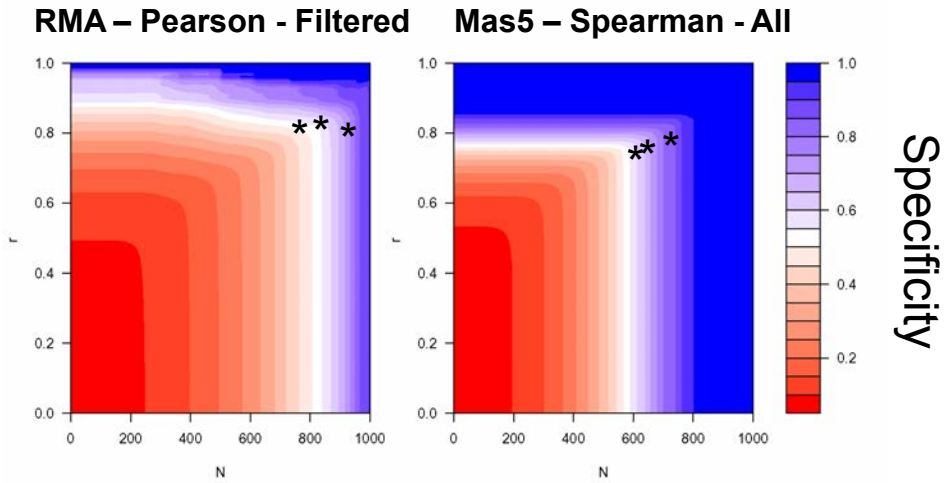




Hi-Fi gene2gene coexpression network

(based in combination of correlation **r** and crossvalidation **N**)

precision obtained for 3 reliable networks at high **r** and **N**



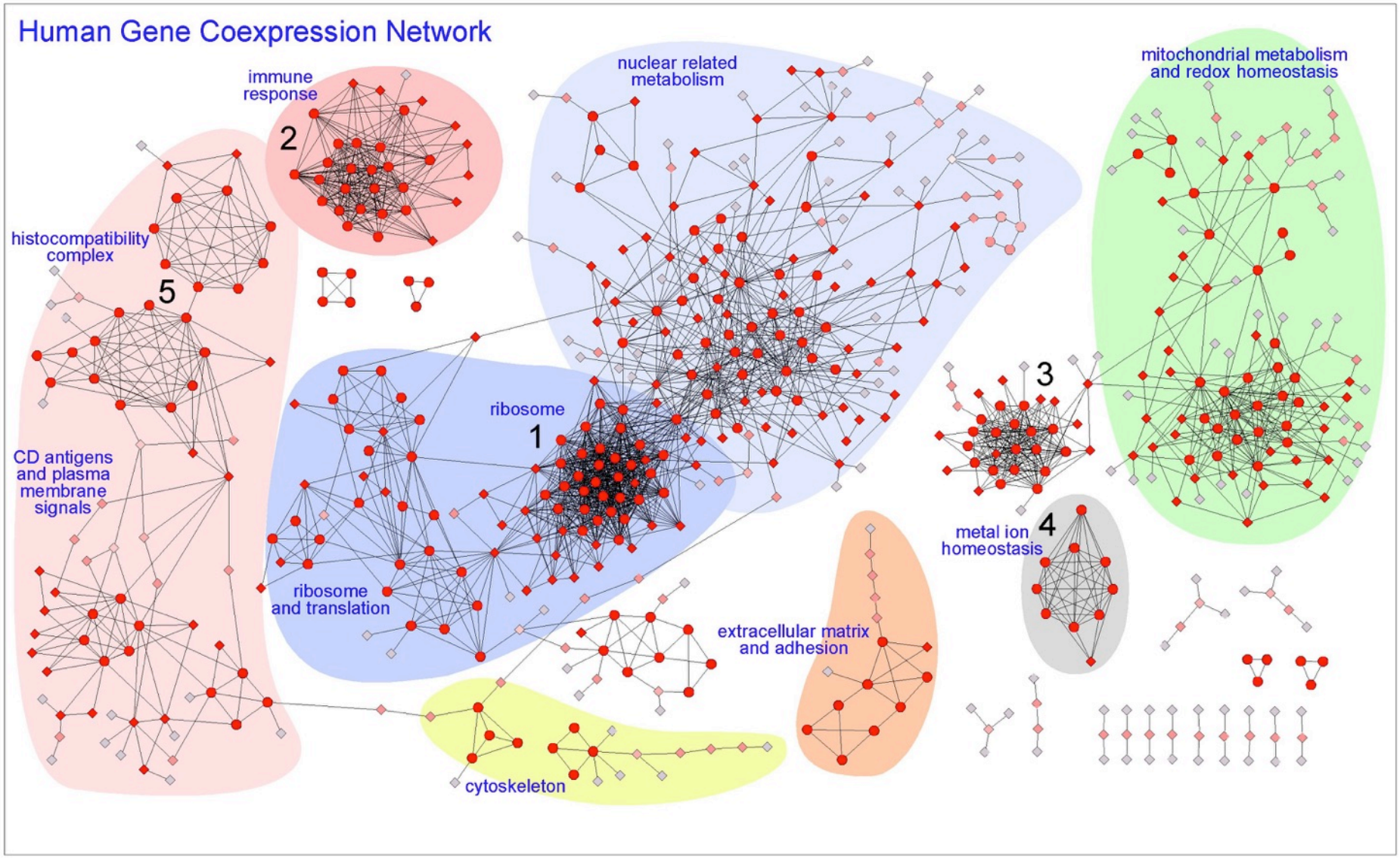
Precision [1]	Coefficients		Number of Nodes [2]	Number of Links [2]
	N	r		
RMA-Pearson (pre-Filtered)				
0.60	765	0.85	1.672	**5.945**
0.70	835	0.87	1.215	3.273
0.80	925	0.84	983	2.423
MAS5-Spearman (non-Filtered)				
0.60	605	0.77	3.052	**12.669**
0.70	645	0.79	2.295	7.874
0.80	695	0.81	1.762	4.910

1. Corresponds to the networks derived for KEGG annotated genes

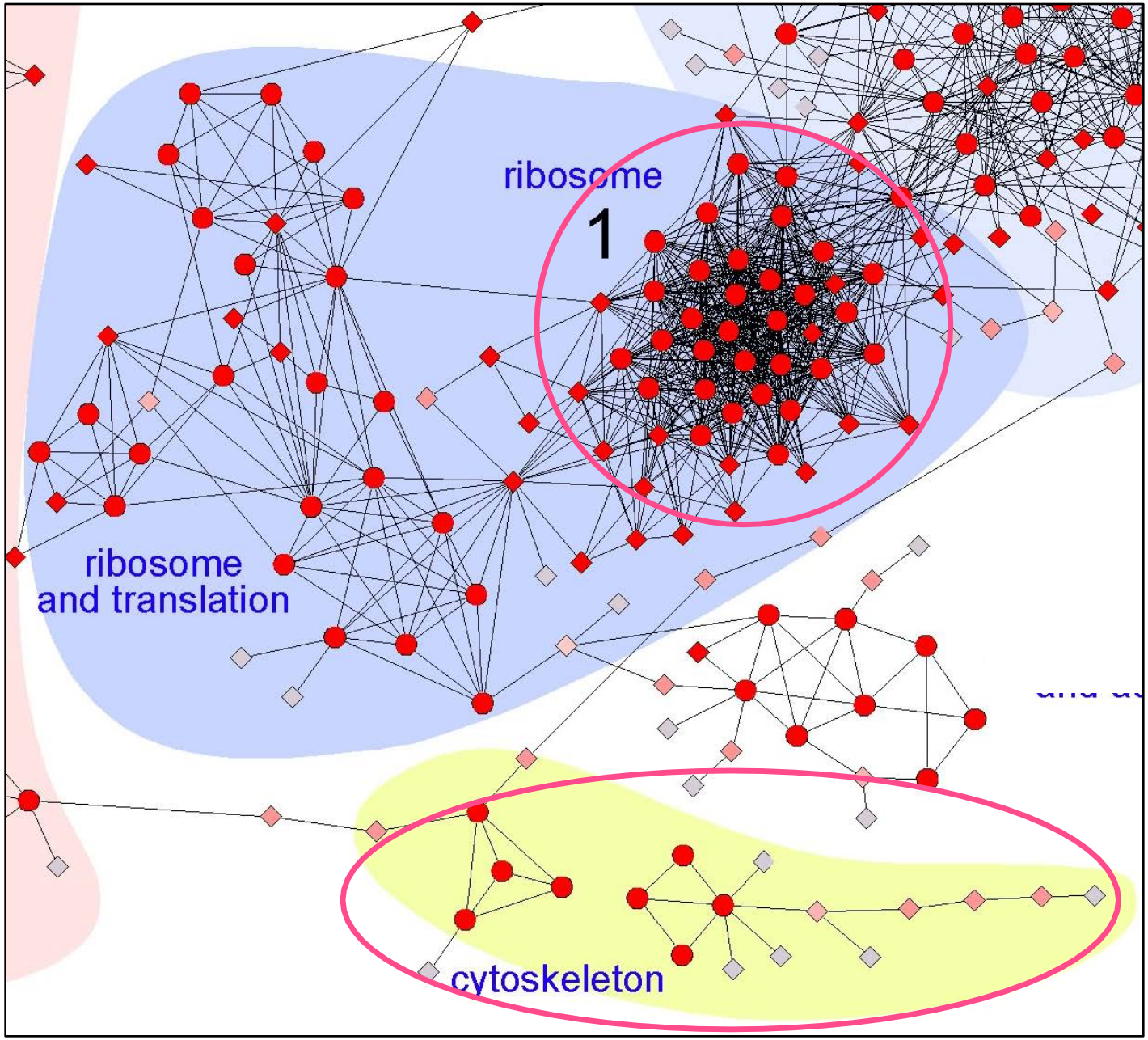
2. Corresponds to the full networks including all genes

Hi-Fi human coexpression network

network = intersection with 2 methods and **precision** ≥ 0.60 (**r** ≥ 0.77, **N** ≥ 605)



Hi-Fi human coexpression network



Analysis done with 2 algorithms
MCODE
MCL

nuclear related metabolism

ribosomal and translation

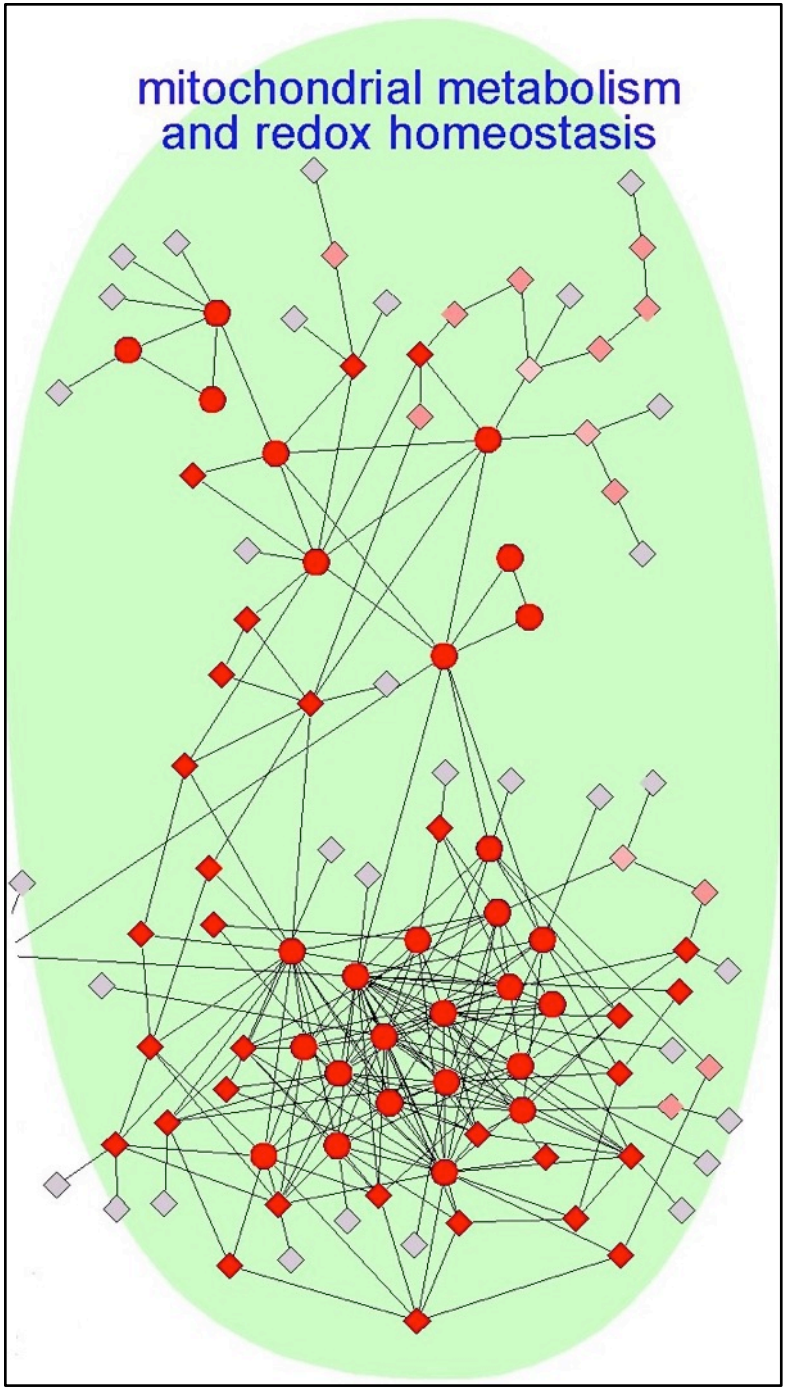
cytoskeleton

Hi-Fi human coexpression network

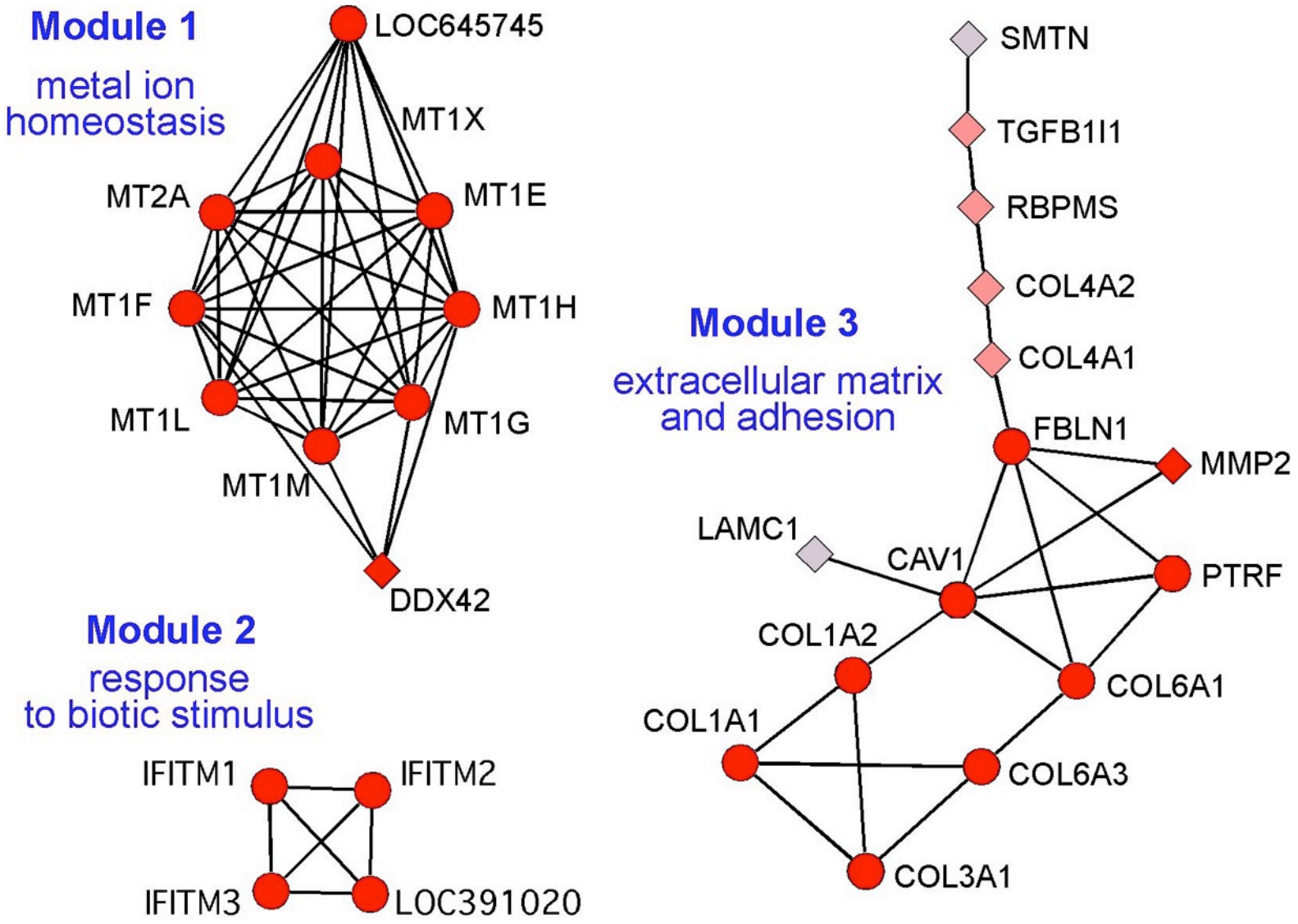
Analysis done
with 2 algorithms
MCODE
MCL

mitochondrial metabolism
and redox homeostasis

most genes of
the COX family,
the NDUF family and
the UQCR family

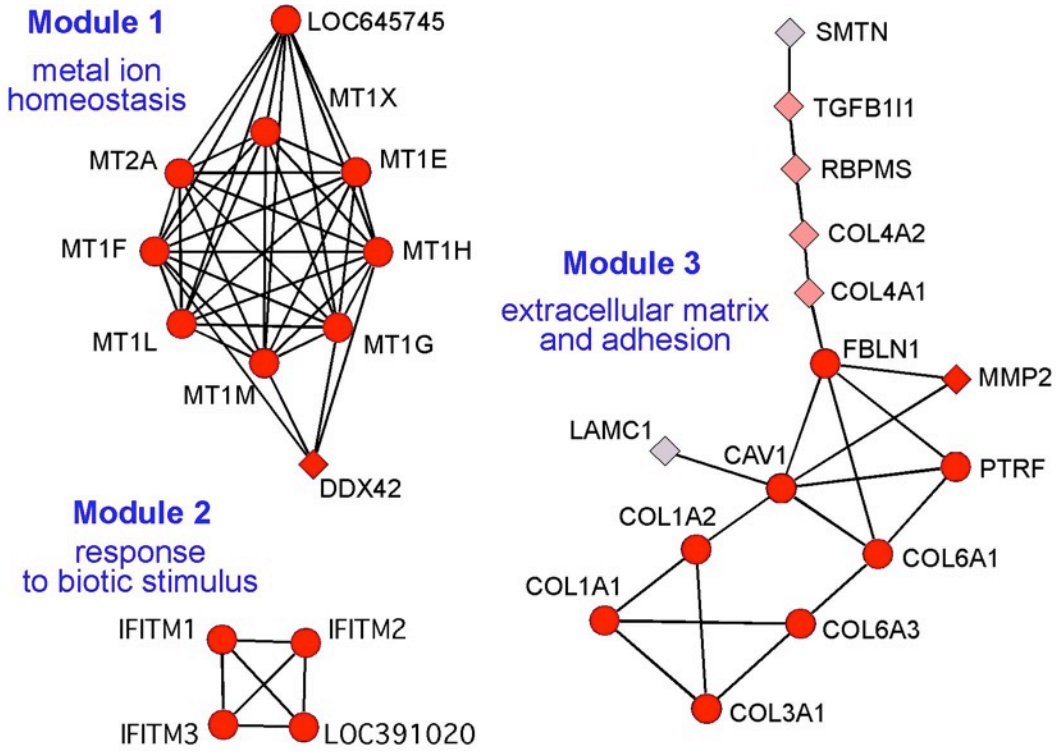


Hi-Fi human coexpression network (functionally coherent)



Hi-Fi human coexpression network

(modules coherent in terms of transcription factor TF regulation)



Coexp Modules	Search in	TF found	p-value	TransFac_db	TF Gene Name
Module 1	*PAP*	MTF-1	**0.001**	T02354	**MTF1** metal-regulatory transcription factor 1
10 genes	*Factory*	–	–		
Module 2	*PAP*	CRE-BP1	**0.0172**	T00167	**ATF2** activating transcription factor 2
4 genes	*Factory*	CRE-BP1	**0.0033**		
Module 3	*PAP*	Sp1	**0.13**	T00759	**SP1** Sp1 transcription factor
15 genes	*Factory*	Sp1	**0.017**		

Human transcriptomic network of normal tissues: a global map without malignant data

We achieved:

1st.- Reliable calculation of human genome-wide (global) expression data

2nd.- Reliable calculation of human gene2gene (global) co-expression data

Networks & Pathways

Comparison and combination of these type of complex data

Wu et al. (2010)

Wu *et al. Genome Biology* 2010, **11**:R53
http://genomebiology.com/2010/11/5/R53

Genome Biology

RESEARCH Open Access

A human functional protein interaction network and its application to cancer data analysis

Guanming Wu*[1], Xin Feng[2,3] and Lincoln Stein[1,2]

Data source	Proteins
Human PPIs	10,287
Fly PPIs	13,383
Worm PPIs	5,223
Yeast PPIs	5,646
Domain interaction	60,569
Lee's Gene Expression	8,250
✦ Prieto's Gene Expression	3,024
GO BP sharing	14,197
PPIs from GeneWays	5,252

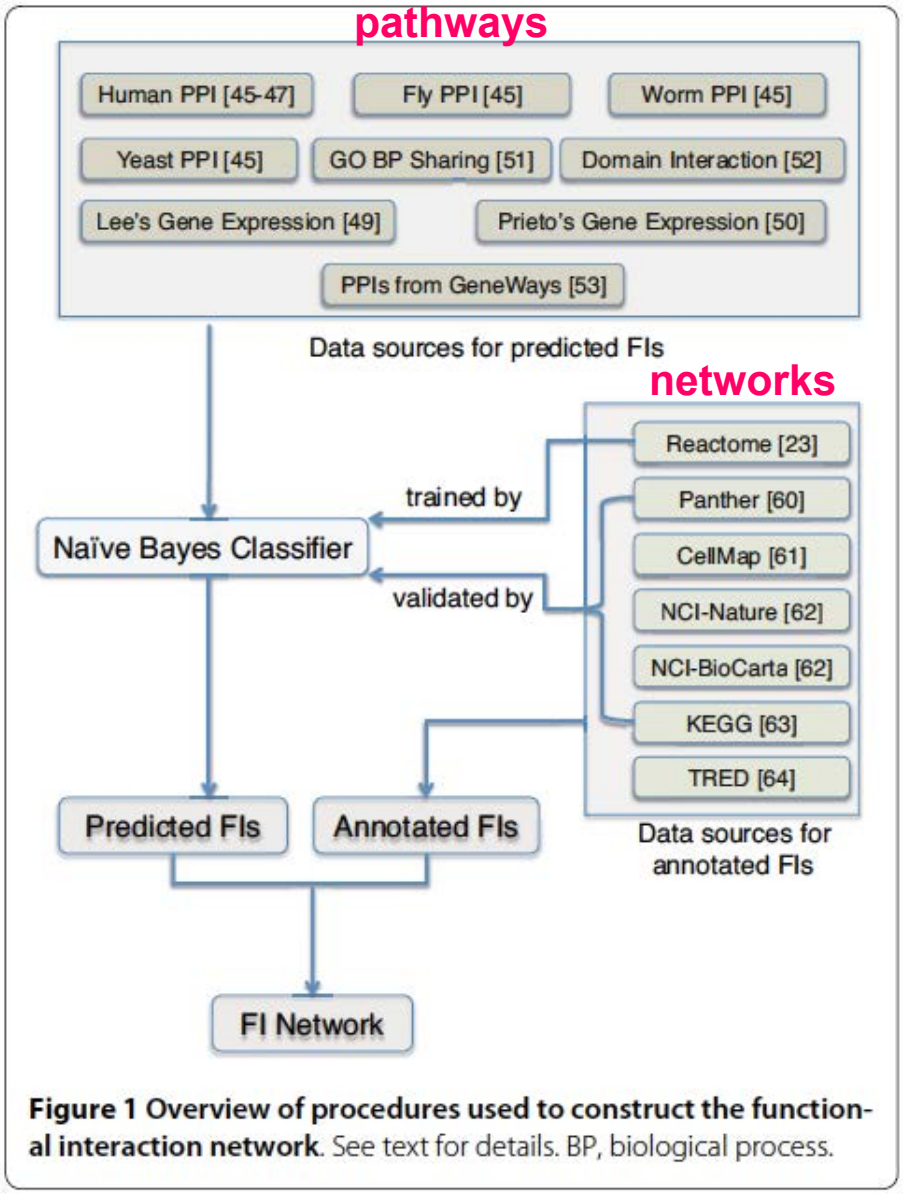


Figure 1 Overview of procedures used to construct the functional interaction network. See text for details. BP, biological process.

Networks & Pathways

Comparison and combination of these type of complex data

Wu et al. (2010)

Wu *et al. Genome Biology* 2010, **11**:R53
http://genomebiology.com/2010/11/5/R53

Genome Biology

RESEARCH Open Access

A human functional protein interaction network and its application to cancer data analysis

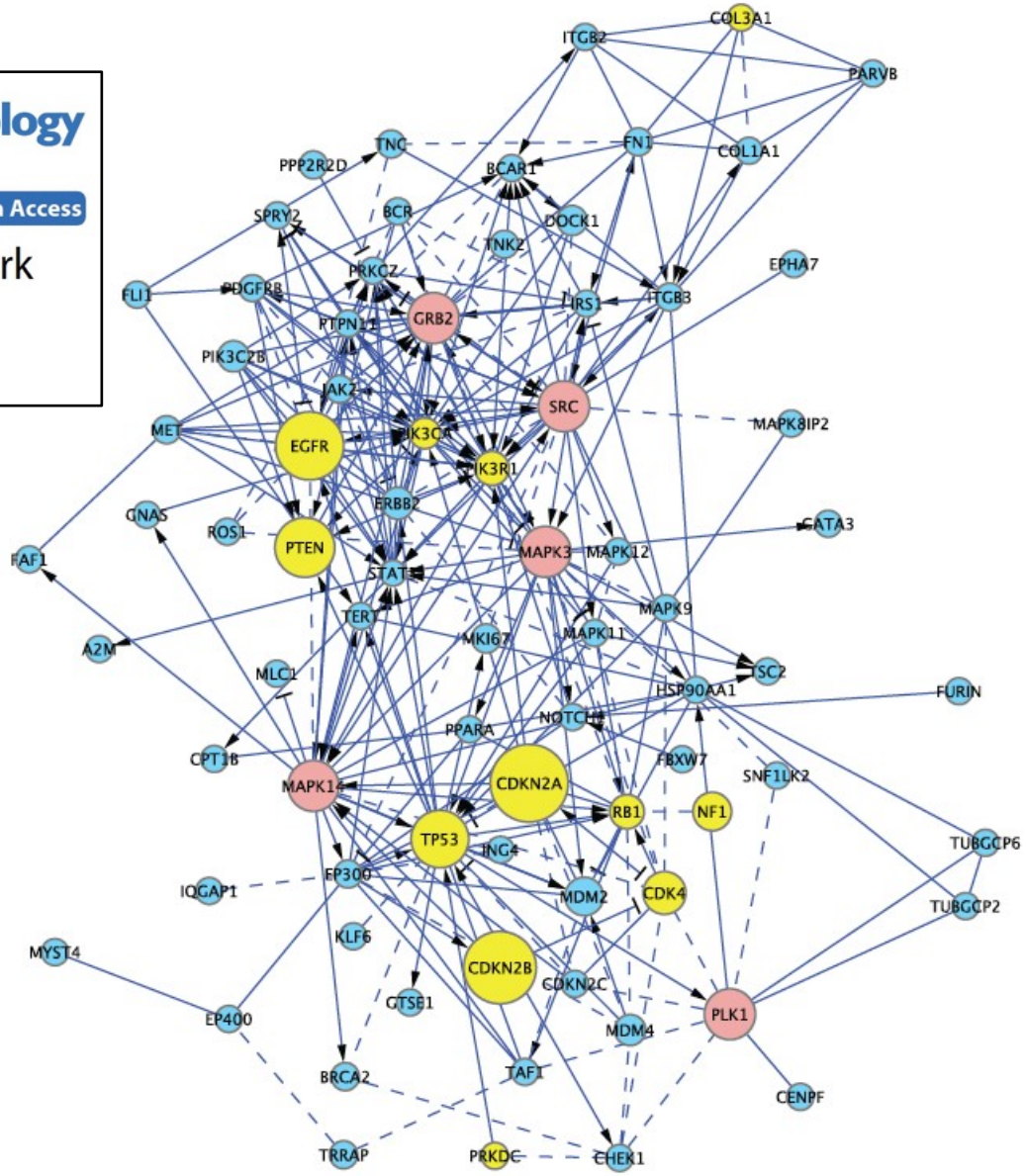
Guanming Wu*[1], Xin Feng[2,3] and Lincoln Stein[1,2]

Subnetwork derived from The Cancer Genome Atlas (TCGA) of somatic mutation data set:

77 cancer genes

and

5 linker genes



CBO Computational Biology @ UCT

CiC

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Hands-on: Practical Examples

Build the coexpression network for a gene list using Cytoscape (plugin ReactomeFI, that includes FI DB)
http://wiki.reactome.orglindex.php/Reactome_FI_Cytoscape_Piugin_4

Microarray DataAnalysis on a Network
(NejmlogRatioNormGiobaiZScore_070111.txt human)

Hands-on: Practical Examples

Build the coexpression network for a gene list using Cytoscape (plugin GeneMANIA)

Protein_

SETs_ 2014.xls

(NOTCH 33p human)

Session 3 (9:30 - 12:30, 3h)

Protein interaction networks

Session 4 (13:30 - 16:30, 3h)

Construction and analysis of gene/protein networks

- From gene expression signatures to gene coexpression networks
- Definition and properties of protein interaction networks
- Visualize and analyse biomolecular networks in Cytoscape
- Using on-line tools to build gene/protein networks: APID, STRING, GeneMANIA, PSICQUIC
- Network medicine: proteins and drugs interactions (STITCH)

Networks

Two major types of networks derived from experimental data

Two major types of networks derived from large-scale *omic* data

1.– **Gene Coexpression Networks**: ***ggcoe***
derived from gene expression profiling and transcriptomic studies

2.– **Protein-Protein Interaction Networks**: ***ppi***
derived from proteomic studies

Protein-Protein Interactions (PPIs)
biological networks

REVIEW

Zhu et al. (2007) Genes Dev.

Getting connected: analysis and principles of biological networks

Xiaowei Zhu,[1,2] Mark Gerstein,[3] and Michael Snyder[1,2,4]

The review shows that **PPI data** are, at present, a major part of the new systematic approaches to large-scale experimental determination of **biomolecular networks**

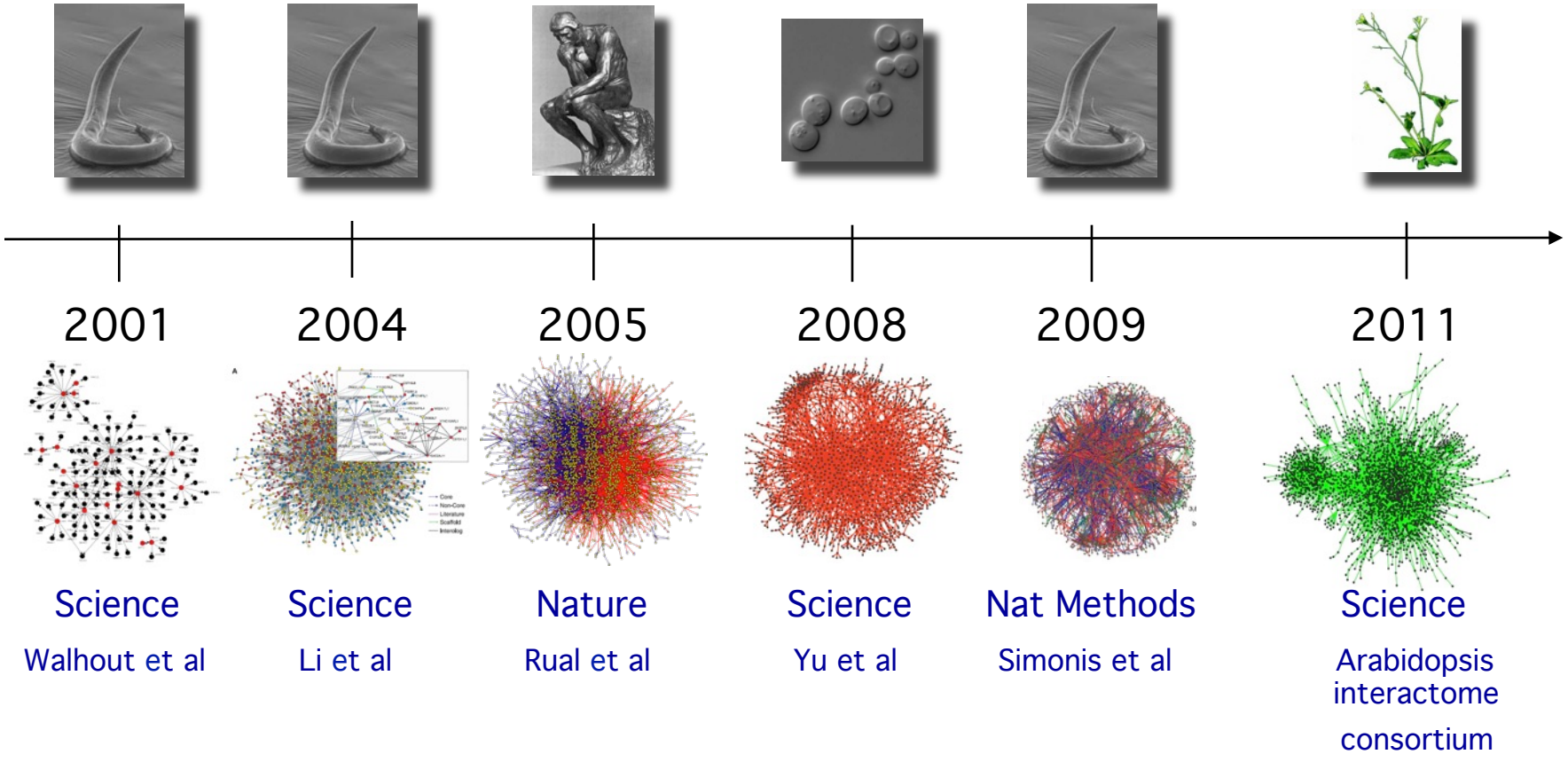
Type of network	Species	Number of nodes	Number of interactions	Reference
Transcription factor-binding network	*S. cerevisiae*	3528	7419	Yu et al. 2003[a]
		3207	11231	Harbison et al. 2004[b]
Protein–protein interaction	*C. elegans*	2788	4441	Stark et al. 2006
	D. melanogaster	7546	25403	
	Homo sapiens	7509	20979	
	Mus musculus	209	393	
	S. cerevisiae	5325	51773	
Phosphorylation network	*S. cerevisiae*	1325	4200	Ptacek et al. 2005
Metabolic network	*E. coli*	473	574	Guimera and Nunes Amaral 2005
	S. cerevisiae	646	1149	Tong et al. 2004
Genetic network	*S. cerevisiae*	3258	13963	Reguly et al. 2006[c]

[a]Transcriptional factor-binding data collected at rich-media condition.
[b]Transcriptional factor-binding data collected at a variety of growth conditions.
[c]Synthetic lethal interactions among nonessential genes.

From ***Zhu et al. (2007) Genes Dev.***

Protein-Protein Interactions (PPIs)

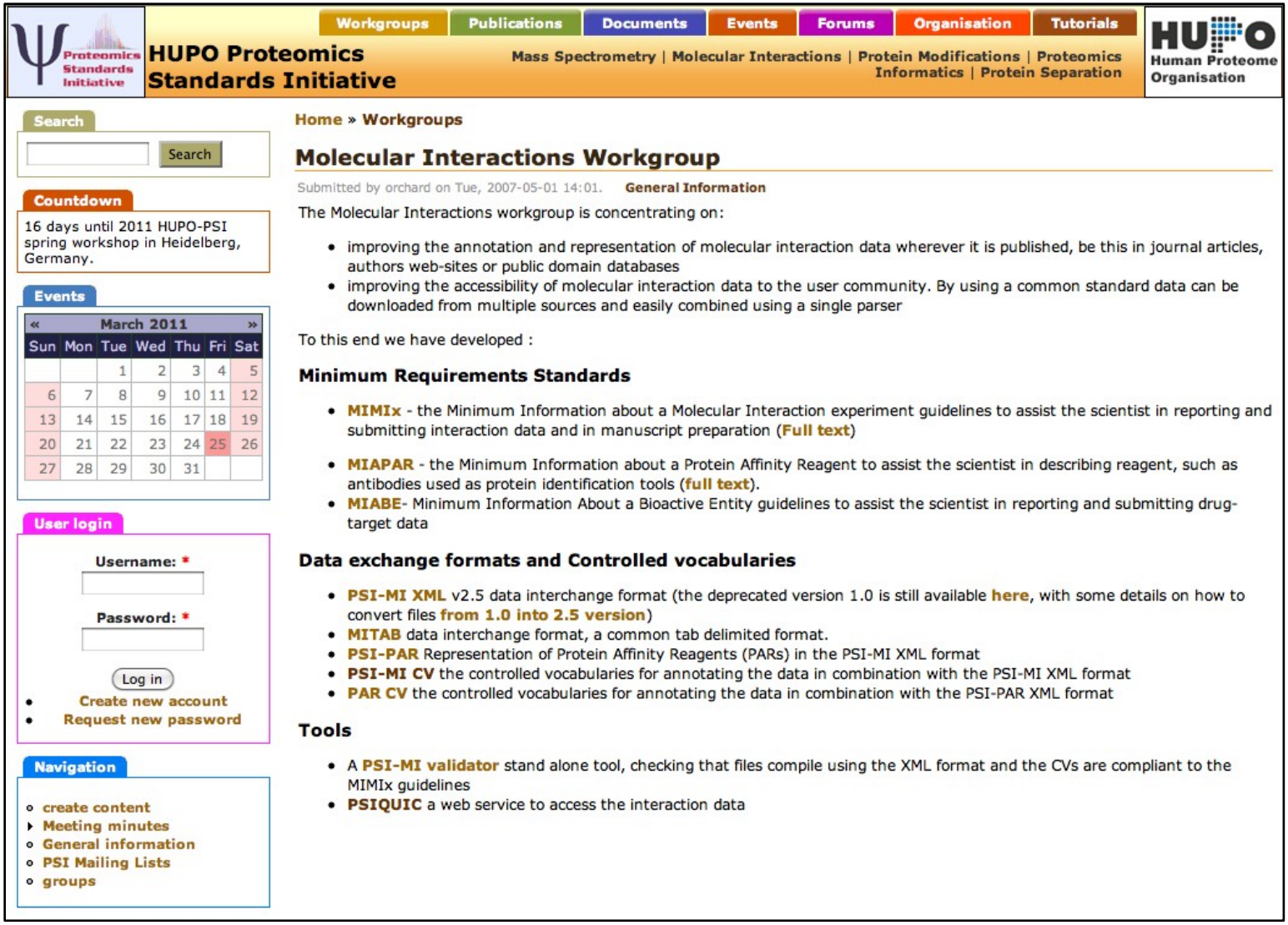
our first decade of interactome mappig: PPI data



Protein-Protein Interactions (PPIs)

international consortiums

Our group participates actively in HUPO **PSI-MI** (Molecular Interactions Workgroup)



Workgroups | Publications | Documents | Events | Forums | Organisation | Tutorials

HUPO Proteomics Standards Initiative

Mass Spectrometry | Molecular Interactions | Protein Modifications | Proteomics Informatics | Protein Separation

HUPO Human Proteome Organisation

Search

Search

Countdown

16 days until 2011 HUPO-PSI spring workshop in Heidelberg, Germany.

Events

« March 2011 »

Sun	Mon	Tue	Wed	Thu	Fri	Sat
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

User login

Username: *

Password: *

Log in

- Create new account
- Request new password

Navigation

- create content
- Meeting minutes
- General information
- PSI Mailing Lists
- groups

Home » Workgroups

Molecular Interactions Workgroup

Submitted by orchard on Tue, 2007-05-01 14:01. **General Information**

The Molecular Interactions workgroup is concentrating on:

- improving the annotation and representation of molecular interaction data wherever it is published, be this in journal articles, authors web-sites or public domain databases
- improving the accessibility of molecular interaction data to the user community. By using a common standard data can be downloaded from multiple sources and easily combined using a single parser

To this end we have developed :

Minimum Requirements Standards

- **MIMIx** - the Minimum Information about a Molecular Interaction experiment guidelines to assist the scientist in reporting and submitting interaction data and in manuscript preparation (**Full text**)
- **MIAPAR** - the Minimum Information about a Protein Affinity Reagent to assist the scientist in describing reagent, such as antibodies used as protein identification tools (**full text**).
- **MIABE**- Minimum Information About a Bioactive Entity guidelines to assist the scientist in reporting and submitting drug-target data

Data exchange formats and Controlled vocabularies

- **PSI-MI XML** v2.5 data interchange format (the deprecated version 1.0 is still available **here**, with some details on how to convert files **from 1.0 into 2.5 version**)
- **MITAB** data interchange format, a common tab delimited format.
- **PSI-PAR** Representation of Protein Affinity Reagents (PARs) in the PSI-MI XML format
- **PSI-MI CV** the controlled vocabularies for annotating the data in combination with the PSI-MI XML format
- **PAR CV** the controlled vocabularies for annotating the data in combination with the PSI-PAR XML format

Tools

- A **PSI-MI validator** stand alone tool, checking that files compile using the XML format and the CVs are compliant to the MIMIx guidelines
- **PSIQUIC** a web service to access the interaction data

Protein-Protein Interactions (PPIs)

international consortiums

There are several primary **PPIs** databases, but at present there is small integration.

PPIs
proteins
&
MIs
biomolecules

EU project
PSIMEx
FP7-HEALTH-2007-223411

Protein-Protein Interactions (PPIs)

review some essential concepts on **PPIs**

OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Education

Protein–Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks

Javier De Las Rivas*, Celia Fontanillo

Bioinformatics & Functional Genomics Research Group, Cancer Research Center (CiC-IBMCC, CSIC/USAL), Salamanca, Spain

PLoS Comp. Bio. (2010)

Protein-Protein Interactions (PPIs)
definition

The advancement of **genome and proteome-wide** experimental technologies have introduced modern biology in the **high complexity of living cells**, where thousands of **biomolecules work together** with many **cross-talks** and **cross-regulations**.

To achieve a first level of understanding of such cellular complexity we need to unravel the **interactions** that occur **between** all the **proteins** that integrate a living cell.

BUT, what do we mean by
protein-protein interaction ?

just
physical contact

or

other level of biomolecular
relation / association

What do we mean by protein interaction?

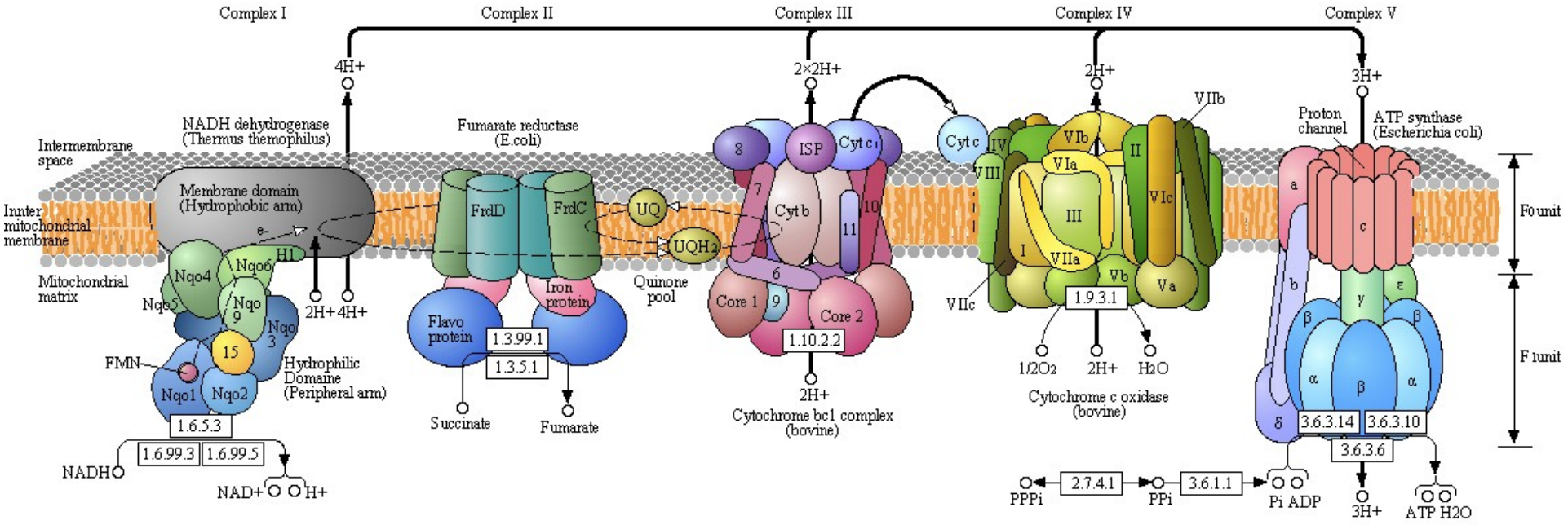
Intuitively, the definition of protein interaction in its more restrictive meaning would only involve the interaction produced by physical contact between the surfaces of two proteins. But most of the methods currently used have a bias towards the detection of higher levels of relation or association between proteins. Such protein relations can be very different: inclusion in *multiprotein complexes*, common *cellular compartments*, same *signalling pathway*, same *metabolic pathway*, *co-expression*, *genetic co-regulation*, or even molecular *co-evolution*.

From ***De Las Rivas et al. (2004) Comp. Funct. Genomics***

Protein-Protein Interactions (PPIs)
definition

The advancement of **genome and proteome-wide** experimental technologies have introduced modern biology in the **high complexity of living cells**, where thousands of **biomolecules work together** with many **cross-talks** and **cross-regulations**.

To achieve a first level of understanding of such cellular complexity we need to unravel the **interactions** that occur **between** all the **proteins** that integrate a living cell.



From ***De Las Rivas et al. (2004) Comp. Funct. Genomics***

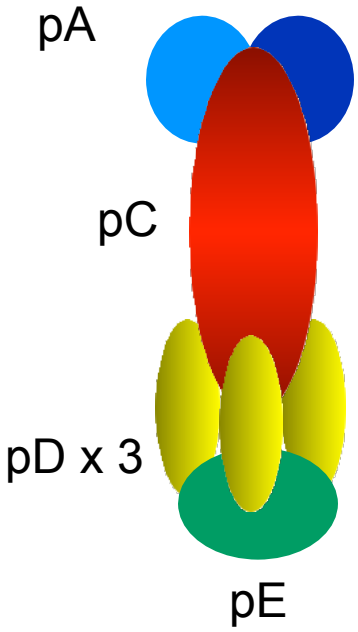
Protein-Protein Interactions (PPIs)
definition

It is important to define the **different types** of **associations between proteins** in order to make clear what are **PPI**.

I.- The **PPI** are proper **physical** interactions (and these can be **direct** or **indirect**)

1. *Co-interacting proteins*, defined as *physical interaction*:
   (a) *Permanent* interaction: proteins forming a stable protein complex that carries out a biomolecular role (structural or functional). These proteins are *protein subunits* of the complex and they work together. Examples include ATPase subunits, subunits of the nuclear pore, and ribosomal proteins within the S and L elements of the ribosome.
   (b) *Transient* interaction: proteins that come together in certain cellular states to undertake a biomolecular function. Examples include the DNA replicative complex, and most of the proteins involved in signal transduction cascades.

pApBpCpD3pE is a **complex**



physical direct
pA with pB
pD with pE
or
physical indirect
pA with pD
pB with pE

complex = stable molecular machine

From ***De Las Rivas et al. (2004) Comp. Funct. Genomics***

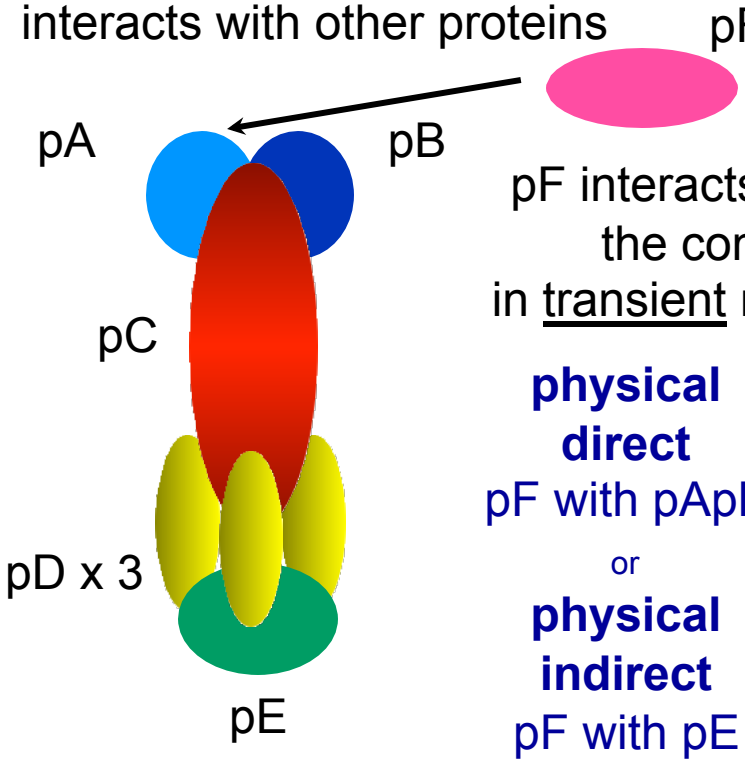
Protein-Protein Interactions (PPIs)
definition

It is important to define the **different types** of **associations between proteins** in order to make clear what are **PPI**.

