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
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

DAY 2

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
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.
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.
Omics era: unraveling biological complexity
the paradox of the "genome alone"

genome

expression track
activation track

stable
legacy

phenome

expression track
activation track

facing
reality
Omics era: unraveling biological complexity
the paradox of the "genome alone"

Genome

Expression track
Activation track

Stable legacy

Phenome

Expression track
Activation track

Facing reality

Genome

Living system
## Genomes ...

### Is there a simple “genome factor”? 

<table>
<thead>
<tr>
<th>Organism</th>
<th>Genome Genes</th>
<th>Genome Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacteria</strong></td>
<td>3.000</td>
<td>x2</td>
</tr>
<tr>
<td><strong>Yeast</strong></td>
<td>6.000</td>
<td>x3 x12</td>
</tr>
<tr>
<td><strong>Worm</strong></td>
<td>18.000</td>
<td></td>
</tr>
<tr>
<td><strong>Human</strong></td>
<td>36.000</td>
<td></td>
</tr>
</tbody>
</table>

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

**interactome**

- expression track
- active track

working-moving *machinery*

**phenome**

- expression track
- active track

facing *reality*

---

**genes**

- proteins

**cellular machinery**

= **living system**
Omics era: unraveling biological complexity

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

genome + cellular machinery = living system
Omics era: unraveling biological complexity
interactome ... network of interacting proteins

Babu et al. (2012) Nature
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

Getting connected: analysis and principles of biological networks

Xiaowei Zhu,¹,² Mark Gerstein,³ and Michael Snyder¹,²,⁴

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)

<table>
<thead>
<tr>
<th>Type of network</th>
<th>Species</th>
<th>Number of nodes</th>
<th>Number of interactions</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcription factor-binding network</td>
<td>S. cerevisiae</td>
<td>3528</td>
<td>7419</td>
<td>Yu et al. 2003⁵</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3207</td>
<td>11231</td>
<td>Harbison et al. 2004⁶</td>
</tr>
<tr>
<td>Protein–protein interaction</td>
<td>C. elegans</td>
<td>2788</td>
<td>4441</td>
<td>Stark et al. 2006⁷</td>
</tr>
<tr>
<td></td>
<td>D. melanogaster</td>
<td>7546</td>
<td>25403</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Homo sapiens</td>
<td>7509</td>
<td>20979</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mus musculus</td>
<td>209</td>
<td>393</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S. cerevisiae</td>
<td>5325</td>
<td>51773</td>
<td></td>
</tr>
<tr>
<td>Phosphorylation network</td>
<td>S. cerevisiae</td>
<td>1325</td>
<td>4200</td>
<td>Ptacek et al. 2005</td>
</tr>
<tr>
<td>Metabolic network</td>
<td>E. coli</td>
<td>473</td>
<td>574</td>
<td>Guimera and Nunes Amaral 2005</td>
</tr>
<tr>
<td>Genetic network</td>
<td>S. cerevisiae</td>
<td>646</td>
<td>1149</td>
<td>Tong et al. 2004</td>
</tr>
<tr>
<td></td>
<td>S. cerevisiae</td>
<td>3258</td>
<td>13963</td>
<td>Reguly et al. 2006⁸</td>
</tr>
</tbody>
</table>

⁵Transcriptional factor-binding data collected at rich-media condition.
⁶Transcriptional factor-binding data collected at a variety of growth conditions.
⁷Synthetic lethal interactions among nonessential genes.

Networks

Two major types of networks derived from experimental data

1. Gene Coexpression Networks: \textit{ggcoe}
derived from gene expression profiling and transcriptomic studies

2. Protein-Protein Interaction Networks: \textit{ppi}
derived from proteomic studies
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.

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.

---

**NETWORK TOPOLOGIES**

- **Random network**
  - Degrees follow Poisson (or peaked) distribution
  - Vulnerability to failure

- **Scale-free network** (Biological/cellular networks)
  - Degrees follow power-law distributions
  - Robustness against random failure
  - Vulnerability to targeted attacks

- **Hierarchical network** (Many types of real networks)
  - Degrees follow power-law distributions
  - Account for modularity, local clustering, and scale-free topology
  - High clustering coefficient (C)

---

*From Seebacher & Gavin (2011) Cell*
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.
**Omics era:** unraveling biological complexity

mapping **biological networks**

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

Representation and visualization of networks

![Network diagram](image)

From *Merico, Gfeller & Bader (2010) Nature Biotechnology*
Omics era: unraveling biological complexity mapping biological networks

Example of *gene to gene coexpression network*: ggcoe
**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
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**
   \[ k_i = \text{number of links connected to node } i \]

B. **Distance**
   \[ d_{ij} = \text{shortest path length between node } i \text{ and } j \]

C. **Diameter**
   \[ D = \max \{ d_{ij} | i, j \in N \} \]
   \[ N : \text{all nodes in the network} \]

D. **Clustering Coefficient**
   \[ c_i = \frac{2e_i}{k_i(k_i-1)} \]
   \[ e_i : \text{number of existing links (labeled in red) among the } k_i \text{ nodes that connect to node } i \]

E. **Betweenness**
   \[ b_i = \sum_{ij} p_{ij}(l) / p_{ij} \]
   \[ p_{ij} : \text{number of shortest paths between } i \text{ and } j \]
   \[ p_{ij}(l) : \text{number of shortest paths between } i \text{ and } j \text{ going through node } l \]

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 neighbor's connectivity

**NETWORK MEASURES**

<table>
<thead>
<tr>
<th>Degree/ connectivity (k)</th>
<th>Clustering coefficient/ interconnectivity (C)</th>
<th>Assortativity/average nearest neighbor’s connectivity (NC)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Diagram" /></td>
<td><img src="image2" alt="Diagram" /></td>
<td><img src="image3" alt="Diagram" /></td>
</tr>
<tr>
<td>$k_A = \text{Nb of edges through A} = 5$</td>
<td>$C_A = \frac{\text{Actual links between A's neighbors (black)}}{\text{Possible links between A's neighbors (orange)}}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$C_A = n_A / [k_A(k_A-1)/2]$</td>
<td>$NC_A = (k_B + k_C + k_D + k_E + k_J) / 5$</td>
</tr>
<tr>
<td></td>
<td>$= 2/[4(4-1)/2] = 0.333$</td>
<td>$= (5+2+2+3+1)/5 = 2.6$</td>
</tr>
</tbody>
</table>

*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

<table>
<thead>
<tr>
<th>Shortest path (SP) between two nodes</th>
<th>Betweenness/centrality (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP_{FH}=(F,D,A,B,H)=4</td>
<td>B_4=Fraction of SPs passing through A =0.090</td>
</tr>
</tbody>
</table>
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/


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)

Important publications:
Bioinformatics (2011)
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>
<thead>
<tr>
<th>Feature</th>
<th>CY</th>
<th>GM</th>
<th>VA</th>
<th>OS</th>
<th>CD</th>
<th>AR</th>
<th>IN</th>
<th>GG</th>
<th>PI</th>
<th>PR</th>
<th>BL</th>
<th>PA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free for academic use</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Free for commercial use</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Open source</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Curated pathway/network content</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Standard file format support</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>User-defined networks/pathways</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Functionality to infer new pathways</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>GO/pathway enrichment analysis</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Automated graph layout</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Complex criteria for visual properties</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Multiple visual styles</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Advanced node selection</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Customizable gene/protein database</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Rich graphical annotation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Statistical network analysis</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Extensible functionality: plugins or API</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Quantitative pathway simulation</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

From *Cline et al. (2007) Nature Protocols*
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
9626 downloads
(October 2015)
Protein-Protein Interactions (PPIs)  
two publications

**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

**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

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


WEB References

http://bioinfow.dep.usal.es/apid
Networks & Pathways
Comparison and combination of these type of complex data

¿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/
Hands-on: Practical Examples

Explore web resources & tools: GeneMANIA, STRING

Protein_SETs_2014.xls
(175g human, 59g5pc yeast)
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**: \textit{ggcoe}
derived from gene expression profiling and transcriptomic studies

2. – **Protein-Protein Interaction Networks**: \textit{ppi}
derived from proteomic studies
Human coexpression studies

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

Joshua M. Stuart,\(^1\star\)† Eran Segal,\(^2\star\) Daphne Koller,\(^2\)†
Stuart K. Kim\(^3\)†

**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

**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\(^a,\star\)
Human coexpression
low signal & high noise

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

Joshua M. Stuart,1*† Eran Segal,2* Daphne Koller,2† Stuart K. Kim3†

Stuart et al. (2003) Science
Human coexpression

low signal & high noise
A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules

Joshua M. Stuart,† Eran Segal,‡ Daphne Koller,‡ Stuart K. Kim

Coexpression Analysis of Human Genes Across Many Microarray Data Sets

Homin K. Lee, Amy K. Hsu,† Jon Sajdak,† Jie Qin,† and Paul Pavlidis†, 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

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

Obi L. Griffith, Erin D. Pleasance, Debra L. Fulton, Mehrdad Oveis, Martin Ester, Asim S. Siddiqui, Steven J.M. Jones,a,*
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 Pavlidis1,3,4

1Columbia Genome Center, 2College of Physicians and Surgeons, and 3Department of Biomedical Informatics, Columbia University, New York, New York 10032, USA
Lee et al. (2004) Genome Research

≈ 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?

