

# Análisis funcional

Genómica Funcional  
Máster en Genética y Evolución

# Motivación

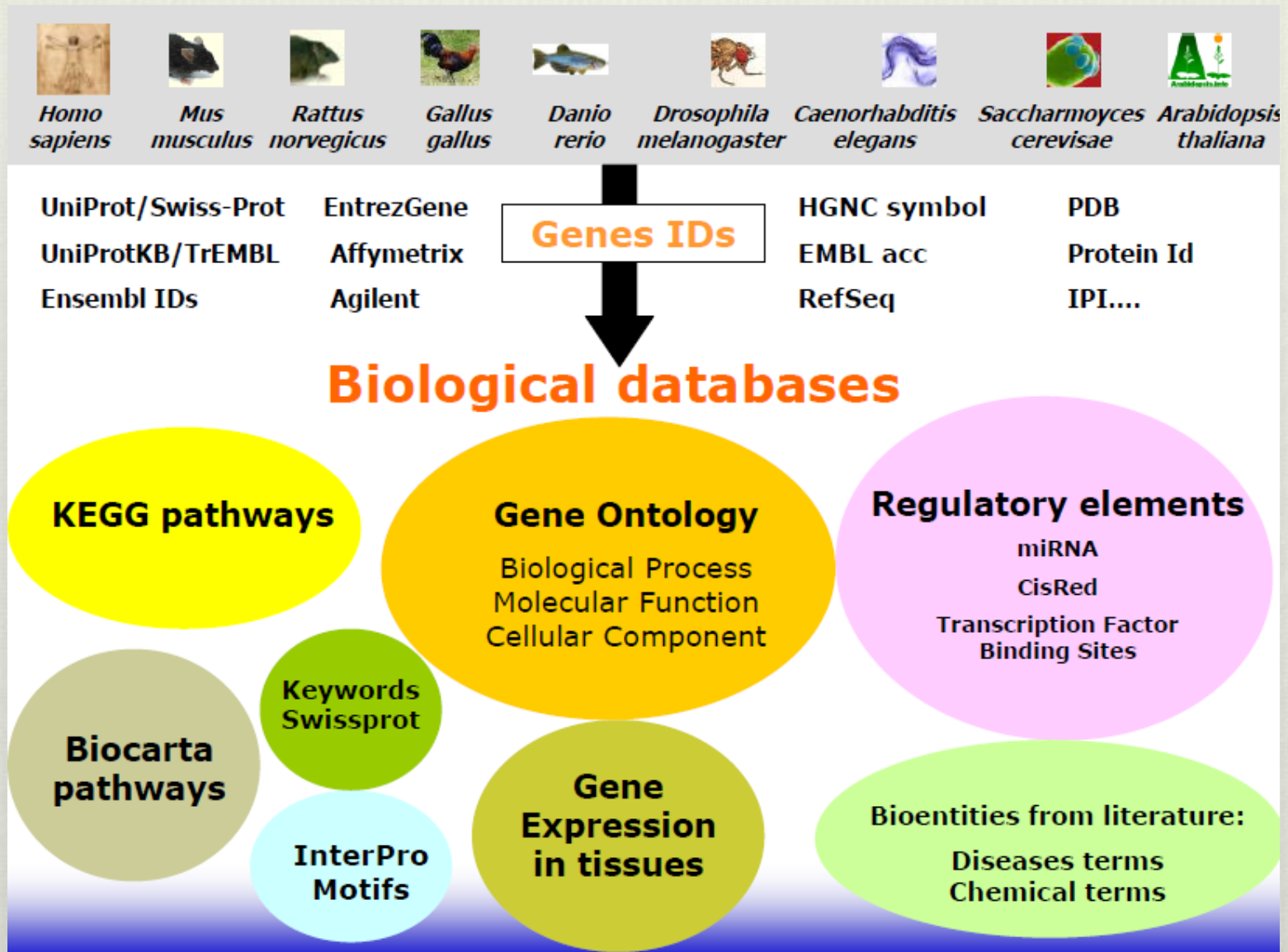
Aparición de técnicas de alto rendimiento para medir los niveles de la expresión génica (chips de ADN, SAGE, secuenciación masiva (mRNA-seq))