II.- **PPI** can be **stable** (i.e. complexes) or **transient** (i.e. in signaling pathways)

1. *Co-interacting proteins*, defined as *physical interaction*:
   (a) *Permanent* interaction: proteins forming a stable protein complex that carries out a biomolecular role (structural or functional). These proteins are *protein subunits* of the complex and they work together. Examples include ATPase subunits, subunits of the nuclear pore, and ribosomal proteins within the S and L elements of the ribosome.
   (b) *Transient* interaction: proteins that come together in certain cellular states to undertake a biomolecular function. Examples include the DNA replicative complex, and most of the proteins involved in signal transduction cascades.

pApBpCpD3pE is a **complex**
interacts with other proteins



pF interacts with the complex in transient mode

physical direct
pF with pApB

or

physical indirect
pF with pE

complex = stable molecular machine

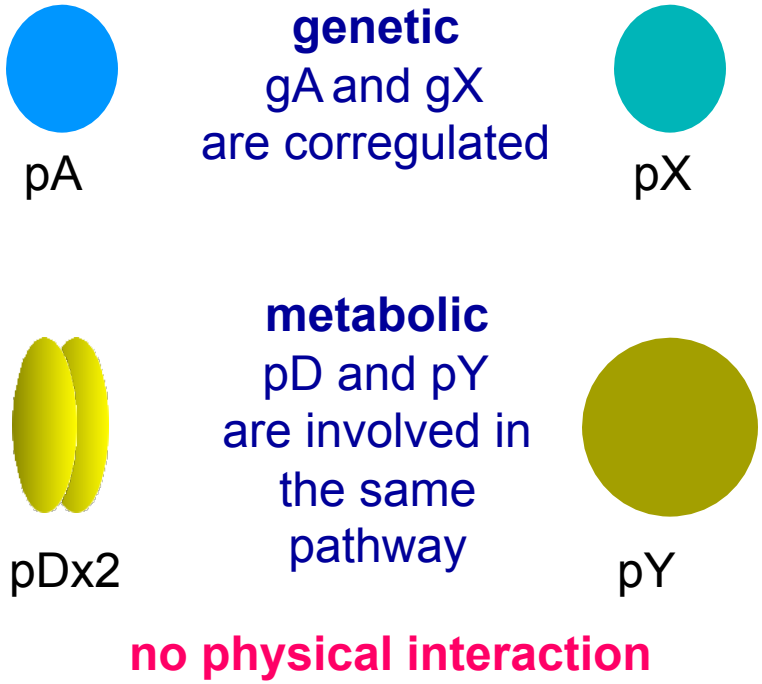
From ***De Las Rivas et al. (2004) Comp. Funct. Genomics***

Protein-Protein Interactions (PPIs)
definition

It is important to define the **different types** of **associations between proteins** in order to make clear what are **PPI**.

III.- Just **associations** but **not PPI** (because there are not **physical** interactions)

2. *Correlated proteins*, defined as proteins that are involved in the *same biomolecular activity* but that do not interact physically:
   (a) *Metabolic* correlation: proteins involved in the same metabolic pathway. These proteins are mostly *enzymes*. Examples include Krebs cycle enzymes, and prostaglandin synthesis enzymes.
   (b) *Genetic* correlation: proteins that are encoded by co-expressed or co-regulated genes. These could be called *operon-type* proteins. Examples include enzymes that regulate the glycolytic pathway, and proteins that regulate a phase of the cell cycle.



From ***De Las Rivas et al. (2004) Comp. Funct. Genomics***

Protein-Protein Interactions (PPIs)
definition

What do we mean by **protein-protein interaction** ?

Protein-to-Protein interactions (**PPIs**) are specific **physical contacts** between **protein pairs** that occur by selective **molecular docking** in a particular **biological context**.

Forward-looking two main challenges remain in the field:

(i)a better **filtering** of **false positives** in the PPI collections

(ii)an adequate **distinction** of the **biological context** that specifies and determines the existence or not of a given PPI at a given biological situation.

From ***De Las Rivas & Fontanillo (2010)***

Protein-Protein Interactions (PPIs)

review some essential concepts on **PPIs**

OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

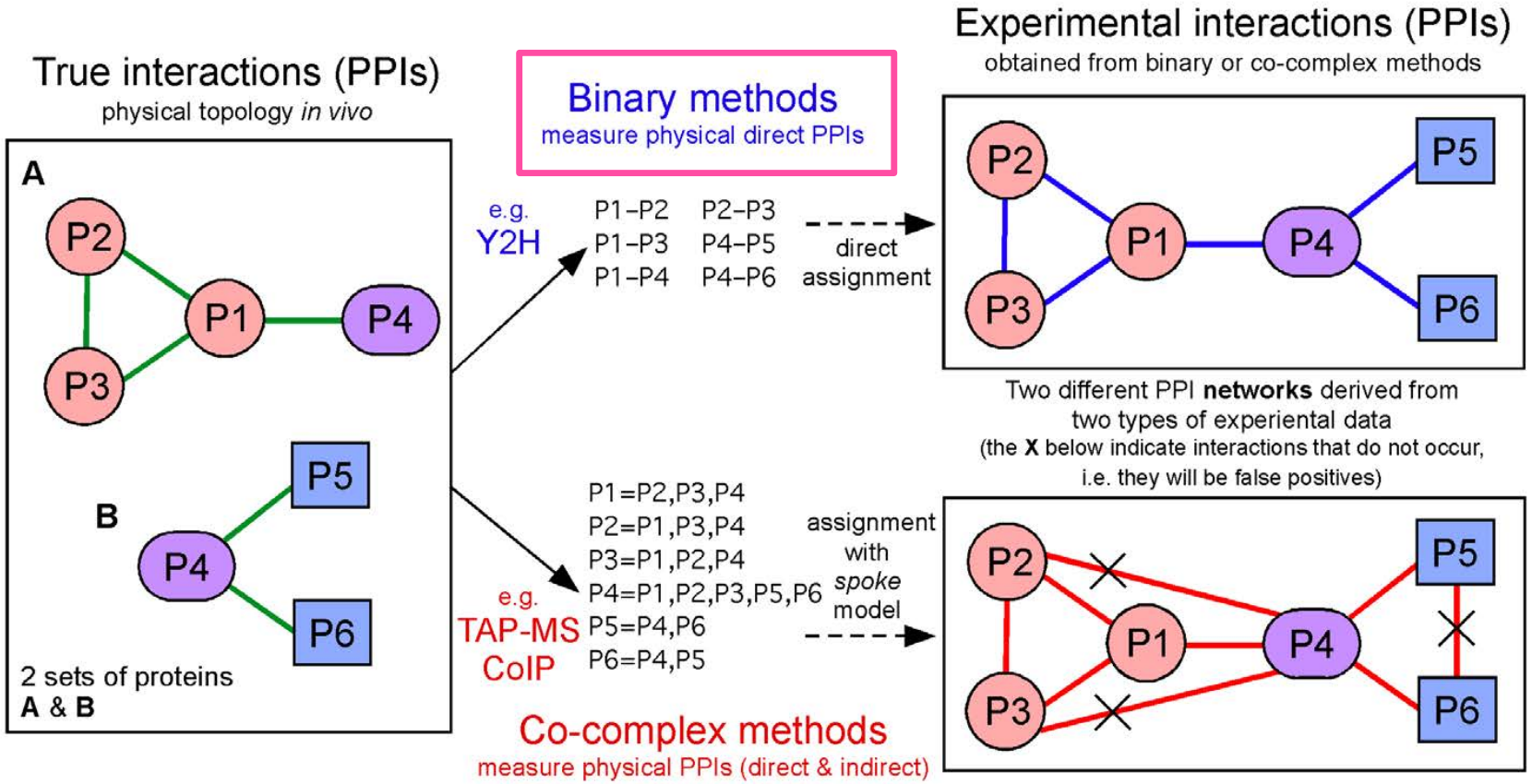
Education

Protein–Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks

Javier De Las Rivas*, Celia Fontanillo

Bioinformatics & Functional Genomics Research Group, Cancer Research Center (CiC-IBMCC, CSIC/USAL), Salamanca, Spain

PLoS Comp. Bio. (2010)



Protein-Protein Interactions (PPIs)
types of experimental methods

Within the last years a large amount of data on protein-protein interactions in cellular systems has been obtained both by the **high-throughput** and **small scale technologies**. A list of most relevant **methods** to is presented:

Complex oriented methods (find *multimeric* PPIs)

- Co-Immunoprecipitation (Co-IP)
- Pull-Down Assays
- Tandem Affinity Purification + Mass Spectrometry (TAP-MS)

Binary oriented methods (find *dimeric* PPIs)

- Two Hybrid systems (Y2H)
- Protein Arrays / Protein Chips

3D-structure based methods (find specific PPI interfaces)

- X-ray Crystallography (X-ray)
- Electro Microscopy (EM)
- Nuclear Magnetic Resonance (NMR)

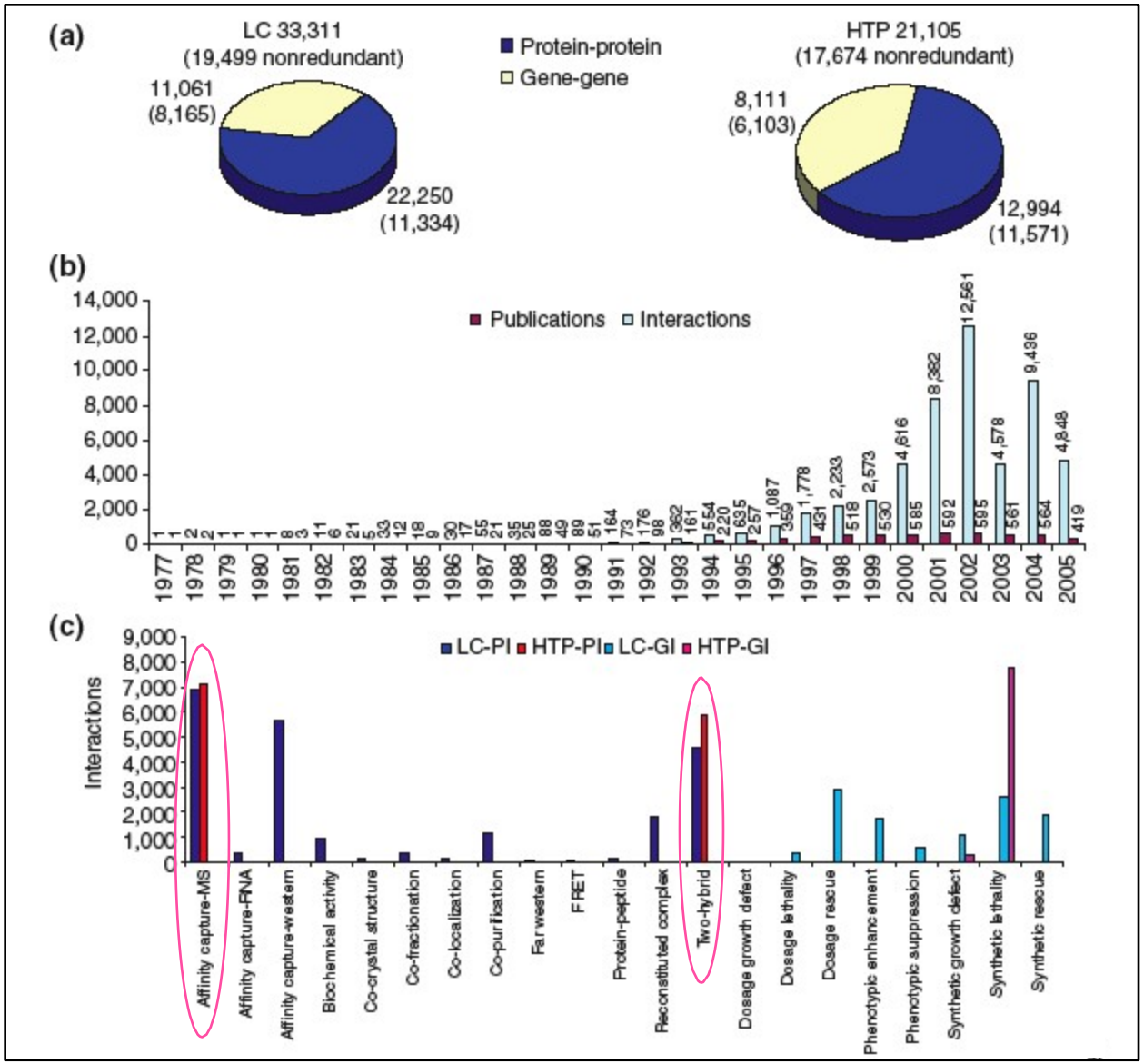
Protein-Protein Interactions (PPIs)

types of experimental methods

Data about the **YEAST** interactome

Two main **high-throughput proteomic techniques** have been applied to determine PPIs:

TAP-MS
&
Y2H

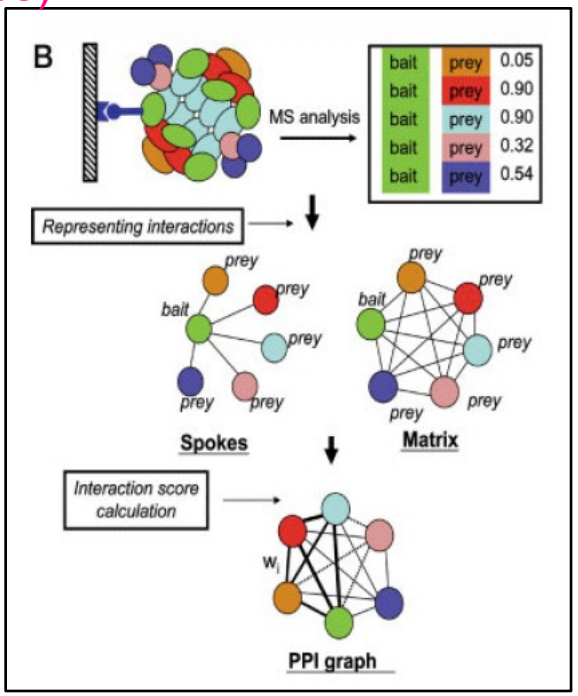


From ***Reguly et al. (2006) Journal of Biology***

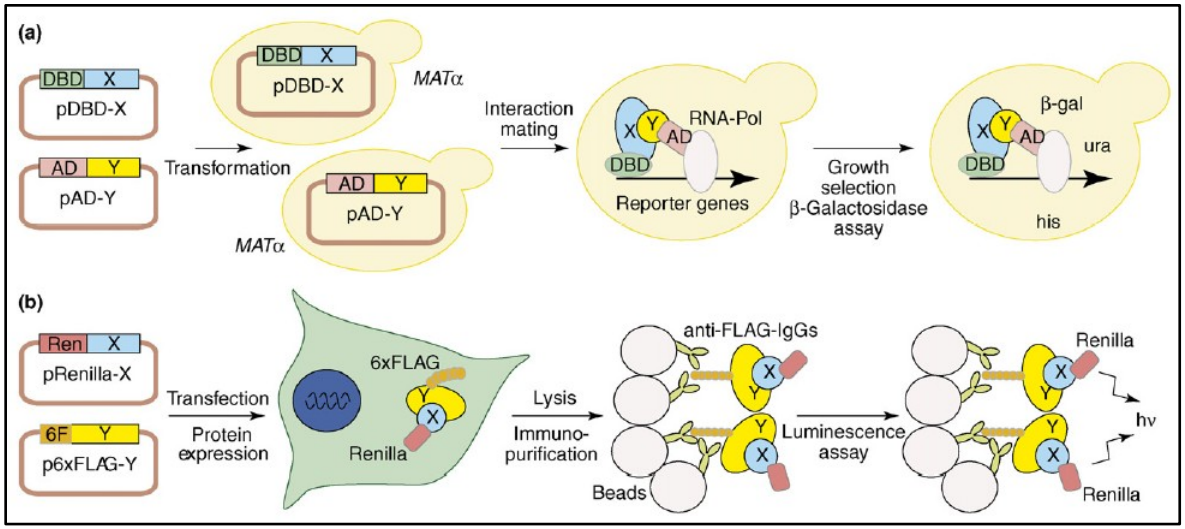
Protein-Protein Interactions (PPIs)
major high-throughput experimental methods

In recent years two main **high-throughput proteomic techniques** have been applied to determine PPIs:

– **Tandem-Affinity Purification and Mass Spectrometry (TAP-MS)** provides multimer interactions (complexes)



– **High-throughput Two-Hybrid systems (Y2H)** provides binary interactions



The Human Interactome

Two major large-scale data types: TAP-MS and Y2H

In recent years two main **high-throughput proteomic techniques** have been applied to determine PPIs:

- **Tandem-Affinity Purification and Mass Spectrometry (TAP-MS)** provides multimer interactions (complexes)

- **High-throughput Two-Hybrid systems (Y2H)** provides binary interactions

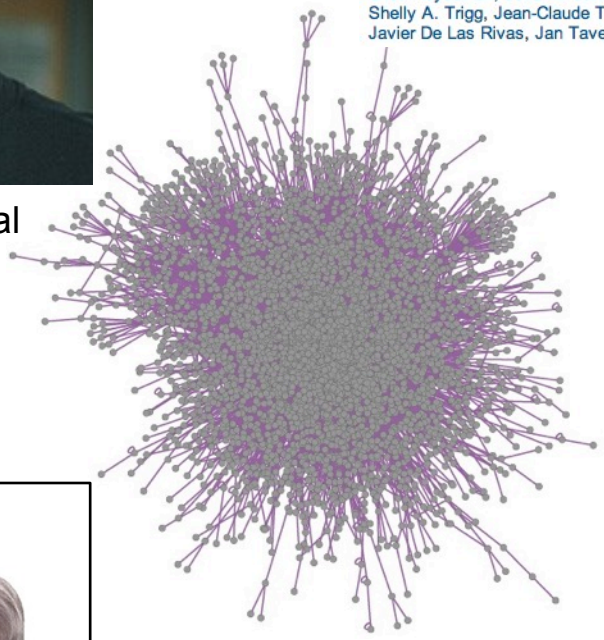
Dr. Marc Vidal
(Boston)

Cell

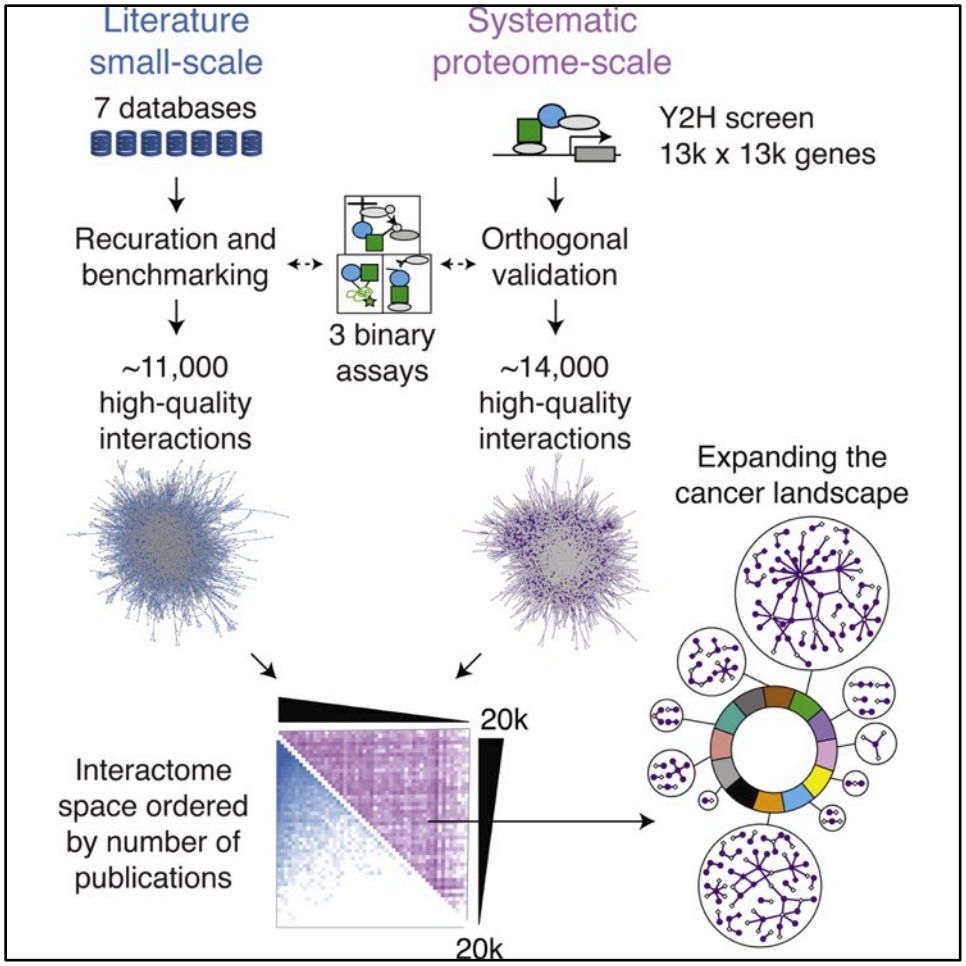
Volume 159, November 2014

A Proteome-Scale Map of the Human Interactome Network

Thomas Rolland[19], Murat Taşan[19], Benoit Charloteaux[19], Samuel J. Pevzner[19], Quan Zhong[19], Nidhi Sahni[19], Song Yi[19], Irma Lemmens, Celia Fontanillo, Roberto Mosca, Atanas Kamburov, Susan D. Ghiassian, Xinping Yang, Lila Ghamsari, Dawit Balcha, Bridget E. Begg, Pascal Braun, Marc Brehme, Martin P. Broly, Anne-Ruxandra Carvunis, Dan Convery-Zupan, Roser Corominas, Jasmin Coulombe-Huntington, Elizabeth Dann, Matija Dreze, Amélie Dricot, Changyu Fan, Eric Franzosa, Fana Gebreab, Bryan J. Gutierrez, Madeleine F. Hardy, Mike Jin, Shuli Kang, Ruth Kiros, Guan Ning Lin, Katja Luck, Andrew MacWilliams, Jörg Menche, Ryan R. Murray, Alexandre Palagi, Matthew M. Poulin, Xavier Rambout, John Rasla, Patrick Reichert, Viviana Romero, Elien Ruyssinck, Julie M. Sahalie, Annemarie Scholz, Akash A. Shah, Amitabh Sharma, Yun Shen, Kerstin Spirohn, Stanley Tam, Alexander O. Tejeda, Shelly A. Trigg, Jean-Claude Twizere, Kerwin Vega, Jennifer Walsh, Michael E. Cusick, Yu Xia, Albert-László Barabási, Lilia M. Iakoucheva, Patrick Aloy, Javier De Las Rivas, Jan Tavernier, Michael A. Calderwood[20], David E. Hill[20], Tong Hao[20], Frederick P. Roth✉, Marc Vidal✉



The network: a systematic map of
*≈ **14,000** interactions between ≈ **4,000** human proteins*



Dr. Javier De Las Rivas
(Salamanca)

(Y2H) binary

Rolland et al. (2014) Cell

The BioPlex Network: A Systematic Exploration of the Human Interactome

Edward L. Huttlin,[1] Lily Ting,[1] Raphael J. Bruckner,[1] Fana Gebreab,[1] Melanie P. Gygi,[1] John Szpyt,[1] Stanley Tam,[1] Gabriela Zarraga,[1] Greg Colby,[1] Kurt Baltier,[1] Rui Dong,[2] Virginia Guarani,[1] Laura Pontano Vaites,[1] Alban Ordureau,[1] Ramin Rad,[1] Brian K. Erickson,[1] Martin Wühr,[1] Joel Chick,[1] Bo Zhai,[1] Deepak Kolippakkam,[1] Julian Mintseris,[1] Robert A. Obar,[1,3] Tim Harris,[3] Spyros Artavanis-Tsakonas,[1,3] Mathew E. Sowa,[1] Pietro De Camilli,[2] Joao A. Paulo,[1] J. Wade Harper,[1,*] and Steven P. Gygi[1,*]

[1]Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA

[2]Department of Cell Biology and Howard Hughes Medical Institute, Yale School of Medicine, New Haven, CT 06519, USA

[3]Biogen, Cambridge, MA 02142, USA

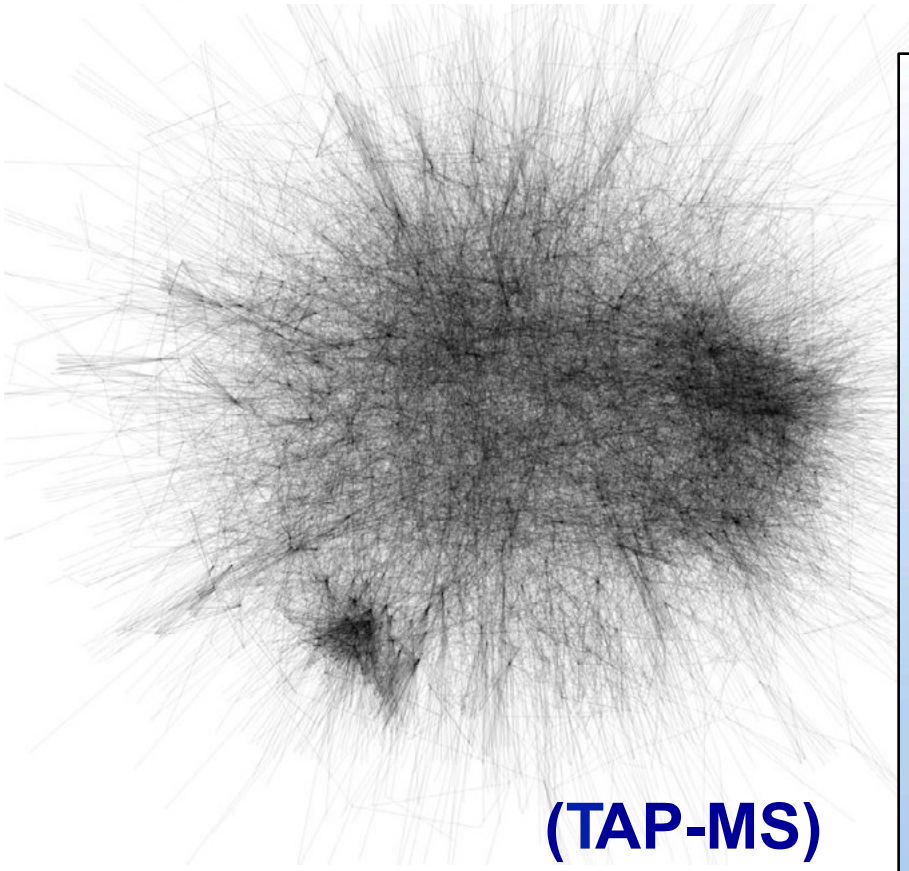
*Correspondence: wade_harper@hms.harvard.edu (J.W.H.), steven_gygi@hms.harvard.edu (S.P.G.)

http://dx.doi.org/10.1016/j.cell.2015.06.043

Cell

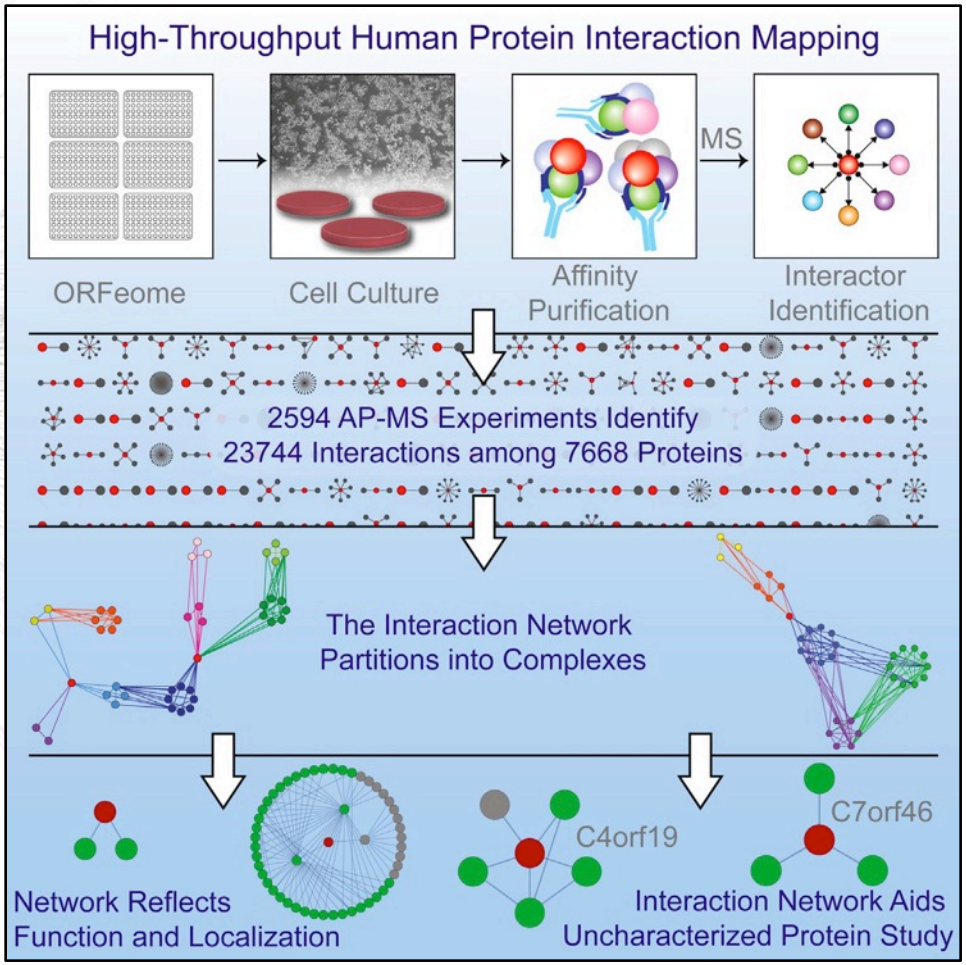
Volume 162, July 2015

The network: a systematic map of
*≈ **23,744** interactions between ≈ **7,668** human proteins*



(TAP-MS) co-complex

Huttlin et al. (2015) Cell



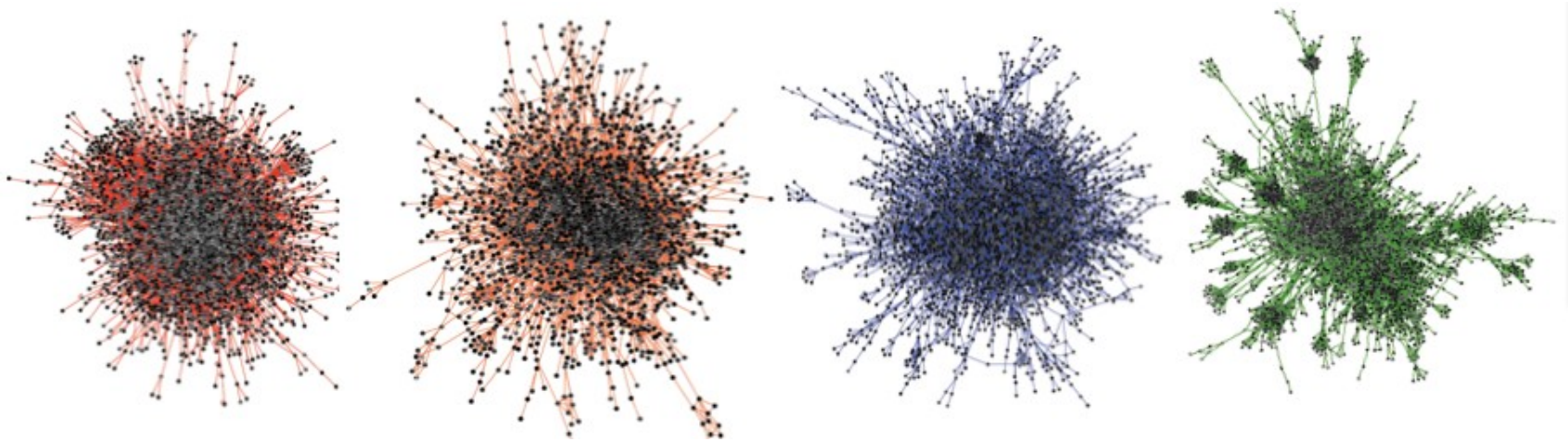
Protein-Protein Interactions (PPIs)

major high-throughput experimental methods

Different human *protein to protein interaction networks*: ***ppi***

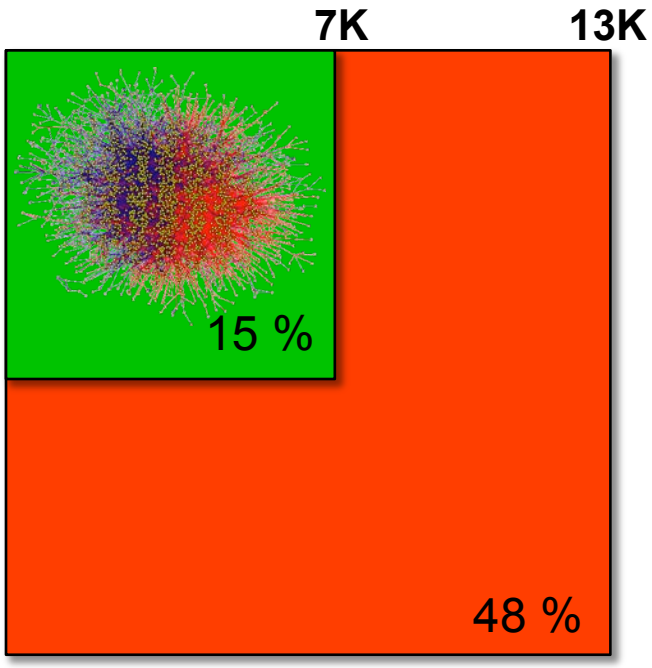
A Proteome-Scale Map of the Human Interactome Network

Thomas Rolland,[1,2,19] Murat Taşan,[1,3,4,5,19] Benoit Charloteaux,[1,2,19] Samuel J. Pevzner,[1,2,6,7,19] Quan Zhong,[1,2,8,19] Nidhi Sahni,[1,2,19] Song Yi,[1,2,19] Irma Lemmens,[9] Celia Fontanillo,[10] Roberto Mosca,[11] Atanas Kamburov,[1,2] Susan D. Ghiassian,[1,12] Xinping Yang,[1,2] Lila Ghamsari,[1,2] Dawit Balcha,[1,2] Bridget E. Begg,[1,2] Pascal Braun,[1,2] Marc Brehme,[1,2] Martin P. Broly,[1,2] Anne-Ruxandra Carvunis,[1,2] Dan Convery-Zupan,[1,2] Roser Corominas,[13] Jasmin Coulombe-Huntington,[1,14] Elizabeth Dann,[1,2] Matija Dreze,[1,2] Amélie Dricot,[1,2] Changyu Fan,[1,2] Eric Franzosa,[1,14] Fana Gebreab,[1,2] Bryan J. Gutierrez,[1,2] Madeleine F. Hardy,[1,2] Mike Jin,[1,2] Shuli Kang,[13] Ruth Kiros,[1,2] Guan Ning Lin,[13] Katja Luck,[1,2] Andrew MacWilliams,[1,2] Jörg Menche,[1,12] Ryan R. Murray,[1,2] Alexandre Palagi,[1,2] Matthew M. Poulin,[1,2] Xavier Rambout,[1,2,15] John Rasla,[1,2] Patrick Reichert,[1,2] Viviana Romero,[1,2] Elien Ruyssinck,[9] Julie M. Sahalie,[1,2] Annemarie Scholz,[1,2] Akash A. Shah,[1,2] Amitabh Sharma,[1,12] Yun Shen,[1,2] Kerstin Spirohn,[1,2] Stanley Tam,[1,2] Alexander O. Tejeda,[1,2] Shelly A. Trigg,[1,2] Jean-Claude Twizere,[1,2,15] Kerwin Vega,[1,2] Jennifer Walsh,[1,2] Michael E. Cusick,[1,2] Yu Xia,[1,14] Albert-László Barabási,[1,12,16] Lilia M. Iakoucheva,[13] Patrick Aloy,[11,17] Javier De Las Rivas,[10] Jan Tavernier,[9] Michael A. Calderwood,[1,2,20] David E. Hill,[1,2,20] Tong Hao,[1,2,20] Frederick P. Roth,[1,3,4,5,18,*] and Marc Vidal[1,2,*]

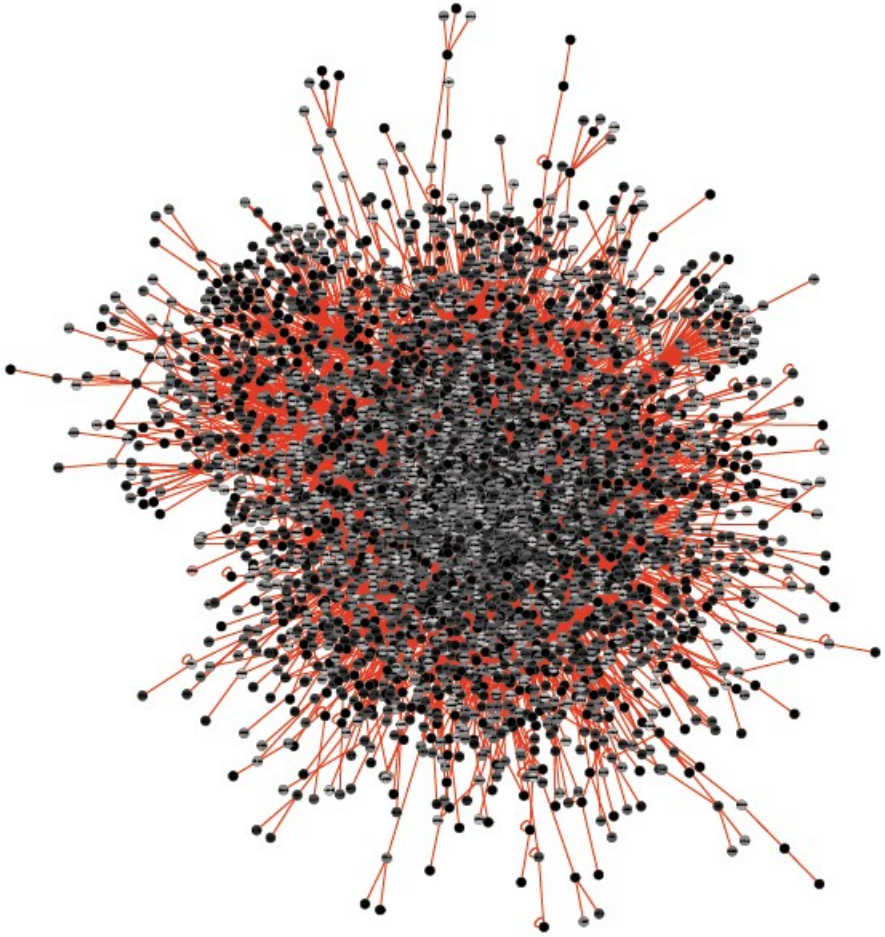


From ***Rolland et al. (2014) Cell***

Binary Human Interactome



HI-II-2014



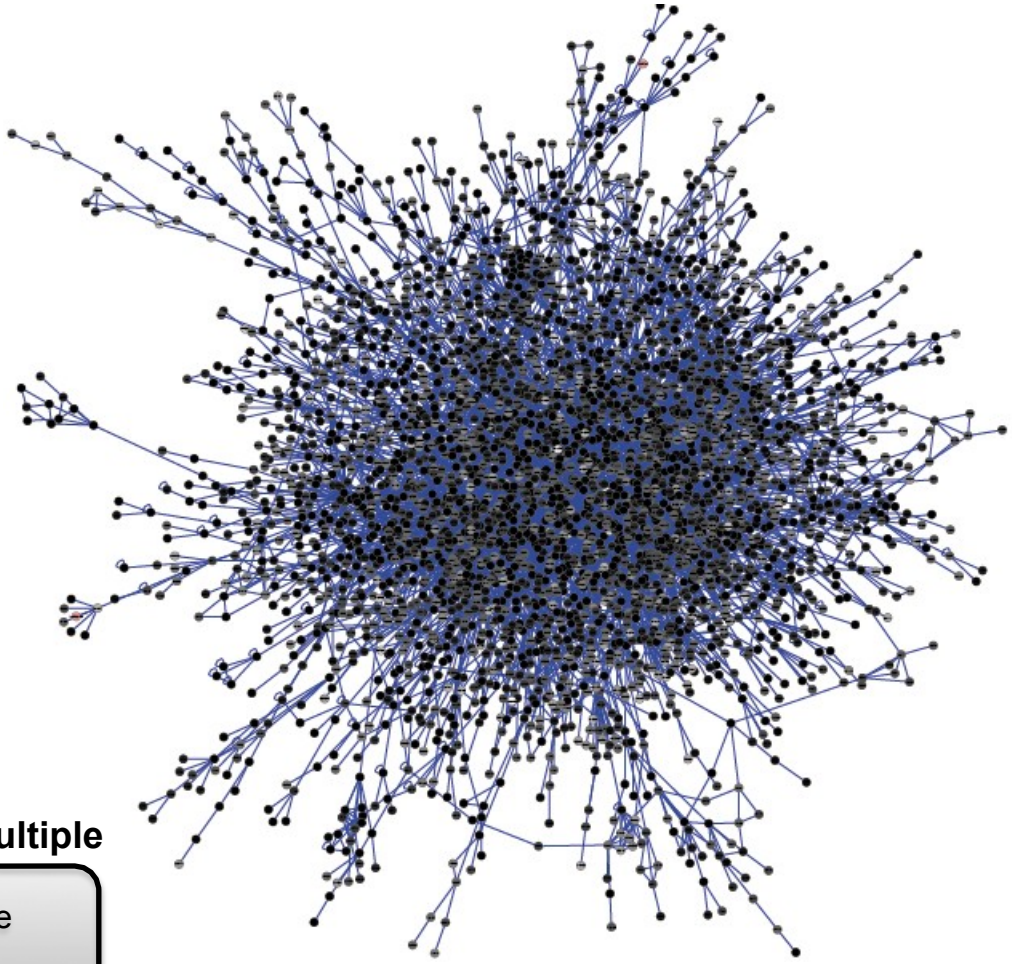
Dataset Name	Release Date	N proteins	N interactions	Search Space
HI-I-2005	2005	1,545	2,750	Space I
HI-2011	2005, 2008, 2011	2,191	3,881	—
HI-II-2014	2014	4,303	13,944	Space II
CCSB-HI-all	**2015 (total union)**	**4,745**	**16,503**	**Space I & II**

Literature Human Interactome

Literature Curated Reference Sets Lit-RS

Validated PRS/RRS

Lit-BM



Random Reference Set RRS
(1000 ppi, 699 heterodimers)
random selected & validated

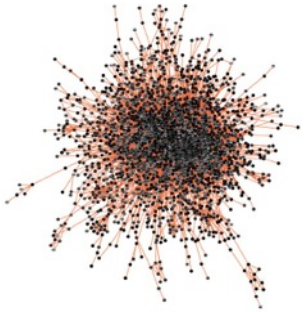
...