<table>
<thead>
<tr>
<th>Reference</th>
<th>Samples</th>
<th>Genes</th>
<th>Raw links</th>
<th>Reference</th>
<th>Samples</th>
<th>Genes</th>
<th>Raw links</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Alizadeh et al. 2000)</td>
<td>96</td>
<td>1759</td>
<td>25748</td>
<td>(Nielsen et al. 2002)</td>
<td>46</td>
<td>3359</td>
<td>25175</td>
</tr>
<tr>
<td>(Allander et al. 2001)</td>
<td>19</td>
<td>1205</td>
<td>1251</td>
<td>(Perou et al. 1999)</td>
<td>26</td>
<td>3027</td>
<td>42105</td>
</tr>
<tr>
<td>(Armstrong et al. 2002)</td>
<td>72</td>
<td>8242</td>
<td>213456</td>
<td>(Perou et al. 2000)</td>
<td>84</td>
<td>5701</td>
<td>167826</td>
</tr>
<tr>
<td>(Bhattacharjee et al. 2001)</td>
<td>203</td>
<td>8243</td>
<td>243303</td>
<td>(Pomeroy et al. 2002)</td>
<td>90</td>
<td>5418</td>
<td>85909</td>
</tr>
<tr>
<td>(Bittner et al. 2000)</td>
<td>38</td>
<td>4382</td>
<td>16141</td>
<td>(Ramaswamy et al. 2001)</td>
<td>255</td>
<td>9528</td>
<td>372500</td>
</tr>
<tr>
<td>(Butte et al. 2000)</td>
<td>68</td>
<td>4906</td>
<td>81755</td>
<td>(Rickman et al. 2001)</td>
<td>51</td>
<td>5418</td>
<td>60169</td>
</tr>
<tr>
<td>(Chang et al. 2001)</td>
<td>32</td>
<td>634</td>
<td>129814</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Chaussabel et al. 2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Chen et al. 2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Cheok et al. 2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Dhanasekaran et al. 2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Diewert et al. 2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Dyrrkjot et al. 2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Dyrrkjot et al. 2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Erraji-BenCheikh et al. 2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Garber et al. 2000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Greenberg et al. 2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Gruber et al. 2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Greenberg et al. 2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Huang et al. 2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Huang et al. 2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Jazaeri et al. 2002)</td>
<td>81</td>
<td>3044</td>
<td>40167</td>
<td>(Welsh et al. 2001)</td>
<td>12</td>
<td>6245</td>
<td>2670</td>
</tr>
<tr>
<td>(Khan et al. 2001)</td>
<td>88</td>
<td>1952</td>
<td>19868</td>
<td>(Welsh et al. 2001)</td>
<td>49</td>
<td>5418</td>
<td>52459</td>
</tr>
<tr>
<td>(Khatua et al. 2003)</td>
<td>13</td>
<td>8257</td>
<td>10072</td>
<td>(Welsh et al. 2001)</td>
<td>55</td>
<td>8258</td>
<td>260155</td>
</tr>
<tr>
<td>(Leung et al. 2002)</td>
<td>126</td>
<td>12657</td>
<td>993195</td>
<td>(West et al. 2001)</td>
<td>49</td>
<td>5418</td>
<td>84842</td>
</tr>
<tr>
<td>(Luo et al. 2001)</td>
<td>25</td>
<td>4354</td>
<td>14873</td>
<td>(Whitfield et al. 2002)</td>
<td>114</td>
<td>12801</td>
<td>1547199</td>
</tr>
<tr>
<td>(Ma et al. 2003)</td>
<td>61</td>
<td>1569</td>
<td>10086</td>
<td>(Yoo et al. 2002)</td>
<td>248</td>
<td>8257</td>
<td>257979</td>
</tr>
<tr>
<td>(MacDonald et al. 2001)</td>
<td>31</td>
<td>1309</td>
<td>3179</td>
<td>(Yoon et al. 2002)</td>
<td>12</td>
<td>5418</td>
<td>53305</td>
</tr>
</tbody>
</table>
Sample bias
“malignant data”

microarray datasets in
Lee et al. (2004) Genome Research

lymphoma
gastrointestinal stromal tumor (GIST)
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

NCI-60 tumor cell lines
lymphoma
prefrontal cortex
prostate cancer
breast cancer
breast cancer
breast cancer
astrocytoma
gastric cancer
prostate cancer
breast cancer
medulloblastoma
sarcoma
breast cancer
breast cancer
brain tumors
brain tumor and normal
glioma
leukemia
breast cancer

colorectal cancer

Javier De Las Rivas - CiC (USAL/CSIC) - 2015  66
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 outliers 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...?
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

22 microarrays from “hematopoietic” samples

34 microarrays from “brain” samples

to achieve transcriptomic global view

it is critical to avoid adequate sample selection

136 microarrays hgu133a Gene Expression Atlas
Sample selection

normal healthy tissues representing the body “evenly” (pvclust algorithm: 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

**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

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 !!!

<table>
<thead>
<tr>
<th>This work (2008)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>pathway name (KEGG ID number)</td>
<td>n°genes</td>
<td>genes coexp / genes</td>
</tr>
<tr>
<td>Proteasome (3050)</td>
<td>31</td>
<td>28 / 28</td>
</tr>
<tr>
<td>Ribosome (3010)</td>
<td>120</td>
<td>52 / 55</td>
</tr>
<tr>
<td>Oxidative phosphorylation (190)</td>
<td>129</td>
<td>88 / 95</td>
</tr>
<tr>
<td>Focal adhesion (4510)</td>
<td>194</td>
<td>154 / 168</td>
</tr>
<tr>
<td>Antigen processing and presentation (4612)</td>
<td>86</td>
<td>71 / 78</td>
</tr>
<tr>
<td>Glycan structures - degradation (1032)</td>
<td>30</td>
<td>20 / 22</td>
</tr>
<tr>
<td>Neuroactive ligand-receptor interact. (4080)</td>
<td>299</td>
<td>227 / 255</td>
</tr>
<tr>
<td>Cell cycle (4110)</td>
<td>114</td>
<td>90 / 102</td>
</tr>
<tr>
<td>Regulation of actin cytoskeleton (4810)</td>
<td>208</td>
<td>141 / 161</td>
</tr>
<tr>
<td>Cytokine-cytokine receptor interact. (4060)</td>
<td>256</td>
<td>196 / 223</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lee et al. (2004)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>pathway name (KEGG ID number)</td>
<td>n°genes</td>
<td>genes coexp / genes</td>
</tr>
<tr>
<td>Ribosome (3010)</td>
<td>120</td>
<td>43 / 44</td>
</tr>
<tr>
<td>Proteasome (3050)</td>
<td>31</td>
<td>19 / 22</td>
</tr>
<tr>
<td>Oxidative phosphorylation (190)</td>
<td>129</td>
<td>31 / 44</td>
</tr>
<tr>
<td>Cell cycle (4110)</td>
<td>114</td>
<td>33 / 47</td>
</tr>
<tr>
<td>ECM-receptor interaction (4512)</td>
<td>87</td>
<td>16 / 23</td>
</tr>
<tr>
<td>Gap junction (4540)</td>
<td>92</td>
<td>9 / 13</td>
</tr>
<tr>
<td>Pathogenic Escherichia coli infection (5130)</td>
<td>49</td>
<td>11 / 16</td>
</tr>
<tr>
<td>Pathogenic Escherichia coli infection (5131)</td>
<td>49</td>
<td>11 / 16</td>
</tr>
<tr>
<td>T cell receptor signaling pathway (4660)</td>
<td>93</td>
<td>15 / 22</td>
</tr>
<tr>
<td>Metabolism of xenobiotics by cytP450 (980)</td>
<td>70</td>
<td>7 / 11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Griffith et al. (2005)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>pathway name (KEGG ID number)</td>
<td>n°genes</td>
<td>genes coexp / genes</td>
</tr>
<tr>
<td>Ribosome (3010)</td>
<td>120</td>
<td>36 / 38</td>
</tr>
<tr>
<td>Proteasome (3050)</td>
<td>31</td>
<td>20 / 24</td>
</tr>
<tr>
<td>Oxidative phosphorylation (190)</td>
<td>129</td>
<td>55 / 67</td>
</tr>
<tr>
<td>Val, Leu and isoleucine degradation (280)</td>
<td>50</td>
<td>15 / 19</td>
</tr>
<tr>
<td>ECM-receptor interaction (4512)</td>
<td>87</td>
<td>16 / 22</td>
</tr>
<tr>
<td>Cell cycle (4110)</td>
<td>114</td>
<td>36 / 51</td>
</tr>
<tr>
<td>Propanoate metabolism (640)</td>
<td>34</td>
<td>9 / 14</td>
</tr>
<tr>
<td>Butanoate metabolism (650)</td>
<td>44</td>
<td>9 / 14</td>
</tr>
<tr>
<td>Hematopoietic cell lineage (4640)</td>
<td>88</td>
<td>18 / 28</td>
</tr>
<tr>
<td>beta-Alanine metabolism (410)</td>
<td>24</td>
<td>7 / 11</td>
</tr>
</tbody>
</table>
Gene2gene coexpression method
(based in combination of correlation $r$ and crossvalidation $N$)

reliable region $\neq$ random signal
signal / noise ratio = high
$r > 0.68$ and $N > 220$ (aprox)

noise region $\approx$ random signal
signal / noise ratio = low

$r$ (correlation factor)

