



Determinación y Análisis de Redes de Interacciones entre Proteínas

Florencio Pazos (CNB-CSIC)

Florencio Pazos Cabaleiro
Grupo de Bioinformática de Sistemas
Centro Nacional de Biotecnología (CNB-CSIC)
pazos@cnb.csic.es
<http://csbg.cnb.csic.es>

Pamplona
Septiembre 2010

Determinación y análisis de redes de interacciones entre proteínas

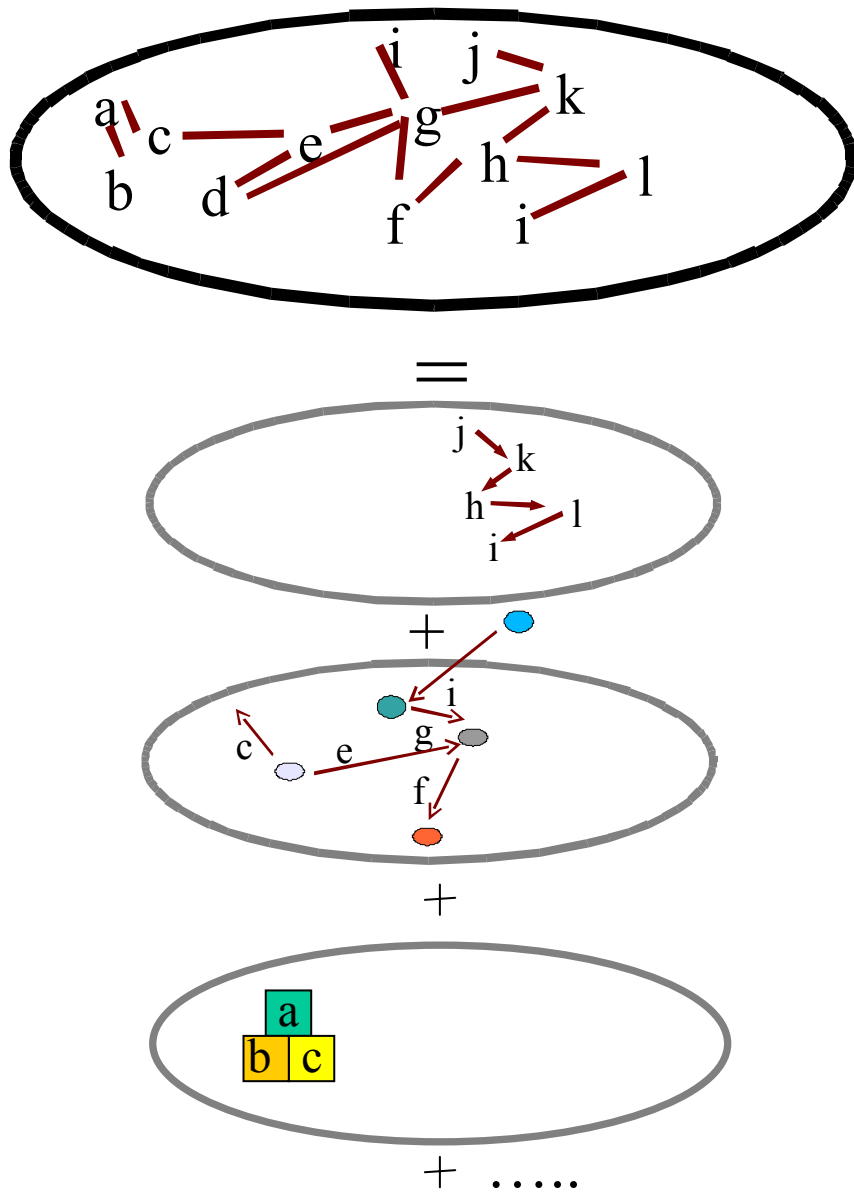
- El Interactoma
 - Determinación experimental masiva del interactoma
 - Calidad de los datos masivos de interacciones
 - Métodos computacionales para predicción de interacciones
 - Repositorios *on-line* de interacciones
 - Análisis del Interactoma
 - Topología
 - Predicción de función
 - Interactoma y enfermedades
 - Resumen
 - Bibliografía
-

Biología de Sistemas

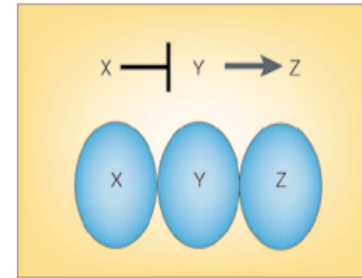
Caracterización masiva de componentes moleculares y sus relaciones

- Genome sequencing (“genome”).
- Transcript level characterization (mRNA) (“transcriptome”)
- Characteristics of the protein repertory (“proteome”)
- Cellular localization of the components (“localizome”)
- Gene regulation network (“regulome”)
- **Protein interaction network (“interactome”)**
- Massive gene-phenotype studies (“phenome”)
- Metabolic networks (“metabolome”)
-

