Departamento Biología Molecular
Universidad Autónoma de Madrid

Redes Biológicas y Biología de Sistemas

21-29 Abril 2008

Coordinación: Florencio Pazos (CNB-CSIC)

Introducción

Florencio Pazos Cabaleiro
Computational Systems Biology Group (CNB-CSIC)
pazos@cnb.csic.es

Departamento Biología Molecular
Universidad Autónoma de Madrid

Redes Biológicas y Biología de Sistemas

21-29 Abril 2008

Coordinación: Florencio Pazos (CNB-CSIC)

- Introducción
    - Obtención e integración de datos masivos (“-omics”)
    - Biología de Sistemas

- Programa del curso

Florencio Pazos Cabaleiro
Computational Systems Biology Group (CNB-CSIC)
pazos@cnb.csic.es

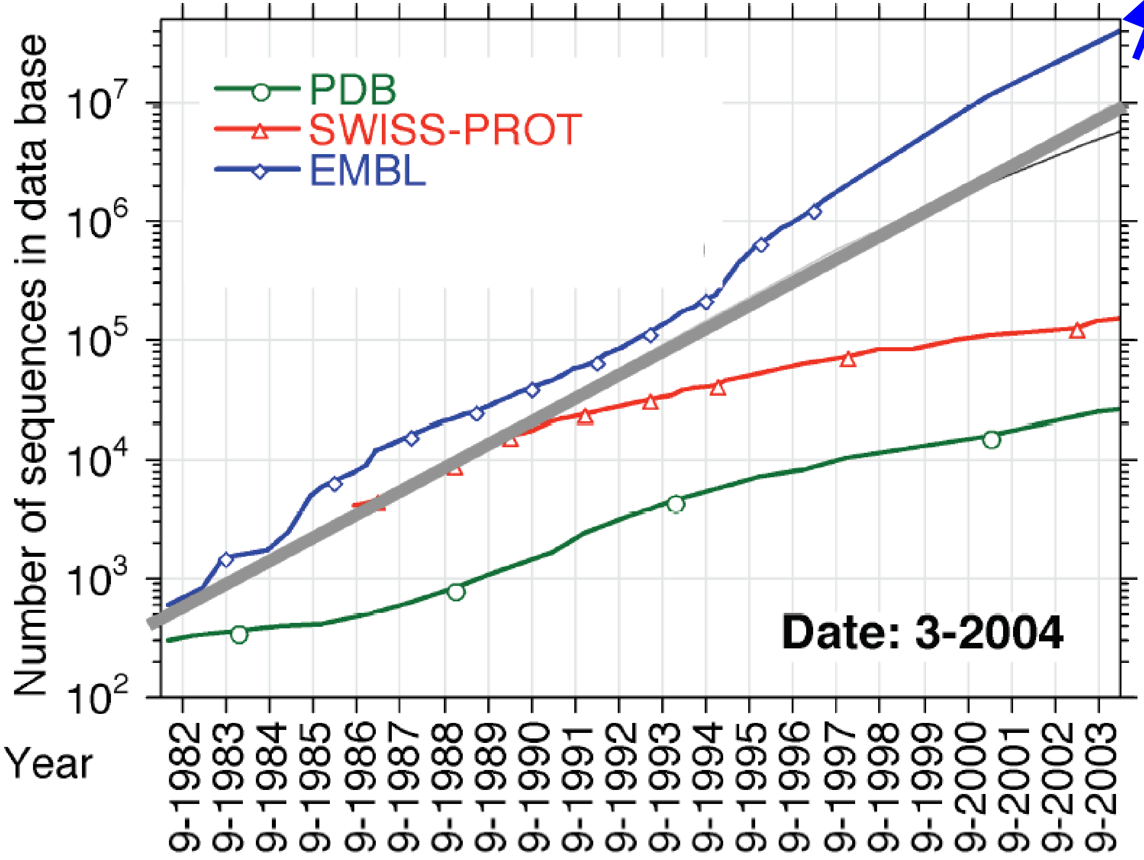
El paradigma “-omics” de la Biología
Necesidad del post-procesamiento computacional de los datos

Era genómica → Era “post-genómica”

(producción masiva de datos biológicos) → (análisis e interpretación)

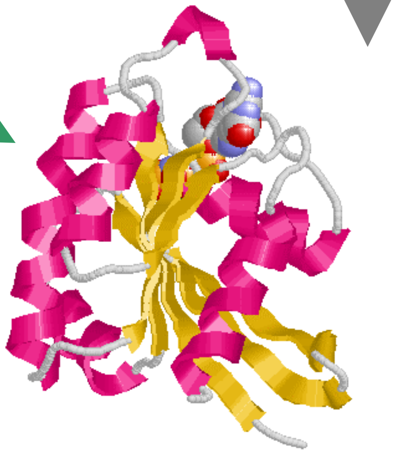
Era pre-genómica: los datos contenían la propia interpretación (conocimiento biológico). Ej. gel.

Interpretando los datos masivos



MREYKLVVLGSGGVGKSALTVQFVQGIFVDEYDPTIEDSY
RKQVEVDCQQCMLEILDTAGTEQFTAMRDLYMKNGQGFAL
VYSITAQSTFNDLQDLREQILRVKDTEDVPMILVGNKCDL
EDERVVGKEQGQNLARQWCNCAFLESSAKSKINVNEIFYD
LVRQINR

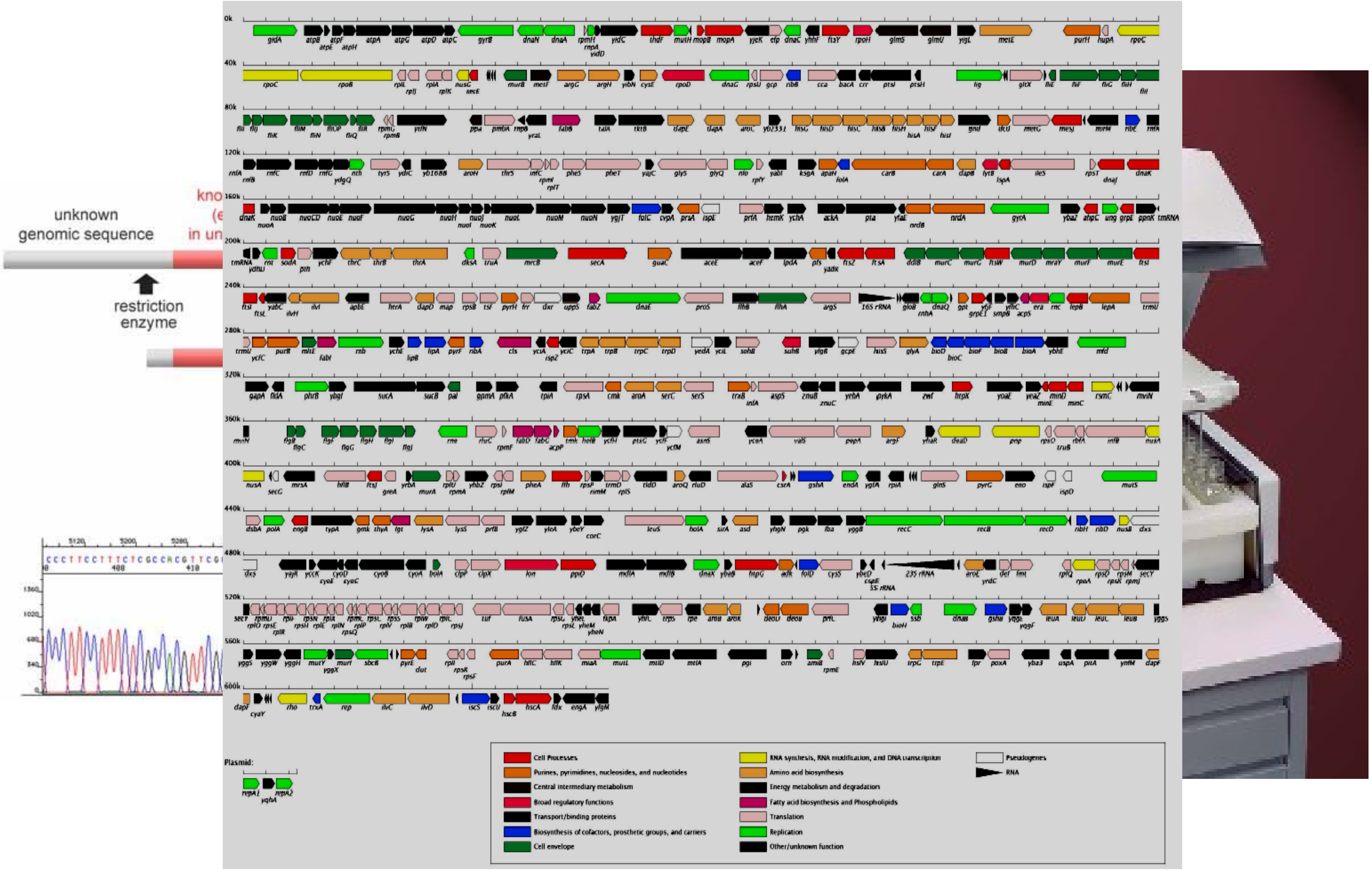
MLEILDTA**GTEQFT**AMRDLYMKNGQGFAL
VYSITAQSTFND**LQDLREQIL**RVKDTEDVPMIL
VGNKCDLEDERV



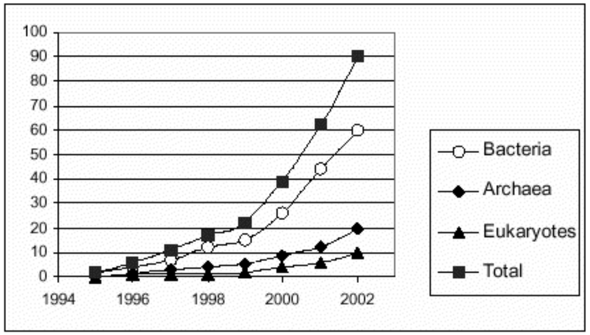
Paradigma “-omics” de la Biología
Caracterización **masiva** de los componentes a varios niveles y sus relaciones

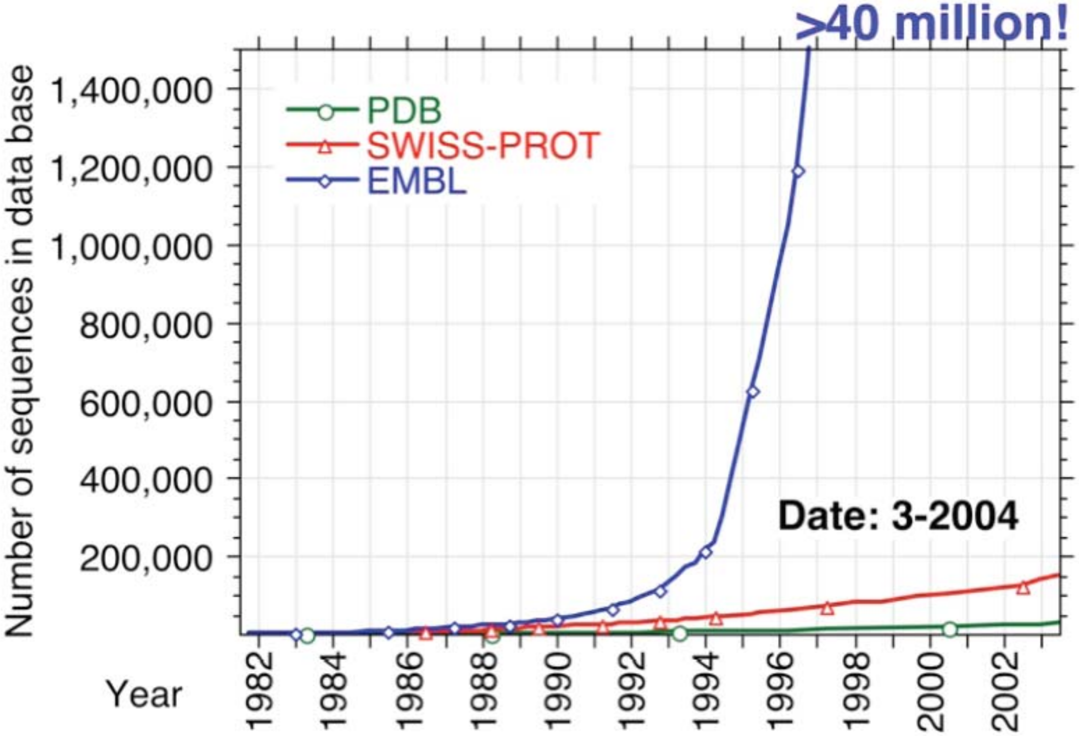
- Secuenciación de genomas completos (“genoma”).
- Caracterización de los transcritos (mRNA) (“transcriptoma”)
- Caracterización del repertorio proteico (“proteoma”)
- Localización celular de los componentes (“localizoma”)
- Redes de regulación génica (“reguloma”)
- Redes de interacciones entre proteínas (“interactoma”)
- Estudios masivos de relaciones gen-fenotipo (“fenoma”)
- Redes metabólicas (“metaboloma”)
- ...... “fosforiloma”, “funcionoma”, ...