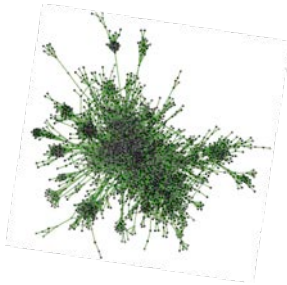
Positive Reference Set PRS (LCI-RS)
(1000 ppi, 699 heterodimers)
random selected & validated

Lit-BM binary multiple

Binary Multiple
7,475

Analysis of Human Interactomes





Cell

Resource

Cell 2012

A Census of Human Soluble Protein Complexes

Pierre C. Havugimana,[1,2,8] G. Traver Hart,[1,2,8] Tamás Nepusz,[4,8] Haixuan Yang,[4,8] Andrei L. Turinsky,[5] Zhihua Li,[6] Peggy I. Wang,[6] Daniel R. Boutz,[6] Vincent Fong,[1] Sadhna Phanse,[1] Mohan Babu,[1] Stephanie A. Craig,[6] Pingzhao Hu,[1] Cuihong Wan,[1] James Vlasblom,[2,5] Vaqaar-un-Nisa Dar,[7] Alexandr Bezginov,[7] Gregory W. Clark,[7] Gabriel C. Wu,[6] Shoshana J. Wodak,[2,3,5] Elisabeth R.M. Tillier,[7] Alberto Paccanaro,[4,*] Edward M. Marcotte,[6,*] and Andrew Emili[1,2,*]

[1]Banting and Best Department of Medical Research, Donnelly Centre for Cellular and Biomolecular Research

[2]Department of Molecular Genetics, Medical Sciences Building

[3]Department of Biochemistry, Medical Sciences Building

University of Toronto, Toronto, Ontario M5S 3E1, Canada

[4]Department of Computer Science, Royal Holloway, University of London, Egham TW20 0EX, UK

[5]Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada

[6]Center for Systems and Synthetic Biology, Institute for Cellular and Molecular Biology, Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, TX 78712, USA

[7]Campbell Family Institute for Cancer Research, Ontario Cancer Institute, University Health Network, University of Toronto, Toronto, Ontario M5G 1L7, Canada

[8]These authors contributed equally to this work

*Correspondence: alberto.paccanaro@cs.rhul.ac.uk (A.P.), marcotte@icmb.utexas.edu (E.M.M.), andrew.emili@utoronto.ca (A.E.)

http://dx.doi.org/10.1016/j.cell.2012.08.011

LETTER

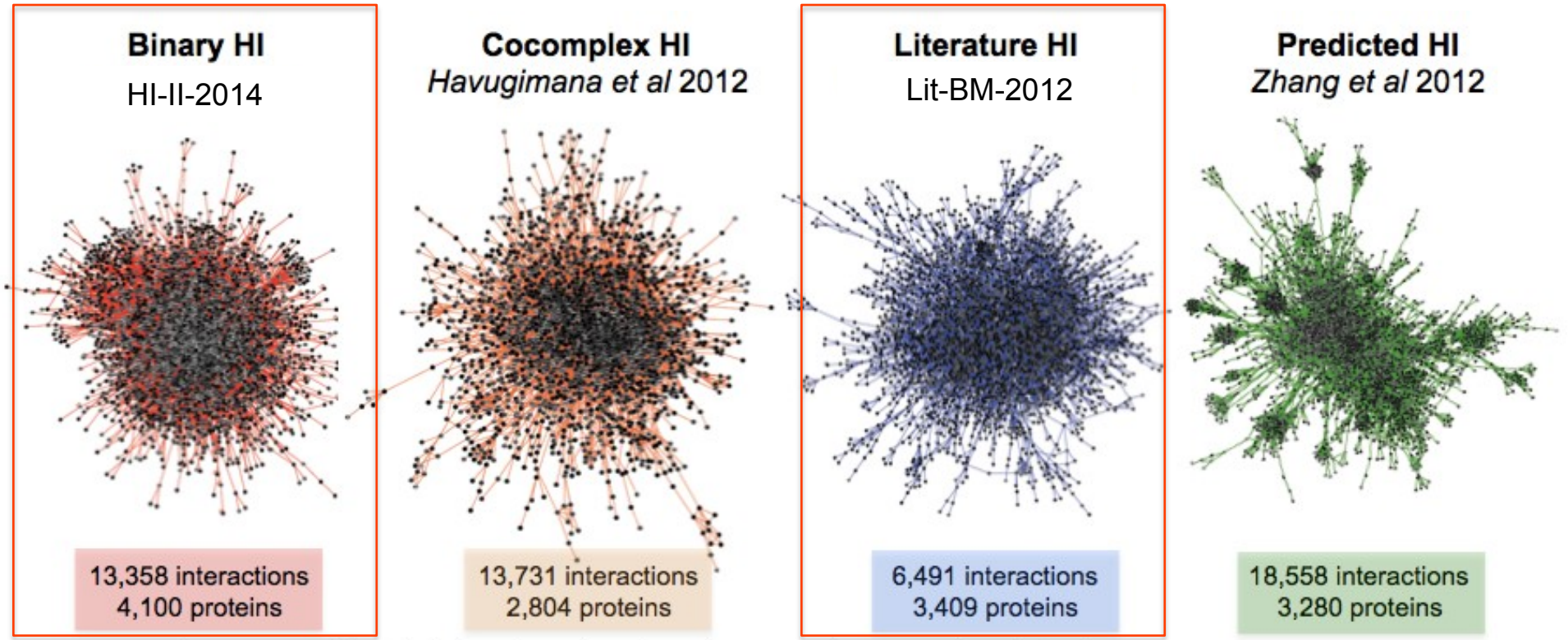
Nature 2012

doi:10.1038/nature11503

Structure-based prediction of protein–protein interactions on a genome-wide scale

Qiangfeng Cliff Zhang[1,2,3]*, Donald Petrey[1,2,3]*, Lei Deng[2,3,4], Li Qiang[5], Yu Shi[6], Chan Aye Thu[2], Brygida Bisikirska[3], Celine Lefebvre[3,7], Domenico Accili[5], Tony Hunter[6], Tom Maniatis[2], Andrea Califano[2,3,7,8] & Barry Honig[1,2,3]

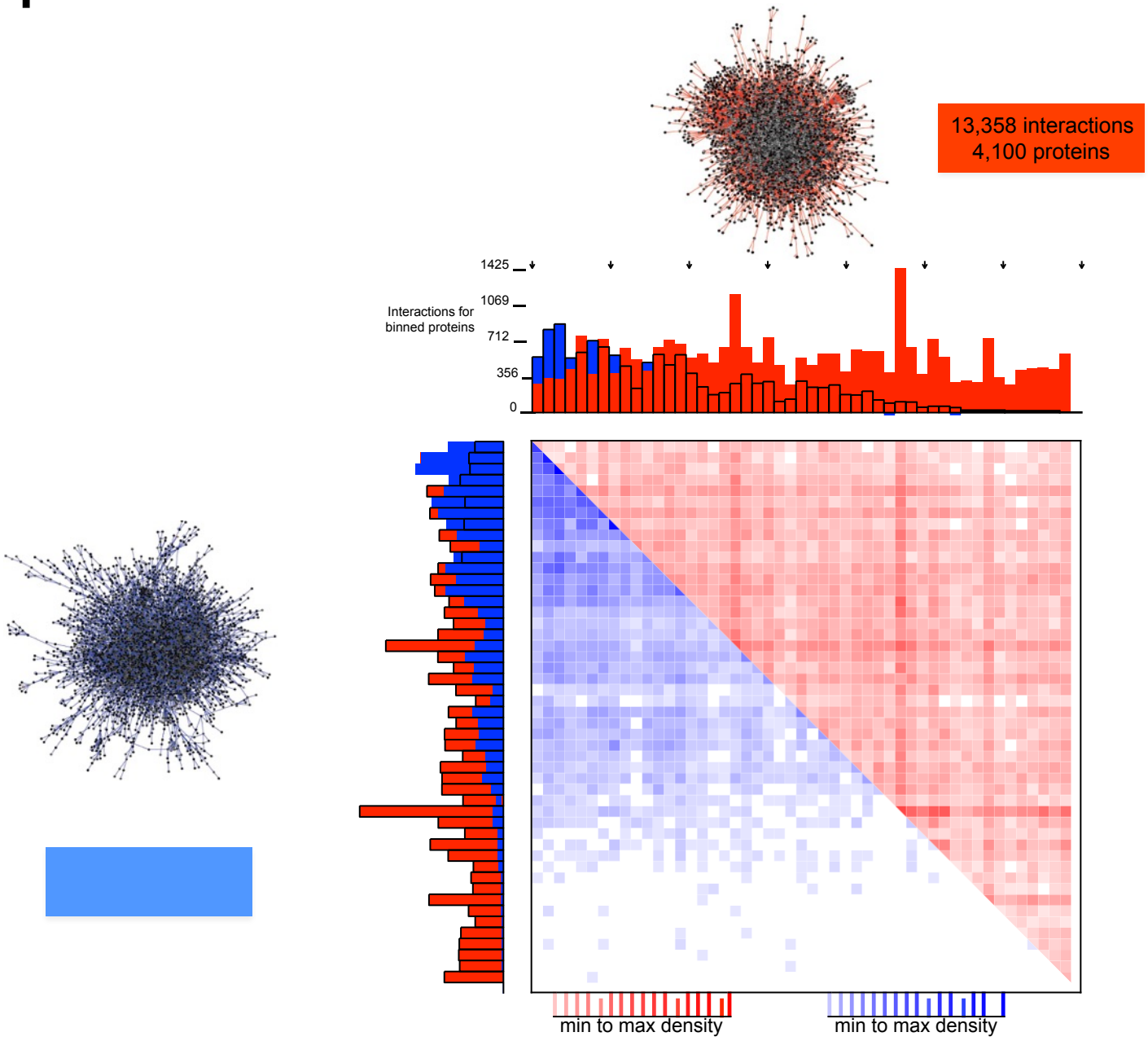
Comparison of 4 Human Interactomes



Network	N proteins in main comp	avg degree (total)	avg degree (neighbors)	avg betweenness	avg closeness	avg cluster coefficient
Binary HI	4,100	6.516	**58.238**	6280.4	**0.0075**	0.086
Cocomplex HI	2,804	9.794	21.606	4921.8	0.0055	**0.210**
Literature HI	3,409	3.808	12.800	8188.5	0.0019	0.204
Predicted HI	3,280	**11.316**	12.953	7444.8	0.0000002	**0.483**

Topological characteristics of the PPI networks (main component, only heterodimers considered)

Comparison of Human Interactomes

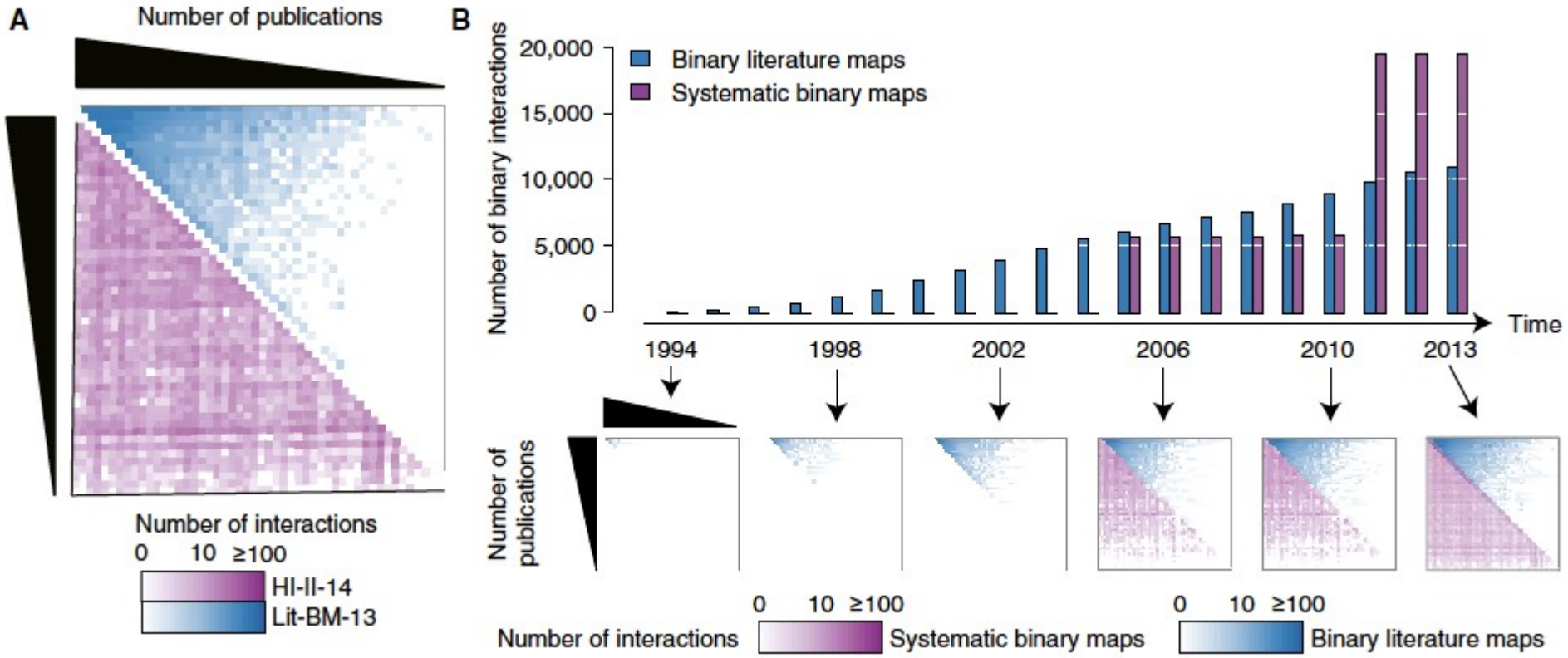


NOTE: excluding homodimers

Pairwise PPIs matrix comparing HIs (in Space III): ordered by **date of 1st publication (PubMed)**

Comparison of Human Interactomes

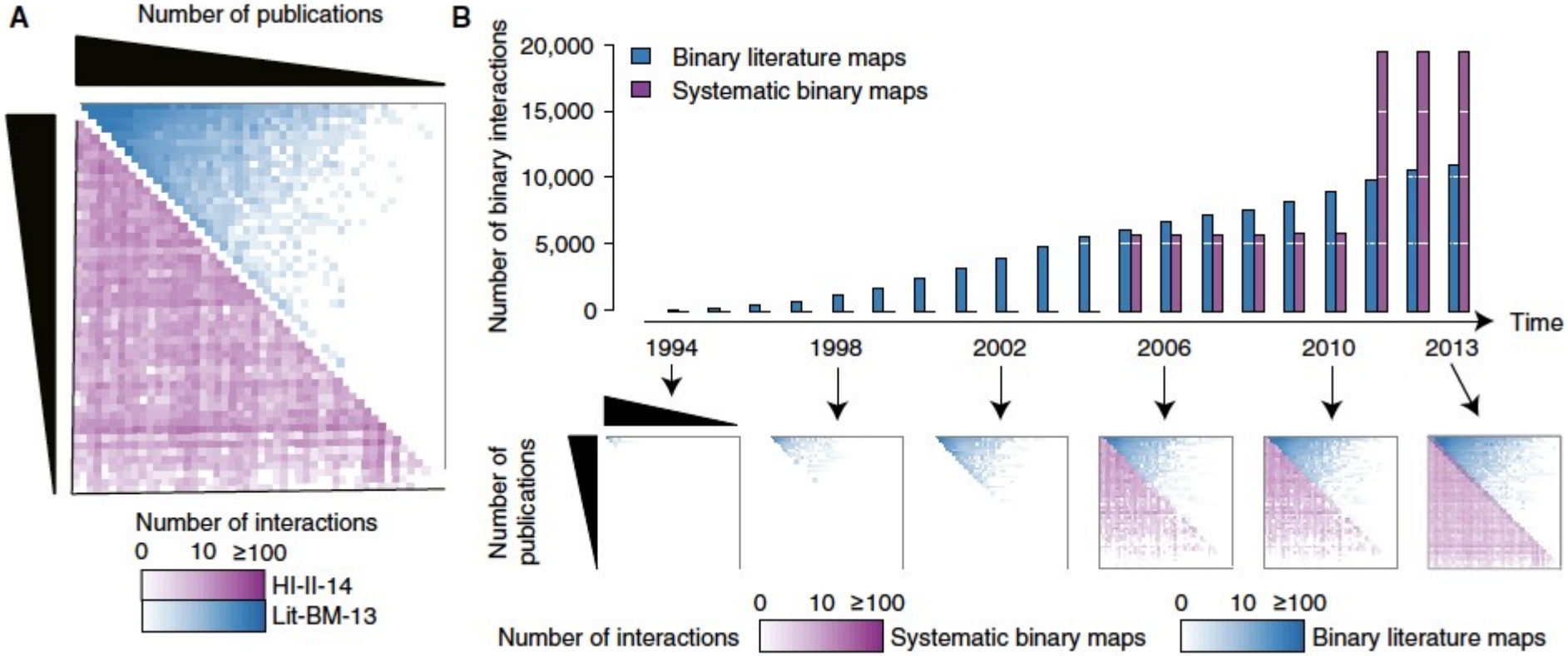
HI-II-2014 : a broader human interactome

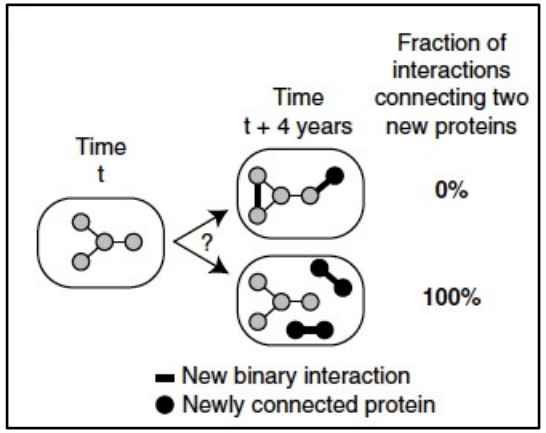


Pairwise PPIs matrix comparing HIs (in Space III): ordered by **number of publication (PubMed)**

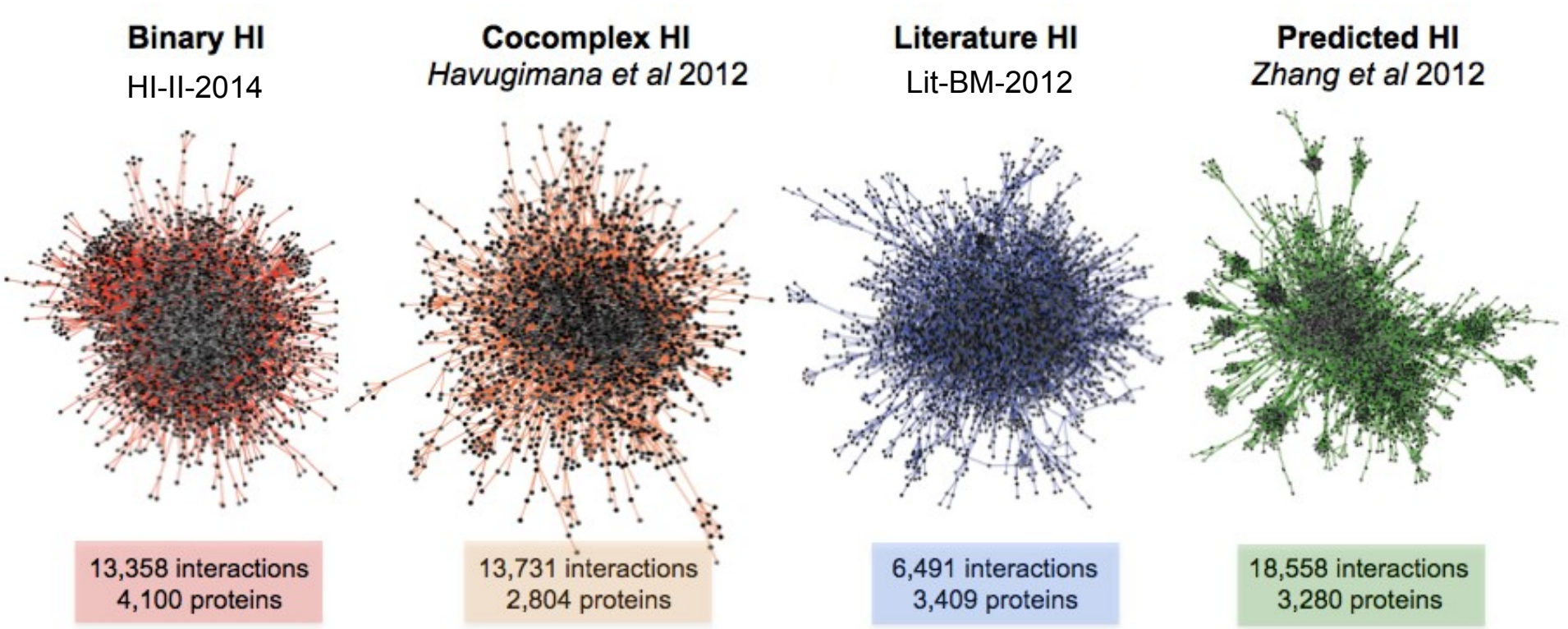
Comparison of Human Interactomes

HI-II-2014 : a broader human interactome





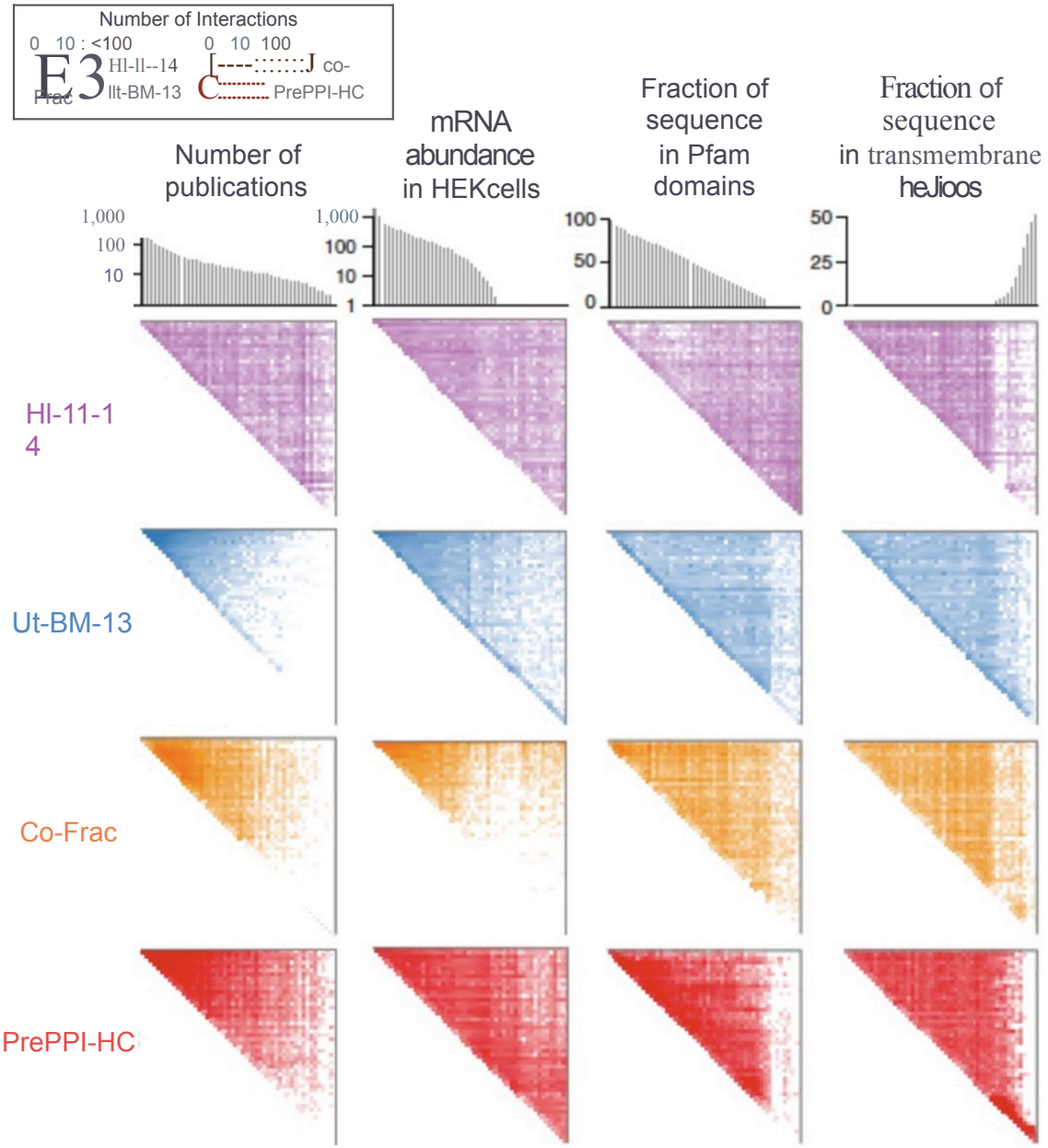
Comparison of 4 Human Interactomes



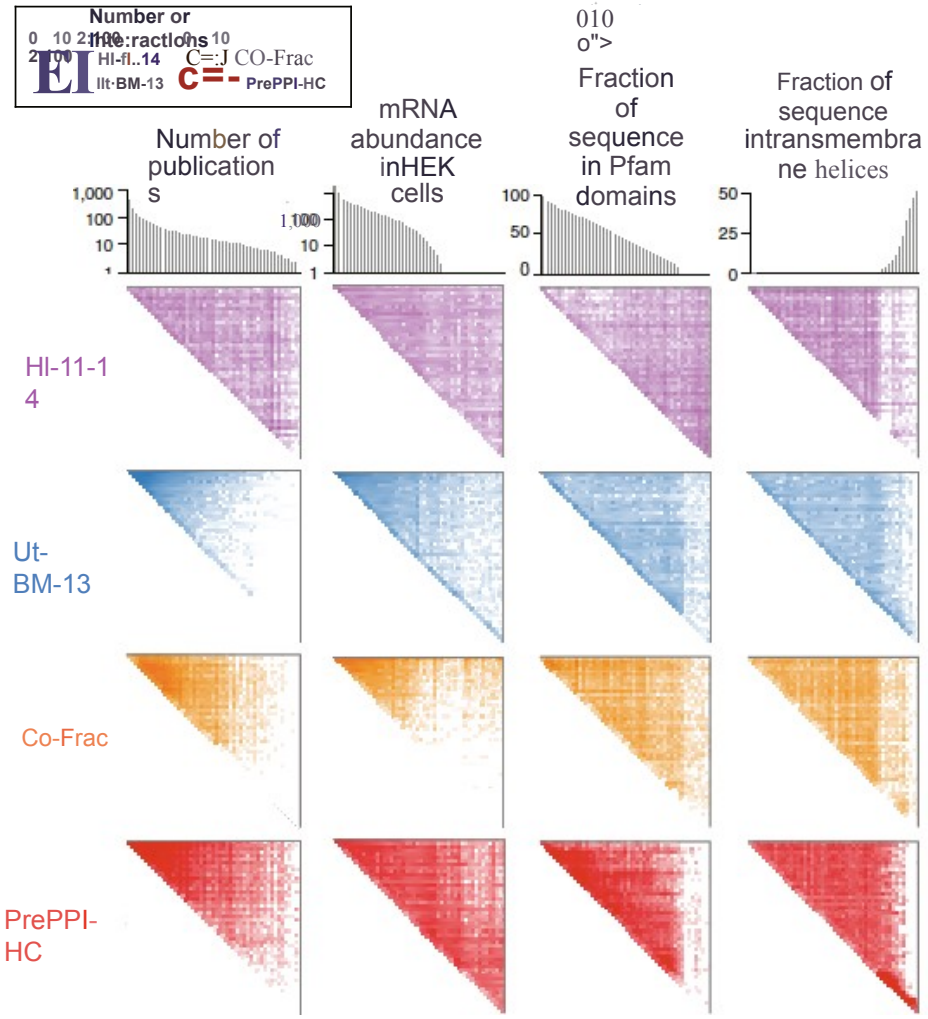
Network	N proteins in main comp	avg degree (total)	avg degree (neighbors)	avg betweenness	avg closeness	avg cluster coefficient
Binary HI	4,100	6.516	**58.238**	6280.4	**0.0075**	0.086
Cocomplex HI	2,804	9.794	21.606	4921.8	0.0055	**0.210**
Literature HI	3,409	3.808	12.800	8188.5	0.0019	0.204
Predicted HI	3,280	**11.316**	12.953	7444.8	0.0000002	**0.483**

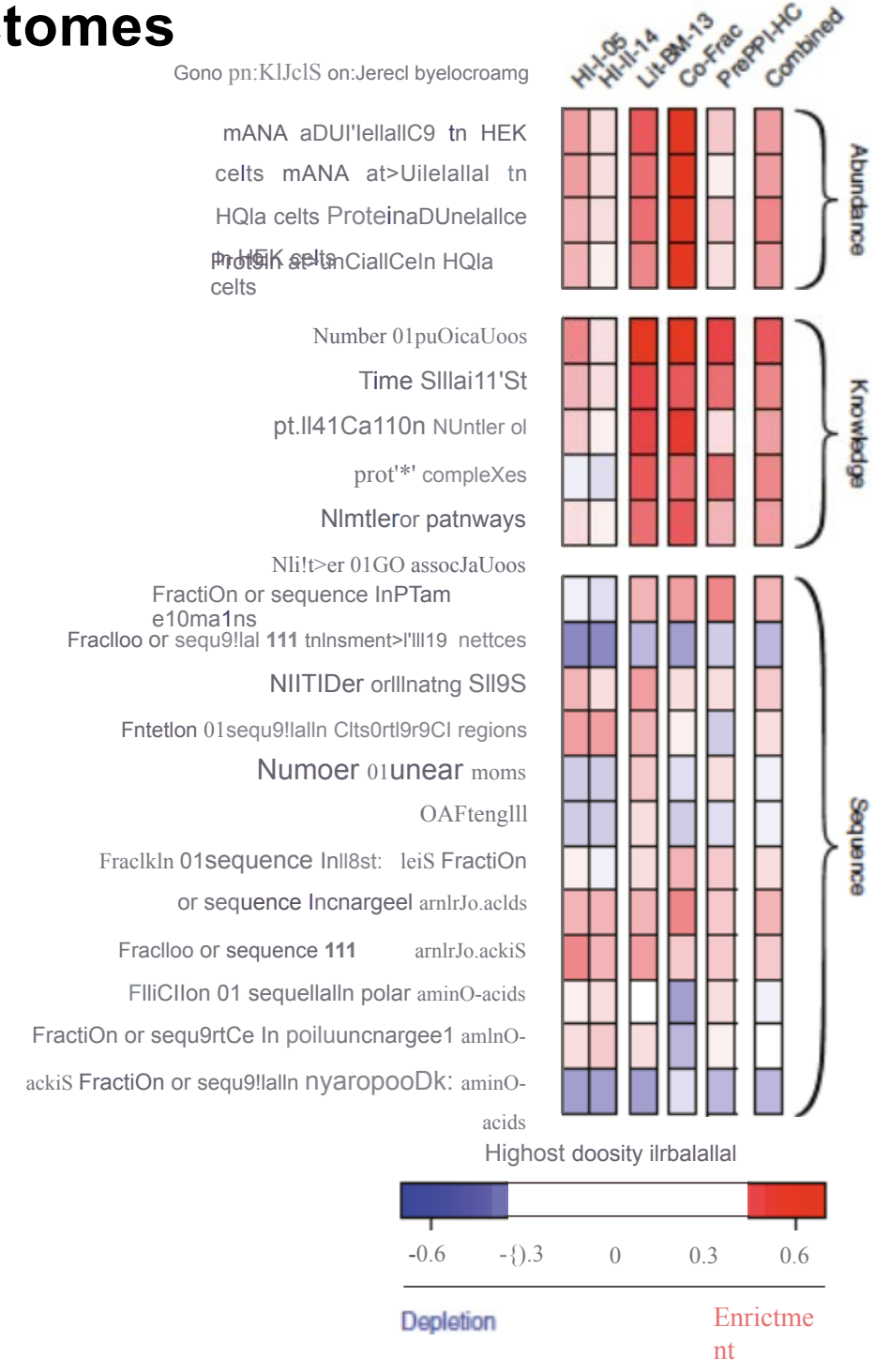
Topological characteristics of the PPI networks (main component, only heterodimers considered)

Comparison of 4 Human Interactomes



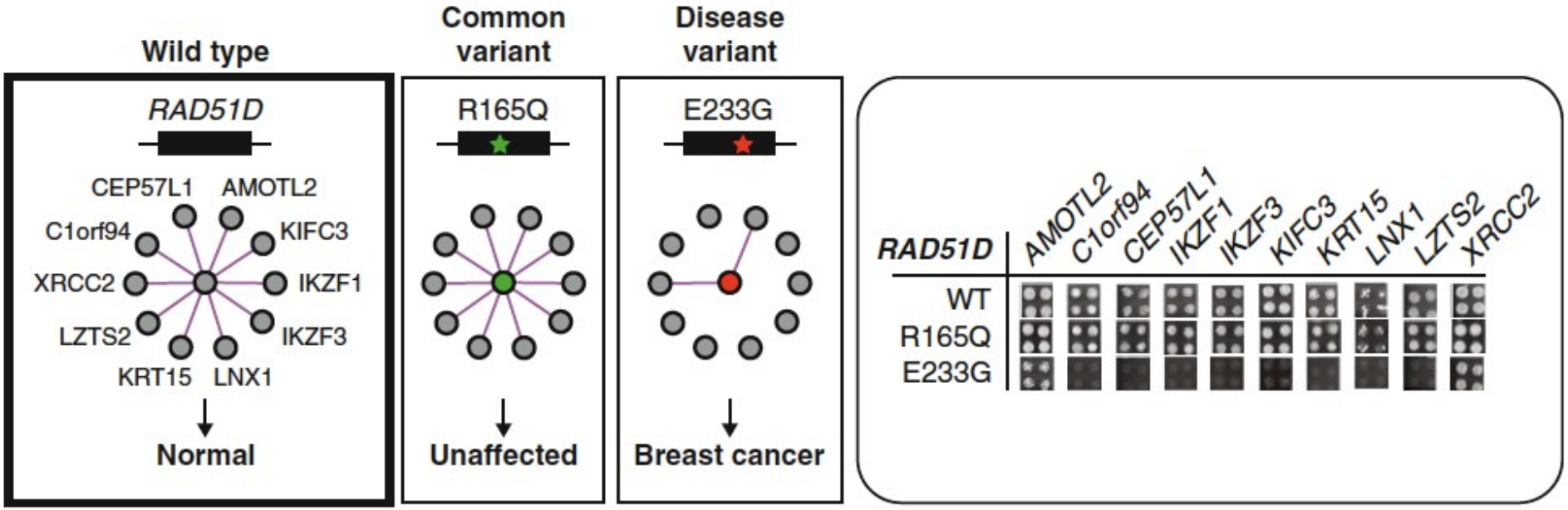
Comparison of 4 Human Interactomes





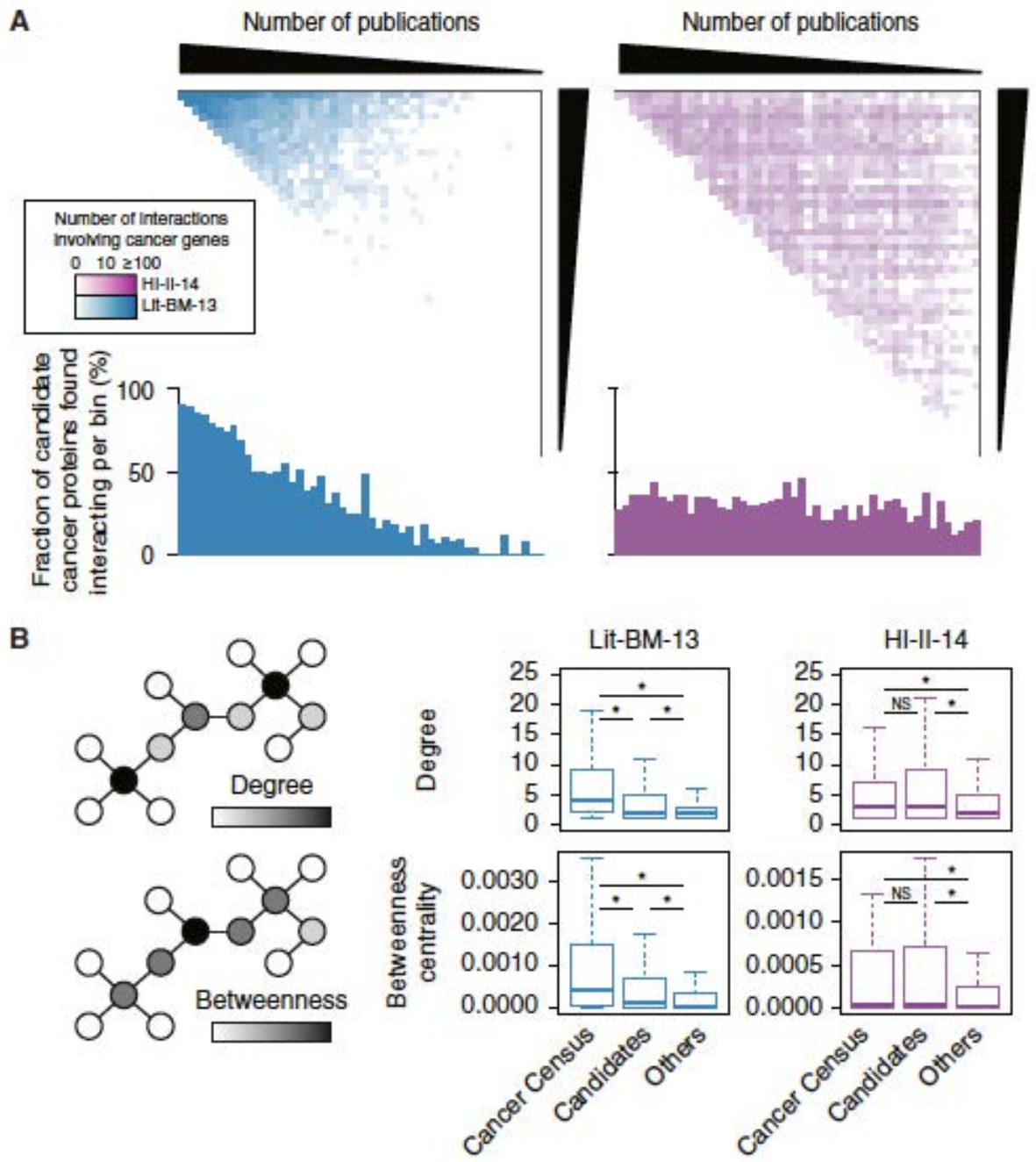
Alteration of the interactome in diseases

RAD51D protein
lost interactions in disease



Finding new disease genes in the interactome

Genes associated with the same disease are believed to be preferentially interconnected in interactome networks:
e.g. *cancer genes*



The Human Interactome

Two major large-scale data types: TAP-MS and Y2H

In recent years two main **high-throughput proteomic techniques** have been applied to determine PPIs:

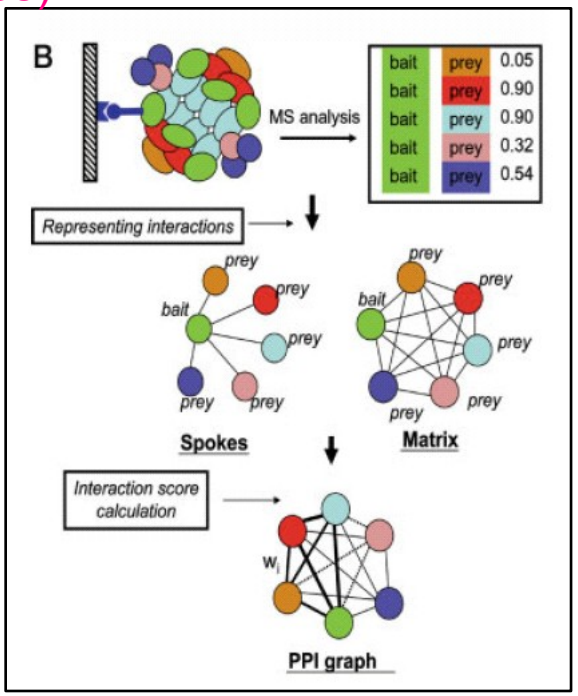
- **Tandem-Affinity Purification and Mass Spectrometry (TAP-MS)** provides multimer interactions (complexes)

- **High-throughput Two-Hybrid systems (Y2H)** provides binary interactions

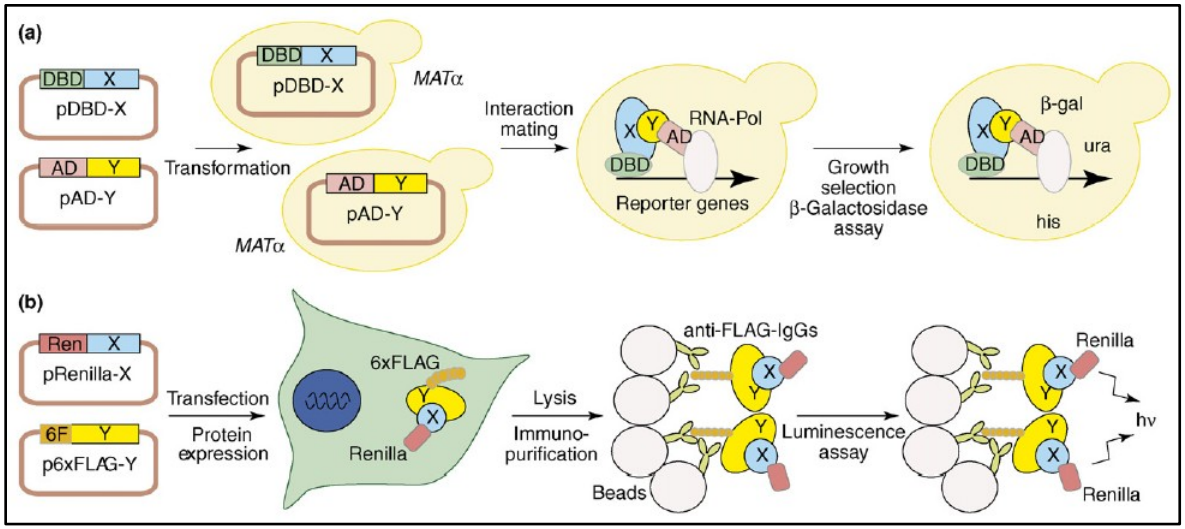
Protein-Protein Interactions (PPIs)
major high-throughput experimental methods

In recent years two main **high-throughput proteomic techniques** have been applied to determine PPIs:

– **Tandem-Affinity Purification and Mass Spectrometry (TAP-MS)** provides multimer interactions (complexes)



– **High-throughput Two-Hybrid systems (Y2H)** provides binary interactions



The BioPlex Network: A Systematic Exploration of the Human Interactome

Edward L. Huttlin,[1] Lily Ting,[1] Raphael J. Bruckner,[1] Fana Gebreab,[1] Melanie P. Gygi,[1] John Szpyt,[1] Stanley Tam,[1] Gabriela Zarraga,[1] Greg Colby,[1] Kurt Baltier,[1] Rui Dong,[2] Virginia Guarani,[1] Laura Pontano Vaites,[1] Alban Ordureau,[1] Ramin Rad,[1] Brian K. Erickson,[1] Martin Wühr,[1] Joel Chick,[1] Bo Zhai,[1] Deepak Kolippakkam,[1] Julian Mintseris,[1] Robert A. Obar,[1,3] Tim Harris,[3] Spyros Artavanis-Tsakonas,[1,3] Mathew E. Sowa,[1] Pietro De Camilli,[2] Joao A. Paulo,[1] J. Wade Harper,[1,*] and Steven P. Gygi[1,*]

[1]Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA

[2]Department of Cell Biology and Howard Hughes Medical Institute, Yale School of Medicine, New Haven, CT 06519, USA

[3]Biogen, Cambridge, MA 02142, USA

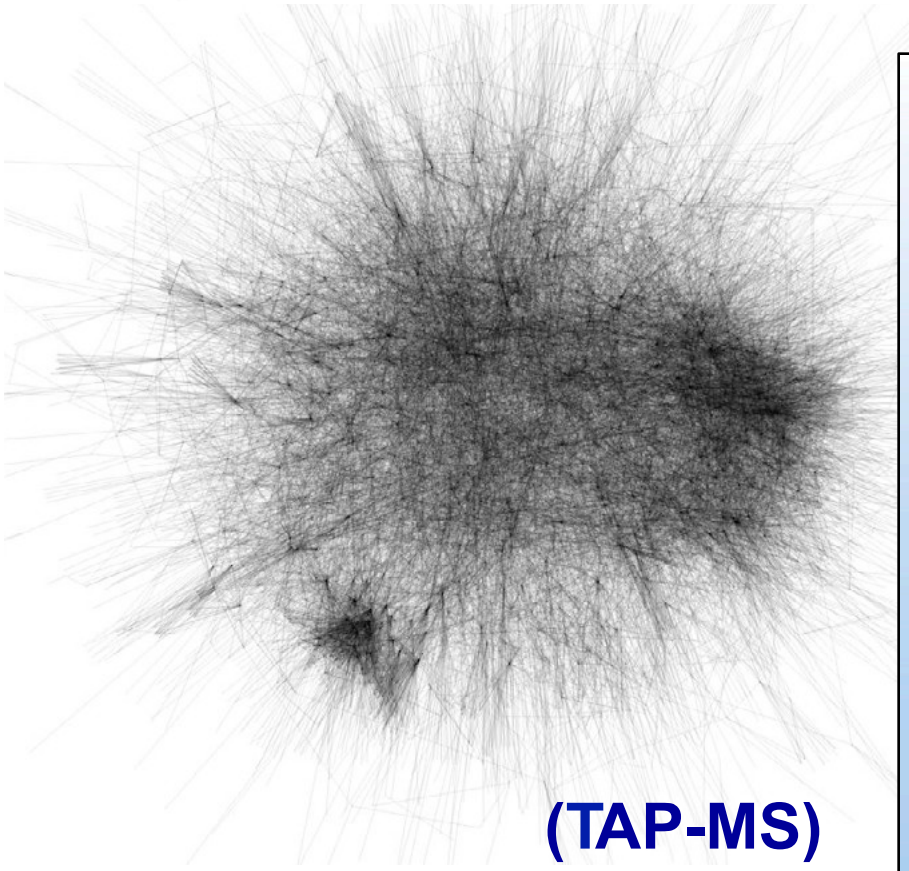
*Correspondence: wade_harper@hms.harvard.edu (J.W.H.), steven_gygi@hms.harvard.edu (S.P.G.)

http://dx.doi.org/10.1016/j.cell.2015.06.043

Cell

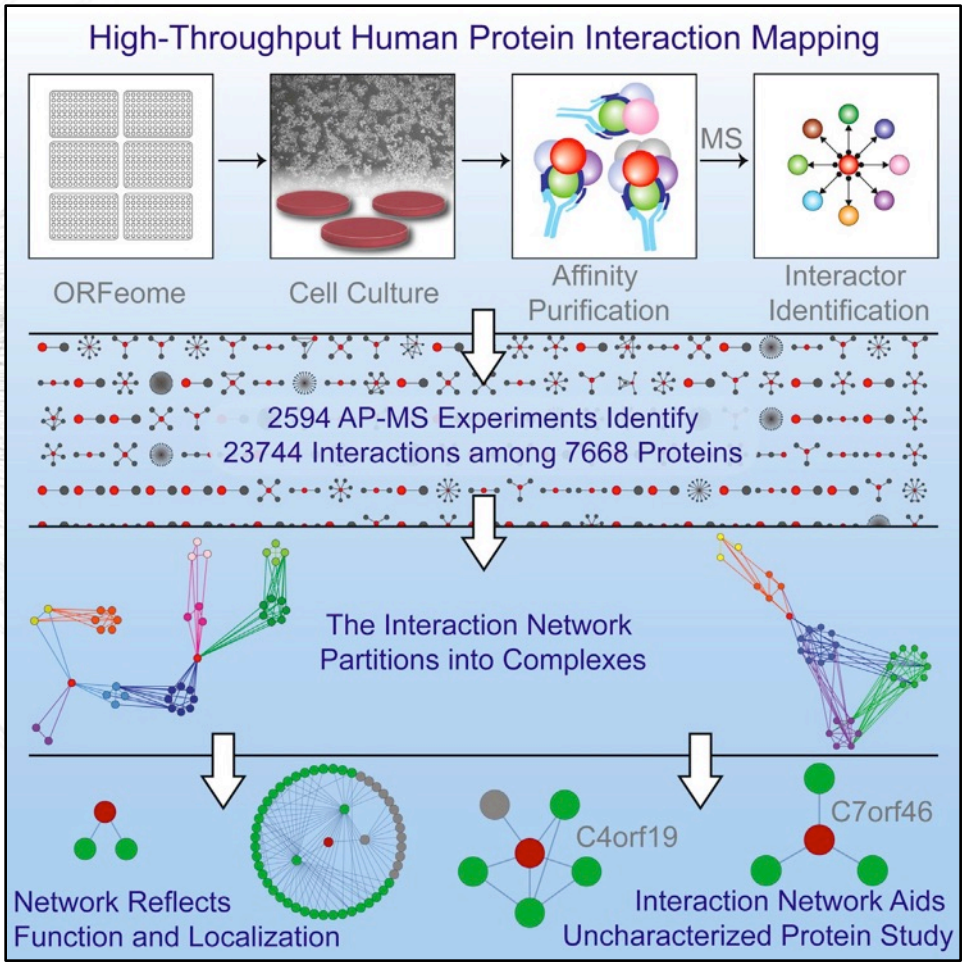
Volume 162, July 2015

The network: a systematic map of
*≈ **23,744** interactions between ≈ **7,668** human proteins*



(TAP-MS) co-complex

Huttlin et al. (2015) Cell



Protein Interactions PPIs

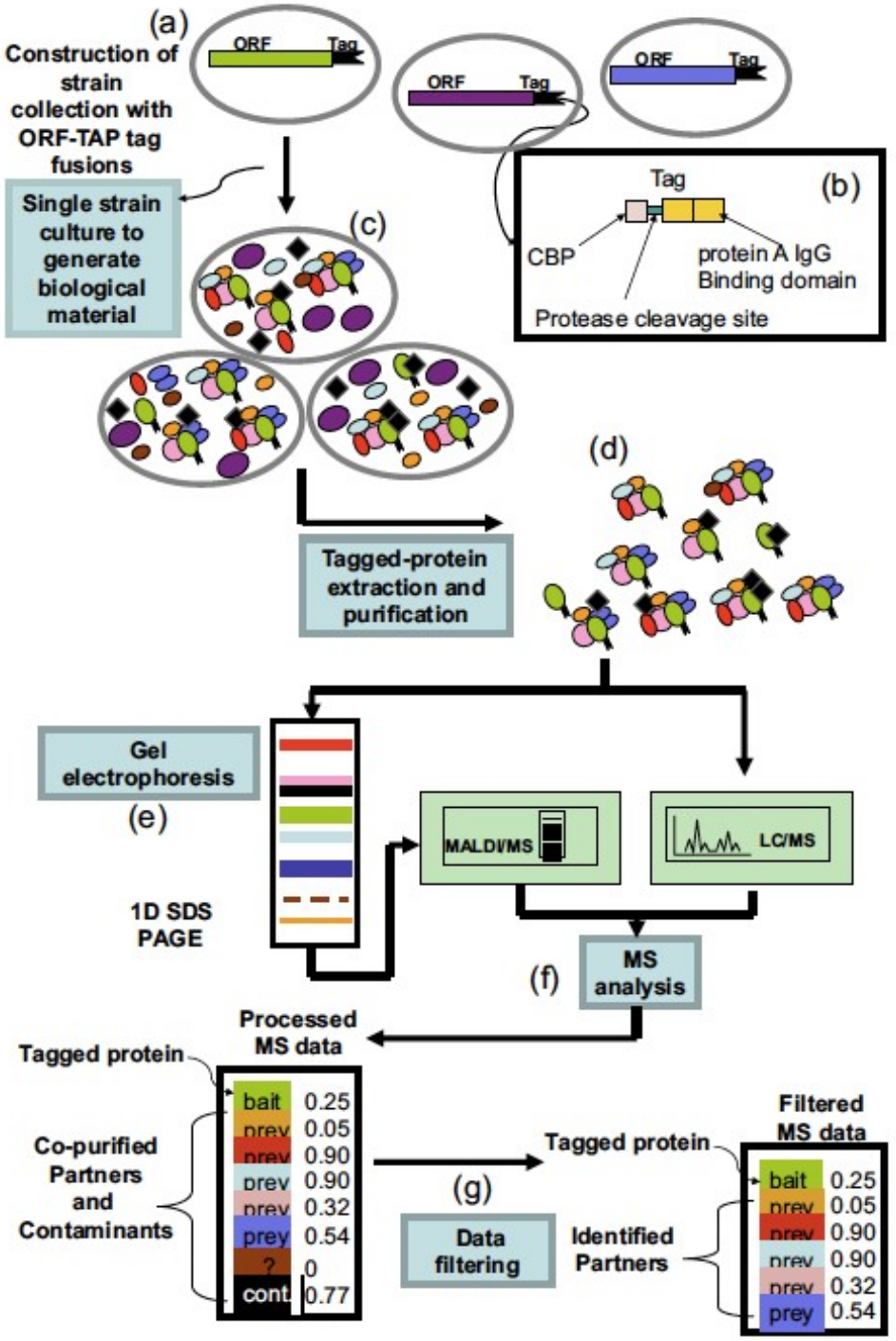
TAP-MS

Tandem-Affinity Purification and Mass Spectrometry (TAP-MS)
provides multimer interactions (complexes)

Bait and **Prey** system

The "bait proteins" are prepared with **tags** in order to fish the "prey proteins"

The **co-purified partners** are identified several times



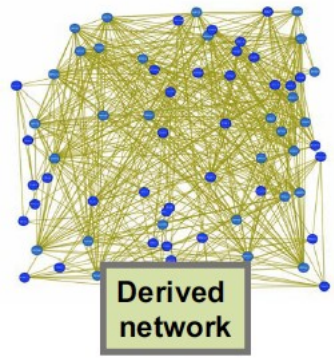
From ***Wodak et al. (2008) Mol Cel Proteomics***