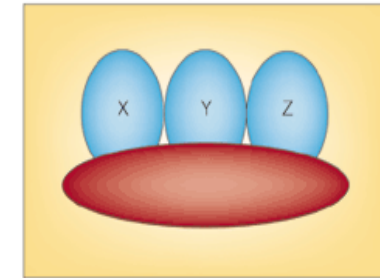
Interactoma



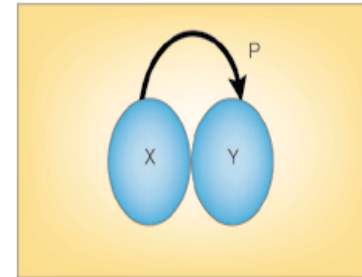
a Genetic pathways



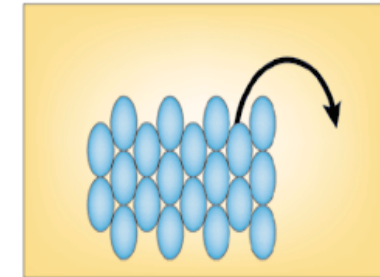
b Pathway scaffolding



c Enzymatic reactions

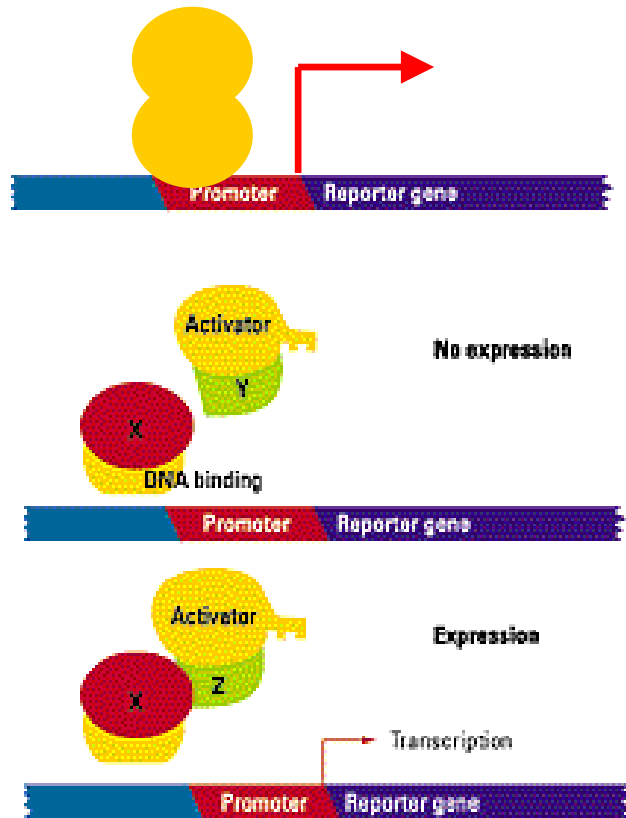


d Molecular machines

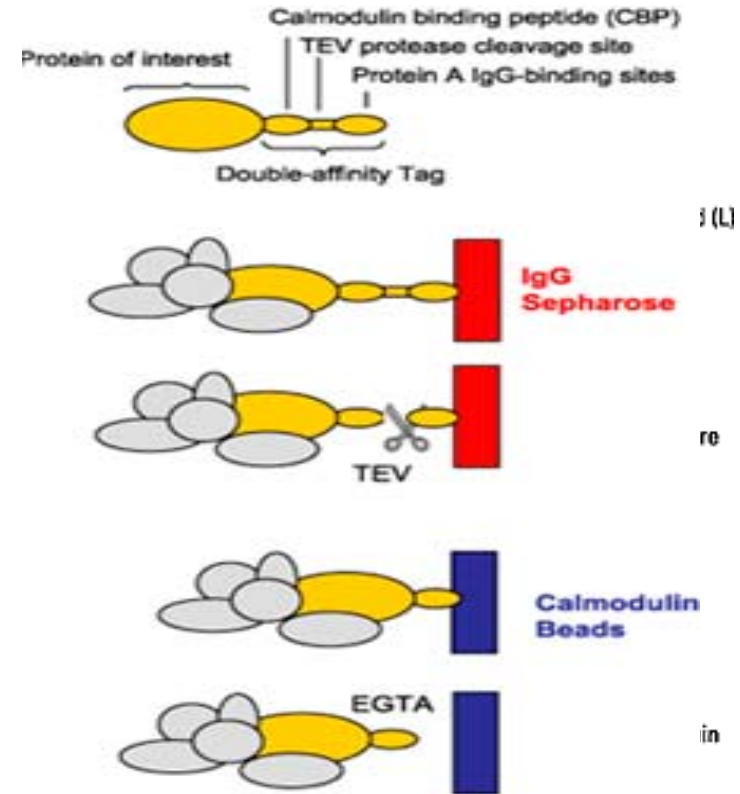


Nature Reviews | Molecular Cell Biology

Determinación Experimental Masiva del Interactoma



Y2H

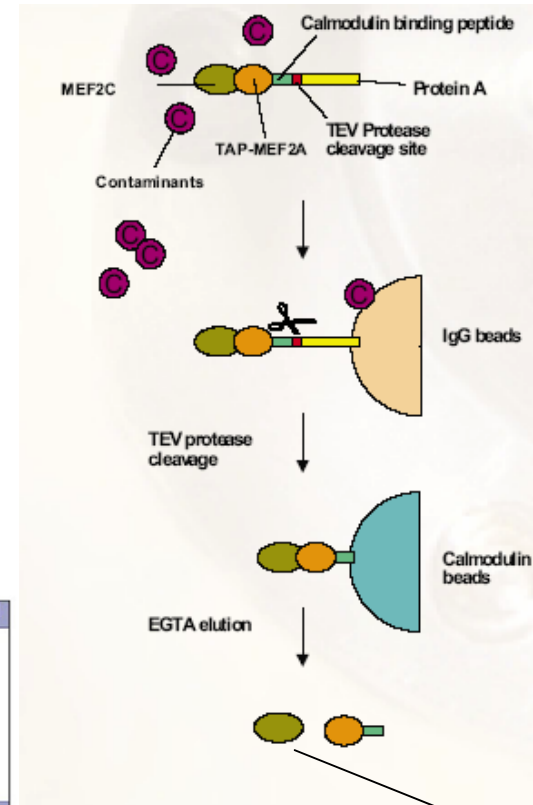
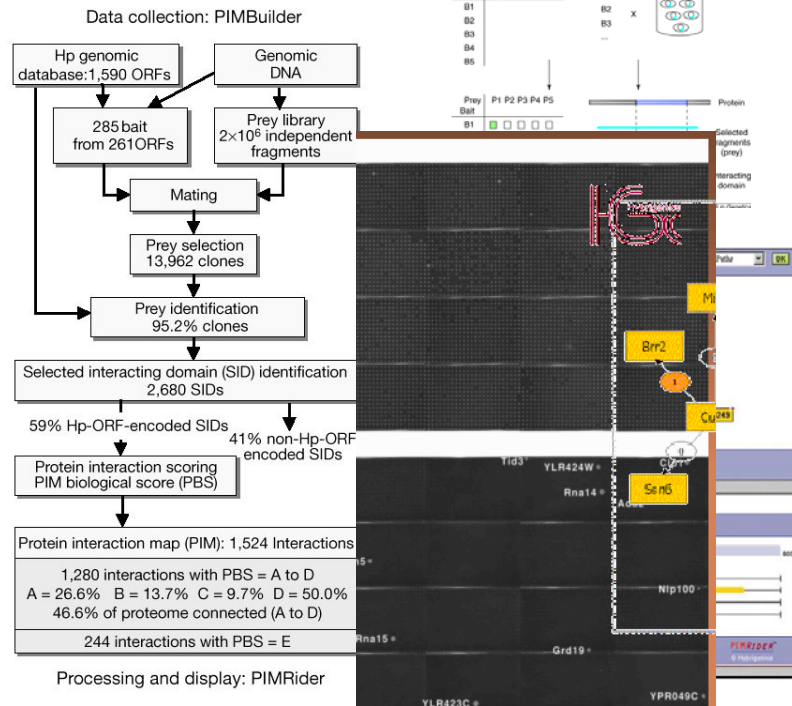
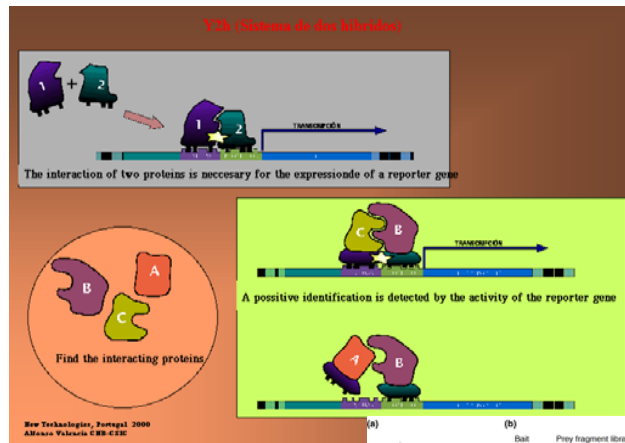


TAP

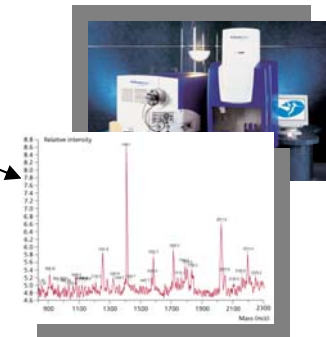
<http://pubs.acs.org/hotartcl/mdd/00/sep/edwards.html>

Determinación Experimental Masiva del Interactoma

Y2H

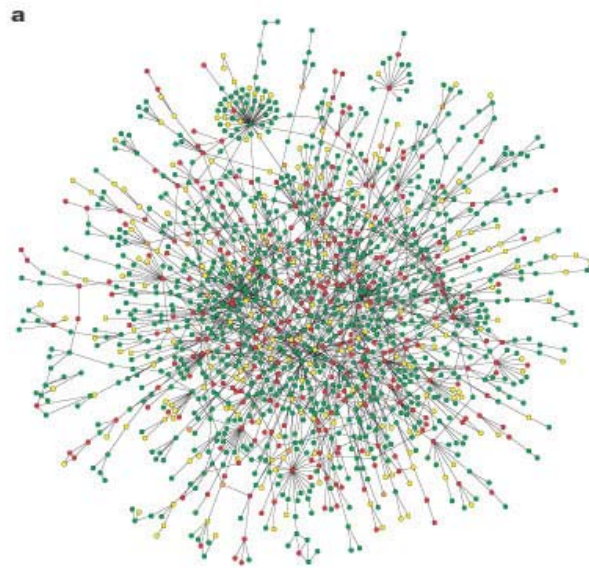


TAP/MS



A.Valencia

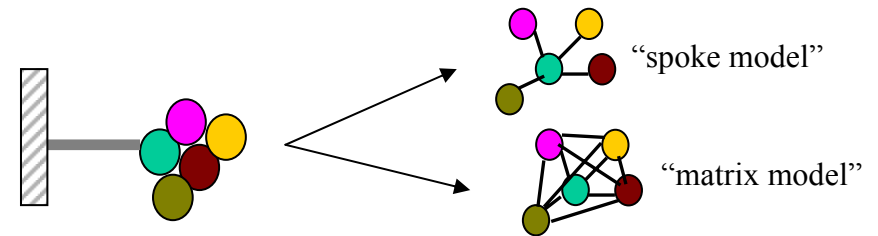
Interactomas determinados experimentalmente (*high throughput*)



- Rain, J.C., Selig, L., De Reuse, H., et al. (2001) The protein-protein interaction map of *Helicobacter pylori*. *Nature*, **409**, 211-215.
- Gavin, A.C., et al. (2002) Functional organisation of the yeast proteome by systematic analysis of protein complexes. *Nature*, **415**, 141-147.
- Ho, Y., et al. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*, **415**, 180-183.
- Ito, T., et al. (2000) Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc Natl Acad Sci USA*, **97**, 1143-1147.
- Uetz, P., et al. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature*, **403**, 623-631.
- Giot, L., Bader, J.S., Brouwer, et al. (2003) A protein interaction map of *Drosophila melanogaster*. *Science*, **302**, 1727-1736.
- Li, S., Armstrong, C.M., Bertin, N., et al. (2004) A map of the interactome network of the metazoan *C. elegans*. *Science*, **303**, 540-543.
- Butland, G., Peregrin-Alvarez, J.M., Li, J., et al. (2005) Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature*, **433**, 531-537.
- Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., et al. (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, **437**, 1173-1178.
- LaCount, D.J., Vignali, M., Chettier, R., Phansalkar, A., Bell, R., Hesselberth, J.R., Schoenfeld, L.W., Ota, I., Sahasrabudhe, S., Kurschner, C., et al. (2005) A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature*, **438**, 103-107.
- Uetz, P., Dong, Y.A., Zeretzke, C., Atzler, C., Baiker, A., Berger, B., Rajagopala, S.V., Roupelieva, M., Rose, D., Fossum, E., et al. (2006) Herpesviral protein networks and their interaction with the human proteome. *Science*, **311**, 239-242.