Secuenciación de ADN



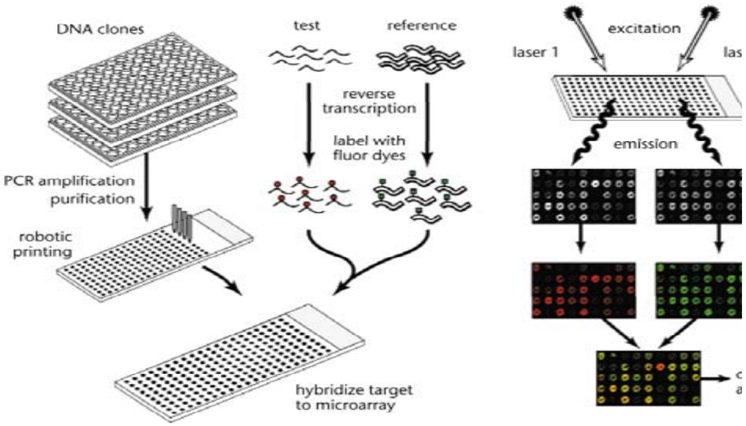
Secuenciación de Genomas Completos y “Metagenomas”

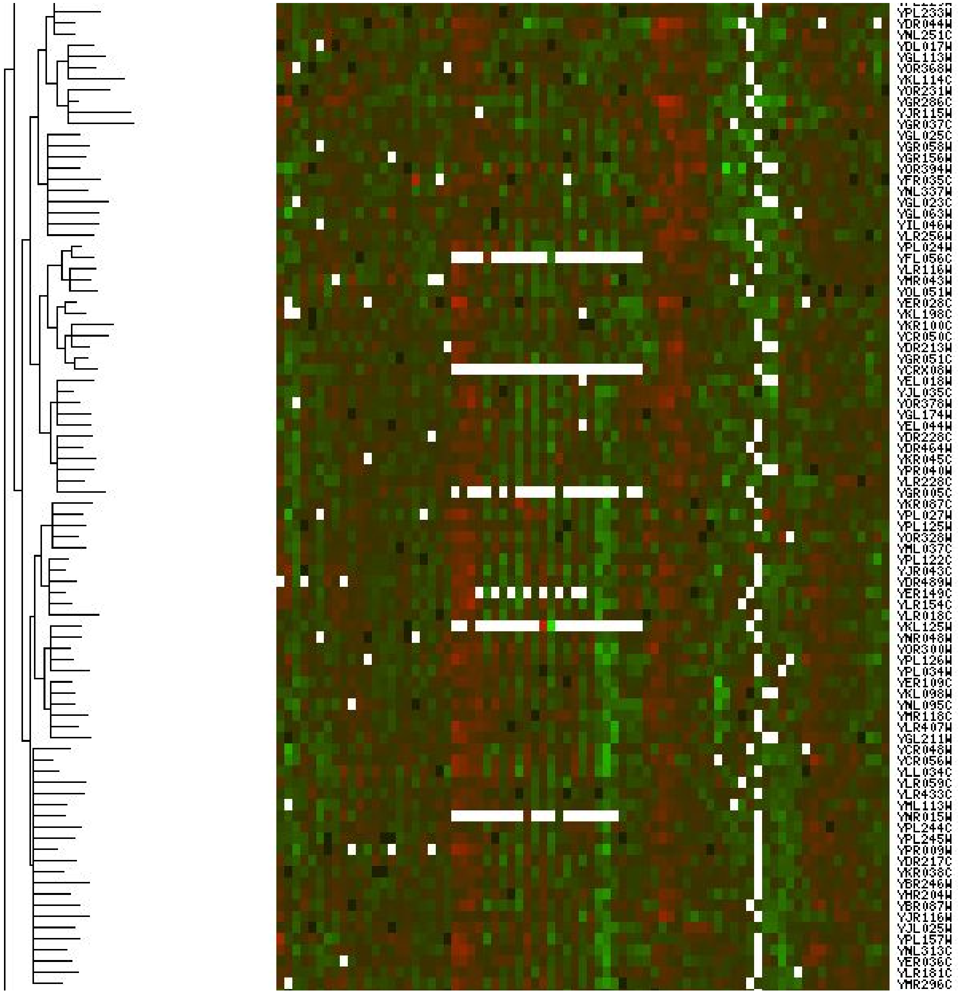




- Collins, F.S., Green, E.D., Guttmacher, A.E. & Guyer, M.S. (2003) A vision for the future of genomic research. *Nature*, **422**, 835-847.
- Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. (2004). *Science* **304,** 66-74.

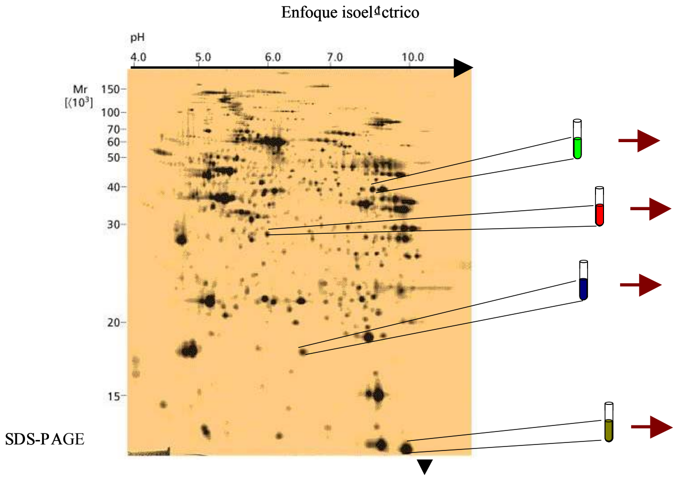
Caracterización de los transcritos (mRNA) (“transcriptoma”)

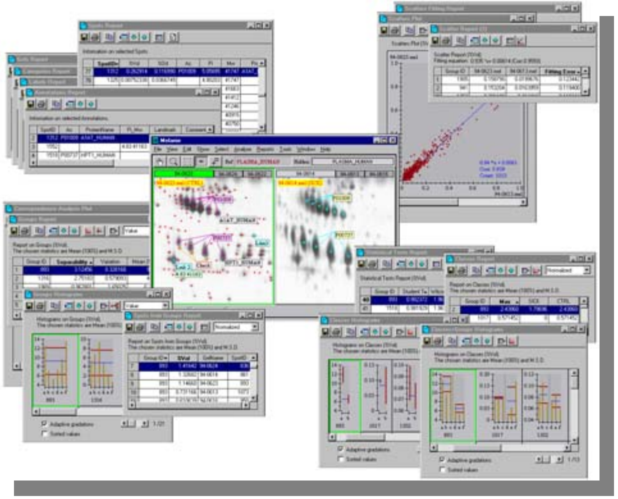


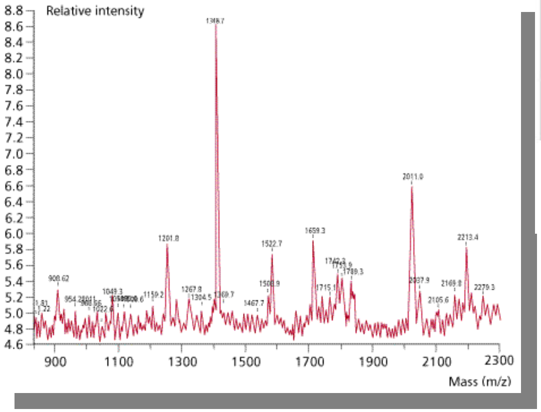


Caracterización del repertorio proteico (“proteoma”)

Gel
2D

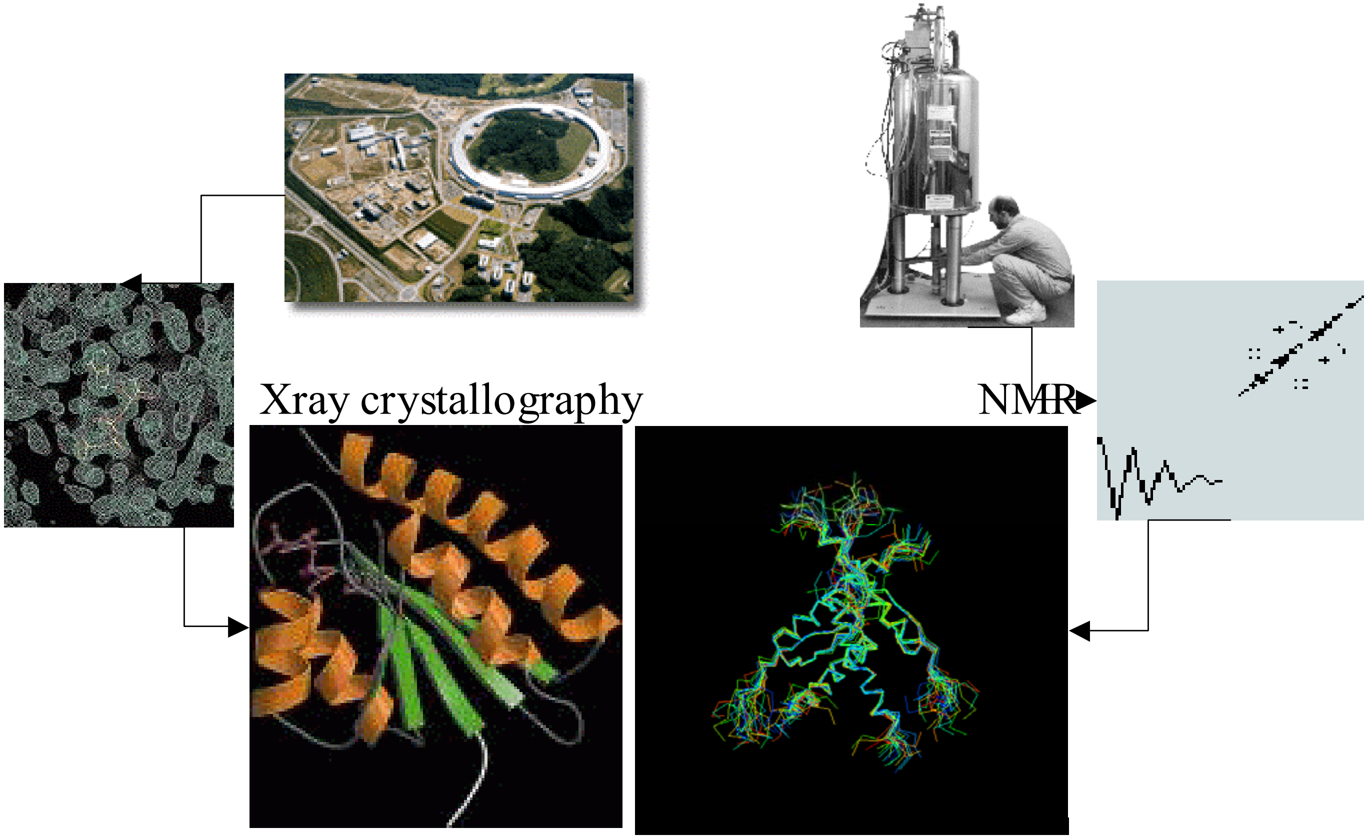






Juan A. G. Ranea

Caracterización del repertorio proteico (“proteoma”)
Estructura 3D

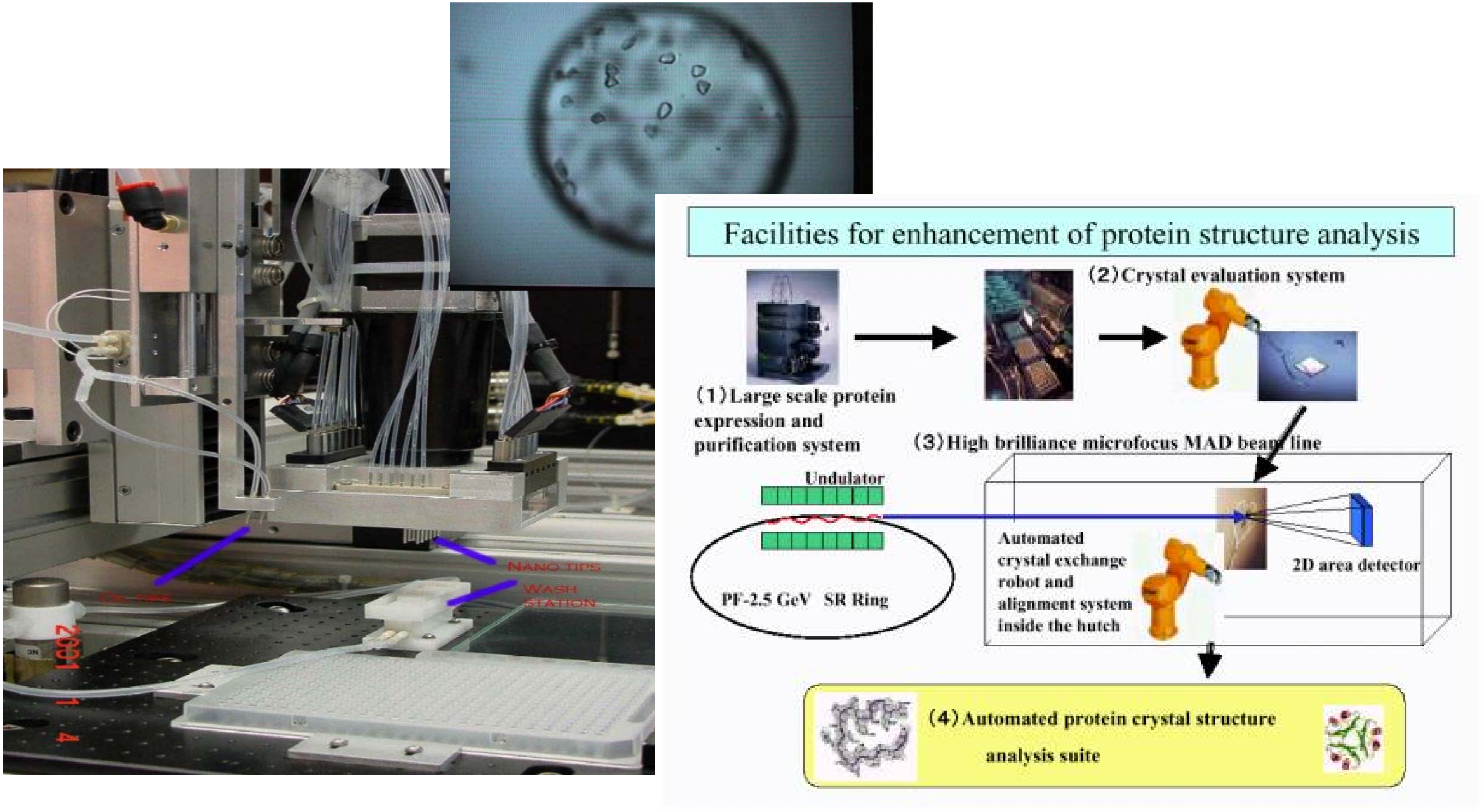


Xray crystallography

NMR

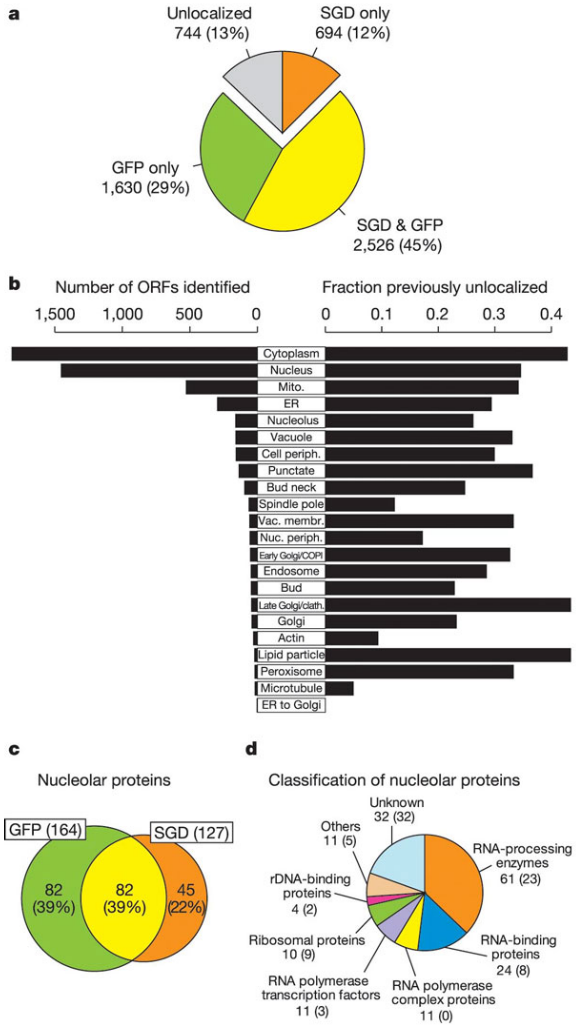
A.Valencia / D. Devos

Caracterización del repertorio proteico (“proteoma”)
Estructura 3D
Genómica Estructural



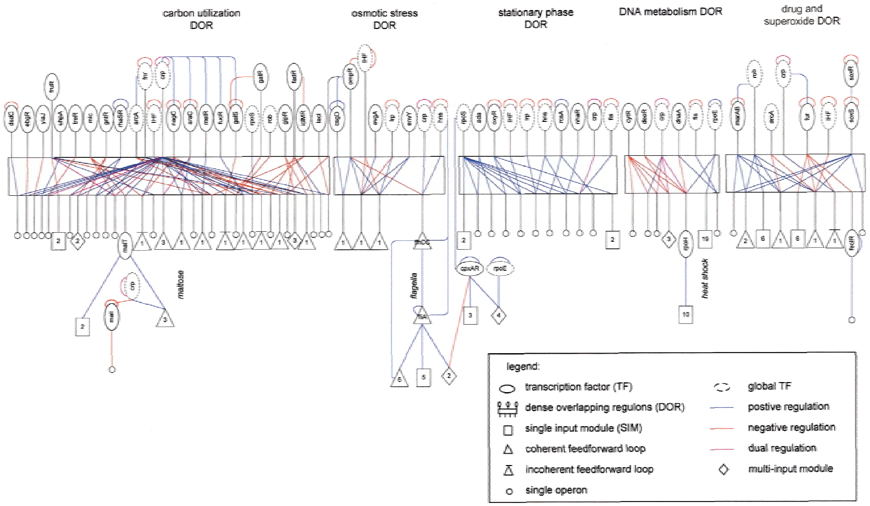
Vitkup, D., Melamud, E., Moult, J. and Sander, C. (2001) Completeness in structural genomics. *Nat Struct Biol*, **8**, 559-566.
Goldsmith-Fischman, S. and Honig, B. (2003) Structural genomics: Computational methods for structure analysis. *Protein Sci*, **12**, 1813-1821.