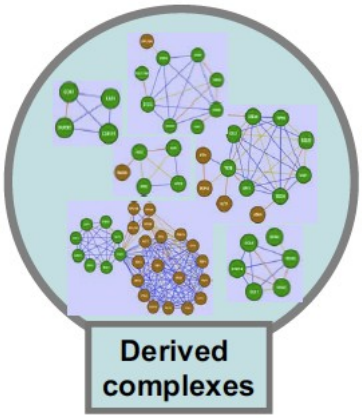
Protein Interactions PPIs

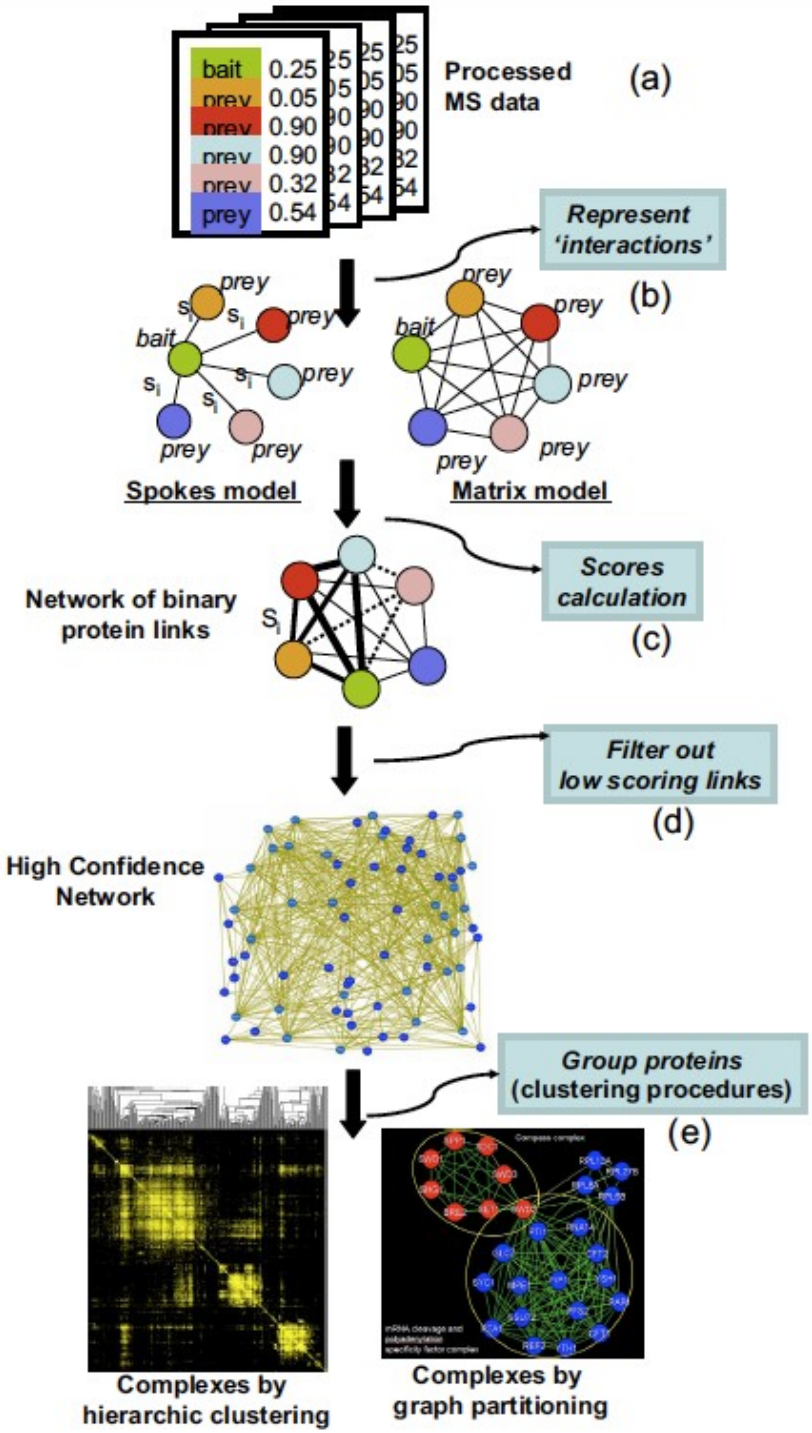
TAP-MS

Tandem-Affinity Purification and Mass Spectrometry (TAP-MS)
provides multimer interactions (complexes)

Once the tables of
co-purified partners
are produced the ***spokes* model**
is applied to estimate
the **binary interactions**







From ***Wodak et al. (2008) Mol Cel Proteomics***

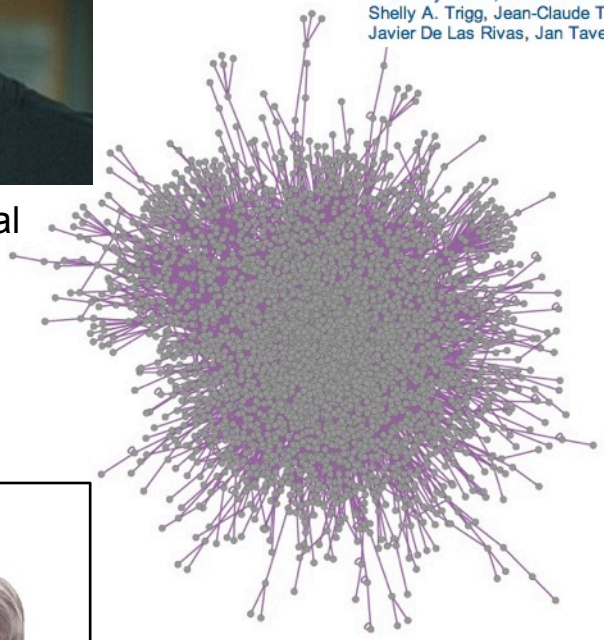
Dr. Marc Vidal
(Boston)

Cell

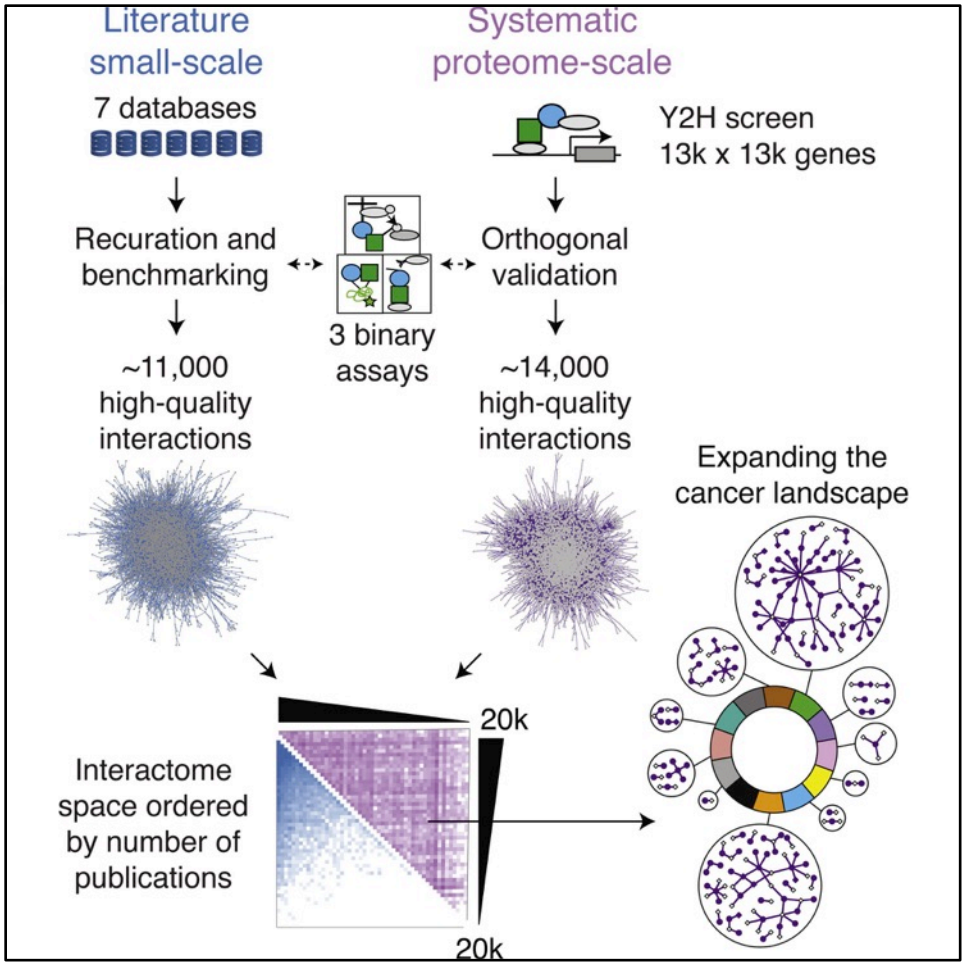
Volume 159, November 2014

A Proteome-Scale Map of the Human Interactome Network

Thomas Rolland[19], Murat Taşan[19], Benoit Charloteaux[19], Samuel J. Pevzner[19], Quan Zhong[19], Nidhi Sahni[19], Song Yi[19], Irma Lemmens, Celia Fontanillo, Roberto Mosca, Atanas Kamburov, Susan D. Ghiassian, Xinping Yang, Lila Ghamsari, Dawit Balcha, Bridget E. Begg, Pascal Braun, Marc Brehme, Martin P. Broly, Anne-Ruxandra Carvunis, Dan Convery-Zupan, Roser Corominas, Jasmin Coulombe-Huntington, Elizabeth Dann, Matija Dreze, Amélie Dricot, Changyu Fan, Eric Franzosa, Fana Gebreab, Bryan J. Gutierrez, Madeleine F. Hardy, Mike Jin, Shuli Kang, Ruth Kiros, Guan Ning Lin, Katja Luck, Andrew MacWilliams, Jörg Menche, Ryan R. Murray, Alexandre Palagi, Matthew M. Poulin, Xavier Rambout, John Rasla, Patrick Reichert, Viviana Romero, Elien Ruyssinck, Julie M. Sahalie, Annemarie Scholz, Akash A. Shah, Amitabh Sharma, Yun Shen, Kerstin Spirohn, Stanley Tam, Alexander O. Tejeda, Shelly A. Trigg, Jean-Claude Twizere, Kerwin Vega, Jennifer Walsh, Michael E. Cusick, Yu Xia, Albert-László Barabási, Lilia M. Iakoucheva, Patrick Aloy, Javier De Las Rivas, Jan Tavernier, Michael A. Calderwood[20], David E. Hill[20], Tong Hao[20], Frederick P. Roth✉, Marc Vidal✉



The network: a systematic map of
*≈ **14,000** interactions between ≈ **4,000** human proteins*



Dr. Javier De Las Rivas
(Salamanca)

(Y2H) binary

Rolland et al. (2014) Cell

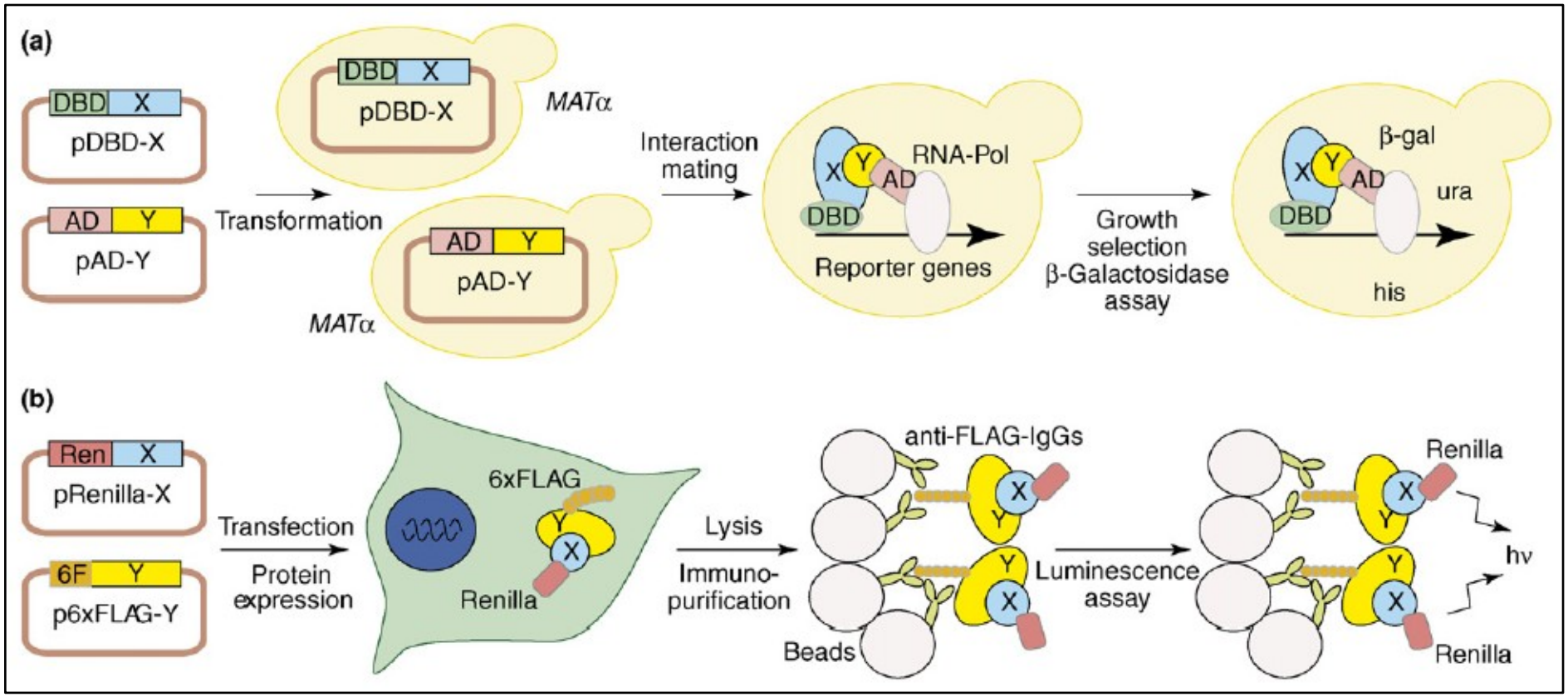
Protein Interactions PPIs

Y2H

High-throughput Two-Hybrid systems
provide binary interactions

(a) Y2H (yeast two hybrid) system, in **yeast** cells

(b) LUMIER system (luciferase), in **mammalian** cells



From ***Stelzl & Wanker (2006) Curr Opin Chem Biol***

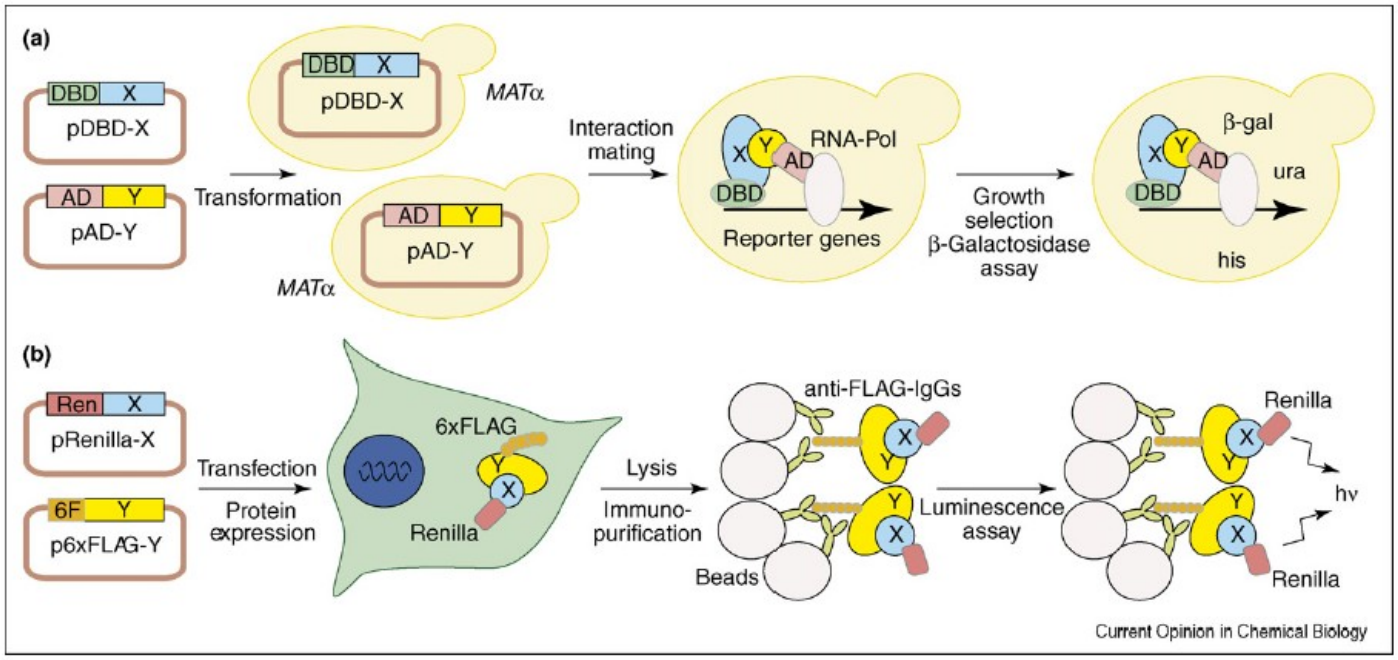
Protein Interactions PPIs

Y2H

High-throughput Two-Hybrid systems
provide binary interactions

Y2H classical system: Coding sequences for a protein X and a protein Y are fused to a DNA binding domain (DBD, i.e. bait plasmid) and a transcription activation domain (AD, i.e. prey plasmid). Upon interaction of protein X and protein Y, transcriptional activity of the DBD and AD domains is reconstituted leading to reporter gene activation.

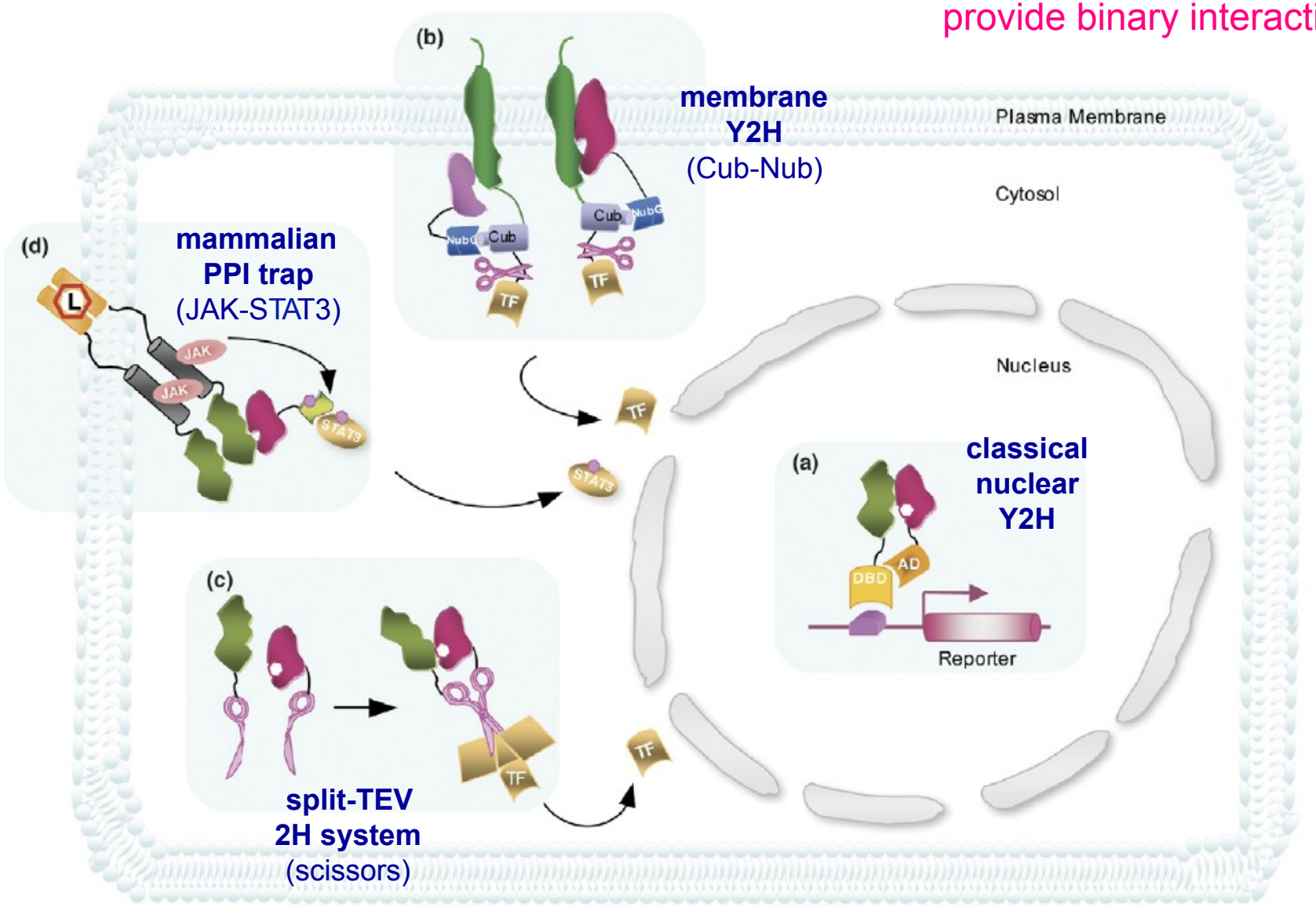
LUMIER system: Coding sequences for a protein X and a protein Y are fused to a 6xFLAG tag sequence and to renilla luciferase and cotransfected in mammalian cells. Upon interaction of protein X and protein Y, the luciferase fusion protein remains bound during the procedure and is detected via light emission.



Protein Interactions PPIs

Y2H

High-throughput Two-Hybrid systems
provide binary interactions



From ***Suter et al. (2008) Curr Opin Biotechnology***

Interactomes (global approaches)

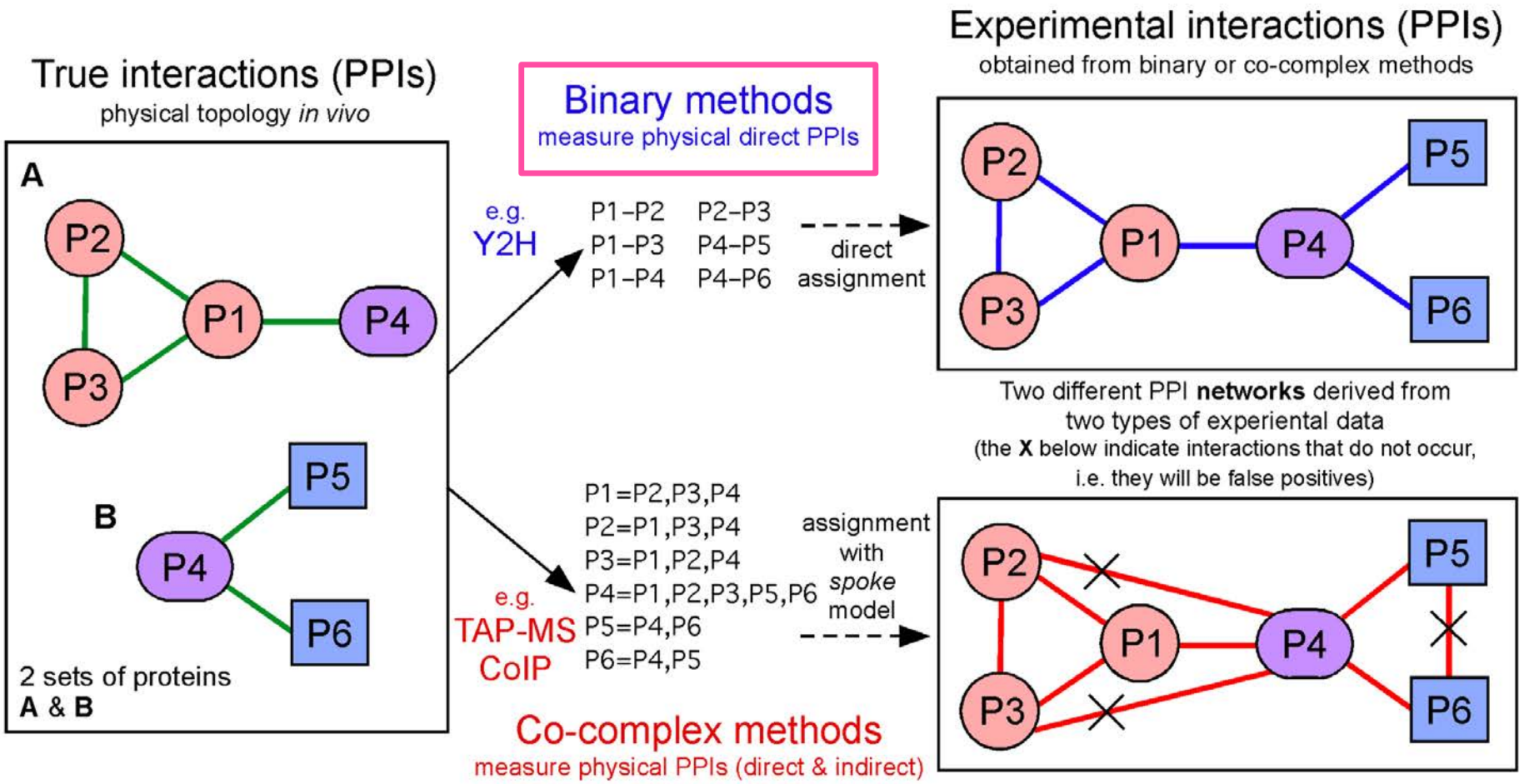
In recent years two main **high-throughput proteomic techniques** have been applied to determine PPIs:

- **Tandem-Affinity Purification and Mass Spectrometry (TAP-MS)** provides multimer interactions (complexes)

- **High-throughput Two-Hybrid systems (Y2H)** provides binary interactions

Protein-Protein Interactions (PPIs)

review some essential concepts on **PPIs**



Protein-Protein Interactions (PPIs)
experimental methods

Within the last years a large amount of data on protein-protein interactions in cellular systems has been obtained both by the **high-throughput** and **small scale technologies**. A list of most relevant **methods** to is presented:

Complex oriented methods (find *multimeric* PPIs)

- Co-Immunoprecipitation (Co-IP)
- Pull-Down Assays
- Tandem Affinity Purification + Mass Spectrometry (TAP-MS)

Binary oriented methods (find *dimeric* PPIs)

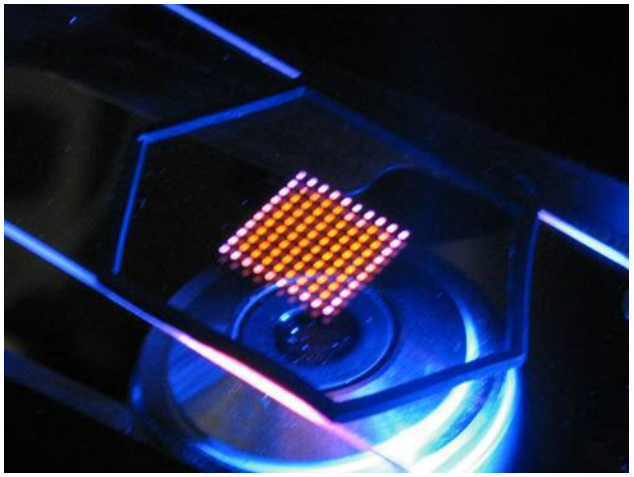
- Two Hybrid systems (Y2H)
- Protein Arrays / Protein Chips

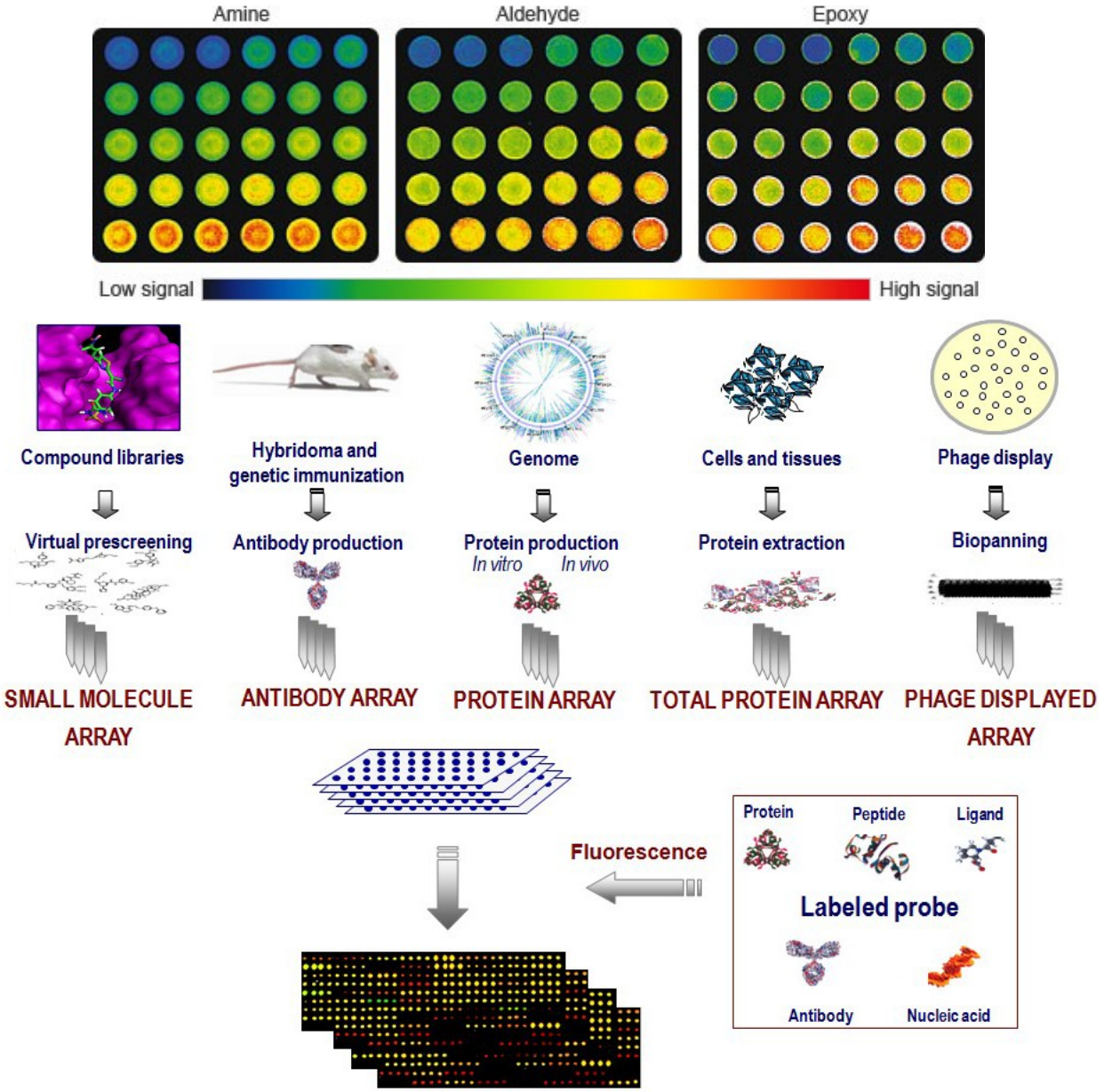
3D-structure based methods (find specific PPI interfaces)

- X-ray Crystallography (X-ray)
- Electro Microscopy (EM)
- Nuclear Magnetic Resonance (NMR)

Protein Interactions (PIs)

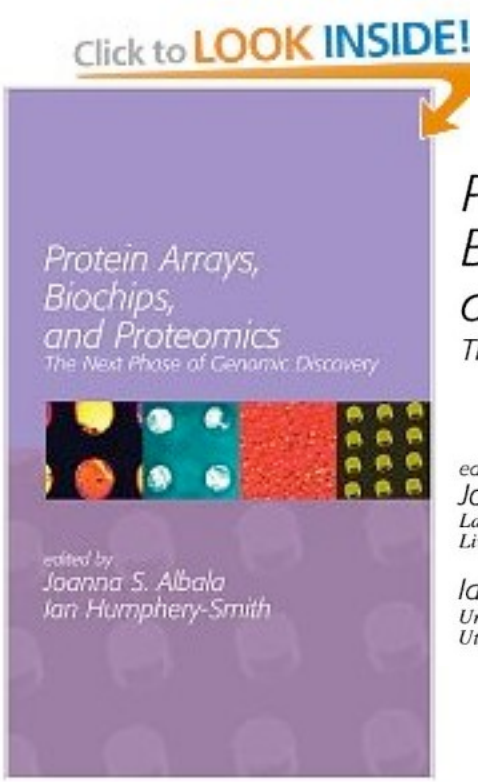
protein arrays/chips: multiple technologies to find **protein interactions**





Protein Interactions (PIs)

protein arrays/chips: multiple technologies to find **protein interactions**

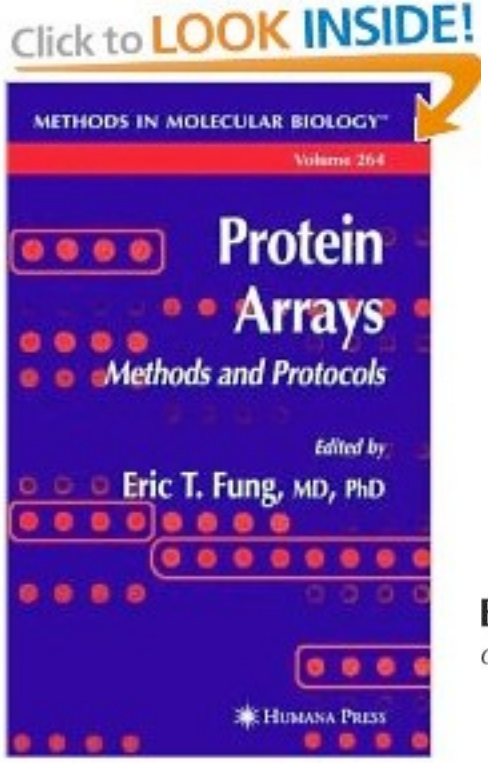


2005

Protein Arrays, Biochips, and Proteomics
The Next Phase of Genomic Discovery

edited by
Joanna S. Albala
Lawrence Livermore National Laboratory
Livermore, California, U.S.A.

Ian Humphery-Smith
University of Utrecht
Utrecht, The Netherlands



2004

Protein Arrays

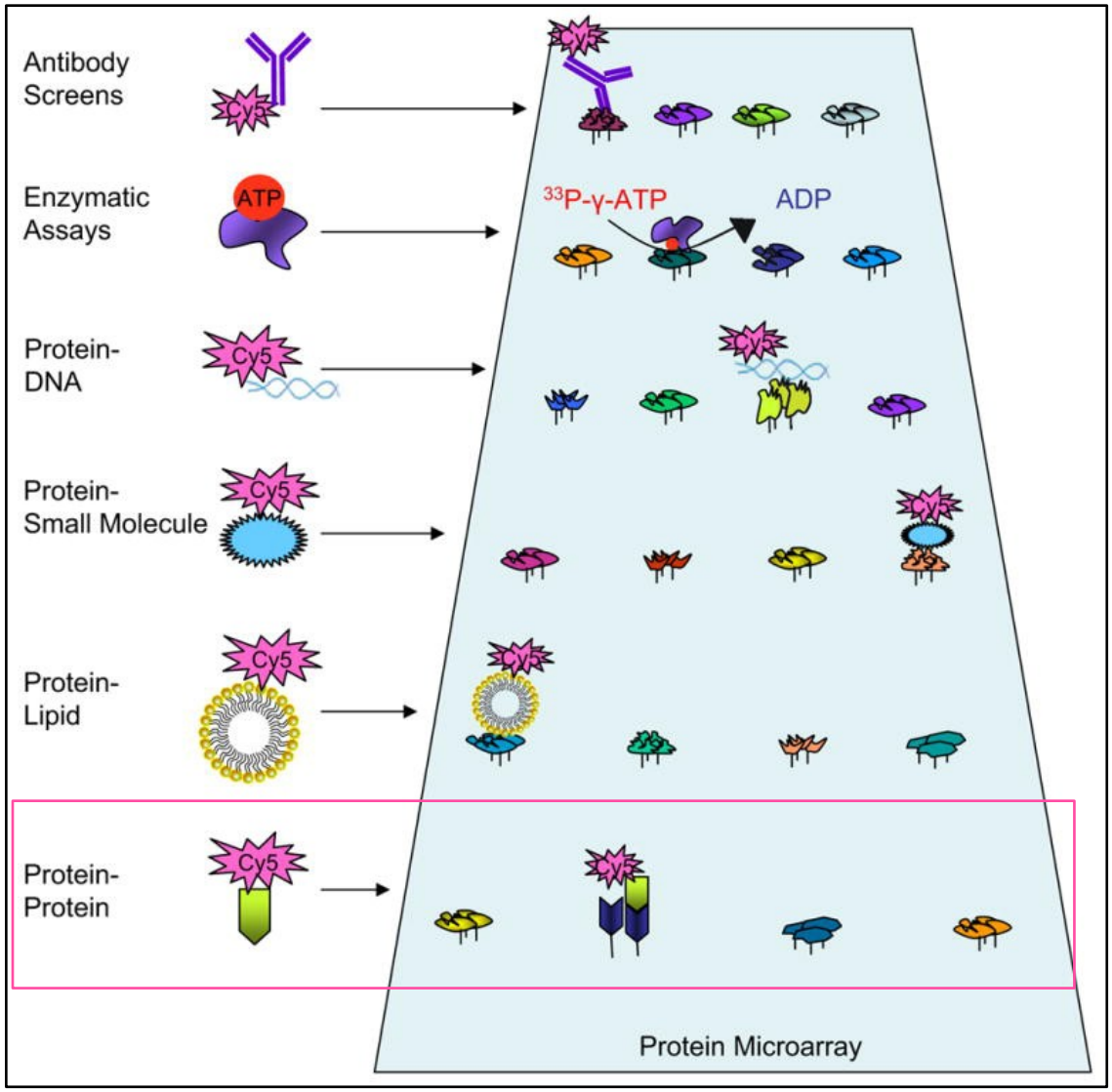
Methods and Protocols

Edited by
Eric T. Fung, MD, PhD
Ciphergen Biosystems Inc., Fremont, CA

Protein Interactions (PIs)

2007

protein arrays/chips: multiple technologies to find **protein interactions**



Multiple types of **protein arrays ≈ protein chips** designed to find different types of protein interactions:

– **protein - ligand** interactions
(ligands ≈ metabolites, drugs, chemicals, ...)

– **protein - antibody** interactions
(the protein is the antigen)

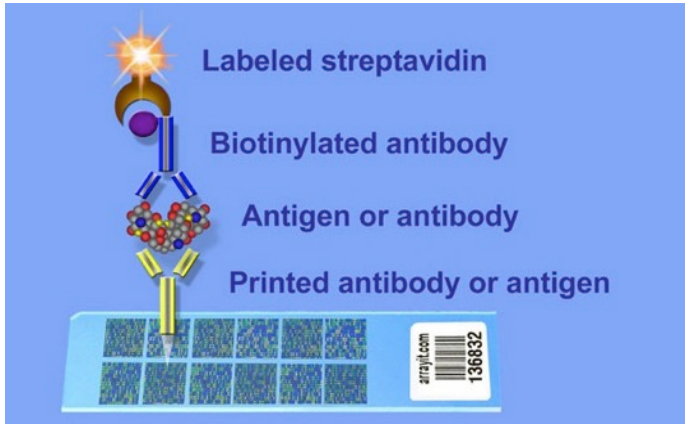
– **protein - DNA/RNA** interactions
(many proteins bind nucleic acids)

– **protein - protein** interactions
(many proteins have specific binding to other proteins in a stable or transient way)

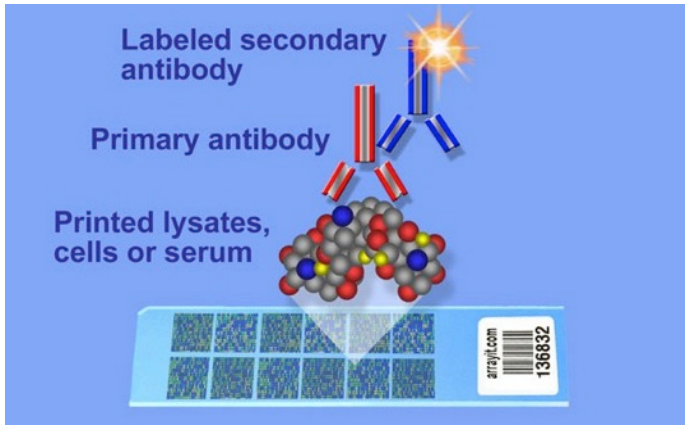
Hall, Ptacek & Snyder (2007)

Protein Interactions (PIs)

protein arrays: 1st data analysis step is the **signal** quantification

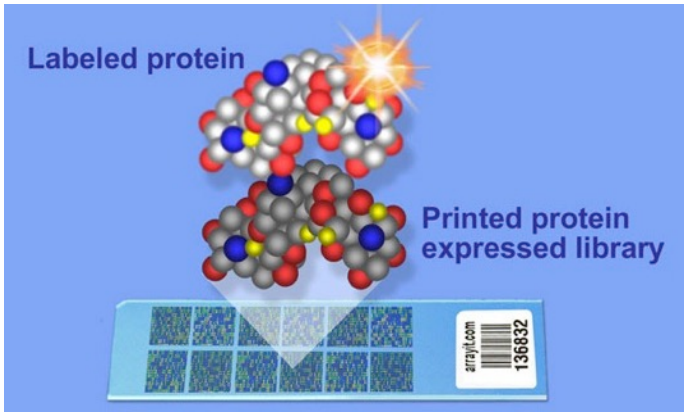


Antibody arrays

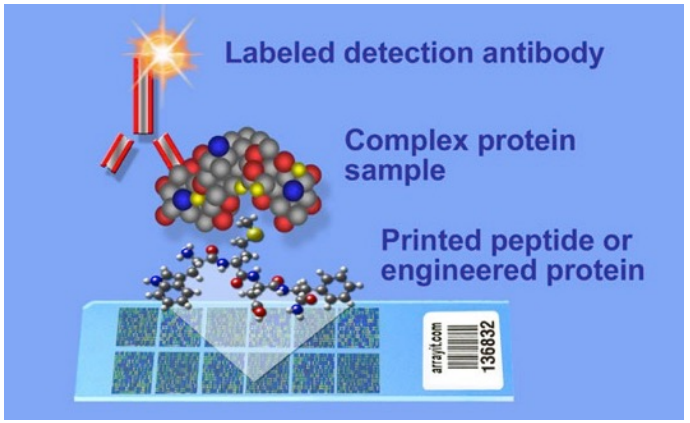


Reverse arrays

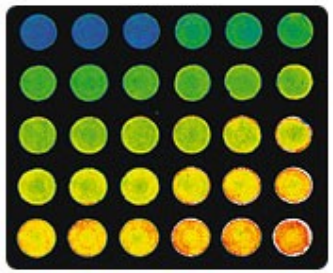
Signal quantification is a **technical problem** that has to be resolved by each platform with maximum **precission** and **accuracy**



Proteome arrays



Peptide arrays



Protein-Protein Interactions (PPIs)
proteins arrays to detect **protein-protein interations**

Within the last years a large amount of data on protein-protein interactions in cellular systems has been obtained both by the **high-throughput** and **small scale technologies**. A list of most relevant **methods** to is presented:

Complex oriented methods (find *multimeric* PPIs)

- Co-Immunoprecipitation (Co-IP)
- Pull-Down Assays
- Tandem Affinity Purification + Mass Spectrometry (TAP-MS)

Binary oriented methods (find *dimeric* PPIs)

- Two Hybrid systems (Y2H)
- Protein Arrays / Protein Chips

3D-structure based methods (find specific PPI interfaces)

- X-ray Crystallography (X-ray)
- Electro Microscopy (EM)
- Nuclear Magnetic Resonance (NMR)

Protein-Protein Interactions (PPIs)

data sources: databases

OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Education

Protein–Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks

Javier De Las Rivas*, Celia Fontanillo

Bioinformatics & Functional Genomics Research Group, Cancer Research Center (CiC-IBMCC, CSIC/USAL), Salamanca, Spain

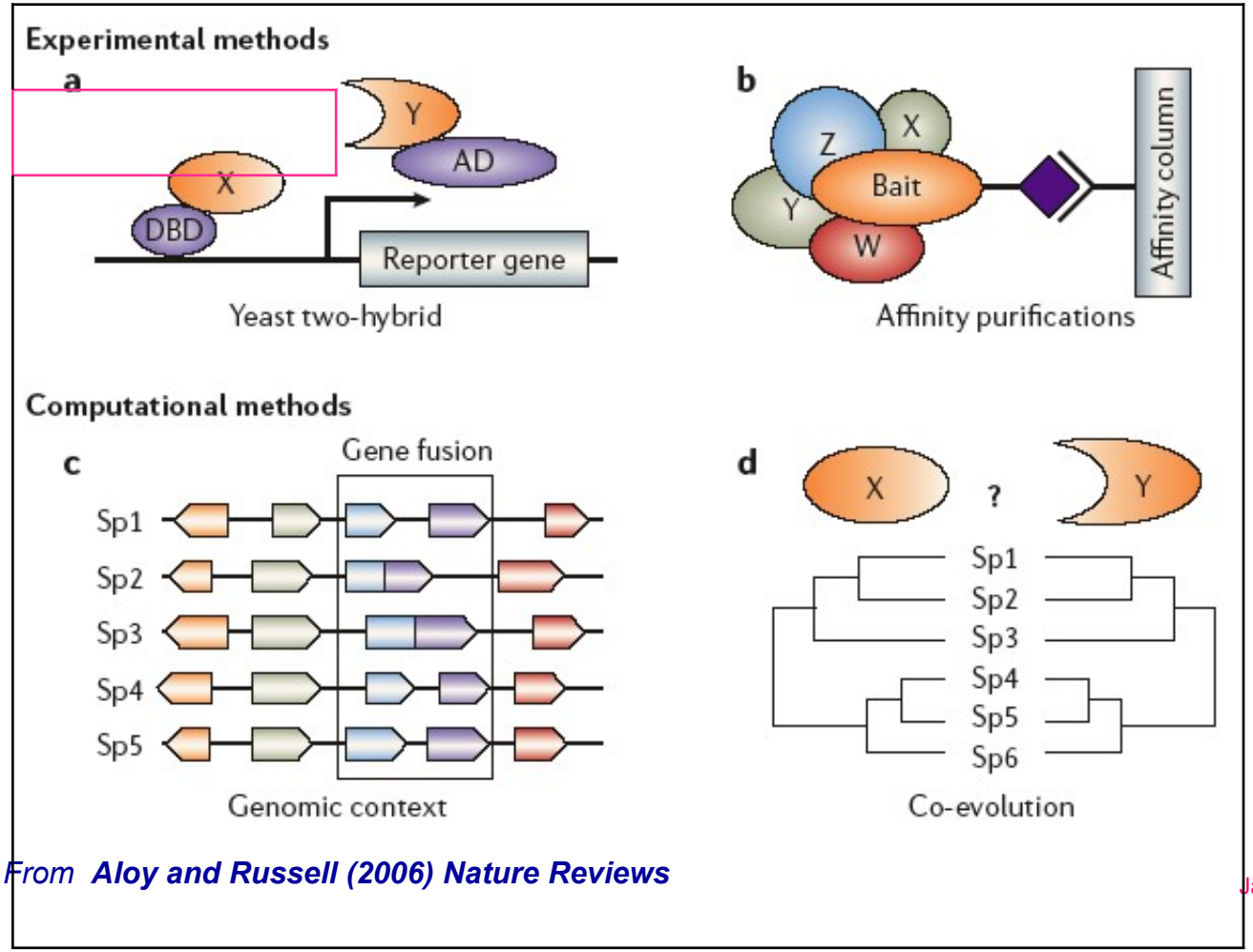
PLoS Comp. Bio. (2010)

Name	DB full name and type	PPIs sources	Type of MI	species	n prot. (Dec.2009)	n interact. (Dec.2009)
Primary Databases: PPI **experimental** data (curated from specific SSc & LSc published studies)						
BIND	Biomolecular Interaction Network Database	Ssc & Lsc published studies (literature-curated)	PPIs & others	all	[31972]	[58266]
BioGRID	General Repository for Interaction Datasets	Ssc & Lsc published studies (literature-curated)	PPIs & others	all	[28717]	[108691]
DIP	Database of Interacting Proteins	Ssc & Lsc published studies (literature-curated)	only PPIs	all	20728	57683
HPRD	Human Protein Reference Database	Ssc & Lsc published studies (literature-curated)	only PPIs	human	27081	38806
IntAct	Database of protein InterAction data	Ssc & Lsc published studies (literature-curated)	PPIs & others	all	[60504]	[202826]
MINT	Molecular INTeractions database	Ssc & Lsc published studies (literature-curated)	only PPIs	all	30089	83744
MIPS-MPact	MIPS protein interaction resource on yeast	derived from CYGD	only PPIs	yeast	1500	4300
MIPS-MPPI	MIPS mammalian protein-protein interaction db	Ssc published studies (literature-curated)	only PPIs	mammalia	982	937
Meta Databases: PPI **experimental** data (integrated and unified from different public repositories)						
APID	Agile Protein Interaction DataAnalyzer	BIND, BioGRID, DIP, HPRD, IntAct, MINT	only PPIs	all	56460	322579
MPIDB	The microbial protein interaction database	BIND, DIP, IntAct, MINT, other sets (exp & litcur)	only PPIs	microbial	7810	24295
PINA	Protein Interaction Network Analysis platform	BioGRID, DIP, HPRD, IntAct, MINT, MPact	only PPIs	all	[?]	188823
Prediction Databases: PPI **experimental & predicted** data ("functional interactions", i.e. interactions ***lato sensu*** derived from different types of data)						
MiMI	Michigan Molecular Interactions	BIND, BioGRID, DIP, HPRD, IntAct & nonPPI dt	PPIs & others	all	[45452]	[391386]
PIPs	Human protein-protein interactions prediction db	BIND, DIP, HPRD, OPHID & nonPPI dt	PPIs & others	human	[?]	[37606]
OPHID	Online Predicted Human Interaction Database	BIND, BioGRID, HPRD, IntAct, MINT, MPact & nonPPI dt	PPIs & others	human	[?]	[424066]
STRING	Known and Predicted Protein-Protein Interactions	BIND, BioGRID, DIP, HPRD, IntAct, MINT & nonPPI dt	PPIs & others	all	[2590259]	[88633860]
UniHI	Unified Human Interactome	BIND, BioGRID, DIP, HPRD, IntAct, MINT & nonPPI dt	PPIs & others	human	[22307]	[200473]

From ***De Las Rivas & Fontanillo (2010)***

Protein-Protein Interactions (PPIs)
experimental *vs* computational

For a proper study of protein-protein interactions it is very important to distinguish and separate the data that come from **experimental methods** (provided PPIs validated in the lab by some technique) & the data coming from **computational methods** (that provided PPIs infered but not really proved).



From ***Aloy and Russell (2006) Nature Reviews***

Many **databases** and **repositories** of **PPIs** include both **experimentally** and **computationally determined interactions** and this **mix** may produce confusion or false expectations in the analyses done on these combined data.

