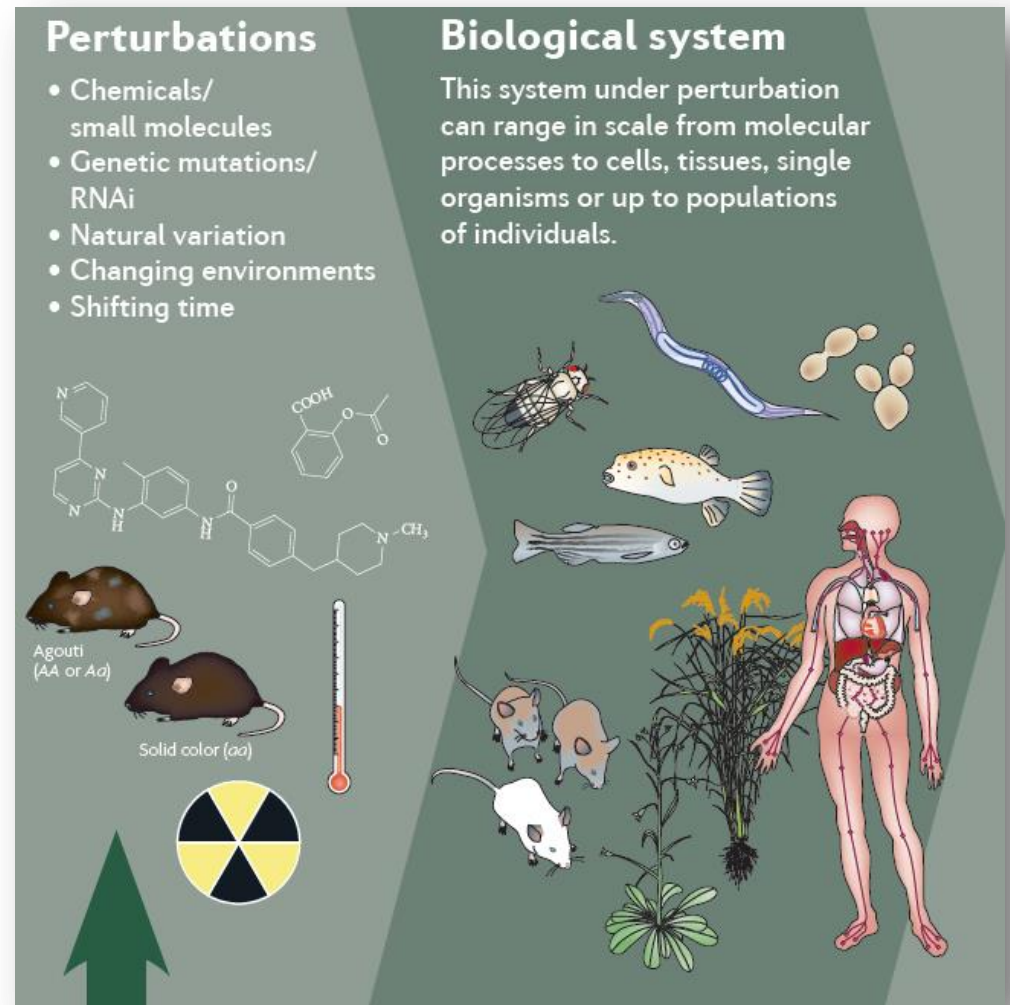


Redes. Clusters. Biologia.

Contexto: Biología de Sistemas Integrativa

Estudio de sistemas biológicos mediante

- Análisis de respuestas globales ante perturbaciones sistemáticas
- Integración de datos en modelos predictivos



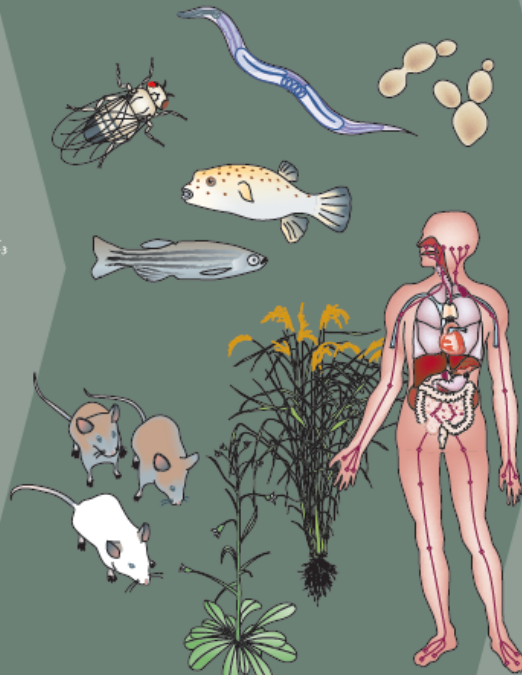
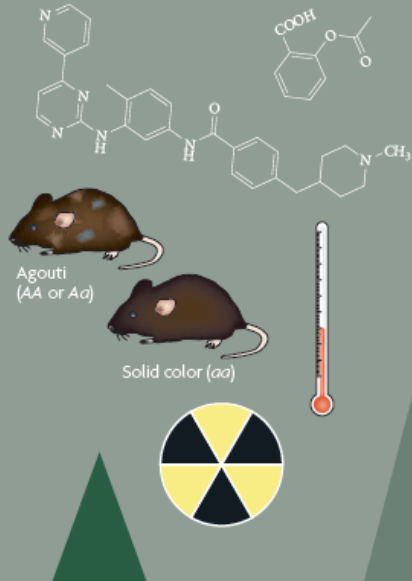
Biología de Sistemas Integrativa

Perturbations

- Chemicals/
small molecules
- Genetic mutations/
RNAi
- Natural variation
- Changing environments
- Shifting time

Biological system

This system under perturbation can range in scale from molecular processes to cells, tissues, single organisms or up to populations of individuals.



Molecular state measurements (nodes)

	Large-scale dataset types	Technologies
Genome	Whole-genome DNA sequences, SNPs and CNVs	DNA sequencing, genotyping microarrays
Epigenome	Chromatin modifications and structure	ChIP-seq, methyl-seq, DHS-chip
Transcriptome	Transcript abundances, translation rate and microRNAs	DNA microarrays, RNA-seq, CAGE, GRO-seq, ribosome profiling
Proteome	Protein abundances and modifications	NMR, mass spectrometry, multiparameter FACS
Metabolome	Metabolite profiling	Mass spectrometry, liquid chromatography

Molecular interaction measurements (edges)

	Data types	Technologies
Physical	Protein-protein	Immuno-precipitation (IP), co-affinity purification, yeast two-hybrid, protein arrays, kinase-substrate measurements
	Protein-DNA, protein-RNA	Genome-wide chromatin immuno-precipitation (ChIP), DNA binding arrays
	Protein-small molecule, reaction fluxes	Isotope labeling, mass spectrometry
Genetic and functional	Synthetic lethality, epistasis	Synthetic genetic arrays (SGA) combinatorial RNAi, population genetics
	Cause-effect relationships	Genetic perturbation (gene knockout, RNAi) followed by phenotyping (microarrays, cellular imaging); <i>trans</i> eQTLs (expression quantitative trait loci)

Biología de Sistemas Integrativa

Molecular state measurements (nodes)

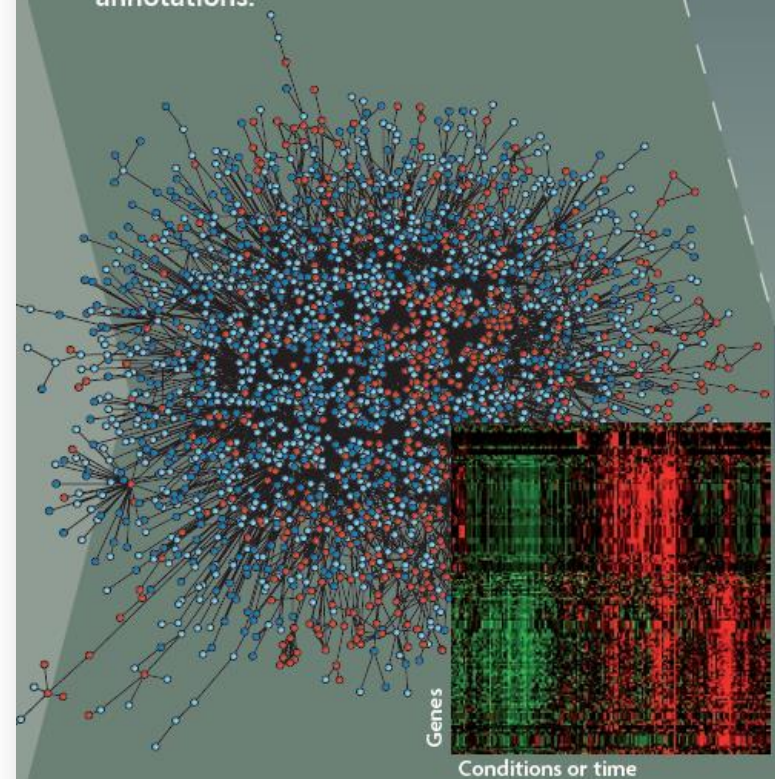
	Large-scale dataset types	Technologies
Genome	Whole-genome DNA sequences, SNPs and CNVs	DNA sequencing, genotyping microarrays
Epigenome	Chromatin modifications and structure	ChIP-seq, methyl-seq, DHS-chip
Transcriptome	Transcript abundances, translation rate and microRNAs	DNA microarrays, RNA-seq, CAGE, GRO-seq, ribosome profiling
Proteome	Protein abundances and modifications	NMR, mass spectrometry, multiparameter FACS
Metabolome	Metabolite profiling	Mass spectrometry, liquid chromatography

Molecular interaction measurements (edges)

	Data types	Technologies
Physical	Protein–protein	Immuno-precipitation (IP), co-affinity purification, yeast two-hybrid, protein arrays, kinase-substrate measurements
	Protein–DNA, protein–RNA	Genome-wide chromatin immuno-precipitation (ChIP), DNA binding arrays
	Protein–small molecule, reaction fluxes	Isotope labeling, mass spectrometry
Genetic and functional	Synthetic lethality, epistasis	Synthetic genetic arrays (SGA) combinatorial RNAi, population genetics
	Cause–effect relationships	Genetic perturbation (gene knockout, RNAi) followed by phenotyping (microarrays, cellular imaging); <i>trans</i> eQTLs (expression quantitative trait loci)

Molecular databases

New measurements are stored alongside existing data, including functional annotations.



	Annotations	States	Interactions
Selected molecular databases	GO , KEGG , REACTOME , NetPath , UCSC Genome Browser	Genbank , GEO , ArrayExpress , Proteome Commons	BioGRID , HPRD , IntAct , TRANSFAC , STRING , iHOP , functionalnet.org

Conocimiento Biológico

	Interaction type(s)	Detection methodologies	Databases
Physical	Protein-protein	Yeast-2-hybrid (Y2H) ¹⁻³ , co-immuno-precipitation (Co-IP) ⁴ , mass spectroscopy ^{5,6} , affinity purification coupled with mass spectroscopy (AP/MS) ^{7,8}	BioGRID ⁹ , IntAct ¹⁰ , APID ¹¹ , STRING ¹² , MINT ¹³ , DIP ¹⁴ , HPRD ¹⁵ , MIPS-MPPI ¹⁶ , Netpath ¹⁷ , DroiD ¹⁸
	Protein-DNA (e.g., regulatory networks)	Yeast-1-hybrid (Y1H) ¹⁹ , chromatin immuno-precipitation based methods (CHIP-CHIP) ²⁰ , DNA-footprinting ²¹	TRANSFAC ²² , UniProbe ²³ , DroiD ¹⁸ , BioGRID ⁹ , TcoF-DB ²⁴ , BIPA ²⁵ , hPDI ²⁶ , EDGEdb ²⁷ , NPIDB ^{28,29}
	Protein-RNA	RNA electro-mobility shift (RNA-EMSA) ³⁰ , RNA-pull down ³¹	PRID ³² , BIPA ²⁵ NPInter ³³ , RBPDB ³⁴ , PRIDB ³⁵ , StarBase ³⁶
	Metabolic (e.g., enzyme-substrate, ligand-receptor)	Mass spectroscopy based selective reaction monitoring (SRM) ^{6,37} , NMR ³⁸ , affinity purification ⁸ , co-IP ³ , fluorescence spectroscopy ³⁹	Reactome ⁴⁰ , KEGG ⁴¹ , BioCyc and MetaCyc ⁴² , HMDB ^{43,44} , EcoCyc ⁴⁵ , HumanCyc ⁴⁶ , ConsensusPathDB ⁴⁷
	Protein/gene-compound (e.g., drug-target, chemical-protein)	Chemical structure ^{48,49} , forward or reverse chemo-genomic/proteomic profiling ⁵⁰⁻⁵² , in silico predictions ⁵³	SuperTarget ⁵⁴ , Matador ⁵⁴ , DrugBank ⁵⁵ , ChemProt ⁵⁶ , STITCH ⁵⁷ , AffinDB ⁵⁸ , MatrixDB ⁵⁹ , PSMDB ⁶⁰ , PDB-Ligand ⁶¹ , ChEMBL ⁶² , ConsensusPathDB ⁴⁷
Functional	Genetic (gene-gene)	Synthetic genetic array (SGA) ⁶³ , Epistatic Miniarray Profiling (E-MAP) ⁶⁴ , co-expression profiling ^{65,66}	BioGRID ⁹ , DRYGIN ⁶⁷ , CYGD ⁶⁸ , DroiD ¹⁸ , ConsensusPathDB ⁴⁷
	Gene-Disease	Literature curation, clinical and sequence information	OMIM ⁶⁹ , HuDiNe ⁷⁰ , Disasome ⁷¹
Omics data type		Detection methodologies	Databases
Transcriptomics		Microarray ⁷² , RNASeq ⁷³	GEO ⁷⁴ , SMD ⁷⁵ , TCGA ⁷⁶ , GXD ⁷⁷ , ONCOMINE ⁷⁸ , ArrayExpress ⁷⁹
RNAi (phenomics)		RNAi interference assay ⁸⁰	RNAiDB ⁸¹ , GenomeRNAi ⁸² , siRecords ⁸³
Epigenomics		Methylation profiling ⁸⁴	DAnCER ⁸⁵ , DiseaseMeth ⁸⁶ , PubMeth ⁸⁷ , MethDB ⁸⁸ , MethCancerDB ⁸⁹ , MethyCancer ⁹⁰
Mutation / SNP		SNP Array ⁹¹ , genome sequencing ⁹²	TCGA ⁷⁶ , dbSNP ⁹³ , dbQSNP ⁹⁴ , GWAS Central ⁹⁵ , OMIM ⁶⁹
Proteomics		CHIP ^{11,2,96} , Mass Spectrometry ^{5,6}	PDB ⁹⁷ , ExPASy ⁹⁸ , InterPro ⁹⁹ , World-2DPage ¹⁰⁰ , JASPAR ¹⁰¹
Phosphorylation profile		Mass Spectrometry ⁵ , literature curation	PhosphoGRID ¹⁰² , PhosphoELM ¹⁰³ , PHOSIDA ¹⁰⁴

Portal de Conocimiento Biológico

The screenshot shows the Pathguide website interface. At the top, there is a search bar with the text "Buscar" and a navigation menu with "Home", "BioPAX", and "cBio" buttons. The main heading is "Pathguide the pathway resource list".

