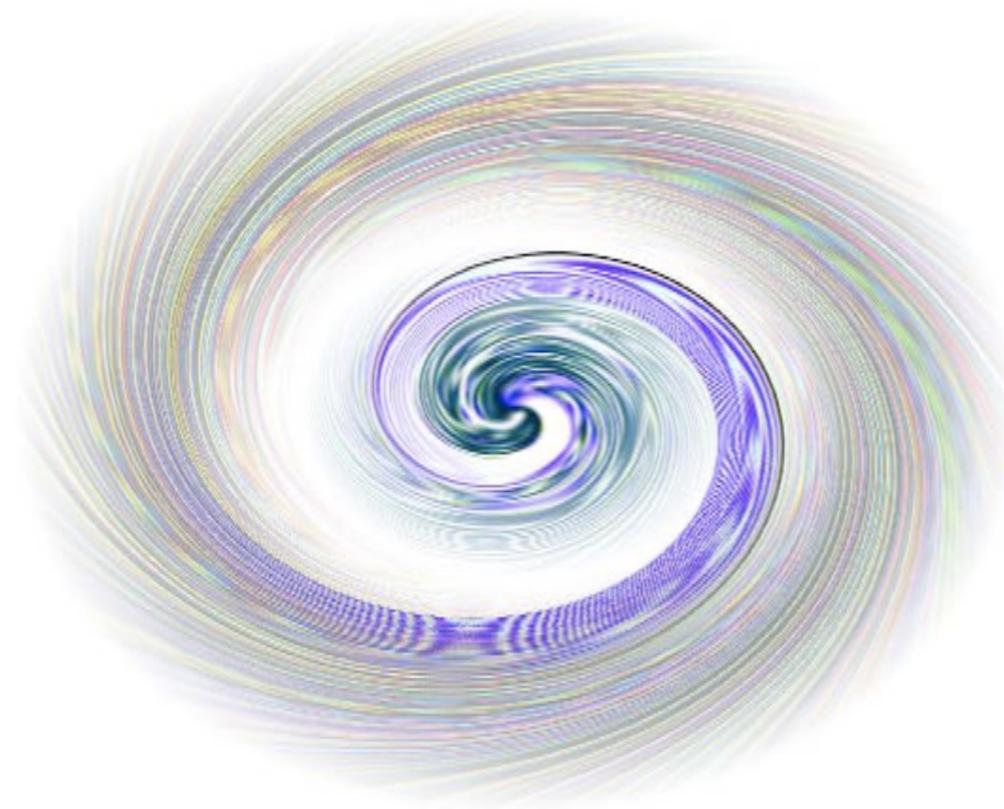


# Redes de Interacción de Proteínas



**David A. Juan**  
**CNIO. Grupo de Biología Computacional y Estructural.**  
[dadejuan@cniio.es](mailto:dadejuan@cniio.es)

# Esquema

1. Introducción
2. Datos Experimentales
3. Datos Computacionales
4. Calidad de los datos
5. Estudio de las redes
6. Más allá de las redes

# Introducción

## Esquema

- Introducción**
- Datos Experimentales
- Datos Computacionales
- Calidad de los datos
- Estudio de las redes
- Más allá de las redes

# Interacciones entre proteínas

- Fundamentales para comprender la **función** de las proteínas.
- Intervienen** en el control del ciclo celular, diferenciación celular, plegamiento de proteínas, señalización, transcripción, traducción, modificación post-traduccionales y transporte.
- Las interacciones **pueden** alterar las propiedades cinéticas de los enzimas, permitir la canalización de secuencias de reacciones, crear nuevos sitios de unión, inactivar o destruir una proteína, cambiar su especificidad, tener un papel regulatorio, etc.

Se pueden clasificar en:

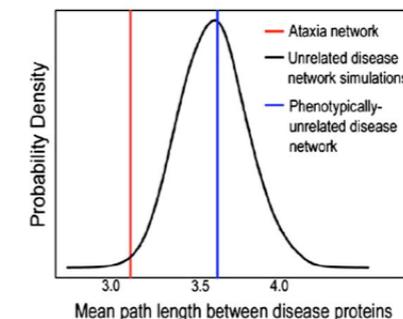
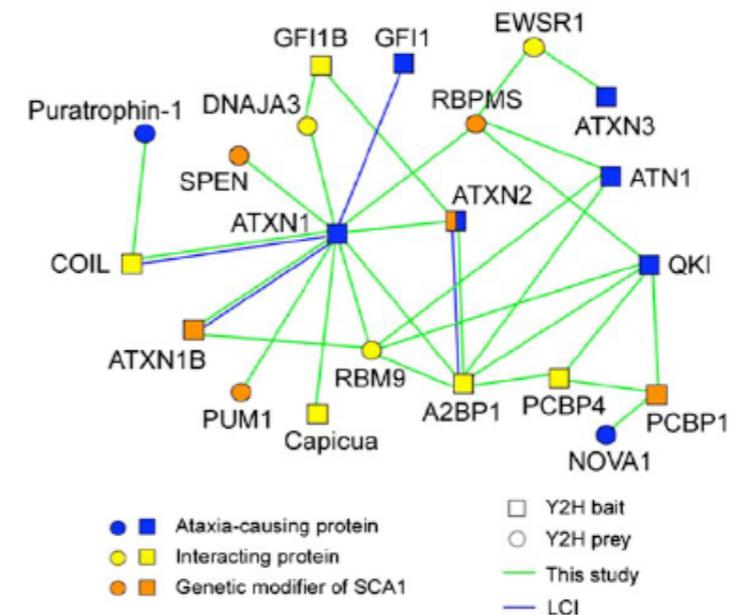
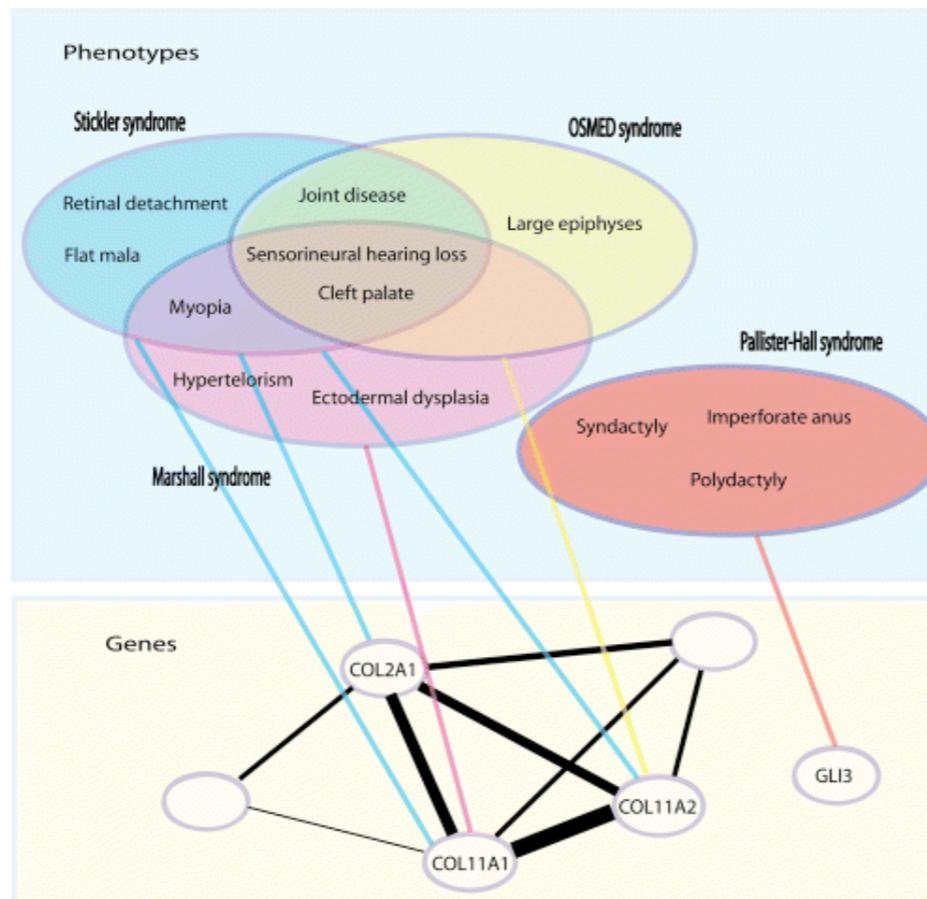
- Interacciones estables:** asociados a los complejos con múltiples subunidades (Hemoglobina, RNA polimerasa, etc).
- Interacciones transitorias:** implicadas en el control de la mayoría de los procesos celulares. Normalmente requieren de un conjunto de condiciones que provocan la reacción.

## Esquema

- Introducción
- Datos Experimentales
- Datos Computacionales
- Calidad de los datos
- Estudio de las redes
- Más allá de las redes

# Interacciones entre proteínas

- El estudio de proteínas o interacciones concretas no es suficiente para explicar algunos **fenotipos o procesos biológicos**.
- En muchos de estos procesos intervienen un número de proteínas, que forman una **red característica de interacciones**.
- La agrupación de enfermedades por **fenotipos** sugiere relaciones inesperadas a nivel de sub-redes de interacciones.



Lim et al. Cell. 2006.

# **Datos Experimentales**

## Esquema

- Introducción
- Datos Experimentales**
- Datos Computacionales
- Calidad de los datos
- Estudio de las redes
- Más allá de las redes

# Conjuntos masivos de interacciones detectadas experimentalmente

## Yeast two-hybrid

Uetz *et al.* Nature. 2000 (YEAST)  
Ito *et al.* PNAS. 2001 (YEAST)  
Giot *et al.* Science. 2003 (FLY)  
Li *et al.* Science. 2004 (WORM)  
Barrios-Rodiles *et al.* Science. 2005 (MAMMALIAN)  
Rual *et al.* Nature. 2005 (HUMAN)  
Stelzl *et al.* Cell. 2005 (HUMAN)

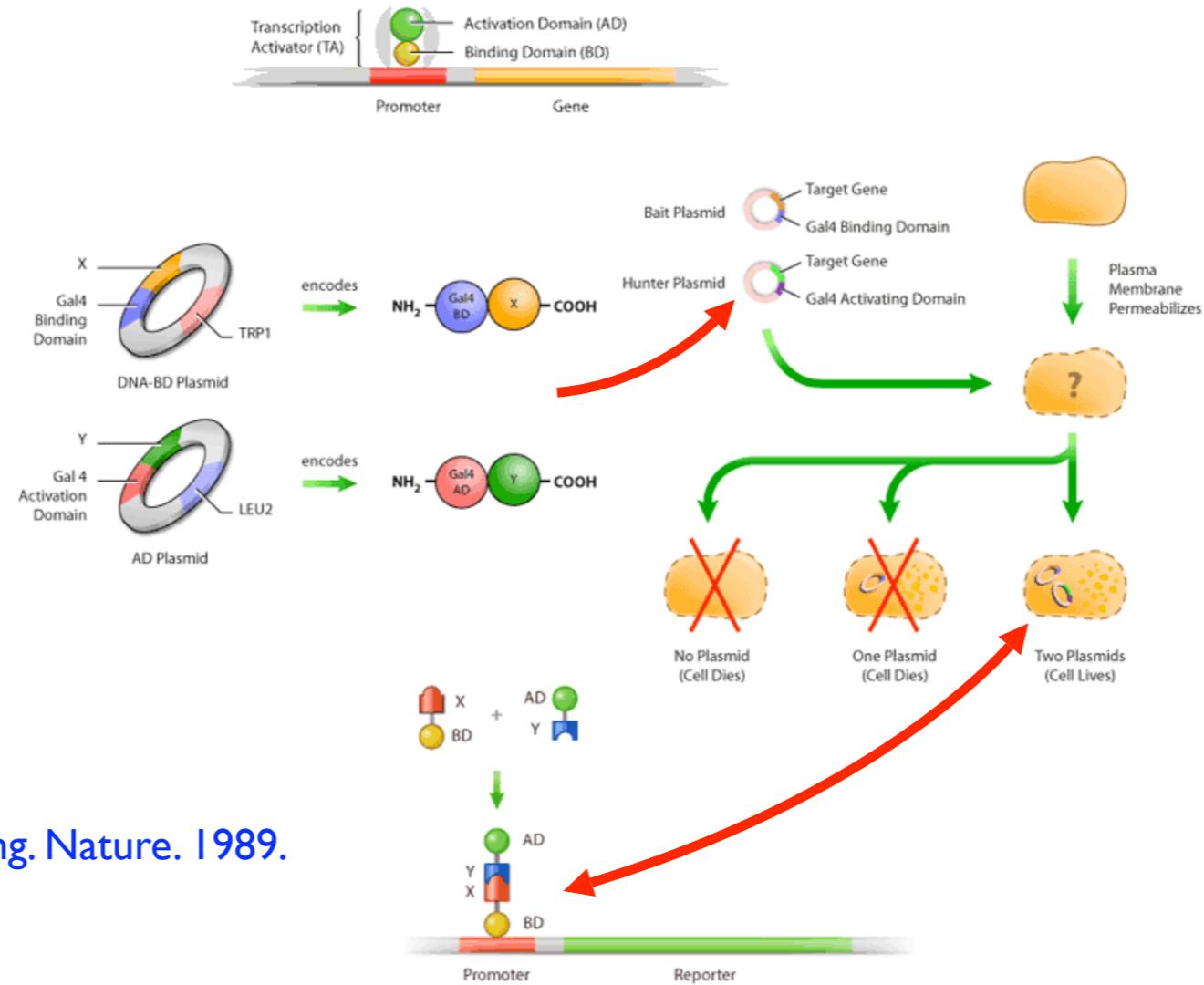
## Purificación de complejos

Gavin *et al.* Nature. 2002 (YEAST)  
Ho *et al.* Nature. 2002 (YEAST)  
Butland *et al.* Nature. 2005 (*E. coli*)  
Arifuzzaman *et al.* Genome Res. 2006 (*E. coli*)  
Ewing *et al.* Mol. Sys. Biol. 2007 (HUMAN)

## Esquema

- Introducción
- Datos Experimentales**
- Datos Computacionales
- Calidad de los datos
- Estudio de las redes
- Más allá de las redes