Interactomas determinados experimentalmente (*high throughput*)

	Number of Proteins	Number of Interactions
<i>H. pylori</i>		
Two-hybrid assays	710 [Rain <i>et al.</i> 2001]	1425
<i>E. coli</i>		
Co-immunoprecipitation/ Mass spectrometry	530 [Butland <i>et al.</i> 2005]	5420 (spoke)
<i>S. cerevisiae</i>		
Two-hybrid assays	934 [Uetz <i>et al.</i> 2000] 4131 [Ito <i>et al.</i> 2001]	854 3986
Co-immunoprecipitation/ Mass spectrometry	1361 [Gavin <i>et al.</i> 2002] 1560 [Ho <i>et al.</i> 2002]	3221 (spoke) 31304 (matrix) 3589 (spoke)
Synthetic lethal assays	1029 [Tong <i>et al.</i> 2004]	3627
DIP (small scale experiments)	1629	5068
<i>C. elegans</i>		
Two-hybrid assays	2898 [Li <i>et al.</i> 2004]	4027
<i>D. melanogaster</i>		
Two-hybrid assays	7048 [Giot <i>et al.</i> 2003]	20405
<i>H. sapiens</i>		
Co-immunoprecipitation/ Mass spectrometry	32 [Bouwmeester <i>et al.</i> 2004]	221
HPRD (small scale experiments)	2750 [Peri <i>et al.</i> 2004]	10534



Problemas con los datos experimentales *high-throughput*

Y2H

Falsos negativos:

- Interferencia de los **dominios fusionados** (AD y DB).
- Interacciones 1 Vs 1, no tiene en cuenta **efectos cooperativos**.

Falsos positivos:

- Interacciones mediadas por **terceras proteínas**.
- Es capaz de obtener **interacciones lábiles**, pero esto lo hace más vulnerable a uniones **inexpecíficas**.
- El ensayo es *in vivo*, pero en **condiciones artificiales** (sobre-expresión, forzado en el núcleo, etc)
- **Baja reproducibilidad**.

Interacciones **binarias, directas y lábiles**

↑% error

TAP/MS

Falsos negativos:

- Interferencia del **TAP-cassette** en la interacción (~18% de las proteínas no son funcionalmente viables).
- Proteínas **no expresadas** en el momento de la lisis (se ha relacionado con la concentración de mRNA).
- Sesgo en contra de **proteínas pequeñas** (<15K).
- Detecta principalmente **interacciones estables** (se pierde las lábiles).
- En mamíferos requiere **sobre-expresión**.

Falsos positivos:

- Proteínas **pegajosas**.
- Se estima que produce un **70% de complejos reproducibles**.
- Interacciones establecidas durante la **lisis**.

Complejos, interacciones estables e indirectas

↓% error

Problemas con los datos experimentales *high-throughput*

Solapamiento

- Complejos gran escala
- Y2H gran escala
- Text Mining
- Experimentos pequeña escala

Data set	Proteins	Interactions	Homodimers
Ho "spoke"	1,578	3,618	0
Ho "matrix"	1,578	28,252	1,578
Gavin "spoke"	1,363	3,225	0
Gavin "matrix"	1,363	18,677	1,363
Uetz	1,001	946	43
Ito "full"	3,274	4,468	82
Ito "core"	796	805	52
PreBIND	859	1,196	0
MIPS	964	1,353	51
YPD	1,538	2,205	283
MIPS + PB + YPD	1,762	3,310	303

Proteins\interactions\homodimers shared by datasets

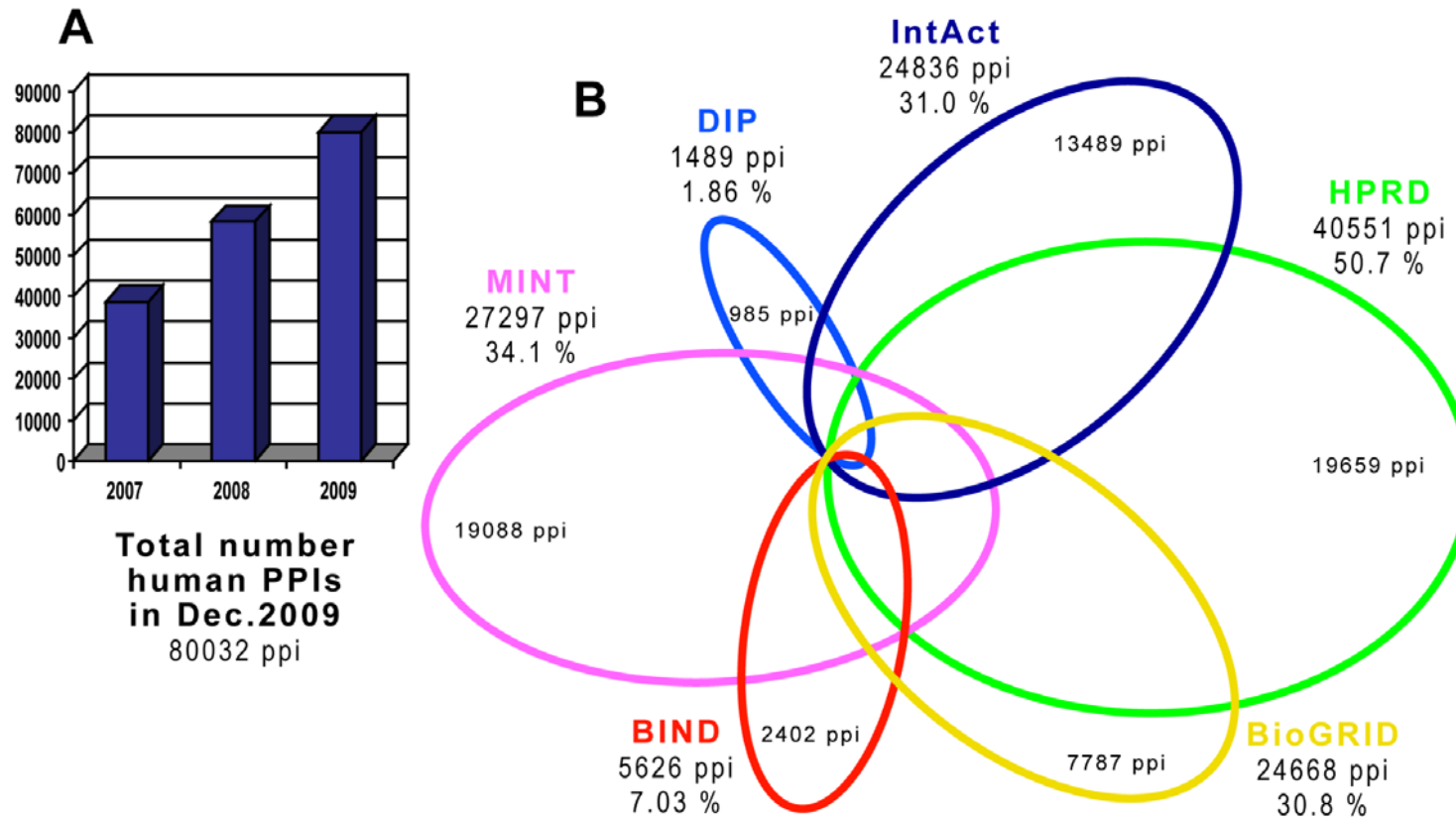
Data set	MIPS+PB+YPD	YPD	MIPS	PreBIND	Ito core	Ito full	Uetz	Gavin matrix	Gavin spoke	Ho matrix
Ho "spoke"	265\210\0	230\168\0	161\119\0	169\113\0	71\41\0	109\64\0	88\55\0	333\366\0	222\198\0	1578\3618\0
Ho "matrix"	448\480\135	385\357\126	226\202\21	246\192\0	101\69\13	162\117\22	120\86\12	658\2230\658	362\549\0	
Gavin "spoke"	361\333\0	276\198\0	249\230\0	163\117\0	71\40\0	97\55\0	78\47\0	1363\3225\0		
Gavin "matrix"	537\691\121	452\418\111	319\412\23	227\188\0	118\73\5	182\122\15	134\91\9			
Uetz	168\106\3	142\86\3	117\70\1	77\47\0	201\133\10	276\187\15				
Ito "full"	205\135\10	175\112\10	114\69\1	94\54\0	796\804\52					
Ito "core"	127\82\7	109\68\7	76\46\1	61\35\0						
PreBIND	859\1196\0	579\554\0	442\402\0							
MIPS	964\1353\51	803\834\31								
YPD	1538\2205\283									