Protein-Protein Interactions (PPIs)
types of databases

There are several types of **PPIs** databases:

- **primary-db**
- **meta-db**
- **prediction-db**

Name	DB full name and type	PPIs sources	Type of MI	species	n prot.	n interact.
Primary Databases: PPI **experimental** data (curated from specific SSc & LSc published studies)					(Dec.2009)	(Dec.2009)
BIND	Biomolecular Interaction Network Database	Ssc & Lsc published studies (literature-curated)	PPIs & others	all	[31972]	[58266]
BioGRID	General Repository for Interaction Datasets	Ssc & Lsc published studies (literature-curated)	PPIs & others	all	[28717]	[108691]
DIP	Database of Interacting Proteins	Ssc & Lsc published studies (literature-curated)	only PPIs	all	20728	57683
HPRD	Human Protein Reference Database	Ssc & Lsc published studies (literature-curated)	only PPIs	human	27081	38806
IntAct	Database of protein InterAction data	Ssc & Lsc published studies (literature-curated)	PPIs & others	all	[60504]	[202826]
MINT	Molecular INTeractions database	Ssc & Lsc published studies (literature-curated)	only PPIs	all	30089	83744
MIPS-MPact	MIPS protein interaction resource on yeast	derived from CYGD	only PPIs	yeast	1500	4300
MIPS-MPPI	MIPS mammalian protein-protein interaction db	Ssc published studies (literature-curated)	only PPIs	mammalia	982	937
Meta Databases: PPI **experimental** data (integrated and unified from different public repositories)						
APID	Agile Protein Interaction DataAnalyzer	BIND, BioGRID, DIP, HPRD, IntAct, MINT	only PPIs	all	56460	322579
MPIDB	The microbial protein interaction database	BIND, DIP, IntAct, MINT, other sets (exp & litcur)	only PPIs	microbial	7810	24295
PINA	Protein Interaction Network Analysis platform	BioGRID, DIP, HPRD, IntAct, MINT, MPact	only PPIs	all	[?]	188823
Prediction Databases: PPI **experimental & predicted** data ("functional interactions", i.e. interactions ***lato sensu*** derived from different types of data)						
MiMI	Michigan Molecular Interactions	BIND, BioGRID, DIP, HPRD, IntAct & nonPPI dt	PPIs & others	all	[45452]	[391386]
PIPs	Human protein-protein interactions prediction db	BIND, DIP, HPRD, OPHID & nonPPI dt	PPIs & others	human	[?]	[37606]
OPHID	Online Predicted Human Interaction Database	BIND, BioGRID, HPRD, IntAct, MINT, MPact & nonPPI dt	PPIs & others	human	[?]	[424066]
STRING	Known and Predicted Protein-Protein Interactions	BIND, BioGRID, DIP, HPRD, IntAct, MINT & nonPPI dt	PPIs & others	all	[2590259]	[88633860]
UniHI	Unified Human Interactome	BIND, BioGRID, DIP, HPRD, IntAct, MINT & nonPPI dt	PPIs & others	human	[22307]	[200473]

From ***De Las Rivas & Fontanillo (2010)***

Protein-Protein Interactions (PPIs)
types of databases

There are several types of **PPIs** databases:

- **primary-db**
- **meta-db**
- **prediction-db**

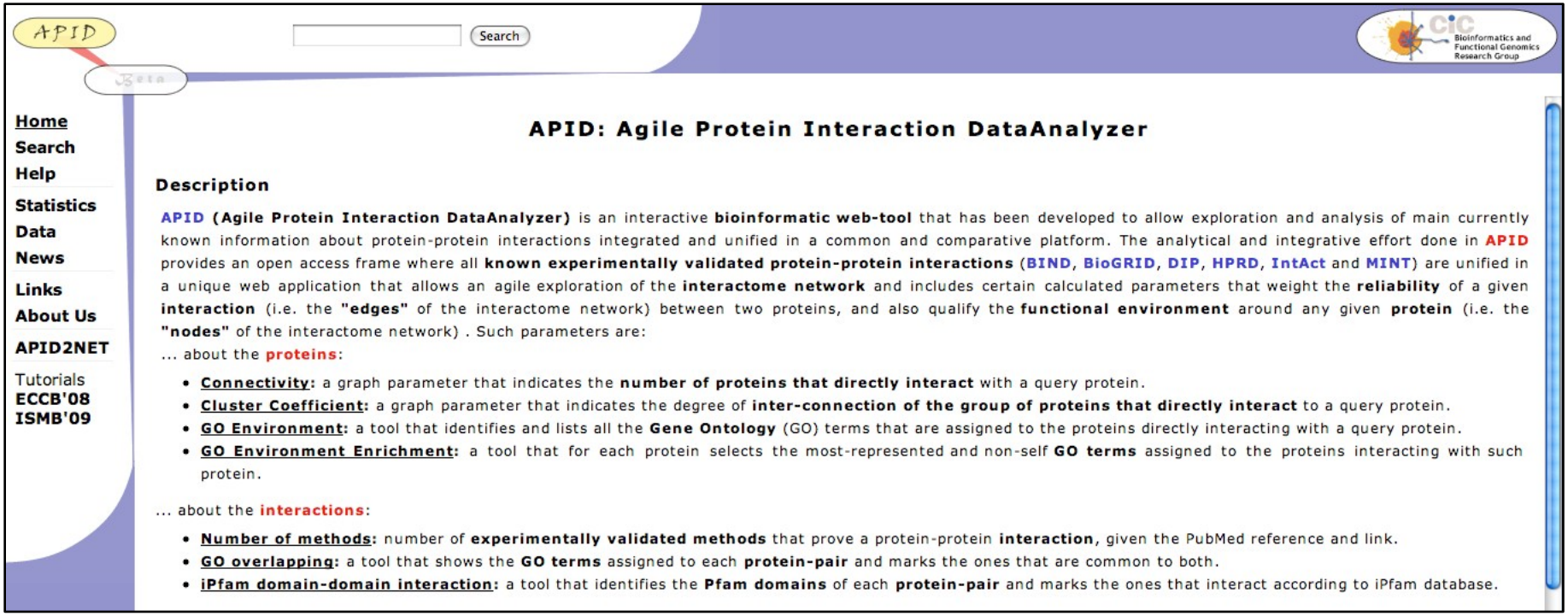
Name	DB full name and type	PPIs sources	Type of MI	species	n prot.	n interact.
Primary Databases: PPI **experimental** data (curated from specific SSc & LSc published studies)					(Dec.2009)	(Dec.2009)
BIND	Biomolecular Interaction Network Database	Ssc & Lsc published studies (literature-curated)	PPIs & others	all	[31972]	[58266]
BioGRID	General Repository for Interaction Datasets	Ssc & Lsc published studies (literature-curated)	PPIs & others	all	[28717]	[108691]
DIP	Database of Interacting Proteins	Ssc & Lsc published studies (literature-curated)	only PPIs	all	20728	57683
HPRD	Human Protein Reference Database	Ssc & Lsc published studies (literature-curated)	only PPIs	human	27081	38806
IntAct	Database of protein InterAction data	Ssc & Lsc published studies (literature-curated)	PPIs & others	all	[60504]	[202826]
MINT	Molecular INTeractions database	Ssc & Lsc published studies (literature-curated)	only PPIs	all	30089	83744
MIPS-MPact	MIPS protein interaction resource on yeast	derived from CYGD	only PPIs	yeast	1500	4300
MIPS-MPPI	MIPS mammalian protein-protein interaction db	Ssc published studies (literature-curated)	only PPIs	mammalia	982	937
Meta Databases: PPI **experimental** data (integrated and unified from different public repositories)						
APID	Agile Protein Interaction DataAnalyzer	BIND, BioGRID, DIP, HPRD, IntAct, MINT	only PPIs	all	56460	322579
MPIDB	The microbial protein interaction database	BIND, DIP, IntAct, MINT, other sets (exp & litcur)	only PPIs	microbial	7810	24295
PINA	Protein Interaction Network Analysis platform	BioGRID, DIP, HPRD, IntAct, MINT, MPact	only PPIs	all	[?]	188823
Prediction Databases: PPI **experimental & predicted** data ("functional interactions", i.e. interactions ***lato sensu*** derived from different types of data)						
MiMI	Michigan Molecular Interactions	BIND, BioGRID, DIP, HPRD, IntAct & nonPPI dt	PPIs & others	all	[45452]	[391386]
PIPs	Human protein-protein interactions prediction db	BIND, DIP, HPRD, OPHID & nonPPI dt	PPIs & others	human	[?]	[37606]
OPHID	Online Predicted Human Interaction Database	BIND, BioGRID, HPRD, IntAct, MINT, MPact & nonPPI dt	PPIs & others	human	[?]	[424066]
STRING	Known and Predicted Protein-Protein Interactions	BIND, BioGRID, DIP, HPRD, IntAct, MINT & nonPPI dt	PPIs & others	all	[2590259]	[88633860]
UniHI	Unified Human Interactome	BIND, BioGRID, DIP, HPRD, IntAct, MINT & nonPPI dt	PPIs & others	human	[22307]	[200473]

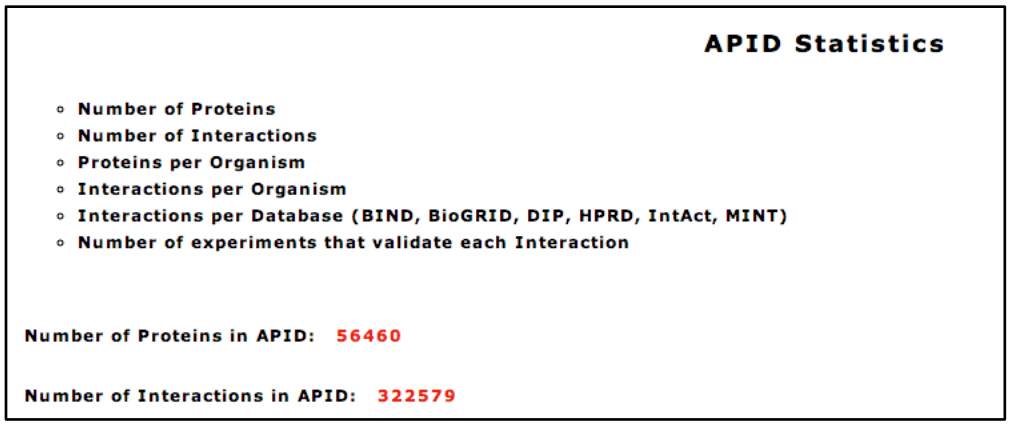
From ***De Las Rivas & Fontanillo (2010)***

From protein interactions to protein networks

integration & unification of **protein interaction data**

We have developed a database that integrates and unifies **PPIs**: **APID** & **APID2NET**





Experimental data unified in APID

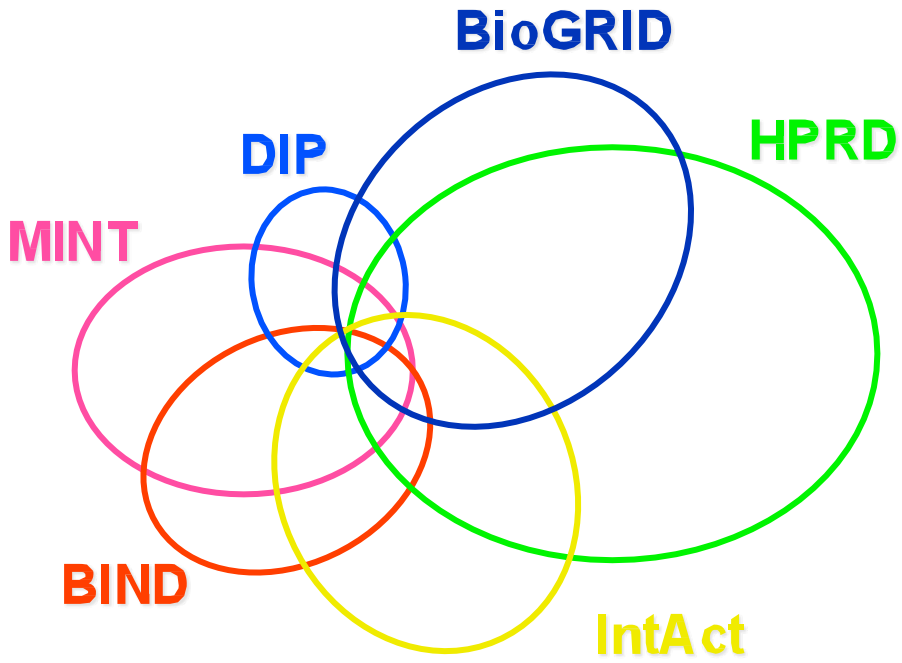
Number of Proteins in APID: 56460

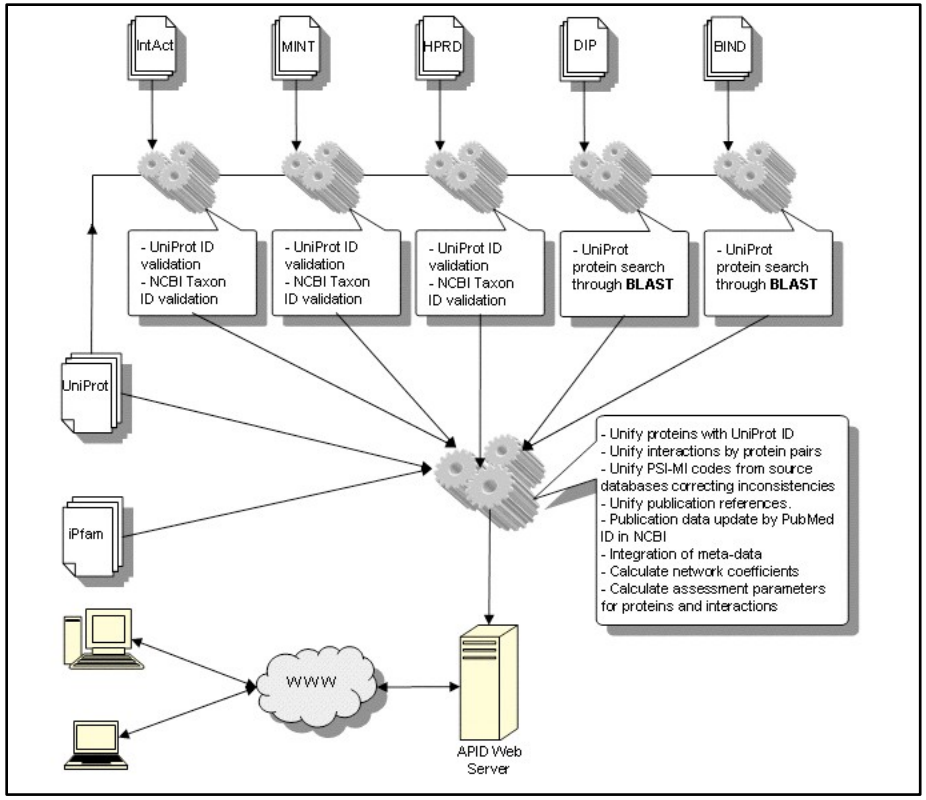
Number of Interactions in APID: 322579

Protein-Protein Interactions (PPIs)

integration & unification of **PPI data**

APID (Agile Protein Interaction DataAnalizer) http://bioinfow.dep.usal.es/apid





At present 6 source PPI DBs were unified:

- **BIND** (Biomolecular Interaction Network DB)
- **BioGRID** (Biological Gral. Repository for Interaction Datasets)
- **DIP** (Database of Interacting Proteins)
- **HPRD** (Human Protein Reference Database)
- **IntAct** (Database system & analysis tools for PI data)
- **MINT** (Molecular Interactions Database)

Data integration & unification by Sequence UniProt_ID PubMed_ID

Protein-Protein Interactions (PPIs)

integration & unification of **PPI data**

APID (Agile Protein Interaction DataAnalizer) http://bioinfow.dep.usal.es/apid

We are developing a new **APID** database that will integrate **PDB** and **sDDI (3D) data**

Number of PROTEINS (total): 63934
Number of INTERACTIONS (total): 333436

Number of Proteins in 3D interactions: 12784
Number of 3D Interactions: 15130

Number of PROTEINS in PPI sources: 56460
Number of INTERACTIONS in PPI sources: 322579

- Proteins per Organism in PPI sources
- Interactions per Organism in PPI sources
- Interactions per Database (BIND, BioGRID, DIP, HPRD, IntAct, MINT)
- Number of experiments that validate each Interaction

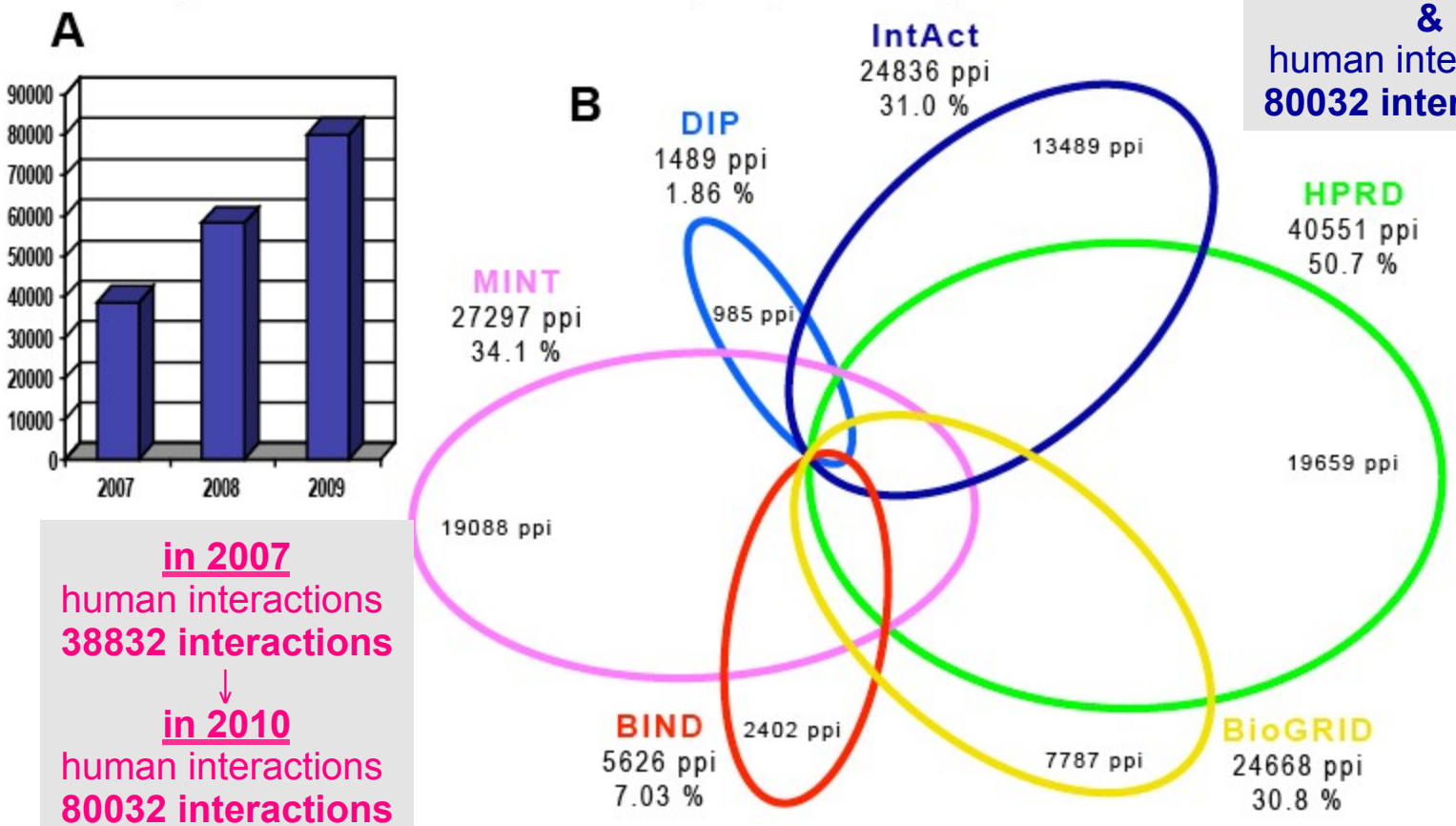
Protein-Protein Interactions (PPIs)

integration & unification of PPI data: **hsPPIs** in **APID**

There are several primary **PPIs** databases, but at present there is **small integration**:

Human Interactome

Coverage of human PPIs on major public repositories



in 2007
human interactions
38832 interactions
↓
in 2010
human interactions
80032 interactions

in 2010
human proteins
11998 proteins
&
human interactions
80032 interactions

CBO Computational Biology @ UCT

CiC

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Hands-on: Practical Examples

Build ppi networks in Cytoscape (plugins APID2NET and PSICQUIC)

Protein_SETs_2015.xls (PreRIBOSOME, Proteasome, NOTCH)

From protein interactions to protein networks

build reliable networks with biological meaning: examples

Challenge: obtain and integrate ***omic* data** to build **biological networks** and solve **biological questions**.

Three examples based in PPI data:

1.– Use of PPI data to build **protein networks** and find different **sub-complexes** and **assembly steps:** the PRE-RIBOSOME example.

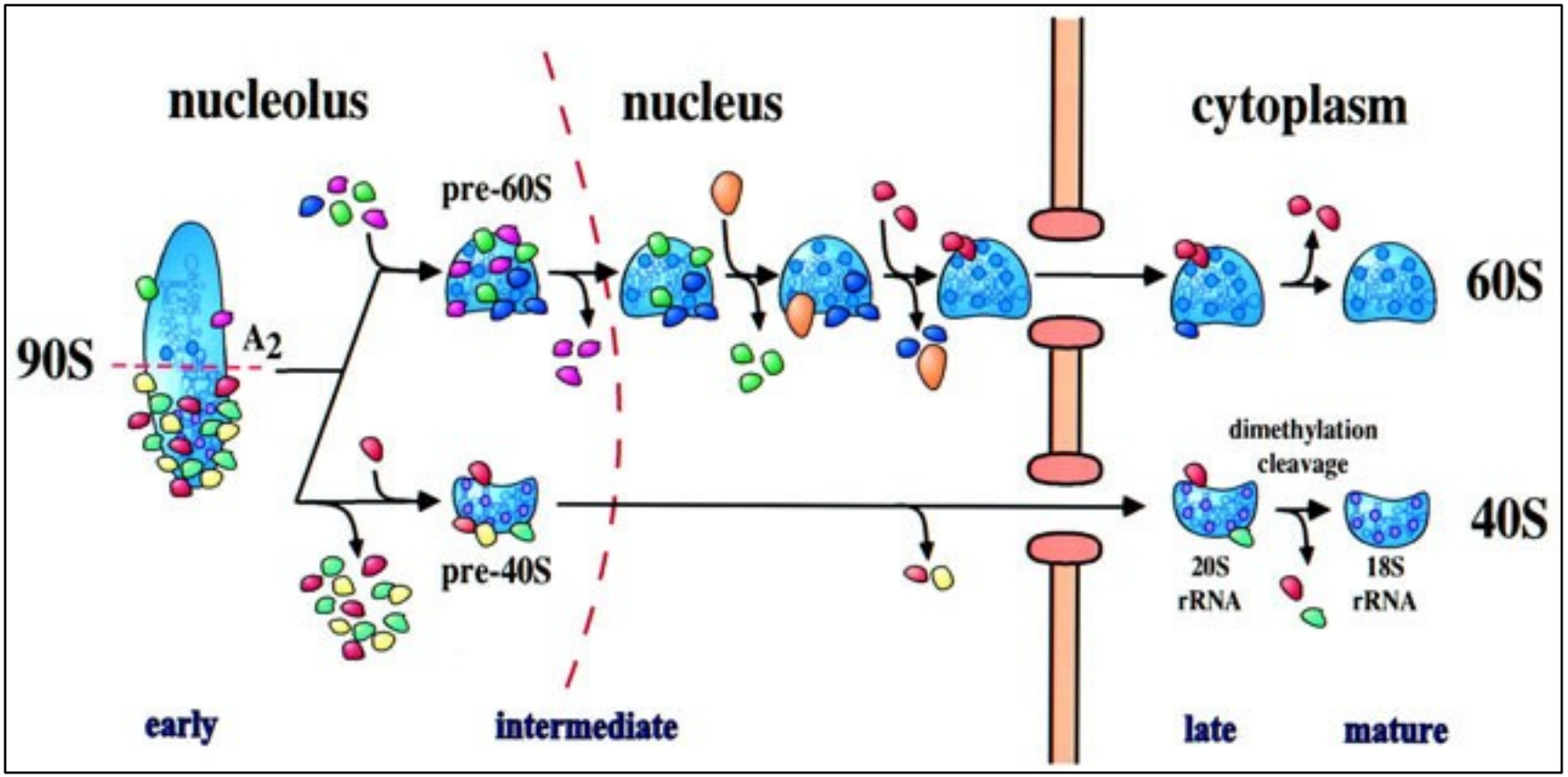
2.– Use of PPI data to build the **protein network** corresponding to a **molecular machine**: the PROTEASOME example.

3.– Use of PPI data and pathways to build integrated **protein networks** and find **specific connectors** and **hubs**: the NOTCH example.

From protein interactions to protein networks

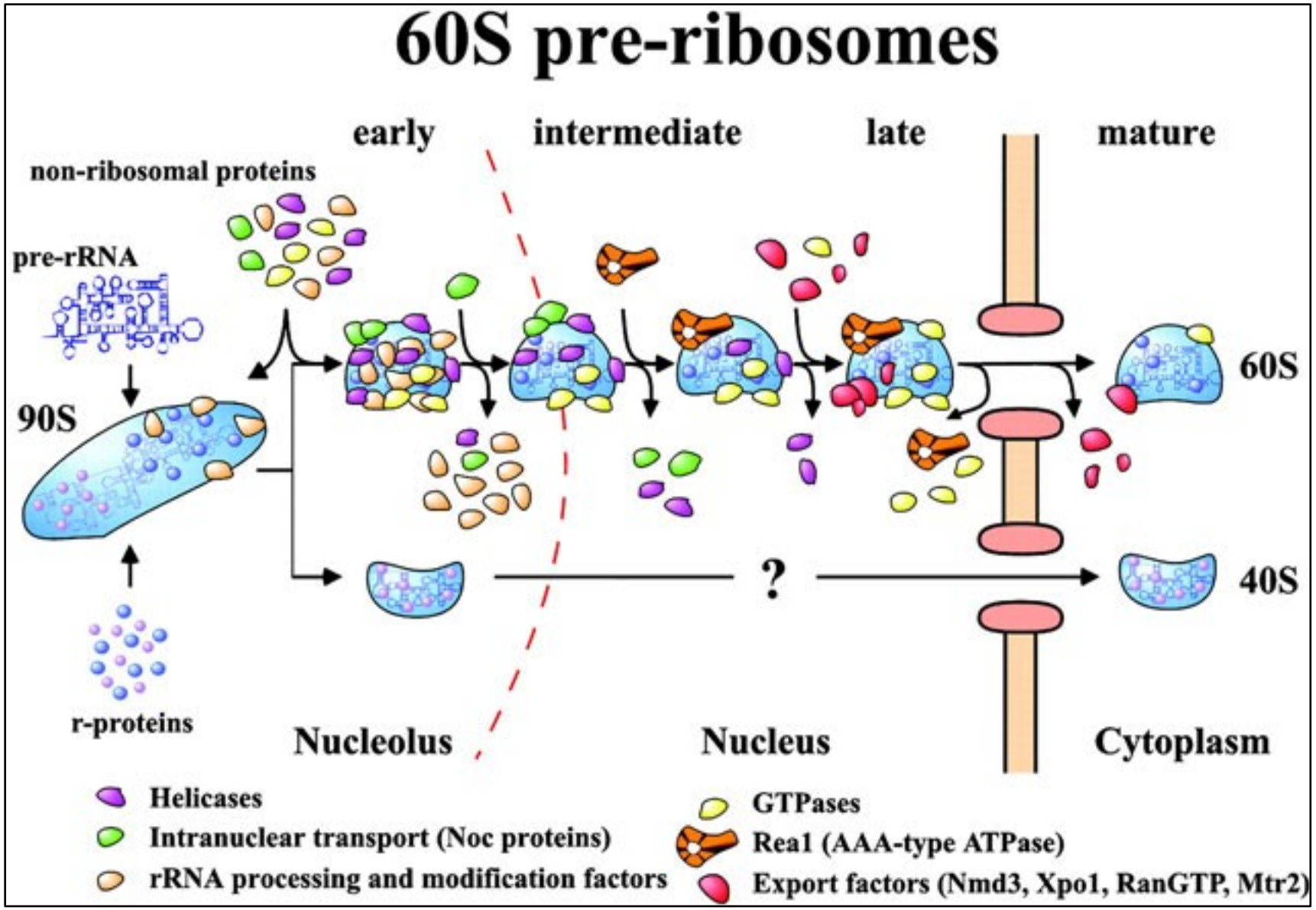
build reliable networks with biological meaning: example 1

Building a molecular machine: **Pre-RIBOSOME** (90S)
steps for the biogenesis and **assemble** of the **ribosome**



From ***Schäfer et al. (2003) EMBO Journal***

Model of the pathway of 60S **pre-ribosome maturation** and **export**



From ***Nissan et al. (2002) EMBO Journal***

From protein interactions to protein networks

build reliable networks with biological meaning: example 1

Many proteins have been involved in the **assemble** of **Pre-RIBOSOME** (90S)

NameSystematic	Uniprot_ID	NameGene	UniProt_Name	Synonyms	MW(kDa)	Study	SubComplex	Description
YJL109c	P42945	Utp10	UTP10_YEAST	na	200.08	1stStudy	UTP-A	U3 small nucleolar RNA-associated protein 10U3 snoRNA-associate
YPL126w	Q02931	Nan1	NAN1_YEAST	Utp17	101.24	1stStudy	UTP-A	Nucleolar protin NAN1U3 small nucleolar RNA-associated protein 17
YDR324c	Q06679	Utp4	UTP4_YEAST	na	87.8	1stStudy	UTP-A	U3 small nucleolar RNA-associated protein 4U3 snoRNA-associated
YGR128c	P53276	Utp8	UTP8_YEAST	na	80.19	1stStudy	UTP-A	U3 small nucleolar RNA-associated protein 8U3 snoRNA-associated
YDR398w	Q04177	Utp5	UTP5_YEAST	na	72	1stStudy	UTP-A	U3 small nucleolar RNA-associated protein 5U3 snoRNA-associated
YHR196w	P38882	Utp9	UTP9_YEAST	na	65.27	1stStudy	UTP-A	U3 small nucleolar RNA-associated protein 9U3 snoRNA-associated
YMR093w	Q04305	Utp15	UTP15_YEAST	na	57.69	1stStudy	UTP-A	U3 small nucleolar RNA-associated protein 15U3 snoRNA-associate
YLR129w	Q12220	Dip2	DIP2_YEAST	na	106.34	1stStudy	UTP-B	DOM34 interacting protein 2U3 small nucleolar RNA-associated prot
YLR409c	Q06078	Utp21	YL09_YEAST	na	104.79	1stStudy	UTP-B	Hypothetical 104.8 kDa Trp-Asp repeats containing protein in RPL31
YCR057c	P25635	Pwp2	PWP2_YEAST	Utp1	103.98	1stStudy	UTP-B	Periodic tryptophan protein 2U3 small nucleolar RNA-associated prot
YLR222c	Q05946	Utp13	UTP13_YEAST	na	91.03	1stStudy	UTP-B	U3 small nucleolar RNA-associated protein 13U3 snoRNA-associate
YJL069c	P40362	Utp18	CG48_YEAST	na	66.42	1stStudy	UTP-B	Hypothetical 66.4 kDa Trp-Asp repeats containing protein in SMC3-M
YDR449c	Q02354	Utp6	UTP6_YEAST	na	52.42	1stStudy	UTP-B	U3 small nucleolar RNA-associated protein 6U3 snoRNA-associated
YGR090w	P53254	Utp22	YG2L_YEAST	na	140.48	1stStudy	UTP-C	Hypothetical 140.5 kDa protein in CTT1-PRP31 intergenic region
YIL035c	P15790	Cka1	CSK21_YEAST	Csk21	44.67	1stStudy	UTP-C	Casein kinase II, alpha chainCK II alpha subunit
YOR061W	P19454	Cka2	CSK22_YEAST	Csk22	39.4	1stStudy	UTP-C	Casein kinase II, alpha' chain (CK II)
YCL031c	P25368	Rrp7	RRP7_YEAST	na	34.47	1stStudy	UTP-C	Ribosomal RNA processing protein 7
YGL019W	P43639	Ckb1	CSK2B_YEAST	Csk2b	32.26	1stStudy	UTP-C	Casein kinase II beta subunitCK II beta
YOR039W	P38930	Ckb2	CSK2C_YEAST	Csk2c	29.84	1stStudy	UTP-C	Casein kinase II beta' subunitCK II beta'
YJR002w	P47083	Mpp10	MPP10_YEAST	na	66.95	1stStudy	MPP10-C	U3 small nucleolar ribonucleoprotein protein MPP10
YNL075w	P53941	Imp4	IMP4_YEAST	na	33.48	1stStudy	MPP10-C	U3 small nucleolar ribonucleoprotein protein IMP4
YHR148w	P32899	Imp3	IMP3_YEAST	na	21.89	1stStudy	MPP10-C	U3 small nucleolar ribonucleoprotein protein IMP3
YPL217c	Q08965	Bms1	BMS1_YEAST	na	135.57	1stStudy	outSubC	Ribosome biogenesis protein BMS1
YGR145w	P48234	Enp2	YG3J_YEAST	na	81.75	1stStudy	outSubC	Hypothetical WD-repeat protein in MOL1-NAT2 intergenic region
YMR290c	Q03532	Has1	HAS1_YEAST	na	56.72	1stStudy	outSubC	Probable ATP-dependent RNA helicase HAS1
YNL132w	P53914	Kre33	YNN2_YEAST	na	119.35	1stStudy	outSubC	Hypothetical UPF0202 protein YNL132w
YCL059c	P25586	Krr1	YCF9_YEAST	na	37.16	1stStudy	outSubC	Hypothetical 37.2 kDa protein in CHA1-PRD1 intergenic region
YPR144c	Q06512	Noc4	NOC4_YEAST	Utp19	63.64	1stStudy	outSubC	Nucleolar complex protein 4U3 small nucleolar RNA-associated prote
YDL014w	P15646	Nop1	FBRL_YEAST	Lot3_FBRL	34.47	1stStudy	outSubC	FibrillarinNucleolar protein 1
YDL148c	Q99207	Nop14	NOP14_YEAST	Utp2	94.3	1stStudy	outSubC	Nucleolar complex protein 14U3 small nucleolar RNA-associated pro
YMR229c	Q05022	Rrp5	RRP5_YEAST	na	193.13	1stStudy	outSubC	rRNA biogenesis protein RRP5
YBL004w	P35194	Utp20	YBA4_YEAST	na	287.56	1stStudy	outSubC	Hypothetical 287.5 kDa protein in PDR3-HTA2 intergenic region

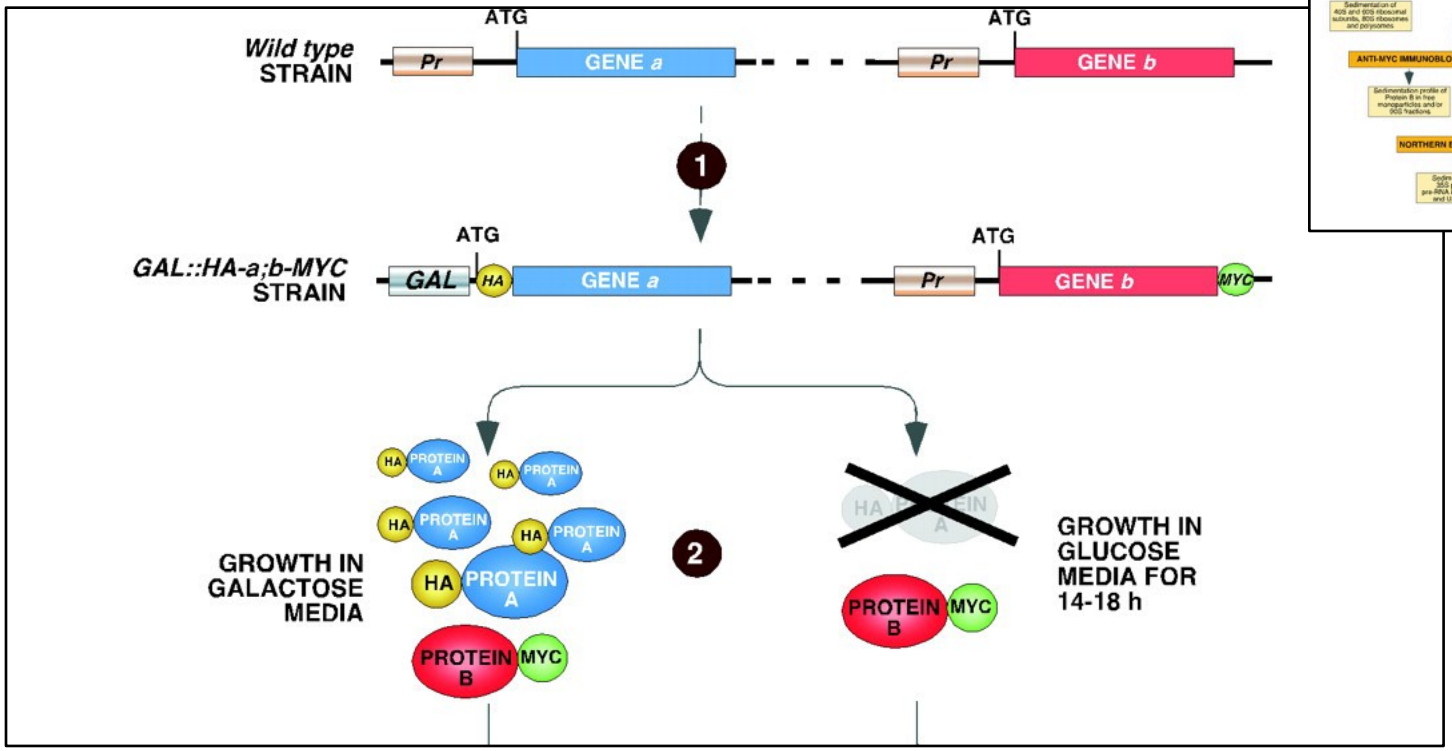
From ***Perez-Fernandez et al. (2007) Mol. Cell. Biol.***

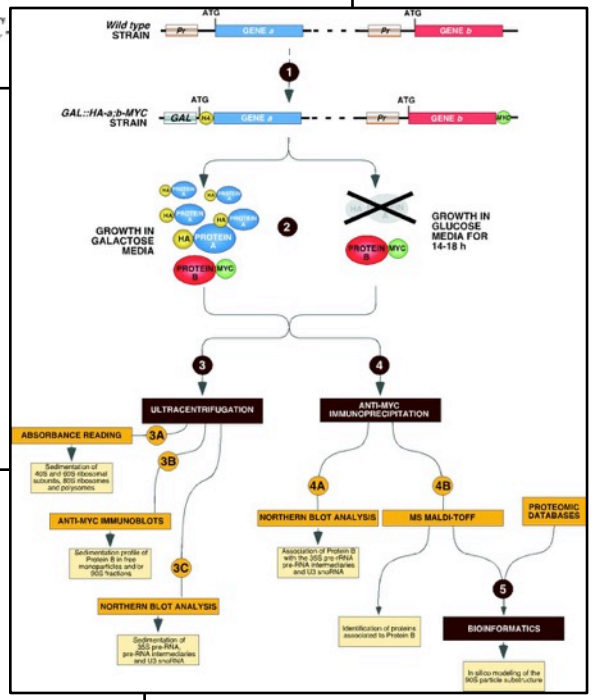
The 90S Preribosome Is a Multimodular Structure That Is Assembled through a Hierarchical Mechanism▽†

Jorge Pérez-Fernández, Ángel Román, Javier De Las Rivas, Xosé R. Bustelo, and Mercedes Dosil*

Centro de Investigación del Cáncer and Instituto de Biología Molecular y Celular del Cáncer, CSIC-Campus Unamuno, E-37007 Salamanca, Spain

Combination **proteomic** techniques, and **bioinformatic** analyses to shed light into the rules of assembly of the yeast 90S preribosome. The results indicate that several protein **subcomplexes** work as **discrete assembly** subunits binding in defined steps.



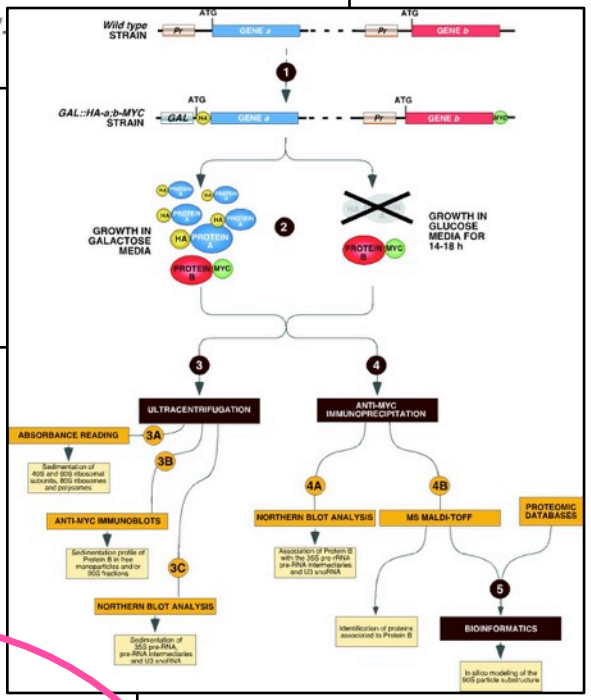


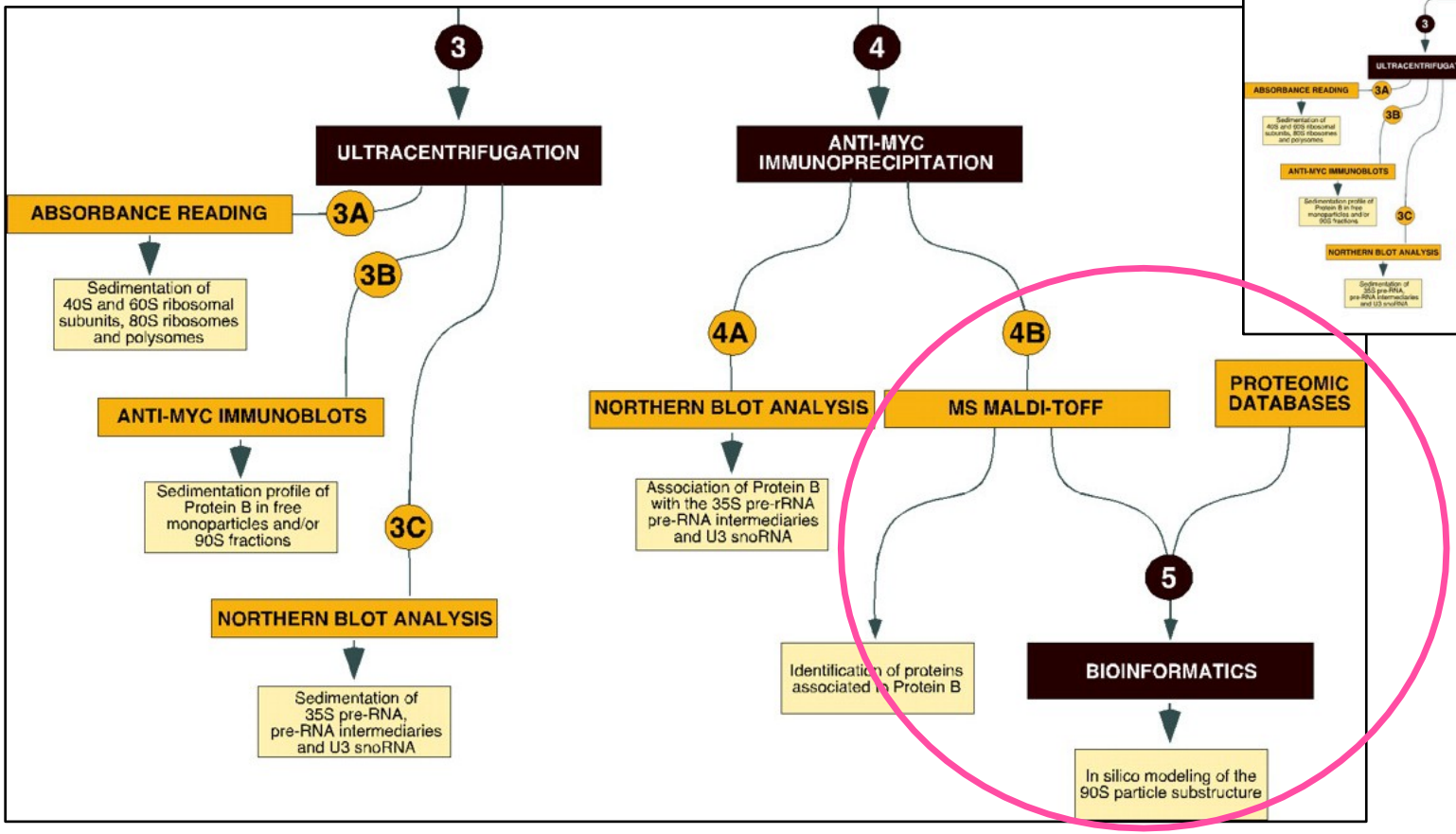
The 90S Preribosome Is a Multimodular Structure That Is Assembled through a Hierarchical Mechanism[▽]†

Jorge Pérez-Fernández, Ángel Román, Javier De Las Rivas, Xosé R. Bustelo, and Mercedes Dosil*

Centro de Investigación del Cáncer and Instituto de Biología Molecular y Celular del Cáncer, CSIC-Campus Unamuno, E-37007 Salamanca, Spain

A bioinformatic approach that provides a model for the **topological arrangement** of protein components within the fully assembled particle.