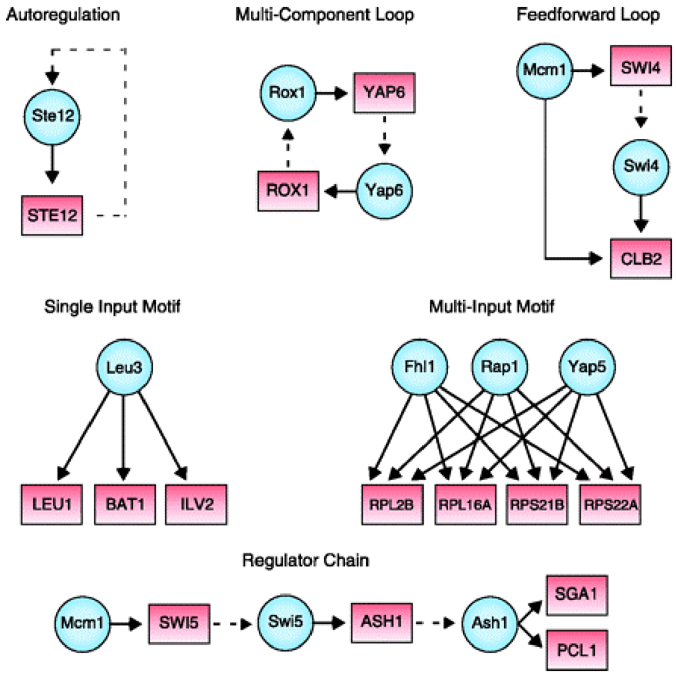
Localización celular de los componentes (“localizoma”)



Huh, W.K., *et al.* (2003) Global analysis of protein localization in budding yeast. *Nature*, **425**, 686-691.

Redes de regulación génica (“reguloma”)

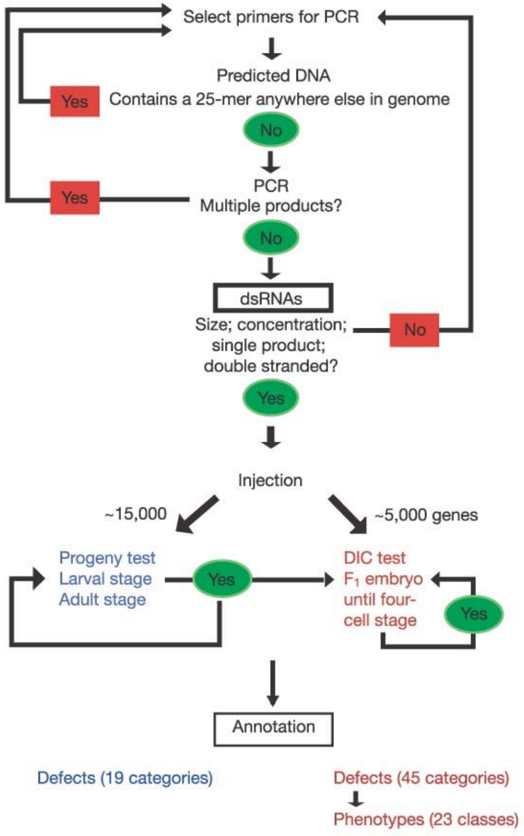


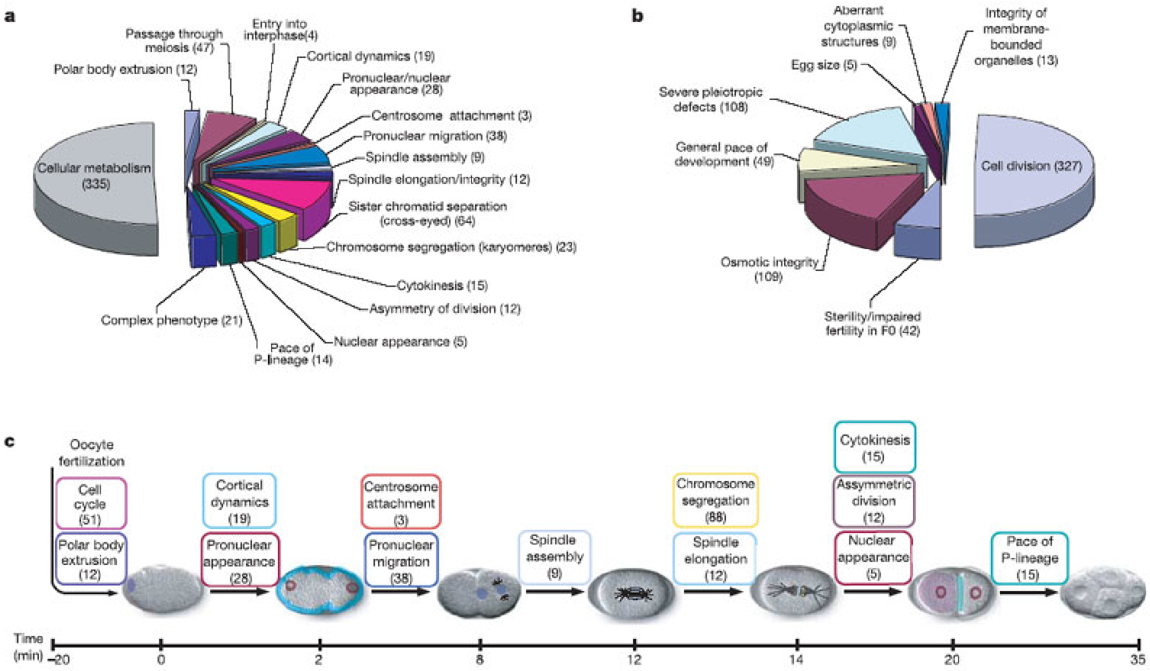


Lee, T.H., *et al.* (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, **298**, 799-804.

Shen-Orr, S.S., Milo, R., Mangan, S. & Alon, U. (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genet*, **31**, 64-68.

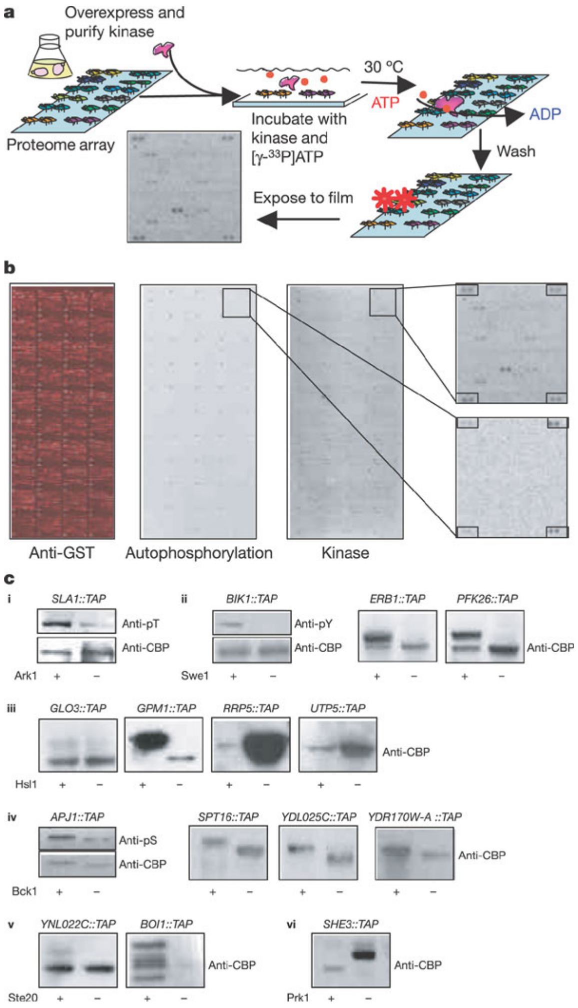
Estudios masivos de relaciones gen-fenotipo (“fenoma”)

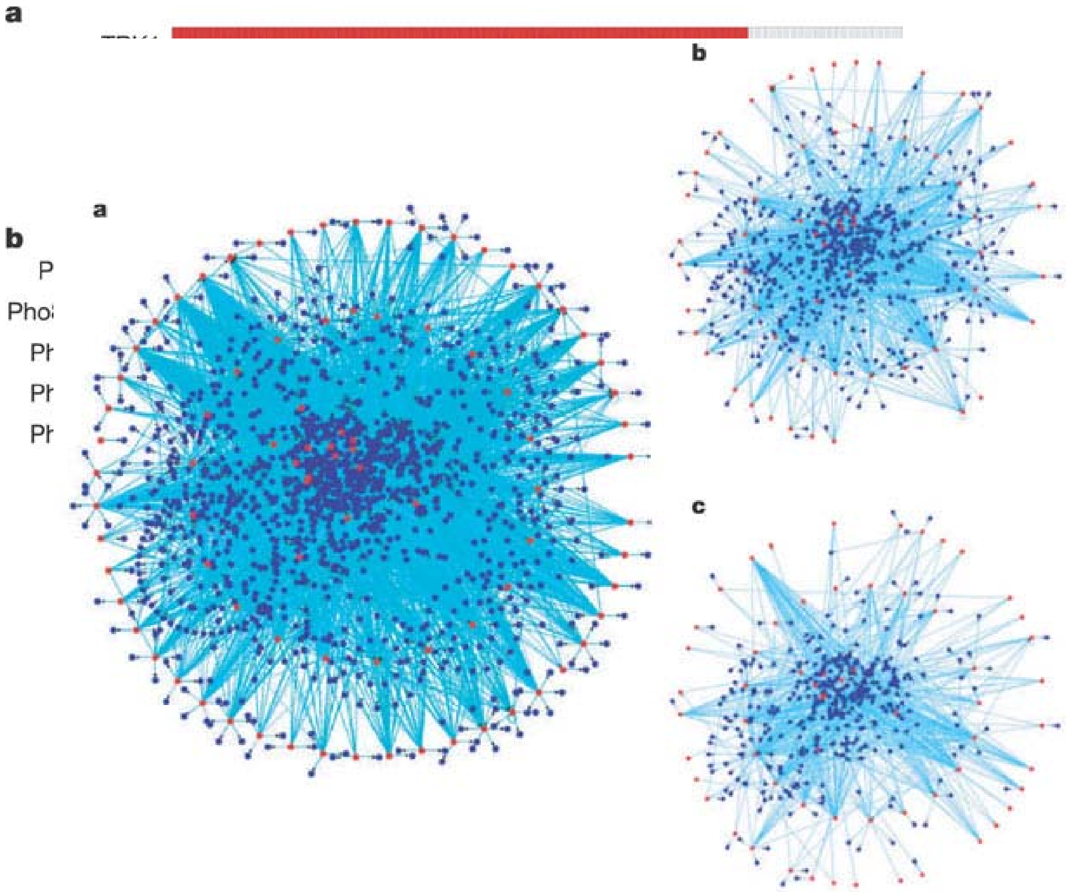




- Winzeler, E.A. & *al.*, *e.* (1999) Functional charazterization of the S. cerevisiae genome by delection and parallel analysis. *Science*, **285**, 901-906
- Kobayashi, K., Ehrlich, S.D., Albertini, A. *et al.* (2003) Essential bacillus subtilis genes. *Proc Natl Acad Sci U S A*, **100**, 4678-4683.
- Sönnichsen, B., Koski, L.B., Walsh, A., *et al.* (2005) Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. *Nature*, **434,** 462-469.