Navigation

- Protein-Protein Interactions
- Metabolic Pathways
- Signaling Pathways
- Pathway Diagrams
- Transcription Factors / Gene Regulatory Networks
- Protein-Compound Interactions
- Genetic Interaction Networks
- Protein Sequence Focused
- Other

Search

Organisms: All

Availability: All

Standards: All

Buttons: Reset, Search

Analysis

- Statistics
- Database Interactions

Contact

Comments, Questions, Suggestions are

Complete Listing of All Pathguide Resources

Pathguide contains information about **547** biological pathway related resources and molecular interaction related resources. Click on a link to go to the resource home page or 'Details' for a description page. Databases that are free and those supporting BioPAX, CellML, PSI-MI or SBML standards are respectively indicated.

If you know of a pathway resource that is not listed here, or have other questions or comments, please [send us an e-mail](#).

Protein-Protein Interactions

Database Name (Order: alphabetically | by [web popularity](#))

Database Name	Full Record	Availability
2P2Idb - The Protein-Protein Interaction Inhibition Database	Details	Free
3D-Interologs - 3D-Interologs	Details	Free
3DID - 3D interacting domains	Details	Free
ADAN - Prediction of protein-protein interaction of modular domains	Details	X
AHD2.0 - Arabidopsis Hormone Database 2.0	Details	Free
AllFuse - Functional Associations of Proteins in Complete Genomes	Details	X
aMAZE - Protein Function and Biochemical Pathways Project	Details	X
ANAP - Arabidopsis Network Analysis Pipeline	Details	Free
AnimalTFDB - Animal Transcription Factor Database	Details	Free
AntiJen - AntiJen a Kinetic, Thermodynamic and Cellular Database	Details	Free
APID - Agile Protein Interaction DataAnalyzer	Details	Free
AS-ALPS - Alternative Splicing - induced ALteration of Protein Structure	Details	Free
ASD - Allosteric Database	Details	Free
ASEdb - Alanine Scanning Energetics Database	Details	Free
ASPD - Artificial Selected Proteins/Peptides Database	Details	Free
ATDB - Animal Toxin Database	Details	Free
AtPID - Arabidopsis thaliana Protein Interactome Database	Details	Free

News

- Major new update of Pathguide August 2010. We now have information about ~550 resources!
- Visual navigation added May 2010. Click the 'Database Interactions' link on the



Volume 48, Issue D1
08 January 2020

[Cover image](#)

ISSN 0305-1048
EISSN 1362-4962

Editorial

Major Multi-Database
Resources

Nucleotide Sequence
Databases

RNA sequence databases

Protein sequence databases

Structure Databases

Genomics Databases (non-
vertebrate)

Metabolic and Signaling
Pathways

Human and other Vertebrate
Genomes

Human Genes and Diseases

Corrigendum

Retraction

Volume 48, Issue D1, 08 January 2020

Page 1 of 2

1 2 Next

EDITORIAL

The 27th annual Nucleic Acids Research database issue and
molecular biology database collection 

[Daniel J Rigden](#), [Xosé M Fernández](#)

Nucleic Acids Research, Volume 48, Issue D1, 08 January 2020, Pages D1–D8, <https://doi.org/10.1093/nar/gkz1161>

[Abstract ▾](#) [View article](#)


MAJOR MULTI-DATABASE RESOURCES

Database resources of the National Center for Biotechnology
Information 

[Eric W Sayers](#), [Jeff Beck](#), [J Rodney Brister](#), [Evan E Bolton](#), [Kathi Canese](#) ...

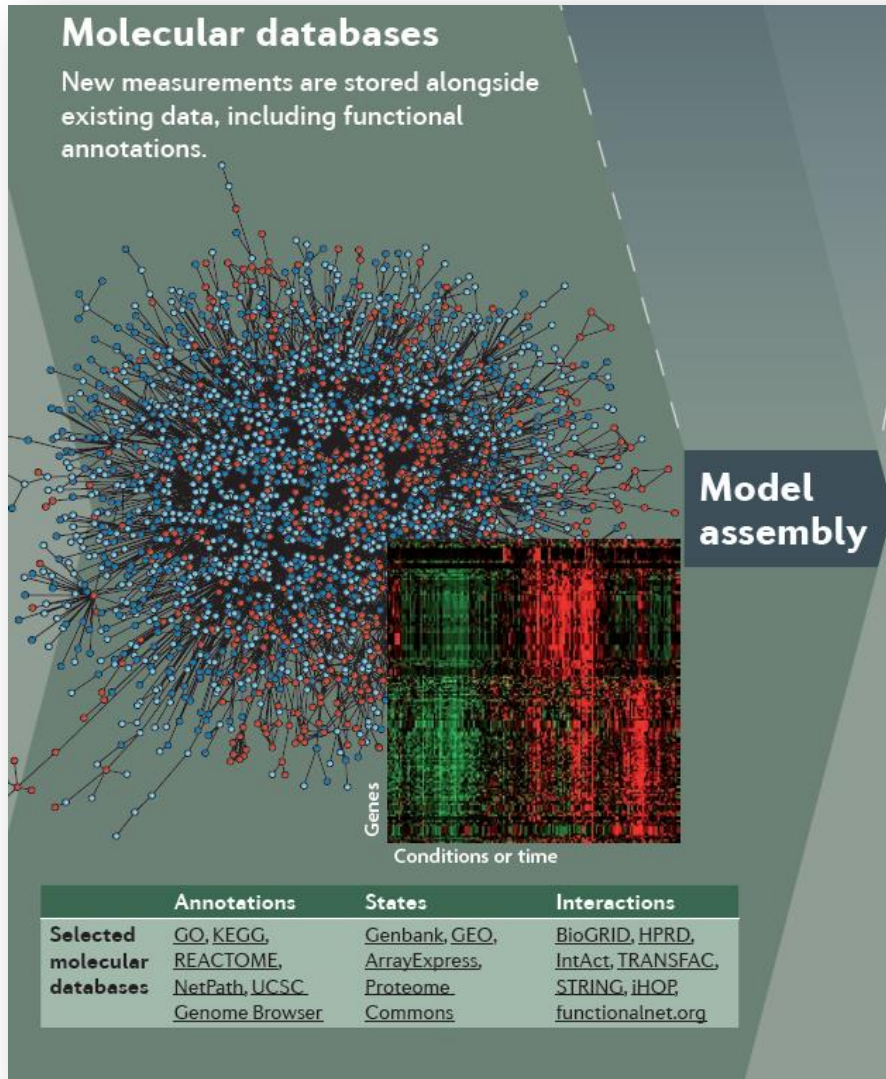
Nucleic Acids Research, Volume 48, Issue D1, 08 January 2020, Pages D9–D16, <https://doi.org/10.1093/nar/gkz899>

[Abstract ▾](#) [View article](#)

The European Bioinformatics Institute in 2020: building a global
infrastructure of interconnected data resources for the life sciences 

[Charles E Cook](#), [Oana Stroe](#), [Guy Cochrane](#), [Ewan Birney](#), [Rolf Apweiler](#)

Biología de Sistemas Integrativa



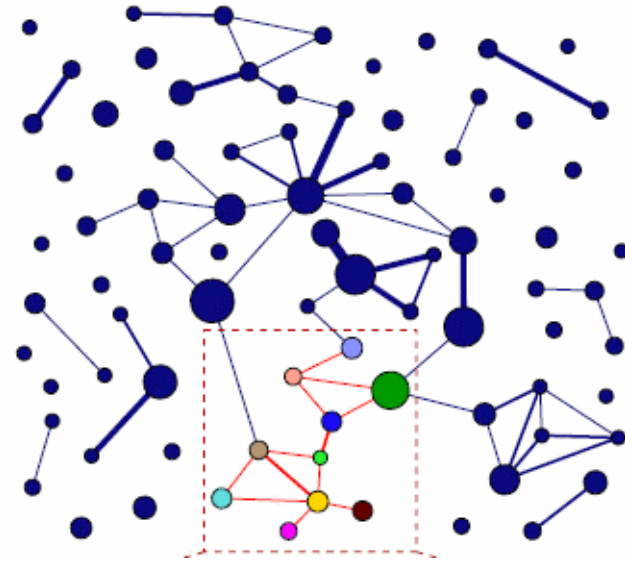
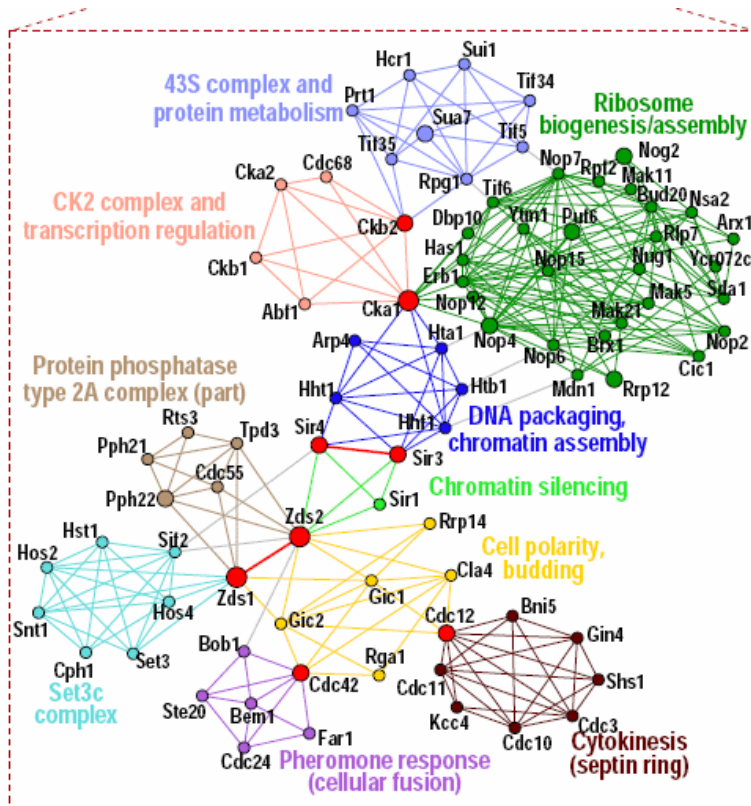
Análisis estadístico

- Clustering de datos
- Overlaps y correlaciones en datasets
- Enriquecimiento funcional

Modelado

- Inferencia estadística de interacciones a partir de datos
- Modelado de sistemas dinámicos y simulaciones biofísicas

Enriquecimiento funcional



- **MsigDB**
- **Gene Ontology**
- **DAVID**



Molecular Signatures DataBase



- Conjunto de **bolsas de genes** armadas bajo algún criterio relevante desde el punto de vista biológico.
- Conocimiento biológico se encuentra embebido en la composición de cada conjunto

The MSigDB gene sets are divided into 8 major collections:

H **hallmark gene sets** are coherently expressed signatures derived by aggregating many MSigDB gene sets to represent well-defined biological states or processes.

C1 **positional gene sets** for each human chromosome and cytogenetic band.

C2 **curated gene sets** from online pathway databases, publications in PubMed, and knowledge of domain experts.

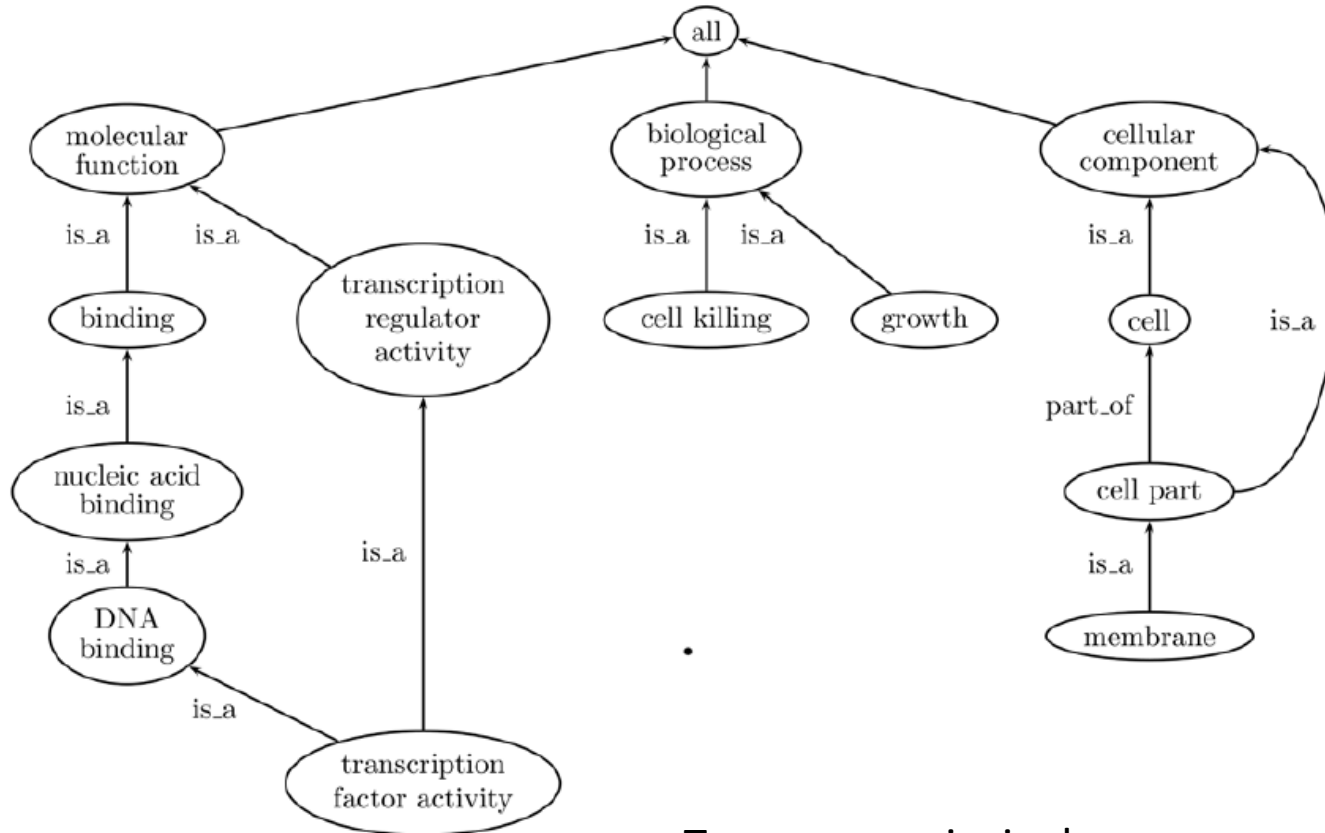
C3 **motif gene sets** based on conserved cis-regulatory motifs from a comparative analysis of the human, mouse, rat, and dog genomes.