Problemas con los datos experimentales *high-throughput*

Solapamiento

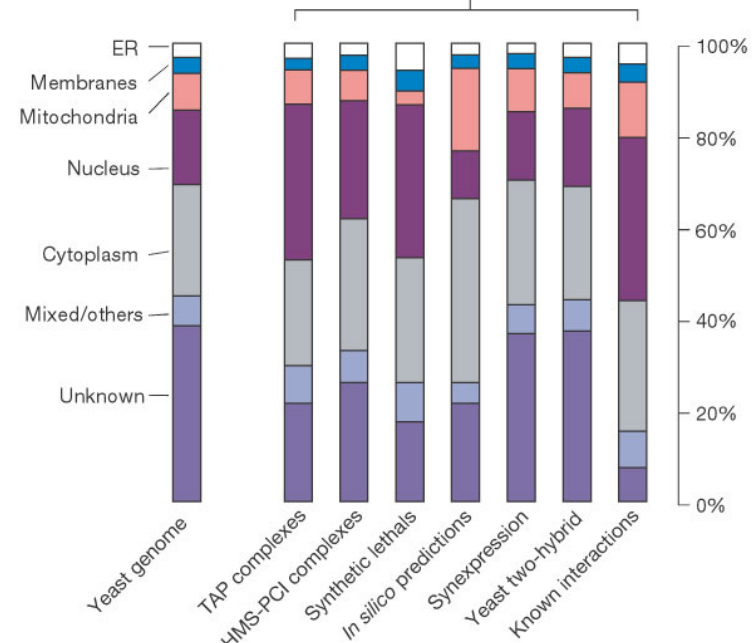
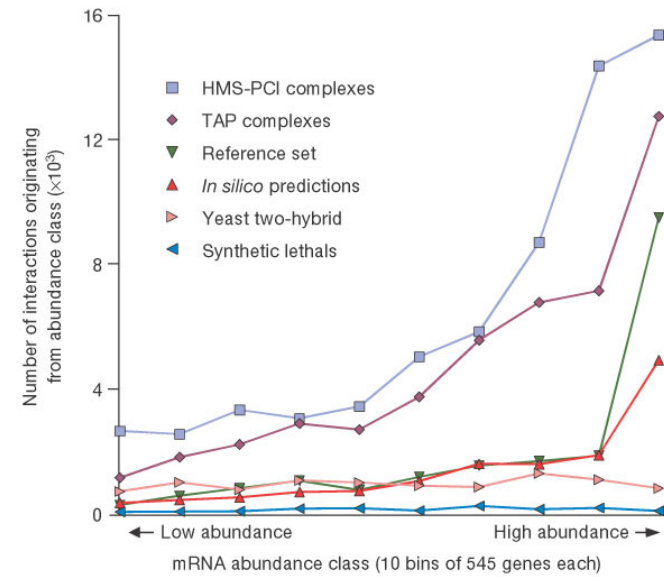
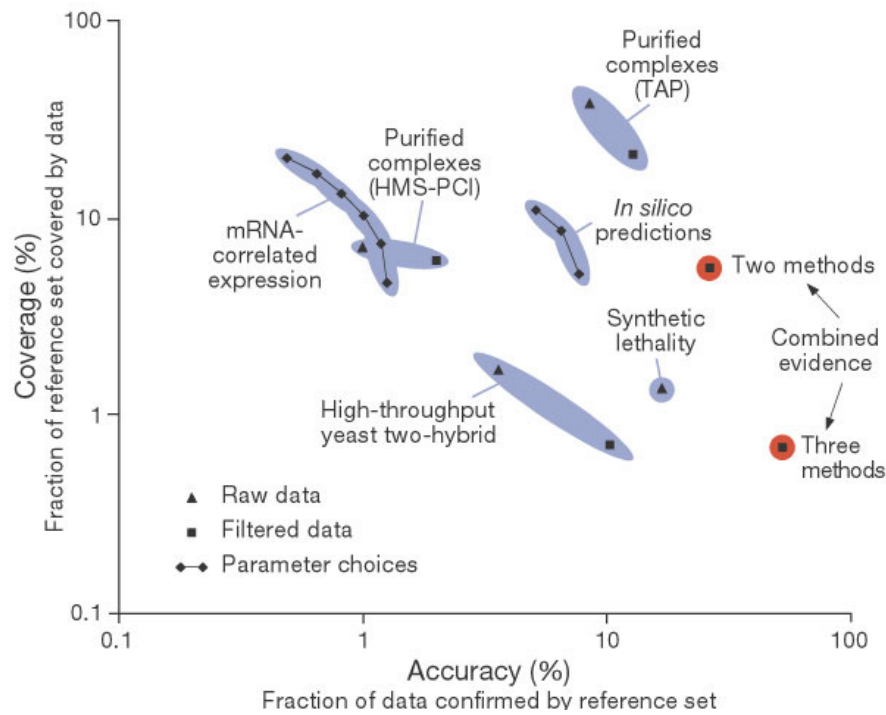
Human Interactome

Coverage of human PPIs on major public repositories



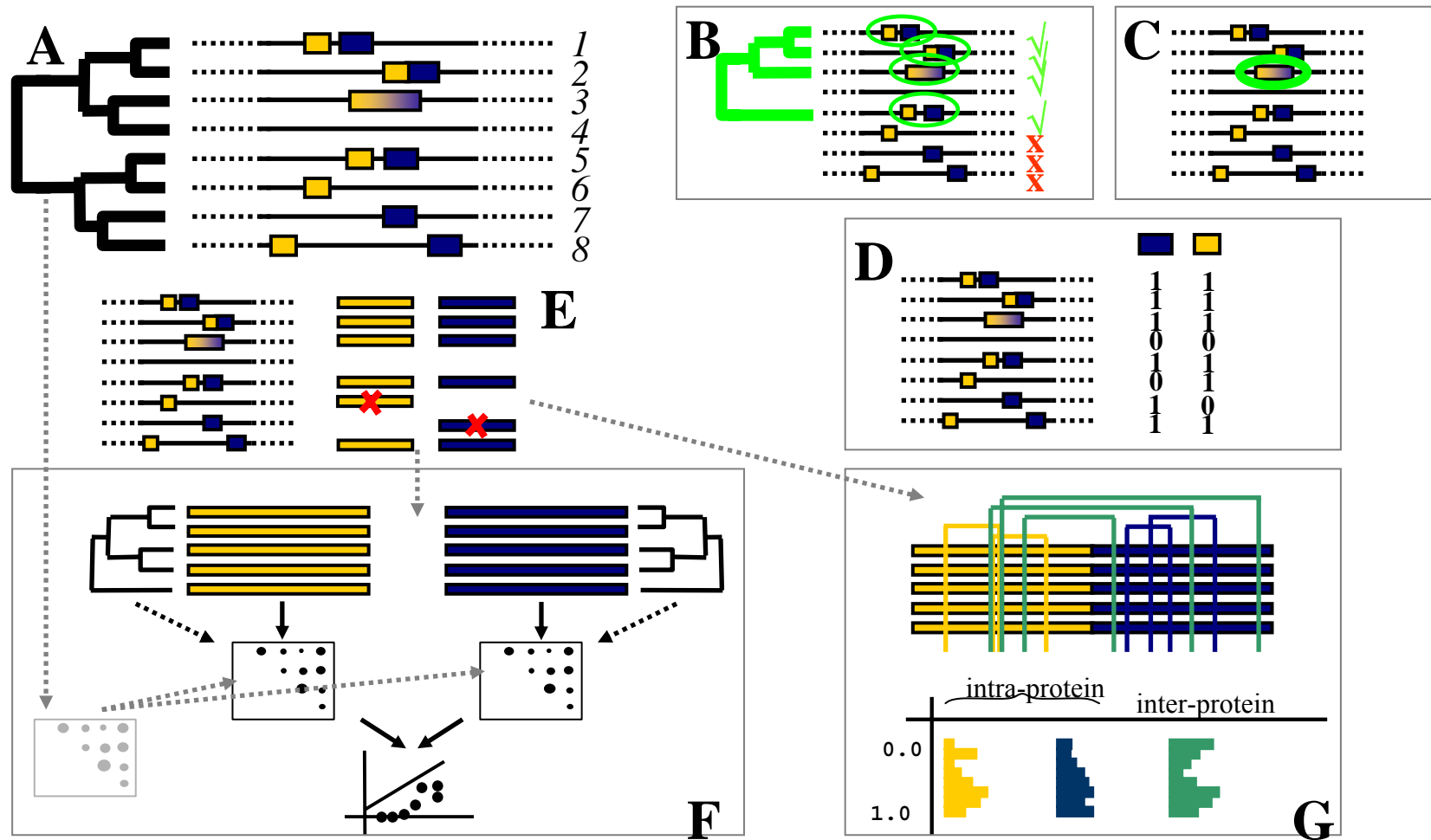
Problemas con los datos experimentales *high-throughput*