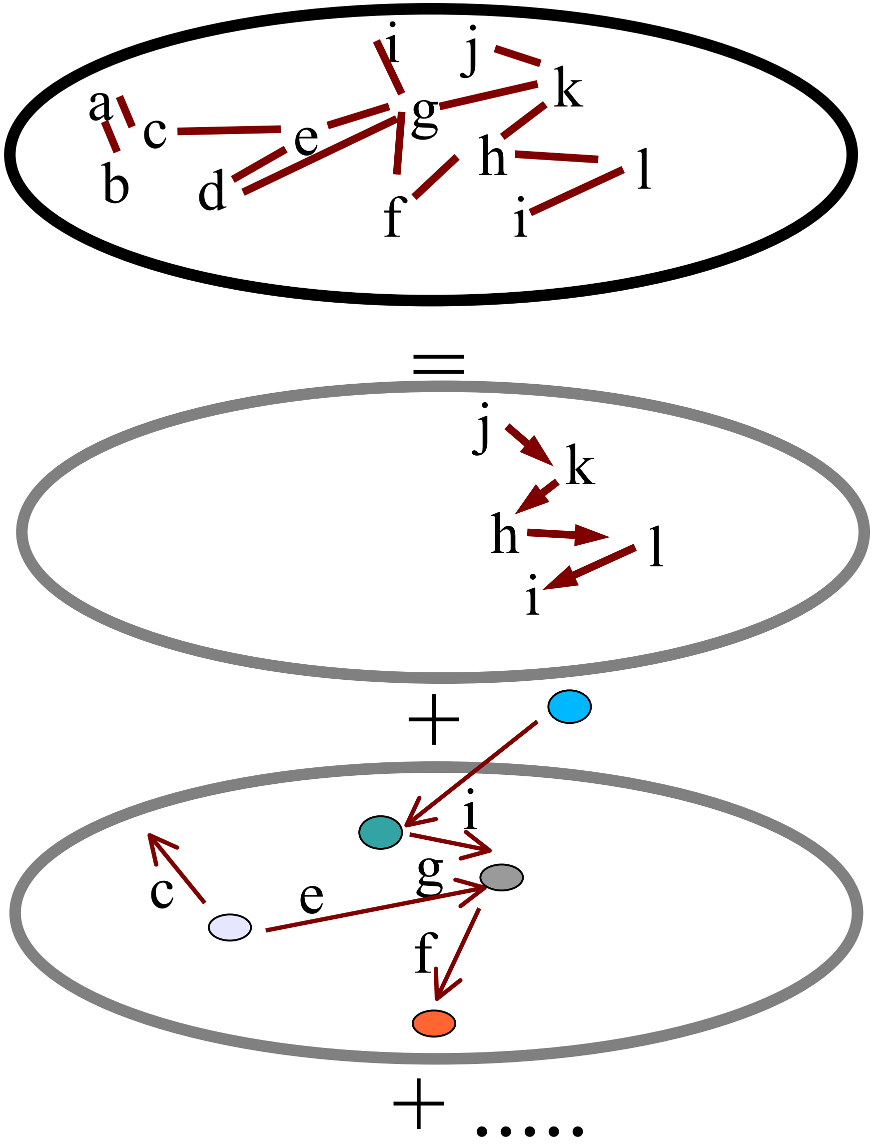
Otros... “fosforiloma”

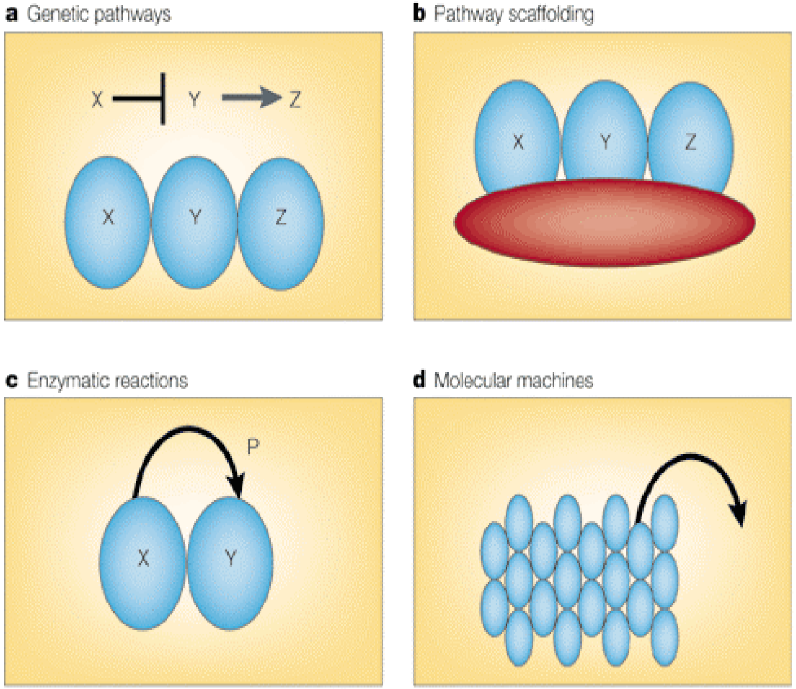




- Ptacek, J., Devgan, G., Michaud, G., Zhu, H., Zhu, X., Fasolo, J., Guo, H., Jona, G., Breitkreutz, A., Sopko, R., *et al.* (2005) Global analysis of protein phosphorylation in yeast. *Nature.*, **438**, 679-684.

Redes de interacciones entre proteínas (“interactoma”)





a Genetic pathways

b Pathway scaffolding

c Enzymatic reactions

d Molecular machines

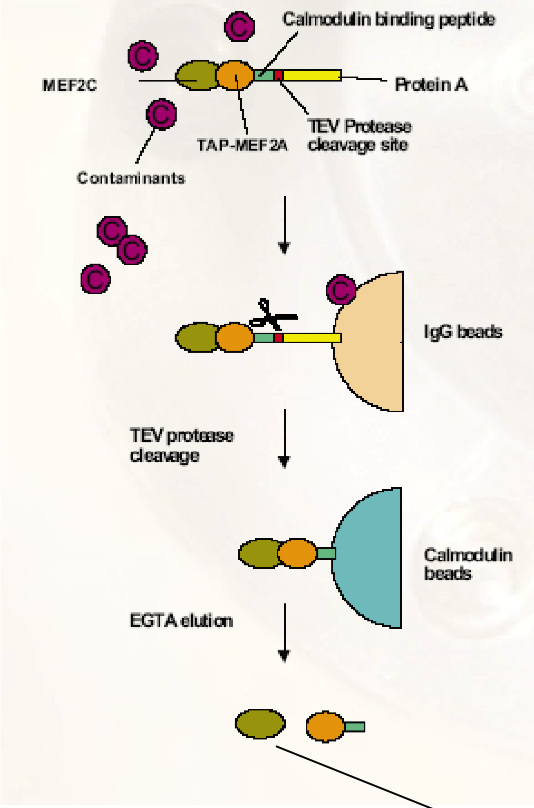
Nature Reviews | **Molecular Cell Biology**

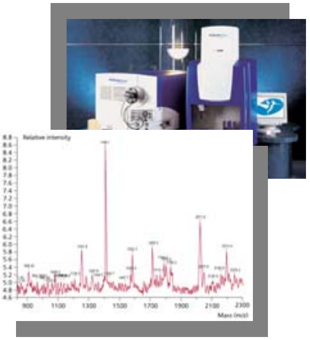
Walhout, A. J. & Vidal, M. (2001). Protein interaction maps for model organisms. *Nat Rev Mol Cell Biol* **2(1),** 55-62.

Redes de interacciones entre proteínas (“interactoma”)

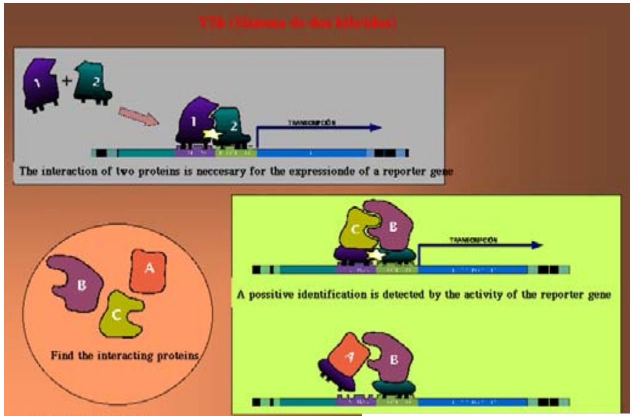
A.Valencia

TAP/MS





Y2H

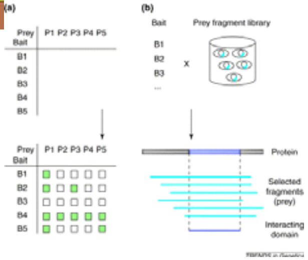

Y2h (Sistema de dos híbridos)

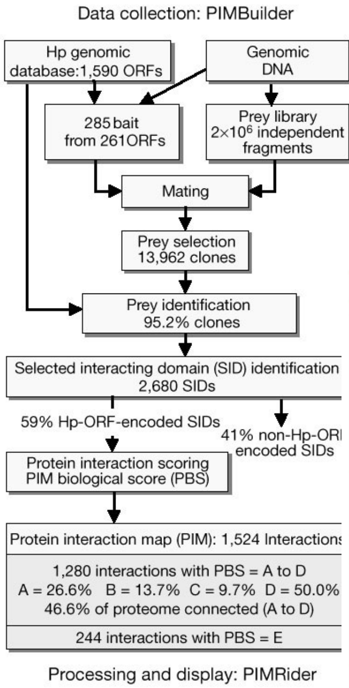
The interaction of two proteins is neccesary for the expressionde of a reporter gene

A possitive identification is detected by the activity of the reporter gene

Find the interacting proteins

New Technologies, Portugal 2000
Alfonso Valencia CNB-CSIC




Data collection: PIMBuilder

Hp genomic database:1,590 ORFs

Genomic DNA

285 bait from 261ORFs

Prey library 2×10^6 independent fragments

Mating

Prey selection 13,962 clones

Prey identification 95.2% clones

Selected interacting domain (SID) identification 2,680 SIDs

59% Hp-ORF-encoded SIDs

41% non-Hp-ORF encoded SIDs

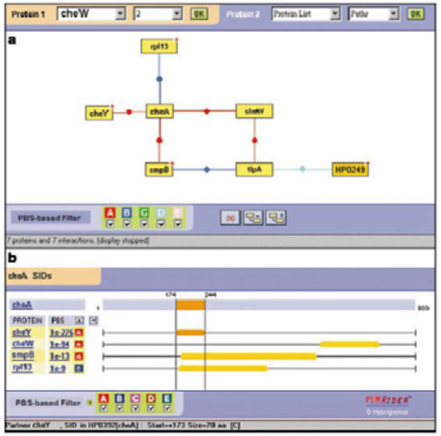
Protein interaction scoring PIM biological score (PBS)

Protein interaction map (PIM): 1,524 Interactions

1,280 interactions with PBS = A to D
A = 26.6% B = 13.7% C = 9.7% D = 50.0%
46.6% of proteome connected (A to D)

244 interactions with PBS = E

Processing and display: PIMRider


Protein 1: cheW, 2, OK
Protein 2: Protein List, Pathw, OK

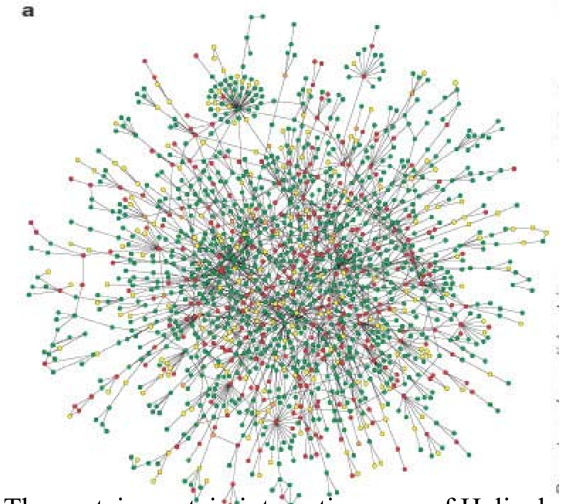
7 proteins and 7 interactions. [display stopped]

cheA SIDs

PROTEIN	PBS
cheY	1e-226 A
cheW	1e-94 A
smpB	1e-13 A
rpl13	1e-9 B

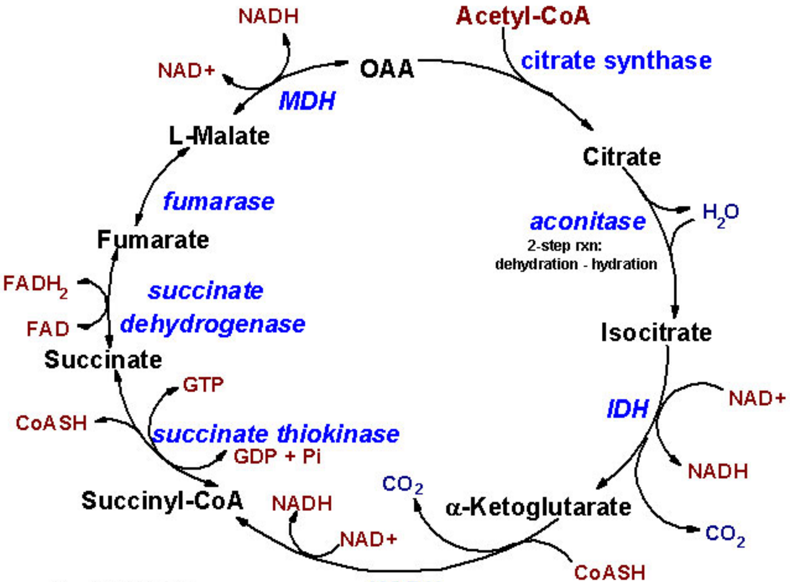
Partner cheY , SID in HP0392[cheA] : Start=+173 Size=70 aa [C]

Redes de interacciones entre proteínas (“interactoma”)

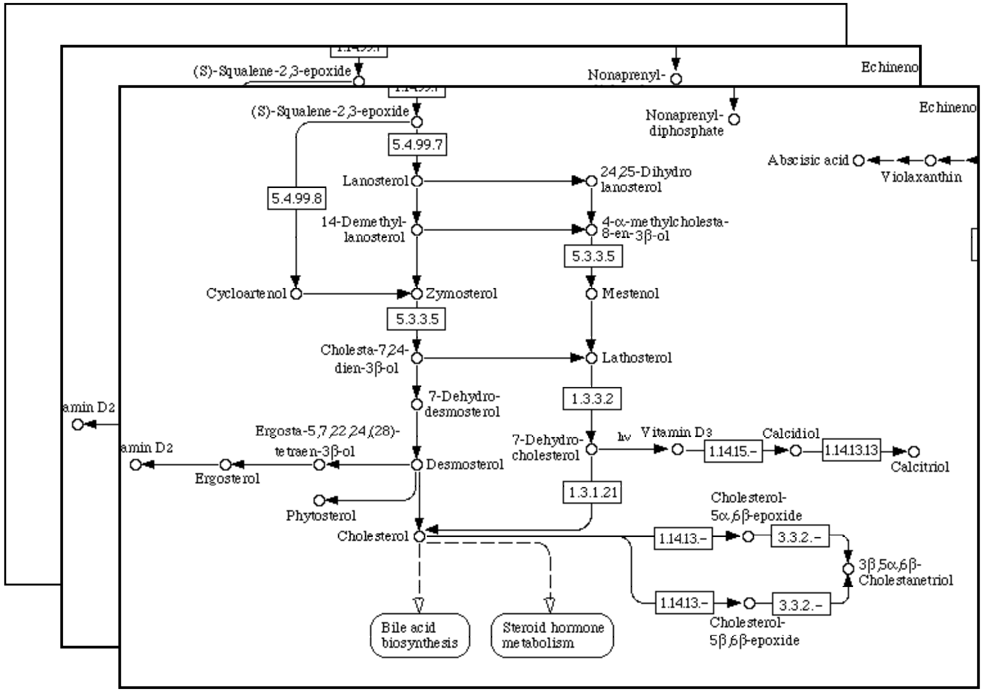


- Rain, J.C., Selig, L., De Reuse, H., et al. (2001) The protein-protein interaction map of Helicobacter pylori. *Nature*, **409,** 211-215.
- Gavin, A.C., *et al.* (2002) Functional organisation of the yeast proteome by systematic analysis of protein complexes. *Nature*, **415**, 141-147.
- Ho, Y., *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.
- Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., et al. (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, **122,** 957-968.
- 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.
- 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.
- 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.
- Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B., *et al.* (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature*., **440**, 631-636.
- Krogan, N.J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A.P., *et al.* (2006) Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature*., **440**, 637-643.