$N$ (number of positives in the random crossvalidation)
Gene2gene coexpression method
mapping coexpression of housekeeping genes and tissue specific genes
(based in KEGG pathways)

A  MAS5 - Spearman

B  RMA - Pearson

House Keeping genes
- Ribosome
- Cytokine-Cytokine receptor interaction
- Oxidative phosphorylation
- Neuroactive ligand-receptor interaction

Tissue Specific genes
- Proteasome
- Complement and coagulation cascades

N (nº of positives in the random crossvalidation)
Hi-Fi gene2gene coexpression network
(based in combination of correlation \( r \) and crossvalidation \( N \))
precision obtained for 3 reliable networks at high \( r \) and \( N \)

<table>
<thead>
<tr>
<th>Precision ¹</th>
<th>Coefficients</th>
<th>Number of Nodes ²</th>
<th>Number of Links ²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( N )</td>
<td>( r )</td>
<td></td>
</tr>
<tr>
<td>RMA-Pearson (pre-Filtered)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.60</td>
<td>765</td>
<td>0.85</td>
<td>1.672</td>
</tr>
<tr>
<td>0.70</td>
<td>835</td>
<td>0.87</td>
<td>1.215</td>
</tr>
<tr>
<td>0.80</td>
<td>925</td>
<td>0.84</td>
<td>983</td>
</tr>
<tr>
<td>MAS5-Spearman (non-Filtered)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.60</td>
<td>605</td>
<td>0.77</td>
<td>3.052</td>
</tr>
<tr>
<td>0.70</td>
<td>645</td>
<td>0.79</td>
<td>2.295</td>
</tr>
<tr>
<td>0.80</td>
<td>695</td>
<td>0.81</td>
<td>1.762</td>
</tr>
</tbody>
</table>

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

ribosome
ribosome 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)

Module 1
metal ion homeostasis

Module 2
response to biotic stimulus

Module 3
extracellular matrix and adhesion
Hi-Fi human coexpression network
(modules coherent in terms of transcription factor TF regulation)

<table>
<thead>
<tr>
<th>Coexp Modules</th>
<th>Search in</th>
<th>TF found</th>
<th>p-value</th>
<th>TransFac_db</th>
<th>TF Gene Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Module 1</td>
<td>PAP</td>
<td>MTF-1</td>
<td>0.001</td>
<td>T02354</td>
<td>MTF1 metal-regulatory transcription factor 1</td>
</tr>
<tr>
<td>10 genes</td>
<td>Factory</td>
<td>=</td>
<td>=</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module 2</td>
<td>PAP</td>
<td>CRE-BP1</td>
<td>0.0172</td>
<td>T00167</td>
<td>ATF2 activating transcription factor 2</td>
</tr>
<tr>
<td>4 genes</td>
<td>Factory</td>
<td>CRE-BP1</td>
<td>0.0033</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Module 3</td>
<td>PAP</td>
<td>Sp1</td>
<td>0.13</td>
<td>T00759</td>
<td>SP1 Sp1 transcription factor</td>
</tr>
<tr>
<td>15 genes</td>
<td>Factory</td>
<td>Sp1</td>
<td>0.017</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Human transcriptomic network of normal tissues: a global map without malignant data

We achieved:

1\textsuperscript{st}. - Reliable calculation of human genome-wide (global) expression data

2\textsuperscript{nd}. - Reliable calculation of human gene2gene (global) co-expression data
Networks & Pathways
Comparison and combination of these type of complex data

**Wu et al. (2010)**

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

<table>
<thead>
<tr>
<th>Data source</th>
<th>Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human PPIs</td>
<td>10,287</td>
</tr>
<tr>
<td>Fly PPIs</td>
<td>13,383</td>
</tr>
<tr>
<td>Worm PPIs</td>
<td>5,223</td>
</tr>
<tr>
<td>Yeast PPIs</td>
<td>5,646</td>
</tr>
<tr>
<td>Domain interaction</td>
<td>60,569</td>
</tr>
<tr>
<td>Lee's Gene Expression</td>
<td>8,250</td>
</tr>
<tr>
<td>Prieto's Gene Expression</td>
<td>3,024</td>
</tr>
<tr>
<td>GO BP sharing</td>
<td>14,197</td>
</tr>
<tr>
<td>PPIs from GeneWays</td>
<td>5,252</td>
</tr>
</tbody>
</table>

**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)*

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

**Subnetwork** derived from The Cancer Genome Atlas (TCGA) of somatic mutation data set:
- **77 cancer genes**
- **5 linker genes**
Hands-on: Practical Examples

Build the coexpression network for a gene list using Cytoscape (plugin ReactomeFI, that includes FI DB)
http://wiki.reactome.org/index.php/Reactome_FI_Cytoscape_Plugin_4

Microarray Data Analysis on a Network
(NejmLogRatioNormGlobalZScore_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

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

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

Two major types of networks derived from large-scale *omic* data
Protein-Protein Interactions (PPIs)
biological networks


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.

Getting connected: analysis and principles of biological networks

Xiaowei Zhu,1,2 Mark Gerstein,3 and Michael Snyder1,2,4

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

aTranscriptional factor-binding data collected at rich-media condition.
bTranscriptional factor-binding data collected at a variety of growth conditions.
cSynthetic lethal interactions among nonessential genes.


Javier De Las Rivas - CiC (USAL/CSIC) - 2015
Protein-Protein Interactions (PPIs)
our first decade of interactome mapping: PPI data


Science Walhout et al
Science Li et al
Nature Rual et al
Science Yu et al
Nat Methods Simonis et al
Science Arabidopsis interactome consortium
Protein-Protein Interactions (PPIs)
international consortia

Our group participates actively in HUPO PSI-MI (Molecular Interactions Workgroup)
There are several primary PPI 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
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)

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.

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
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.
Protein-Protein Interactions (PPIs) review some essential concepts on PPIs

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-IBIMC, CSIC-USAL), Salamanca, Spain

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)

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

**A Proteome-Scale Map of the Human Interactome Network**

Rolland et al. (2014) Cell

Volume 159, November 2014

**Rolland et al.**

**Y2H**

**Binary**
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 Zaragoza,1 Greg Colby,1 Kurt Baltier,1 Rui Dong,2 Virginia Guarani,1 Laura Pontano Vaite,1 Alban Orduereau,2 Ramin Rad,1 Brian K. Erickson,1 Martin Wühr,1 Joel Chick,1 Bo Zhai,1 Deepak Kollipakkam,1 Julian Minelteris,1 Robert A. Obar,1,3 Tim Harris,3 Spyros Artavanis-Tsakonas,1,2 Mathew E. Sowa,1 Pietro De Camilli,2 Joao A. Paulo,1 J. Wade Harper,1,1 and Steven P. Gygi1,*

1Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA
2Department of Cell Biology and Howard Hughes Medical Institute, Yale School of Medicine, New Haven, CT 06519, USA
3Biogen, 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

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

High-Throughput Human Protein Interaction Mapping

ORFeome → Cell Culture → Affinity Purification → Interactor Identification → MS

2594 AP-MS Experiments Identify
23744 Interactions among 7668 Proteins

The Interaction Network Partitions into Complexes

Network Reflects Function and Localization

Interaction Network Aids Uncharacterized Protein Study

(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
## Binary Human Interactome

![Graphical representation of the human interactome](image)

### Dataset Summary

<table>
<thead>
<tr>
<th>Dataset Name</th>
<th>Release Date</th>
<th>N proteins</th>
<th>N interactions</th>
<th>Search Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>HI-I-2005</td>
<td>2005</td>
<td>1,545</td>
<td>2,750</td>
<td>Space I</td>
</tr>
<tr>
<td>HI-II-2014</td>
<td>2014</td>
<td>4,303</td>
<td>13,944</td>
<td>Space II</td>
</tr>
<tr>
<td>CCSB-HI-all</td>
<td>2015 (total union)</td>
<td>4,745</td>
<td>16,503</td>
<td>Space I &amp; II</td>
</tr>
</tbody>
</table>
Literature Curated Reference Sets
Lit-RS

Validated PRS/RRS

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

A Census of Human Soluble Protein Complexes

Pierro C. Havgmanna,1,2,8 G. Traver Hart,1,2,8 Tamás Nepusz,4,8 Haikuan Yang,4,8 Andrei L. Turinsky,6 Zhihua Li,6 Peggy I. Wang,6 Daniel R. Boutz,6 Vincent Pong,7 Sadhna Phanse,1 Mohan Babu,7 Stephanie A. Craig,6 Pingzheo Hu,1 Cuifong Wan,1 James Vlasblom,2,8 Vaqar-un-Nisa Dar,7 Alexandr Bezginov,7 Gregory W. Clark,7 Gabriel C. Wu,6 Shoshana J. Wodak,5,3,5 Elisabeth R.M. Tillier,7 Alberto Paccanaro,6,8 Edward M. Marcotte,6,8 and Andrew Emil1,2,*