# Yeast two-hybrid

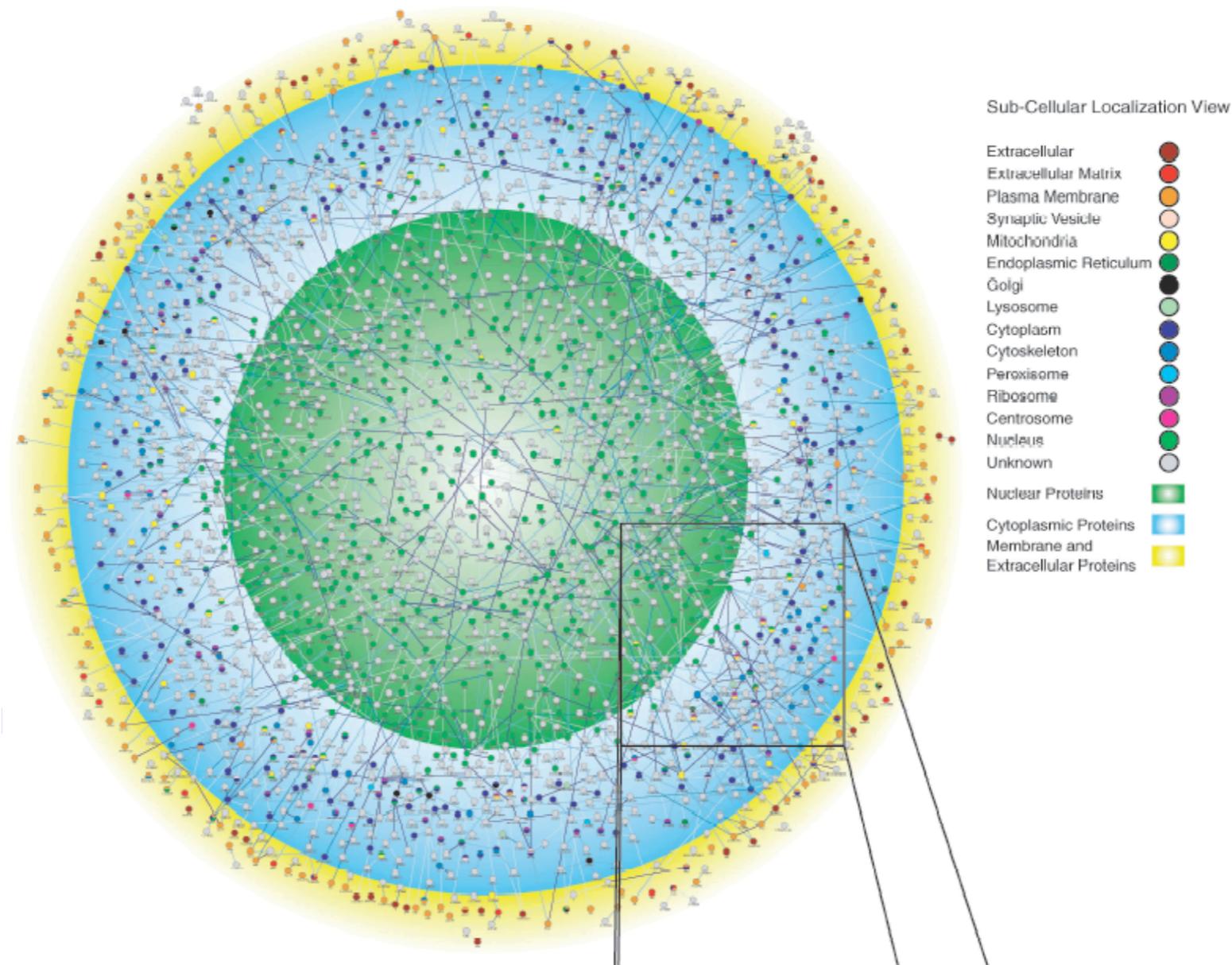


Fields & Song. Nature. 1989.

## Esquema

- Introducción
- Datos Experimentales**
- Datos Computacionales
- Calidad de los datos
- Estudio de las redes
- Más allá de las redes

# Yeast two-hybrid



## Esquema

- Introducción
- Datos Experimentales**
- Datos Computacionales
- Calidad de los datos
- Estudio de las redes
- Más allá de las redes

# Yeast two-hybrid

## Algunos problemas

### **Falsos negativos:**

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

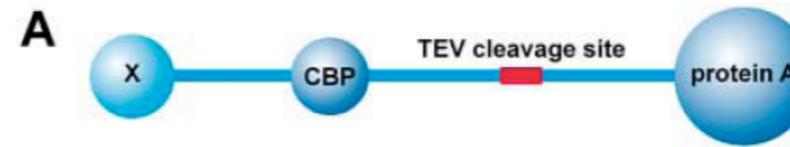
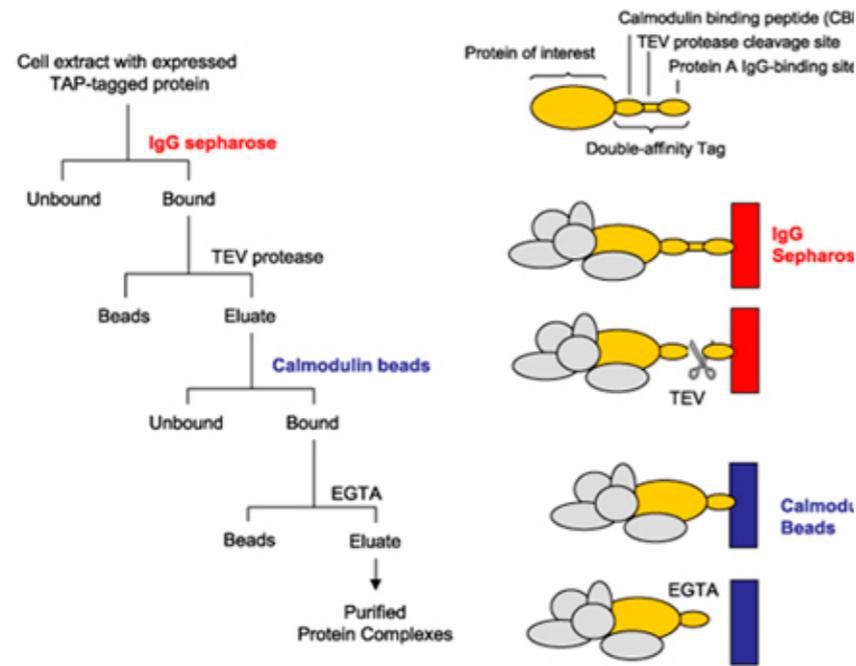
### **Falsos positivos:**

- 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.**

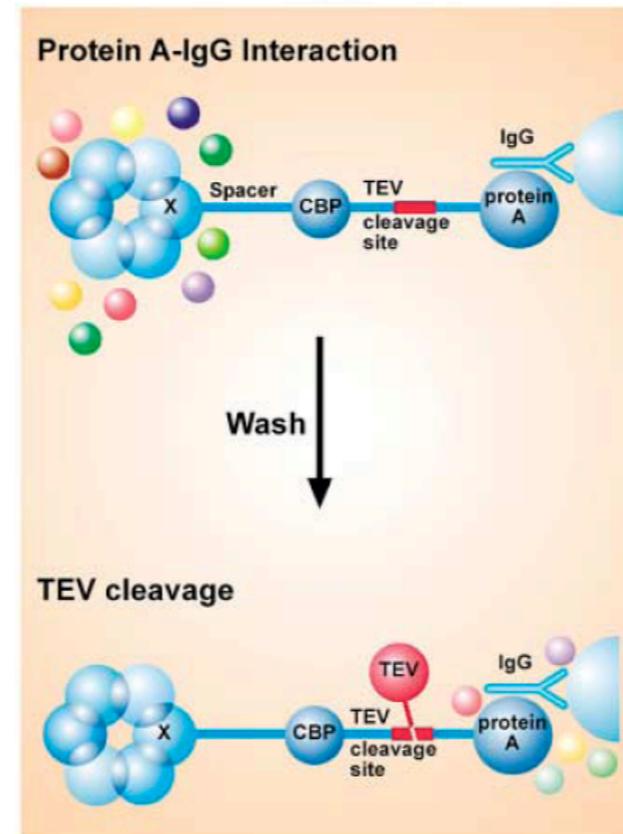
## Esquema

- Introducción
- Datos Experimentales**
- Datos Computacionales
- Calidad de los datos
- Estudio de las redes
- Más allá de las redes

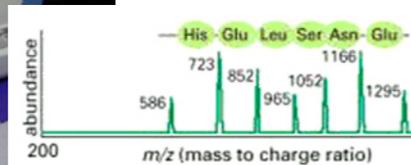
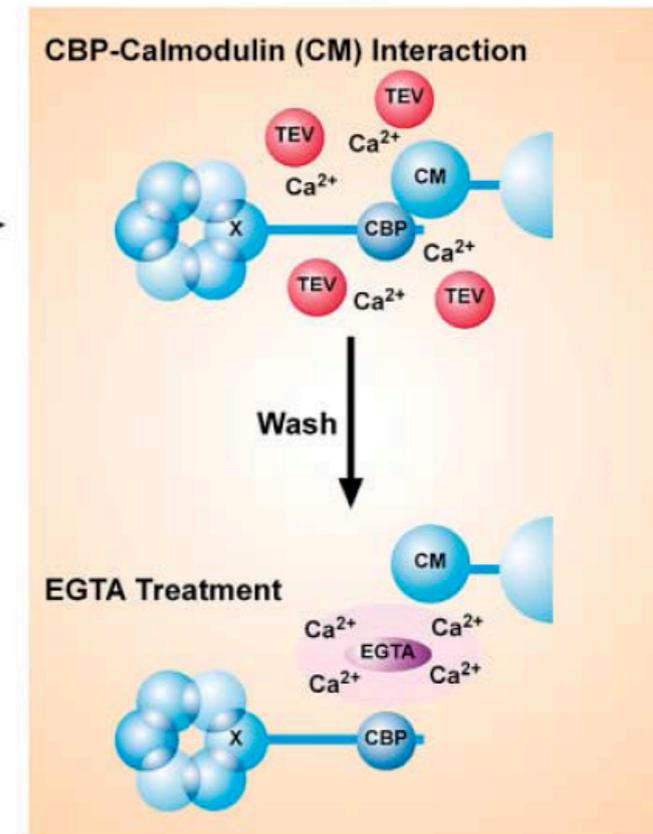
# Purificación de complejos (TAP-MS y HMS-PCI)



**B** First affinity purification



Second affinity purification



## Esquema

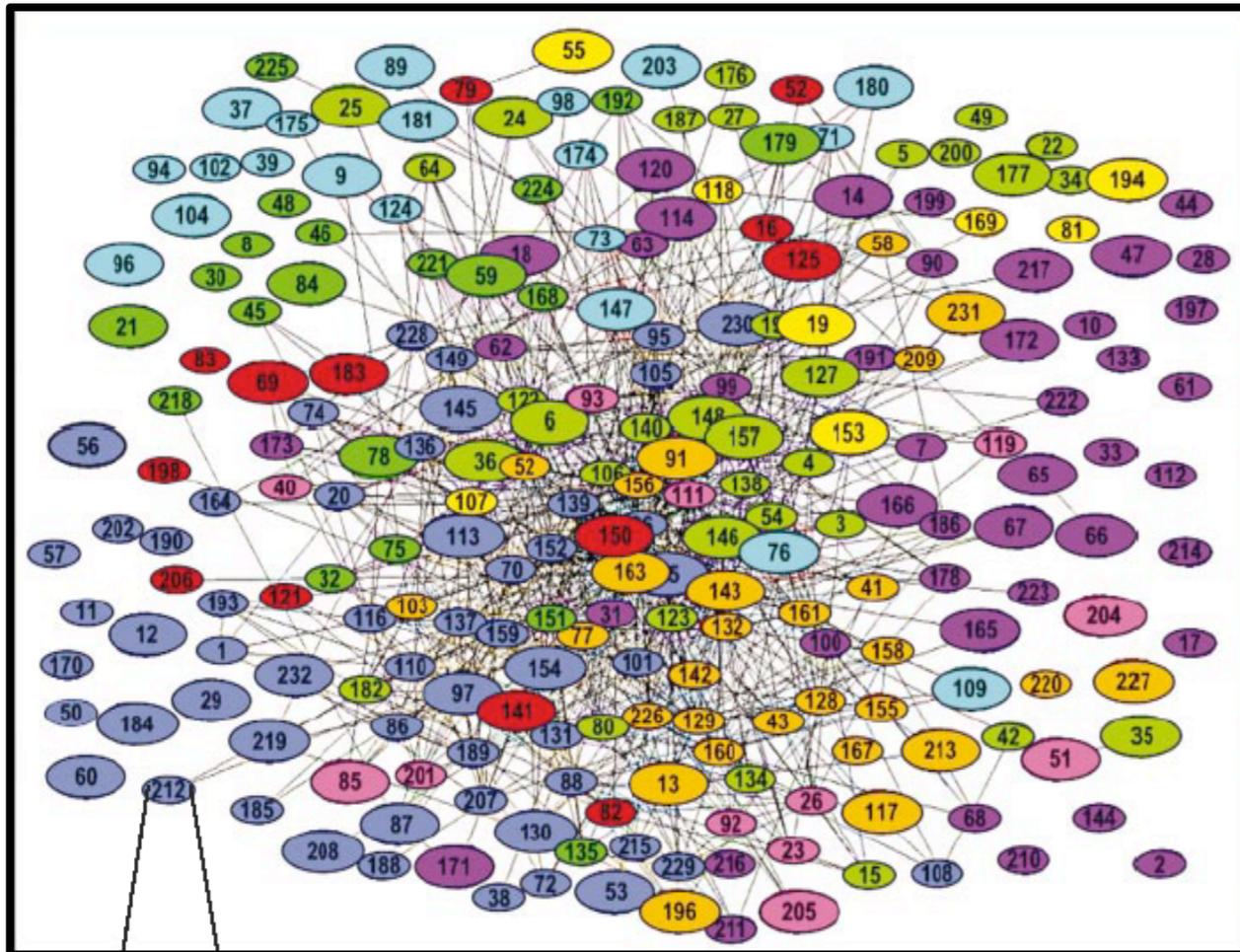
- Introducción
- Datos Experimentales**
- Datos Computacionales
- Calidad de los datos
- Estudio de las redes
- Más allá de las redes

# Purificación de complejos (TAP-MS y HMS-PCI)

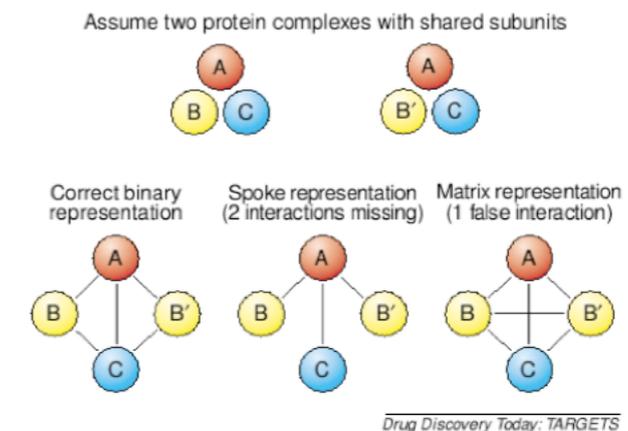
Se generan redes **diferentes** de las de y2h

**Nodos = complejos; Enlaces=comparten elementos**

Se desconoce la **topología** (interacciones prot-prot) de los complejos.



Modos de **representación binaria** de los datos obtenidos de complejos



## Esquema

- Introducción
- Datos Experimentales**
- Datos Computacionales
- Calidad de los datos
- Estudio de las redes
- Más allá de las redes

# Purificación de complejos

## Algunos problemas

### **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** (<15KD).
- 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**.