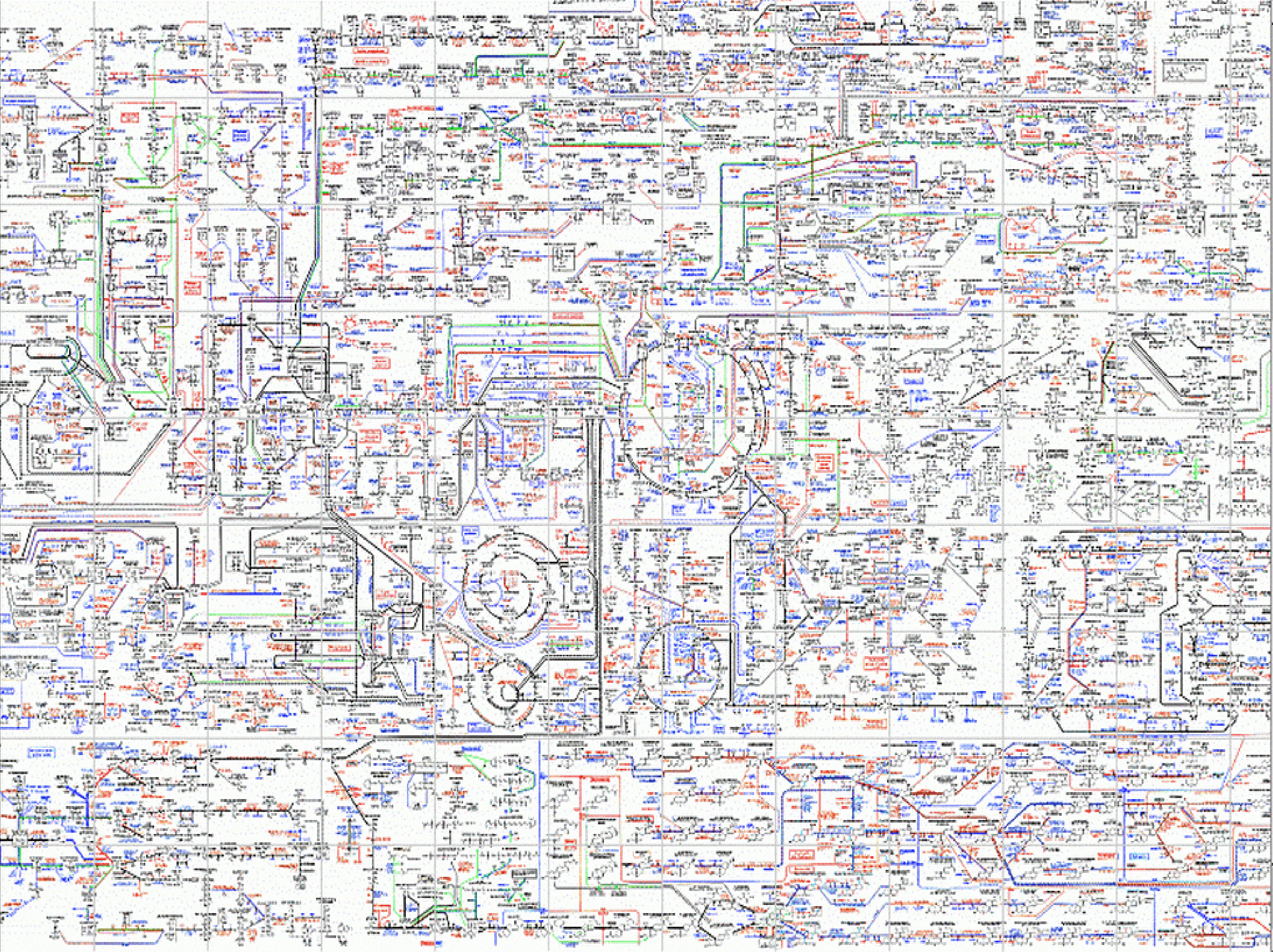
Redes metabólicas (“metaboloma”).



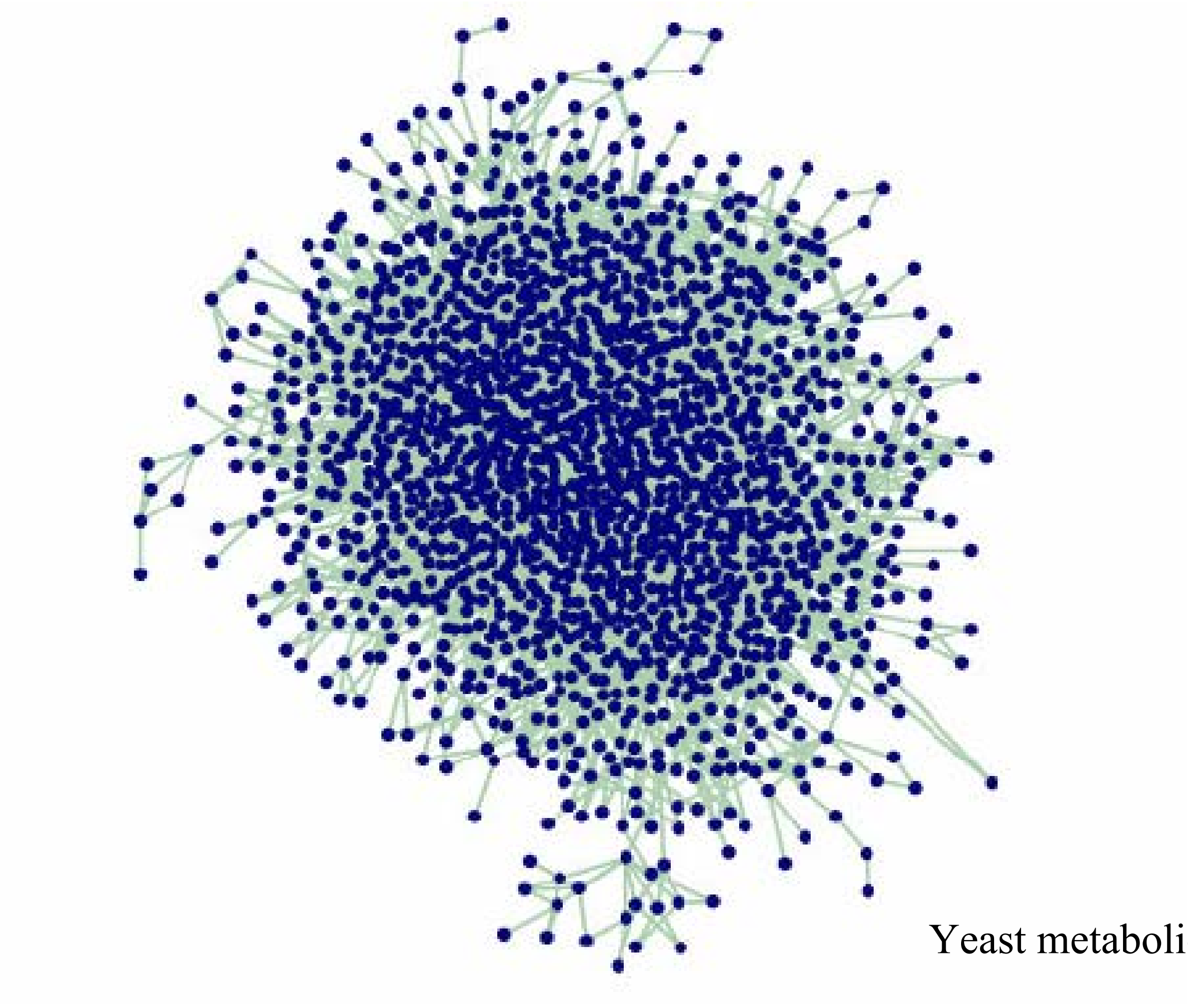
copyright 1996 M.W.King



Kanehisa, M., Goto, S., Kawashima, S. & Nakaya, A. (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res*, **30**, 42-46.

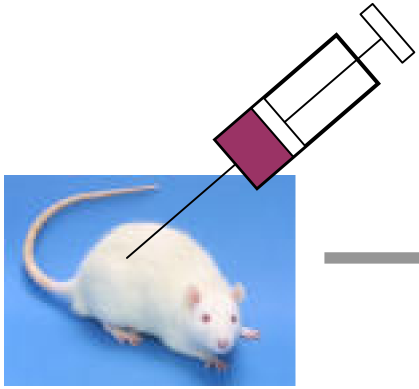


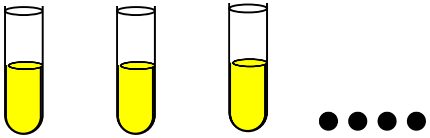
Redes metabólicas (“metaboloma”).

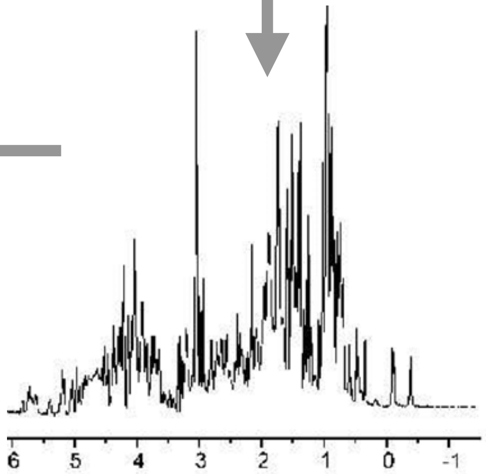


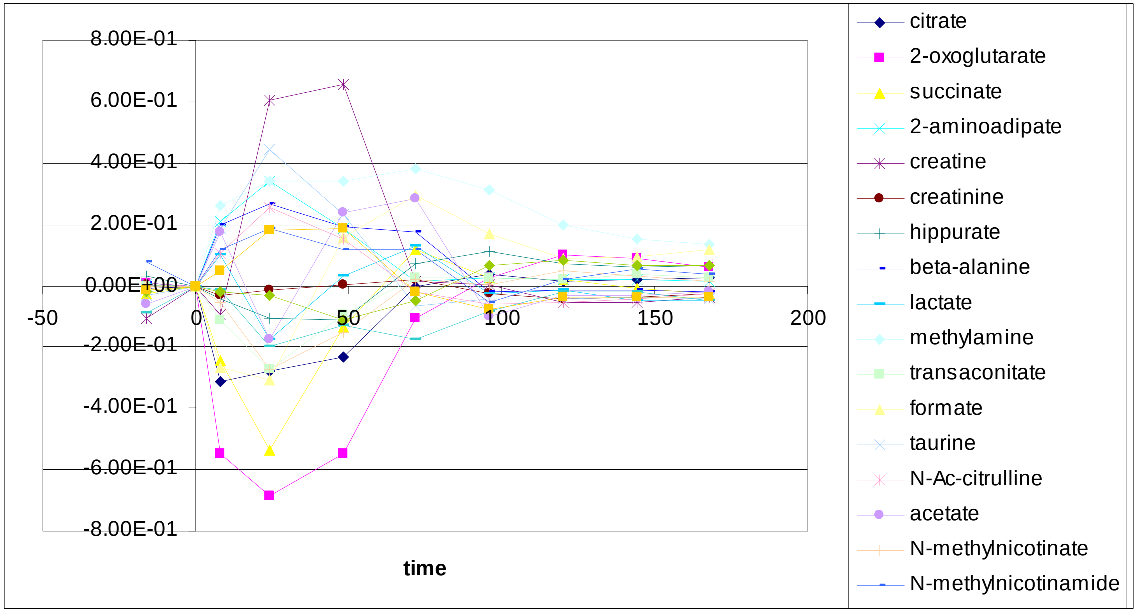
Yeast metabolic network

“Expresión” de Redes metabólicas (“metabonoma”).





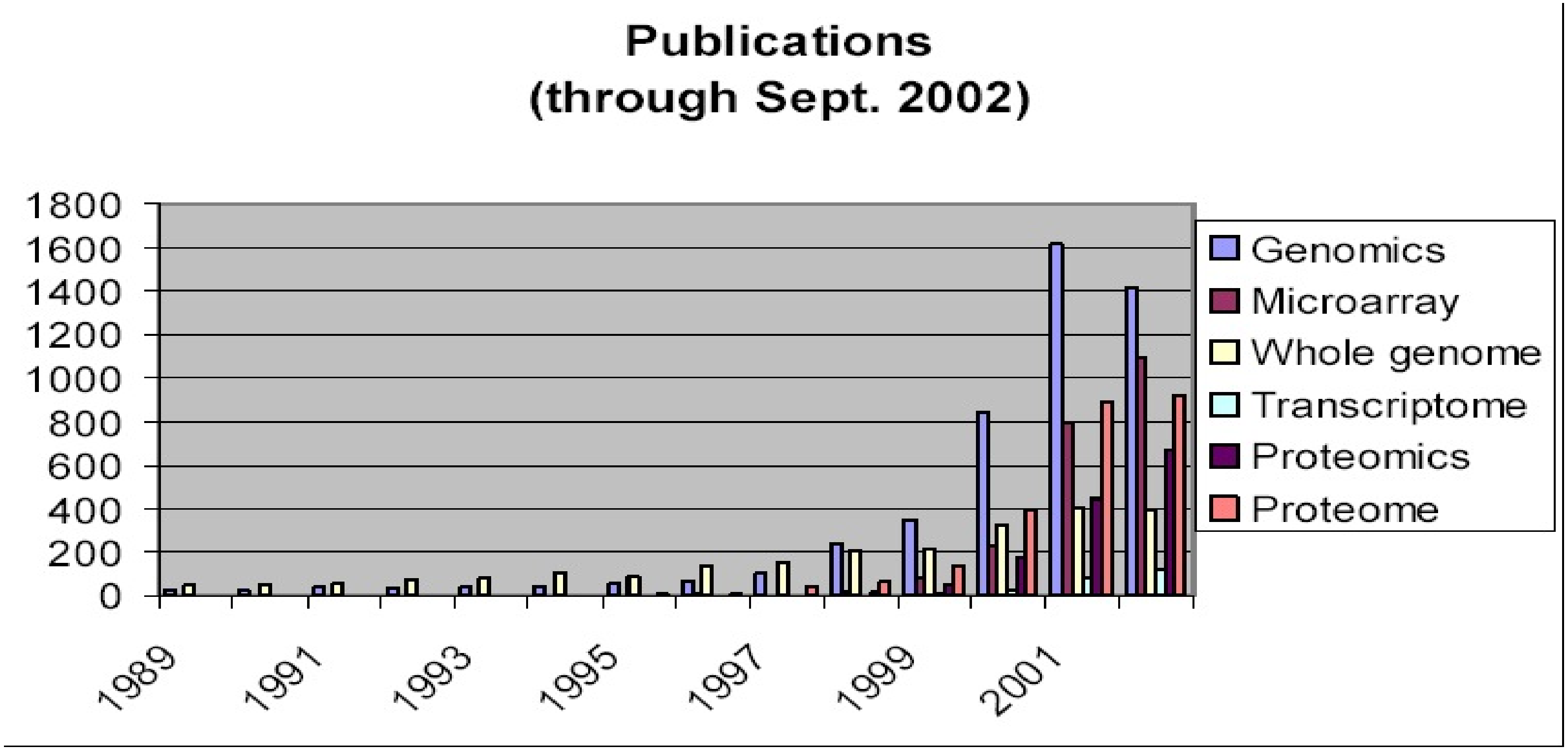




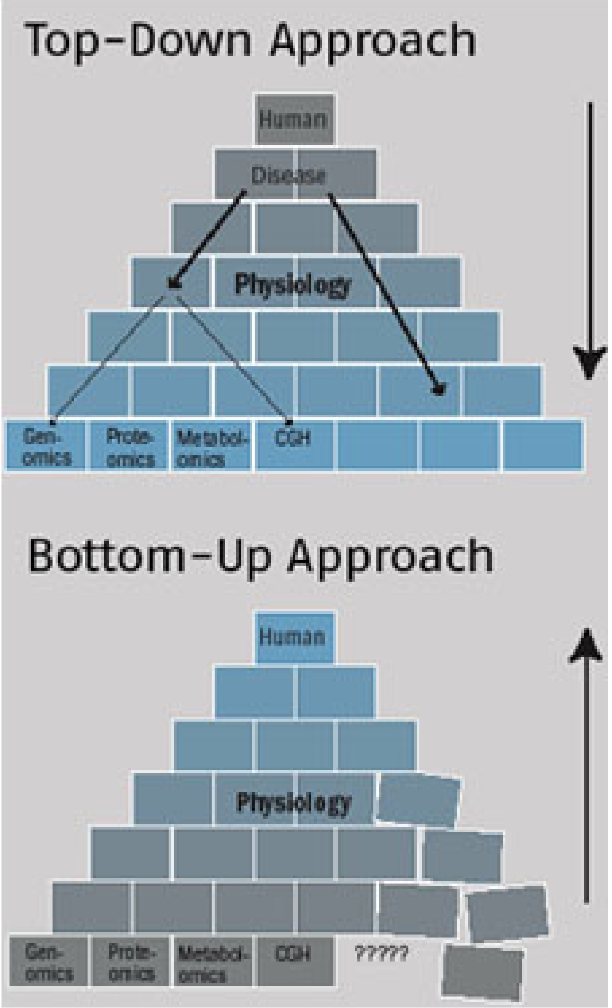
Lindon, J.C., Nicholson, J.K., Holmes, E., et al. (2003) Contemporary issues in toxicology. The role of metabonomics in toxicology and its evaluation by the COMET project. *Tox Appl Pharm*, **187,** 137-146.