C4 **computational gene sets** defined by mining large collections of cancer-oriented microarray data.

C5 **GO gene sets** consist of genes annotated by the same GO terms.

C6 **oncogenic signatures** defined directly from microarray gene expression data from cancer gene perturbations.

C7 **immunologic signatures** defined directly from microarray gene expression data from immunologic studies.

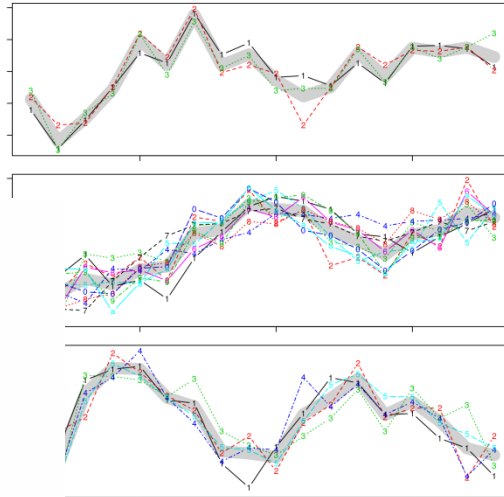
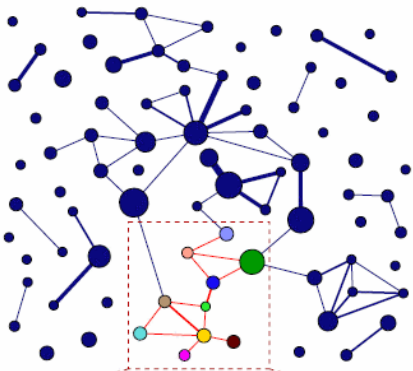


- Tres ramas principales
 - **Molecular Function** (Qué hace?)
 - **Cellular Component** (Dónde lo hace?)
 - **Biological Process** (Por qué lo hace?)

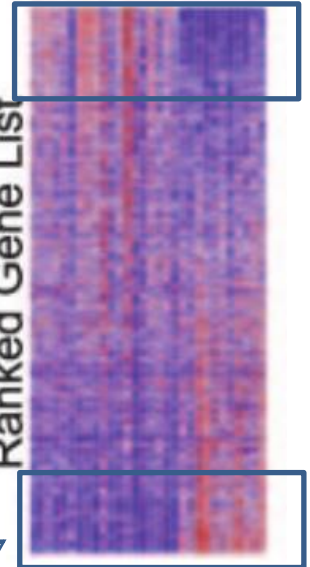
Analisis de sobre-representacion

Algún concepto de **GO** o conjunto de **MSigDB** esta sobrerrepresentado en los **genes diferencialmente expresados** que encuentro en mi experimento?

Algún concepto de **GO** o conjunto de **MSigDB** esta sobrerrepresentado en alguna comunidad o cluster detectado en mis datos?



A Phenotype Classes
A B



gen1
gen2
gen3
gen4
gen5
gen6
gen7
gen8
gen9
gen10

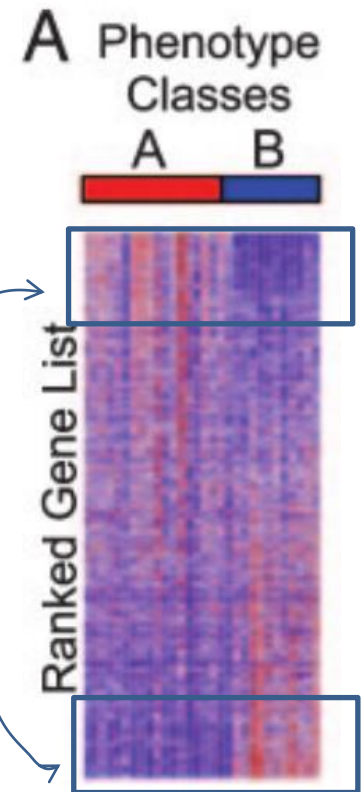
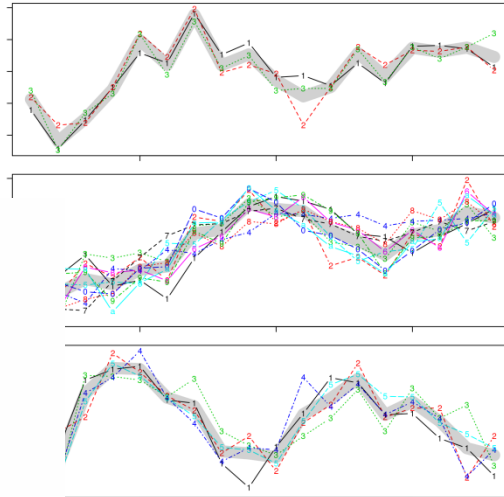
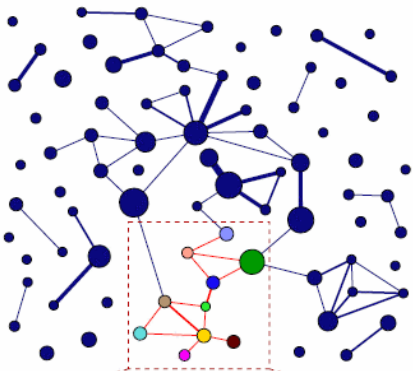
Apoptosis

Respuesta Inmune

Analisis de sobre-representacion

Algún concepto de **GO** o conjunto de **MSigDB** esta sobrerrepresentado en los **genes diferencialmente expresados** que encuentro en mi experimento?

Algún concepto de **GO** o conjunto de **MSigDB** esta sobrerrepresentado en alguna comunidad o cluster detectado en mis datos?

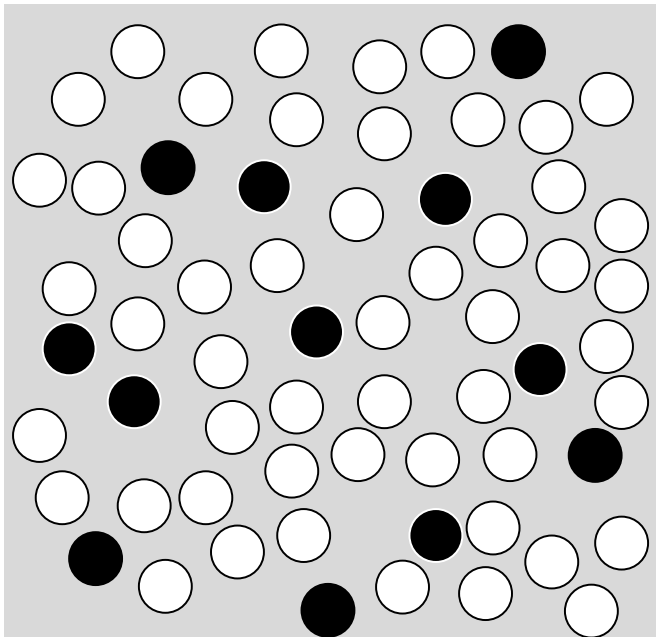
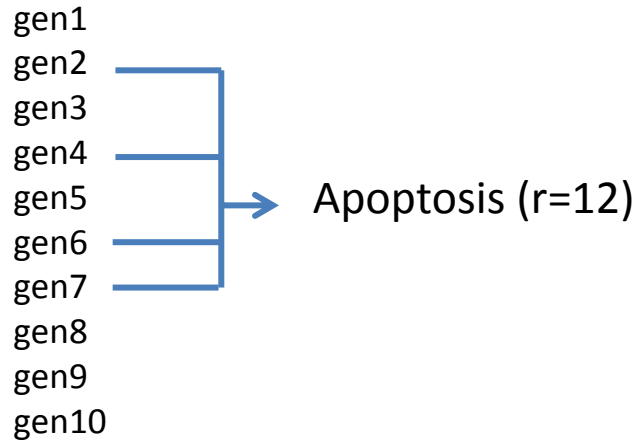


Existe un **vinculo** entre pertenecer a una **cjto de genes identificado en mi experimento** y estar anotado dentro de un **tema biologico**?

Test de Fisher Exacto

Que tan probable es observar un dado numero de hits en la lista solo por azar

ORA - Test de Fisher



N (100) genes examinados

Test de Fisher Exacto

Que tan probable es observar un dado numero de hits en la lista solo por azar

O, alternativamente...

existe una relacion en el sentido estadistico entre pertenecer a la lista de genes detectados y pertenecer a la categoria apoptosis?

ORA - Test de Fisher

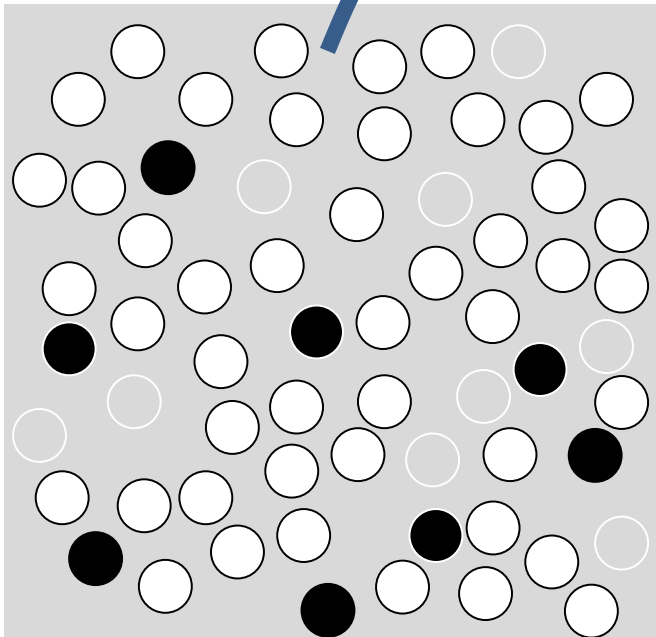
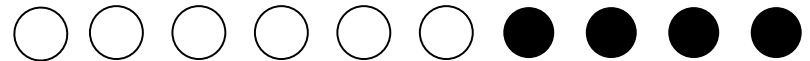
gen1
gen2
gen3
gen4
gen5
gen6
gen7
gen8
gen9
gen10



Apoptosis (r=12)

Test de Fisher Exacto

Que tan probable es observar un dado numero de hits en la lista solo por azar

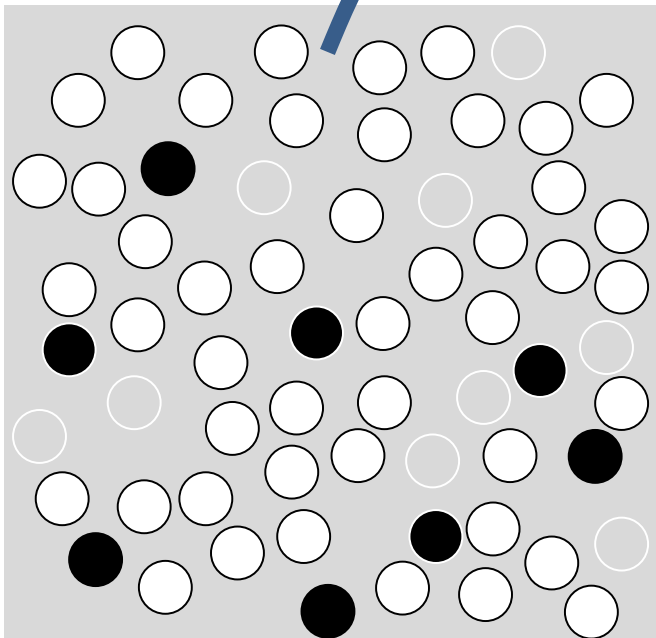


N (100) genes examinados

ORA - Test de Fisher

gen1
gen2
gen3
gen4
gen5
gen6
gen7
gen8
gen9
gen10

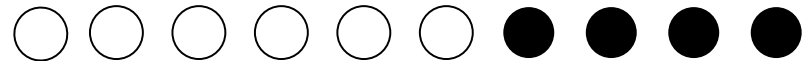
Apoptosis (r=12)



N (100) genes examinados

Test de Fisher Exacto

Que tan probable es observar un dado numero de hits en la lista solo por azar



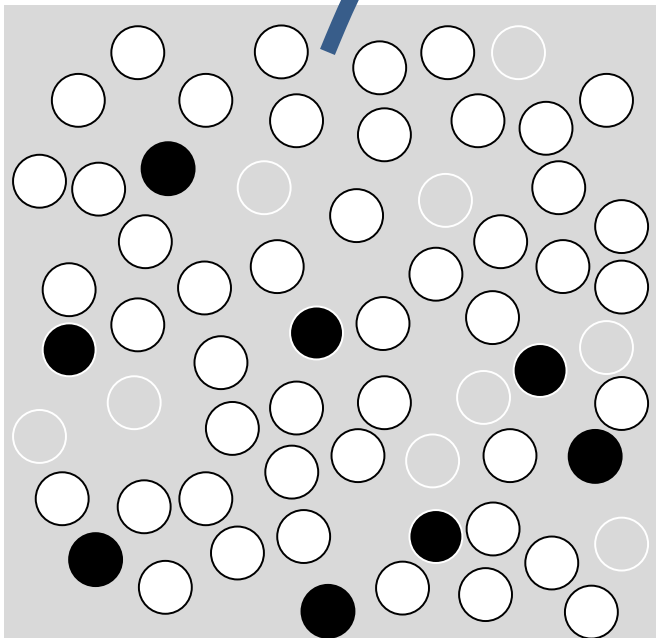
k (10) genes en la comunidad/cluster

m (4) genes del cluster en la categoria biológica

	categoria bio	categoria bio	
cluster	4	6	10
<u>cluster</u>	8	82	90
	12	88	100
	categoria bio	categoria bio	
cluster	m	k-m	k
<u>cluster</u>	r-m	N-k+m-r	N-k
	r	N-r	N

ORA - Test de Fisher

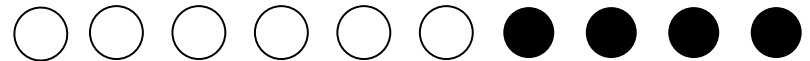
gen1
gen2
gen3
gen4
gen5
gen6
gen7
gen8
gen9
gen10



N (100) genes examinados

Test de Fisher Exacto

Que tan probable es observar un dado numero de hits en la lista solo por azar



k (10) genes en la comunidad/cluster

m (4) genes del cluster en la categoria biológica

	categoria bio	<u>categoria bio</u>	
cluster	m	k-m	k
<u>cluster</u>	r-m	N-k+m-r	N-k
	r	N-r	N

Cual es la probabilidad de ver esta tabla o alguna **mas extrema** si saco las bolas al azar? (i.e. no hay coneccion alguna entre pertenencia al cluster y a la categoria biológica?)