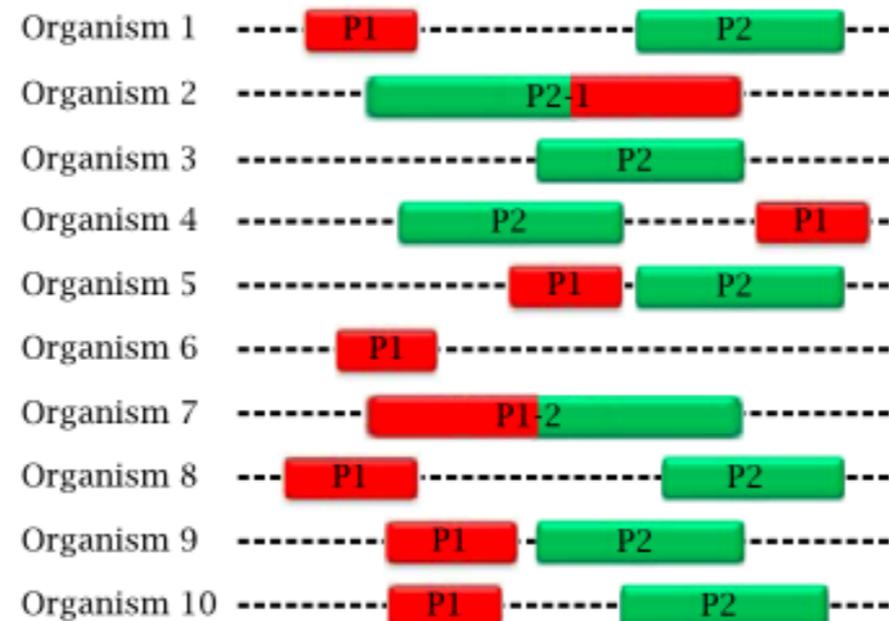
# **Datos Computacionales**

## Esquema

- Introducción
- Datos Experimentales
- Datos Computacionales**
- Calidad de los datos
- Estudio de las redes
- Más allá de las redes

# Interacción y Evolución

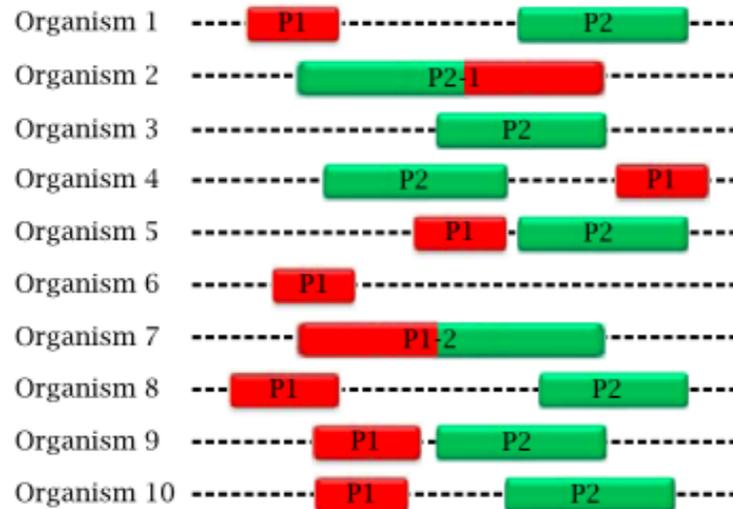
- Proteínas funcionalmente relacionadas tienden a presentar una **evolución coordinada**.
- Podemos considerar a la interacción como una **relación funcional** muy intensa.
- Esta evolución coordinada se detecta por comparación de determinadas **características génicas** a lo largo de un elevado número de organismos.
- Para ello comparamos grupos de proteínas homólogas, que desarrollan la **misma función en diferentes organismos**(equiválogas).



## Esquema

- Introducción
- Datos Experimentales
- Datos Computacionales**
- Calidad de los datos
- Estudio de las redes
- Más allá de las redes

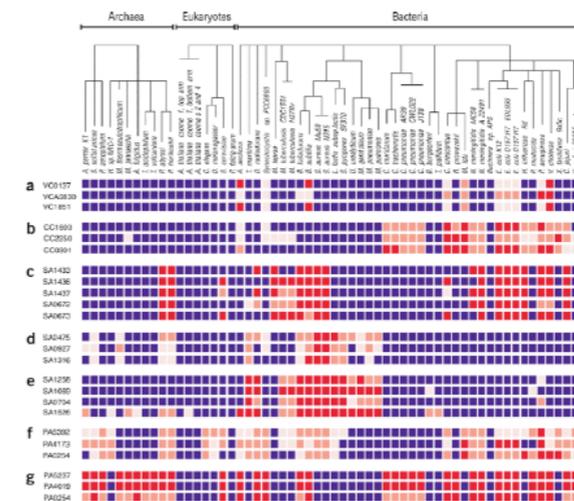
# Perfiles filogenéticos



(a) Phylogenetic Profiling



- Un perfil filogenético es un vector que define la **ausencia/presencia** de un representante de un conjunto de equivalentes en cada organismo.
- Las proteínas que interactúan han de estar en los mismos organismos
- La evolución tiende a **eliminar proteínas innecesarias**.
- Estos perfiles se pueden extender a parecidos entre una secuencia de referencia y sus ortólogos.



•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.

•Bowers, P.M., Cokus, S.J., Eisenberg, D. and Yeates, T.O. (2004) Use of logic relationships to decipher protein network organization. *Science*, **306**, 2246-2249.

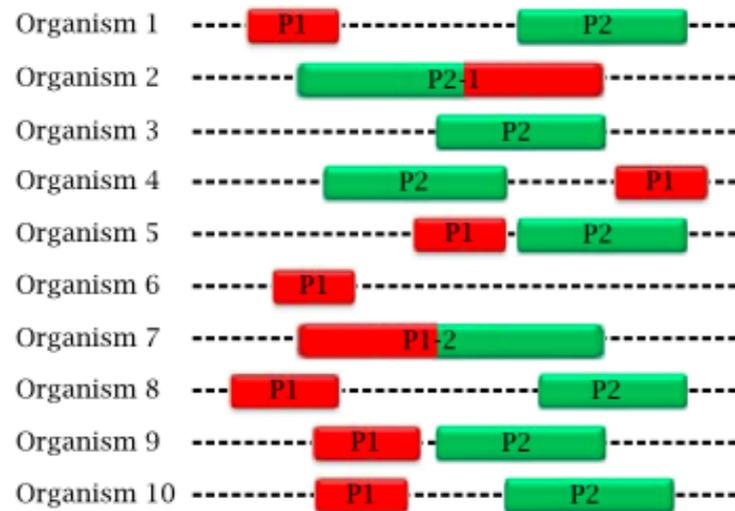
$$p_{ij} = -1/\log E_{ij}$$

$$H(A) = - \sum p(a) \ln p(a)$$

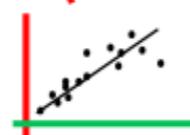
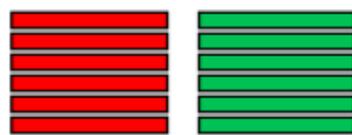
$$MI(A,B) = H(A) + H(B) - H(A,B)$$

## Esquema

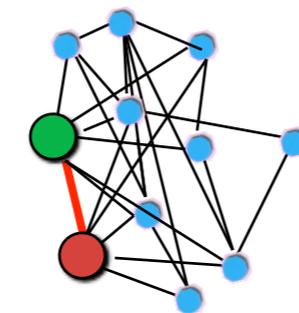
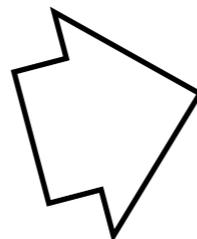
- Introducción
- Datos Experimentales
- Datos Computacionales**
- Calidad de los datos
- Estudio de las redes
- Más allá de las redes



(b) Similarity of Trees

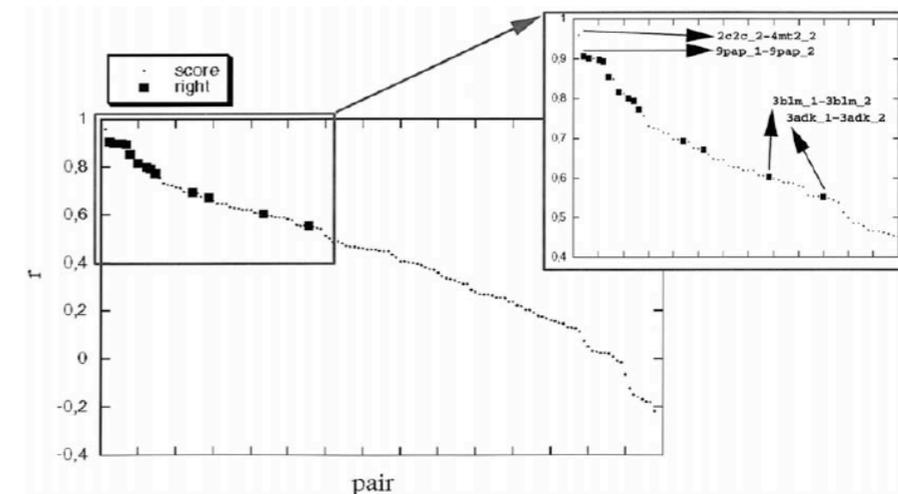


$$r_{xy} = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}}$$



# MirrorTree

- Las proteínas que interactúan tienden a compartir un conjunto de **restricciones evolutivas comunes**.
- Este método intenta detectar la **coevolución** al nivel de secuencias comparando una simplificación de los **árboles evolutivos** de pares de alineamientos.

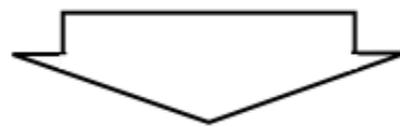
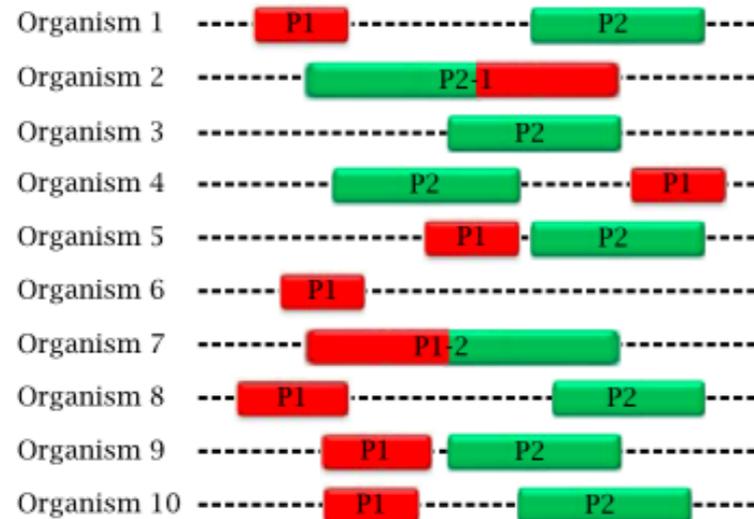


Pazos & Valencia. *Prot. Eng.* 2001.

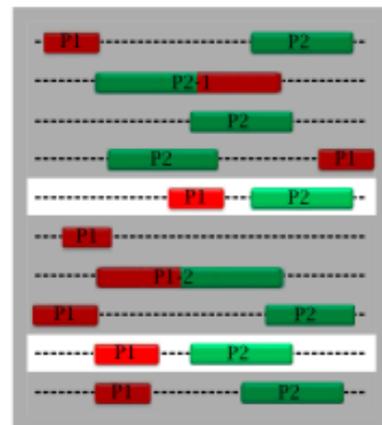
## Esquema

- Introducción
- Datos Experimentales
- Datos Computacionales**
- Calidad de los datos
- Estudio de las redes
- Más allá de las redes

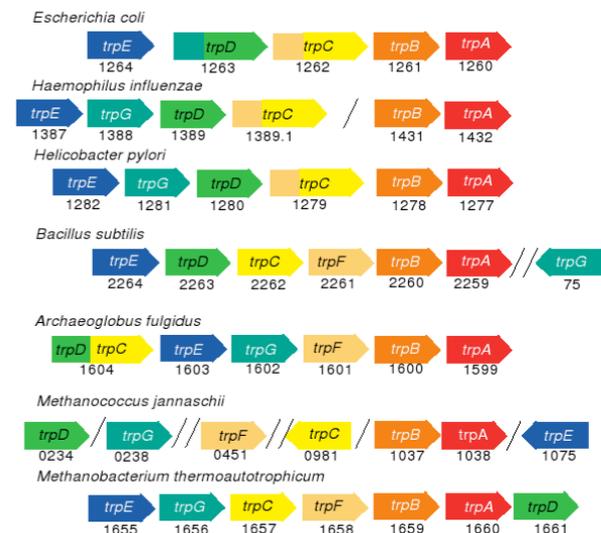
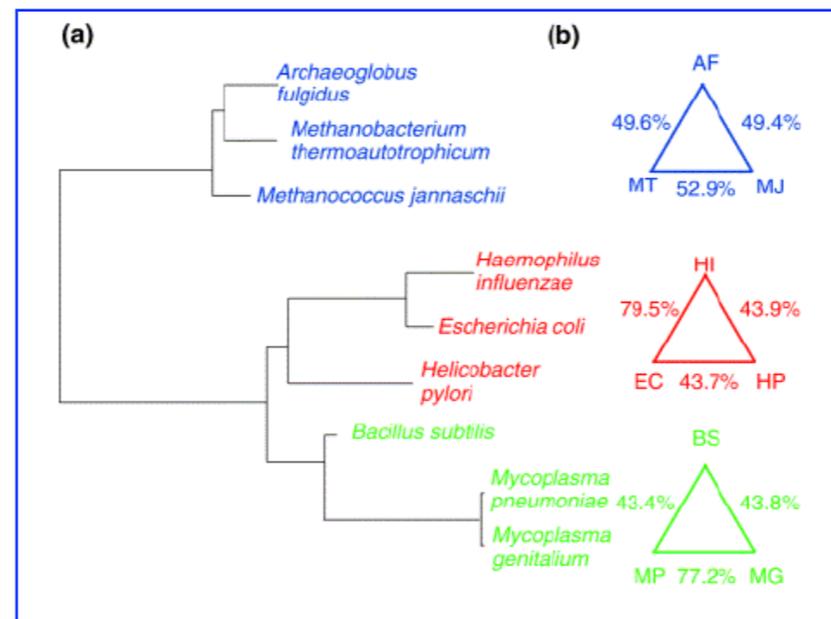
# Gene neighbourhood



(c) Gene Neighbourhood



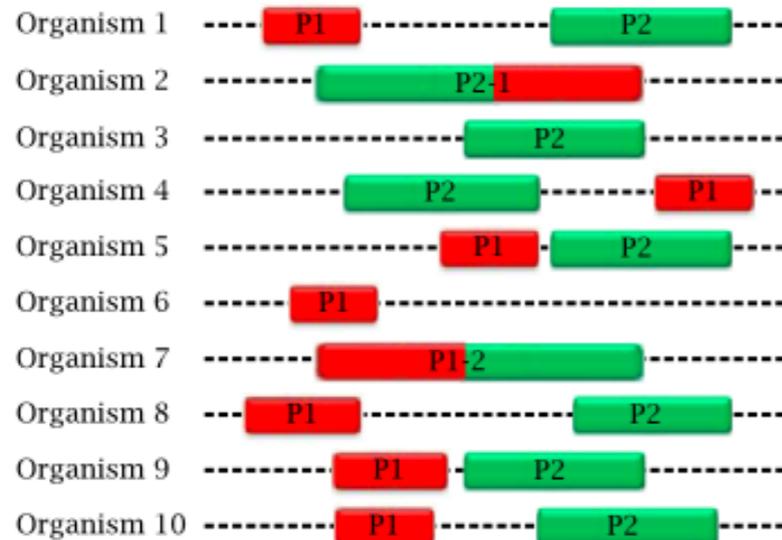
- Dos genes se consideran vecinos cuando están **próximos en el genoma** (menos de 600bp)
- Se sabe que en procariontas esta vecindad se usa para optimizar la **coordinación de su expresión**.
- Además genes próximos pueden ser eliminados y **transferidos juntos**.
- La **conservación** de esta proximidad a lo largo de diferentes organismos es una señal de evolución coordinada.