Nicholson J. K., Connelly J., Lindon J. C. & Holmes E. (2002) Metabonomics: a platform for studying drug toxicity and gene function. *Nature Reviews Drug Discovery* **1,** 153 – 161

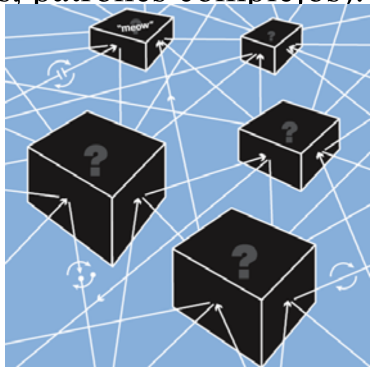
“-omics”


Publications
(through Sept. 2002)
1800
1600
1400
1200
1000
800
600
400
200
0
1989
1991
1993
1995
1997
1999
2001
Genomics
Microarray
Whole genome
Transcriptome
Proteomics
Proteome

Bioinformática y Sistemas Complejos en Biología
Biología de Sistemas vs. Biología Molecular



- Visión desde el punto de vista de sistemas complejos
- Objeto de estudio: redes, relaciones, propiedades emergentes (no propiedades de componentes individuales).
- Biología de Sistemas
- Propiedades “globales” (emergentes, etc.) -> conocimiento biológico
- Enfermedades (dianas/marcadores = redes, patrones complejos).



- Visión reduccionista
- Objeto de estudio: componentes (genes proteínas).
- Biología Molecular
- Propiedades de los componentes -> conocimiento biológico
- Enfermedades (dianas/marcadores = proteínas/genes).

Michael Liebman (www.bioitworld.com)

¿Es suficiente la visión reduccionista?

- Reduccionismo en Biología muy exitoso (Biología Molecular). *"The ultimate aim of the modern movement in biology is to explain* ***all*** *biology in terms of physics and chemistry". F. Crick (1966)*

- Sistemas biológicos: prototipo de sistemas complejos. => Muchos fenómenos biológicos nunca podrán explicarse a partir de las propiedades de los componentes ("el todo es mas que la suma de las partes").

- Determinación de "repertorios de componentes" y sus caracteristicas (secuenciación de genomas, proteómica, genómica estructural ...): Ni el número ni las características de genes y proteínas dan cuenta de muchas características de los organismos:
    - Similar número de genes en Drosophila y C. elegans.
    - Alta similaridad de secuencia entre humano y ratón.
    - Genoma del erizo: #receptores de luz y olor similar a vertebrados (1000); proteínas que en humano están en el oído; 20 veces más receptores Toll (resp. Inmune) que en humano; ...
    - ...

•Van Regenmortel, M.H. (2004) Reductionism and complexity in molecular biology. Scientists now have the tools to unravel biological and overcome the limitations of reductionism. *EMBO Rep*, **5**, 1016-1020.

¿Es suficiente la visión reduccionista?

- Fallo en tratamiento de cáncer, ... En parte debido al enfoque reduccionista extremo.

- No vacuna HIV, ... “

- Fracaso vacunas de péptidos. “

- Reducción de fármacos en el mercado a pesar de la creciente inversión. (1 fármaco <–> 1 diana).

- No resultados esperados para técnicas terapéuticas prometedoras de base reduccionista (terapia génica, RNA antisentido, ...).

- No mejora esperada de estas aproximaciones con secuenciación de genomas, etc.

•Van Regenmortel, M.H. (2004) Reductionism and complexity in molecular biology. Scientists now have the tools to unravel biological and overcome the limitations of reductionism. *EMBO Rep*, **5**, 1016-1020.

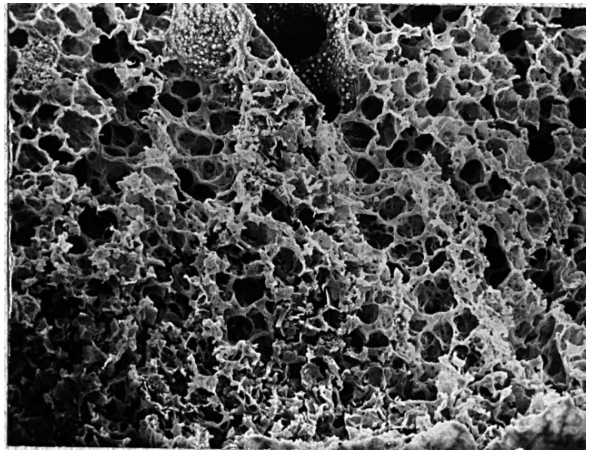
¿Es suficiente la visión reduccionista?

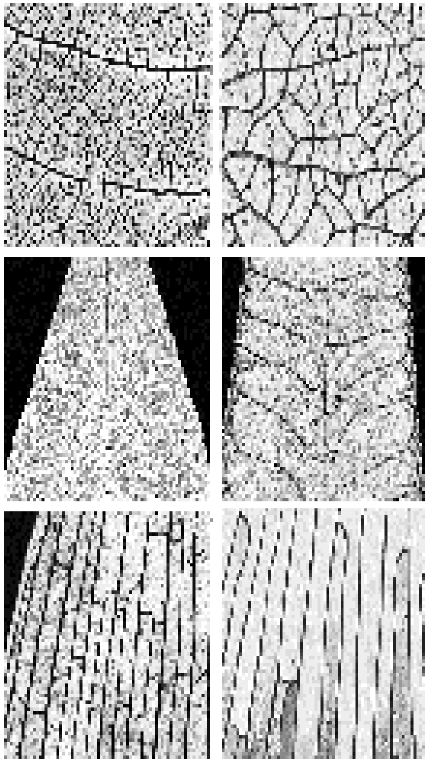
Fallos en aproximaciones *in-vitro* e *in-silico*.

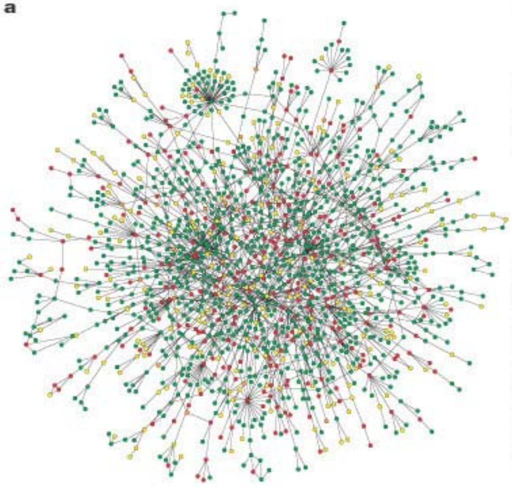
Fallos en técnicas experimentales de base reduccionista (delecciones, *knockout, ...).*
Knockout: no efecto, efecto distinto al esperado, o efecto "inespecífico" (cambio expresión 100's genes).

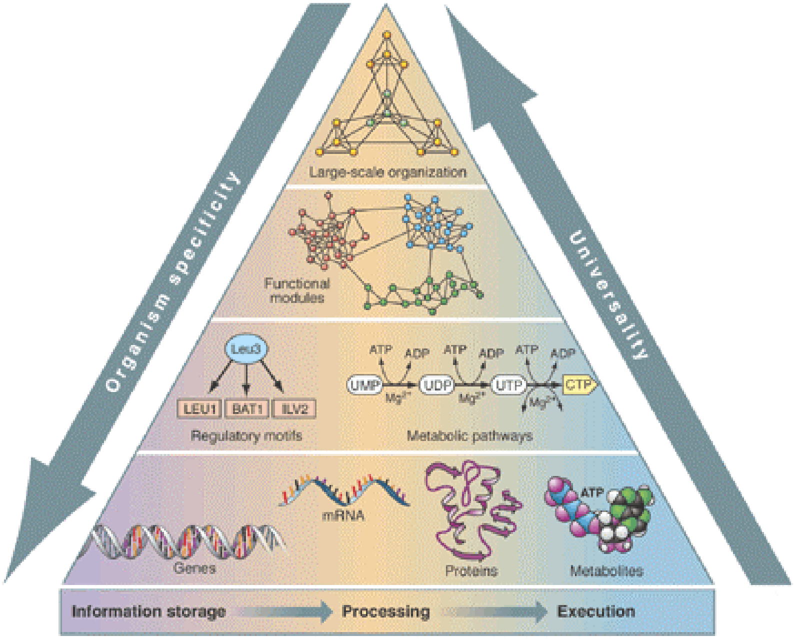
> *"Some mice should, by rights, be dead. At the very least, Teyumuras Kurzchalia expected his to be critically ill. But the most prominent symptom of his genetically engineered mice was a persistent erection"*
>
> Pearson, H. (2002) Surviving a knockout blow. *Nature*, **415**, 8-9.

Características de sistemas complejos biológicos que escapan a una explicación reduccionista *Emergencia*





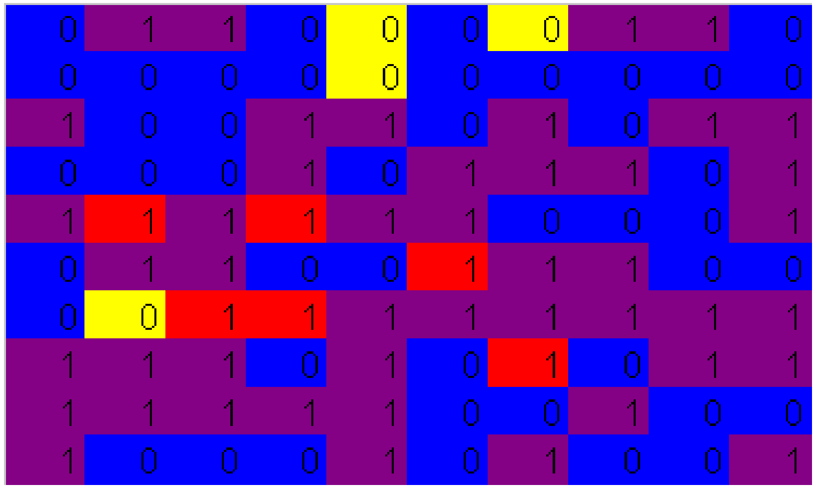




Oltvai, Z.N. and Barabási, A.-L. (2002)
Life’s complexity pyramid. *Science*, **298**, 763–764.

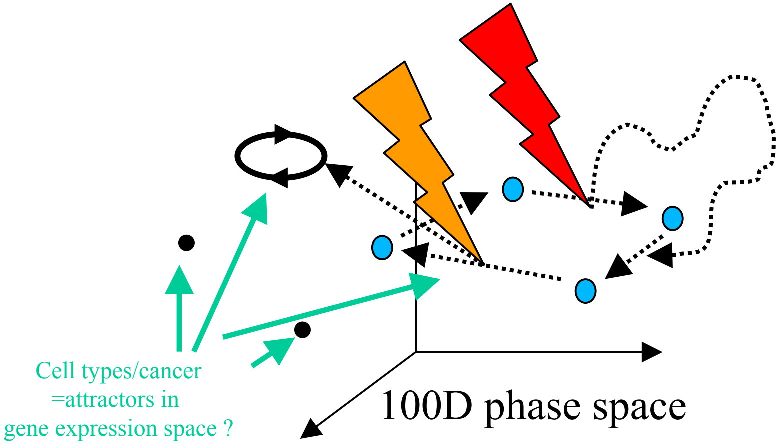
Características de sistemas complejos biológicos que escapan a una explicación reduccionista