- permiten monitorizar la expresión de miles de genes simultáneamente
- esto permite detectar los genes que se expresan diferencialmente en dos condiciones (por ejemplo, tejido sano frente a tejido cancerígeno)
- el resultado es una **lista de genes** que se expresan diferencialmente entre dos condiciones

Surge la cuestión: ¿Cómo se puede traducir **una lista de genes en conocimiento biológico?**

- Análisis estadístico de las propiedades funcionales de los genes para detectar aquellas funciones que están o enriquecidas o empobrecidas en la lista de genes

# Anotaciones funcionales



# GO ontology

<http://geneontology.org/docs/ontology-documentation/>

## Las tres categorías de términos GO

### **Función Molecular**

Describen las tareas que llevan a cabo los productos génicos. Ejemplos: *factor de transcripción, ADN helicasa, etc.*

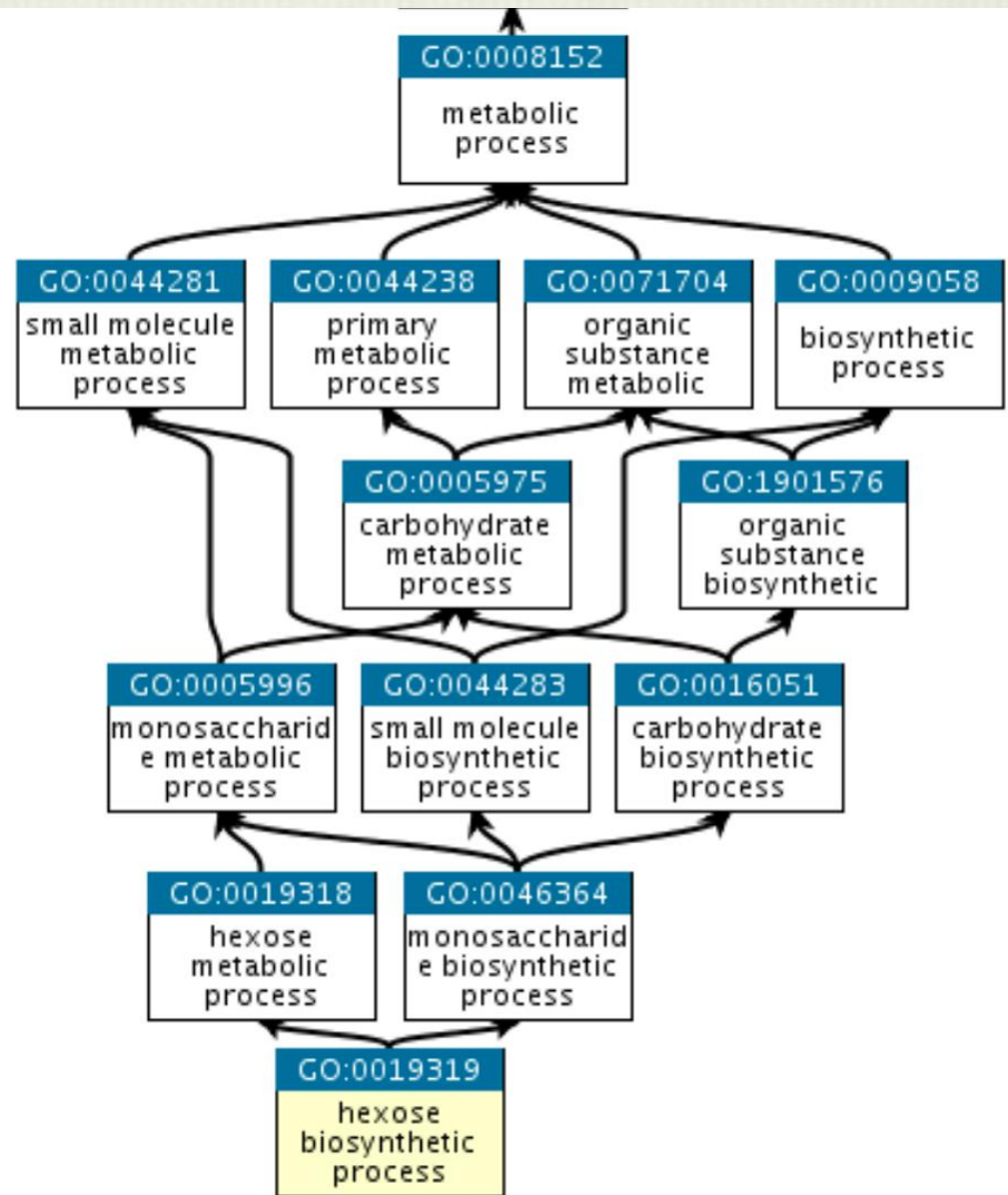
### **Proceso Biológico**

Se trata de los grandes procesos biológicos, como la *mitosis* o el *metabolismo de las purinas*, que son llevados a cabo por conjuntos ordenados de funciones moleculares.

### **Componente Celular**

Estructuras subcelulares, localizaciones y complejos macromoleculares. Ejemplos: *núcleo, telómero, origen de replicación, etc.*

# Organización en grafo acíclico dirigido



# Análisis funcional

Mediante experimentos se pueden asignar los términos GO a los genes

→Mediante estas anotaciones se puede caracterizar los genes

Objetivo:

**¿Que términos funcionales están sobre o infrarepresentados en una lista de genes?**

→Se puede comparar una lista de genes con un conjunto de genes de referencia (todos los genes del genoma por ejemplo)