Calidad



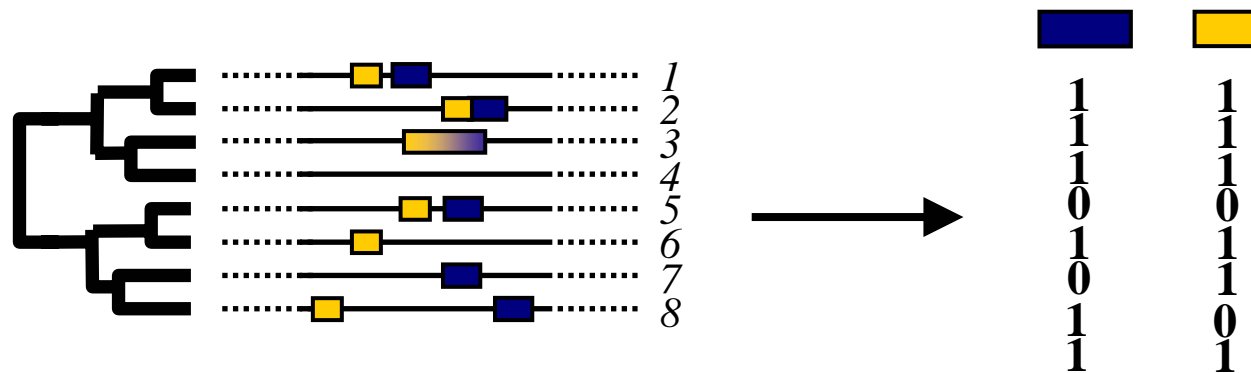
von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S. and Bork, P. (2002) Comparative assessment of large scale data sets of protein-protein interactions. *Nature*, **417**, 399-403.

Métodos Computacionales de Predicción de Compañeros de Interacción



- Alfonso Valencia & Florencio Pazos (2002). Prediction of Protein Interactions with Computational Methods. *Curr Op Str Biol.* **12(3)**: 368-373.
- Salwinski, L. & Eisenberg, D. (2003). Computational methods of analysis of protein-protein interactions. *Curr Opin Struct Biol.* **13**, 377-382.
- Shoemaker, B.A. and Panchenko, A.R. (2007) Deciphering protein-protein interactions. Part II. Computational methods to predict protein and domain interaction partners. *PLoS Comput Biol.*, **3**, e43

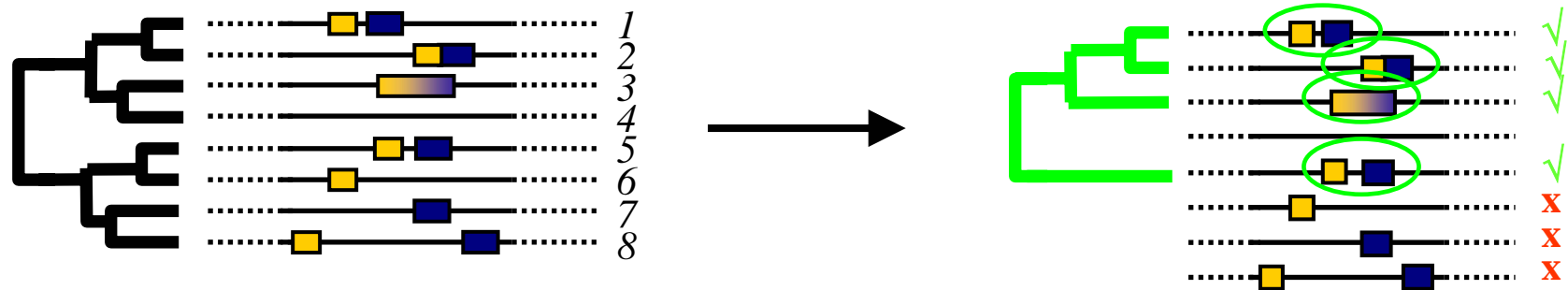
Perfiles Filogenéticos



•Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. (1999). Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. *Proc Natl Acad Sci USA*. **96**, 4285-4288.

•Date, S. V. & Marcotte, E. M. (2003). Discovery of uncharacterized cellular systems by genome-wide analysis of functional linkages. *Nat Biotechnol*. **21**, 1055-1062.

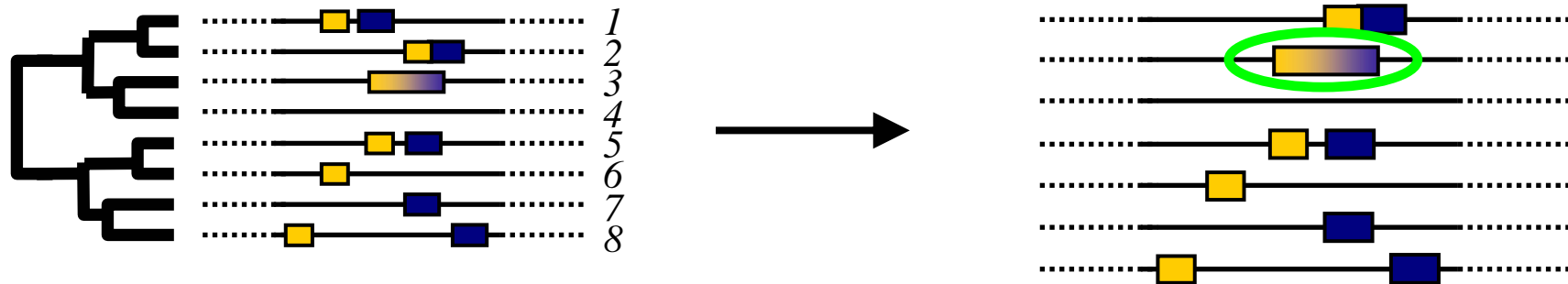
Conservación de Cercanía Genómica



Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D. & Maltsev, N. (1999). Use of contiguity on the chromosome to predict functional coupling. *In Silico Biol.* **1**, 93-108.

Dandekar, T., Snel, B., Huynen, M. & Bork, P. (1998). Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci.* **23**, 324-328.