Robustez

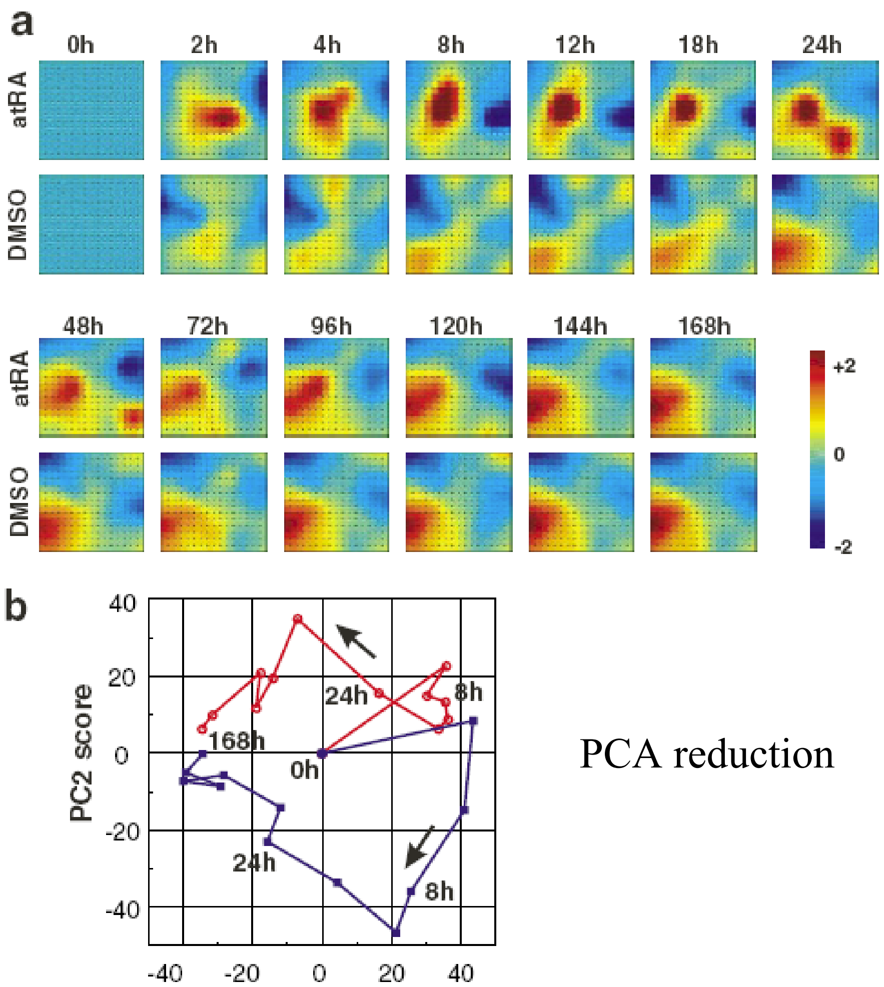


0	1	1	0	0	0	0	1	1	0
0	0	0	0	0	0	0	0	0	0
1	0	0	1	1	0	1	0	1	1
0	0	0	1	0	1	1	1	0	1
1	1	1	1	1	1	0	0	0	1
0	1	1	0	0	1	1	1	0	0
0	0	1	1	1	1	1	1	1	1
1	1	1	0	1	0	1	0	1	1
1	1	1	1	1	0	0	1	0	0
1	0	0	0	1	0	1	0	0	1

http://www.bion.si/BoolMrezeE.htm



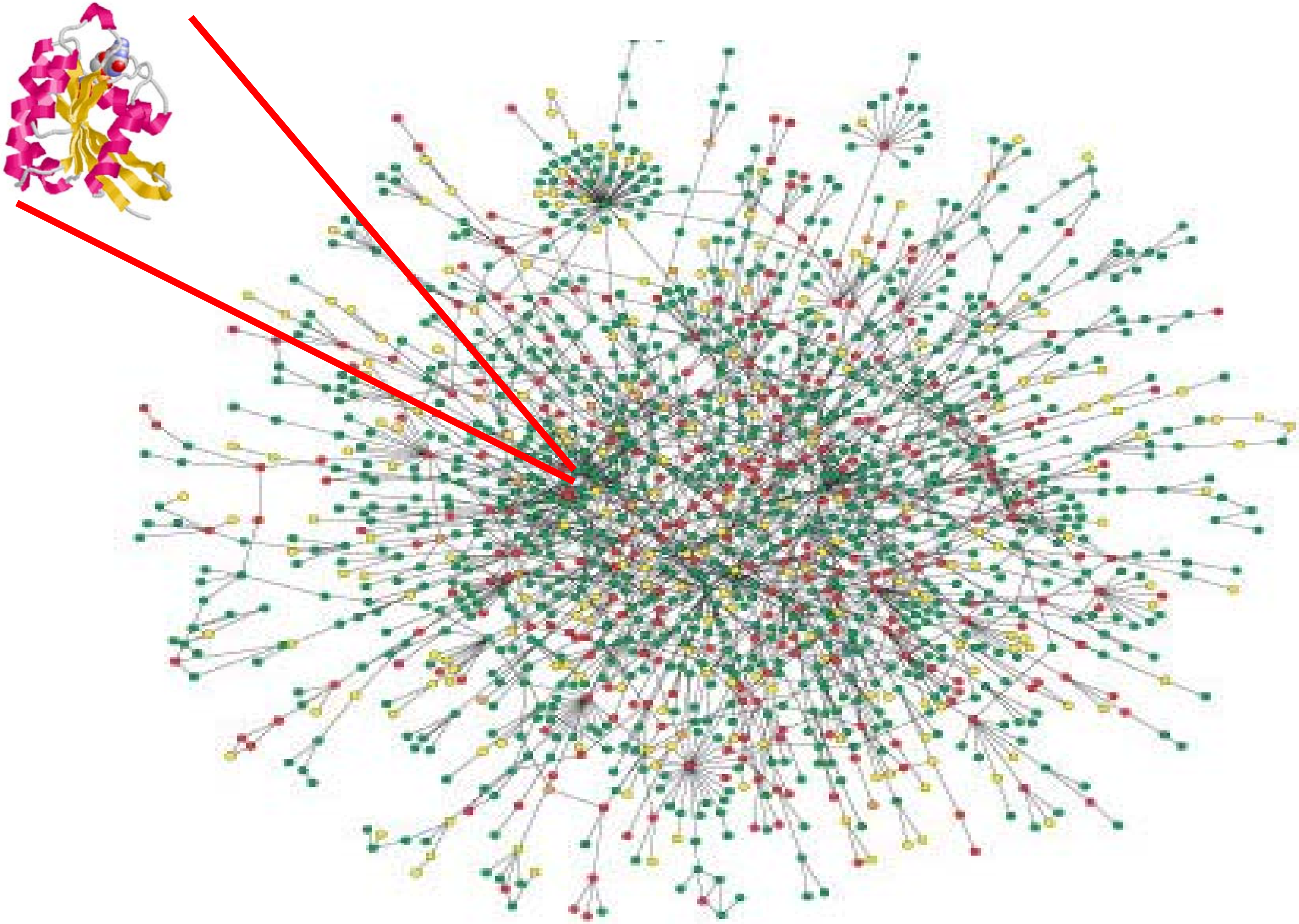
2773D phase space



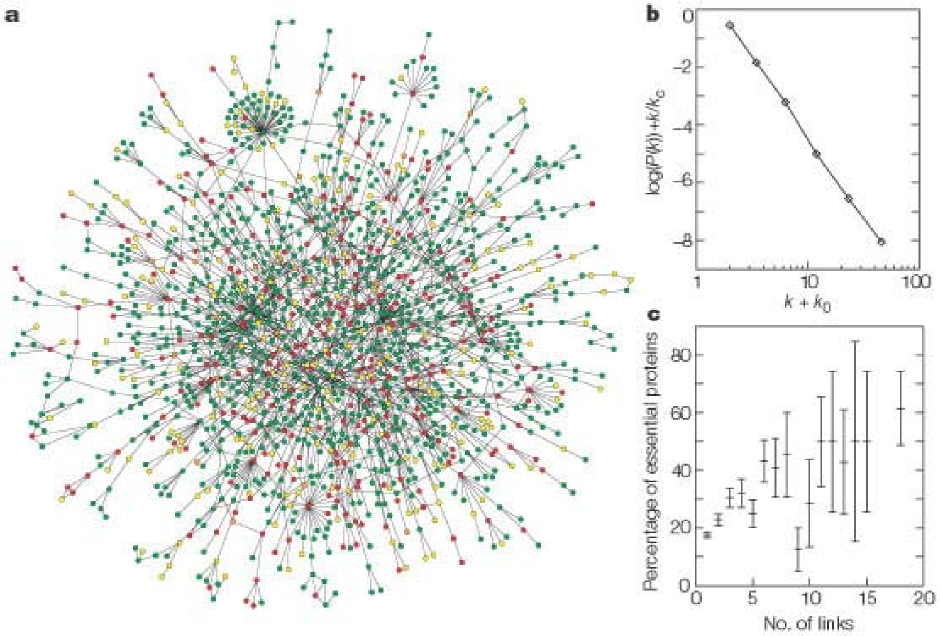
PCA reduction

Kauffman, S.A. (1993) *The Origins of Order*. Oxford University Press, New York.

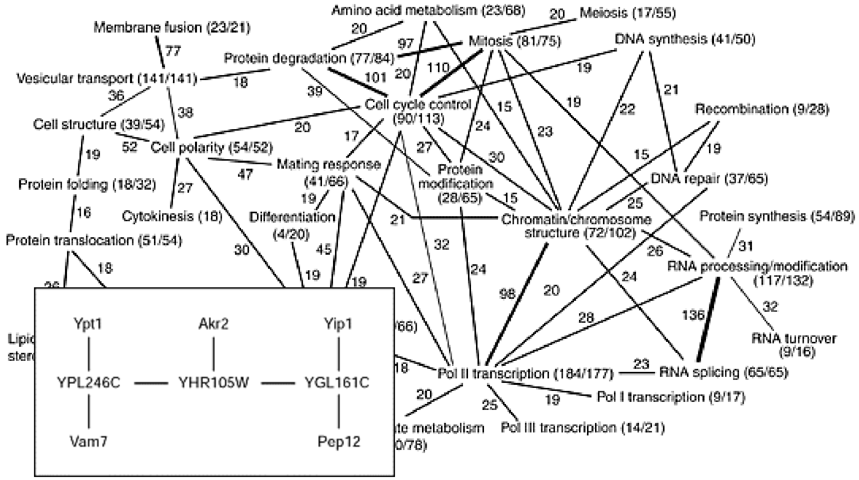
Huang, S., Eichler, G., Bar-Yam, Y. and Ingber, D.E. (2005) Cell Fates as High-Dimensional Attractor States of a Complex Gene Regulatory Network. *Phys Rev Letters*, **94**, 128701.



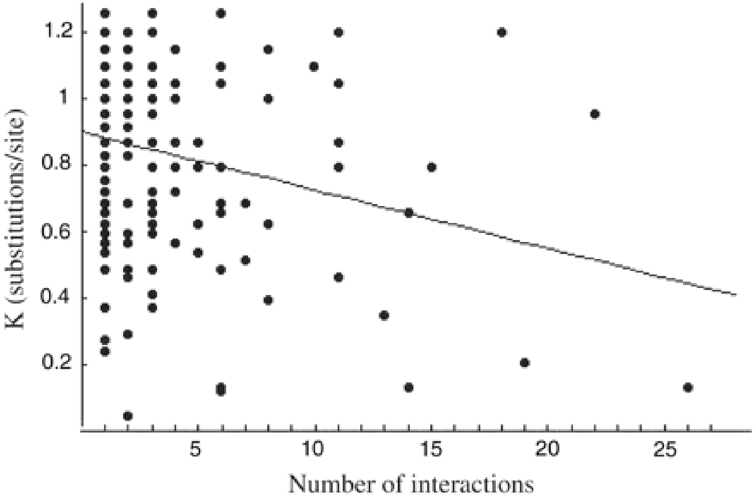
Estudio desde un enfoque de “Sistemas”
“Interactoma”



Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature* **411**, 41-42.

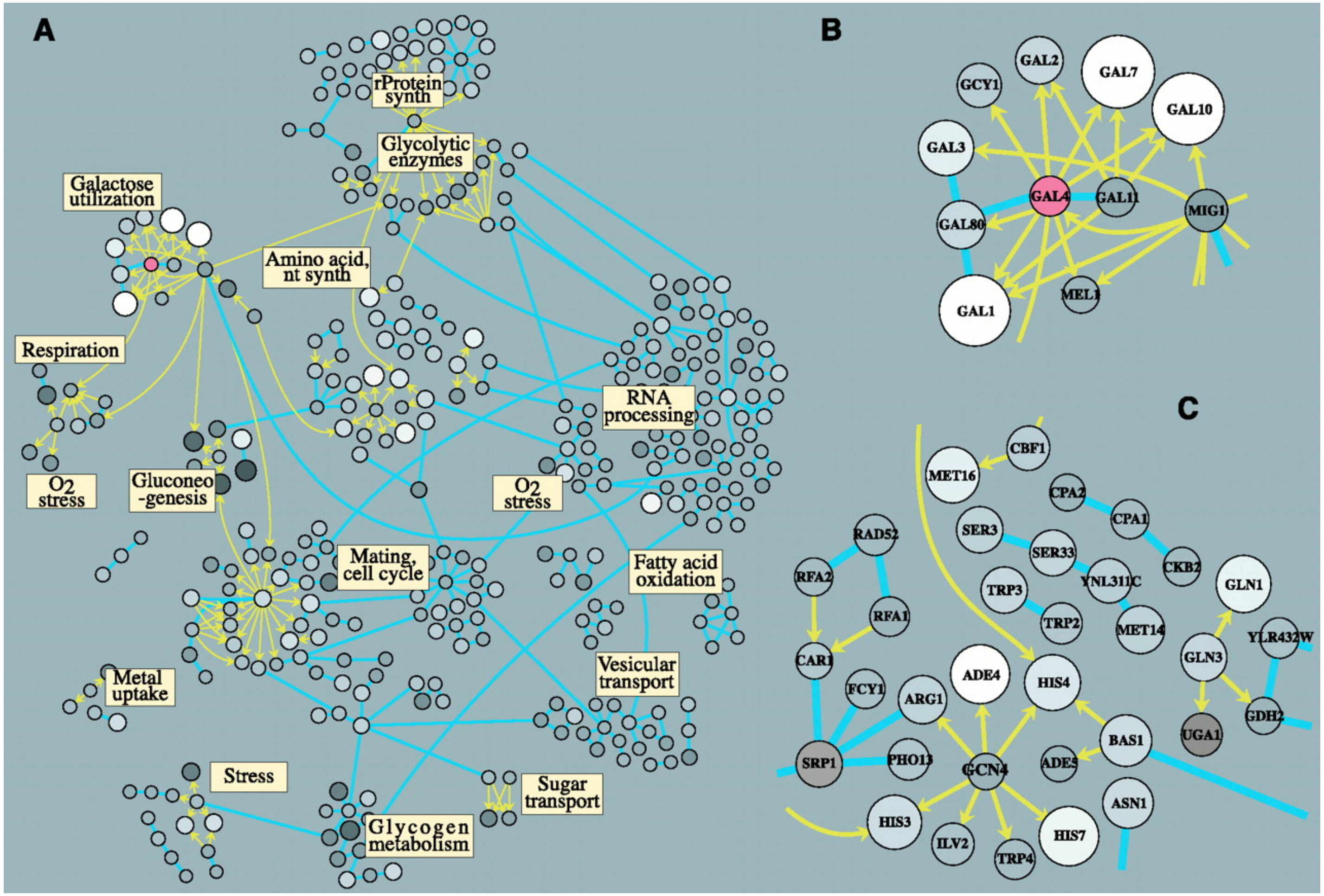


Schwikowski, B., Uetz, P. & Fields, S. (2002). A network of protein-protein interactions in yeast**. *Nature Biotech* 18**, 1257-1261.