→De una forma muy parecida se comparan dos listas de genes una frente a la otra

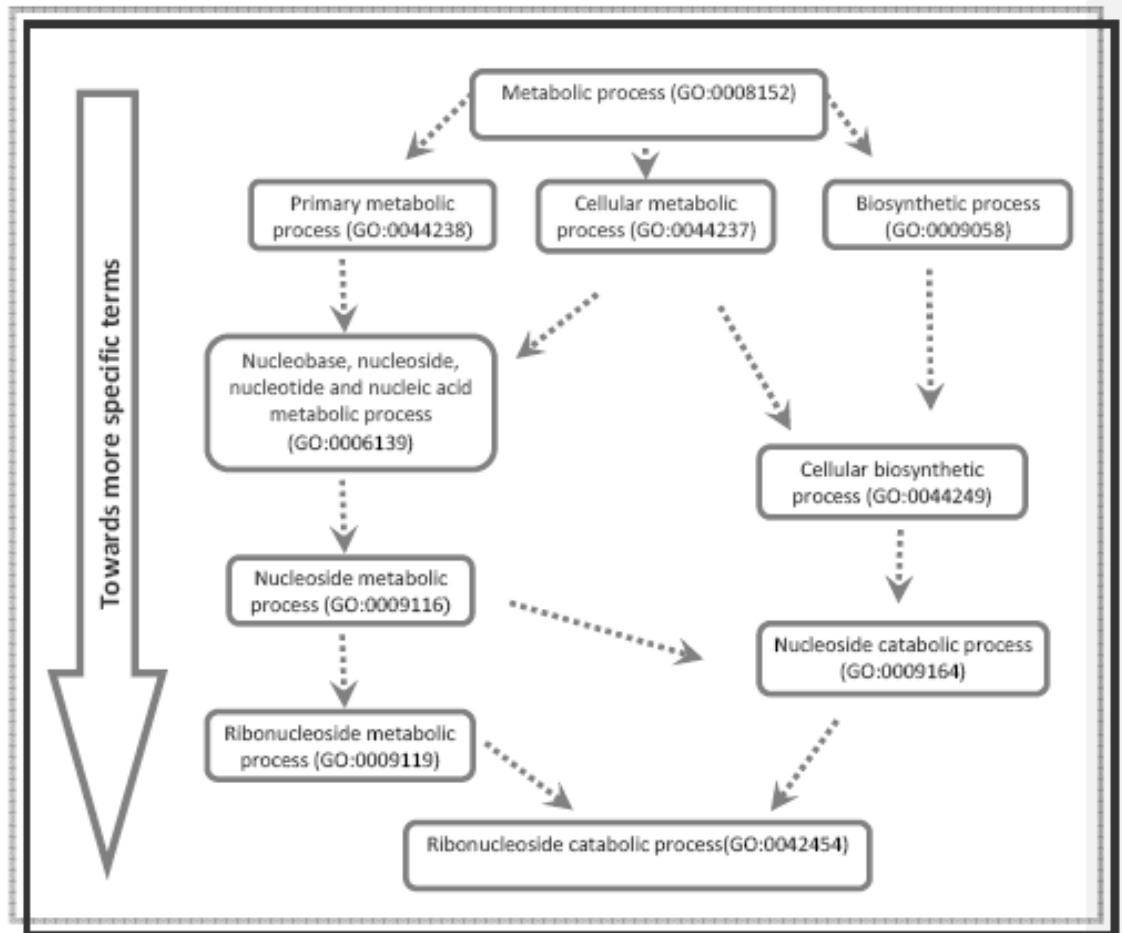


Figure 1: The figure illustrates the structure in which the functional terms are organized in the Gene Ontology by means of a subgraph of the GO term "metabolic process". The terms are ordered in a hierarchical structure called DAG (direct acyclic graph). The categories are ordered from more general (top of the graphic) to more specific terms (bottom of the graphic).

# Enriquecimiento relativo

Enriquecimiento relativo:

$$R_e(A_i) = \frac{N_p}{N} \cdot \frac{n}{n_p}$$

$N_p$ : número de genes asignados a la anotación A en la lista de genes objeto

$N$ : número de genes en la lista objeto

$n$ : número de genes en la referencia (background gene set)