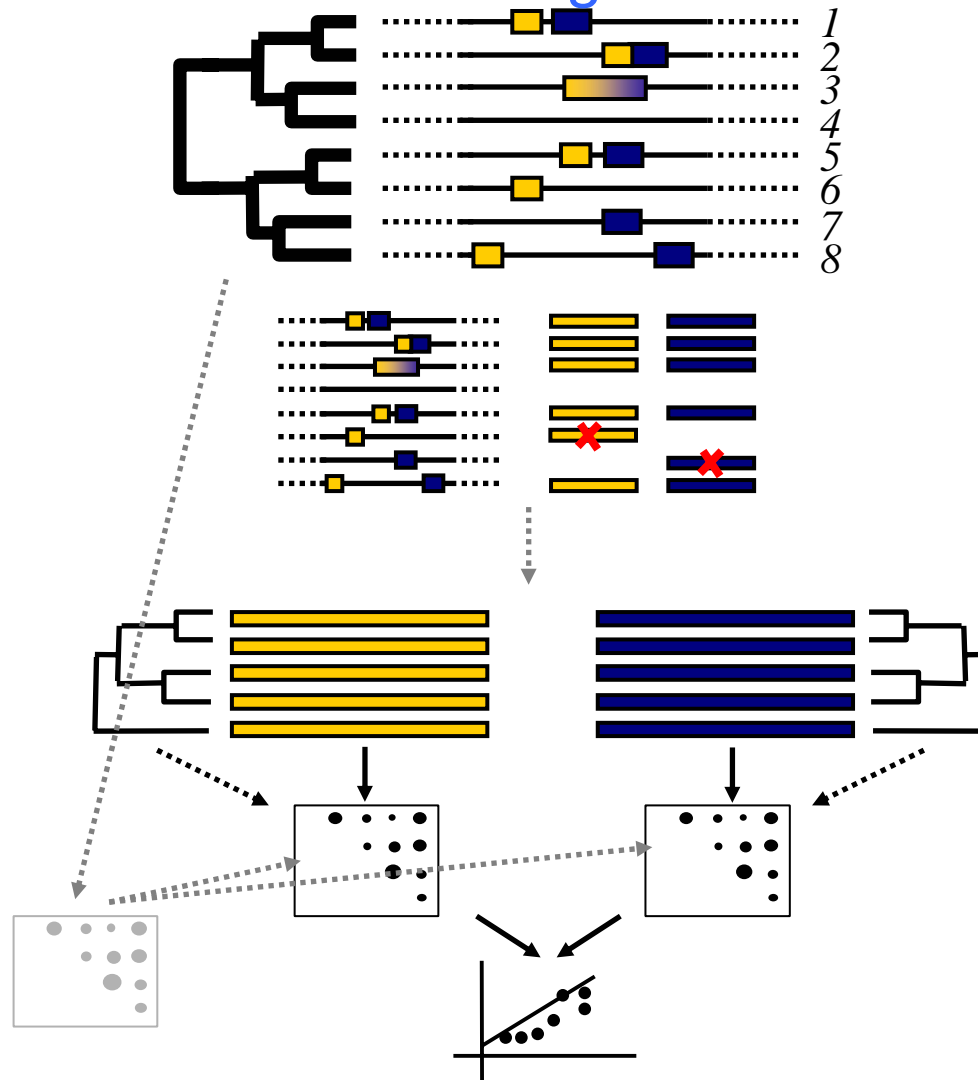
Fusion Génica



Enright, A. J., Iliopoulos, I., Kyripides, N. C. & Ouzounis, C. A. (1999). Protein interaction maps for complete genomes based on gene fusion events. *Nature*. **402**, 86-90.

Marcotte, E. M., Pellegrini, M., Ho-Leung, N., Rice, D. W., Yeates, T. O. & Eisenberg, D. (1999). Detecting protein function and protein-protein interactions from genome sequences. *Science*. **285**, 751-753.

Similitud de Árboles Filogenéticos - *MirrorTree*



Pazos, F. and Valencia, A. (2001) Similarity of phylogenetic trees as indicator of protein-protein interaction. *Protein Eng*, **14**, 609-614.

Pazos, F., Ranea, J.A.G., Juan, D. and Sternberg, M.J.E. (2005) Assessing Protein Co-evolution in the Context of the Tree of Life Assists in the Prediction of the Interactome. *J Mol Biol*, **352**, 1002-1015.

Juan, D., Pazos, F. and Valencia, A. (2008) High-confidence prediction of global interactomes based on genome-wide coevolutionary networks. *Proc Natl Acad Sci U S A.*, **105**, 934-939.

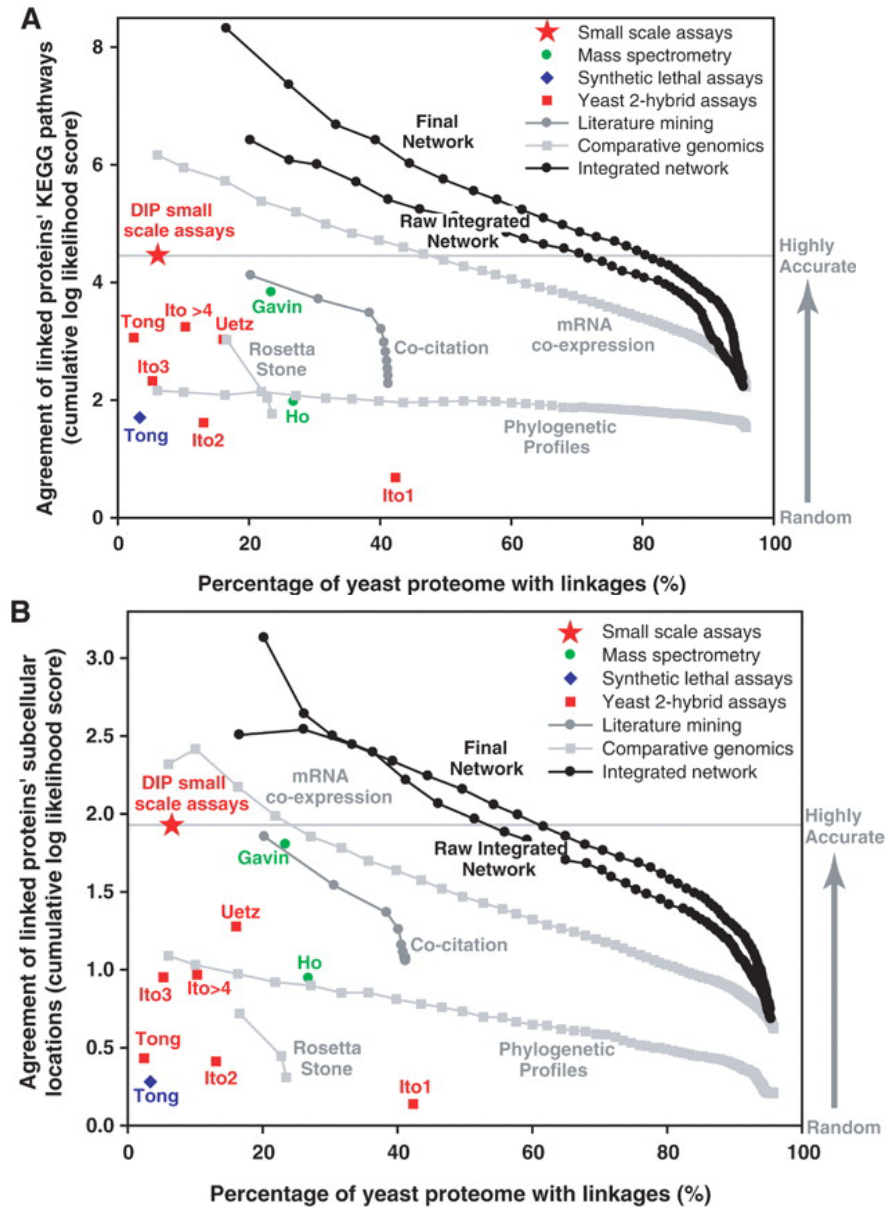
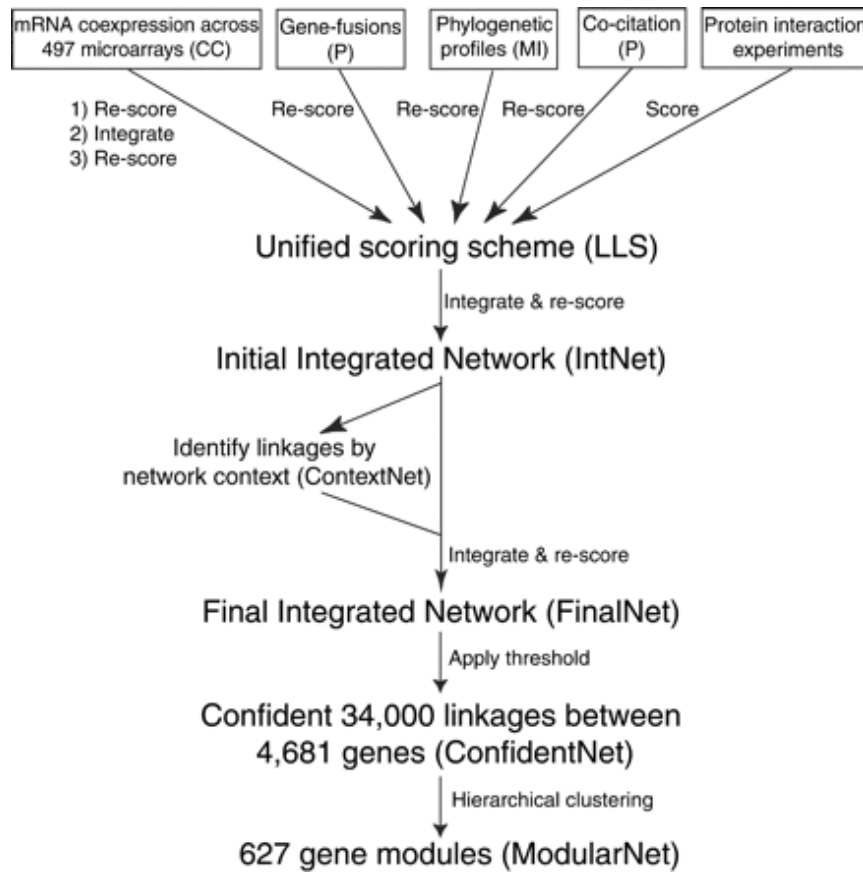
Problemas con las predicciones computacionales de interacción

- **La mayoría predicen interacciones entre “familias”, en vez de proteínas concretas => asumen “interologos”**
- Se requiere una señal clara a lo largo de varios organismos.
- Si la detección de proteínas ortólogas falla, no se encuentra la señal.
- **Algunos de estos métodos son difíciles de aplicar a eucariotas (problemas en la correcta localización de ortólogos, etc.).**
- Una interacción dada no necesariamente debe mostrar ninguno de los indicios usados.
- Las relaciones filogenéticas entre los organismos, suponen sesgos que pueden producir señales erróneas.
- **La mayoría de estos métodos no distinguen entre interacción física y asociación funcional.**

Otras evidencias de interacción o relación funcional

- Co-expresión
- Interacciones genéticas (*synthetic lethality*, etc.)
- Co-citación en la literatura (*text-mining*) – Trabajos describiendo interacciones obtenidas por métodos “clásicos” *low-throughput* (co-cristalización, biacore, ...).
- ...