The 90S Preribosome Is a Multimodular Structure That Is Assembled through a Hierarchical Mechanismvt

Jorge Perez-Fernandez, Angel Roman, Javier De Las Rivas, Xose R. Bustelo, and Mercedes Dosil*

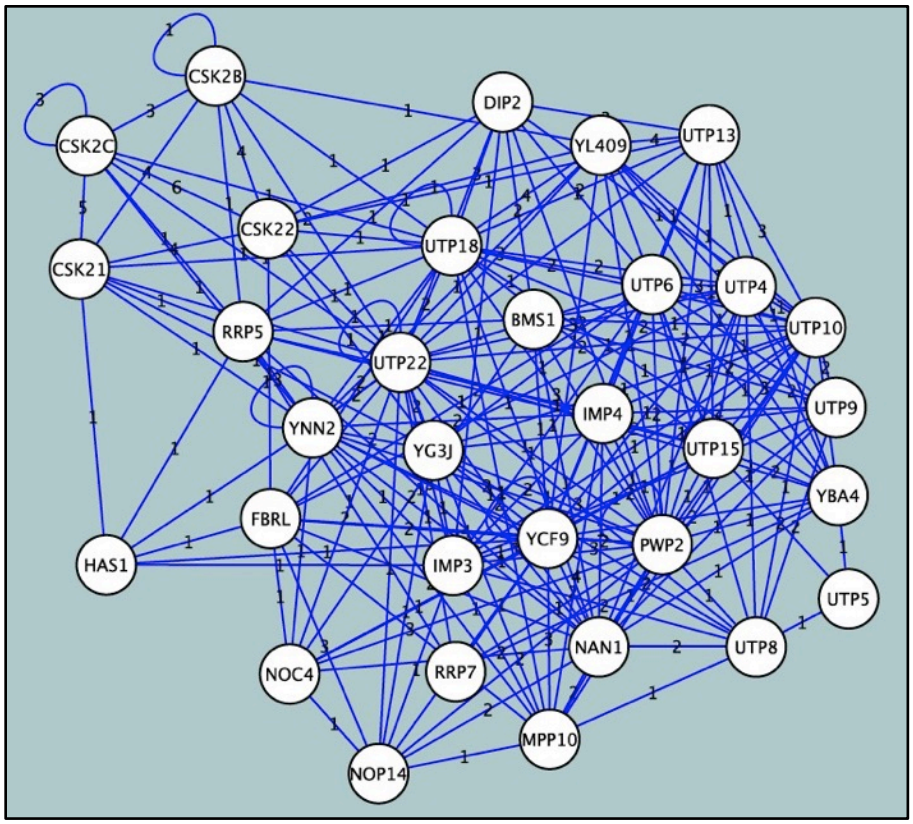
Centro de lnvestigaciOn del Cancer and lnstituto de Biologfa Molecular y Celular del Cancer, CSI C-University of Salamanca, Campus Unamuno, E-37007 Salamanca, Spain

			Pwp2p-MYC			Rrp7p-MYC			Nan1p-MYC			Utp4p-MYC
			Nan1p depletion	Rrp7p depletion	Rrp5p depletion	Nan1p depletion	Pwp2p depletion	Rrp5p depletion	Pwp2p depletion	Rrp7p depletion	Rrp5p depletion	Pwp2p depletion
UTP-AI t-UTP	Utp10p	YJL109c										
	Nan1p	YPL126w										
	Utp4p	YDR324c										
	Utp8p	YGR128c										
	Utp5p	YDR398w										
	Utp9p	YHR196w										
	Utp15p	YMR093w										
Pwp2p/ UTP-B	Pwp2p	YCR057c										
	Dlp2p	YLR129w										
	Utp21p	YLR409c										
	Utp13p	YLR222c										
	Utp18p	YJL069c										
	Utp6p	YDR449c										
UTP-C	Utp22p	YGR090w										
	Rrp7p	YCL03 1c										
	Cka1p	YIL035c										
Mpp10	Mpp10p	YJR002w										
	Imp4p	YNL075w										
OTHER 90S PROTEINS	Utp20p	YBL004w										
	Rrp5p	YMR229c										
	Bms1p	YPL217c										
	Kre33p	YNL132w										
	Noo14o	YDL148c										
	Enp2p	YGR145w										
	Noc4p	YPR144c										
	Has1p	YMR290c										
	Krr1p	YCL059c										
	Nop1p	YDL014w										

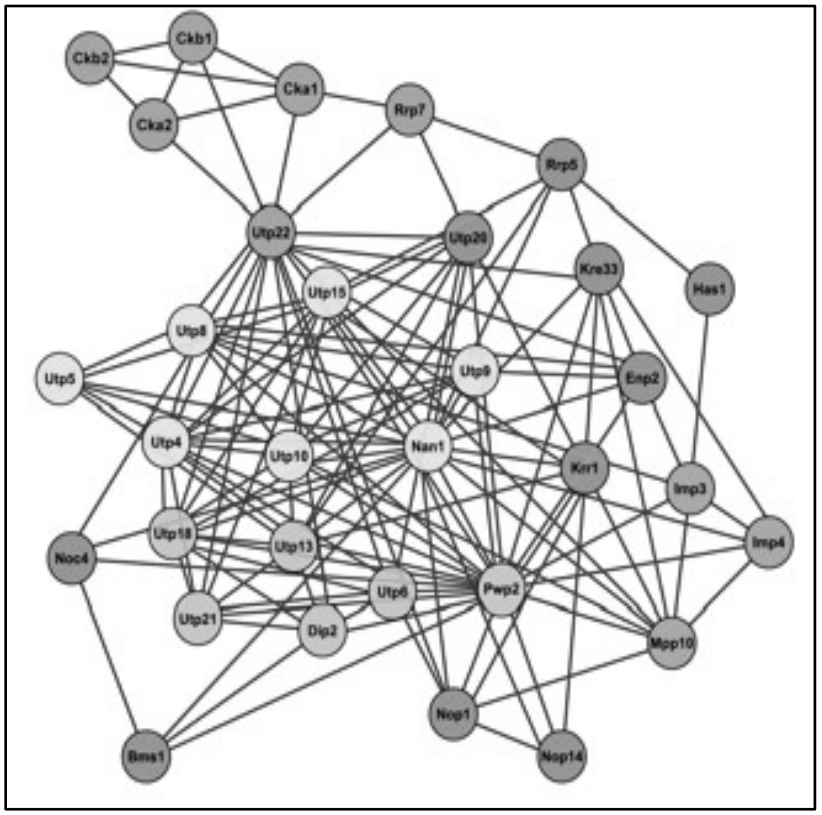
From protein interactions to protein networks

build reliable networks with biological meaning: example 1

Proteomics finds 32 proteins involved in the **assemble** of **Pre-RIBOSOME** (90S)



interactions validated by
≥ 1 experimental method



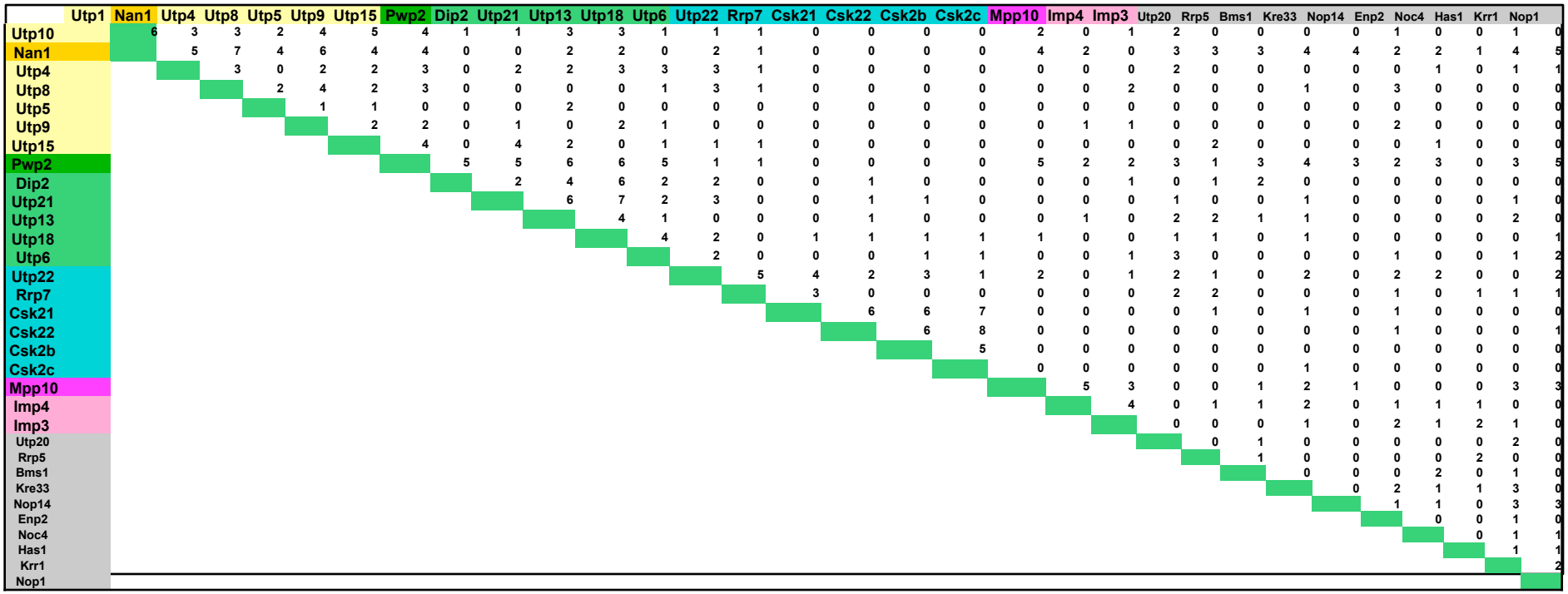
interactions validated by
≥ 2 experimental methods

From ***Perez-Fernandez et al. (2007) Mol. Cell. Biol.***

From protein interactions to protein networks

build reliable networks with biological meaning: example 1

Proteomics finds 32 proteins involved in the **assemble** of **Pre-RIBOSOME** (90S)



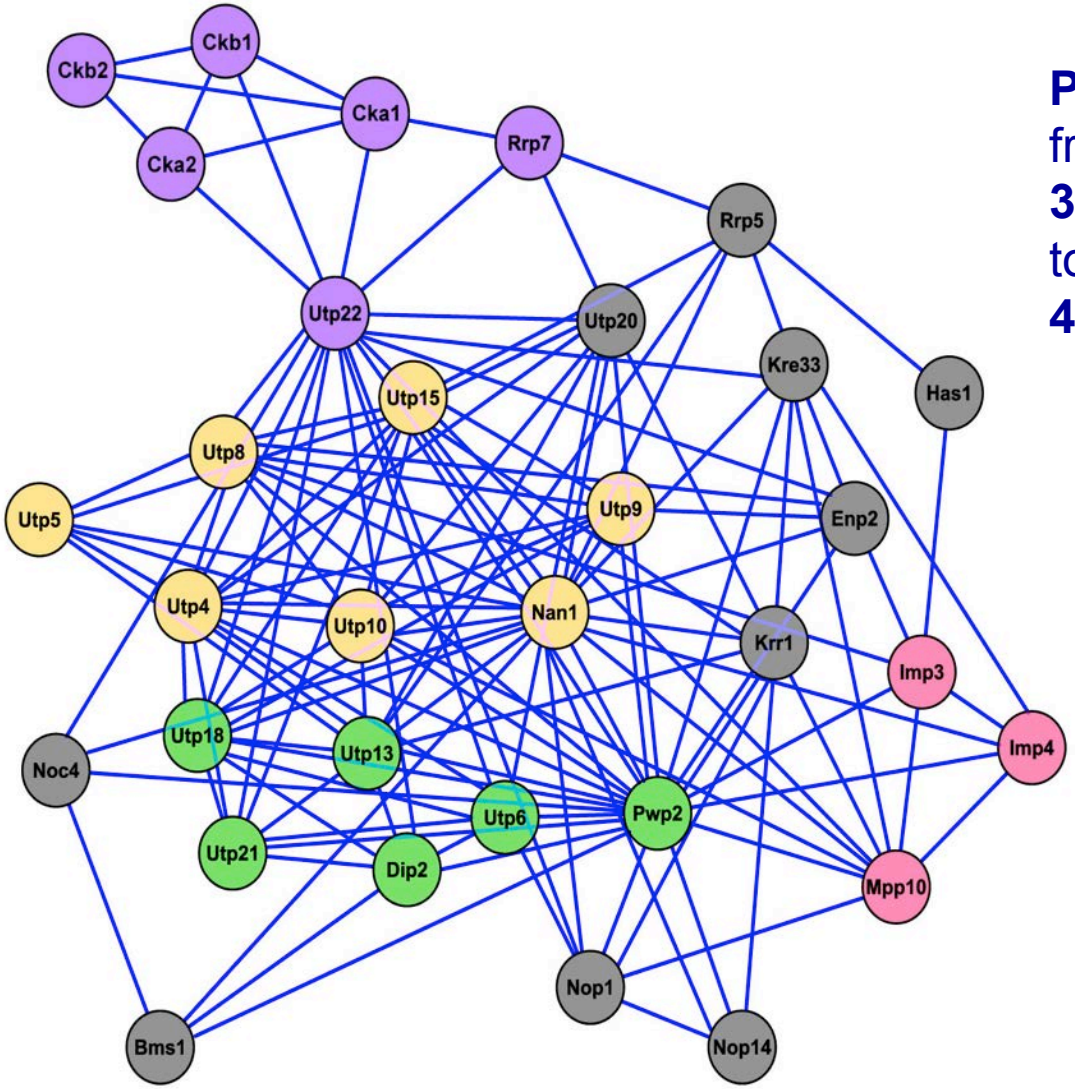
	Utp1	Nan1	Utp4	Utp8	Utp5	Utp9	Utp15	Pwp2	Dip2	Utp21	Utp13	Utp18	Utp6	Utp22	Rrp7	Csk21	Csk22	Csk2b	Csk2c	Mpp10	Imp4	Imp3	Utp20	Rrp5	Bms1	Kre33	Nop14	Enp2	Noc4	Has1	Krr1	Nop1
Utp10	6	3	3	2	4	5	4	1	1	3	3	1	1	1	0	0	0	0	2	0	1	2	0	0	0	0	1	0	0	1	0
Nan1	5	7	4	6	4	4	0	0	2	2	0	2	1	0	0	0	0	0	4	2	0	3	3	3	4	4	2	2	1	4	5
Utp4	3	0	2	2	3	0	2	2	3	3	3	1	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	1	0	1	1
Utp8	2	4	2	3	0	0	0	0	1	3	1	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0	3	0	0	0	0
Utp5	1	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Utp9	2	2	0	1	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	2	0	0	0	0
Utp15	4	0	4	2	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	1	0	0	0
Pwp2	5	5	6	6	5	1	1	0	0	0	0	0	0	5	2	2	3	1	3	4	3	2	3	0	3	5
Dip2	2	4	6	2	2	0	0	1	0	0	0	0	0	0	0	1	0	1	2	0	0	0	0	0	0	0
Utp21	6	7	2	3	0	0	1	1	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	1	0
Utp13	4	1	0	0	0	1	0	0	0	1	0	0	0	1	0	2	2	1	1	0	0	0	0	0	0	2	0
Utp18	4	2	0	1	1	1	1	1	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1
Utp6	2	0	0	0	1	1	0	0	1	3	0	0	0	0	1	0	0	1	2
Utp22	5	4	2	3	1	2	0	1	2	1	0	2	0	2	2	0	0	2
Rrp7	3	0	0	0	0	0	0	2	2	0	0	0	1	0	1	1	1
Csk21	6	6	7	0	0	0	0	1	0	1	0	1	0	0	0	0
Csk22	6	8	0	0	0	0	0	0	0	0	0	1	0	0	0	1
Csk2b	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Csk2c	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Mpp10	5	3	0	0	1	2	1	0	0	0	3	3
Imp4	4	0	1	1	2	0	1	1	1	0	0
Imp3	0	0	0	1	0	2	1	2	1	0
Utp20	0	1	0	0	0	0	0	2	0
Rrp5	1	0	0	0	0	2	0	0
Bms1	0	0	0	2	0	1	0
Kre33	0	2	1	1	3	0
Nop14	1	1	0	3	3
Enp2	0	0	1	0
Noc4	0	1	1
Has1	1	1
Krr1	2
Nop1	

symmetric matrix of binary protein-protein interactions,
weighted by the number of experimental methods
that validate each interaction

From ***Perez-Fernandez et al. (2007) Mol. Cell. Biol.***

From protein interactions to protein networks

build reliable networks with biological meaning: example 1



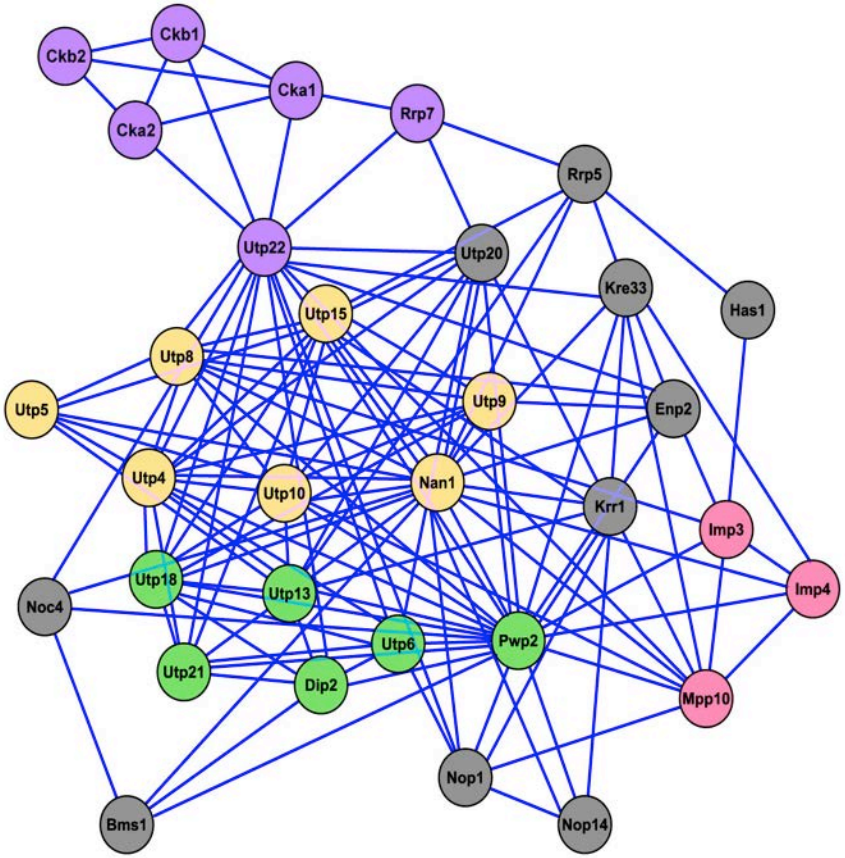
Pre-RIBOSOME
from
32 proteins
to
4 groups

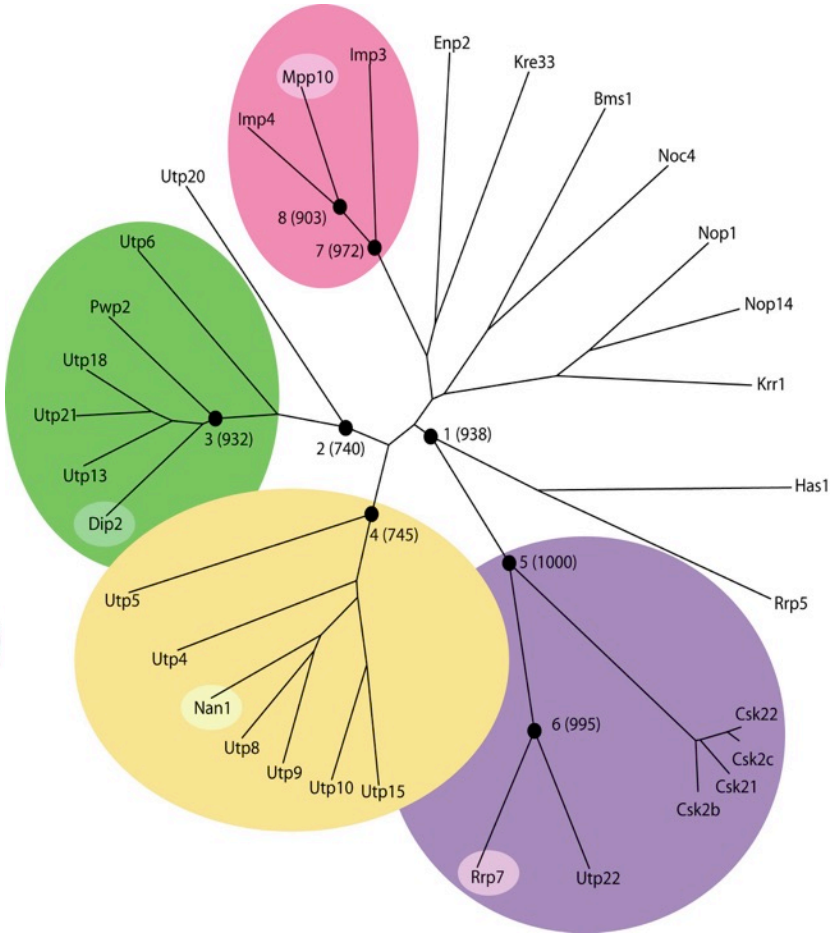
From ***Perez-Fernandez et al. (2007) Mol. Cell. Biol.***

From protein interactions to protein networks

build reliable networks with biological meaning: example 1

Proteomics finds 32 proteins involved in the **assemble** of **Pre-RIBOSOME** (90S)



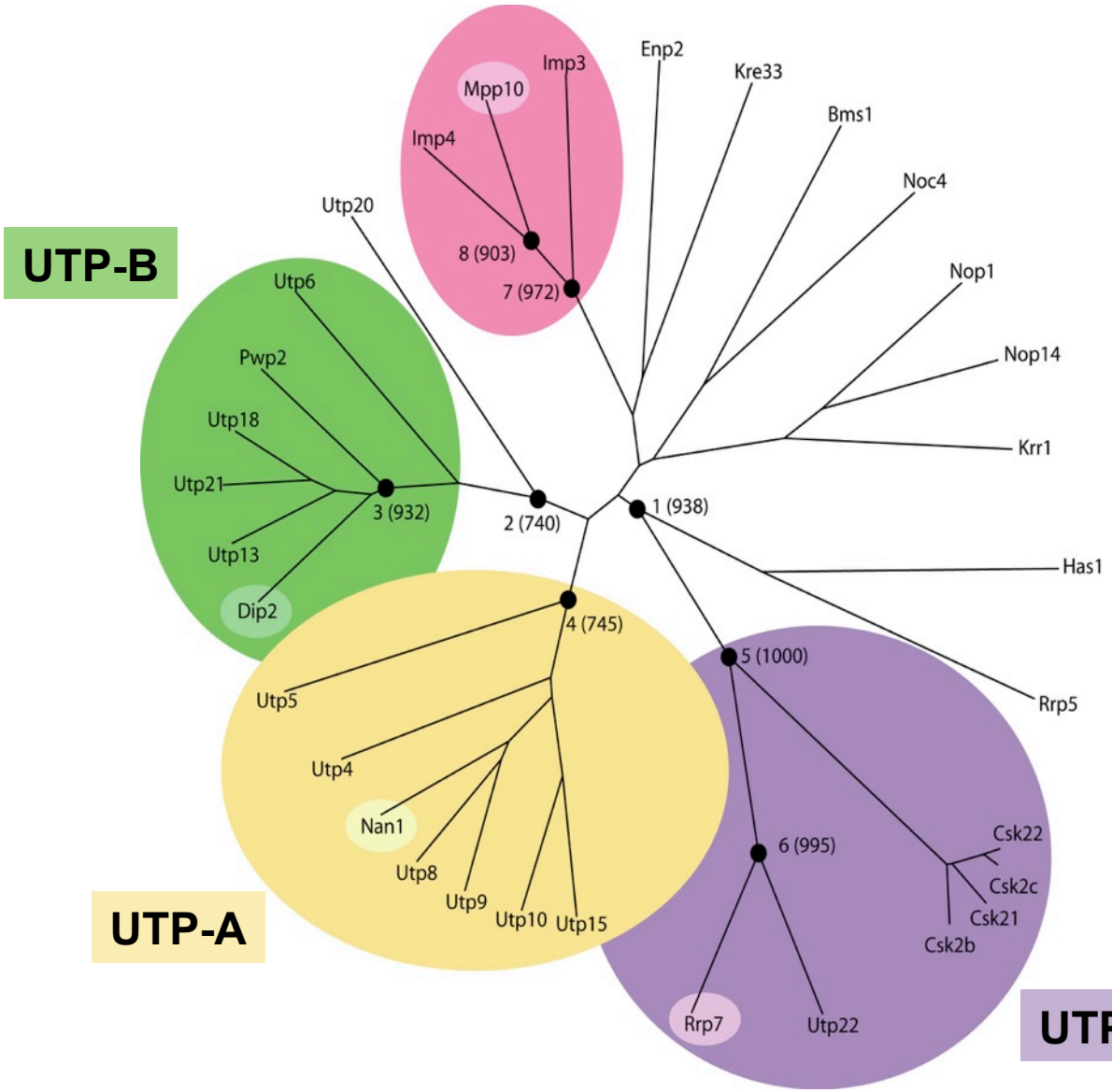


using former matrix we calculate the binary distances and we generate a tree

From ***Perez-Fernandez et al. (2007) Mol. Cell. Biol.***

From protein interactions to protein networks

build reliable networks with biological meaning: example 1



Pre-RIBOSOME
from
32 proteins
to
4 sub-complexes

We discover protein groups that correspond to subcomplexes experimentally found

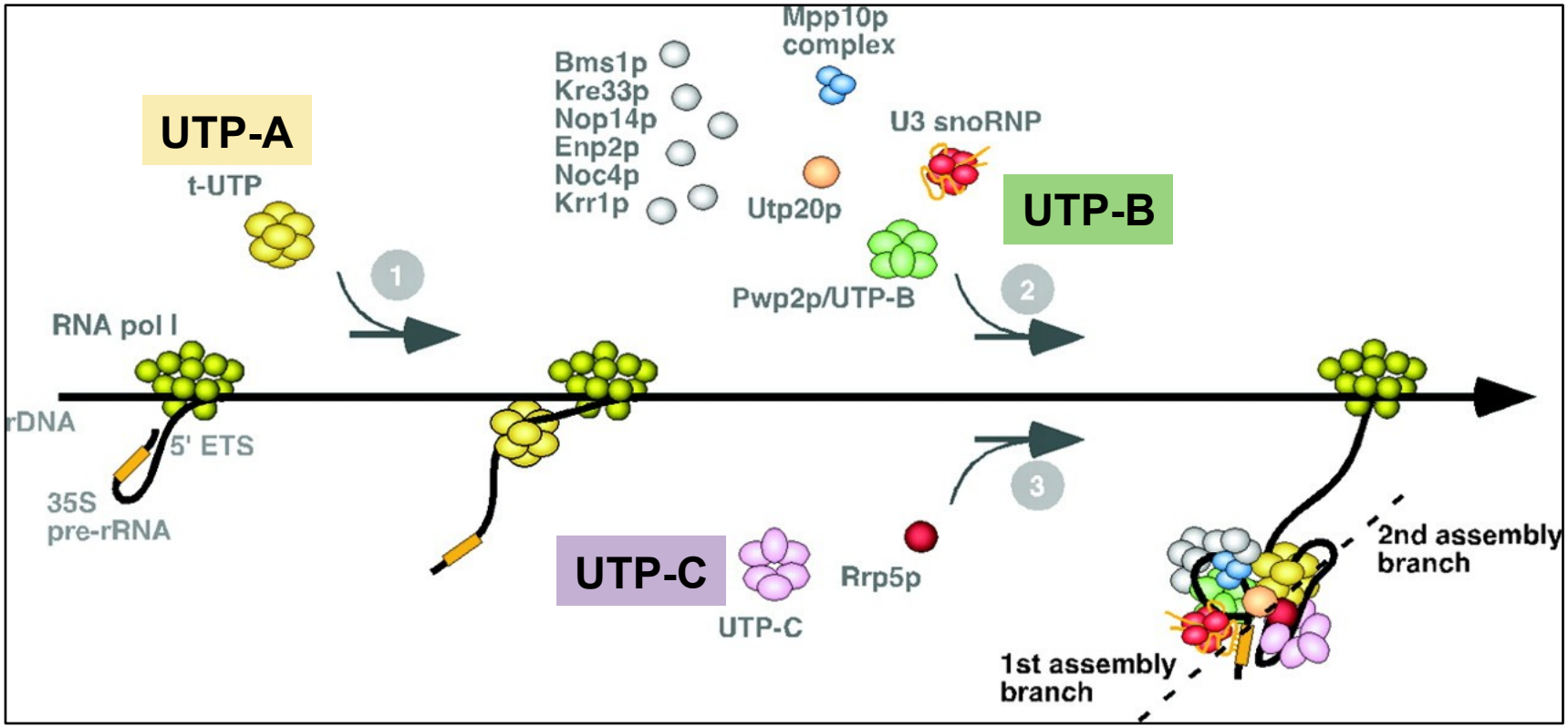
From ***Perez-Fernandez et al. (2007) Mol. Cell. Biol.***

From protein interactions to protein networks

build reliable networks with biological meaning: example 1

Building a molecular machine: **Pre-RIBOSOME** (90S),
steps for the biogenesis and **assemble** of the **ribosome**

The **90S pre-ribosomal assembly particle** includes several **subunits**
UTP-A, UTP-B, UTP-C, etc.



From ***Perez-Fernandez et al. (2007) Mol. Cell. Biol.***

From protein interactions to protein networks

build reliable networks with biological meaning: examples

Challenge: obtain and integrate ***omic* data** to build **biological networks** and solve **biological questions**.

Three examples based in PPI data:

1.– Use of PPI data to build **protein networks** and find different **sub-complexes** and **assembly steps:** the PRE-RIBOSOME example.

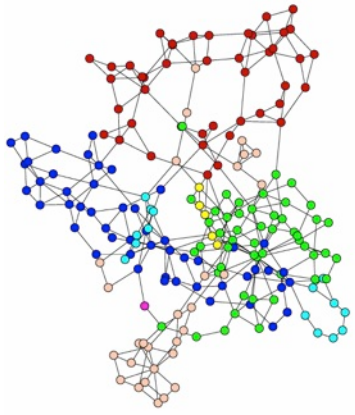
2.– Use of PPI data to build the **protein network** corresponding to a **molecular machine**: the PROTEASOME example.

3.– Use of PPI data and pathways to build integrated **protein networks** and find **specific connectors** and **hubs**: the NOTCH example.

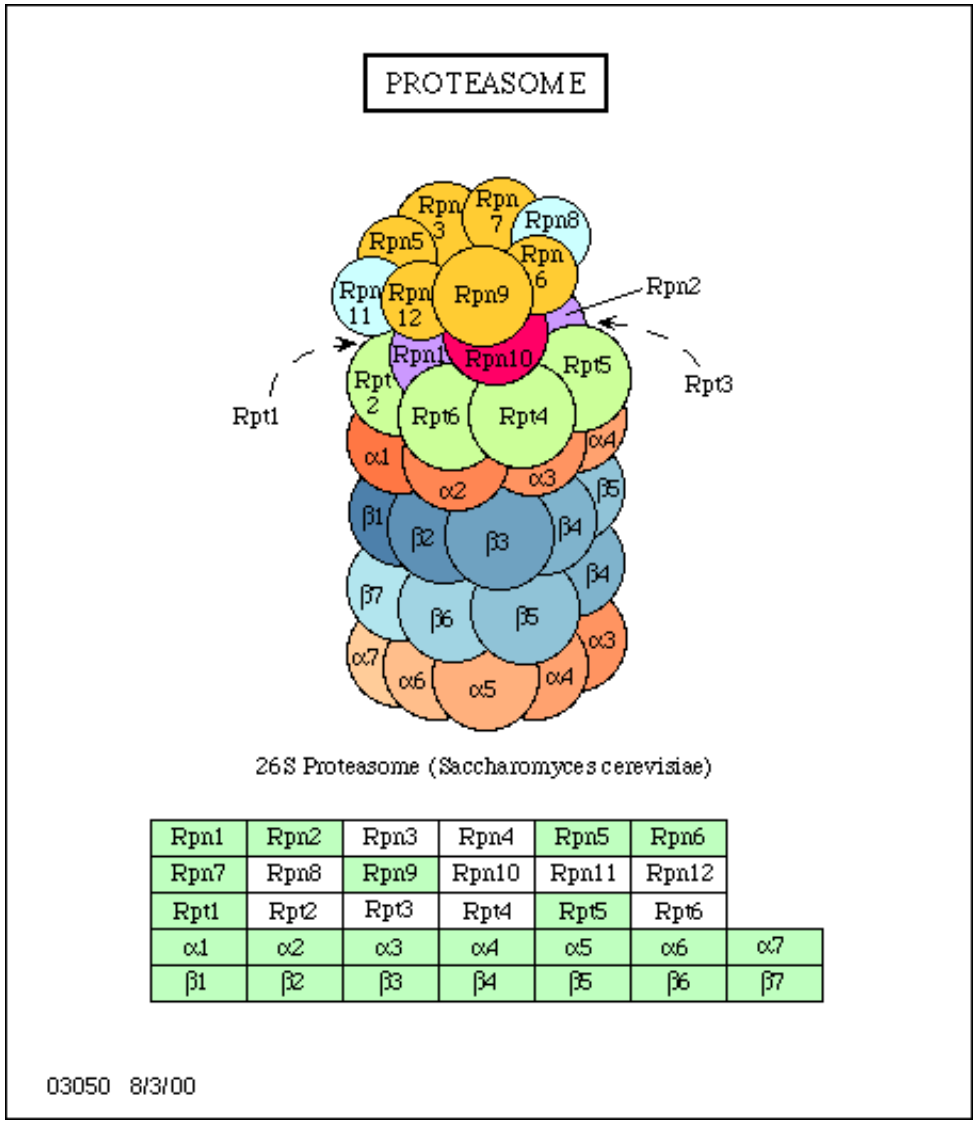
From protein interactions to protein networks

analyse interaction networks to discover biology: example 2

A molecular machine within the **PPI network**: the **PROTEASOME**



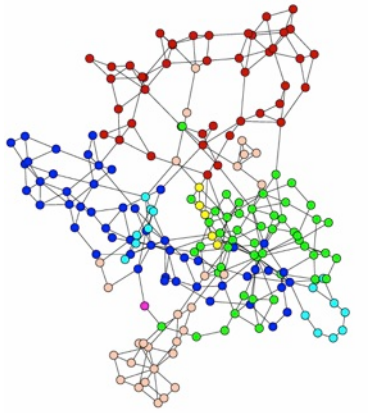
complex
Have all the subunits the same biological role?



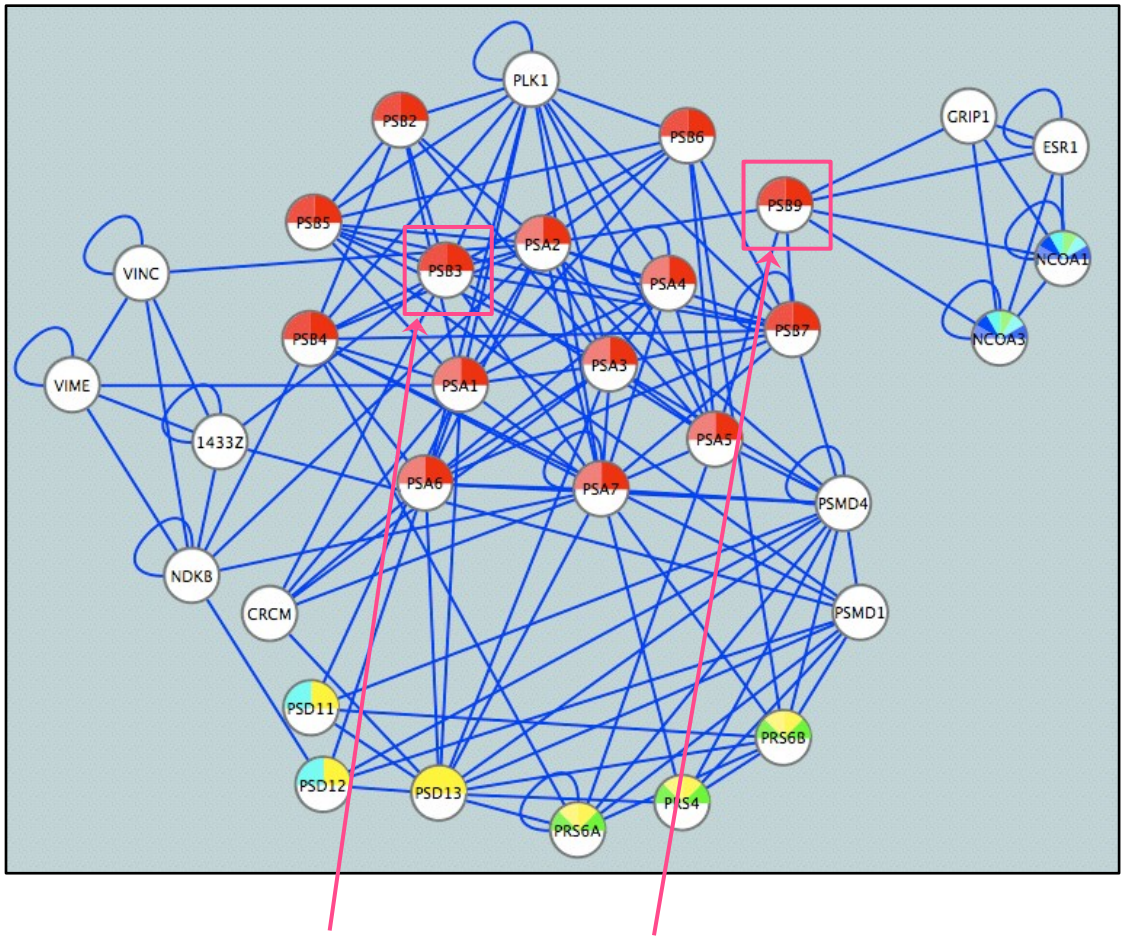
From protein interactions to protein networks

analyse interaction networks to discover biology: example 2

A molecular machine within the **PPI network**: the **PROTEASOME**



network
All the subunits in a complex do not have the same biological role

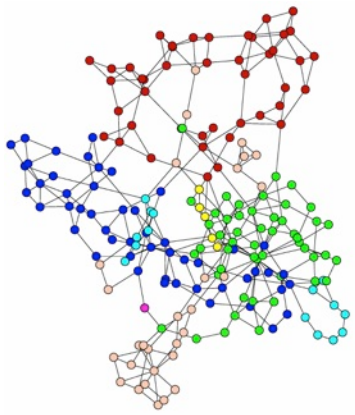


Intramodular hubs *vs* **Intermodular** hubs

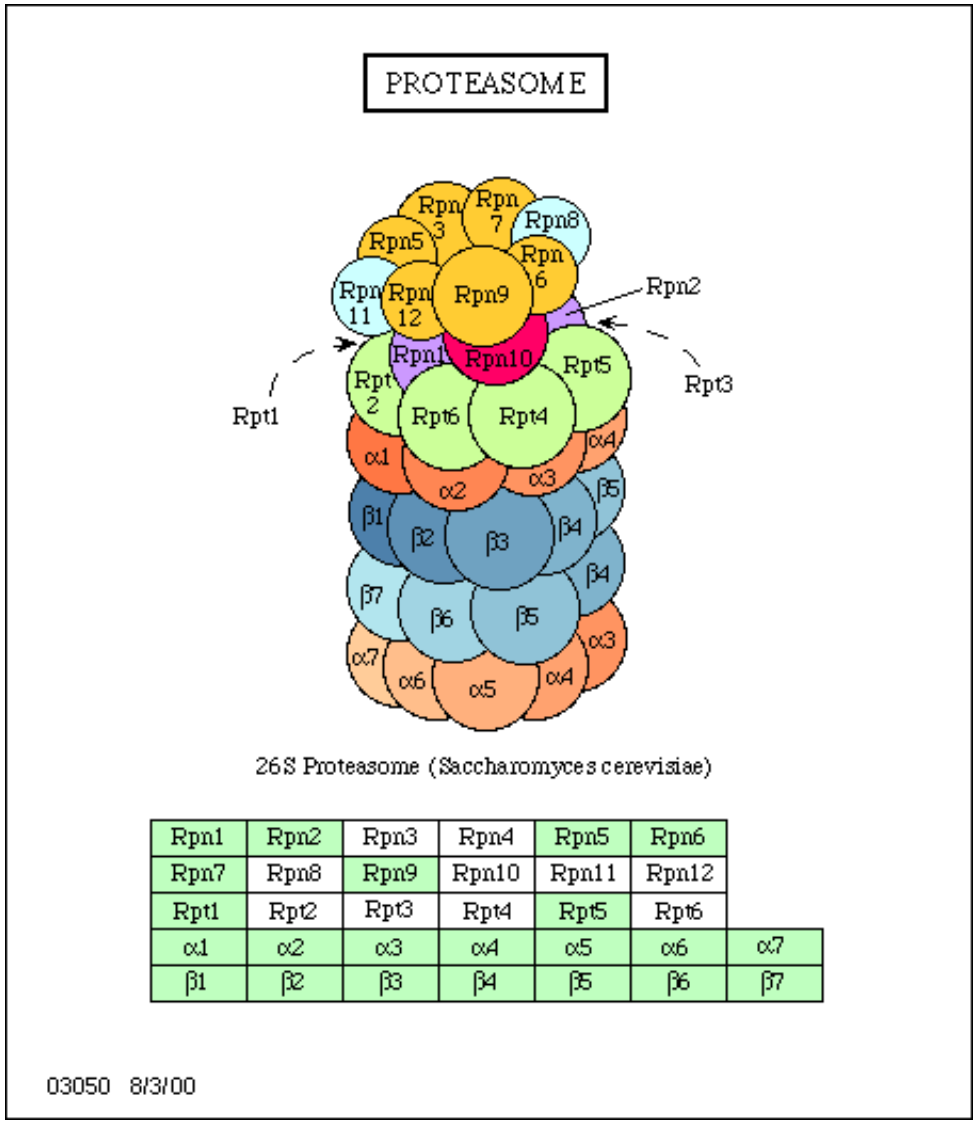
From protein interactions to protein networks

analyse interaction networks to discover biology: example 2

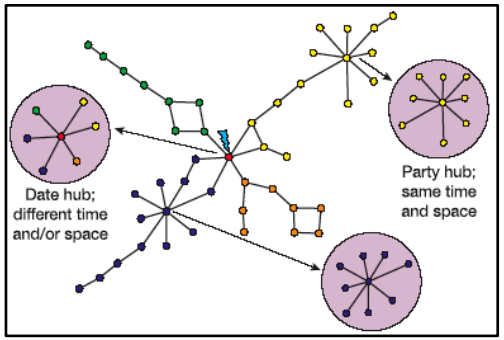
A molecular machine within the **PPI network**: the **PROTEASOME**



complex
Have all the subunits the same biological role?

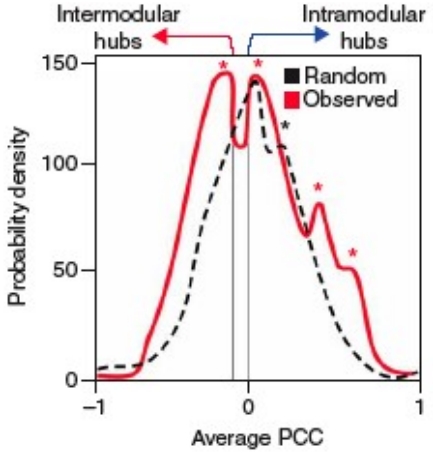


Party hubs *vs* **Date** hubs



Han *et al.* (2004) *Nature*

Intramodular hubs *vs* **Intermodular** hubs

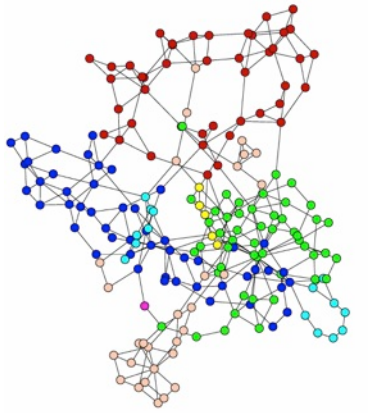


Taylor *et al.* (2009) *Nat. Biotech.*

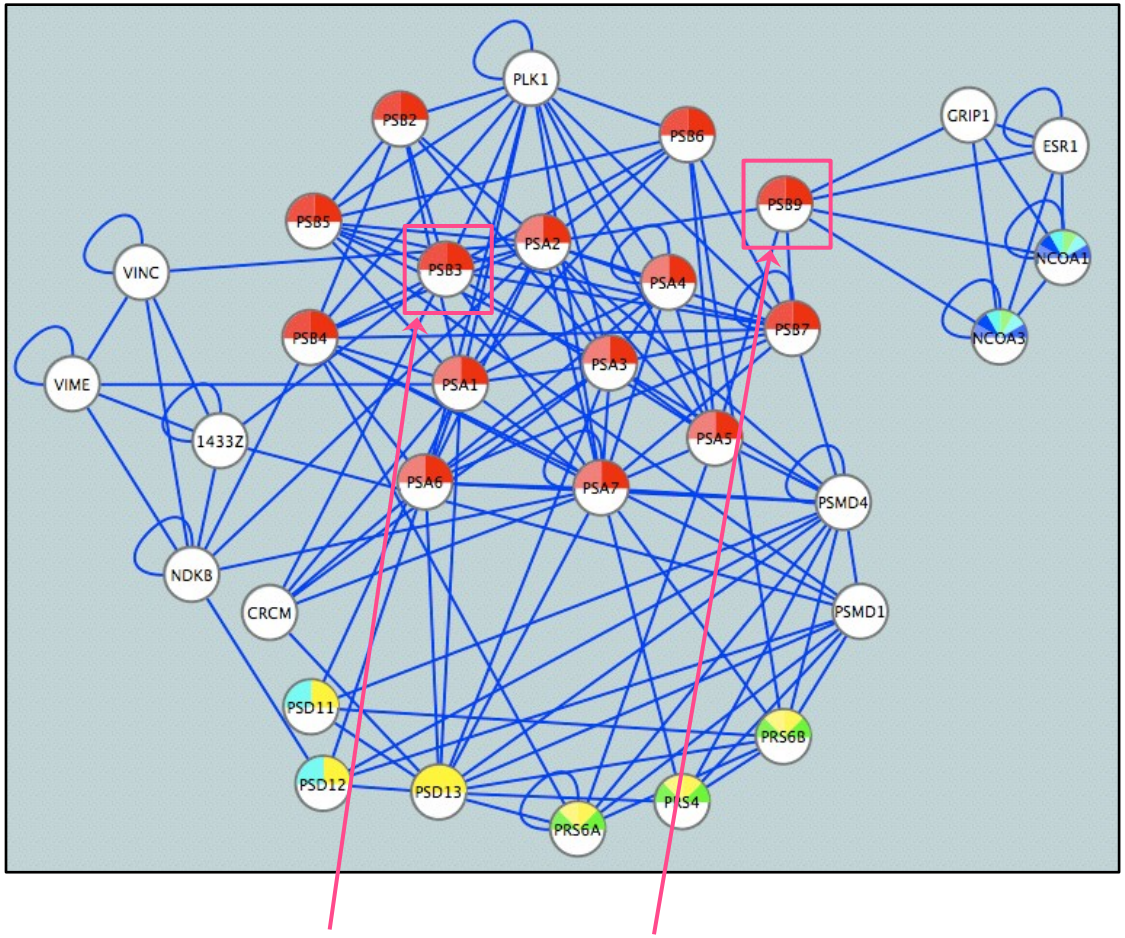
From protein interactions to protein networks

analyse interaction networks to discover biology: example 2

A molecular machine within the **PPI network**: the **PROTEASOME**



network
All the subunits in a complex do not have the same biological role



Intramodular hubs *vs* **Intermodular** hubs

From protein interactions to protein networks

build reliable networks with biological meaning: examples

Challenge: obtain and integrate ***omic* data** to build **biological networks** and solve **biological questions**.

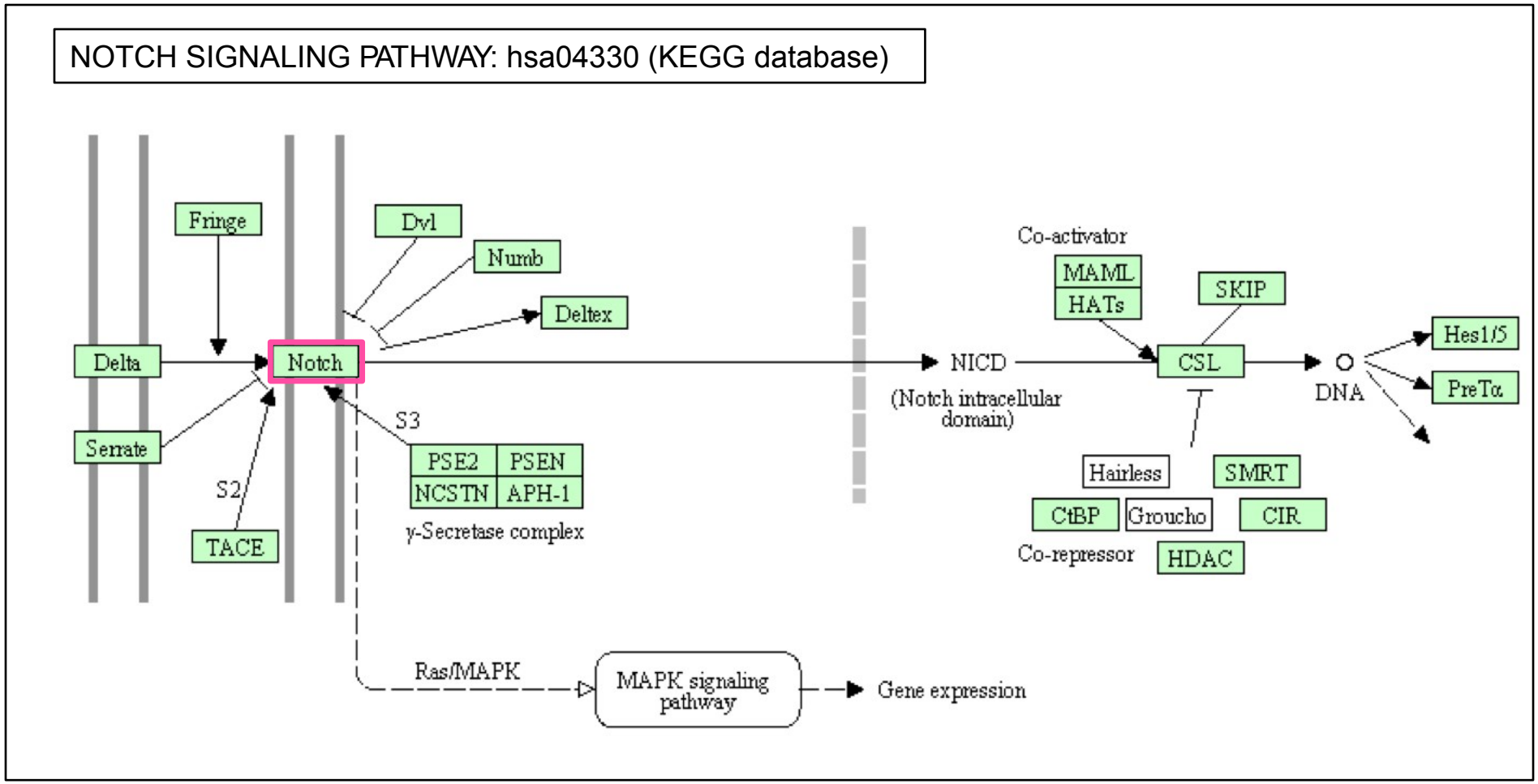
Three examples based in PPI data:

1.– Use of PPI data to build **protein networks** and find different **sub-complexes** and **assembly steps:** the PRE-RIBOSOME example.

2.– Use of PPI data to build the **protein network** corresponding to a **molecular machine**: the PROTEASOME example.

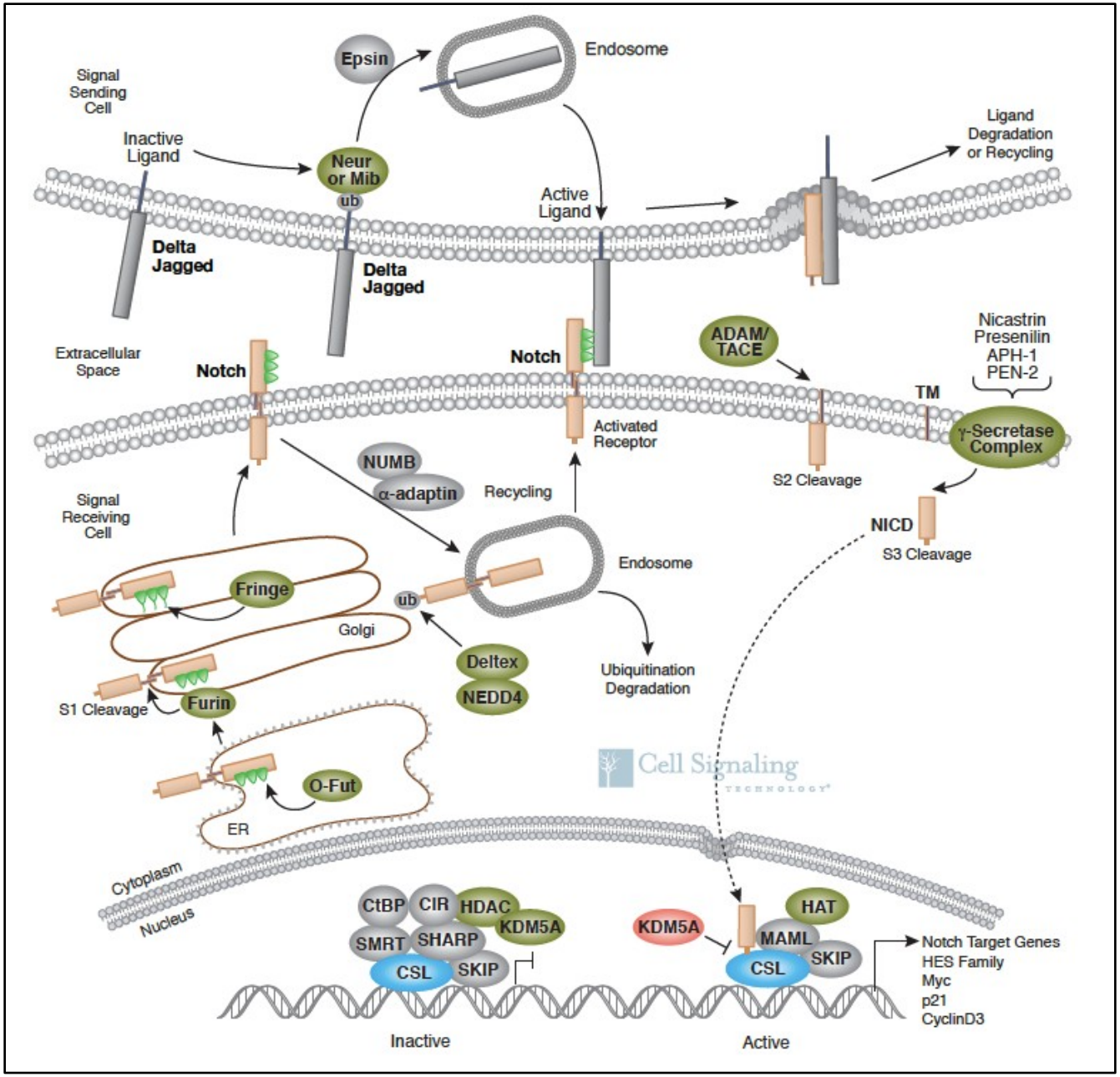
3.– Use of PPI data and pathways to build integrated **protein networks** and find **specific connectors** and **hubs**: the NOTCH example.