1Banting and Best Department of Medical Research, Donnelly Centre for Cellular and Biomolecular Research
2Department of Molecular Genetics, Medical Sciences Building
3Department of Biochemistry, Medical Sciences Building
4Department of Computer Science, Royal Holloway, University of London, Egham TW20 0EX, UK
5Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada
6Center 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
7Campbell Family Institute for Cancer Research, Ontario Cancer Institute, University Health Network, University of Toronto, Toronto, Ontario M5G 1L7, Canada
8These authors contributed equally to this work
9Correspondence: Alberto Paccanaro@cs.ru.nl.ac.uk (A.P.), marcotte@icmb.utexas.edu (E.M.M.), andrew.emil@utoronto.ca (A.E.)
http://dx.doi.org/10.1016/j.cell.2012.06.011

LETTER

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

Qiangfeng Cliff Zhang1,2,3,*, Donald Petrey1,2,3,*, Lei Deng2,3,4, Li Qiang5, Yu Shi6, Chan Aye Thu2, Brygida Bisikirsk3, Celine Lefebvre3,7, Domenico Accili8, Tony Hunter4, Tom Maniatis4, Andrea Califano2,4,7,8,9 & Barry Honig1,2,3

Nature 2012

doi:10.1038/nature11503
Comparison of 4 Human Interactomes

<table>
<thead>
<tr>
<th>Network</th>
<th>N proteins in main comp</th>
<th>avg degree (total)</th>
<th>avg degree (neighbors)</th>
<th>avg betweenness</th>
<th>avg closeness</th>
<th>avg cluster coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary HI</td>
<td>4,100</td>
<td>6.516</td>
<td>58.238</td>
<td>6280.4</td>
<td>0.0075</td>
<td>0.086</td>
</tr>
<tr>
<td>Cocomplex HI</td>
<td>2,804</td>
<td>9.794</td>
<td>21.806</td>
<td>4921.8</td>
<td>0.0055</td>
<td>0.210</td>
</tr>
<tr>
<td>Literature HI</td>
<td>3,409</td>
<td>3.808</td>
<td>12.800</td>
<td>8188.5</td>
<td>0.0019</td>
<td>0.204</td>
</tr>
<tr>
<td>Predicted HI</td>
<td>3,280</td>
<td>11.316</td>
<td>12.953</td>
<td>7444.8</td>
<td>0.0000002</td>
<td>0.483</td>
</tr>
</tbody>
</table>

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

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

NOTE: excluding homodimers
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

A) Number of publications

B) Number of binary interactions
- Binary literature maps
- Systematic binary maps

Time
- 1994
- 1998
- 2002
- 2006
- 2010
- 2013

Number of publications

Number of interactions
- 0
- 10
- ≥100

HI-II-14
Lit-BM-13

Systematic binary maps
Binary literature maps

Time t
Time t + 4 years
Fraction of interactions connecting two new proteins
- 0%
- 100%

New binary interaction
Newly connected protein
Comparison of 4 Human Interactomes

<table>
<thead>
<tr>
<th>Network</th>
<th>N proteins in main comp</th>
<th>avg degree (total)</th>
<th>avg degree (neighbors)</th>
<th>avg betweenness</th>
<th>avg closeness</th>
<th>avg cluster coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary HI</td>
<td>4,100</td>
<td>6.516</td>
<td>58.238</td>
<td>6280.4</td>
<td>0.0075</td>
<td>0.086</td>
</tr>
<tr>
<td>Cocomplex HI</td>
<td>2,804</td>
<td>9.794</td>
<td>21.806</td>
<td>4921.8</td>
<td>0.0055</td>
<td>0.210</td>
</tr>
<tr>
<td>Literature HI</td>
<td>3,409</td>
<td>3.888</td>
<td>12.800</td>
<td>8188.5</td>
<td>0.0019</td>
<td>0.204</td>
</tr>
<tr>
<td>Predicted HI</td>
<td>3,280</td>
<td>11.316</td>
<td>12.953</td>
<td>7444.8</td>
<td>0.0000002</td>
<td>0.483</td>
</tr>
</tbody>
</table>

Topological characteristics of the PPI networks (main component, only heterodimers considered)
Comparison of 4 Human Interactomes

<table>
<thead>
<tr>
<th>Number of Interactions</th>
<th>mRNA abundance in HEKcells</th>
<th>Fraction of sequence in Pfam domains</th>
<th>Fraction of sequence in transmembrane heJioos</th>
</tr>
</thead>
<tbody>
<tr>
<td>E3</td>
<td>Hl-Il-14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PrePPI-HC</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frac</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hi-BM-13</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Co-Frac</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PrePPI-HC</td>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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
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 Orduereau,1 Ramin Rad,1 Brian K. Erickson,1 Martin Wühr,1 Joel Chick,1 Bo Zhai,1 Deepak Kollipakkam,1 Julian Mitrissis,1 Robert A. Obar,1,3 Tim Harris,3 Spyros Artavanis-Tsakonas,1,2 Mathew E. Sowa,1 Pietro De Camilli,2 Joao A. Paulo,1 J. Wade Harper,1,4 and Steven P. Gygi1,4

1Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA
2Department of Cell Biology and Howard Hughes Medical Institute, Yale School of Medicine, New Haven, CT 06519, USA
3Biogen, Cambridge, MA 02142, USA
4Correspondence: 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

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

(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

**Protein Interactions PPIs**

**Y2H**

**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
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

True interactions (PPIs)
physical topology in vivo

A

P2
P1
P4
P3

B

P4
P5
P6

2 sets of proteins A & B

Binary methods
measure physical direct PPIs

e.g. Y2H

P1–P2
P1–P3
P1–P4

P2–P3
P4–P5

P1–P4
P4–P6

direct assignment

Experimental interactions (PPIs)

obtained from binary or co-complex methods

Co-complex methods
measure physical PPIs (direct & indirect)

E.g. TAP-MS
CoIP

P1=P2,P3,P4
P2=P1,P3,P4
P3=P1,P2,P4
P4=P1,P2,P3,P5,P6
P5=P4,P6
P6=P4,P5

assignment with spoke model

two different PPI networks derived from

two types of experiential data
(the X below indicate interactions that do not occur,
i.e. they will be false positives)

Javier De Las Rivas - CiC (USAL/CSIC) - 2015131
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**
Protein Interactions (PIs)

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

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

Antibody arrays

Proteome arrays

Peptide arrays

Reverse arrays
Protein-Protein Interactions (PPIs)

Proteins arrays to detect protein-protein interactions

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

## Primary Databases: PPI experimental data (curated from specific SSc & LSc published studies)

<table>
<thead>
<tr>
<th>Name</th>
<th>DB full name and type</th>
<th>PPIs sources</th>
<th>Type of MI</th>
<th>species</th>
<th>n prot.</th>
<th>n interact.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIND</td>
<td>Biomolecular Interaction Network Database</td>
<td>Ssc &amp; Lsc published studies (literature-curated)</td>
<td>PPIs &amp; others</td>
<td>all</td>
<td>[31972]</td>
<td>[58266]</td>
</tr>
<tr>
<td>BioGRID</td>
<td>General Repository for Interaction Datasets</td>
<td>Ssc &amp; Lsc published studies (literature-curated)</td>
<td>PPIs &amp; others</td>
<td>all</td>
<td>[28717]</td>
<td>[108691]</td>
</tr>
<tr>
<td>DIP</td>
<td>Database of Interacting Proteins</td>
<td>Ssc &amp; Lsc published studies (literature-curated)</td>
<td>only PPIs</td>
<td>all</td>
<td>20728</td>
<td>57683</td>
</tr>
<tr>
<td>HPRD</td>
<td>Human Protein Reference Database</td>
<td>Ssc &amp; Lsc published studies (literature-curated)</td>
<td>only PPIs</td>
<td>human</td>
<td>27081</td>
<td>38806</td>
</tr>
<tr>
<td>IntAct</td>
<td>Database of protein InterAction data</td>
<td>Ssc &amp; Lsc published studies (literature-curated)</td>
<td>PPIs &amp; others</td>
<td>all</td>
<td>[60504]</td>
<td>[202826]</td>
</tr>
<tr>
<td>MINT</td>
<td>Molecular INTeractions database</td>
<td>Ssc &amp; Lsc published studies (literature-curated)</td>
<td>only PPIs</td>
<td>all</td>
<td>30089</td>
<td>83744</td>
</tr>
<tr>
<td>MIPS-MPact</td>
<td>MIPS protein interaction resource on yeast</td>
<td>derived from CYGD</td>
<td>only PPIs</td>
<td>yeast</td>
<td>1500</td>
<td>4300</td>
</tr>
<tr>
<td>MIPS-MPPI</td>
<td>MIPS mammalian protein-protein interaction db</td>
<td>Ssc published studies (literature-curated)</td>
<td>only PPIs</td>
<td>mammalia</td>
<td>982</td>
<td>937</td>
</tr>
</tbody>
</table>