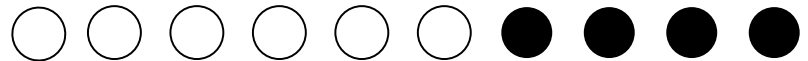
ORA - Test de Fisher

gen1
gen2
gen3
gen4
gen5
gen6
gen7
gen8
gen9
gen10

Apoptosis (r=12)

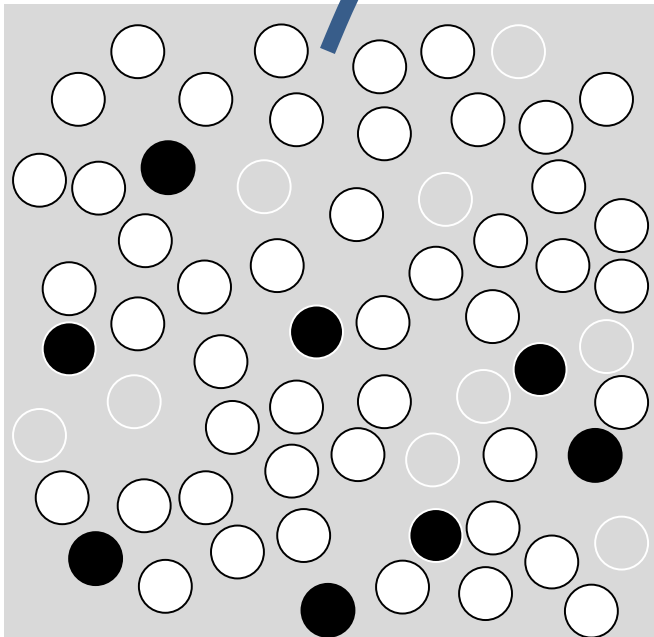
Test de Fisher Exacto

Que tan probable es observar un dado numero de hits en la lista solo por azar



k (10) genes en la comunidad/cluster

m (4) genes del cluster en la categoria biológica



N (100) genes examinados

	categoria bio	categoria bio	
cluster	m	k-m	k
cluster	r-m	N-k+m-r	N-k
	r	N-r	N

Cual es la probabilidad de ver esta tabla?

de maneras de elegir las bolas negras

$$p = \frac{\binom{r}{m} \binom{N-r}{k-m}}{\binom{N}{r}}$$

de maneras de elegir las bolas blancas

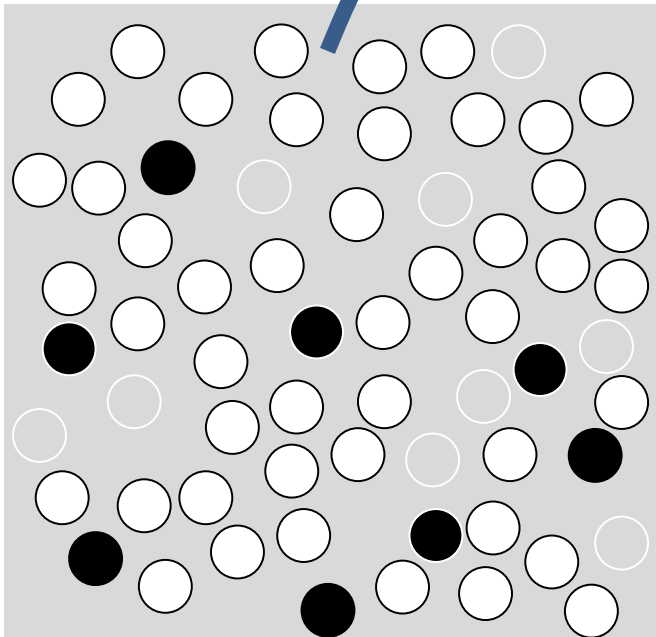
maneras de elegir r elementos de un total de N

ORA - Test de Fisher

gen1
gen2
gen3
gen4
gen5
gen6
gen7
gen8
gen9
gen10



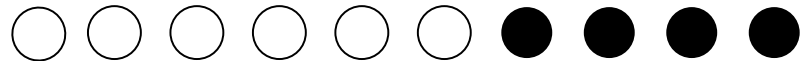
Apoptosis (r=12)



N (100) genes examinados

Test de Fisher Exacto

Que tan probable es observar un dado numero de hits en la lista solo por azar



k (10) genes en la comunidad/cluster

m (4) genes del cluster en la categoria biológica

	categoria bio	categoria bio	
cluster	m	k-m	k
cluster	r-m	N-k+m-r	N-k
	r	N-r	N

Cual es la probabilidad de ver esta tabla u **otra más extrema?**

$$p_{valor} = \sum_{m'=m}^r \frac{\binom{r}{m'} \binom{N-r}{k-m'}}{\binom{N}{r}}$$

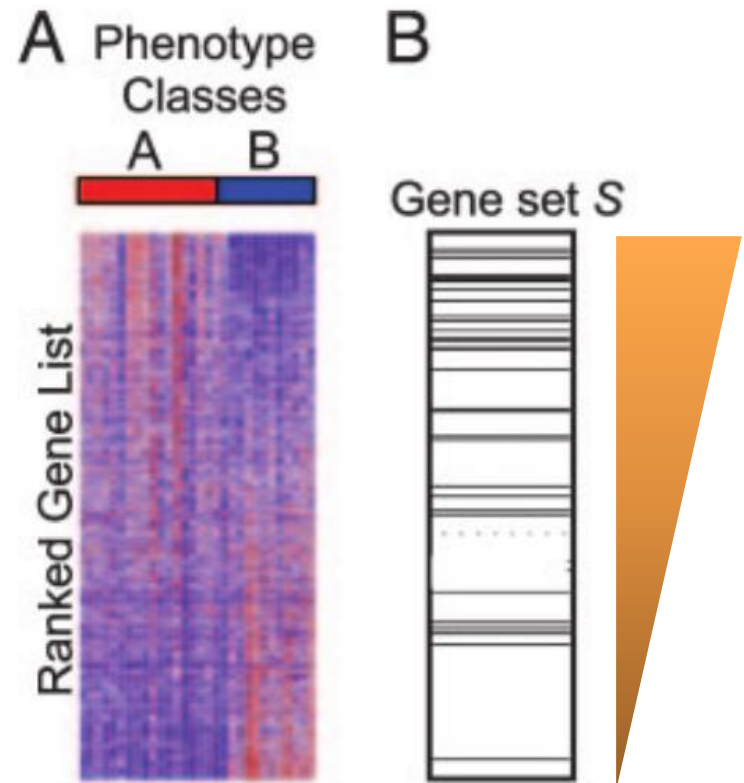
Es chica esta cantidad?

$$p_v < p_{umbral}$$

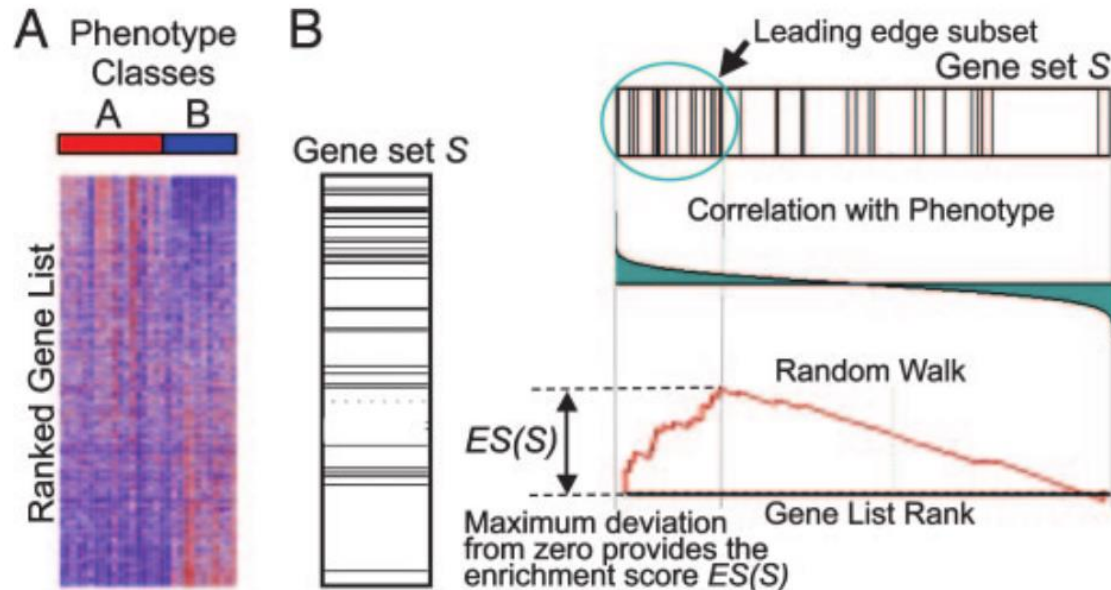
Gene Set Enrichment Analysis

- Mootha et al., Nature Genetics 34, 267–273 (2003)
- Subramanian et al., PNAS 102(43). 15545–15550 (2005).

- Estimé la expresión de un conjunto de genes y los rankeo según la correlación que presenten con algún fenotipo de interés
- Con estos datos, me pregunto si el fenotipo estudiado se relaciona con una dada categoría biológica de interés (CBI) previamente asociada un dado cjto de genes (por ej una categoría GO)
- Si hay una asociación positiva espero que los genes de la CBI *estén rankeados arriba* en el ranking...como cuantifico esto?

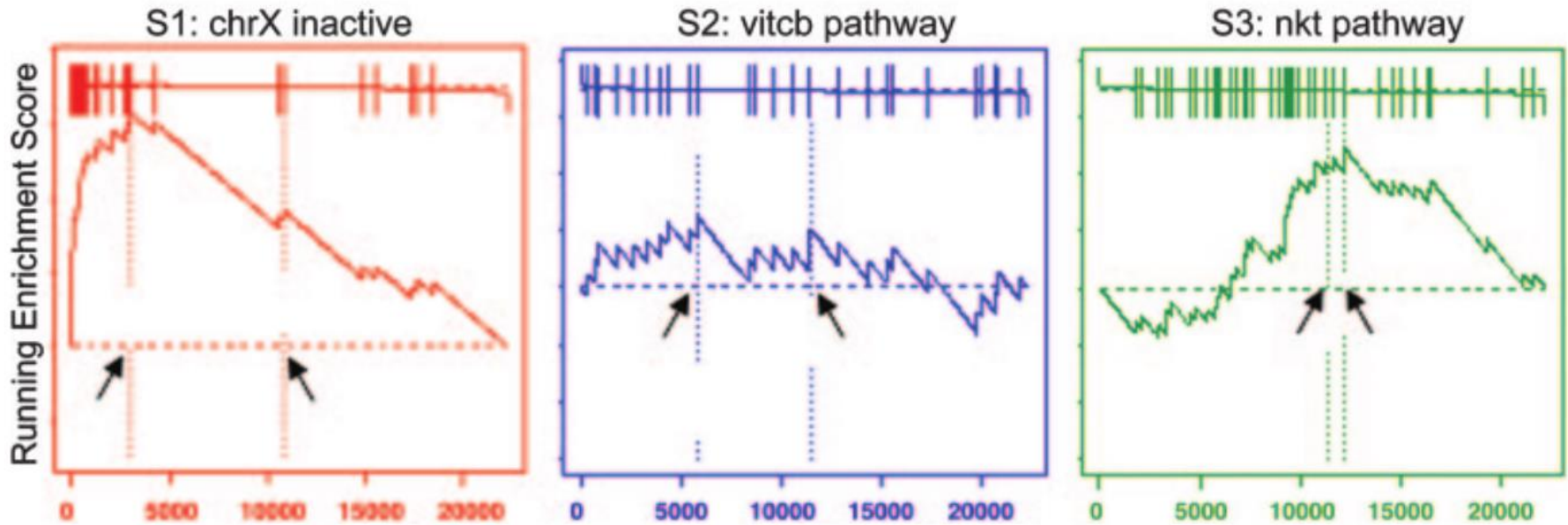


GSEA



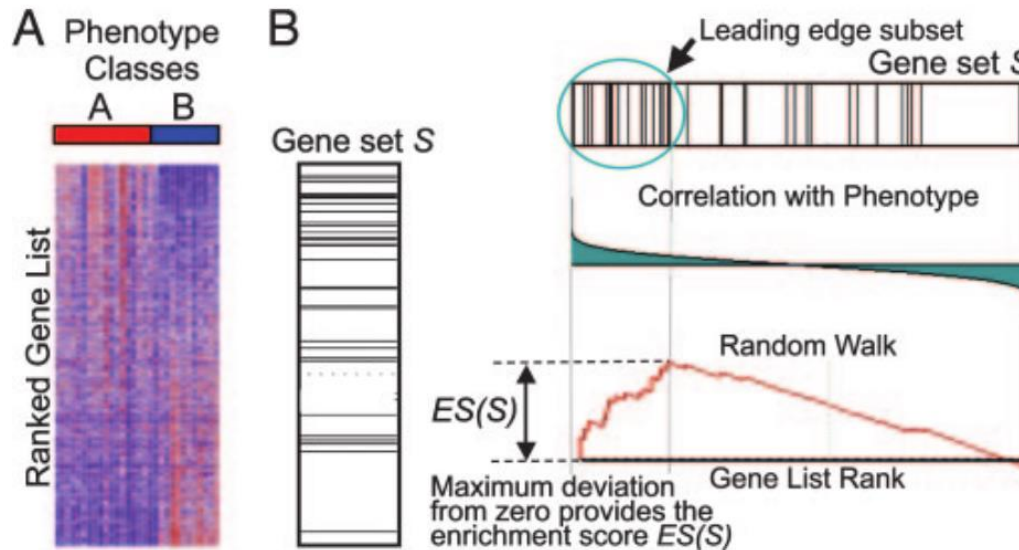
- Computo suma acumulativa sobre el ranking: Sumo un poroto cuando encuentro un gen del CBI, sino resto uno (la magnitud del incremento depende de la correlacion del gen con el fenotipo)
- La maxima desviacion respecto a cero es el Enrichment Score (ES) asociado al CBI