Pathways
KEGG: NOTCH signaling



Pathways

NOTCH signaling pathway



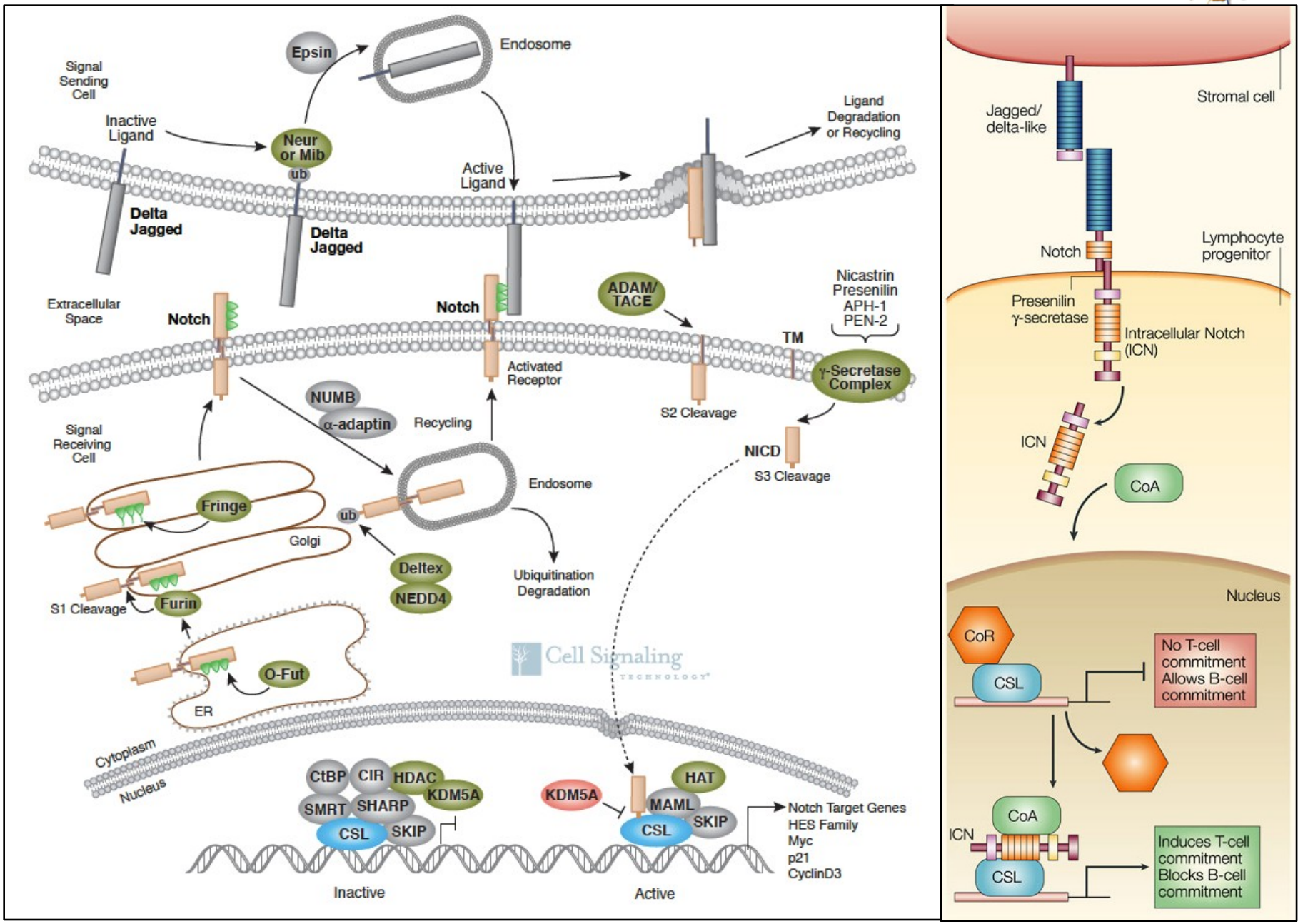
The **notch signaling** pathway is important for **cell-cell communication**, which involves gene regulation mechanisms that control **multiple cell differentiation processes** during embryonic and adult life.

The notch cascade consists of notch and notch ligands, as well as intracellular proteins transmitting the notch signal to the cell's nucleus.

Notch signaling is dysregulated in many cancers.

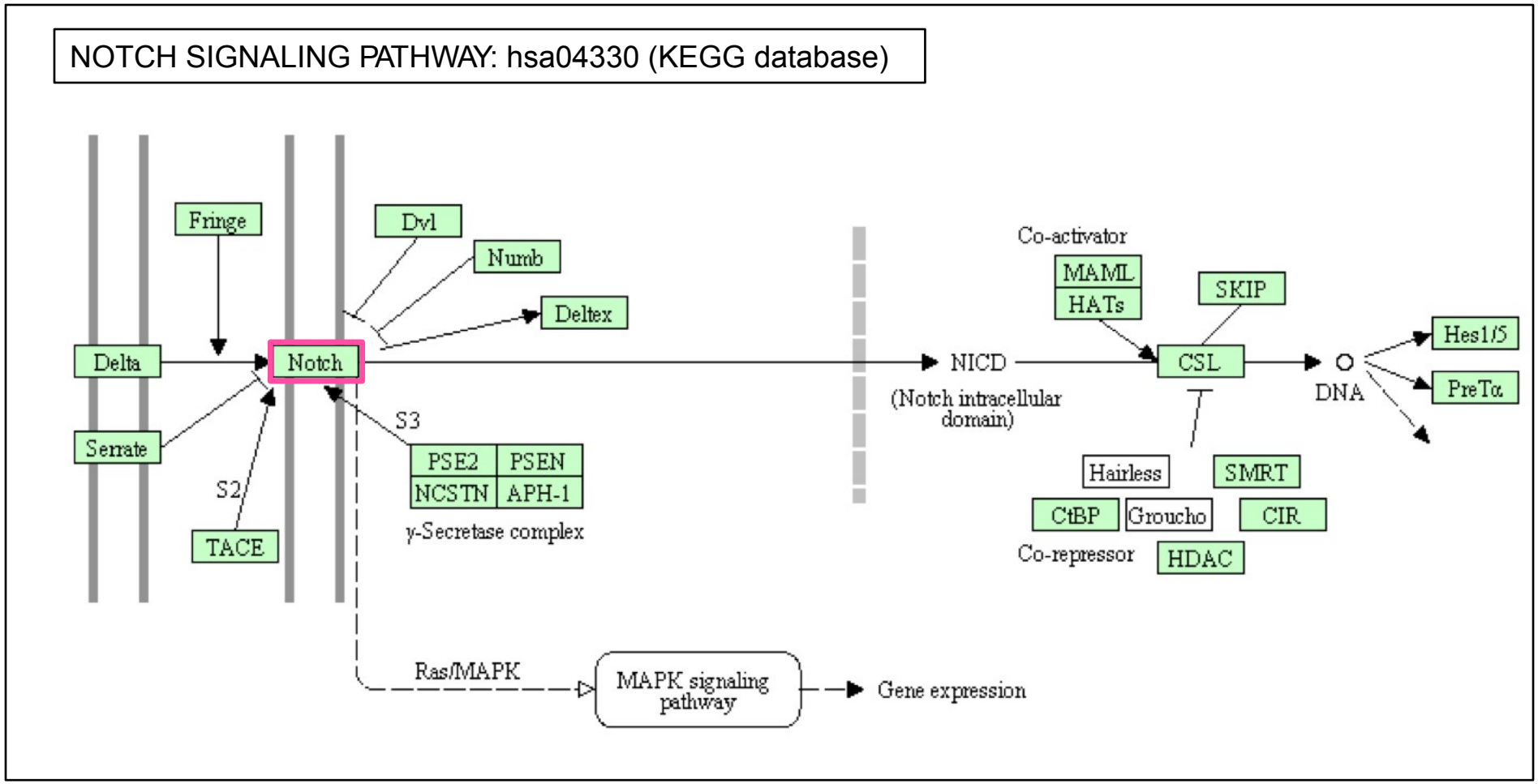
Pathways

NOTCH signaling pathway



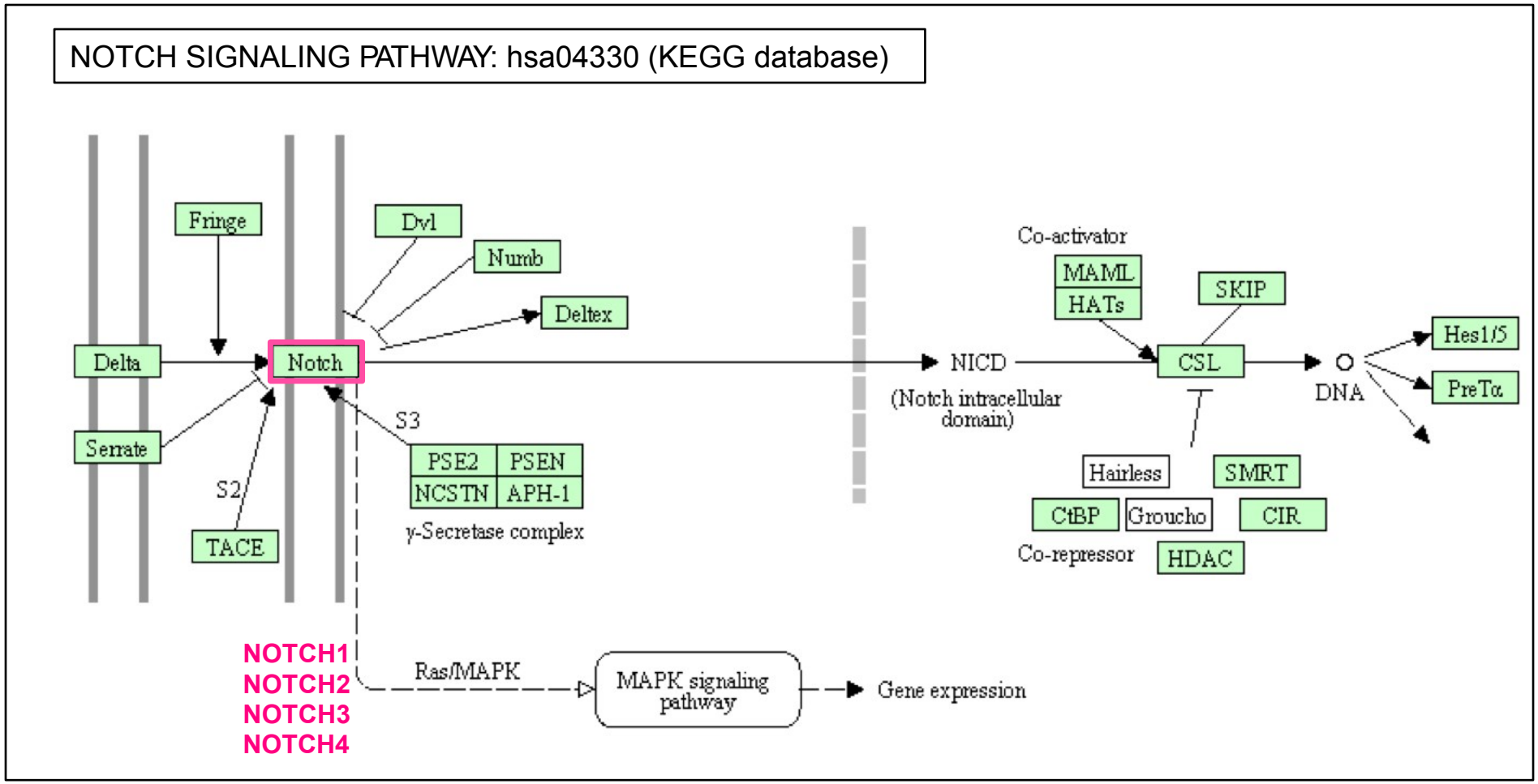
From PPI & pathways to protein networks

build reliable networks with biological meaning: example 3



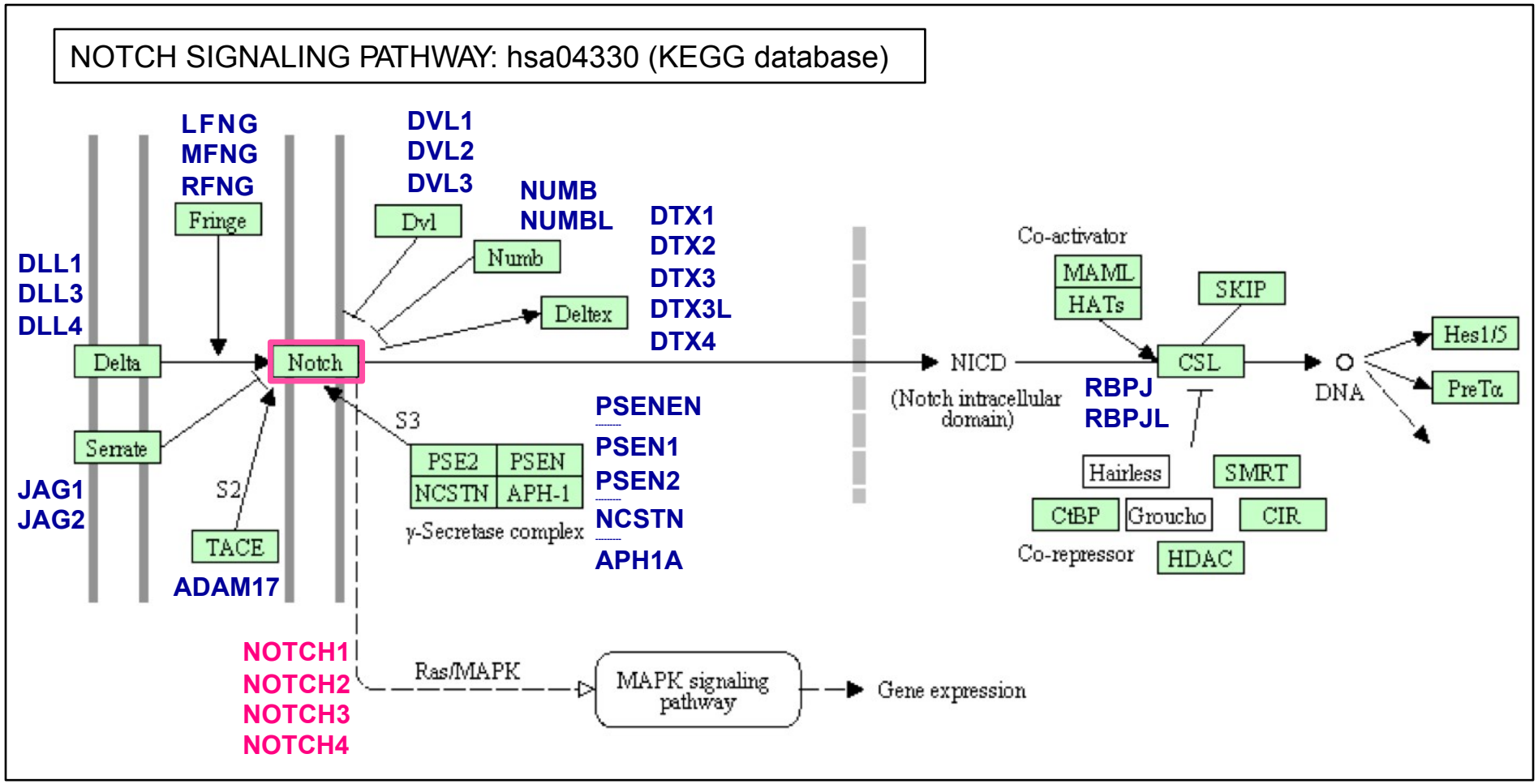
From PPI & pathways to protein networks

build reliable networks with biological meaning: example 3



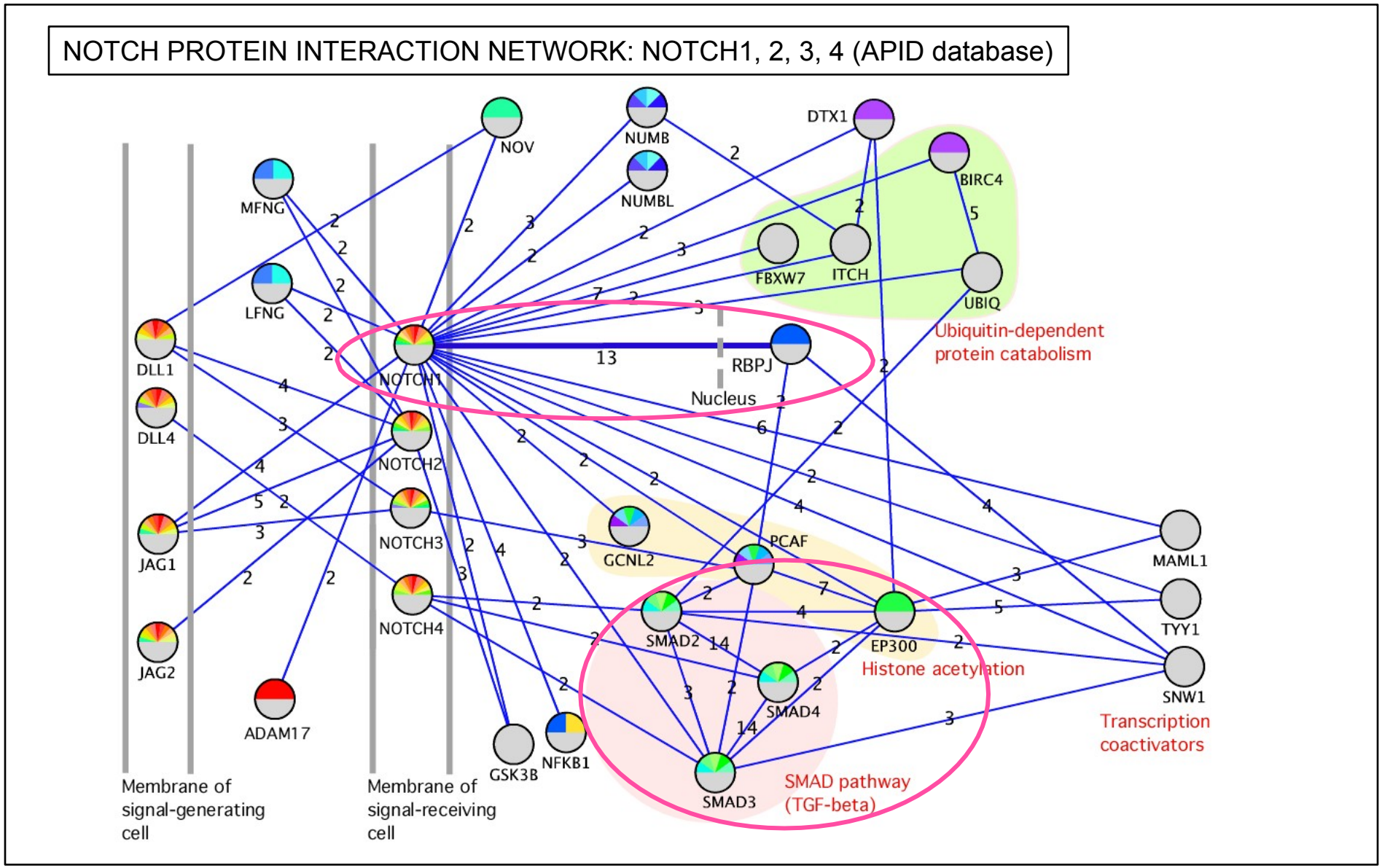
From PPI & pathways to protein networks

build reliable networks with biological meaning: example 3



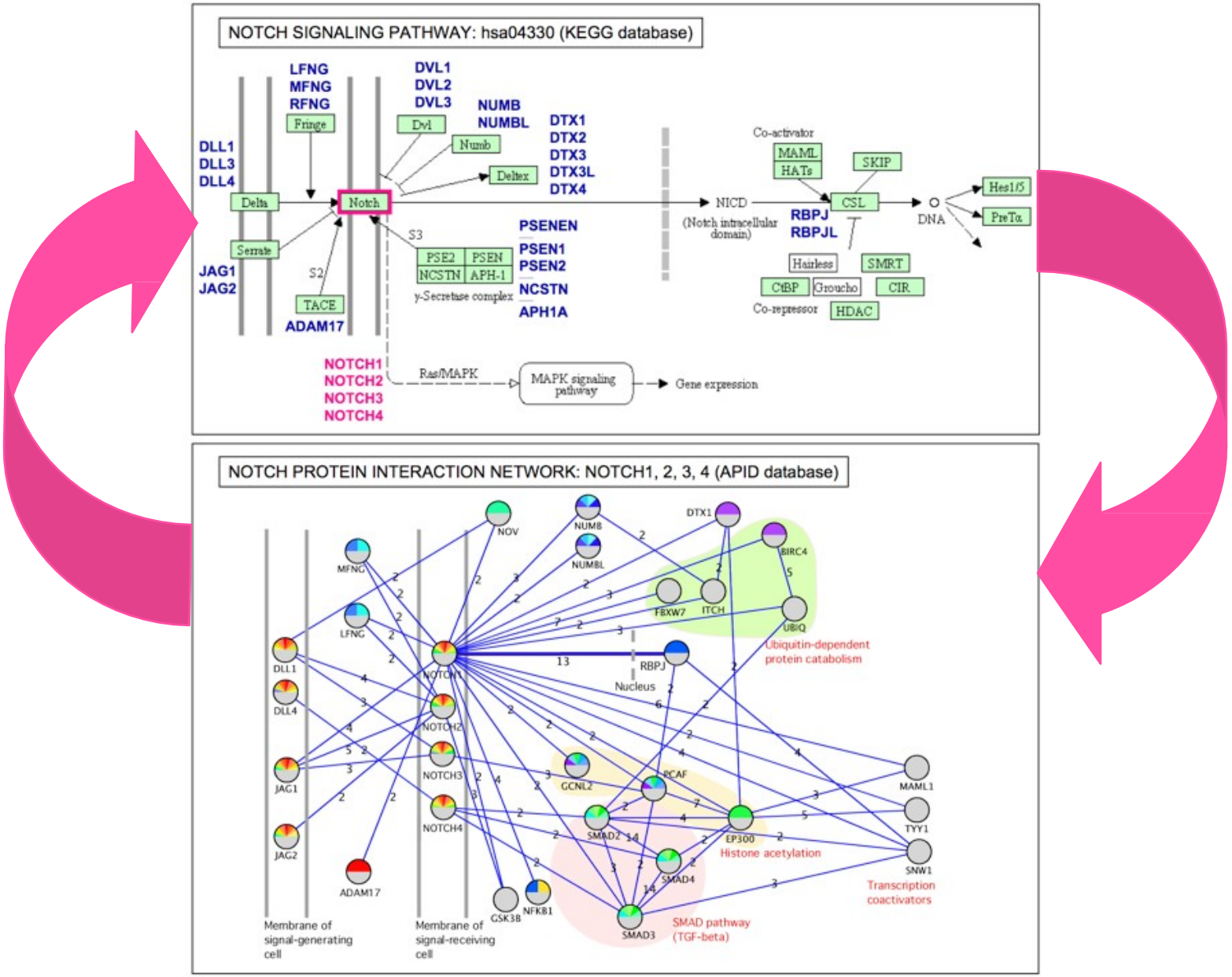
From PPI & pathways to protein networks

build reliable networks with biological meaning: example 3



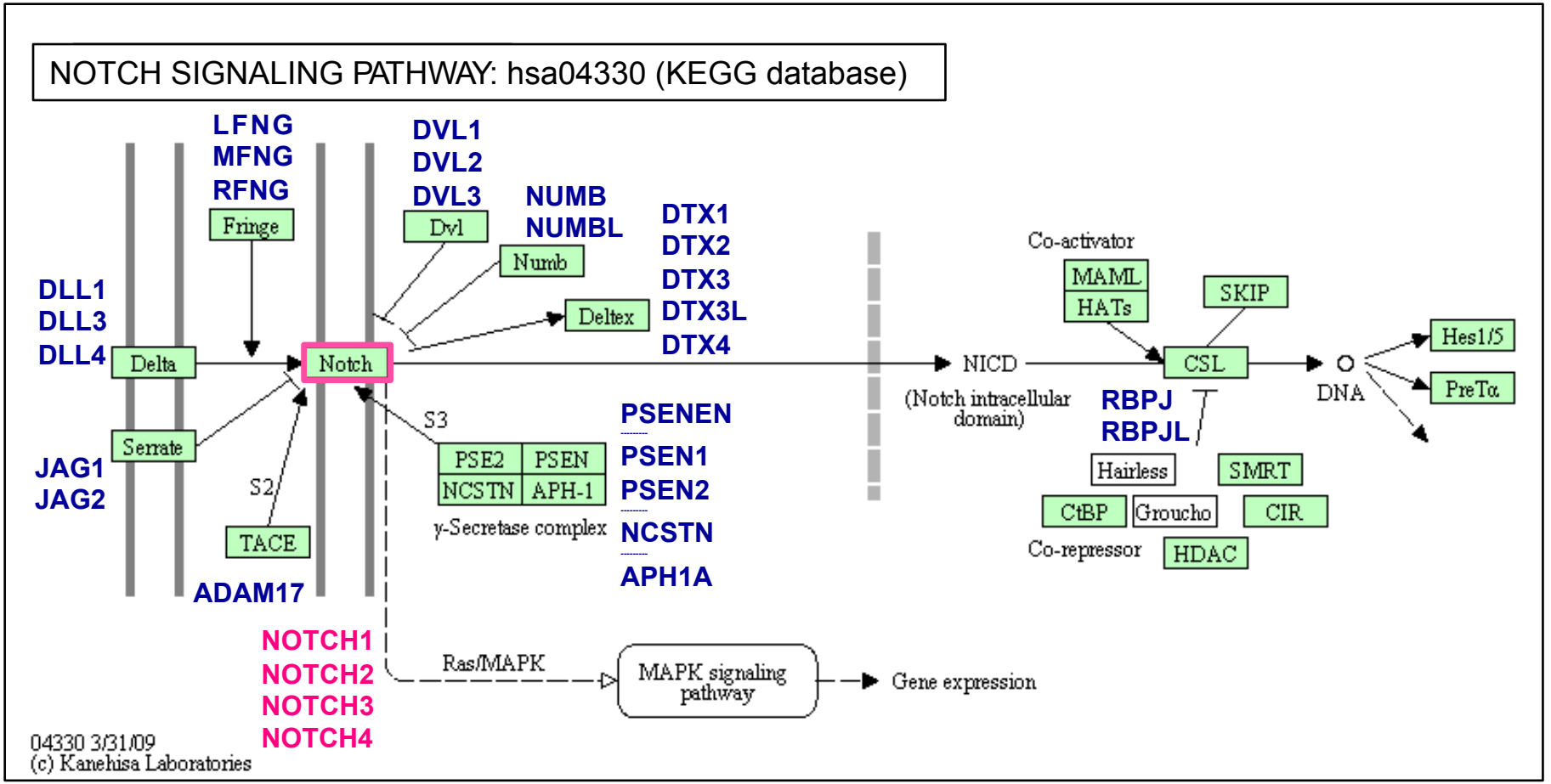
Networks & Pathways

Comparison and combination of these type of complex data



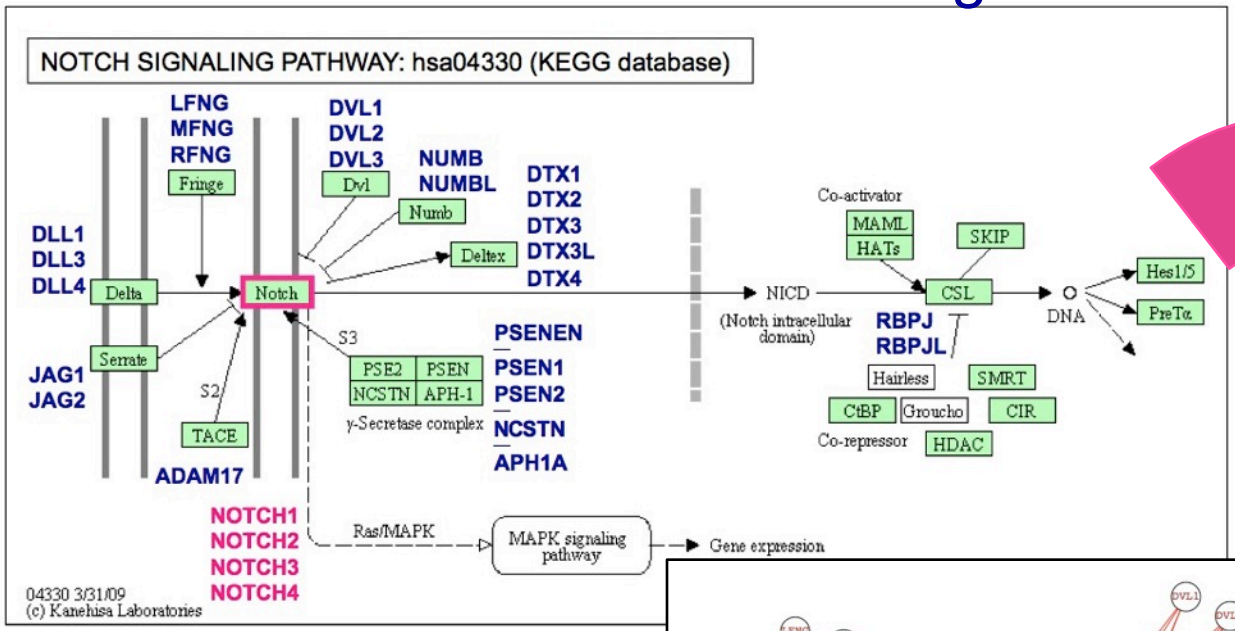
From PPI & pathways to protein networks

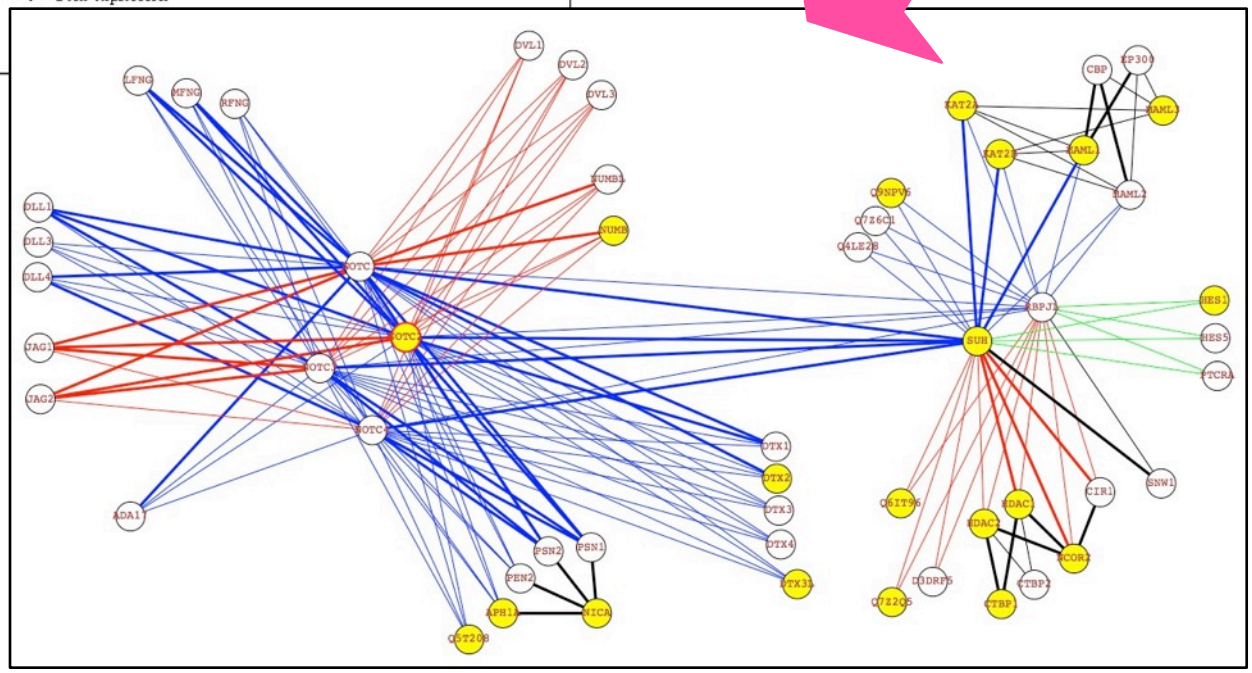
build reliable networks with biological meaning: example 3



From PPI & pathways to protein networks

build reliable networks with biological meaning: example 3



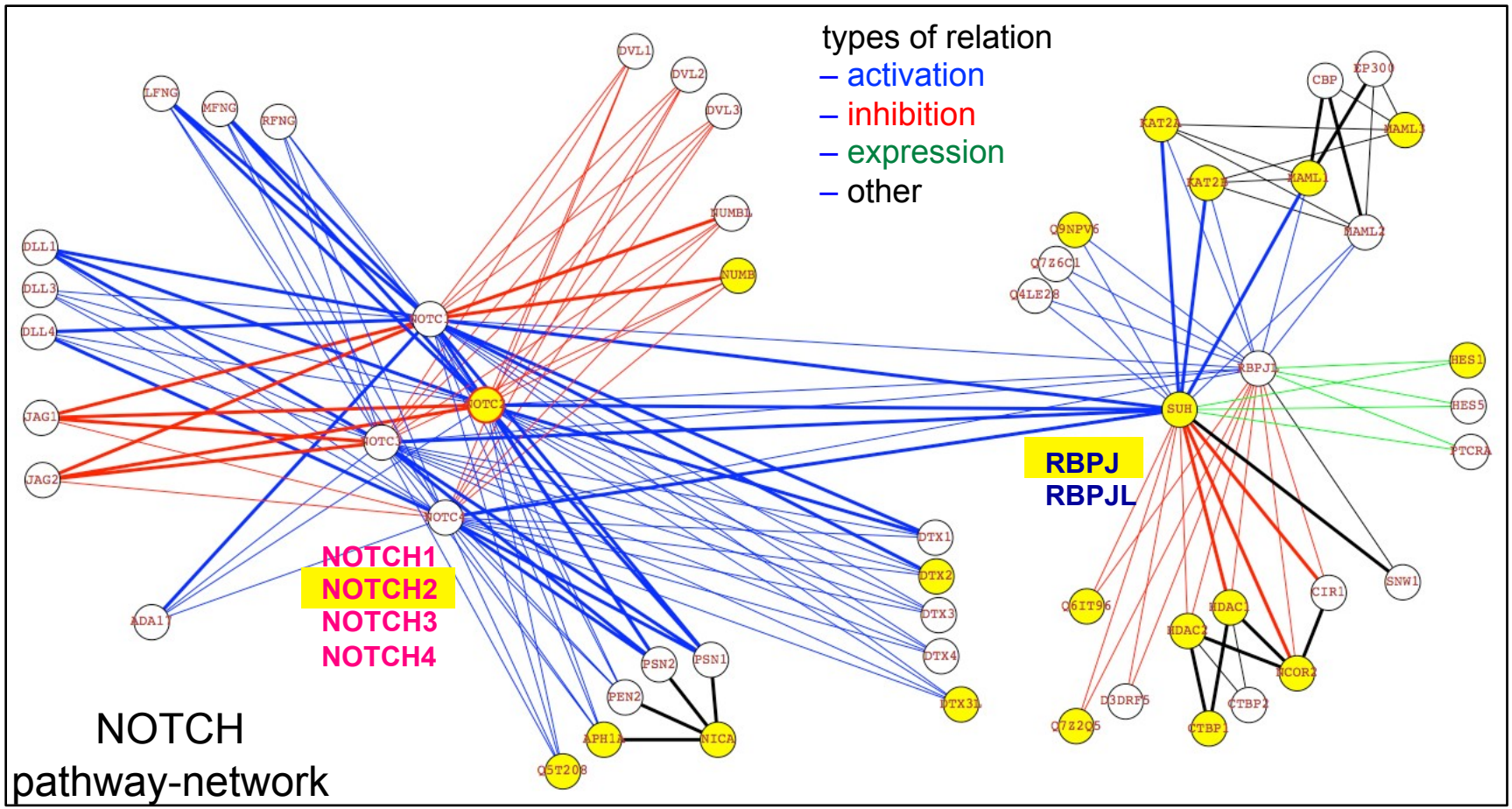


New method:
transfors
a **pathway** in a **network**
showing
the paralogous proteins,
the type of relation
plus
the physical interactions &
the tissue-specificity

From PPI & pathways to protein networks

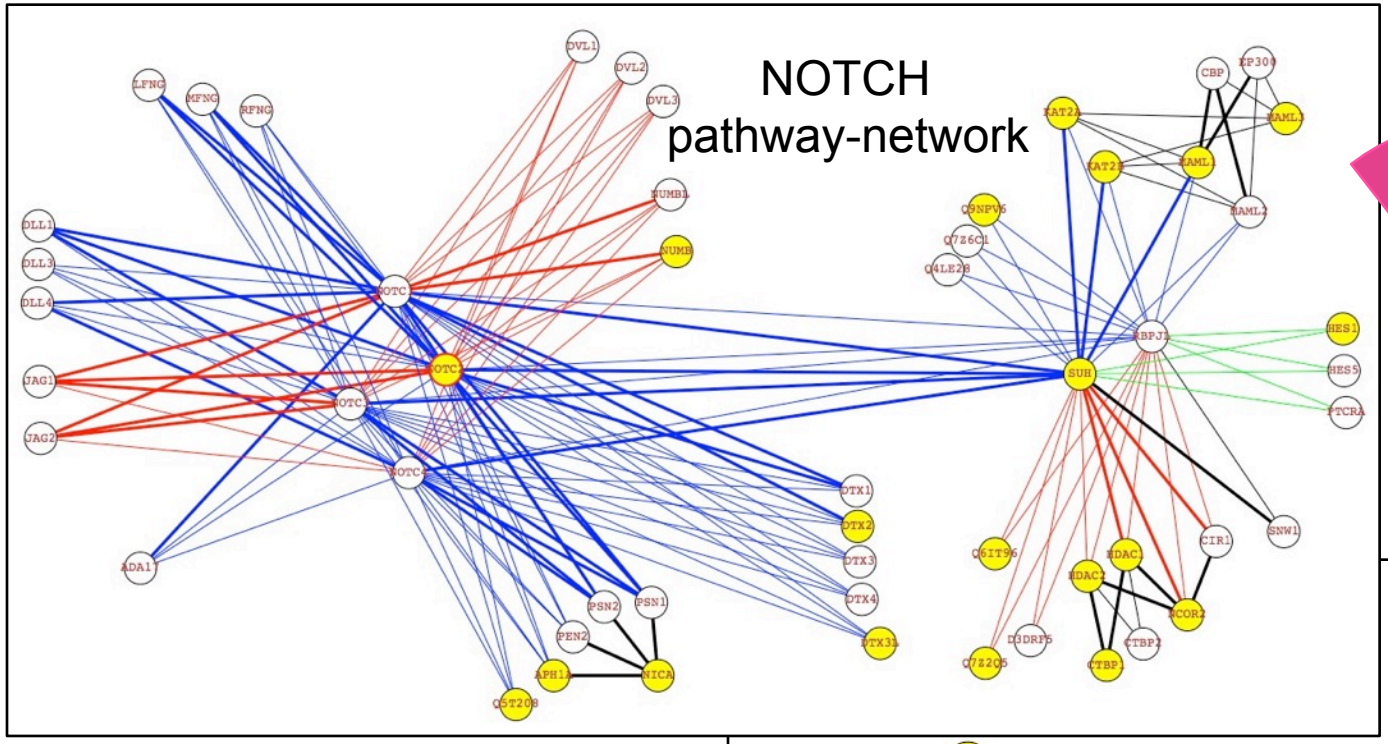
build reliable networks with biological meaning: example 3

NOTCH SIGNALING PATHWAY: hsa04330 (KEGG database) transformed in a NETWORK

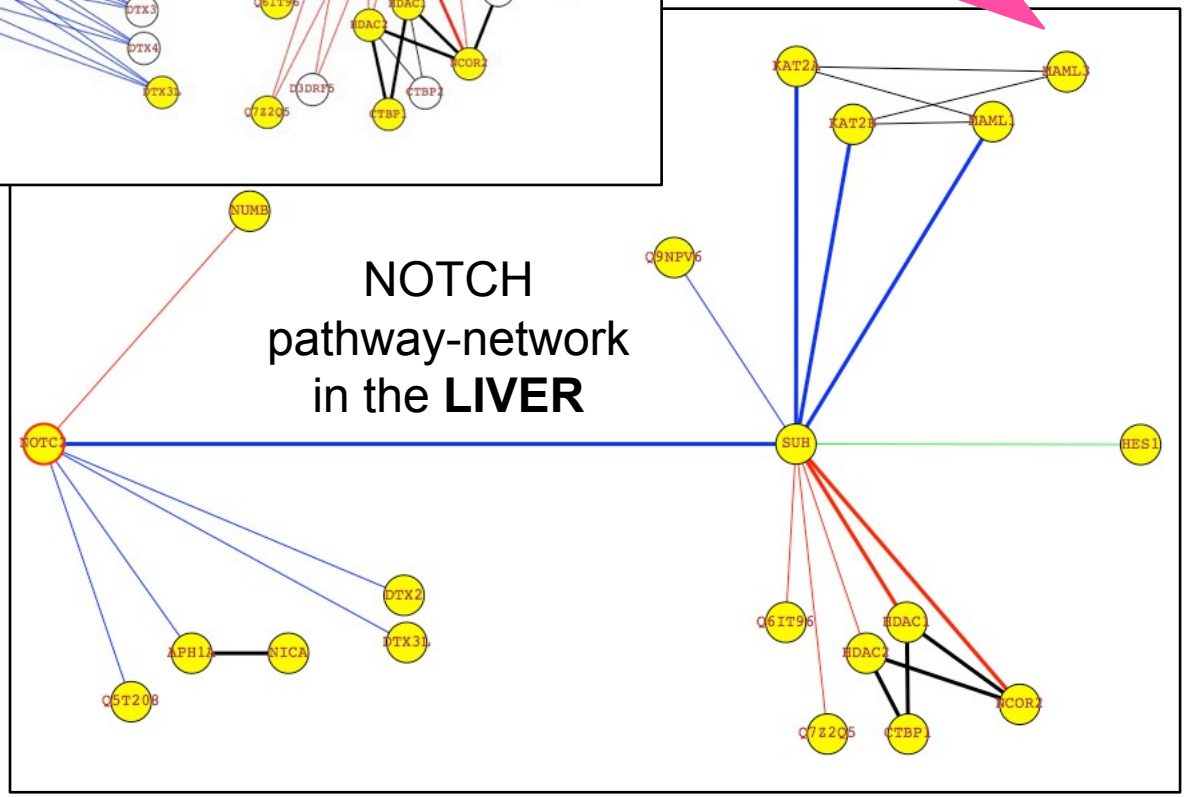


tissue-specificity:
the nodes in yellow correspond to proteins expressed in liver

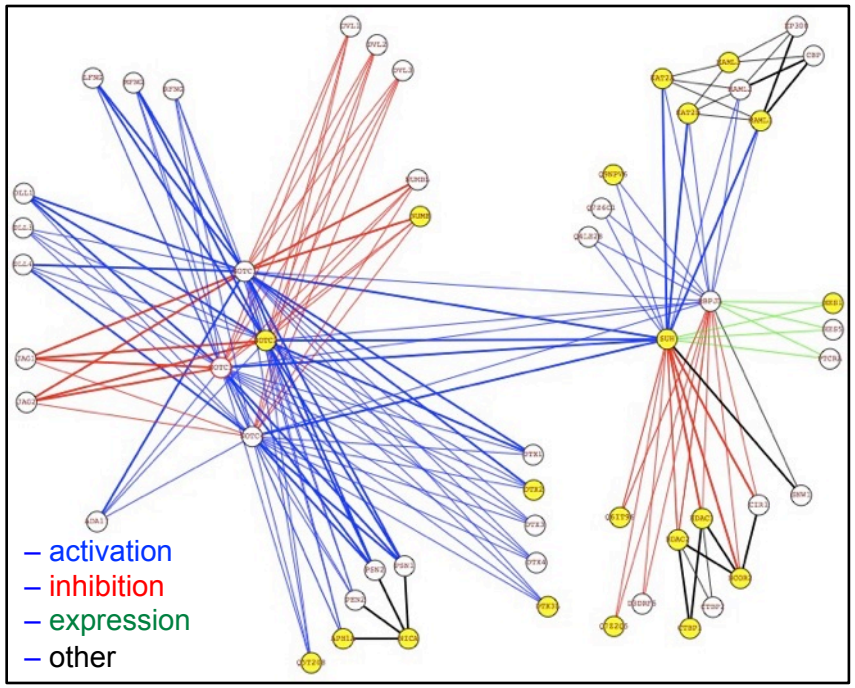
Tissue-specificity: nodes in yellow correspond to proteins expressed in liver

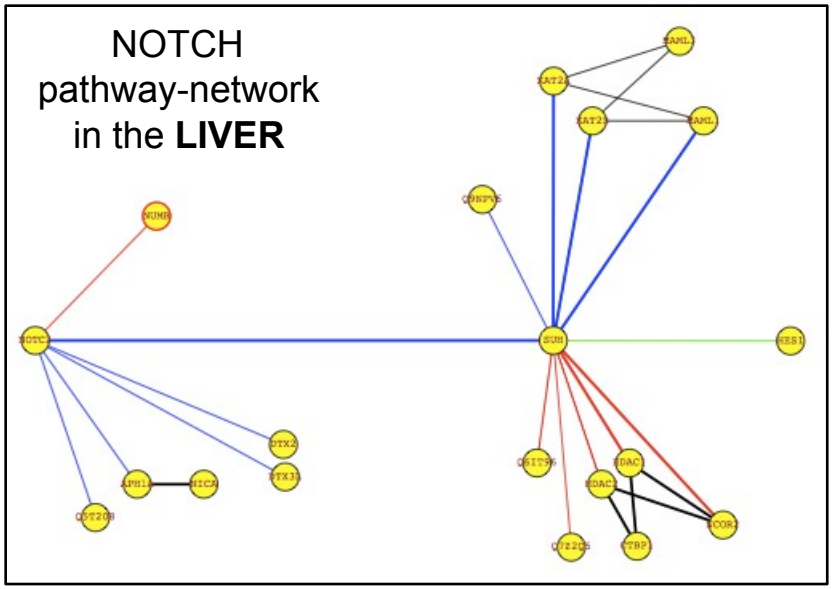


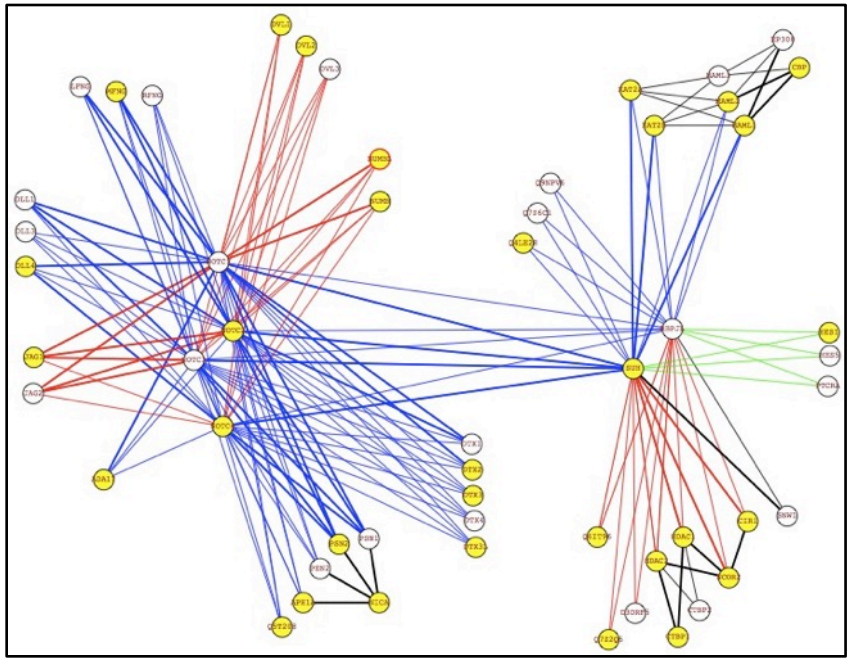
types of relation
– activation
– inhibition
– expression
– other