## Meta Databases: PPI experimental data (integrated and unified from different public repositories)

<table>
<thead>
<tr>
<th>Name</th>
<th>DB full name and type</th>
<th>PPIs sources</th>
<th>Type of MI</th>
<th>species</th>
<th>n prot.</th>
<th>n interact.</th>
</tr>
</thead>
<tbody>
<tr>
<td>APID</td>
<td>Agile Protein Interaction DataAnalyzer</td>
<td>BIND, BioGRID, DIP, HPRD, IntAct, MINT</td>
<td>only PPIs</td>
<td>all</td>
<td>56460</td>
<td>322579</td>
</tr>
<tr>
<td>MPIDB</td>
<td>The microbial protein interaction database</td>
<td>BIND, DIP, IntAct, MINT, other sets (exp &amp; litcur)</td>
<td>only PPIs</td>
<td>microbial</td>
<td>7810</td>
<td>24295</td>
</tr>
<tr>
<td>PINA</td>
<td>Protein Interaction Network Analysis platform</td>
<td>BioGRID, DIP, HPRD, IntAct, MINT, MPact</td>
<td>only PPIs</td>
<td>all</td>
<td>[?]</td>
<td>188823</td>
</tr>
</tbody>
</table>

## Prediction Databases: PPI experimental & predicted data ("functional interactions", i.e. interactions *lato sensu* derived from different types of data)

<table>
<thead>
<tr>
<th>Name</th>
<th>DB full name and type</th>
<th>PPIs sources</th>
<th>Type of MI</th>
<th>species</th>
<th>n prot.</th>
<th>n interact.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MiMI</td>
<td>Michigan Molecular Interactions</td>
<td>BIND, BioGRID, DIP, HPRD, IntAct &amp; nonPPI dt</td>
<td>PPIs &amp; others</td>
<td>all</td>
<td>[45452]</td>
<td>[391386]</td>
</tr>
<tr>
<td>PPIs</td>
<td>Human protein-protein interactions prediction db</td>
<td>BIND, DIP, HPRD, OPHID &amp; nonPPI dt</td>
<td>PPIs &amp; others</td>
<td>human</td>
<td>[?]</td>
<td>[37606]</td>
</tr>
<tr>
<td>OPHID</td>
<td>Online Predicted Human Interaction Database</td>
<td>BIND, BioGRID, HPRD, IntAct, MINT, MPact &amp; nonPPI dt</td>
<td>PPIs &amp; others</td>
<td>human</td>
<td>[?]</td>
<td>[424066]</td>
</tr>
<tr>
<td>STRING</td>
<td>Known and Predicted Protein-Protein Interactions</td>
<td>BIND, BioGRID, DIP, HPRD, IntAct, MINT &amp; nonPPI dt</td>
<td>PPIs &amp; others</td>
<td>all</td>
<td>[2590259]</td>
<td>[88633860]</td>
</tr>
<tr>
<td>UniHI</td>
<td>Unified Human Interactome</td>
<td>BIND, BioGRID, DIP, HPRD, IntAct, MINT &amp; nonPPI dt</td>
<td>PPIs &amp; others</td>
<td>human</td>
<td>[22307]</td>
<td>[200473]</td>
</tr>
</tbody>
</table>

---

From *De Las Rivas & Fontanillo (2010)*  
Javier De Las Rivas - CiC (USAL/CSIC) - 2015139
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 inferred but not really proved).

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**

<table>
<thead>
<tr>
<th>Name</th>
<th>DB full name and type</th>
<th>PPIs sources</th>
<th>Type of MI</th>
<th>species</th>
<th>n prot.</th>
<th>n interact.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary Databases:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIND</td>
<td>Biomolecular Interaction Network Database</td>
<td>Ssc &amp; Lsc published studies (literature-curated)</td>
<td>PPIs &amp; others</td>
<td>all</td>
<td>[31972]</td>
<td>[58266]</td>
</tr>
<tr>
<td>BioGRID</td>
<td>General Repository for Interaction Datasets</td>
<td>Ssc &amp; Lsc published studies (literature-curated)</td>
<td>PPIs &amp; others</td>
<td>all</td>
<td>[28717]</td>
<td>[108691]</td>
</tr>
<tr>
<td>DIP</td>
<td>Database of Interacting Proteins</td>
<td>Ssc &amp; Lsc published studies (literature-curated)</td>
<td>only PPIs</td>
<td>all</td>
<td>[20728]</td>
<td>[57683]</td>
</tr>
<tr>
<td>HPRD</td>
<td>Human Protein Reference Database</td>
<td>Ssc &amp; Lsc published studies (literature-curated)</td>
<td>only PPIs</td>
<td>human</td>
<td>27081</td>
<td>38806</td>
</tr>
<tr>
<td>IntAct</td>
<td>Database of protein InterAction data</td>
<td>Ssc &amp; Lsc published studies (literature-curated)</td>
<td>PPIs &amp; others</td>
<td>all</td>
<td>[60504]</td>
<td>[202826]</td>
</tr>
<tr>
<td>MINT</td>
<td>Molecular INTeractions database</td>
<td>Ssc &amp; Lsc published studies (literature-curated)</td>
<td>only PPIs</td>
<td>all</td>
<td>[30089]</td>
<td>[83744]</td>
</tr>
<tr>
<td>MIPS-MPact</td>
<td>MIPS protein interaction resource on yeast</td>
<td>derived from CYGD</td>
<td>only PPIs</td>
<td>yeast</td>
<td>1500</td>
<td>4300</td>
</tr>
<tr>
<td>MIPS-MPPI</td>
<td>MIPS mammalian protein-protein interaction db</td>
<td>Ssc published studies (literature-curated)</td>
<td>only PPIs</td>
<td>mammalia</td>
<td>982</td>
<td>937</td>
</tr>
<tr>
<td><strong>Meta Databases:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APID</td>
<td>Agile Protein Interaction DataAnalyzer</td>
<td>BIND, BioGRID, DIP, HPRD, IntAct, MINT</td>
<td>only PPIs</td>
<td>all</td>
<td>56460</td>
<td>322579</td>
</tr>
<tr>
<td>MPIDB</td>
<td>The microbial protein interaction database</td>
<td>BIND, DIP, IntAct, MINT, other sets (exp &amp; litcur)</td>
<td>only PPIs</td>
<td>microbial</td>
<td>7810</td>
<td>24295</td>
</tr>
<tr>
<td>PINA</td>
<td>Protein Interaction Network Analysis platform</td>
<td>BioGRID, DIP, HPRD, IntAct, MInt, MPact</td>
<td>only PPIs</td>
<td>all</td>
<td>[?]</td>
<td>188823</td>
</tr>
<tr>
<td><strong>Prediction Databases:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MiMI</td>
<td>Michigan Molecular Interactions</td>
<td>BIND, BioGRID, DIP, HPRD, IntAct &amp; nonPPI dt</td>
<td>PPIs &amp; others</td>
<td>all</td>
<td>[45452]</td>
<td>[391386]</td>
</tr>
<tr>
<td>PIPs</td>
<td>Human protein-protein interactions prediction db</td>
<td>BIND, DIP, HPRD, OPHID &amp; nonPPI dt</td>
<td>PPIs &amp; others</td>
<td>human</td>
<td>[?]</td>
<td>[37606]</td>
</tr>
<tr>
<td>OPHID</td>
<td>Online Predicted Human Interaction Database</td>
<td>BIND, BioGRID, HPRD, IntAct, MInt, MPact &amp; nonPPI dt</td>
<td>PPIs &amp; others</td>
<td>human</td>
<td>[?]</td>
<td>[424066]</td>
</tr>
<tr>
<td>STRING</td>
<td>Known and Predicted Protein-Protein Interactions</td>
<td>BIND, BioGRID, DIP, HPRD, IntAct, MINT &amp; nonPPI dt</td>
<td>PPIs &amp; others</td>
<td>all</td>
<td>[2590259]</td>
<td>[88633860]</td>
</tr>
<tr>
<td>UniHI</td>
<td>Unified Human Interactome</td>
<td>BIND, BioGRID, DIP, HPRD, IntAct, MINT &amp; nonPPI dt</td>
<td>PPIs &amp; others</td>
<td>human</td>
<td>[22307]</td>
<td>[200473]</td>
</tr>
</tbody>
</table>