Combinación de datos HT con otras fuentes de información



Lee, I., Date, S.V., Adai, A.T. and Marcotte, E.M. (2004) A probabilistic functional network of yeast genes. *Science*, **306**, 1555-1558.

Repositorio *on-line* de interacciones (físicas/funcionales-experimentales/predichas)

String: functional protein association networks - Mozilla

File Edit View Go Bookmarks Tools Window Help

Back Forward Reload Stop http://string.embl.de/ Search Print

Home Bookmarks SUSE LINUX The Mozilla Org...

Home · Download · Help/Info **STRING**

STRING - Search Tool for the Retrieval of Interacting Gen

Enter your gene/protein of interest ...

identifier: e.g. 'trpB', 'ANP1_YEAST', ...
you may also upload a list

alternatively, paste an amino-acid sequence:

interactors wanted:
GO! Reset COGs Proteins

What it does ...

STRING is a database of kn protein-protein interactions. The interactions include dir (functional) associations; th sources:

Genomic Context High-throughput Experiments

STRING quantitatively integ these sources for a large n transfers information betw applicable. The database cu proteins in 179 species.

References / Info ...

STRING uses orthology information from the excellent COG database (Ref). Up-to-date genomes and proteins are maintained at SWISSPROT and ENSEMBL. STRING references: von Mering et.al. 2005 / Snel et.al. 2000. Miscellaneous: Access Statistics, Robot Access Guide, Supported Browsers.

conserved genomic neighborhood

Escherichia coli (2 species)
Bifidobacterium longum
Leptospira interrogans
Borrelia burgdorferi
Treponema pallidum

phylogenetic co-occurrence

Cyanobacteria (3 species)
Thermotoga spirulina
Desulfohalobium radiodurans
Fusobacterium nucleatum
Haemifys. melioides
Chlorobium tepidum
Crenarchaeota (4 species)

literature co-occurrence

... inactivation of the epsilon subunit does not eliminate th assembly or stability of F₁ in the wild-type, *atp2* (Δ 1) and *atp2* (Δ 2) mutants. It is suggested that assembly or stability of F₁ in the wild-type, *atp2* (Δ 1) and *atp2* (Δ 2) mutants is suppressed by the *atp1* (Δ 1), *atp2* (Δ 1) and *atp2* (Δ 2) mutations respectively ...

database imports (knowledge)

annotated protein complex (APC) submitted info
ATPA_YEAST ATPG_YEAST ATPQ_YEAST ATPR_YEAST
ATPH_YEAST ATPK_YEAST [... and 12 other proteins]

high-throughput experiments

Chromatin immunoprecipitation submitted info
ATPA_YEAST UCR2_YEAST ATPB_YEAST ATPD_YEAST
UCR1_YEAST UCR3_YEAST ATPQ_YEAST ATPR_YEAST
CYL1_YEAST UCR1_YEAST UCR3_YEAST [... and 125 other proteins]

co-expression analysis

Heatmap showing expression levels for various genes across different conditions.

Your Input:

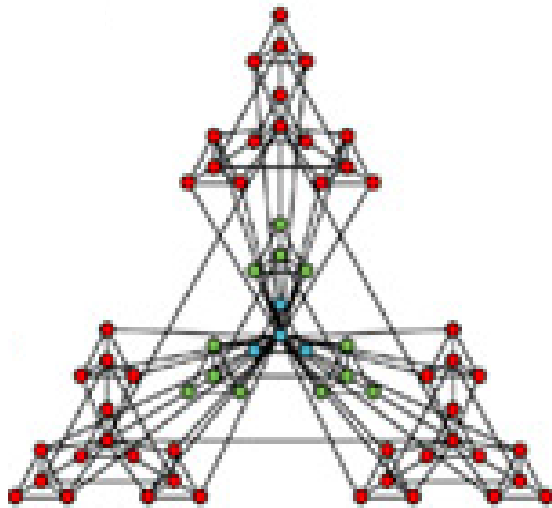
- ATPA_YEAST ATP synthase alpha chain, mitochondrial precursor (EC 3.6.3.14) [...]
- UCR2_YEAST Ubiquinol-cytochrome C reductase complex core protein 2, mitoch[...]

Predicted Functional Associations:

Association	Neighborhood	Gene Fusion	Co-occurrence	Co-expression	Experiments	Databases	Text Mining	ChemoLogit	Score
ATPB_YEAST ATP synthase beta chain, mitochondrial precursor (EC 3.6.3.14) [...]	●	●	●	●	●	●	●	●	0.999
ATPG_YEAST ATP synthase gamma chain, mitochondrial precursor (EC 3.6.3.14) [...]	●	●	●	●	●	●	●	●	0.999
UCR1_YEAST Ubiquinol-cytochrome C reductase complex core protein 1, mitoch[...]	●	●	●	●	●	●	●	●	0.999
UCR3_YEAST Ubiquinol-cytochrome C reductase iron-sulfur subunit, mitochond[...]	●	●	●	●	●	●	●	●	0.995
ATPD_YEAST ATP synthase oligomycin sensitivity conferral protein, mitochond[...]	●	●	●	●	●	●	●	●	0.986

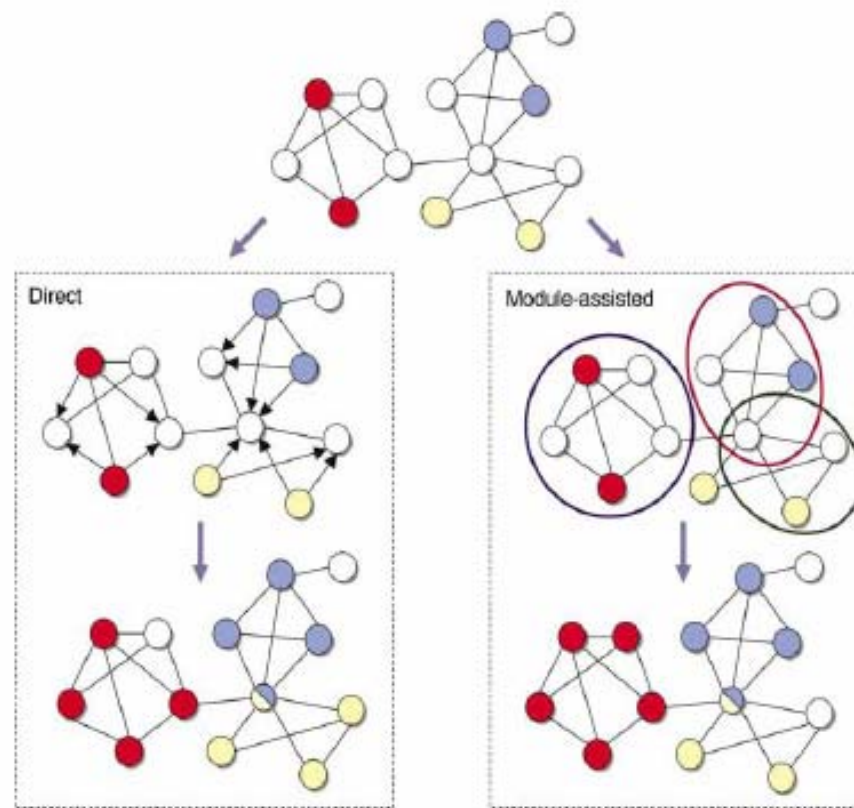
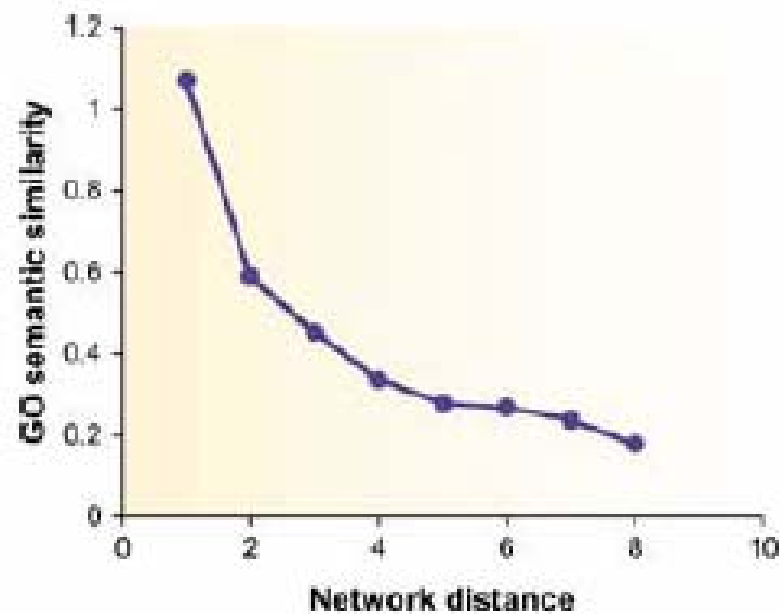
von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P. and Snel, B. (2003) STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res*, **31**, 258-261.