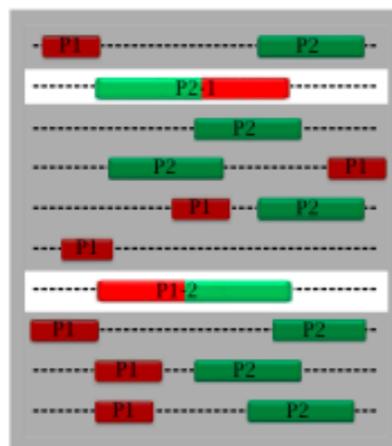
Dandekar et al. TIBS. 1998.  
Overbeek et al. PNAS. 1999.

## Esquema

- Introducción
- Datos Experimentales
- Datos Computacionales**
- Calidad de los datos
- Estudio de las redes
- Más allá de las redes



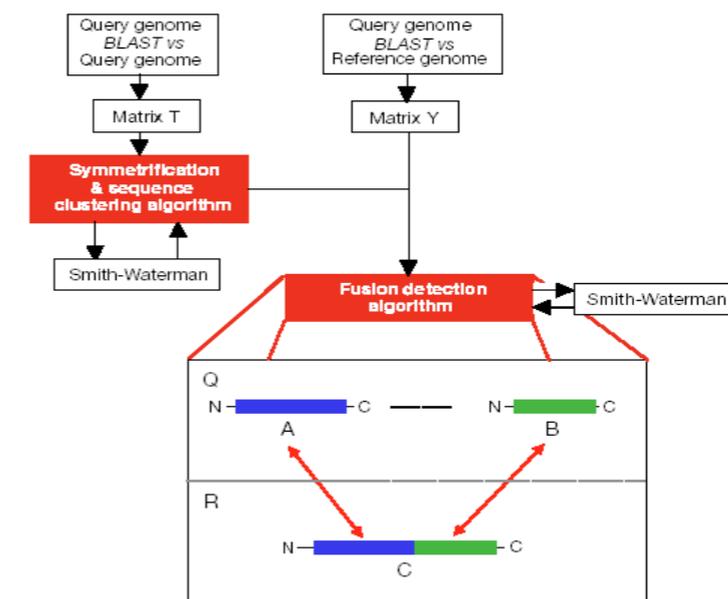
(d) Gene Fusion



# Gene fusion

- La evolución genera **secuencias híbridas** por fusión de otras más simples.
- Esto permite una mayor **coordinación de las funciones** desempeñadas por ambas proteínas.
- La **combinación y especialización de dominios** supone un incremento de la complejidad de los organismos (eucariotas).
- La presencia de estas fusiones sugiere una **interacción** entre las secuencias homólogas no fusionadas.

Marcotte *et al.* Science 1999  
Marcotte *et al.* Nature 1999  
Enright *et al.* Nature 1999.



## Esquema

- Introducción
- Datos Experimentales
- Datos Computacionales**
- Calidad de los datos
- Estudio de las redes
- Más allá de las redes

# Problemas comunes de estos métodos de predicción

## Falsos negativos:

- Se requiere una señal clara a lo largo de **varios organismos**.
- Si la detección de **proteínas equívologas** falla, no se encuentra la señal.
- Una interacción dada **no necesariamente** debe mostrar ninguno de los indicios usados.

## Falsos positivos:

- Las **relaciones filogenéticas** entre los organismos, suponen sesgos que pueden producir señales erróneas.
- La evolución coordinada tiene problemas para distinguir entre interacción física y **asociación funcional**.
- La evolución coordinada sufre de cierta **transitividad** (si a-b y b-c entonces a-c).

## Esquema

- Introducción
- Datos Experimentales
- Datos Computacionales
- Calidad de los datos
- Estudio de las redes
- Más allá de las redes

# Otras redes relacionadas con interacción (minería de textos)



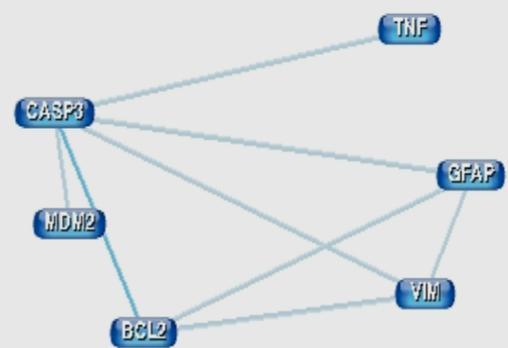
Information hyperlinked Over proteins

Search Gene

Clear model

Print version

Edges in the graph correspond to genes that concur in at least 1 2 3 phrases.



MDM2 is **cleaved** by Caspase 3 (CPP32) during apoptosis after aspartic acid-361, generating a 60 kd fragment.

These findings indicate that IR-induced apoptosis involves activation of CPP32 and that this CrmA-insensitive apoptotic pathway is distinct from those **induced** by TNF and certain other stimuli.

At the end of the experiment, the whole CNS of each animal was collected for histopathology and immunohistochemistry for apoptotic markers (BAX, BCL2 and CPP32) and for glial fibrillary acid protein (GFAP; vimentin).

Fas-induced activation of the ce



Information hyperlinked Over proteins

Search Gene

Show associations of SNF1 with genes from...

Human

Mouse

Drosophila

Zebrafish

C. elegans

Arabidopsis

S. Cerevisiae

E. Coli

Filter and options

Gene Model

Print version

Help

Symbol	Name	Synonyms	Organism
SNF1		CAT1, CCR1, GLC2, HAF3, PAS14	Saccharomyces cerevisiae
NCBI Protein	NP_010765		

The **snf1** mutation also **suppresses** the glucose repression defects of **reg1**.

The **SIP1** protein **co-immunoprecipitated** with **SNF1** and was phosphorylated in vitro.

Here we show that **Reg1** **interacts** with the **Snf1** catalytic domain in the two-hybrid system.

Previous studies showed that **Reg1** **regulates** the **Snf1** protein kinase in response to glucose.

The **SNF4** protein is physically **associated** with **SNF1** and positively affects the kinase activity.

The **Sip1** protein is known to undergo phosphorylation when **associated** in vitro with the **Snf1** protein kinase.

Genetic evidence indicated that the catalytic activity of **Snf1** negatively **regulates** its interaction with **Reg1**.

The **SNF1** protein kinase and the **associated** **SNF4** protein are required for release of glucose repression in *Saccharomyces cerevisiae*.

The **SIP1** gene of *Saccharomyces cerevisiae* is a carbon-catabolite-specific negative regulator of GAL gene transcription and acts as a multicopy suppressor of growth defects **associated** with impaired **Snf1p** protein kinase activity.

We show that different sequences of **Reg1** **interact** with **Glc7** and **Snf1**.

In two-hybrid assays, one **SNF4** mutation **enhances** the interaction between **Snf4** and **Snf1**.

Previously, we identified **SIP1** and **SIP2** as proteins that **interact** with **SNF1** in vivo by the two-hybrid system.

Previous experimental evidence had indicated that **Reg1** might **target** **Glc7** to nuclear substrates such as the **Snf1** kinase complex.

The catalytic subunits of Arabidopsis SnRKs, AKIN10 and AKIN11, interact with **Snf4** and **suppress** the **snf1** and **snf4** mutations in yeast.

**Pak1** **associates** with the **Snf1** kinase in vivo, and the association is greatly enhanced under glucose-limiting conditions when **Snf1** is active.

We show that **SNF4** **binds** to the **SNF1** regulatory domain in low glucose, whereas in high glucose the regulatory domain binds to the kinase domain of **SNF1** itself.

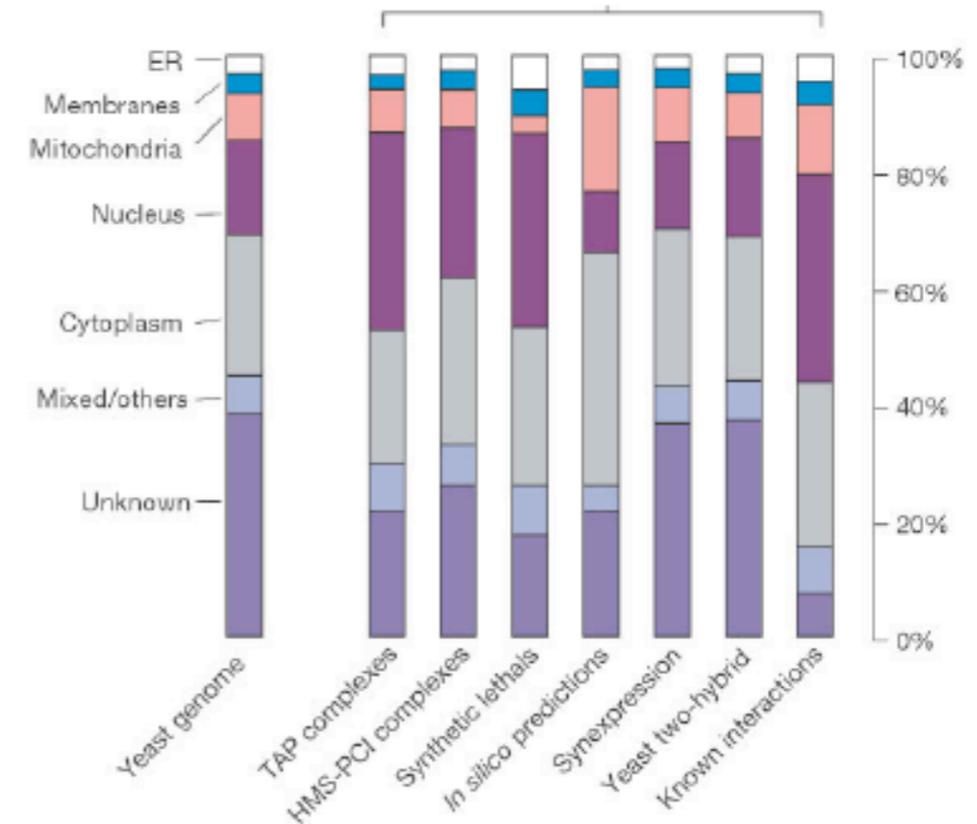
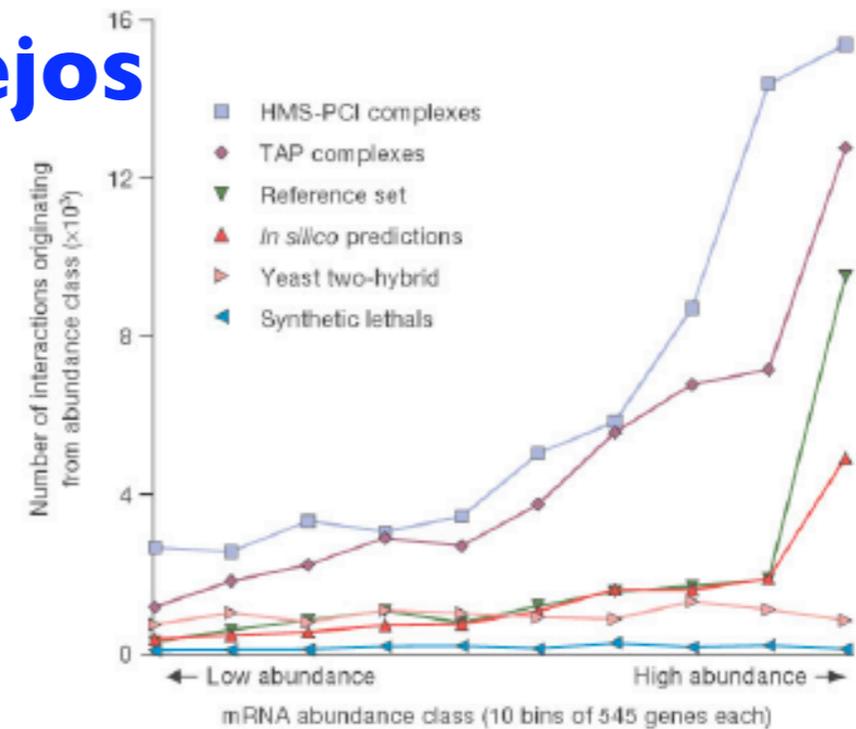
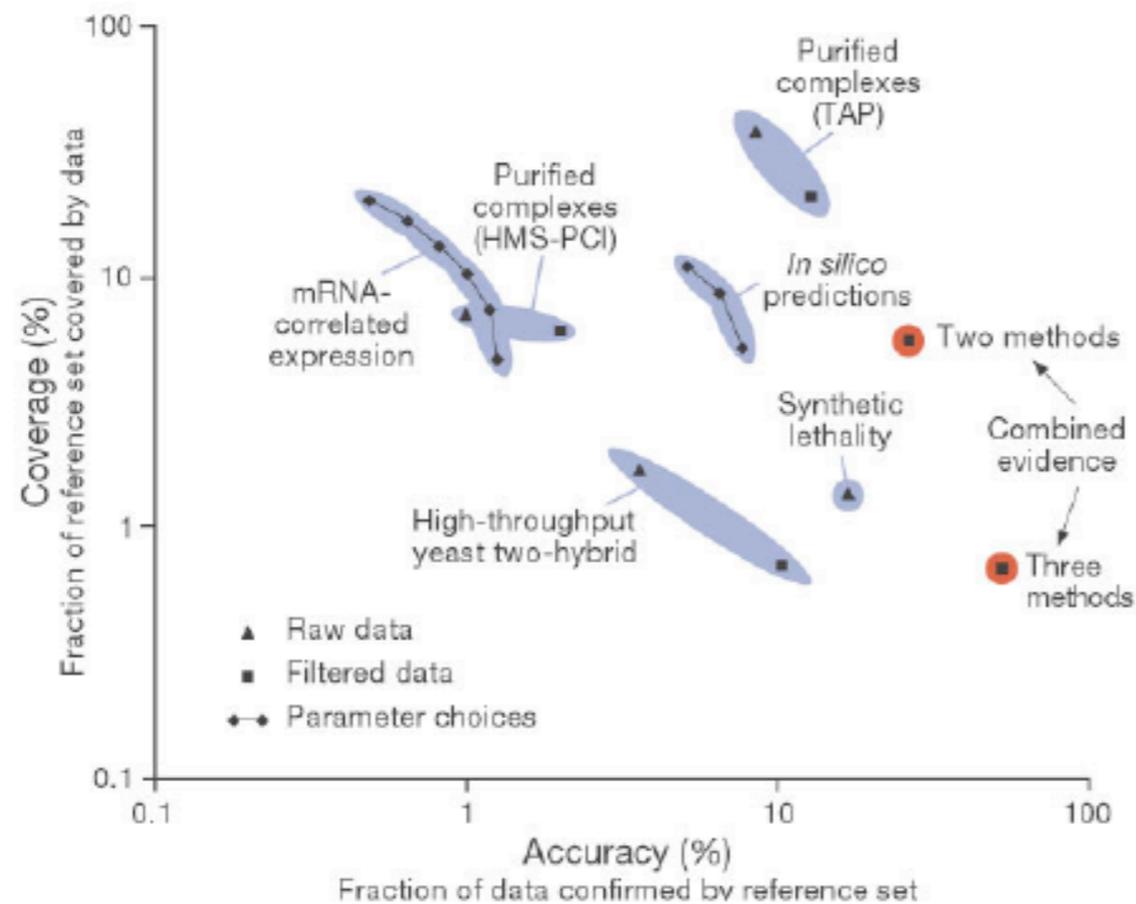
Blaschke & Valencia. Genome Inform Ser Workshop Genome Inform. 2001  
 Hoffmann & Valencia. Nat. Genetics. 2004

# **Calidad de los datos**

## Esquema

- Introducción
- Datos Experimentales
- Datos Computacionales
- Calidad de los datos**
- Estudio de las redes
- Más allá de las redes