*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

<table>
<thead>
<tr>
<th>Name</th>
<th>DB full name and type</th>
<th>PPIs sources</th>
<th>Type of MI</th>
<th>species</th>
<th>n prot.</th>
<th>n interact.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary Databases:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIND</td>
<td>Biomolecular Interaction Network Database</td>
<td>Ssc &amp; Lsc published studies (literature-curated)</td>
<td>PPIs &amp; others</td>
<td>all</td>
<td>[31972]</td>
<td>[58266]</td>
</tr>
<tr>
<td>BioGRID</td>
<td>General Repository for Interaction Datasets</td>
<td>Ssc &amp; Lsc published studies (literature-curated)</td>
<td>PPIs &amp; others</td>
<td>all</td>
<td>[28717]</td>
<td>[108691]</td>
</tr>
<tr>
<td>DIP</td>
<td>Database of Interacting Proteins</td>
<td>Ssc &amp; Lsc published studies (literature-curated)</td>
<td>only PPIs</td>
<td>all</td>
<td>20728</td>
<td>57683</td>
</tr>
<tr>
<td>HPRD</td>
<td>Human Protein Reference Database</td>
<td>Ssc &amp; Lsc published studies (literature-curated)</td>
<td>only PPIs</td>
<td>human</td>
<td>27081</td>
<td>38806</td>
</tr>
<tr>
<td>IntAct</td>
<td>Database of protein Interaction data</td>
<td>Ssc &amp; Lsc published studies (literature-curated)</td>
<td>PPIs &amp; others</td>
<td>all</td>
<td>[60504]</td>
<td>[202826]</td>
</tr>
<tr>
<td>MINT</td>
<td>Molecular INTeractions database</td>
<td>Ssc &amp; Lsc published studies (literature-curated)</td>
<td>only PPIs</td>
<td>all</td>
<td>30089</td>
<td>83744</td>
</tr>
<tr>
<td>MIPS-MPact</td>
<td>MIPS protein interaction resource on yeast</td>
<td>derived from CYGD</td>
<td>only PPIs</td>
<td>yeast</td>
<td>1500</td>
<td>4300</td>
</tr>
<tr>
<td>MIPS-MPPi</td>
<td>MIPS mammalian protein-protein interaction db</td>
<td>Ssc published studies (literature-curated)</td>
<td>only PPIs</td>
<td>mammalia</td>
<td>982</td>
<td>937</td>
</tr>
<tr>
<td><strong>Meta Databases:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APID</td>
<td>Agile Protein Interaction DataAnalyzer</td>
<td>BIND, BioGRID, DIP, HPRD, IntAct, MINT</td>
<td>only PPIs</td>
<td>all</td>
<td>56460</td>
<td>322579</td>
</tr>
<tr>
<td>MPIDB</td>
<td>The microbial protein interaction database</td>
<td>BIND, DIP, IntAct, MINT, other sets (exp &amp; litcur)</td>
<td>only PPIs</td>
<td>microbial</td>
<td>7810</td>
<td>24295</td>
</tr>
<tr>
<td>PINA</td>
<td>Protein Interaction Network Analysis platform</td>
<td>BioGRID, DIP, HPRD, IntAct, MINT, MPact</td>
<td>only PPIs</td>
<td>all</td>
<td>[?]</td>
<td>188823</td>
</tr>
<tr>
<td><strong>Prediction Databases:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MiMI</td>
<td>Michigan Molecular Interactions</td>
<td>BIND, BioGRID, DIP, HPRD, IntAct &amp; nonPPI dt</td>
<td>PPIs &amp; others</td>
<td>all</td>
<td>[45452]</td>
<td>[391386]</td>
</tr>
<tr>
<td>PIPs</td>
<td>Human protein-protein interactions prediction db</td>
<td>BIND, DIP, HPRD, OPHID &amp; nonPPI dt</td>
<td>PPIs &amp; others</td>
<td>human</td>
<td>[?]</td>
<td>[37606]</td>
</tr>
<tr>
<td>OPHID</td>
<td>Online Predicted Human Interaction Database</td>
<td>BIND, BioGRID, HPRD, IntAct, MINT, MPact &amp; nonPPI dt</td>
<td>PPIs &amp; others</td>
<td>human</td>
<td>[?]</td>
<td>[424066]</td>
</tr>
<tr>
<td>STRING</td>
<td>Known and Predicted Protein-Protein Interactions</td>
<td>BIND, BioGRID, DIP, HPRD, IntAct, MINT &amp; nonPPI dt</td>
<td>PPIs &amp; others</td>
<td>all</td>
<td>[2590259]</td>
<td>[88633860]</td>
</tr>
<tr>
<td>UniHI</td>
<td>Unified Human Interactome</td>
<td>BIND, BioGRID, DIP, HPRD, IntAct, MINT &amp; nonPPI dt</td>
<td>PPIs &amp; others</td>
<td>human</td>
<td>[22307]</td>
<td>[200473]</td>
</tr>
</tbody>
</table>

*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

**APID (Agile Protein Interaction DataAnalyzer)** is an interactive bioinformatic web-tool that has been developed to allow exploration and analysis of main currently known information about protein-protein interactions integrated and unified in a common and comparative platform. The analytical and integrative effort done in APID provides an open access frame where all known experimentally validated protein-protein interactions (BIND, BioGRID, DIP, HPRD, IntAct and MINT) are unified in a unique web application that allows an agile exploration of the interactome network and includes certain calculated parameters that weight the reliability of a given interaction (i.e. the "edges" of the interactome network) between two proteins, and also qualify the functional environment around any given protein (i.e. the "nodes" of the interactome network). Such parameters are:

... about the **proteins**:
- **Connectivity**: a graph parameter that indicates the number of proteins that directly interact with a query protein.
- **Cluster Coefficient**: a graph parameter that indicates the degree of inter-connection of the group of proteins that directly interact to a query protein.
- **GO Environment**: a tool that identifies and lists all the Gene Ontology (GO) terms that are assigned to the proteins directly interacting with a query protein.
- **GO Environment Enrichment**: a tool that for each protein selects the most-represented and non-self GO terms assigned to the proteins interacting with such protein.

... about the **interactions**:
- **Number of methods**: number of experimentally validated methods that prove a protein-protein interaction, given the PubMed reference and link.
- **GO overlapping**: a tool that shows the GO terms assigned to each protein-pair and marks the ones that are common to both.
- **IPfam domain-domain interaction**: a tool that identifies the Pfam domains of each protein-pair and marks the ones that interact according to IPfam database.

**APID Statistics**

- Number of Proteins: 56460
- Number of Interactions: 322579

**Experimental data unified in APID**

**Number of Proteins in APID**: 56460
**Number of Interactions in APID**: 322579
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](http://bioinfow.dep.usal.es/apid)

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

<table>
<thead>
<tr>
<th>Category</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of PROTEINS (total)</td>
<td>63934</td>
</tr>
<tr>
<td>Number of INTERACTIONS (total)</td>
<td>333436</td>
</tr>
<tr>
<td>Number of Proteins in 3D interactions</td>
<td>12784</td>
</tr>
<tr>
<td>Number of 3D Interactions</td>
<td>15130</td>
</tr>
<tr>
<td>Number of PROTEINS in PPI sources</td>
<td>56460</td>
</tr>
<tr>
<td>Number of INTERACTIONS in PPI sources</td>
<td>322579</td>
</tr>
</tbody>
</table>

- 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
Hands-on: Practical Examples

Build ppi networks in Cytoscape
(plugin 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)