GSEA



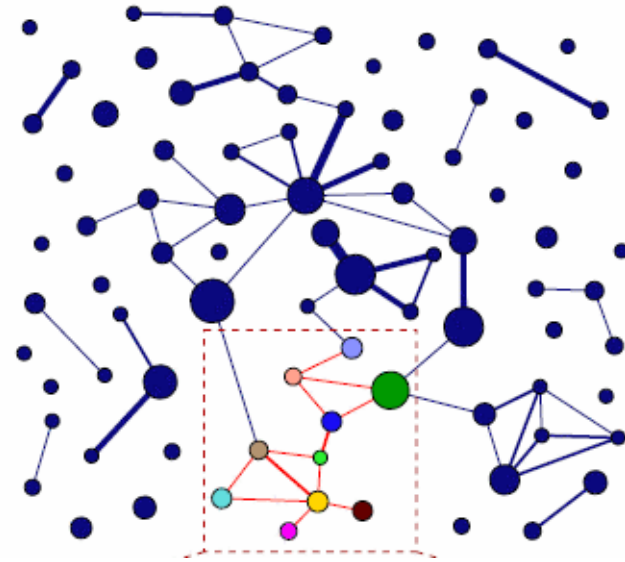
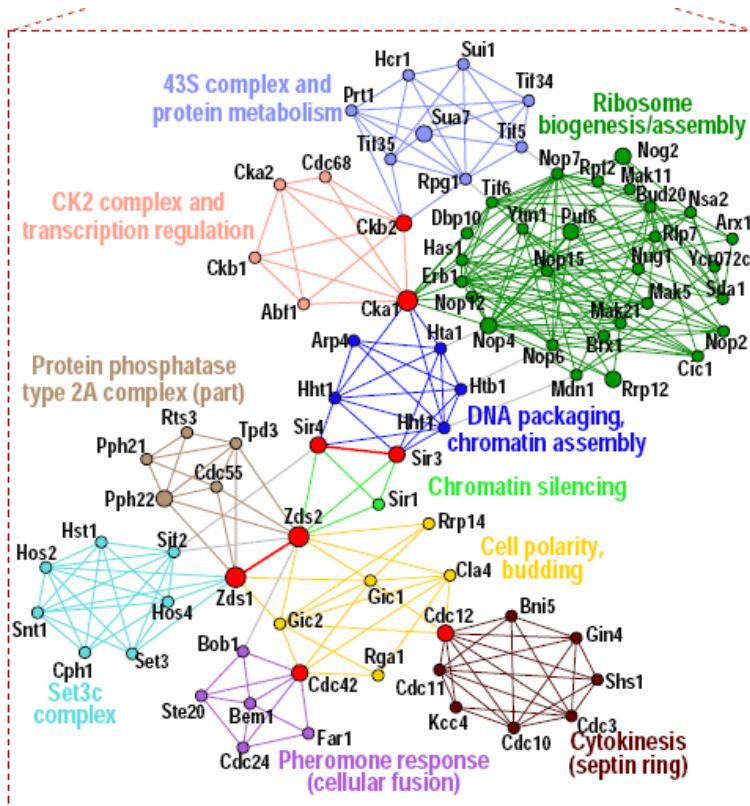
Subramanian et al., PNAS 102(43), 15545–15550 (2005).

GSEA



- Para asignar significancia: se permutan las etiquetas de fenotipos 1000 veces y se computa el ES para cada permutacion.
- Se reporta adicionalmente un Score Normalizado (NES) por la media observada en el dataset random, para alivianar la dependencia con el tamaño del CBI

Enriquecimiento funcional

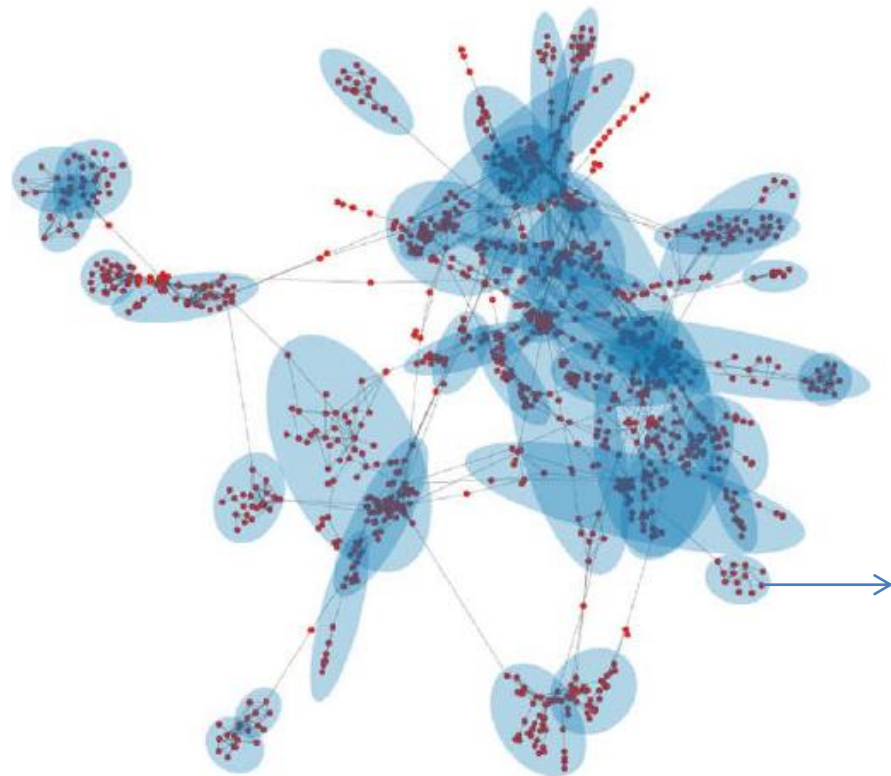
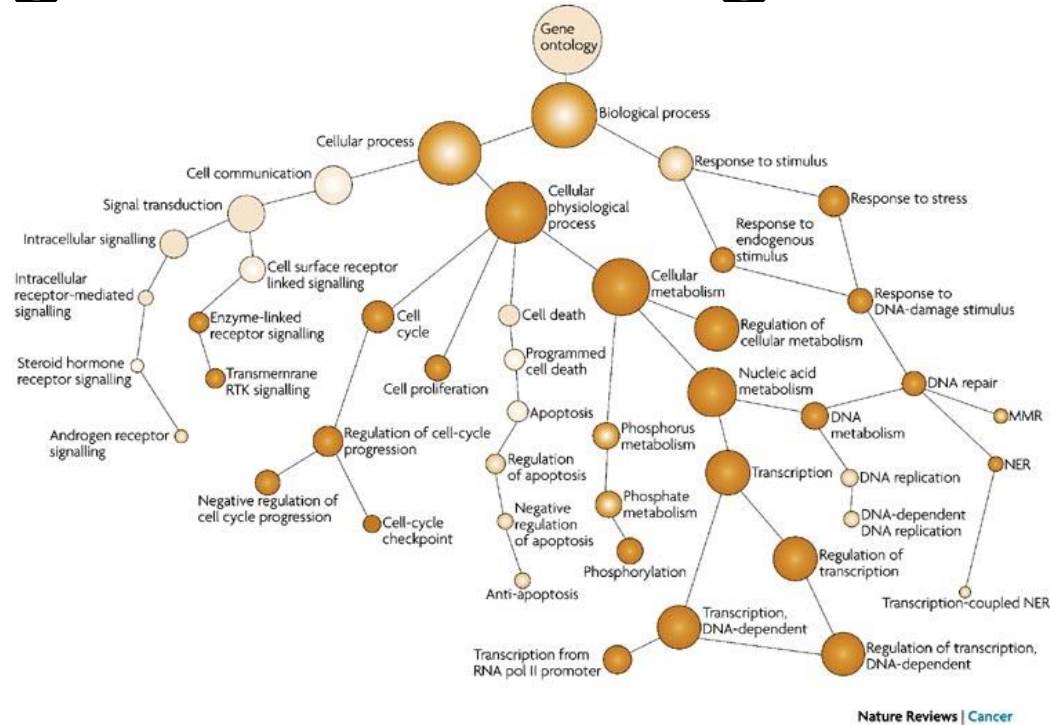


- **MsigDB**
- **Gene Ontology**
- **DAVID**



Índice de Homogeneidad Biológica

Validación de comunidades utilizando **conocimiento externo**

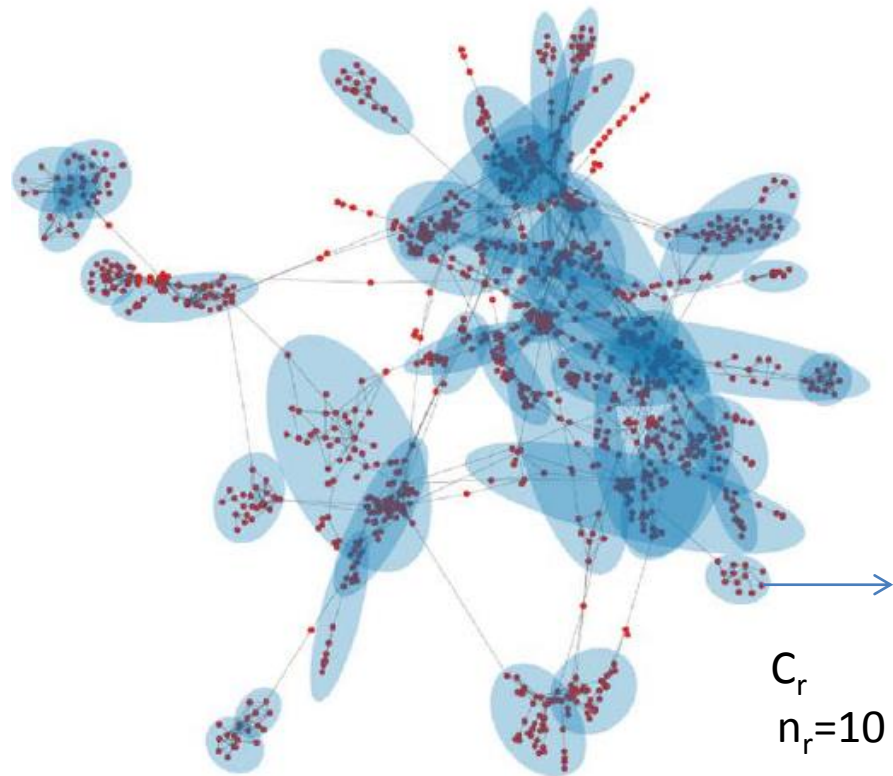
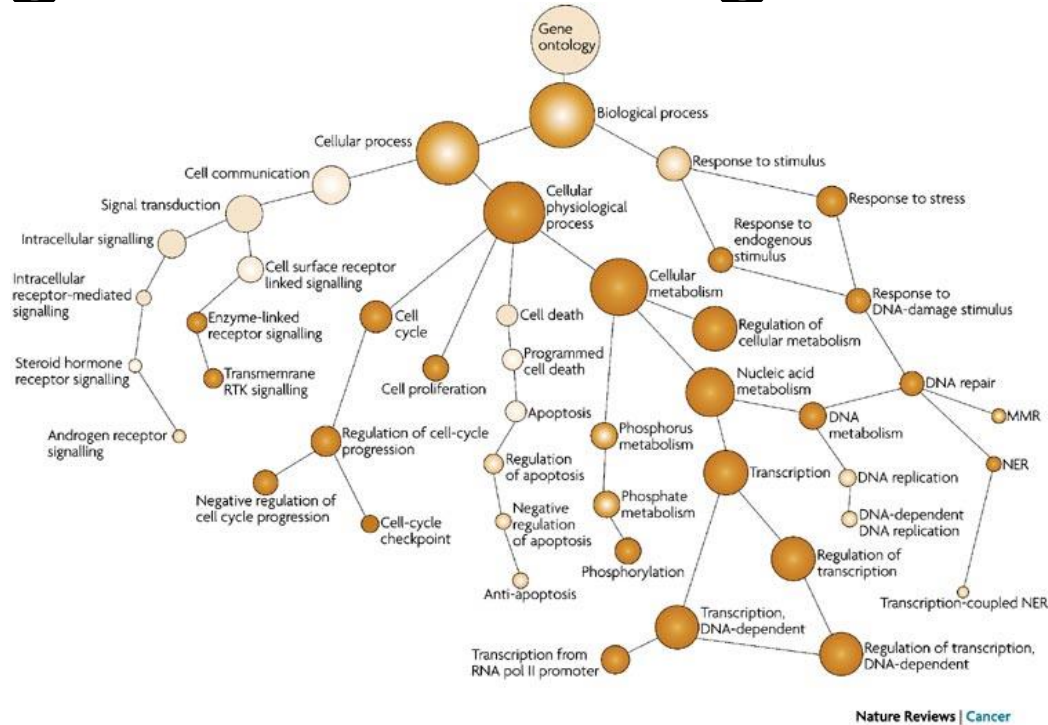


gen1
gen2
gen3
gen4
gen5
gen6
gen7
gen8
gen9
gen10

Grupos de genes/proteínas **cercanos** en el espacio de conocimiento capturado por la **red**, qué tan **compactos/homogéneos** son en **otro espacio de conocimiento** relevante?