# Evaluación respecto a complejos

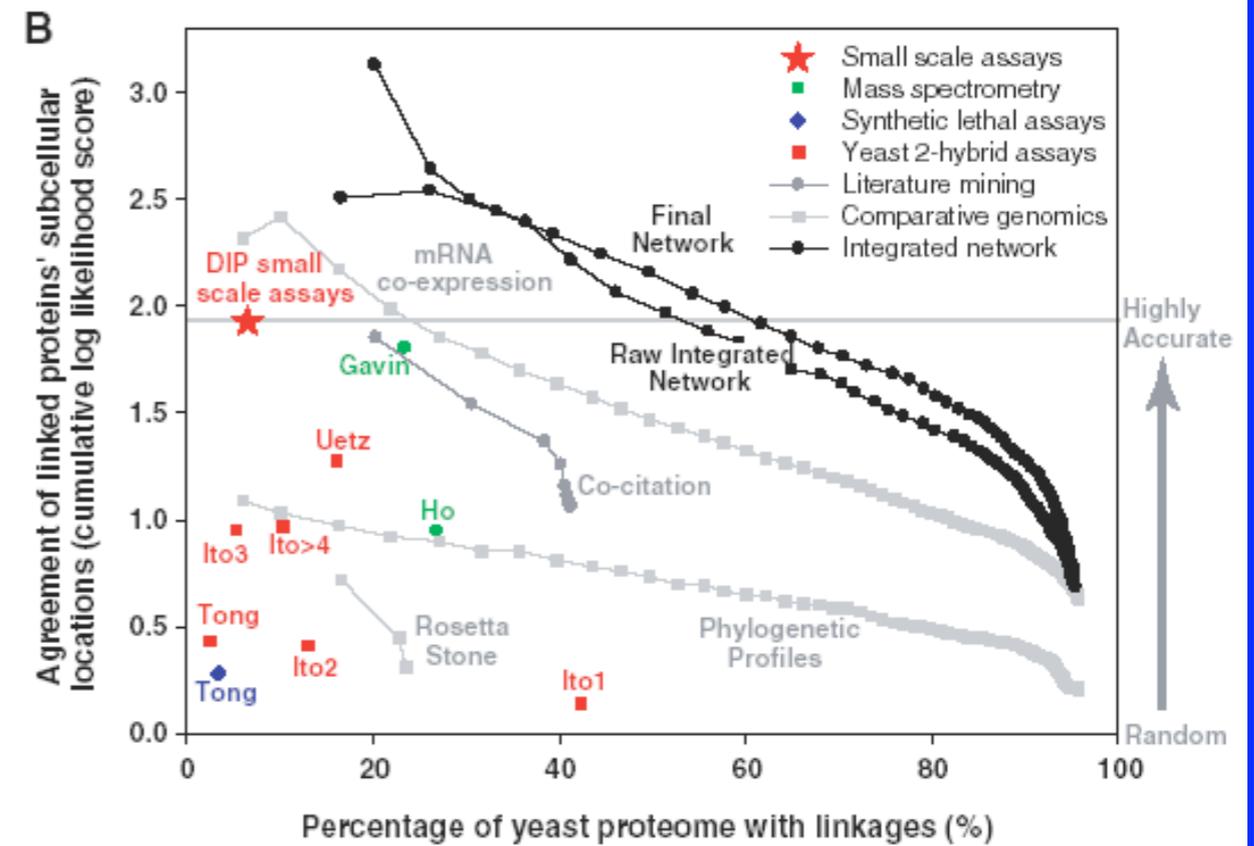
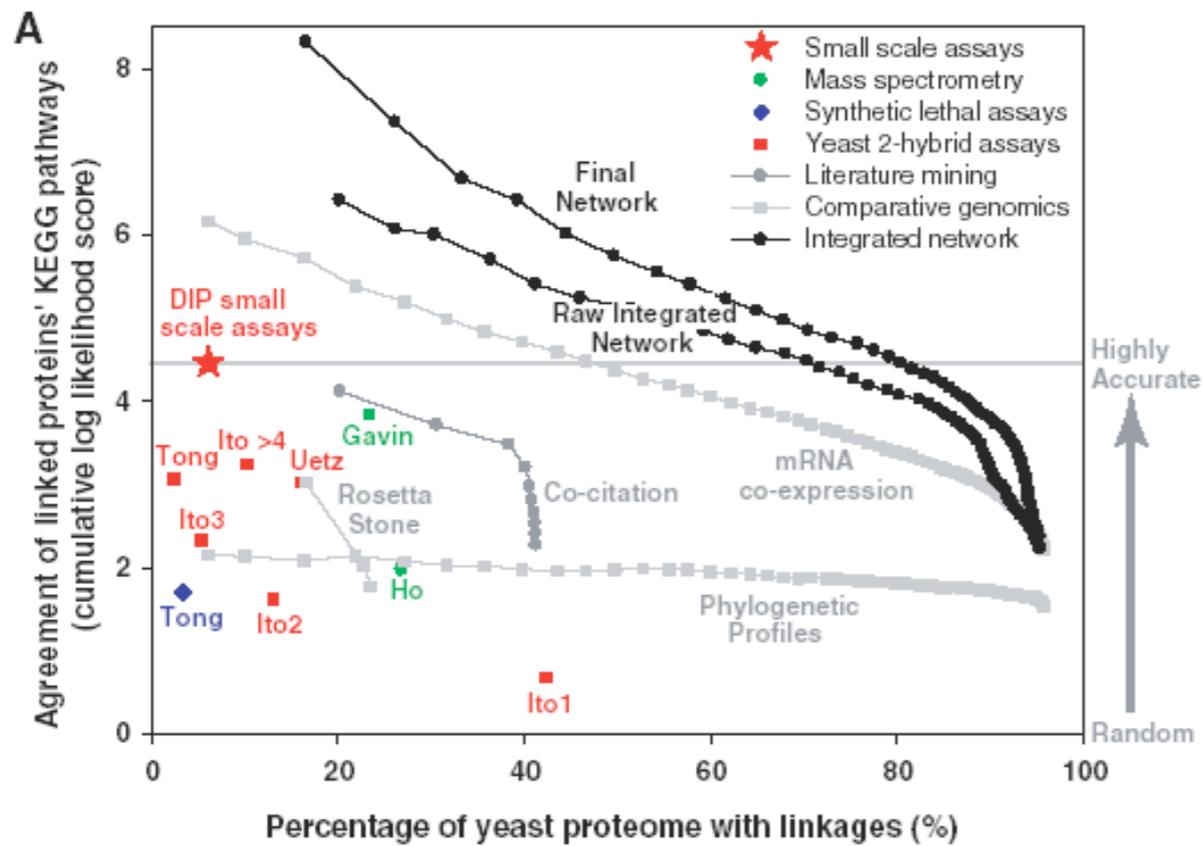


von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S. and Bork, P. (2002) Comparative assessment of large scale data sets of protein-protein interactions. *Nature*, **417**, 399-403.

## Esquema

- Introducción
- Datos Experimentales
- Datos Computacionales
- Calidad de los datos**
- Estudio de las redes
- Más allá de las redes

# Evaluación respecto a rutas metabólicas y localización subcelular



Lee *et al.* Science. 2004

# **Estudio de las redes**

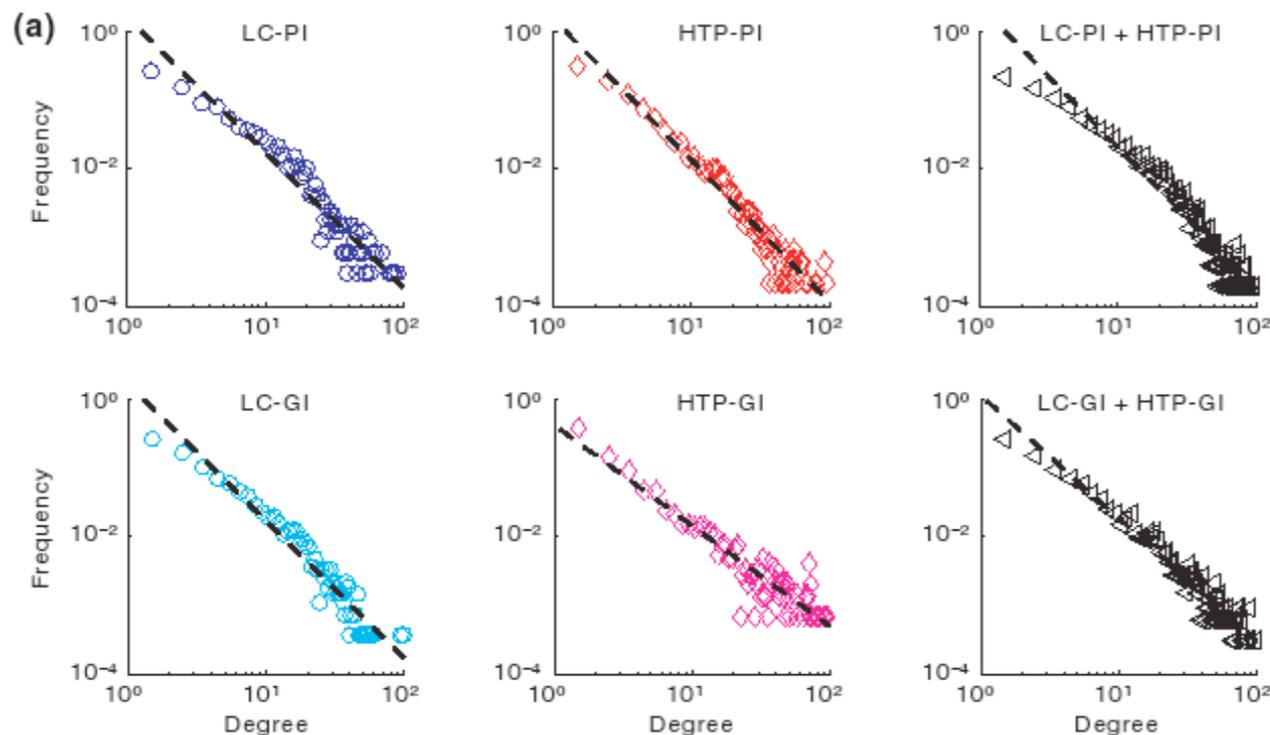
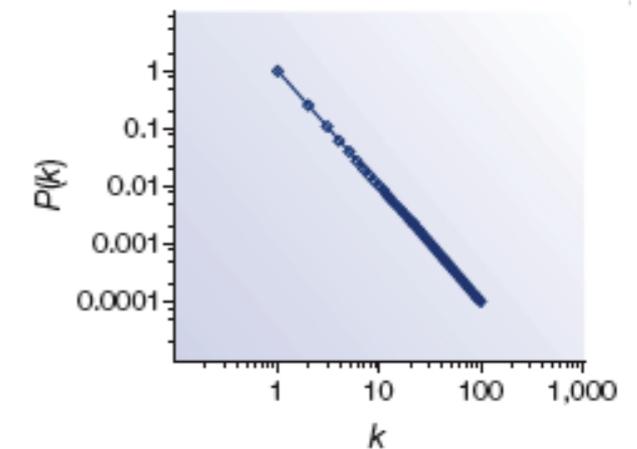
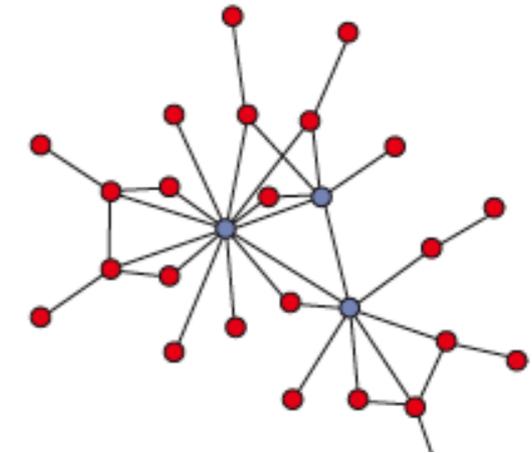
## Esquema

- Introducción
- Datos Experimentales
- Datos Computacionales
- Calidad de los datos
- Estudio de las redes**
- Más allá de las redes

# Estructura de la red

- Camino mínimo:** el mínimo número de enlaces que hay que recorrer para ir de un nodo a otro.
- Camino mínimo medio es mucho más corto que en redes con topología generada al azar (**Small world**).
- Conectividad:** número de enlaces de un nodo.
- Distribución **Power-Law** de conectividades  $\rightarrow p(k) \sim k^{-\gamma}$   
**Scale-free**
- Assortativity:** tendencia de un elemento a asociarse con otro similar a él (en conectividad).
- Los nodos más conectados suelen estar unidos a otros con pocas interacciones (**disassortativity**  $\rightarrow$  ¿ruido?).

## Red Scale-free (Barabasi-Albert)



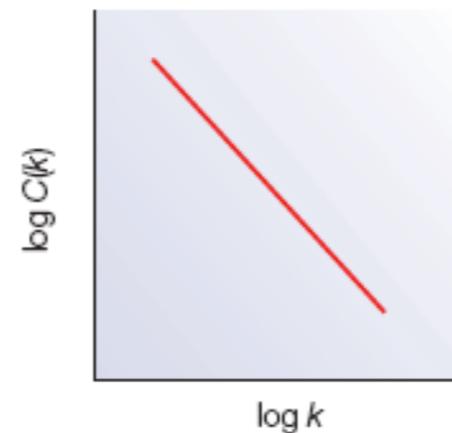
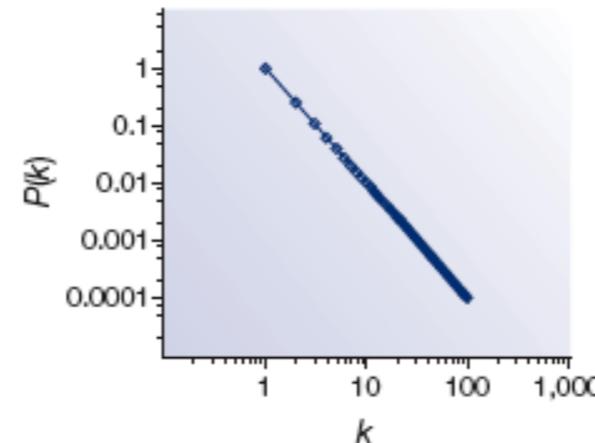
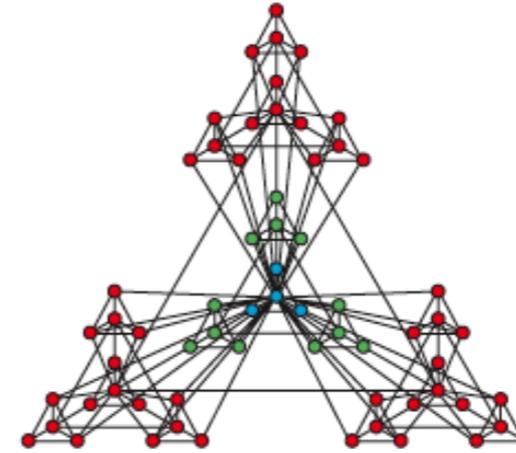
## Esquema

- Introducción
- Datos Experimentales
- Datos Computacionales
- Calidad de los datos
- Estudio de las redes**
- Más allá de las redes