Topología del Interactoma

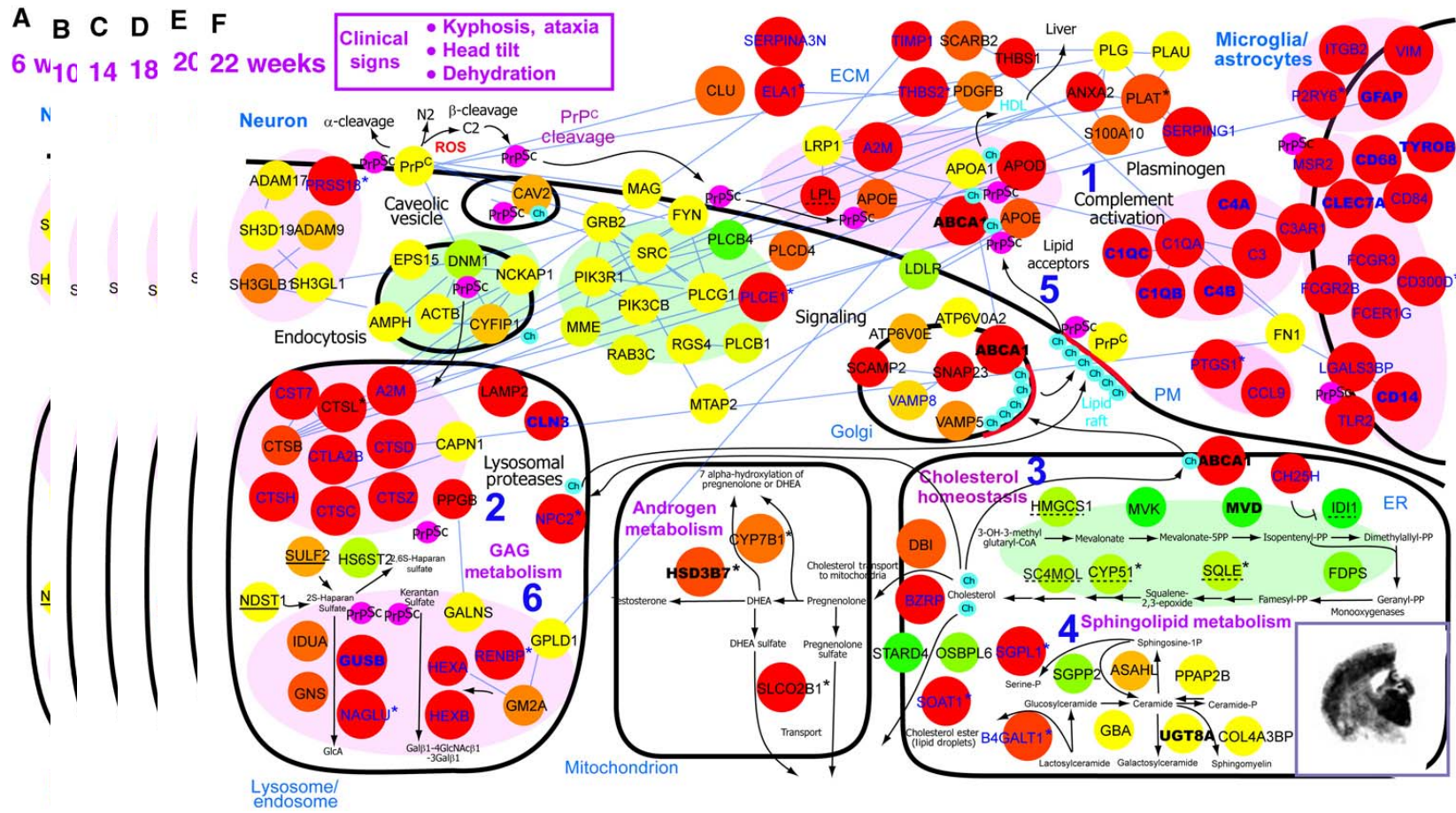


- scale-free / *small world* / jerarquica
=> resistente a fallos aleatorios; caminos cortos
Topología scale-free explicable por duplicaciones
- hubs: esenciales/conservados
(date/party)
- modulos topológicos $\langle \rangle$ modulos funcionales
- motivos (pequeños) funcionales conservados

Predicción de función basada en interacciones



Redes de Interacciones y Terapia/diagnóstico



Hwang D, Lee IY, Yoo H, Gehlenborg N, Cho JH, Petritis B, Baxter D, Pitstick R, Young R, Spicer D, Price ND, Hohmann JG, Dearmond SJ, Carlson GA, Hood LE. (2009). A systems approach to prion disease. *Mol Sys Biol.* 5:252

Resumen

- Es importante conocer el Interactoma de un organismo pues este engloba información muy relevante sobre su biología y permite realizar estudios sistémicos.
- Hay dos aproximaciones experimentales principales para la determinación experimental masiva de interactomas (Y2H y TAP/MS). Los resultados de ambas técnicas son diferentes y complementarios, y están sujetos a un alto grado de error (FP y FN).
- Existe una serie de técnicas computacionales de base genómica que predicen interacciones físicas y relaciones funcionales entre proteínas. Estas técnicas tienen sus propios problemas, limitaciones y rango de aplicación.
- Existen muchas otras evidencias indicativas de interacción (o relación funcional) entre proteínas. La tendencia es a combinar toda esta información (experimental, computacional, evidencias indirectas, ...) para obtener redes de proteínas de alta fiabilidad y cobertura.
- Existen repositorios on-line donde se puede consultar toda esta información relacionada con interacciones.
- El conocimiento de varios interactomas ha permitido los primeros estudios sistémicos de redes biológicas. De la topología y patrones de conectividad de estos interactomas se obtuvo conocimiento biológico no evidente en las propiedades individuales de las proteínas.
- Actualmente se están usando interactomas y otras redes biológicas como marcadores y posibles dianas para el tratamiento de enfermedades, sustituyendo a genes y proteínas individuales.

Bibliografía

- Alm, E. and Arkin, A.P. (2003) Biological networks. *Curr Opin Struct Biol*, **13**, 193-202.
- Xia, Y., Yu, H., Jansen, R., Seringhaus, M., Baxter, S., Greenbaum, D., Zhao, H. and Gerstein, M. (2004) Analyzing cellular biochemistry in terms of molecular networks. *Annu Rev Biochem*, **73**, 1051-1087.
- Uetz, P. and Finley, R.L., Jr. (2005) From protein networks to biological systems. *FEBS Lett*, **579**, 1821-1827.
- Barabasi, A.L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet*, **5**, 101-113.
- Bork, P., Jensen, L.J., von Mering, C., Ramani, A.K., Lee, I. and Marcotte, E.M. (2004) Protein interaction networks from yeast to human. *Curr Opin Struct Biol*, **14**, 292-299.
- Valencia, A. & Pazos, F. (2002) Computational methods for the prediction of protein interactions. *Curr Opin Struct Biol*, **12**, 368-373.
- Salwinski, L. & Eisenberg, D. (2003). Computational methods of analysis of protein-protein interactions. *Curr Opin Struct Biol*. **13**, 377-382.
- Shoemaker, B.A. and Panchenko, A.R. (2007) Deciphering protein-protein interactions. Part II. Computational methods to predict protein and domain interaction partners. *PLoS Comput Biol.*, **3**, e43.
- Shoemaker, B.A. and Panchenko, A.R. (2007) Deciphering protein-protein interactions. Part I. Experimental techniques and databases. *PLoS Comput Biol*, **3**, e42.
- Panchenko, A. R. and Przytycka, T. M. (eds.), Protein-protein interactions and networks (2008). Springer-Verlag, London.

Agradecimientos

- Grupo de Bioinformática de Sistemas (CNB-CSIC)
- D. Juan, A. Valencia (CNIO, Madrid)
- M. J. E. Sternberg, J-W Bang
(Imperial College, Londres)
- J. A. G. Ranea (U. Málaga)



Florencio Pazos Cabaleiro
<http://csbg.cnb.csic.es>
pazos@cnb.csic.es