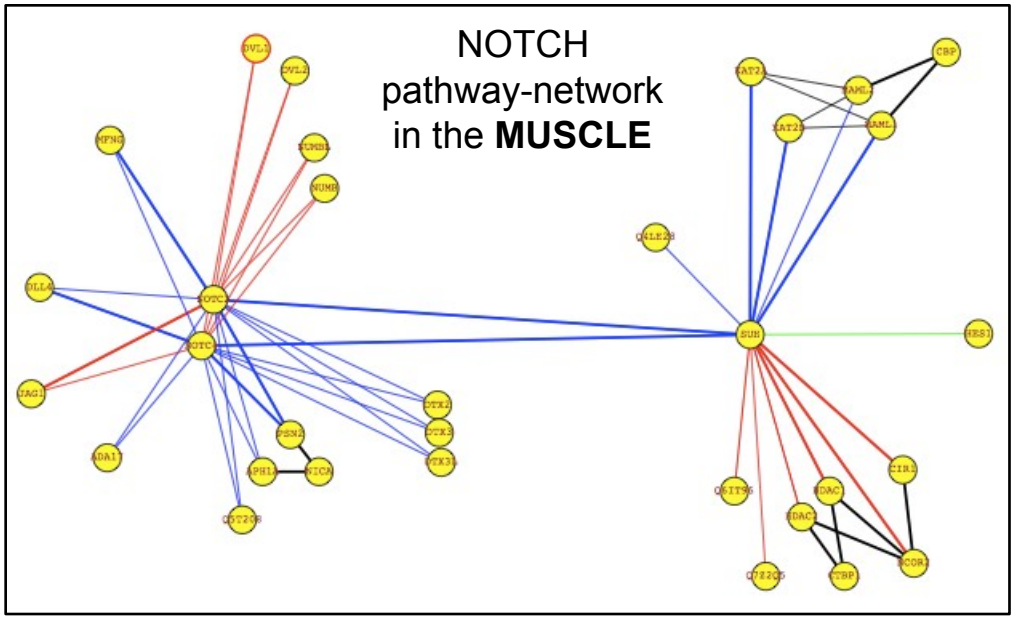


Tissue-specificity: nodes in yellow correspond to proteins expressed in …

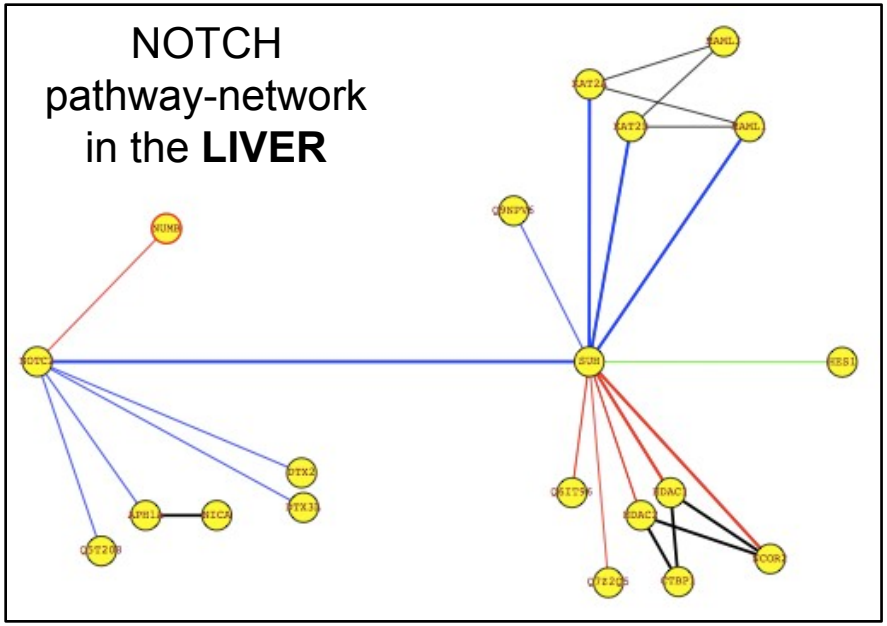


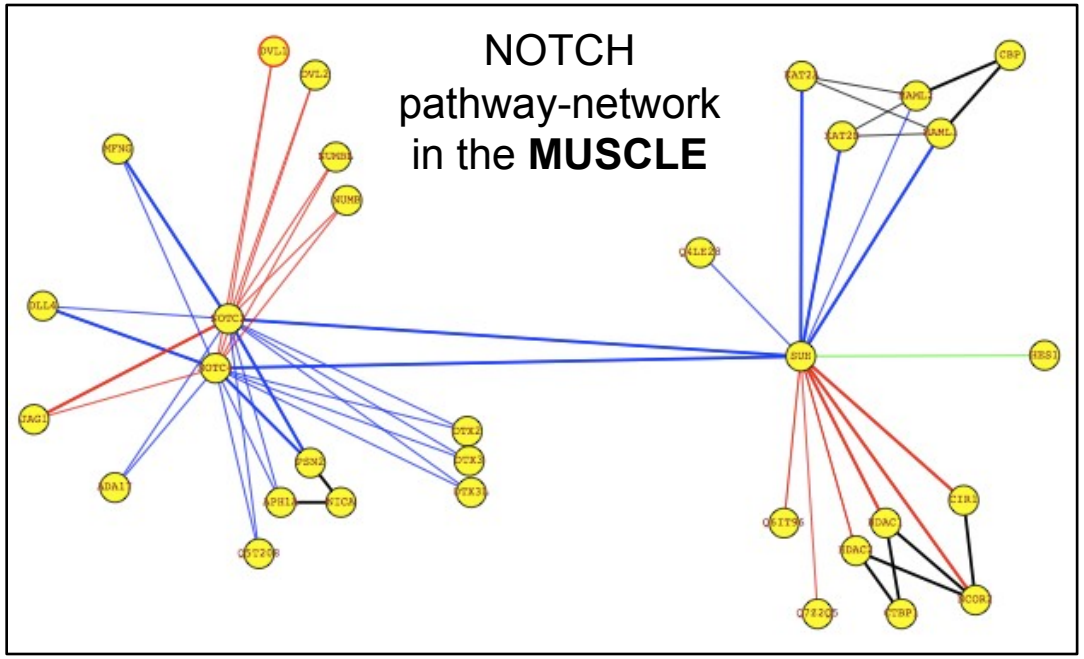






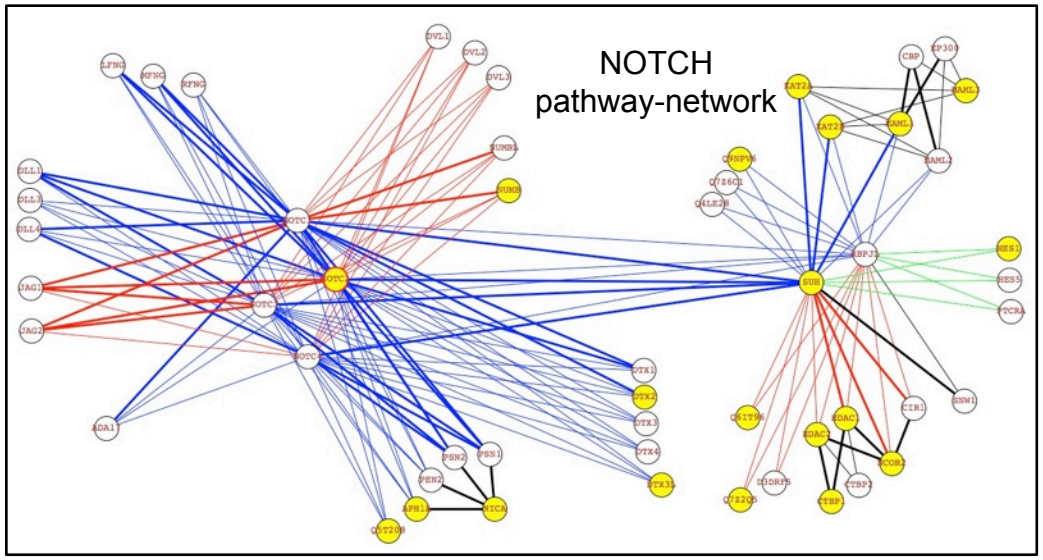
Tissue-specificity: nodes in yellow correspond to proteins expressed in …

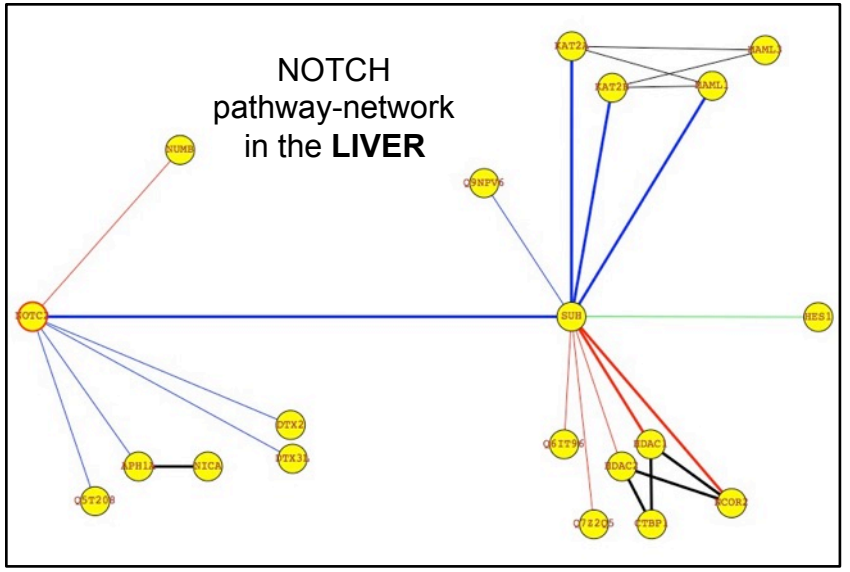


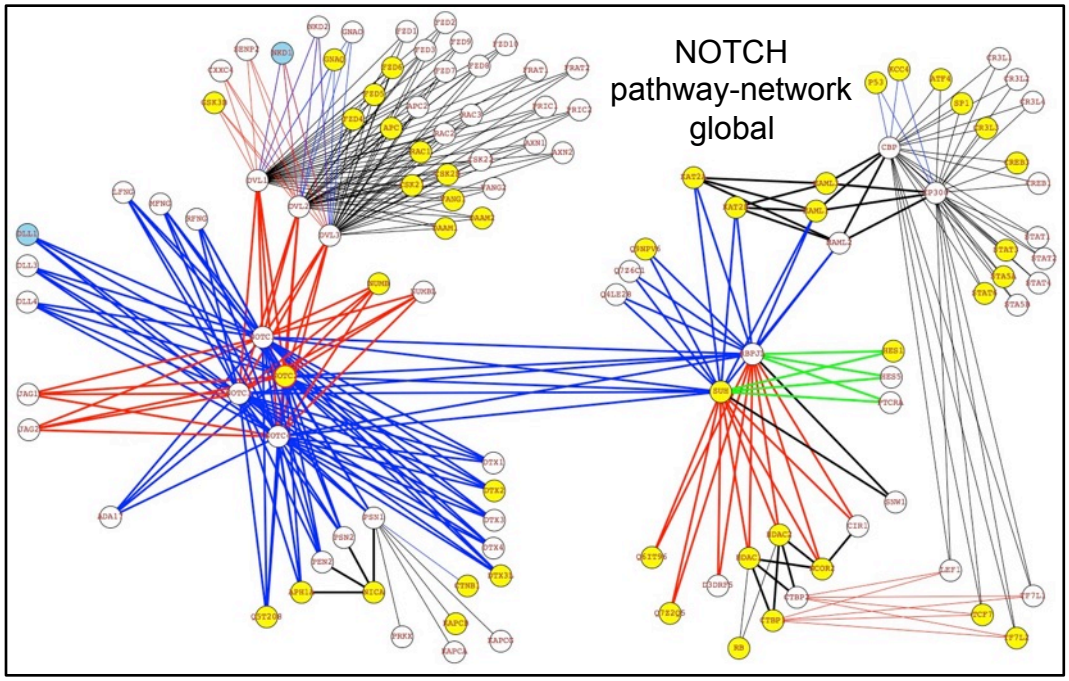


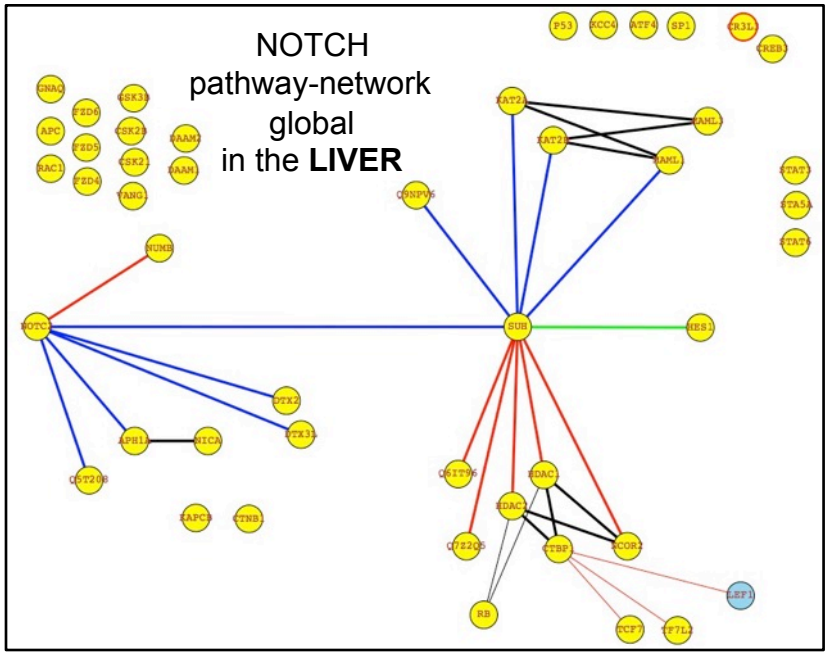
– activation
– inhibition
– expression
– other

From PPI & pathways to protein networks



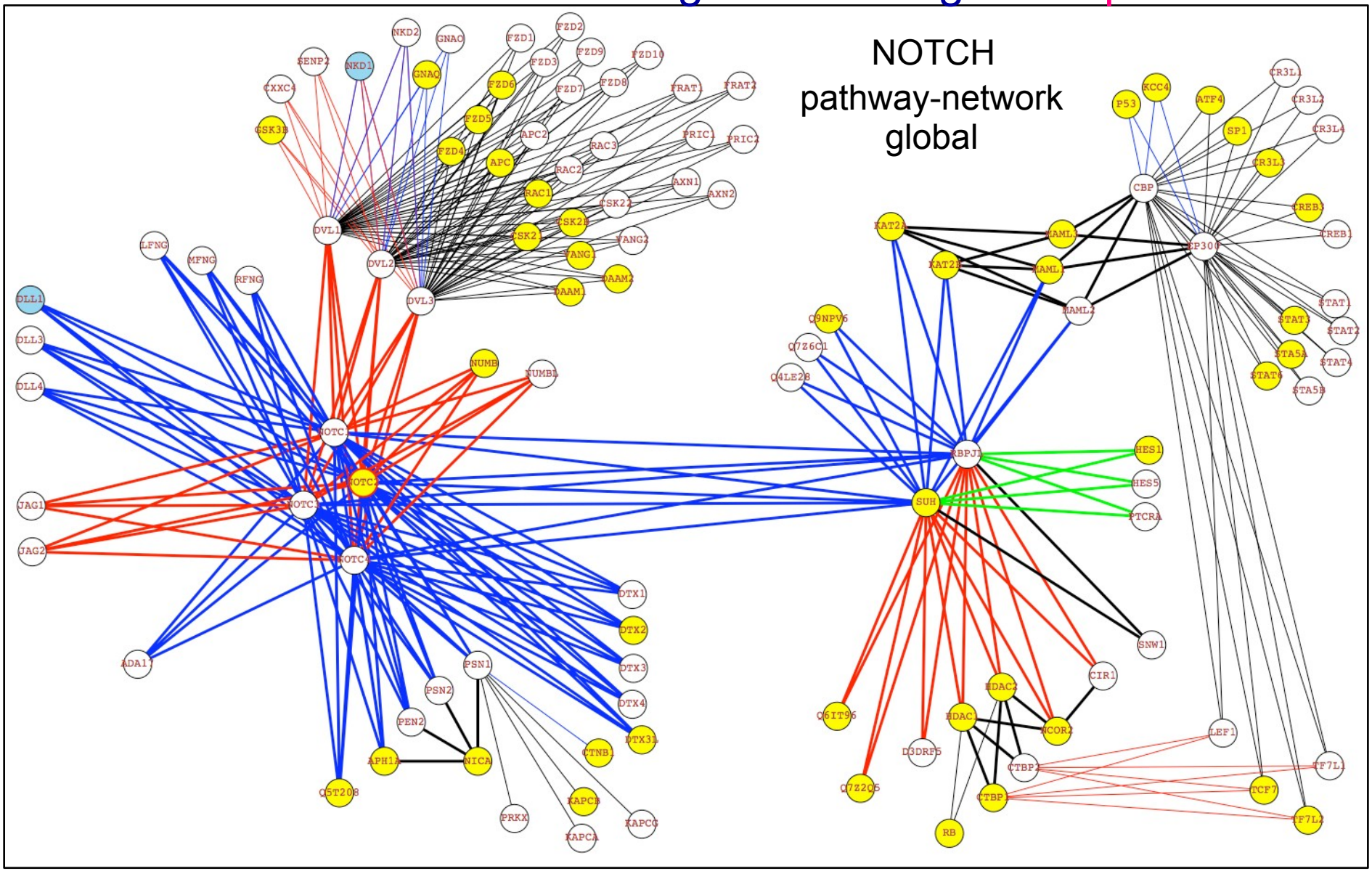






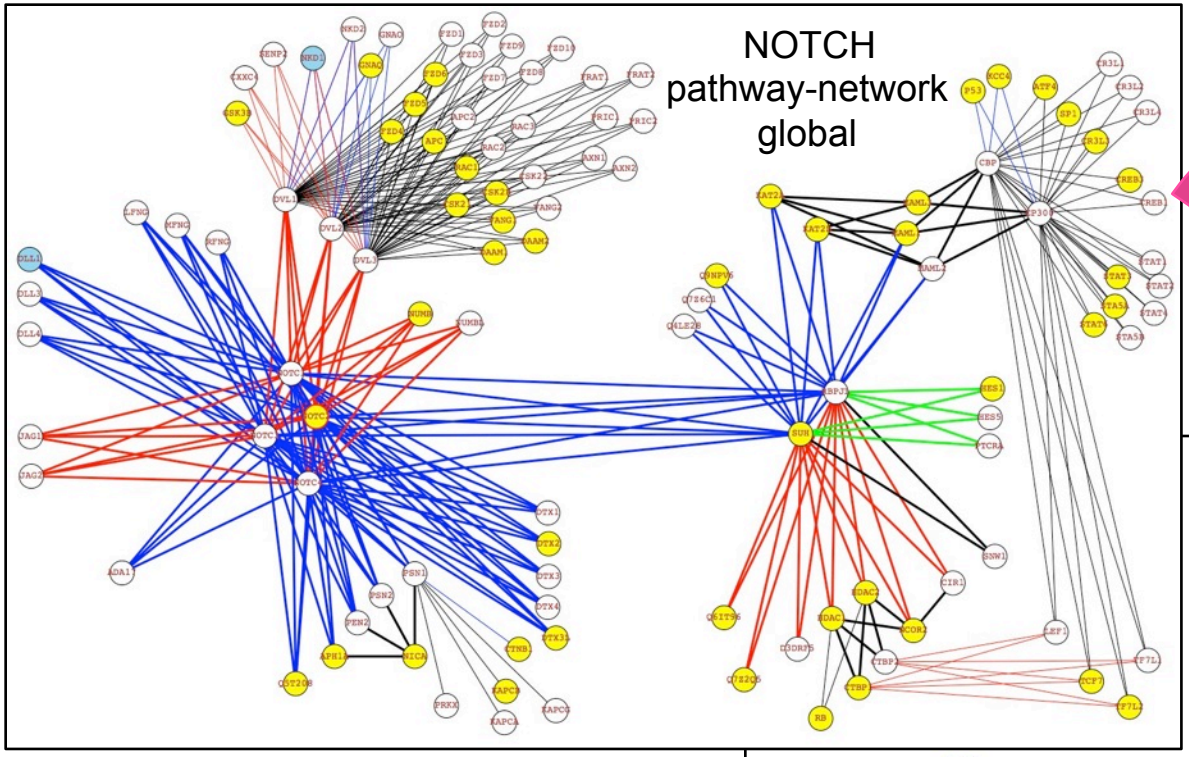
From PPI & pathways to protein networks

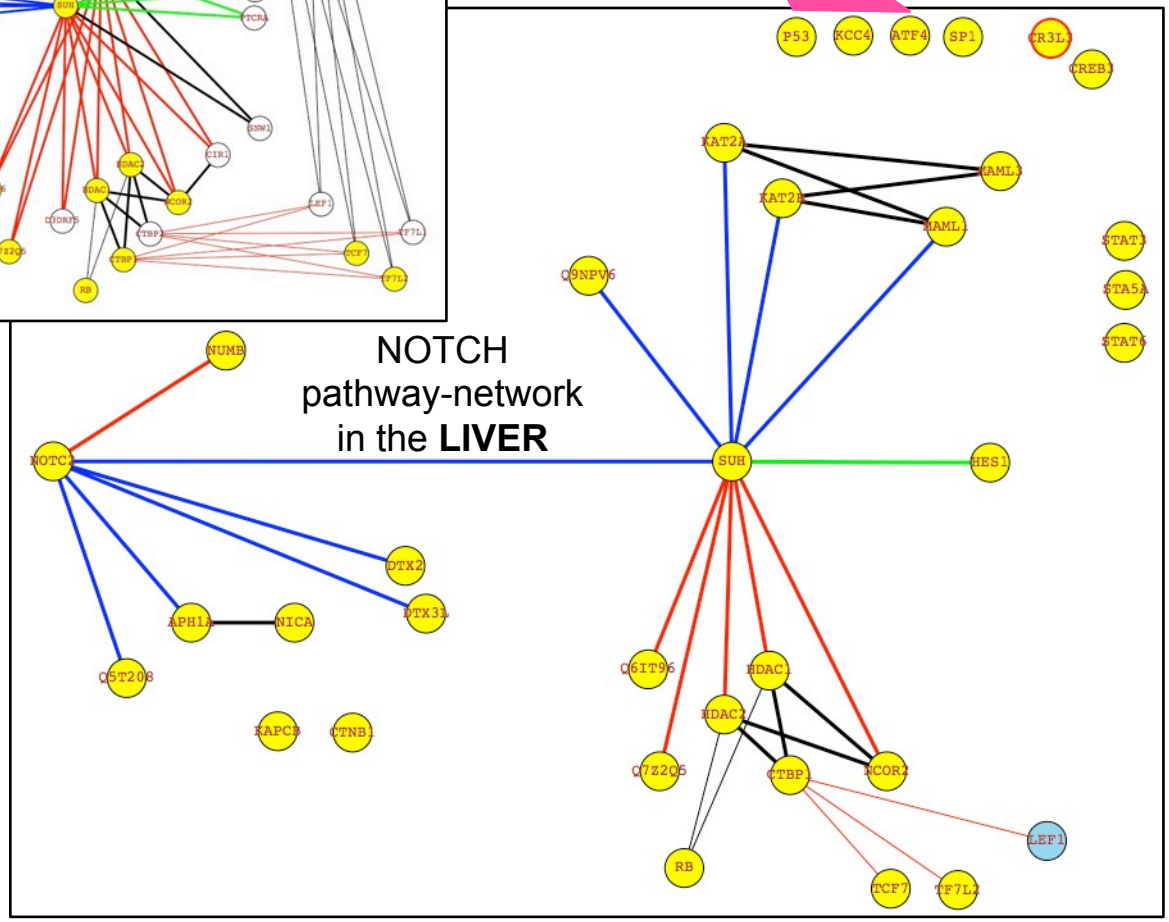
build reliable networks with biological meaning: example 3



includes relations from all signaling pathways

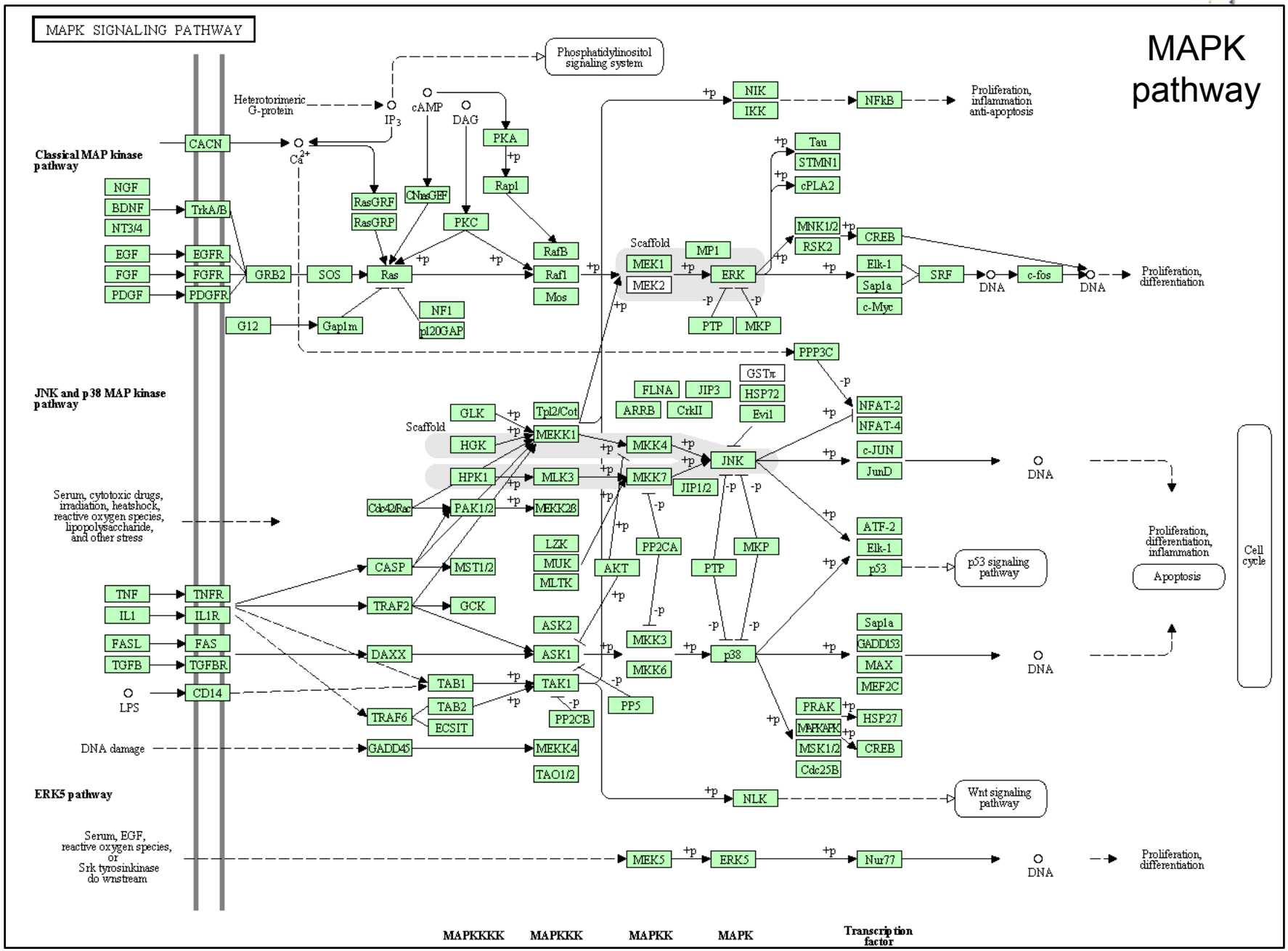
Tissue-specificity: nodes in yellow correspond to proteins expressed in liver





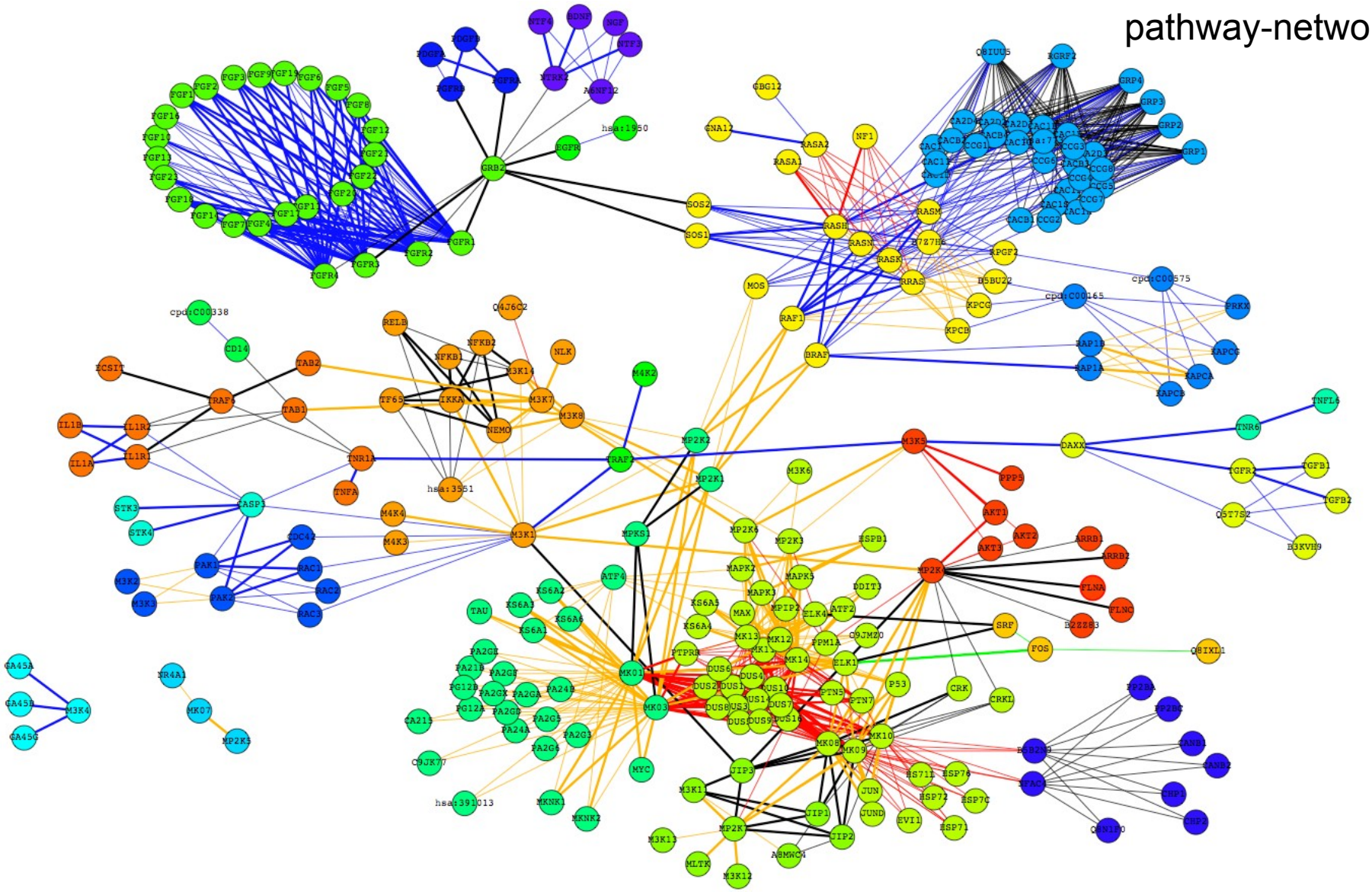
types of relation
– activation
– inhibition
– expression
– other

From PPI & pathways to protein networks



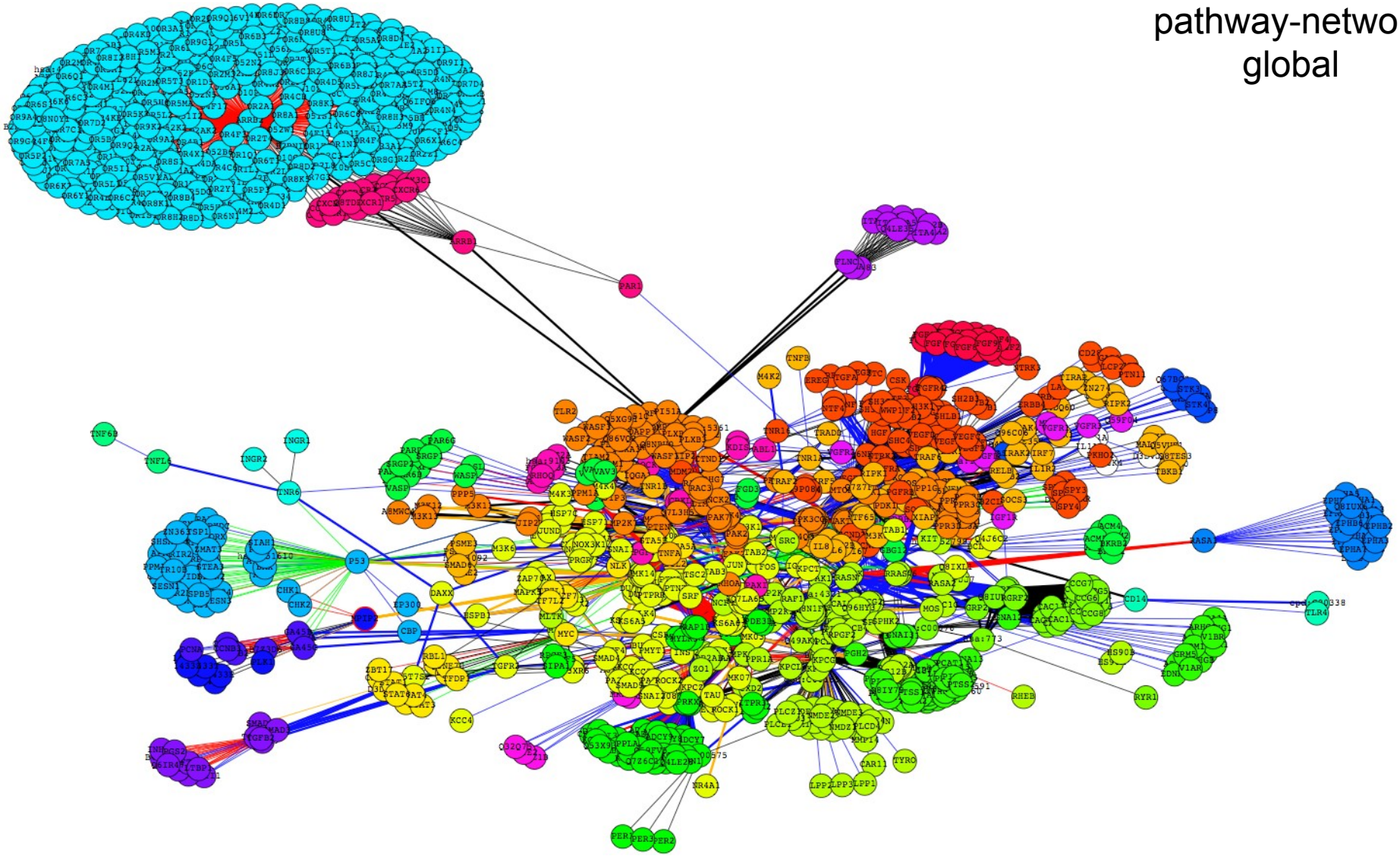
From PPI & pathways to protein networks

MAPK
pathway-network

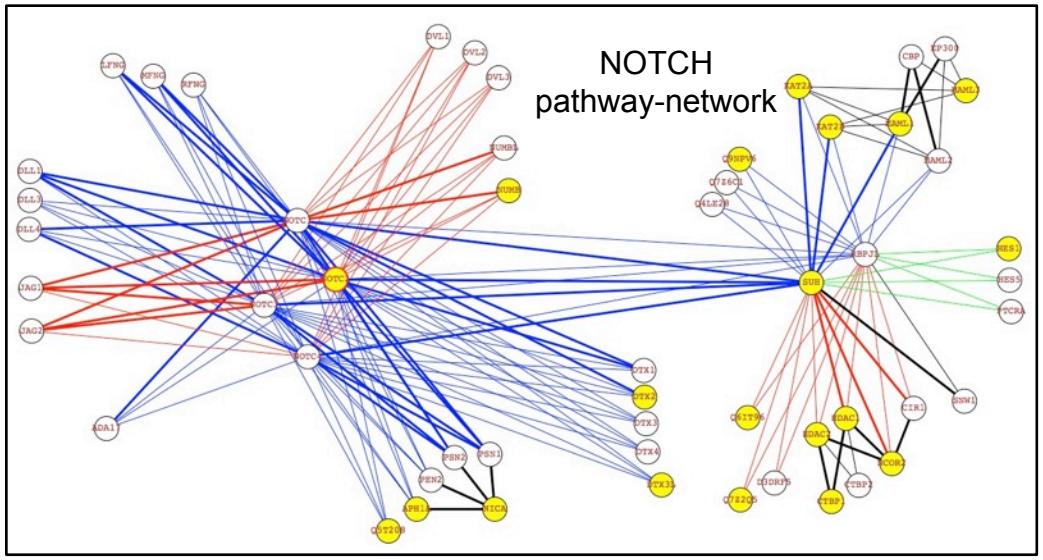


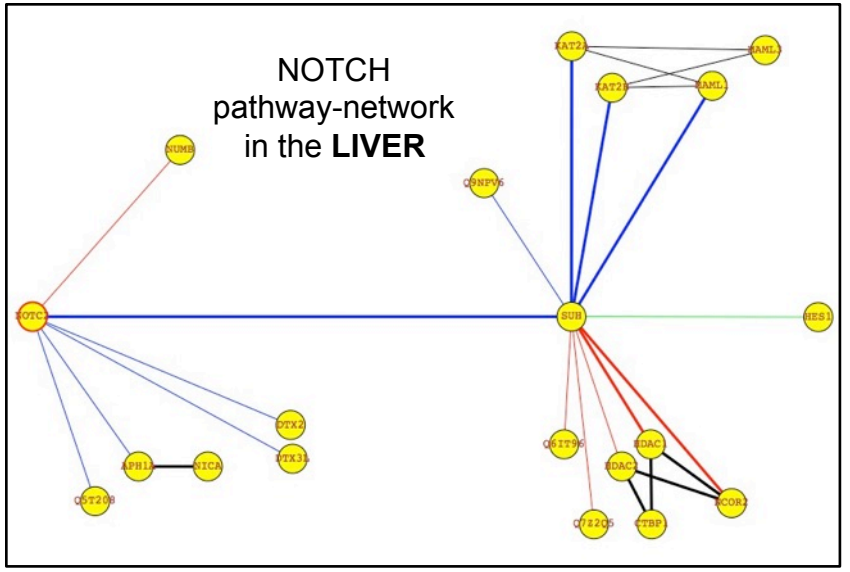
From PPI & pathways to protein networks

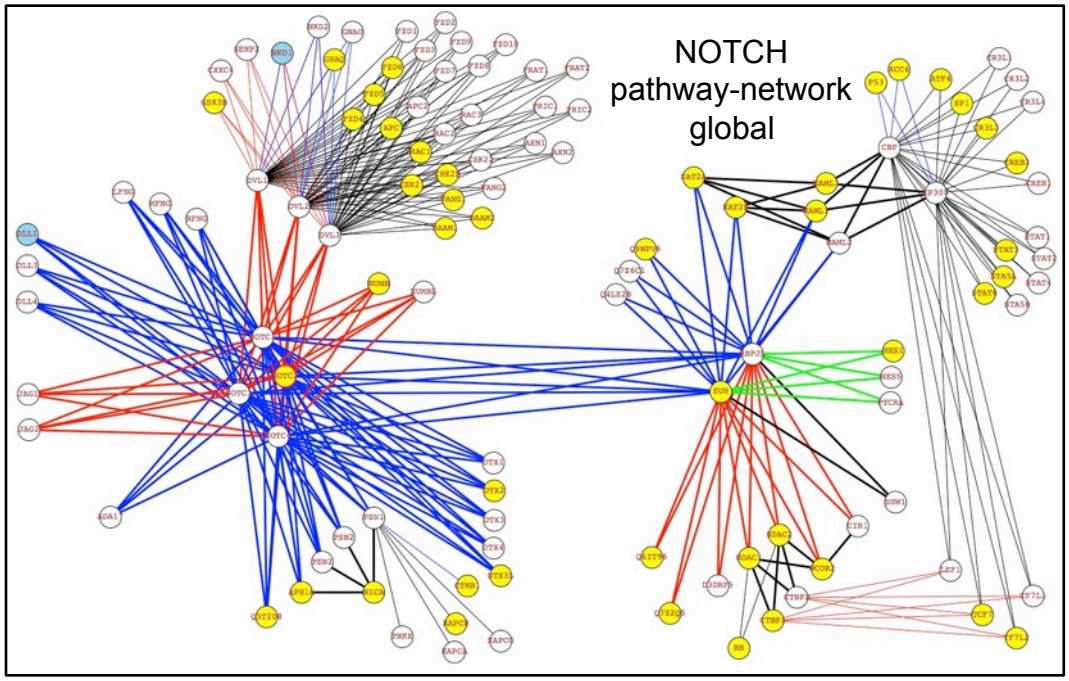
MAPK
pathway-network
global

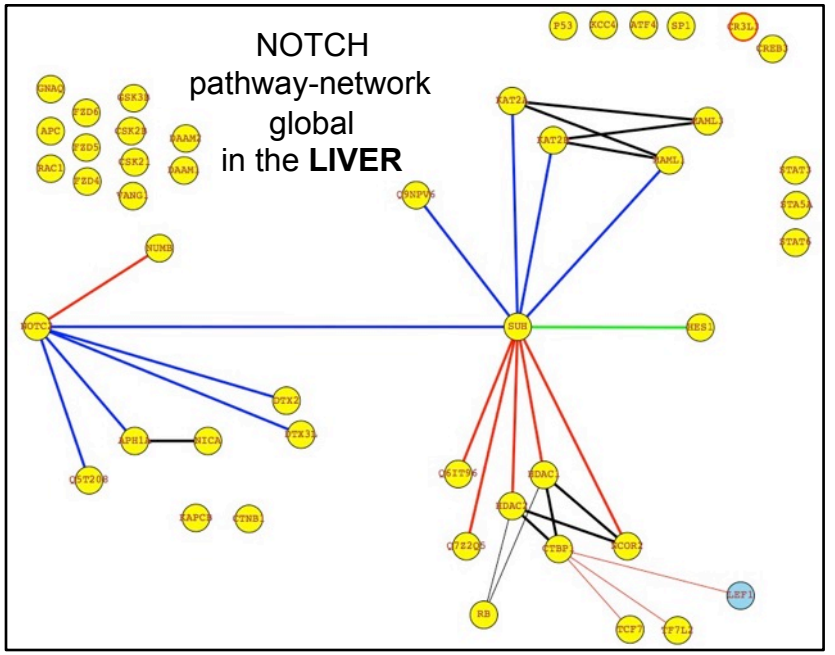


From PPI & pathways to protein networks

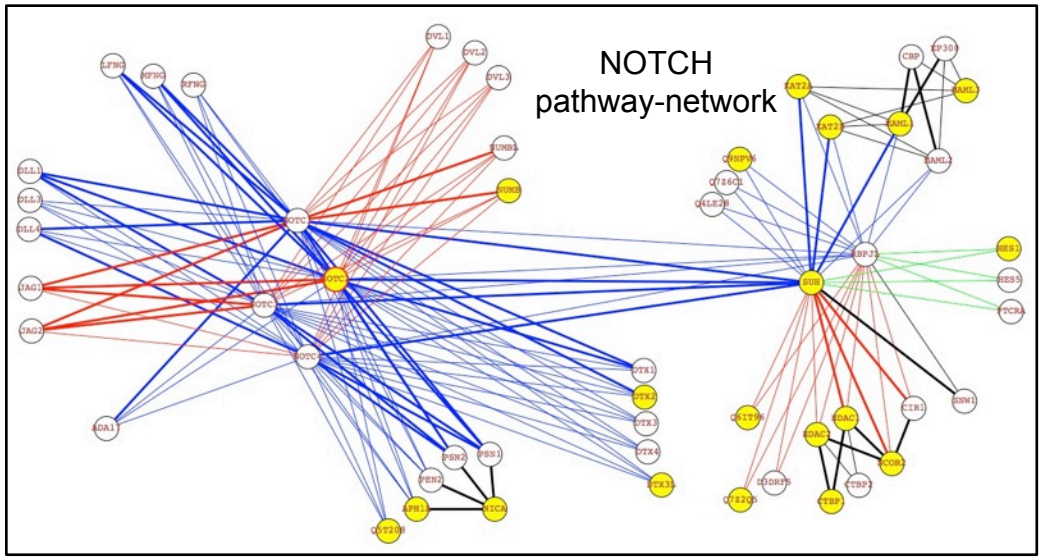








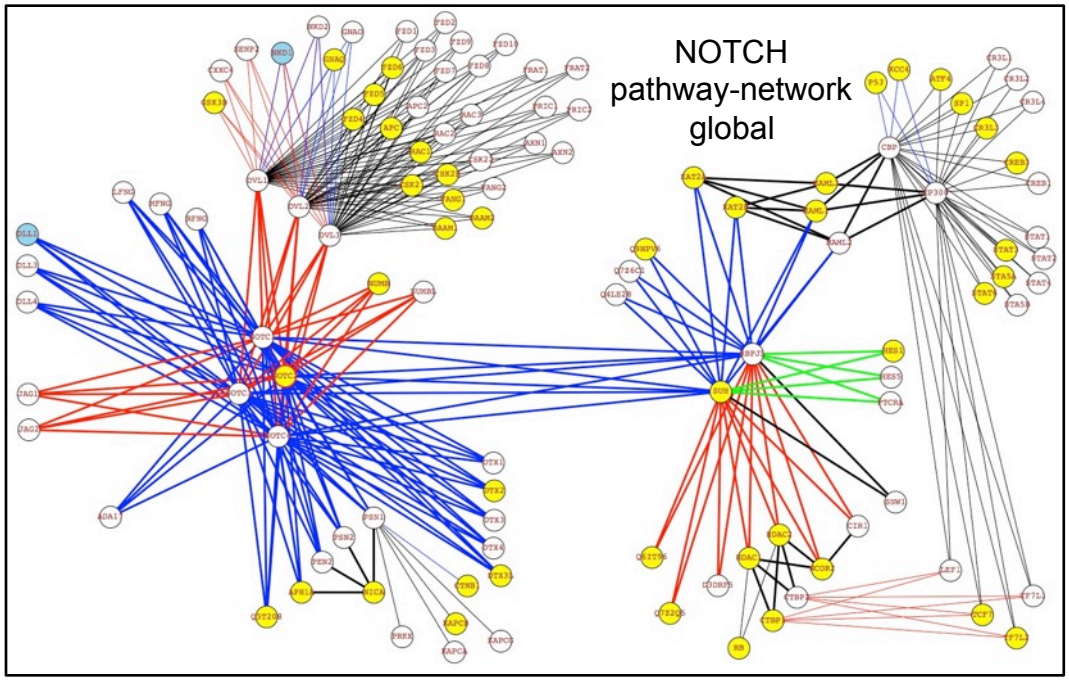
From PPI & pathways to protein networks



The network analysis confirms the **central nodes** of the pathway

Protein_ID	Degree	Betweenness	Eigenvector
[1,] RBPJL_HUMAN	**13**	**187.97**	**0.529**
[3,] **NOTC1_HUMAN**	22	**122.57**	1.0
[6,] **NOTC4_HUMAN**	22	**122.57**	1.0
[7,] HDAC1_HUMAN	5	68.57	0.119
[8,] HDAC2_HUMAN	5	68.57	0.119
[9,] NCOR2_HUMAN	3	37.20	0.025
[10,] NICA_HUMAN	4	6.00	0.000

SUH & **NOTCH2** central nodes of the network



Protein_ID	Degree	Betweenness	Eigenvector
[1,] RBPJL_HUMAN	**13**	**1372.43**	**0.193**
[5,] CTBP1_HUMAN	6	1023.66	0.005
[6,] CTBP2_HUMAN	6	1023.66	0.005
[7,] CBP_HUMAN	24	981.75	0.001
[8,] EP300_HUMAN	24	981.75	0.001
[9,] DVL1_HUMAN	41	956.94	1.0
[10,] DVL2_HUMAN	41	956.94	1.0

HDAC1/2 is enhanced in the global view

From protein interactions to protein networks

build reliable networks with biological meaning: examples

Challenge: obtain and integrate ***omic* data** to build **biological networks** and solve **biological questions**.

Three examples based in PPI data:

1.– Use of PPI data to build **protein networks** and find different **sub-complexes** and **assembly steps:** the PRE-RIBOSOME example.

2.– Use of PPI data to build the **protein network** corresponding to a **molecular machine**: the PROTEASOME example.

3.– Use of PPI data and pathways to build integrated **protein networks** and find **specific connectors** and **hubs**: the NOTCH example.

Networks & Pathways

Comparison and combination of these type of complex data

genes/proteins in networks and **in pathways**

Conclusions

–There are clear links between the proteins working in a **pathway** and the interaction **network** corresponding to such proteins.

–There are useful **databases** and **tools** to explore **pathways** and **networks** using **query sets**: Reactome, KEGG, GeneMANIA, STRING.

–The integration and functional analysis of **pathways** and **networks** can help to find **key genes/proteins** involved in a studied biological state.

THANKS

Bioinformatics and Functional Genomics Research Group
Cancer Research Center (CiC, CSIC/USAL), Salamanca, Spain
http://bioinfow.dep.usal.es

University of Salamanca
founded in 1130
universal chartered in 1216