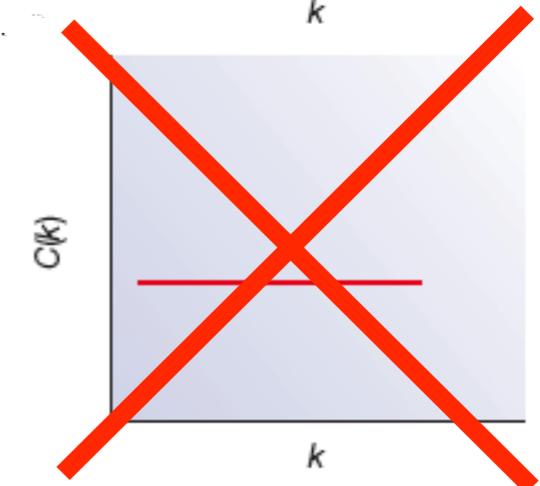
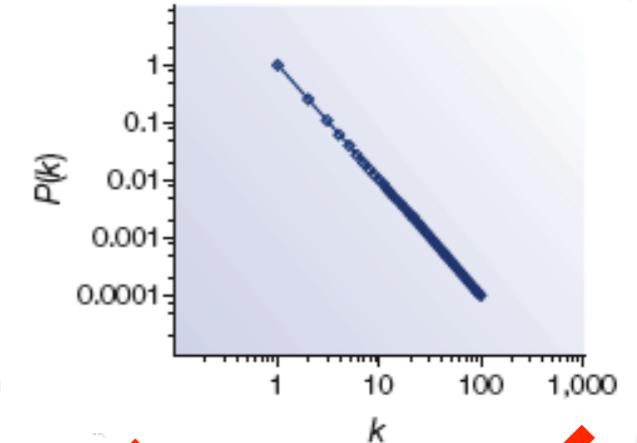
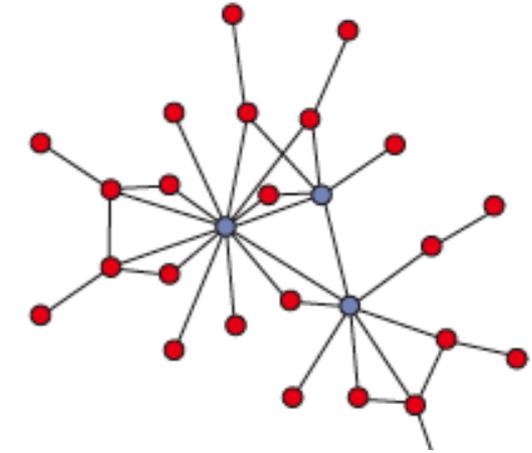
# Estructura de la red

- Coefficiente de clustering:** mide la densidad de conexiones alrededor de un nodo:  $2n_1/(k*(k-1))$
- Los nodos más conectados (**hubs**), parecen tener **menores coeficientes de clustering** (¿redes jerárquicas?).
- Presenta **módulos difíciles** de detectar (¿jerárquica?, ¿ruido?. ¿?).
- Modelo poco verosímil -> **Modelos mixtos**

**Red Scale-free Jerárquica**



**Red Scale-free (Barabasi-Albert)**

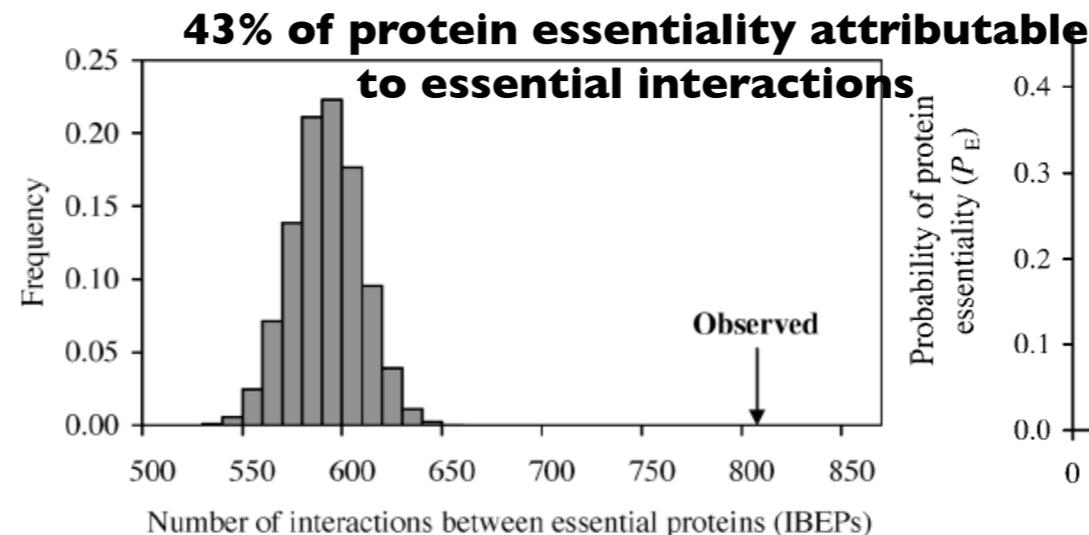
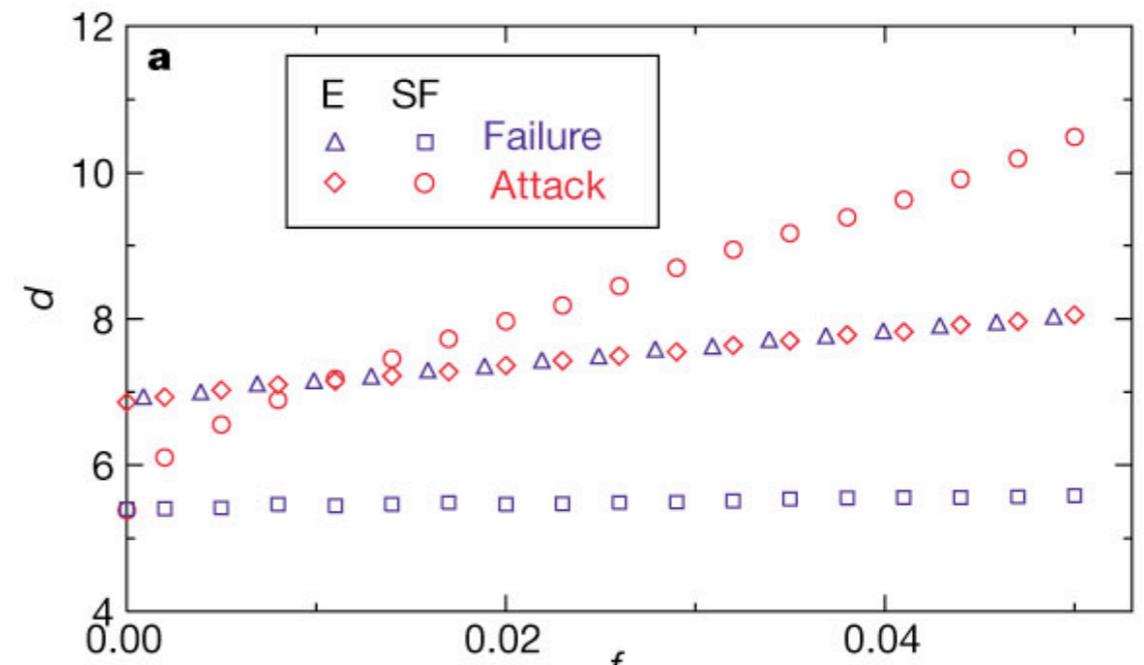
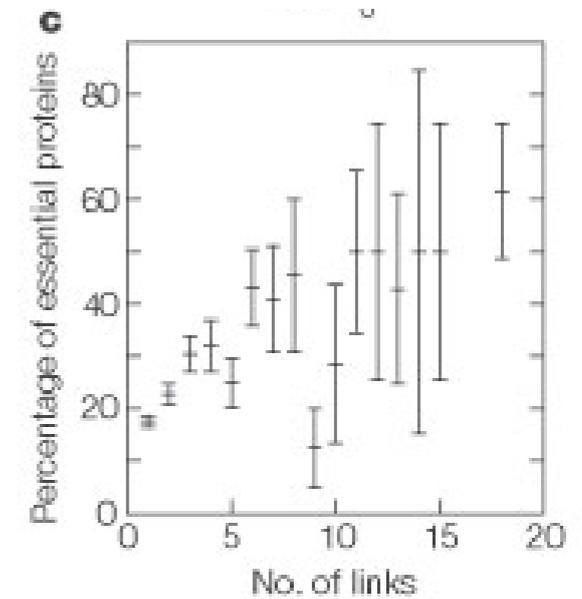
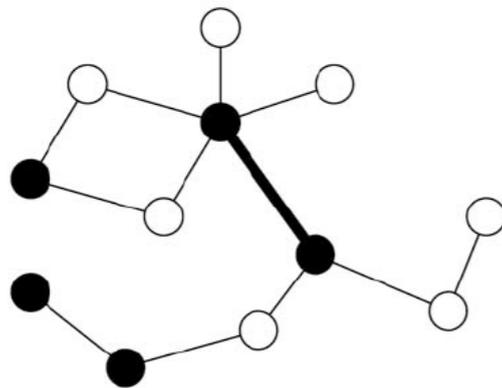


## Esquema

- Introducción
- Datos Experimentales
- Datos Computacionales
- Calidad de los datos
- Estudio de las redes**
- Más allá de las redes

# Estructura de la red

- Robustez:** capacidad para mantener su estructura global ante deleciones de nodos o enlaces.
- Las redes scale-free tienen:
  - Mayor robustez ante **fallos aleatorios** (eliminación de nodos al azar)
  - Menor robustez ante **ataques dirigidos** a los hubs.
- En las redes de interacciones los hubs tienden a ser **esenciales**.
- Explicación **alternativa:** los hubs tienen más probabilidades de estar implicados en una **interacción esencial**.



**43% of protein essentiality attributable to essential interactions**

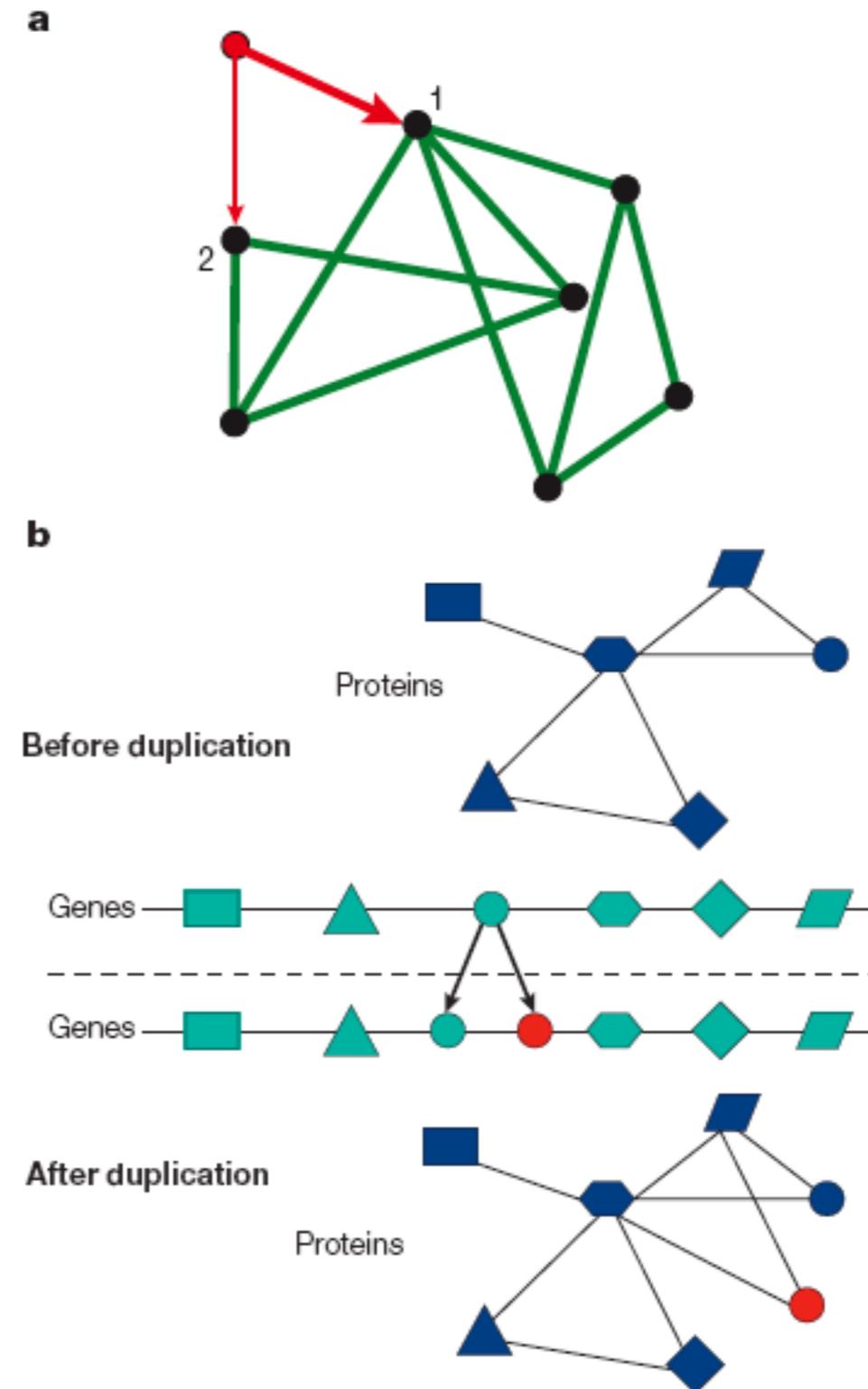
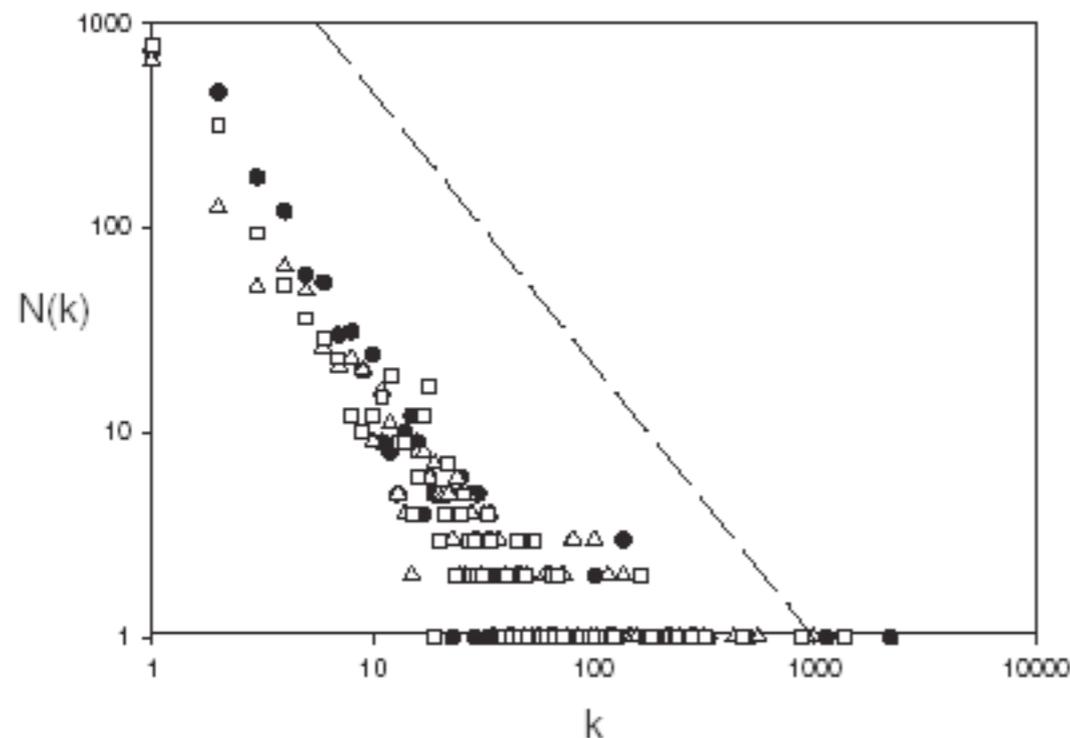
# **Más allá de las redes**

## Esquema

- ✓ Introducción
- ✓ Datos Experimentales
- ✓ Datos Computacionales
- ✓ Calidad de los datos
- ✓ Estudio de las redes
- ✓ **Más allá de las redes**

# Evolución de las redes

- ❑ Estructura de la red: **scale-free**.
- ❑ Coherente con un crecimiento por **unión preferencial**.
- ❑ Se han desarrollado simulaciones incluyendo crecimiento de la red por **duplicación génica**.