<table>
<thead>
<tr>
<th>NameSystematic</th>
<th>Uniprot_ID</th>
<th>NameGene</th>
<th>UniProt_Name</th>
<th>Synonyms</th>
<th>MW(kDa)</th>
<th>Study</th>
<th>SubComplex</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>YJL109c</td>
<td>P42945</td>
<td>Utp10</td>
<td>UTP10_YEAST</td>
<td>na</td>
<td>200.08</td>
<td>1stStudy</td>
<td>UTP-A</td>
<td>U3 small nucleolar RNA-associated protein 10U3 snoRNA-associate</td>
</tr>
<tr>
<td>YPL126w</td>
<td>Q02931</td>
<td>Nan1</td>
<td>NAN1_YEAST</td>
<td>Utp17</td>
<td>101.24</td>
<td>1stStudy</td>
<td>UTP-A</td>
<td>Nucleolar protin NAN1U3 small nucleolar RNA-associated protein 17</td>
</tr>
<tr>
<td>YDR324c</td>
<td>Q06679</td>
<td>Utp4</td>
<td>UTP4_YEAST</td>
<td>na</td>
<td>87.81</td>
<td>1stStudy</td>
<td>UTP-A</td>
<td>U3 small nucleolar RNA-associated protein 4U3 snoRNA-associated</td>
</tr>
<tr>
<td>YGR128c</td>
<td>P53276</td>
<td>Utp8</td>
<td>UTP8_YEAST</td>
<td>na</td>
<td>80.19</td>
<td>1stStudy</td>
<td>UTP-A</td>
<td>U3 small nucleolar RNA-associated protein 8U3 snoRNA-associated</td>
</tr>
<tr>
<td>YDR398w</td>
<td>Q04177</td>
<td>Utp5</td>
<td>UTP5_YEAST</td>
<td>na</td>
<td>72.18</td>
<td>1stStudy</td>
<td>UTP-A</td>
<td>U3 small nucleolar RNA-associated protein 5U3 snoRNA-associated</td>
</tr>
<tr>
<td>YHR196w</td>
<td>P38882</td>
<td>Utp9</td>
<td>UTP9_YEAST</td>
<td>na</td>
<td>65.27</td>
<td>1stStudy</td>
<td>UTP-A</td>
<td>U3 small nucleolar RNA-associated protein 9U3 snoRNA-associated</td>
</tr>
<tr>
<td>YMR093w</td>
<td>P04350</td>
<td>Utp15</td>
<td>UTP15_YEAST</td>
<td>na</td>
<td>57.69</td>
<td>1stStudy</td>
<td>UTP-A</td>
<td>U3 small nucleolar RNA-associated protein 15U3 snoRNA-associate</td>
</tr>
<tr>
<td>YLR129w</td>
<td>P12220</td>
<td>Dip2</td>
<td>DIP2_YEAST</td>
<td>na</td>
<td>106.34</td>
<td>1stStudy</td>
<td>UTP-B</td>
<td>DOM34 interacting protein 2U3 small nucleolar RNA-associated prot</td>
</tr>
<tr>
<td>YLR409c</td>
<td>Q06078</td>
<td>Utp21</td>
<td>YL09_YEAST</td>
<td>na</td>
<td>104.79</td>
<td>1stStudy</td>
<td>UTP-B</td>
<td>Hypothetical 104.8 kDa Trp-Asp repeats containing protein in RPL31</td>
</tr>
<tr>
<td>YCR057c</td>
<td>P25635</td>
<td>Pwp2</td>
<td>PWP2_YEAST</td>
<td>Utp1</td>
<td>103.98</td>
<td>1stStudy</td>
<td>UTP-B</td>
<td>Periodic tryptophan protein 2U3 small nucleolar RNA-associated prot</td>
</tr>
<tr>
<td>YLR222c</td>
<td>Q05946</td>
<td>Utp13</td>
<td>UTP13_YEAST</td>
<td>na</td>
<td>91.03</td>
<td>1stStudy</td>
<td>UTP-B</td>
<td>U3 small nucleolar RNA-associated protein 13U3 snoRNA-associate</td>
</tr>
<tr>
<td>YJL069c</td>
<td>P40362</td>
<td>Utp18</td>
<td>CG48_YEAST</td>
<td>na</td>
<td>66.42</td>
<td>1stStudy</td>
<td>UTP-B</td>
<td>Hypothetical 66.4 kDa Trp-Asp repeats containing protein in SMC3-M</td>
</tr>
<tr>
<td>YDR445c</td>
<td>Q02354</td>
<td>Utp6</td>
<td>UTP6_YEAST</td>
<td>na</td>
<td>52.42</td>
<td>1stStudy</td>
<td>UTP-B</td>
<td>U3 small nucleolar RNA-associated protein 6U3 snoRNA-associated</td>
</tr>
<tr>
<td>YGR090w</td>
<td>P53254</td>
<td>Utp22</td>
<td>YG2L_YEAST</td>
<td>na</td>
<td>140.48</td>
<td>1stStudy</td>
<td>UTP-C</td>
<td>Hypothetical 140.5 kDa protein in CTT1-PRP31 intergenic region</td>
</tr>
<tr>
<td>YIL035c</td>
<td>P15790</td>
<td>Cka1</td>
<td>CSK21_YEAST</td>
<td>Csk21</td>
<td>44.67</td>
<td>1stStudy</td>
<td>UTP-C</td>
<td>Casein kinase II, alpha chain CK II alpha subunit</td>
</tr>
<tr>
<td>YOR061W</td>
<td>P19454</td>
<td>Cka2</td>
<td>CSK22_YEAST</td>
<td>Csk22</td>
<td>39.47</td>
<td>1stStudy</td>
<td>UTP-C</td>
<td>Casein kinase II, alpha chain (CK II)</td>
</tr>
<tr>
<td>YCL031c</td>
<td>P25368</td>
<td>Rrp7</td>
<td>RRP7_YEAST</td>
<td>na</td>
<td>34.47</td>
<td>1stStudy</td>
<td>UTP-C</td>
<td>Ribosomal RNA processing protein 7</td>
</tr>
<tr>
<td>YGL019W</td>
<td>P43639</td>
<td>Ckb1</td>
<td>CSK2B_YEAST</td>
<td>Csk2b</td>
<td>32.26</td>
<td>1stStudy</td>
<td>UTP-C</td>
<td>Casein kinase II beta subunit CK II beta</td>
</tr>
<tr>
<td>YOR039W</td>
<td>P38930</td>
<td>Ckb2</td>
<td>CSK2C_YEAST</td>
<td>Csk2c</td>
<td>29.84</td>
<td>1stStudy</td>
<td>UTP-C</td>
<td>Casein kinase II beta subunit CK II beta'</td>
</tr>
<tr>
<td>YJR002w</td>
<td>P47083</td>
<td>Mpp10</td>
<td>MPP10_YEAST</td>
<td>na</td>
<td>66.95</td>
<td>1stStudy</td>
<td>MPP10-C</td>
<td>U3 small nucleolar ribonucleoprotein protein MPP10</td>
</tr>
<tr>
<td>YNL075w</td>
<td>P53941</td>
<td>Imp4</td>
<td>IMP4_YEAST</td>
<td>na</td>
<td>33.48</td>
<td>1stStudy</td>
<td>MPP10-C</td>
<td>U3 small nucleolar ribonucleoprotein protein IMP4</td>
</tr>
<tr>
<td>YHR148w</td>
<td>P32899</td>
<td>Imp3</td>
<td>IMP3_YEAST</td>
<td>na</td>
<td>21.89</td>
<td>1stStudy</td>
<td>MPP10-C</td>
<td>U3 small nucleolar ribonucleoprotein protein IMP3</td>
</tr>
<tr>
<td>YPL217c</td>
<td>Q08965</td>
<td>Bms1</td>
<td>BMS1_YEAST</td>
<td>na</td>
<td>135.57</td>
<td>1stStudy</td>
<td>outSubC</td>
<td>Ribosome biogenesis protein BMS1</td>
</tr>
<tr>
<td>YGR145w</td>
<td>P48234</td>
<td>Enp2</td>
<td>YG3J_YEAST</td>
<td>na</td>
<td>81.75</td>
<td>1stStudy</td>
<td>outSubC</td>
<td>Hypothetical WD-repeat protein in MOL1-NAT2 intergenic region</td>
</tr>
<tr>
<td>YMR290c</td>
<td>Q03532</td>
<td>Has1</td>
<td>HAS1_YEAST</td>
<td>na</td>
<td>56.72</td>
<td>1stStudy</td>
<td>outSubC</td>
<td>Probable ATP-dependent RNA helicase HAS1</td>
</tr>
<tr>
<td>YNL132w</td>
<td>P53914</td>
<td>Kre33</td>
<td>YN22_YEAST</td>
<td>na</td>
<td>119.35</td>
<td>1stStudy</td>
<td>outSubC</td>
<td>Hypothetical UPF0202 protein YNL132w</td>
</tr>
<tr>
<td>YCL059c</td>
<td>P25586</td>
<td>Krr1</td>
<td>YCF9_YEAST</td>
<td>na</td>
<td>37.16</td>
<td>1stStudy</td>
<td>outSubC</td>
<td>Hypothetical 37.2 kDa protein in CHA1-PRD1 intergenic region</td>
</tr>
<tr>
<td>YPR144c</td>
<td>Q06512</td>
<td>Noc4</td>
<td>NOC4_YEAST</td>
<td>Utp19</td>
<td>63.64</td>
<td>1stStudy</td>
<td>outSubC</td>
<td>Nucleolar complex protein 4U3 small nucleolar RNA-associates probe</td>
</tr>
<tr>
<td>YDL014w</td>
<td>P15646</td>
<td>Nop1</td>
<td>FBR1_YEAST</td>
<td>Lot3_FBRL</td>
<td>34.47</td>
<td>1stStudy</td>
<td>outSubC</td>
<td>FibrillarinNucleolar protein 1</td>
</tr>
<tr>
<td>YDL148c</td>
<td>Q99207</td>
<td>Nop14</td>
<td>NOP14_YEAST</td>
<td>Utp2</td>
<td>94.3</td>
<td>1stStudy</td>
<td>outSubC</td>
<td>Nucleolar complex protein 14U3 small nucleolar RNA-associated protein</td>
</tr>
<tr>
<td>YMR229c</td>
<td>Q05022</td>
<td>Rrp5</td>
<td>RR55_YEAST</td>
<td>na</td>
<td>193.13</td>
<td>1stStudy</td>
<td>outSubC</td>
<td>rRNA biogenesis protein RRP5</td>
</tr>
<tr>
<td>YBL004w</td>
<td>P35194</td>
<td>Utp20</td>
<td>YBA4_YEAST</td>
<td>na</td>
<td>287.56</td>
<td>1stStudy</td>
<td>outSubC</td>
<td>Hypothetical 287.5 kDa protein in PDR3-HTA2 intergenic region</td>
</tr>
</tbody>
</table>

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.
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 Mechanism

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

Centro de Investigación del Cáncer and Instituto de Biología Molecular y Celular del Cáncer, CSI C-University of Salamanca, Campus Unamuno, E-37007 Salamanca, Spain