Índice de Homogeneidad Biológica

Validación de comunidades utilizando conocimiento externo



gen1
gen2
gen3
gen4
gen5
gen6
gen7
gen8
gen9
gen10

C_r
 $n_r=10$

Así como es posible inferir una métrica a partir de la red también es posible inferir una **métrica** a partir de la estructura de **GO** que permita cuantificar el nivel de **similitud** entre proteínas en un **espacio de conocimiento biológico**.

Hacia una métrica GO

Cuanta información gano si me entero que una proteína está anotada en una clase GO_i?

Contenido de información

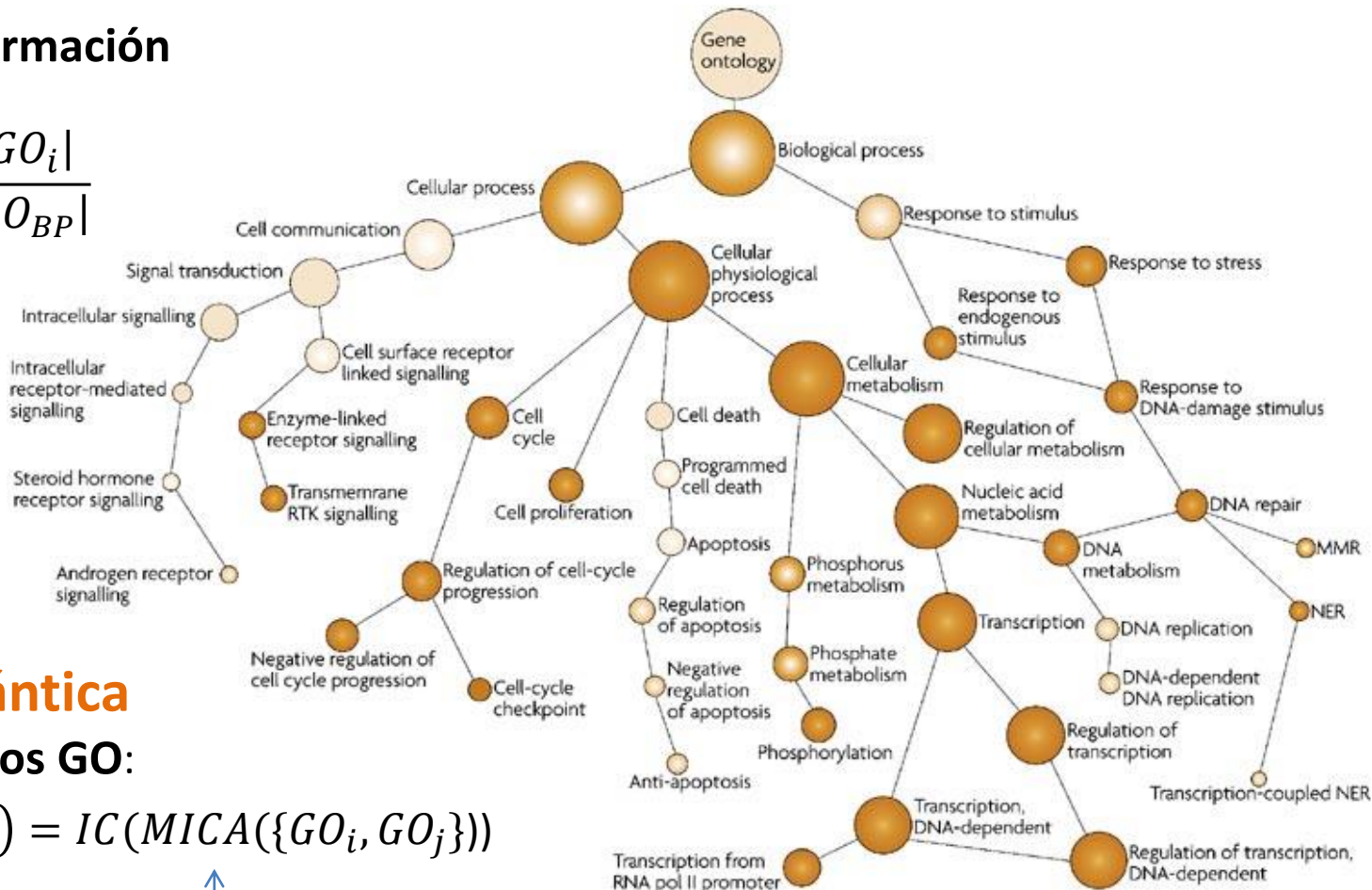
de un concepto GO:

$$IC(GO_i) = -\log \frac{|GO_i|}{|GO_{BP}|}$$

Similaridad **semántica**
entre dos **conceptos GO**:

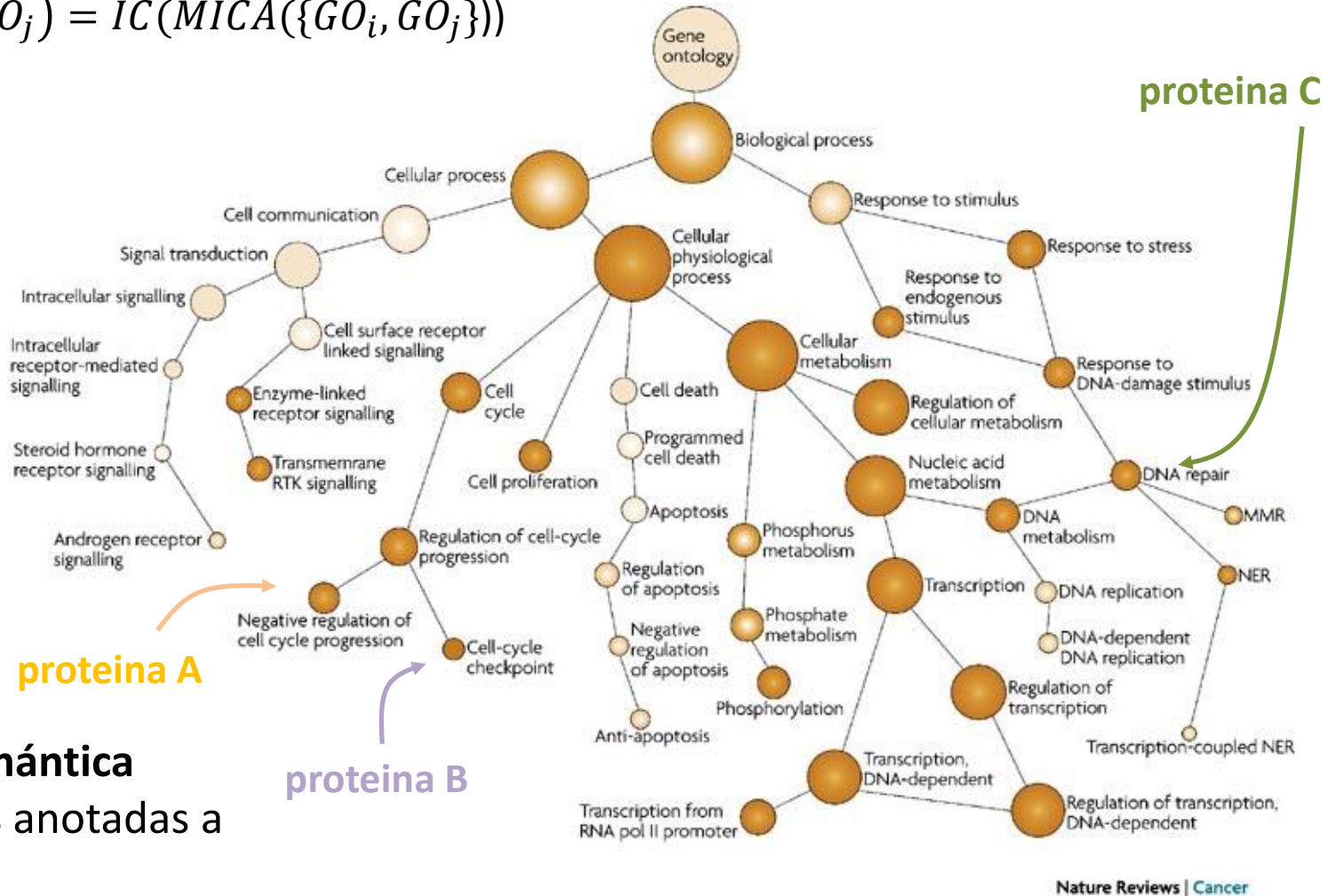
$$sim^{Resnik}(GO_i, GO_j) = IC(MICA(\{GO_i, GO_j\}))$$

Most Informative Common Ancestor



Similaridad semántica

$$sim^{Resnik}(GO_i, GO_j) = IC(MICA(\{GO_i, GO_j\}))$$



Similaridad semántica entre proteínas anotadas a conceptos GO.

$$sim(prot_A, prot_B) = f(\{GO_i\}_A, \{GO_i\}_B)$$

Similaridad semántica

Similaridad semántica entre **conceptos**

$$sim^{Resnik}(GO_i, GO_j) = IC(MICA(\{GO_i, GO_j\}))$$

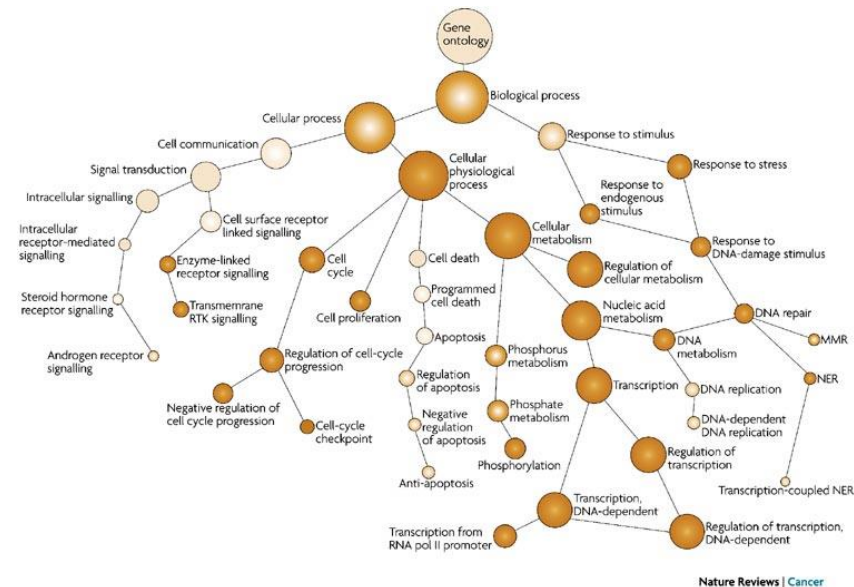
Similaridad semántica entre **proteínas**