## Esquema

- ✓ Introducción
- ✓ Datos Experimentales
- ✓ Datos Computacionales
- ✓ Calidad de los datos
- ✓ Estudio de las redes
- ✓ **Más allá de las redes**

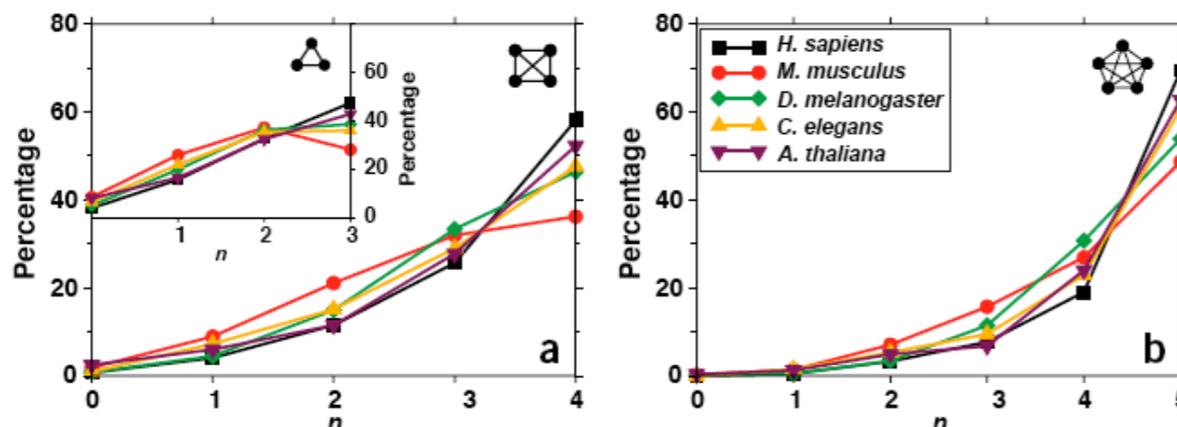
# Motivos, Función y Conservación

- Se pueden describir **motivos** de un número pequeño de nodos y unas conexiones determinadas entre ellos.
- Algunos de estos motivos están **sobre-representados** en las redes de interacción (y2h).
- Los motivos sobre-representados dependen del tipo de **proceso celular**.
- Además, los motivos **más conectados están más conservados** entre organismos (relacionado con la robustez de la red)

Functional class	Overrepresented motifs
Transport facilitation	(10)
Subcellular localization	(21)  (21)  (26)  (15)  (27)  (23) (29)  (20)  (63)  (45)
Regulation	(10)
Protein fate	(14)  (16)  (13)  (33)  (27)  (20) (26)  (24)  (16)  (60)  (41)
Cell cycle	(11)  (14)  (13)  (11)  (14)
Cellular transport	(11)  (12)
Transcription	(12)  (16)  (17)  (13)  (16)  (19) (17)  (15)  (14)  (21)  (23)
Protein synthesis	(12)  (11)  (17)  (11)  (24)

Table 1 Evolutionary conservation of motif constituents

#	Motifs	Number of yeast motifs	Natural conservation rate	Random conservation rate	Conservation ratio
1		9,266	13.67%	4.63%	2.94
2		167,304	4.99%	0.81%	6.15
3		3,846	20.51%	1.01%	20.28
4		3,649,591	0.73%	0.12%	5.87
5		1,763,891	2.64%	0.18%	14.67
6		9,646	6.71%	0.17%	40.44
7		164,075	7.67%	0.17%	45.56
8		12,423	18.68%	0.12%	157.89
9		2,339	32.53%	0.08%	422.78
10		25,749	14.77%	0.05%	279.71
11		1,433	47.24%	0.02%	2,256.67



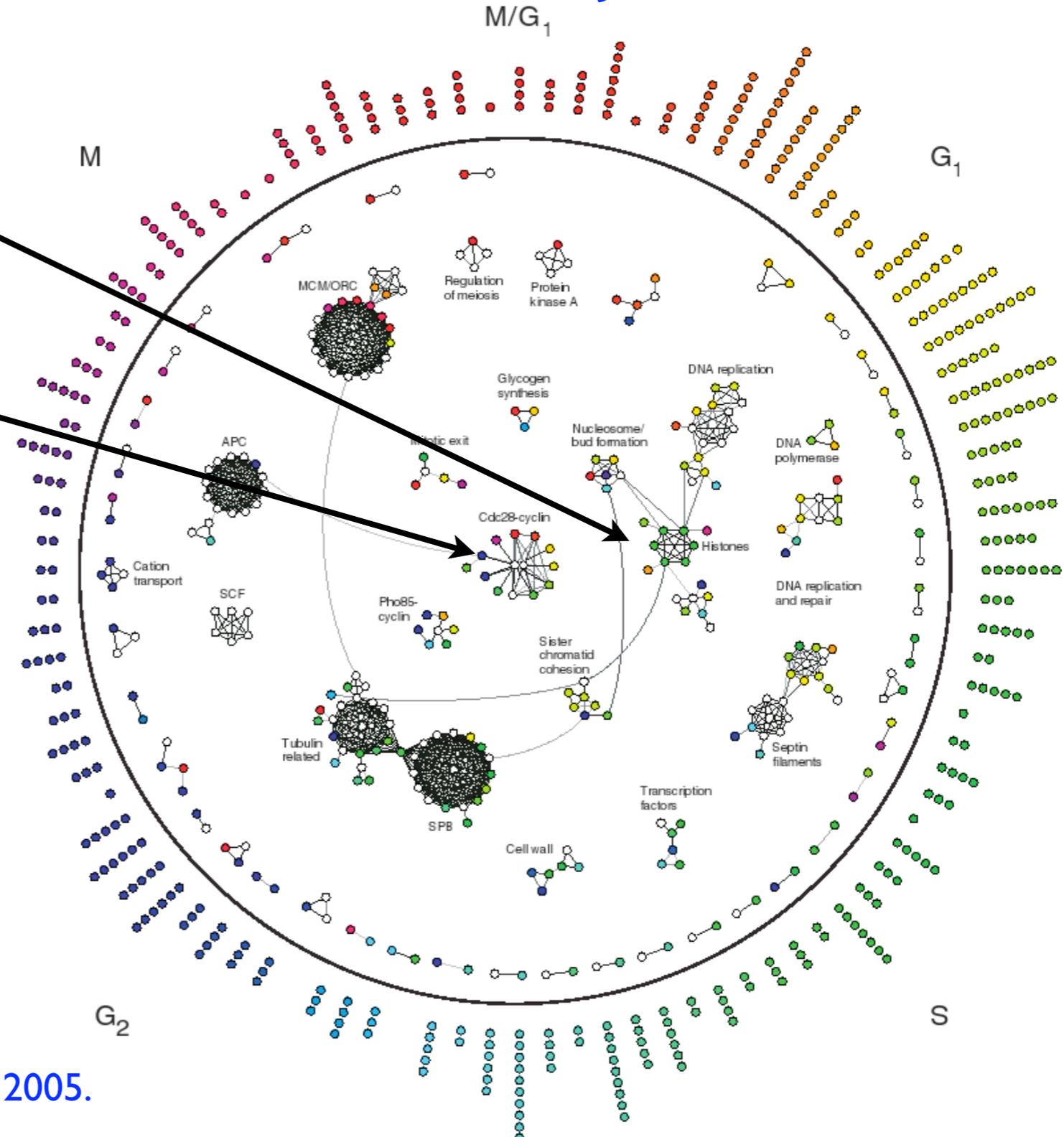
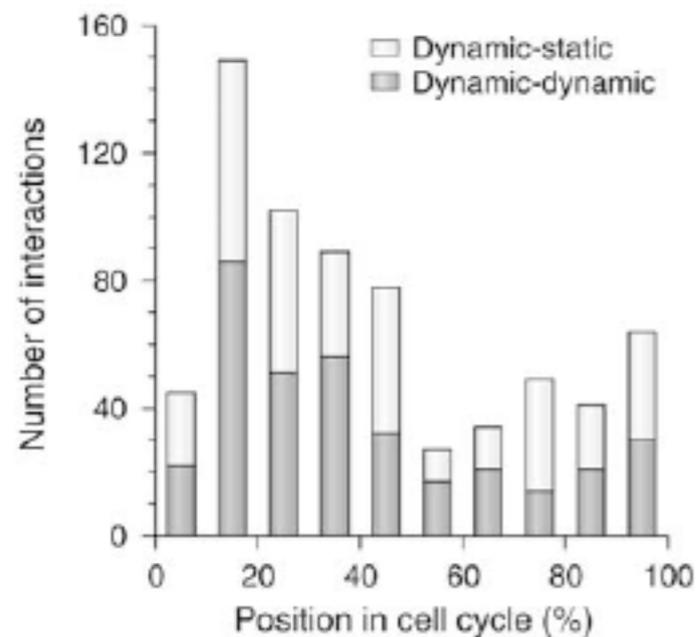
## Esquema

- ✓ Introducción
- ✓ Datos Experimentales
- ✓ Datos Computacionales
- ✓ Calidad de los datos
- ✓ Estudio de las redes
- ✓ **Más allá de las redes**

# Las proteínas ... tienen horarios de trabajo

□ Los “party hubs” se co-expresan con sus interactores.

□ Los “date hubs” no lo hacen.



Ulrik de Lichtenberg, et al. Science. 2005.

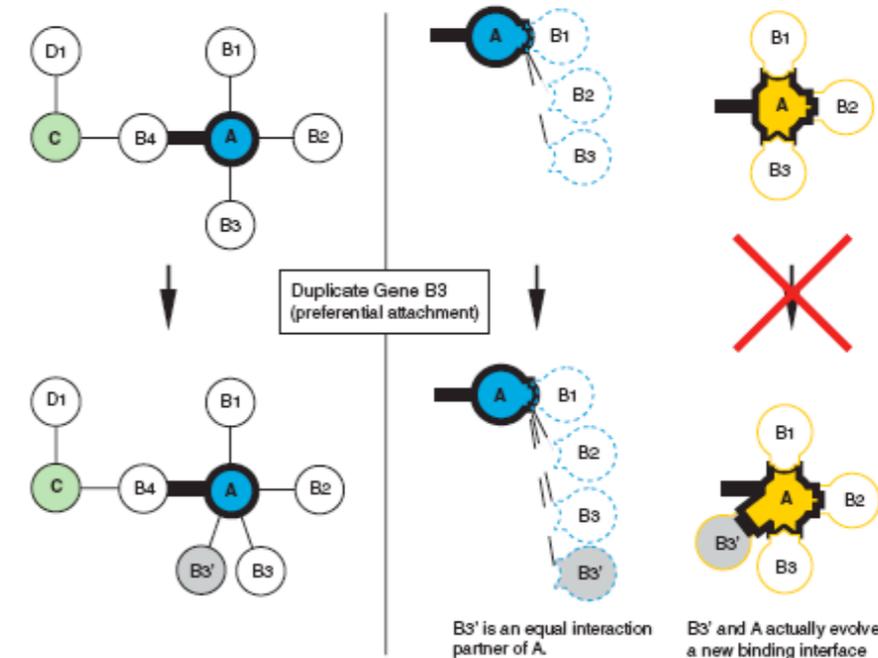
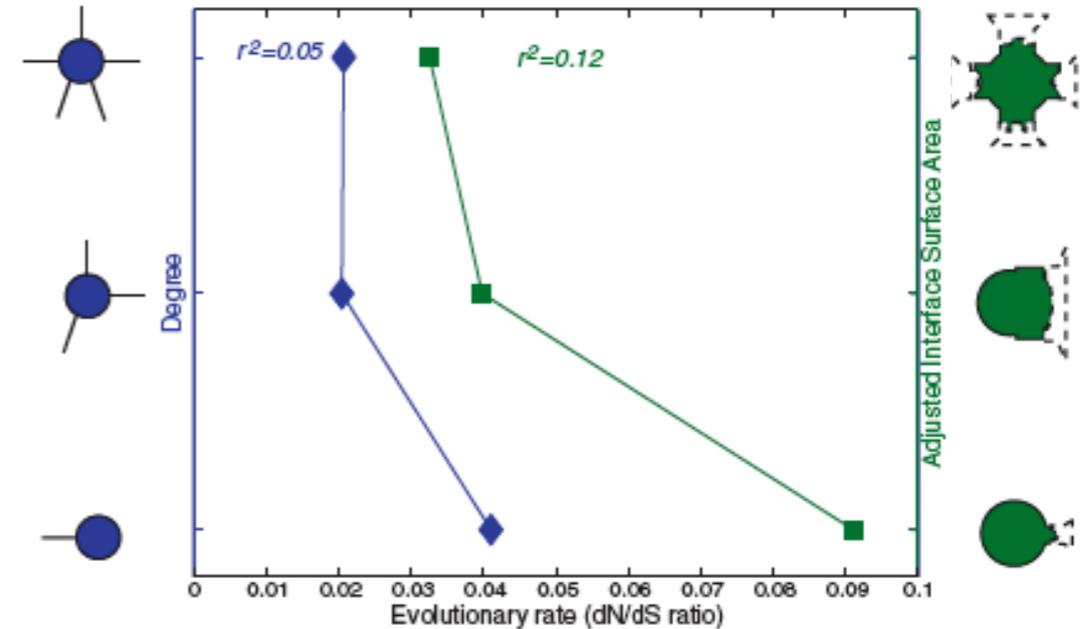
## Esquema

- ✓ Introducción
- ✓ Datos Experimentales
- ✓ Datos Computacionales
- ✓ Calidad de los datos
- ✓ Estudio de las redes
- ✓ **Más allá de las redes**

# Las proteínas ... tienen estructura

- ☐ Las interacciones no son sólo relaciones entre nodos.
- ☐ Las proteínas interactúan a través de **superficies** de distinto tamaño que pueden solapar (mutuamente excluyentes)
- ☐ Teniendo esto en cuenta se ve que:
  - Mayor número de superficies de interacción (no solapantes) supone menor **velocidad de evolución**.
  - Los “**date hubs**” son aquellos con interacciones mutuamente excluyentes.
  - La **duplicación** sólo explica el crecimiento de los “date hubs”.