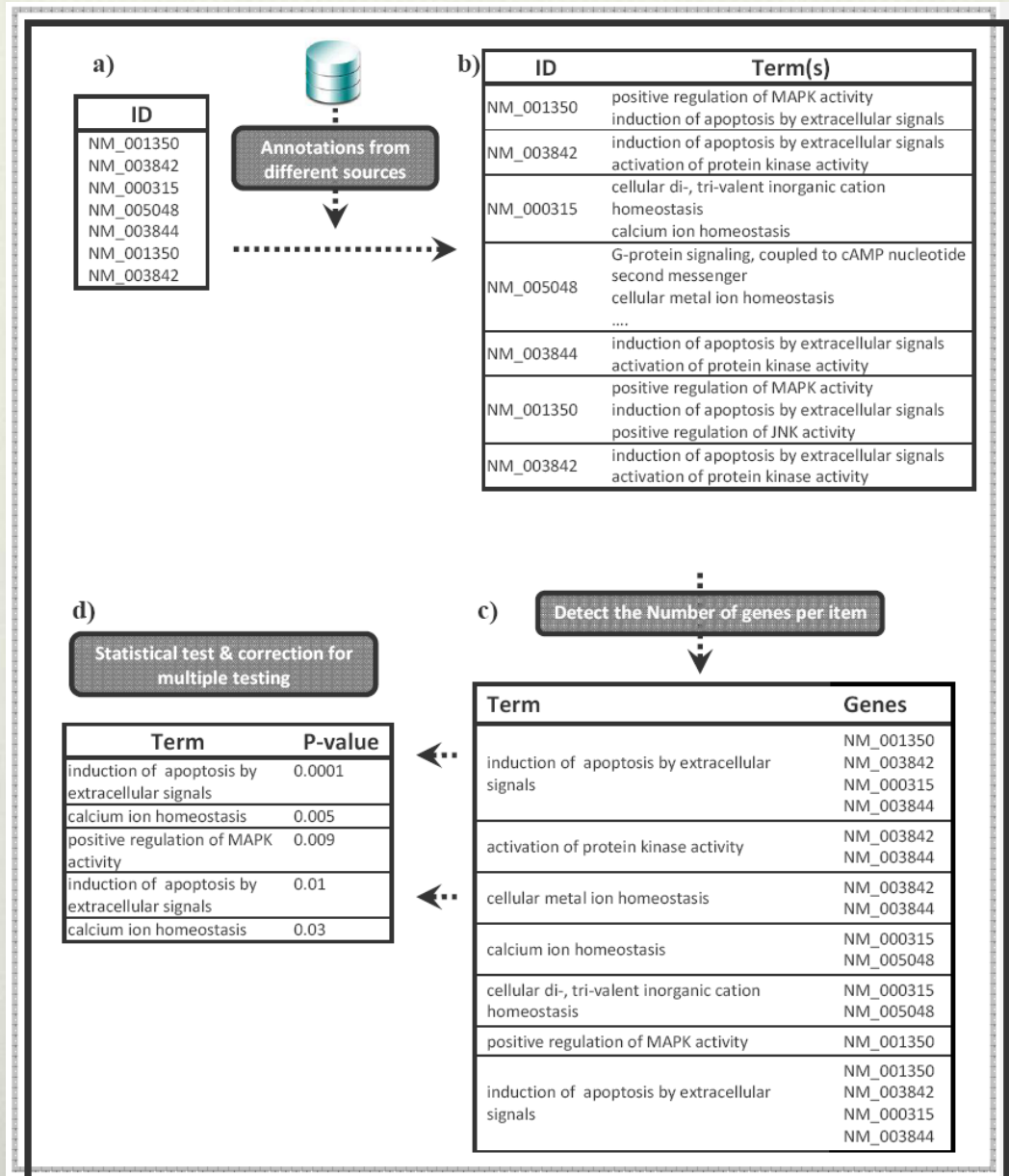
$n_p$ : número de genes asignados a la anotación A en la lista de los genes de referencia

$Re=1$ : no hay diferencias entre la lista analizada y la referencia

$Re>1$ : hay más genes asignados a la anotación en la lista analizada que esperado por azar (enriquecimiento)

$Re<1$ : hay menos genes asignados a la anotación en la lista analizada que esperado por azar (empobrecimiento)

→ Se calcula un valor p para comprobar si el  $Re$  está estadísticamente significativo diferente a 1



# el valor p

Explorar términos GO:

<https://www.ebi.ac.uk/QuickGO>

**Ejemplo:**

$$R_e(A_i) = \frac{N_p}{N} \cdot \frac{n}{n_p}$$

**Termino GO: p53 binding**

Genes en la base de datos: 100 positivos ||| Número total de genes: 20000  
Lista de entrada: 20 positivos ||| Número total de genes: 1000

**Enriquecimiento relativo:**  $(20/1000) \cdot (20000/100) = 4$

→ 4 veces mas genes con 'p53 binding' en la lista de genes que esperado por azar

**Valor p: tabla de contingencia & Fisher Exact Test o Chi2**

<https://www.socscistatistics.com/tests/fisher/default2.aspx>

Enter Your Data Below		
	lista	rest
postive	<input type="text" value="20"/>	<input type="text" value="80"/>
negative	<input type="text" value="980"/>	<input type="text" value="18920"/>

Results			
	lista	rest	<i>Marginal Row Totals</i>
postive	20	80	100
negative	980	18920	19900
<i>Marginal Column Totals</i>	1000	19000	20000 (Grand Total)

The Fisher exact test statistic value is < 0.00001. The result is significant at  $p < .05$ .