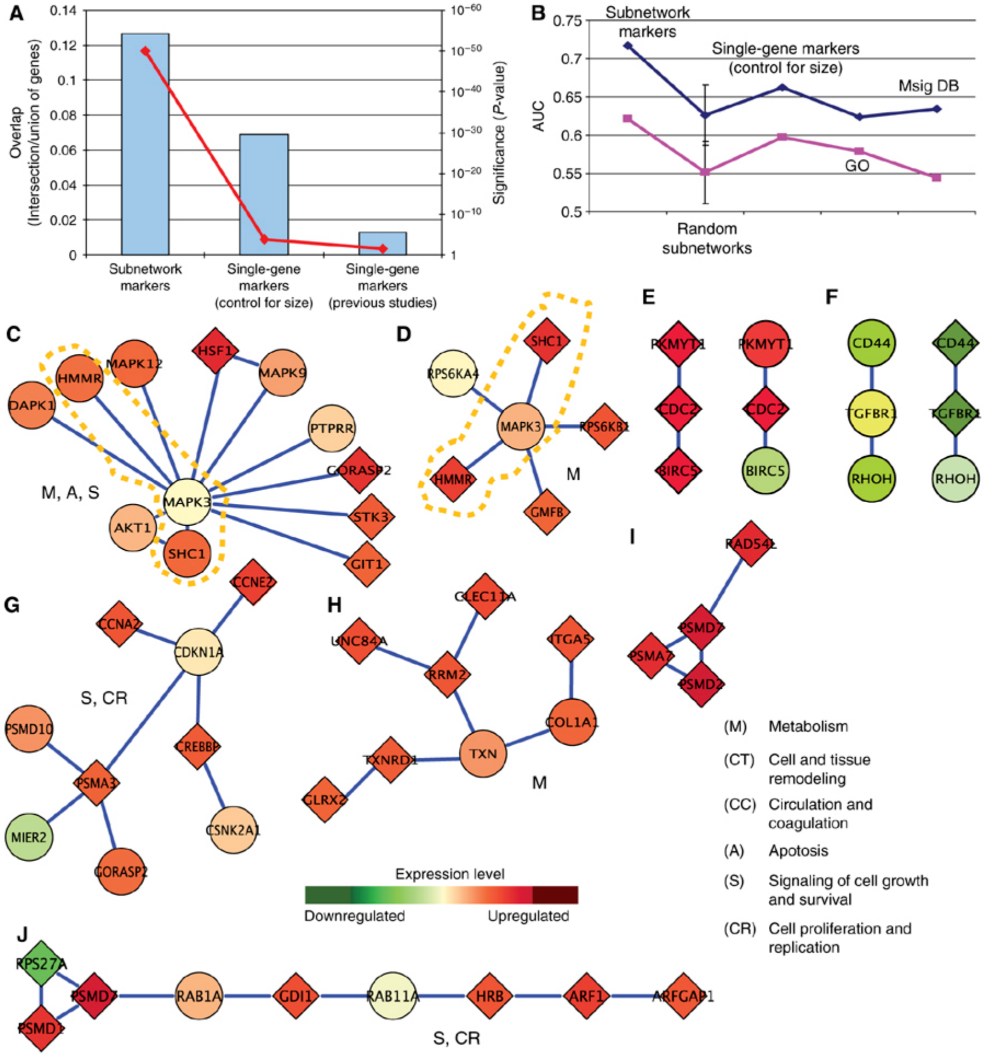
Fraser, H.B., Hirsh, A.E., Steinmetz, L.M., Scharfe, C. and Feldman, M.W. (2002) Evolutionary rate in the protein interaction network. *Science*, **296**, 750-752.

Biología de Sistemas y Terapia



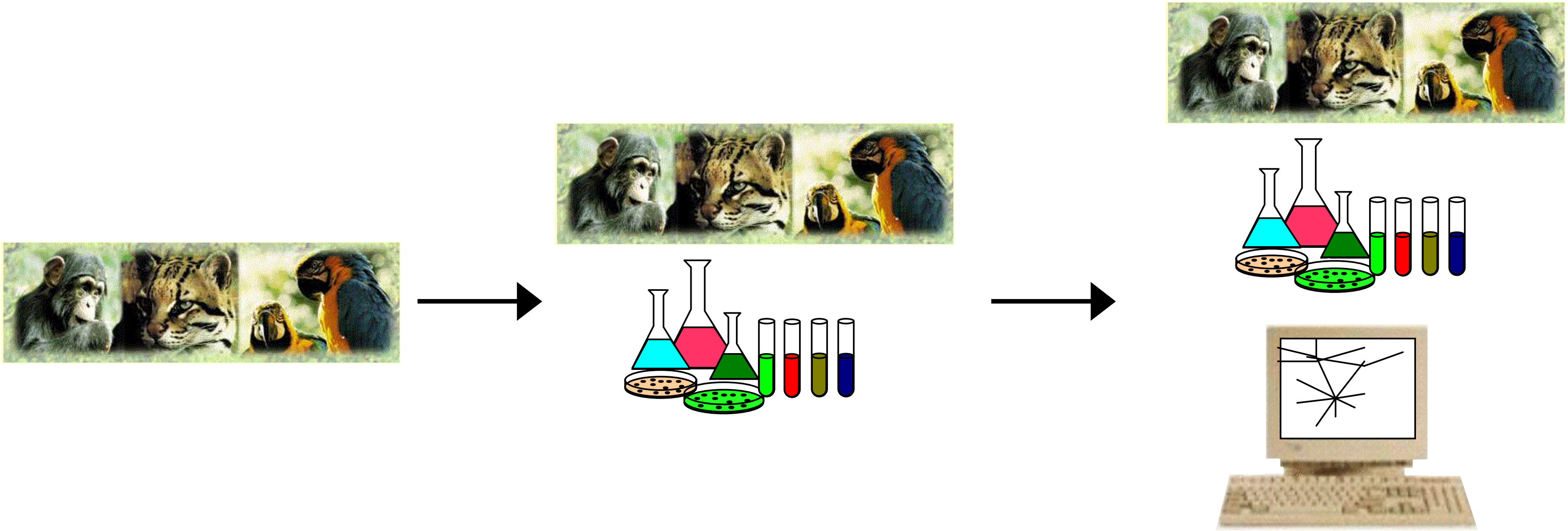
Ideker, T., V. Thorsson, et al. (2001). "Integrated genomic and proteomic analyses of a systematically perturbed metabolic network." Science **292**: 929-934.

Biología de Sistemas y Terapia/diagnóstico



Chuang, H.Y., Lee, E., Liu, Y.T., Lee, D. and Ideker, T. (2007) Network-based classification of breast cancer metastasis. *Mol Syst Biol.*, **3**, 140.

Aproximaciones al estudio de los sistemas vivos



In vivo

Características del
sistema

In vivo + in vitro

Características de los
componentes (moléculas)

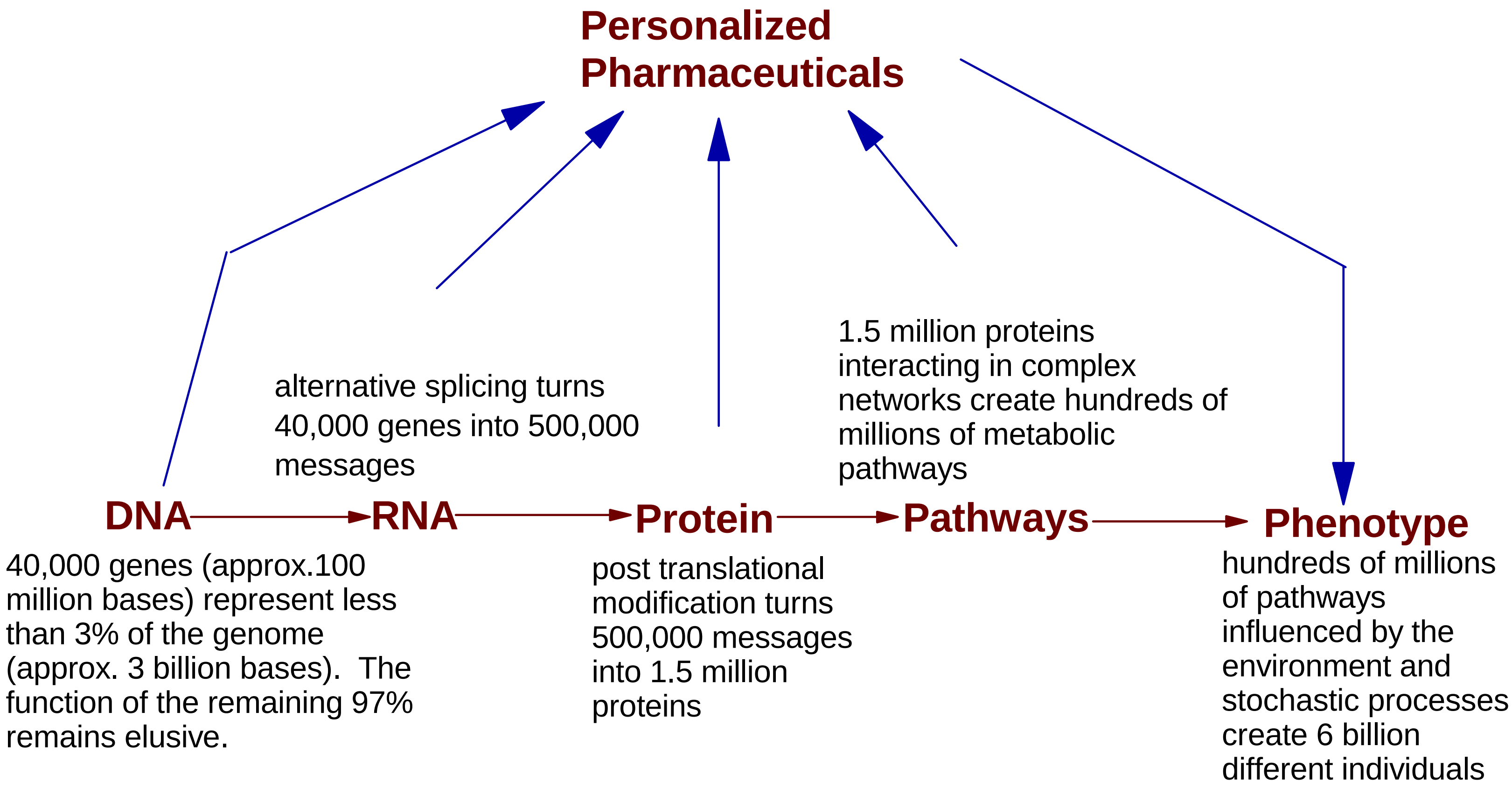
In vivo + in vitro + in silico

Características del
sistema desde el pto.
vista de los componentes y sus
relaciones
Propiedades emergentes, ...

Interés “Comercial” de la Biología de Sistemas

New Sources of Value Creation

Clinical drug delivery requires managing enormous complexity. Academic medical Institutions are beginning to differentiate themselves on the power of their information technology platforms. IT Platform intellectual property is likely to be more valuable than content (gene sequences, metabolic pathways, protein structures, etc.)



Personalized Pharmaceuticals

alternative splicing turns 40,000 genes into 500,000 messages

1.5 million proteins interacting in complex networks create hundreds of millions of metabolic pathways

DNA → **RNA** → **Protein** → **Pathways** → **Phenotype**

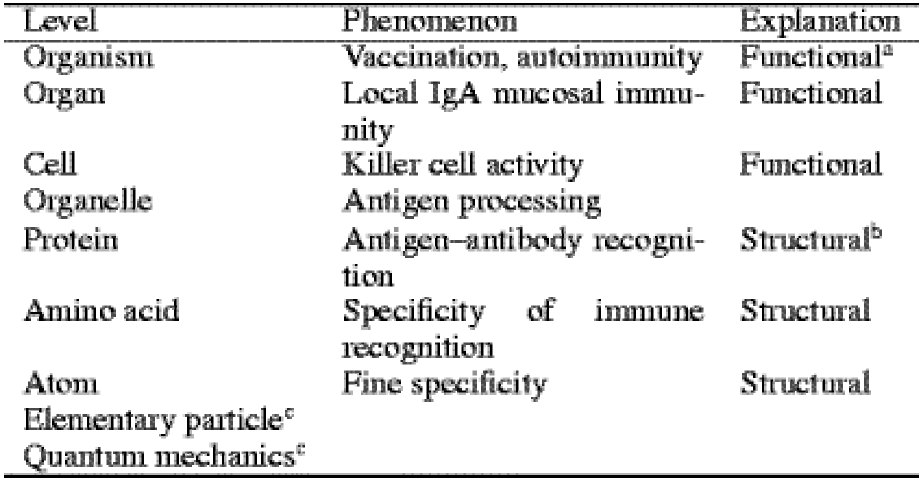
40,000 genes (approx.100 million bases) represent less than 3% of the genome (approx. 3 billion bases). The function of the remaining 97% remains elusive.

post translational modification turns 500,000 messages into 1.5 million proteins

hundreds of millions of pathways influenced by the environment and stochastic processes create 6 billion different individuals

Source: IBM corporation

Biología Molecular y Biología de Sistemas



Level	Phenomenon	Explanation
Organism	Vaccination, autoimmunity	Functional[a]
Organ	Local IgA mucosal immunity	Functional
Cell	Killer cell activity	Functional
Organelle	Antigen processing	
Protein	Antigen–antibody recognition	Structural[b]
Amino acid	Specificity of immune recognition	Structural
Atom	Fine specificity	Structural
Elementary particle[c]		
Quantum mechanics[c]		

Medicina

reactiva
generalista

↓

predictiva
preventiva
personalizada

“Biology is an autonomous discipline that requires its own vocabulary and concepts, that are not found in chemistry and physics”.
M.H. V. Van Regenmortel

- Van Regenmortel, MHV. (2001). Pitfalls of reductionism in the design of peptide-based vaccines. Vacine. **19,** 2369-2374
- Hood, L., Heath, J.R., Phelps, M.E. and Lin, B. (2004) Systems biology and new technologies enable predictive and preventative medicine. *Science*, **306**, 640-643.

Redes Biológicas y Biología de Sistemas

21-29 Abril 2008

L. 21 **10:00-12:15** Introducción Descripción del curso *F. Pazos (CNB)* **12:30-14:00** Teoría de Grafos (I) *C. Aguirre (UAM)*	**M. 22** **10:00-12:00** Redes de interacciones entre proteínas *D. Juan (CNIO)* **12:00-14:00** Manejo y visualización de redes *D. Juan (CNIO)*	**Mi. 23** **10:00-12:15** Redes de regulación génica *I. Cases (CNIO)* **12:30-14:00** Teoría de Grafos (II) *C. Aguirre (UAM)*	J. 24	**V. 25** **10:00-12:00** Bacterias para el medio ambiente: de la Bioremediación a la Biología Sintética *V. Lorenzo (CNB)* **12:30-14:00** Modelado cuantitativo de circuitos biológicos *S. Moreno (CNB)*	S. 26	D. 27
L. 28 **10:00-12:00** Redes metabólicas *F. Pazos (CNB)* **12:00-14:00** Otras redes en Biología Molecular *F. Pazos (CNB)*	**M. 29** **10:00-14:00** Ejercicios prácticos y trabajos de evaluación *F. Pazos (CNB)*	Las clases serán en el Aula de Informática 0 de la Facultad de Ciencias de la UAM. Excepto el primer dia (L.21) que serán en la sala de seminarios "B" del CNB (Hall). **http://pdg.cnb.csic.es/pazos/cursos/bionet_UAM**				