$$prot_A \rightarrow \{GO_i\}_A \quad prot_B \rightarrow \{GO_i\}_B$$

$GO_3 \quad GO_5 \quad GO_{223} \quad GO_{542} \quad GO_{23}$

GO_1
 GO_{12}
 GO_{23}

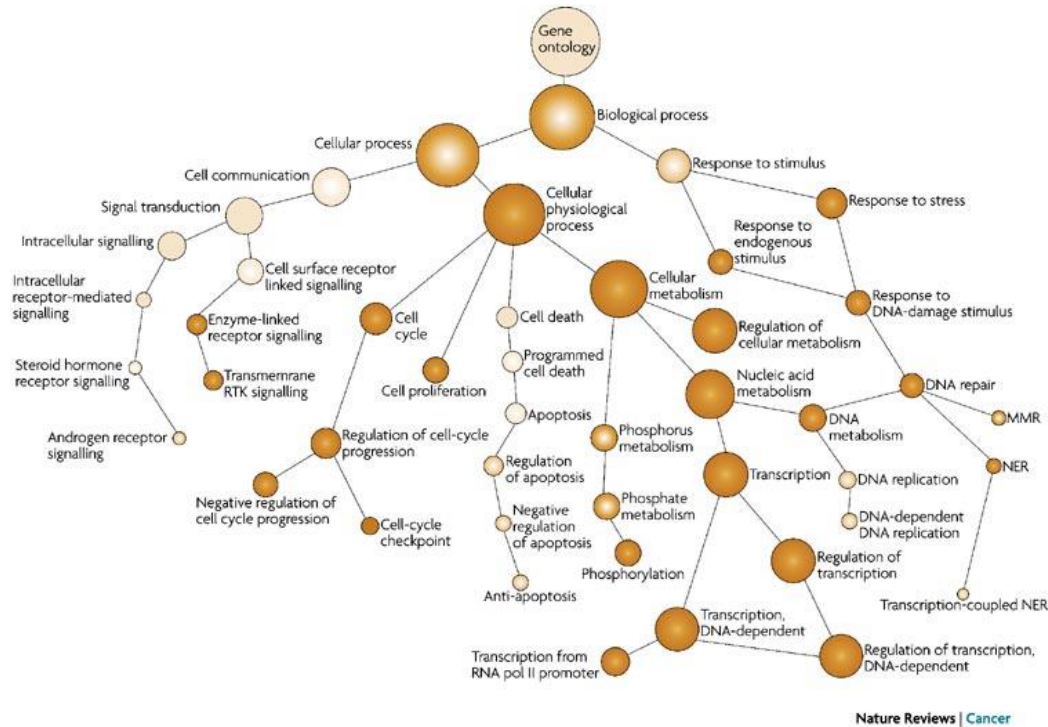
Matriz de sim GO's



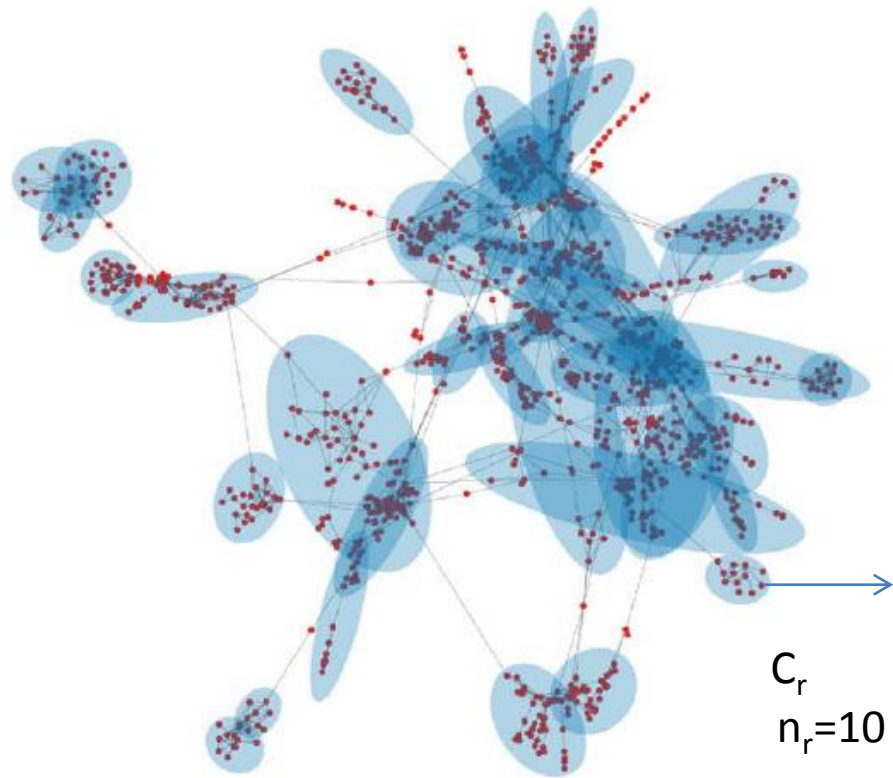
$$sim(prot_A, prot_B) = f(\{GO_i\}_A, \{GO_i\}_B)$$

Indice de Homogeneidad Biológica Reloaded

Validación de comunidades utilizando conocimiento externo



Nature Reviews | Cancer



gen1
gen2
gen3
gen4
gen5
gen6
gen7
gen8
gen9
gen10

C_r
 $n_r=10$

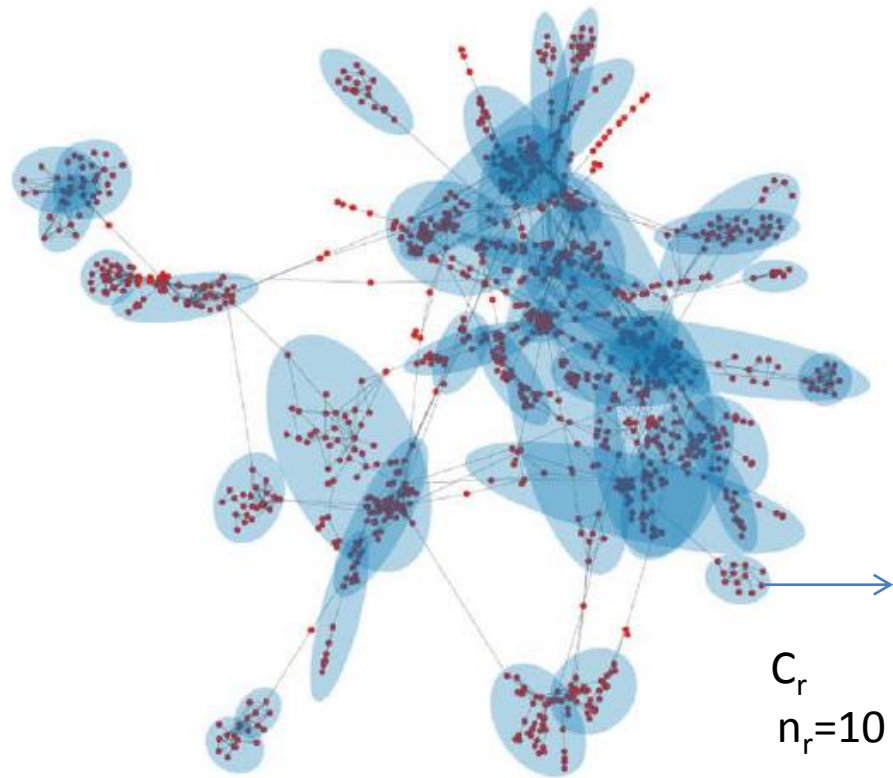
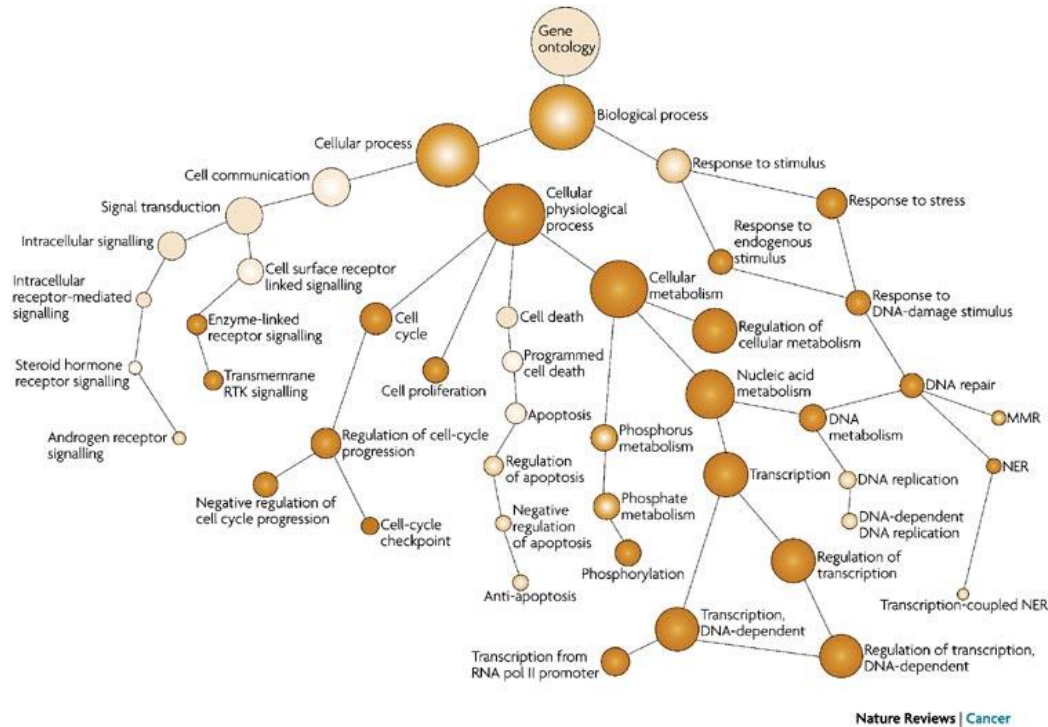
$$BHI_r = \frac{2}{n_r(n_r - 1)} \sum_{i \neq j \in C_r} \delta_{GO_i, GO_j}$$

$$BHI = \frac{1}{k} \sum_{r=1}^k BHI_r$$

↑
criterio binario de
coincidencia en
espacio GO

Indice de Homogeneidad Biológica Reloaded

Validación de comunidades utilizando conocimiento externo

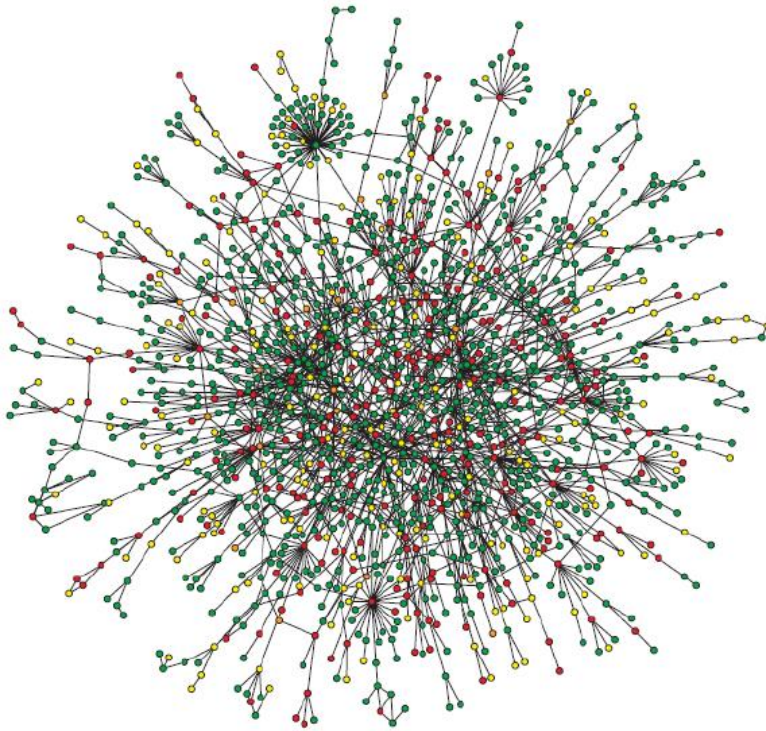


gen1
gen2
gen3
gen4
gen5
gen6
gen7
gen8
gen9
gen10

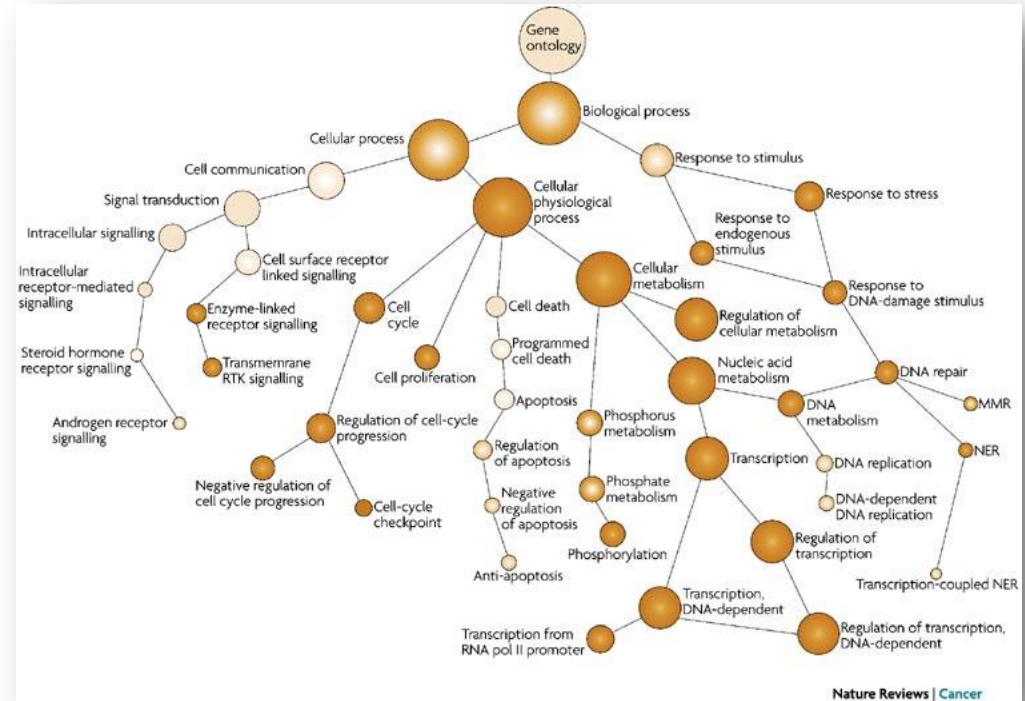
$$BHI_r = \frac{2}{n_r(n_r - 1)} \sum_{i \neq j \in C_r} sim_{GO}(i, j)$$

$$BHI = \frac{1}{k} \sum_{r=1}^k BHI_r$$

Integrando métricas



Gene Ontology



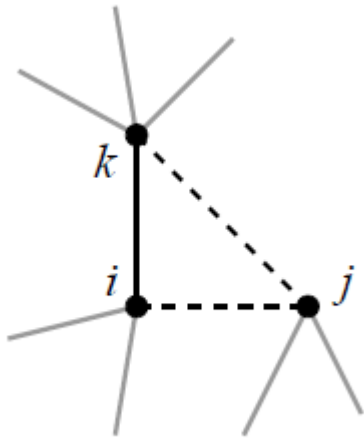
d_{red}

$d_{GO} = d_{semantic}$

Mixed Metrics

Buscamos estructuras **similares** en cuanto a la información embebida en la topología de la red **Y** el espacio de conocimiento biológico

Recordemos...