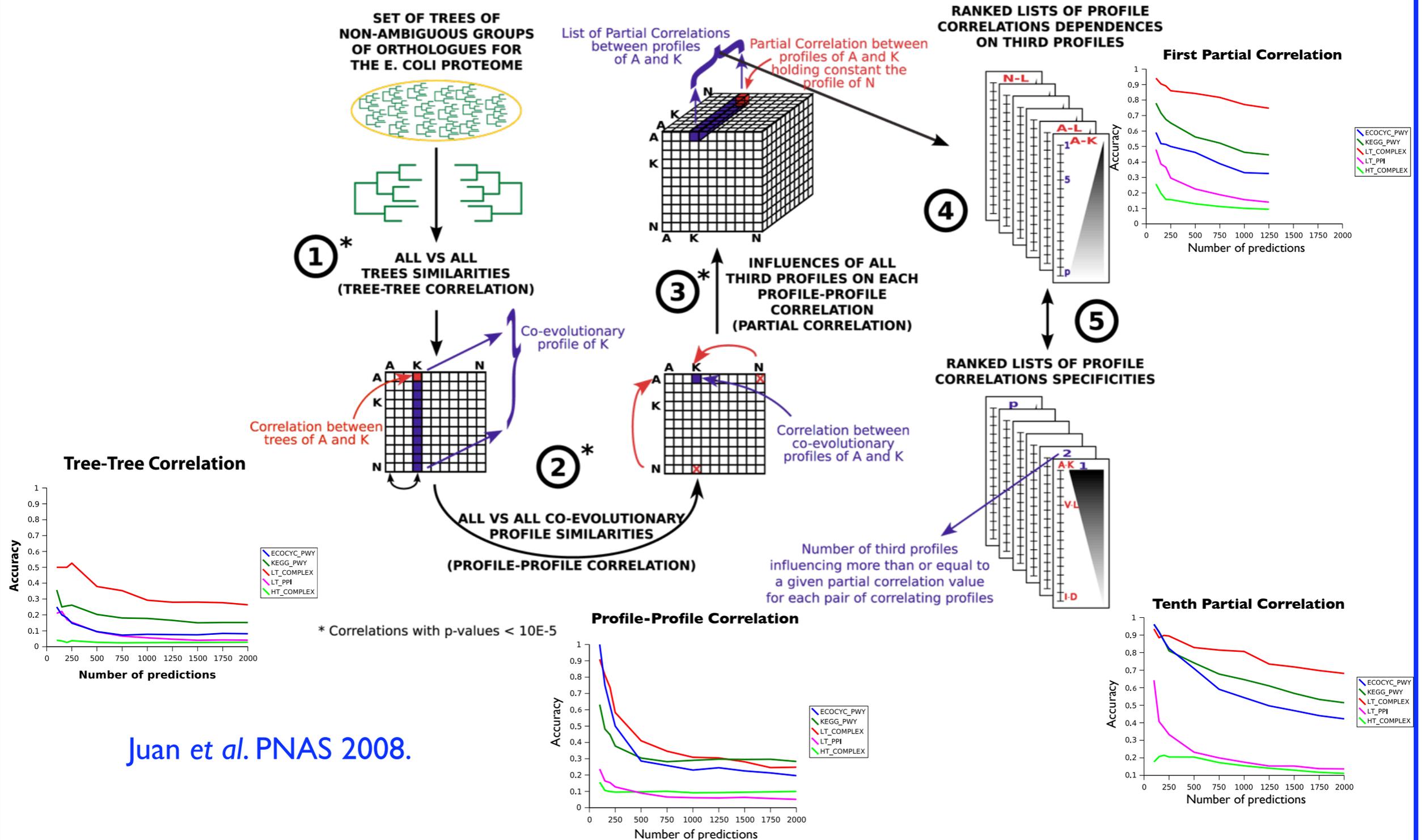
Kim *et al.* Science 2006.



## Esquema

- ✓ Introducción
- ✓ Datos Experimentales
- ✓ Datos Computacionales
- ✓ Calidad de los datos
- ✓ Estudio de las redes
- ✓ **Más allá de las redes**

# Las proteínas ... co-evolucionan



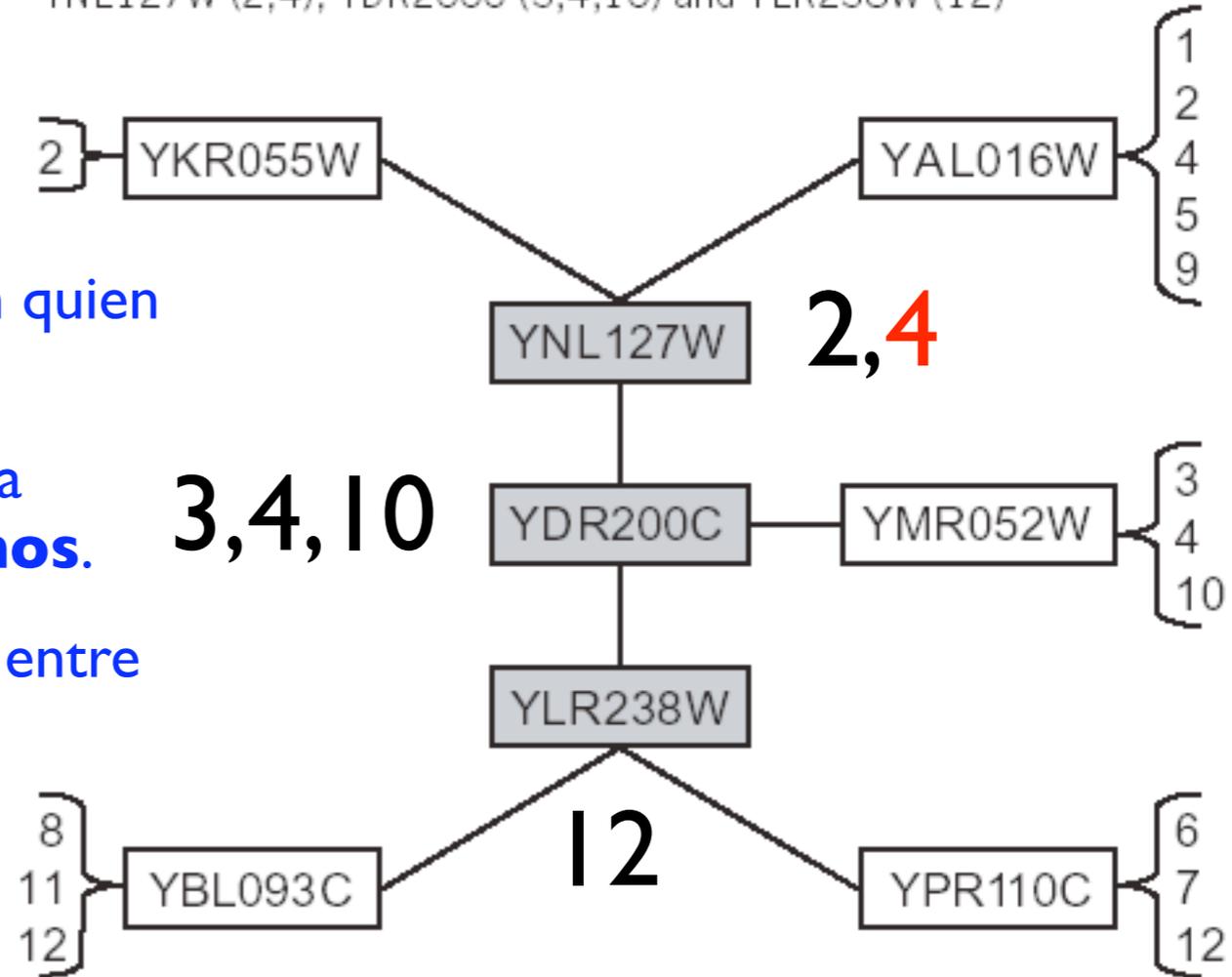
Juan et al. PNAS 2008.

## Esquema

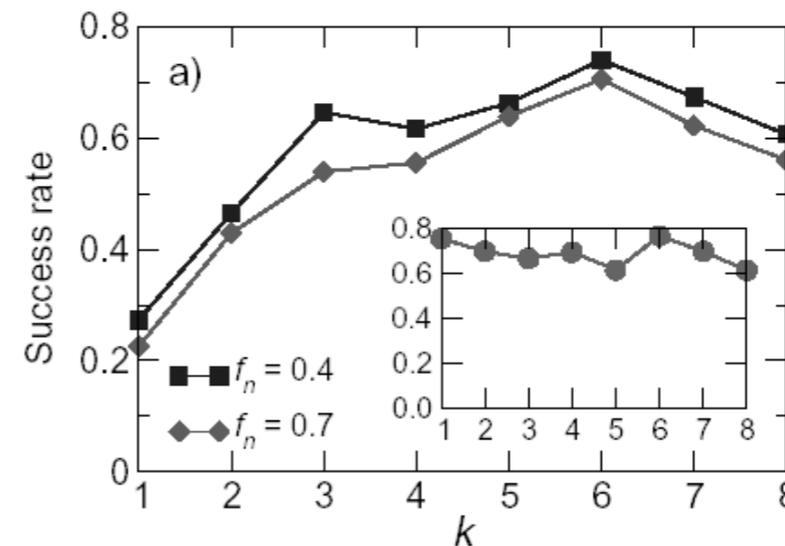
- Introducción
- Datos Experimentales
- Datos Computacionales
- Calidad de los datos
- Estudio de las redes
- Más allá de las redes**

# Prediciendo función con redes de interacción

YNL127W (2,4), YDR200C (3,4,10) and YLR238W (12)



- Contexto de red** o dime con quien andas y te diré quien eres.
- Se asigna función basándose en la **función de los nodos vecinos.**
- Se reduce el número de enlaces entre proteínas con función diferente.



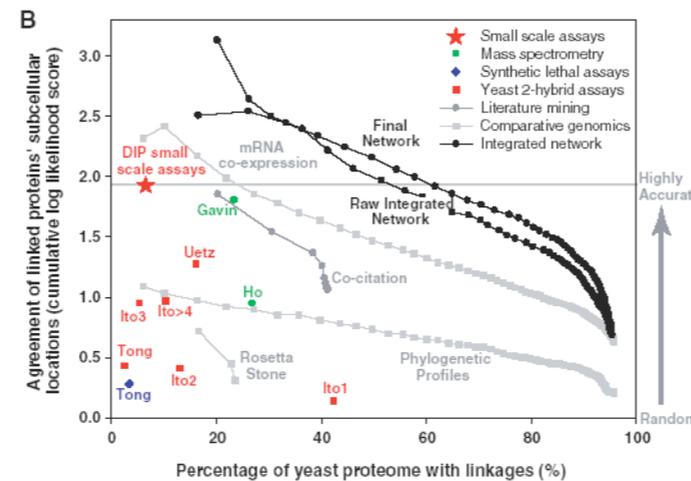
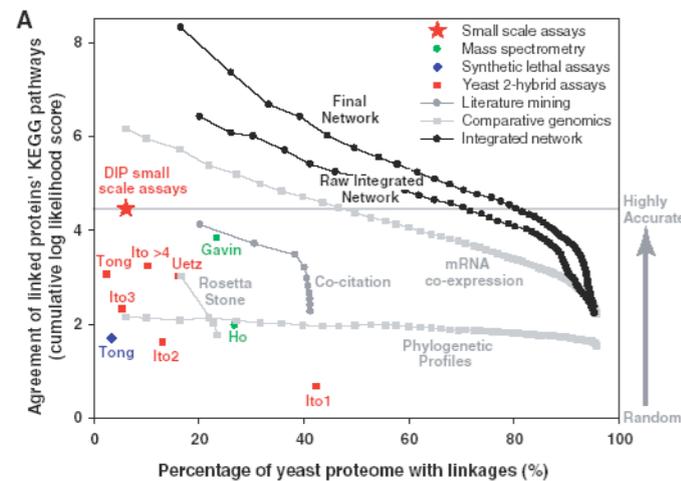
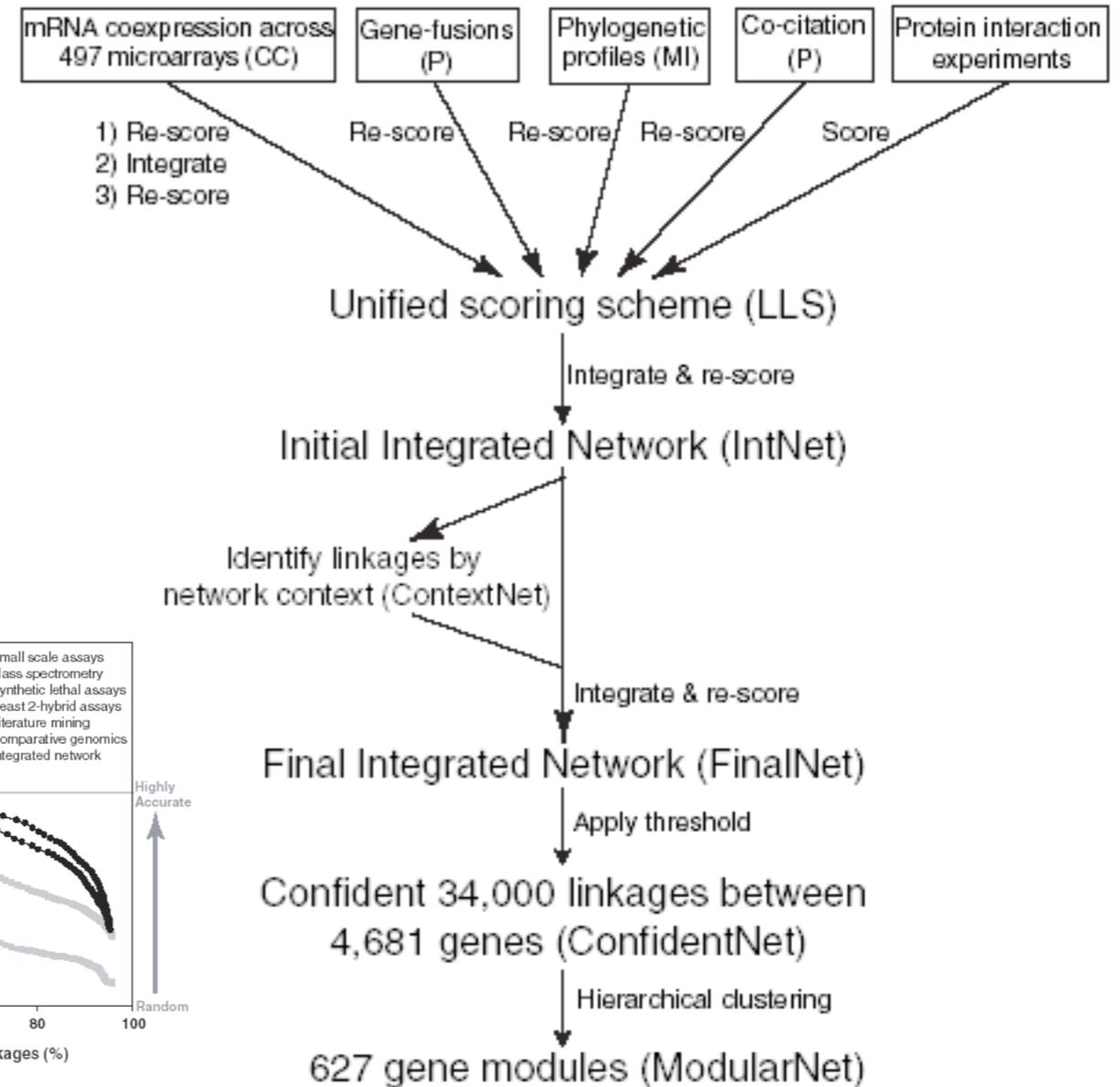
Vazquez et al, Nat Biotech. 2003

## Esquema

- ✓ Introducción
- ✓ Datos Experimentales
- ✓ Datos Computacionales
- ✓ Calidad de los datos
- ✓ Estudio de las redes
- ✓ **Más allá de las redes**

# Predicción de función integrando información

Aproximación bayesiana estableciendo confianzas en función de rutas metabólicas.



Lee *et al.* Science. 2004.