<table>
<thead>
<tr>
<th>UTP-AI t-UTP</th>
<th>Pwp2p-MYC</th>
<th>Rrp7p-MYC</th>
<th>Nan1p-MYC</th>
<th>Utp4p-MYC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utp10p</td>
<td>YJL109c</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nan1p</td>
<td>YPL126w</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utp4p</td>
<td>YDR324c</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utp8p</td>
<td>YGR128c</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utp5p</td>
<td>YDR398w</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utp9p</td>
<td>YHR196w</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utp15p</td>
<td>YMR093w</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pwp2p/UTP-B</th>
<th>Pwp2p-MYC</th>
<th>Rrp7p-MYC</th>
<th>Nan1p-MYC</th>
<th>Utp4p-MYC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pwp2p</td>
<td>YCR057c</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dlp2p</td>
<td>YLR129w</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utp21p</td>
<td>YLR409c</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utp13p</td>
<td>YLR222c</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utp18p</td>
<td>YJL069c</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utp6p</td>
<td>YDR449c</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>UTP-C</th>
<th>Utp22p</th>
<th>Rrp7p</th>
<th>Nan1p</th>
<th>Utp21p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utp22p</td>
<td>YGR090w</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rrp7p</td>
<td>YCL031c</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cka1p</td>
<td>YIL035c</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mpp10</th>
<th>Mpp10p</th>
<th>Imp4p</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mpp10p</td>
<td>YJR002w</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imp4p</td>
<td>YNL075w</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OTHER 90S PROTEINS</th>
<th>Utp20p</th>
<th>Rrp5p</th>
<th>Bms1p</th>
<th>Kre33p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utp20p</td>
<td>YBL004w</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rrp5p</td>
<td>YMR229c</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bms1p</td>
<td>YPL217c</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kre33p</td>
<td>YNL132w</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nan1p</th>
<th>Utp20p</th>
<th>Rrp5p</th>
<th>Bms1p</th>
<th>Kre33p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utp20p</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rrp5p</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bms1p</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kre33p</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nan1p</th>
<th>Utp20p</th>
<th>Rrp5p</th>
<th>Bms1p</th>
<th>Kre33p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utp20p</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rrp5p</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bms1p</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kre33p</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Javier De Las Rivas- CIC (USAL/CSIC)- 2015155
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 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)

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

From protein interactions to protein networks
build reliable networks with biological meaning: example 1

Pre-RIBOSOME
from
32 proteins
to
4 groups

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

We discover protein groups that correspond to subcomplexes experimentally found.

Pre-RIBOSOME from 32 proteins to 4 sub-complexes

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 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?

26S Proteasome (Saccharomyces cerevisiae)

<table>
<thead>
<tr>
<th>Rpn1</th>
<th>Rpn2</th>
<th>Rpn3</th>
<th>Rpn4</th>
<th>Rpn5</th>
<th>Rpn6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rpn7</td>
<td>Rpn8</td>
<td>Rpn9</td>
<td>Rpn10</td>
<td>Rpn11</td>
<td>Rpn12</td>
</tr>
<tr>
<td>Rpt1</td>
<td>Rpt2</td>
<td>Rpt3</td>
<td>Rpt4</td>
<td>Rpt5</td>
<td>Rpt6</td>
</tr>
<tr>
<td>α1</td>
<td>α2</td>
<td>α3</td>
<td>α4</td>
<td>α5</td>
<td>α6</td>
</tr>
<tr>
<td>β1</td>
<td>β2</td>
<td>β3</td>
<td>β4</td>
<td>β5</td>
<td>β6</td>
</tr>
</tbody>
</table>
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?

26S Proteasome (Saccharomyces cerevisiae)

<table>
<thead>
<tr>
<th>Rpn1</th>
<th>Rpn2</th>
<th>Rpn3</th>
<th>Rpn4</th>
<th>Rpn5</th>
<th>Rpn6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rpn7</td>
<td>Rpn8</td>
<td>Rpn9</td>
<td>Rpn10</td>
<td>Rpn11</td>
<td>Rpn12</td>
</tr>
<tr>
<td>Rpt1</td>
<td>Rpt2</td>
<td>Rpt3</td>
<td>Rpt4</td>
<td>Rpt5</td>
<td>Rpt6</td>
</tr>
<tr>
<td>α1</td>
<td>α2</td>
<td>α3</td>
<td>α4</td>
<td>α5</td>
<td>α6</td>
</tr>
<tr>
<td>β1</td>
<td>β2</td>
<td>β3</td>
<td>β4</td>
<td>β5</td>
<td>β6</td>
</tr>
<tr>
<td>β7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Party hubs vs Date hubs

Han et al. (2004) Nature

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

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

Intramodular hubs vs Intermodular hubs
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.
NOTCH SIGNALING PATHWAY: hsa04330 (KEGG database)
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

- Signal Sensing Cell
- Notch or Mib-like
- Ligand
- Endosome
- Delta J-agged
- Extracellular Space
- Notch
- Numb or Numb-like
- Recycling
- Cleavage
- NICD
- S1 Cleavage
- Furin
- S2 Cleavage
- S3 Cleavage
- Ubiquitination Degradation
- Notch Target Genes
- HES Family
- Myc
- p21
- Cyclin D3
- CSL
- SKIP
- MAML
- HAT
- KDM5a
- HDAC
- SMRT
- SHARP
- CIIAP
- CIDP
- CIR
- Cytoplasm
- Nucleus
- Extracellular Space
- 171

Stromal cell
- Jagged/ Delta-like
- Presenilin
- Notch
- γ-secretase
- Intracellular Notch (ICN)
- CoA
- CoR
- CSL
- Intracellular Commitment
- Induces T-cell commitment
- Blocks B-cell commitment
- No T-cell commitment
- Allows B-cell commitment
From PPI & pathways to protein networks
build reliable networks with biological meaning: example 3

NOTCH SIGNALING PATHWAY: hsa04330 (KEGG database)
NOTCH SIGNALING PATHWAY: hsa04330 (KEGG database)
From PPI & pathways to protein networks
build reliable networks with biological meaning: example 3

NOTCH SIGNALING PATHWAY: hsa04330 (KEGG database)
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:**
transforms 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

types of relation
– activation
– inhibition
– expression
– other

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

NOTCH pathway-network

Javier De Las Rivas - CiC (USAL/CSIC) - 2015
Tissue-specificity: **nodes in yellow** correspond to proteins expressed in liver

**NOTCH pathway-network**

**types of relation**
- activation
- inhibition
- expression
- other

**NOTCH pathway-network in the LIVER**
Tissue-specificity: nodes in yellow correspond to proteins expressed in ...
Tissue-specificity: **nodes in yellow** correspond to proteins expressed in ...
From PPI & pathways to protein networks

NOTCH pathway-network

NOTCH pathway-network in the LIVER

NOTCH pathway-network global

NOTCH pathway-network global in the LIVER
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

**NOTCH pathway-network**

global

Types of relation:
- activation
- inhibition
- expression
- other

**NOTCH pathway-network in the LIVER**
From PPI & pathways to protein networks
From PPI & pathways to protein networks
From PPI & pathways to protein networks
From PPI & pathways to protein networks

NOTCH pathway-network

NOTCH pathway-network in the LIVER

NOTCH pathway-network global

NOTCH pathway-network global in the LIVER
From PPI & pathways to protein networks

The network analysis confirms the central nodes of the pathway

<table>
<thead>
<tr>
<th>Protein_ID</th>
<th>Degree</th>
<th>Betweenness</th>
<th>Eigenvector</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1.] RBPJL_HUMAN</td>
<td>13</td>
<td>187.97</td>
<td>0.529</td>
</tr>
<tr>
<td>[3.] NOTC1_HUMAN</td>
<td>22</td>
<td>122.57</td>
<td>1.0</td>
</tr>
<tr>
<td>[4.] NOTC2_HUMAN</td>
<td>22</td>
<td>122.57</td>
<td>1.0</td>
</tr>
<tr>
<td>[6.] NOTC4_HUMAN</td>
<td>22</td>
<td>122.57</td>
<td>1.0</td>
</tr>
<tr>
<td>[7.] HDAC1_HUMAN</td>
<td>5</td>
<td>68.57</td>
<td>0.119</td>
</tr>
<tr>
<td>[8.] HDAC2_HUMAN</td>
<td>5</td>
<td>68.57</td>
<td>0.119</td>
</tr>
<tr>
<td>[9.] NCOA2_HUMAN</td>
<td>3</td>
<td>37.20</td>
<td>0.025</td>
</tr>
<tr>
<td>[10.] NICA_HUMAN</td>
<td>4</td>
<td>6.00</td>
<td>0.000</td>
</tr>
</tbody>
</table>

SUH & NOTCH2 central nodes of the network

<table>
<thead>
<tr>
<th>Protein_ID</th>
<th>Degree</th>
<th>Betweenness</th>
<th>Eigenvector</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1.] RBPJL_HUMAN</td>
<td>13</td>
<td>1372.43</td>
<td>0.193</td>
</tr>
<tr>
<td>[5.] CTBP1_HUMAN</td>
<td>6</td>
<td>1023.66</td>
<td>0.005</td>
</tr>
<tr>
<td>[6.] CTBP2_HUMAN</td>
<td>6</td>
<td>1023.66</td>
<td>0.005</td>
</tr>
<tr>
<td>[7.] CBP_HUMAN</td>
<td>24</td>
<td>981.75</td>
<td>0.001</td>
</tr>
<tr>
<td>[8.] EP300_HUMAN</td>
<td>24</td>
<td>981.75</td>
<td>0.001</td>
</tr>
<tr>
<td>[9.] DVL1_HUMAN</td>
<td>41</td>
<td>956.94</td>
<td>1.0</td>
</tr>
<tr>
<td>[10.] DVL2_HUMAN</td>
<td>41</td>
<td>956.94</td>
<td>1.0</td>
</tr>
</tbody>
</table>

HDAC1/2 is enhanced in the global view
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