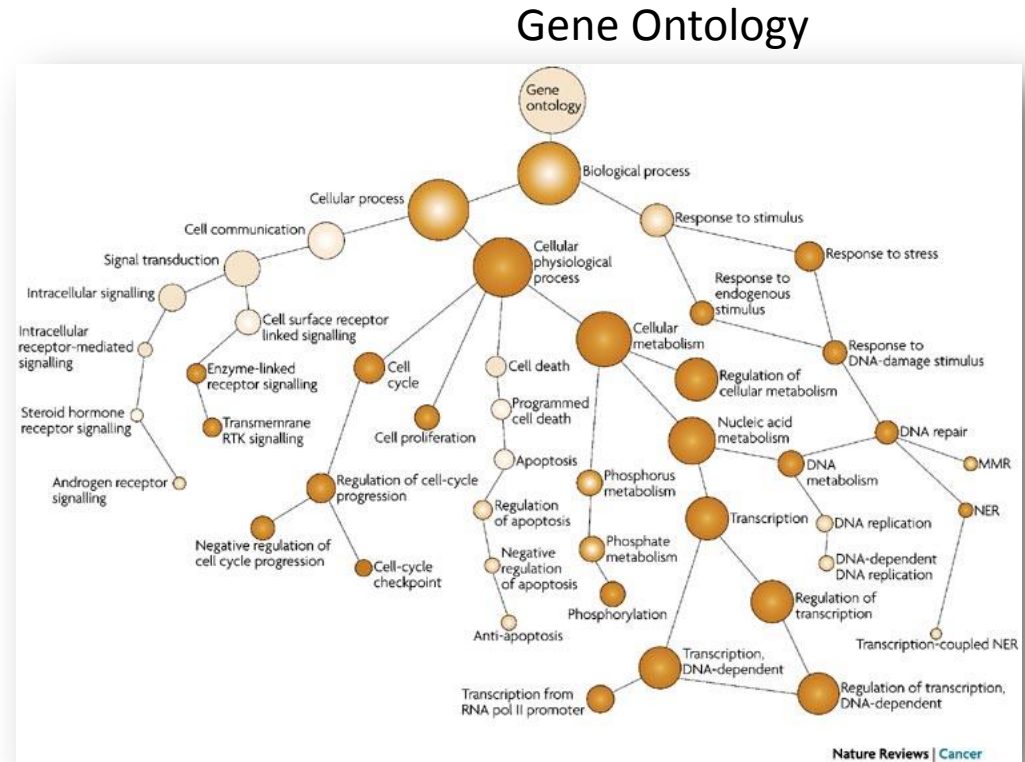
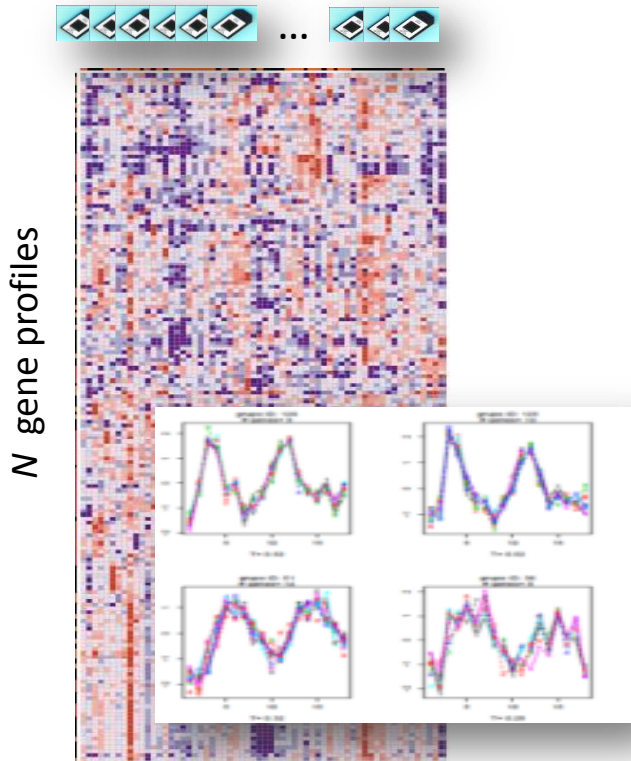
Equivalencia estructural **extendida**:

Dos vértices, i y j son similares si i tiene como vecino a k , quien a su vez es similar a j .

$$\sigma_{ij} = \alpha \sum_k A_{ik} \sigma_{kj} + \underline{s^{ext}}_{ij}$$

Nodos con alto grado tienen más vecinos, que pueden ser similares a terceros nodos, por lo que tienen más chances de aumentar su nivel de similaridad general. Si se desea mitigar este efecto...

Integrando métricas



$$dx = d_{\text{coexpression}}$$

$$d_{GO} = d_{\text{semantic}}$$

Mixed Metrics

Buscamos estructuras **similares** en cuanto a la información transcripcional **Y** el espacio de conocimiento biológico

Agrupando genes con métricas mixtas

El uso de métricas mixtas en algoritmos de reconocimiento de clusters / comunas permite buscar estructura simultáneamente coherente en ambos espacios (por ejemplo: espacio transcripcional y GO)

Las métricas pueden combinarse de diferente forma

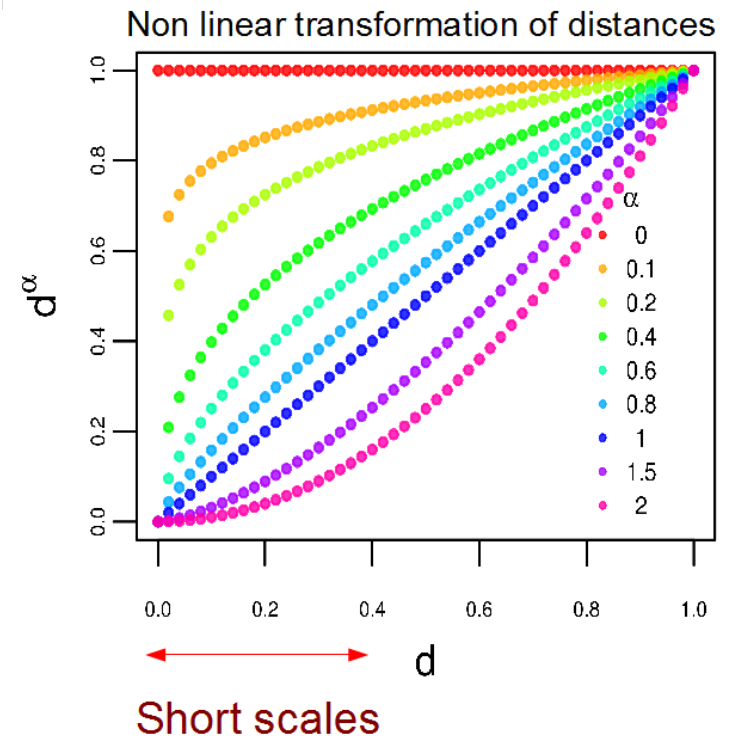
$$* d_{mix}^A = \sqrt{(1 - \alpha)d_x^2 + \alpha d_{go}^2}$$

$$* d_{mix}^B = d_x \cdot d_{go}^\alpha$$

$$* d_{mix}^C = d_x^{1-\alpha} \cdot d_{go}^\alpha$$

- El parámetro α controla la medida en la que el espacio GO es tenido en cuenta en la estimación de similitudes entre objetos.

- d^B **acerca** genes **biológicamente relacionados** en el sentido de GO

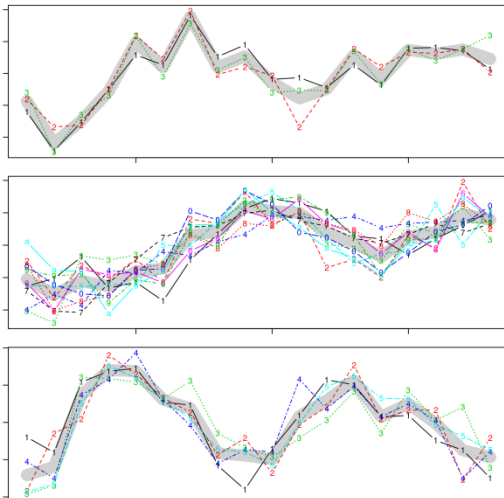


La transformación no-lineal altera de manera diferente la resolución a diferentes escalas

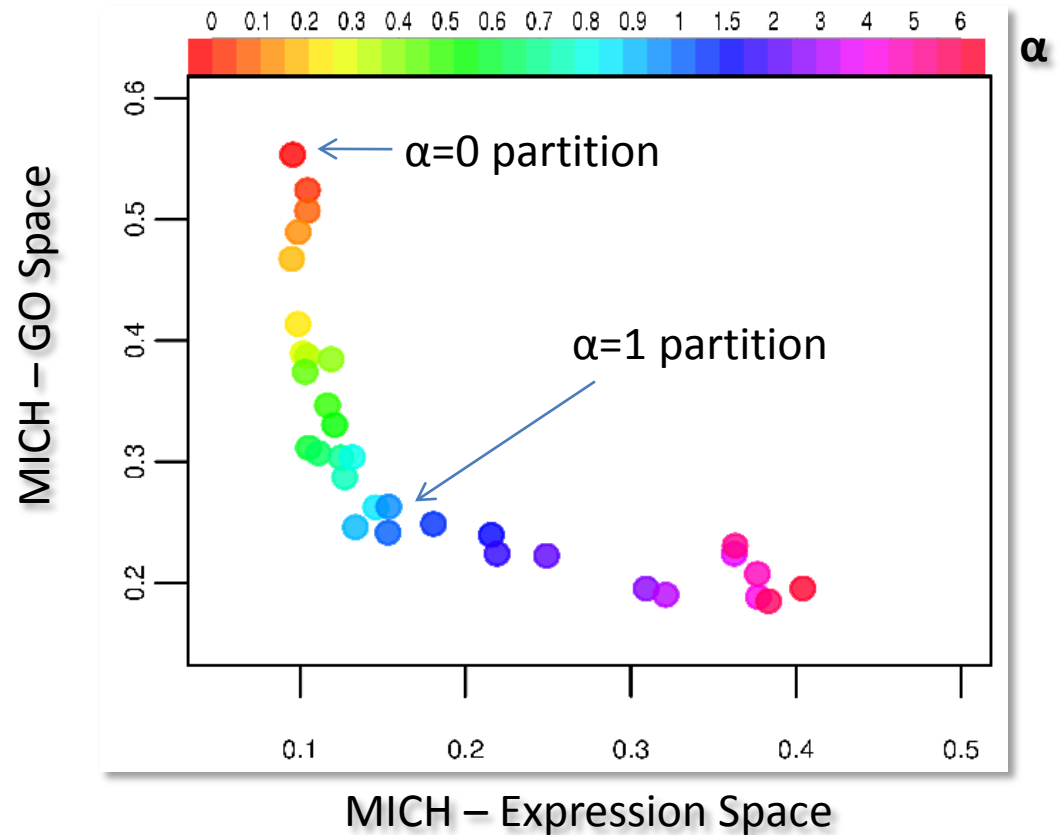
Agrupando genes con métricas mixtas

$$d_{mix} = d_x \cdot d_{go}^\alpha$$

Cluster structure



Mean Intra Cluster Heterogeneity



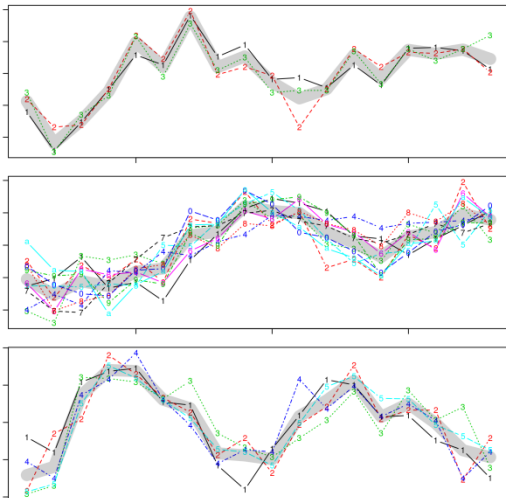
$$MICH(Y) = \frac{1}{N_{cl}} \sum_{cl_k}^{N_{cl}} \frac{2}{n_k(n_k-1)} \sum_{i < j \in cl_k}^{n_k} d_{ij}^Y$$

Agrupando genes con métricas mixtas

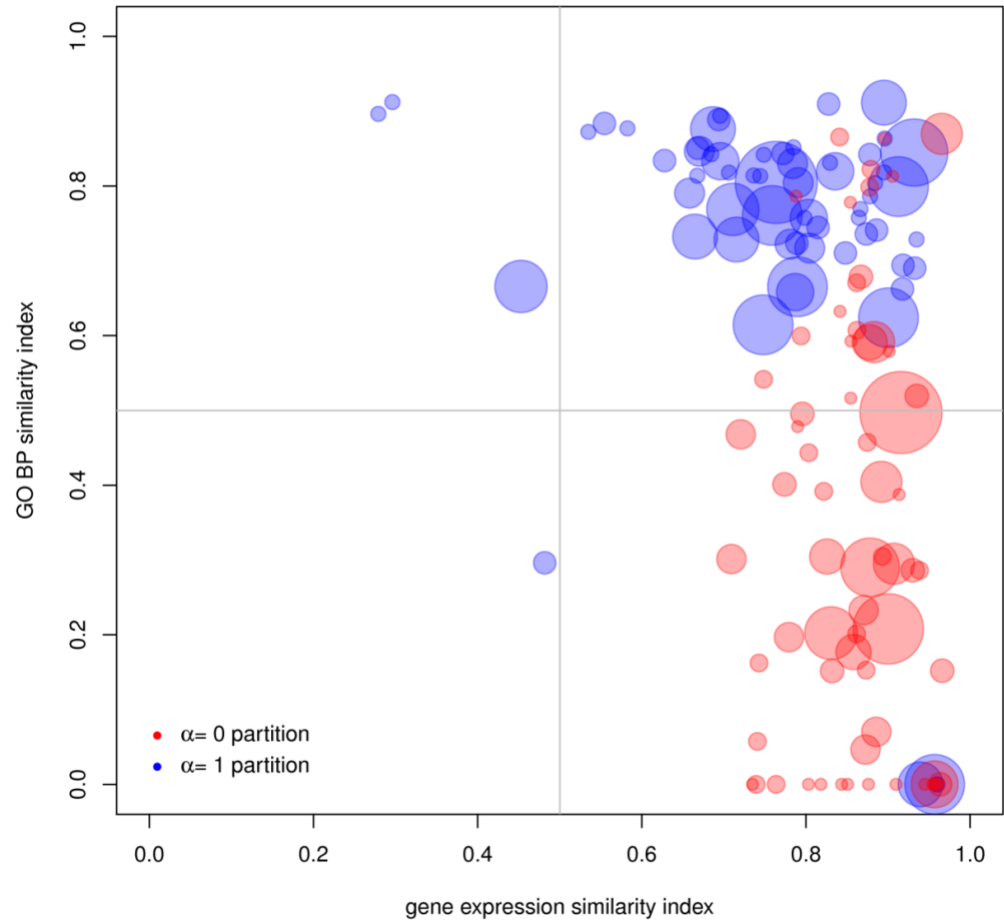
$$d_{mix} = d_x \cdot d_{go}^\alpha$$



Cluster structure



Clusters in similarity plane





SAMBA + CLICK