

# Gene Ontology: GO

- ...los bioquímicos caracterizan las proteínas por su actividad y concentración
- ...los genéticos caracterizan los genes por sus fenotipos y sus mutaciones
- ...todos los biólogos reconocen ahora que probablemente existe un **universo limitado de genes y proteínas**, conservado en la mayor parte de las formas de vida

Esto ha promovido la **gran unificación de la biología**; la información acerca de los genes y proteínas compartidos contribuye a comprender mejor las diversas formas de vida

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*, 25(1), 25-9. doi: 10.1038/75556.



*Homo sapiens*

*Mus musculus*

*Rattus norvegicus*

*Gallus gallus*

*Danio rerio*

*Drosophila melanogaster*

*Caenorhabditis elegans*

*Saccharomyces cerevisiae*

*Arabidopsis thaliana*

UniProt/Swiss-Prot

EntrezGene

UniProtKB/TrEMBL

Affymetrix

Ensembl IDs

Agilent

**Genes IDs**

HGNC symbol

PDB

EMBL acc

Protein Id

RefSeq

IPI....

## Biological databases

**KEGG pathways**

**Gene Ontology**

Biological Process  
Molecular Function  
Cellular Component

**Regulatory elements**

miRNA

CisRed

Transcription Factor  
Binding Sites

**Biocarta pathways**

Keywords  
Swissprot

**Gene Expression  
in tissues**

**InterPro  
Motifs**

**Bioentities from literature:**

Diseases terms  
Chemical terms



# Gene Ontology CONSORTIUM

<http://www.geneontology.org>

- Su objetivo es elaborar vocabularios controlados de **términos** para la descripción de la función molecular, el proceso biológico y el componente celular de los productos génicos
- Estos términos GO se usan como atributos de los productos génicos por parte de las distintas bases de datos, facilitando así búsquedas estandarizadas
- Los vocabularios de términos GO están estructurados

# Las tres categorías de términos GO

## Función Molecular

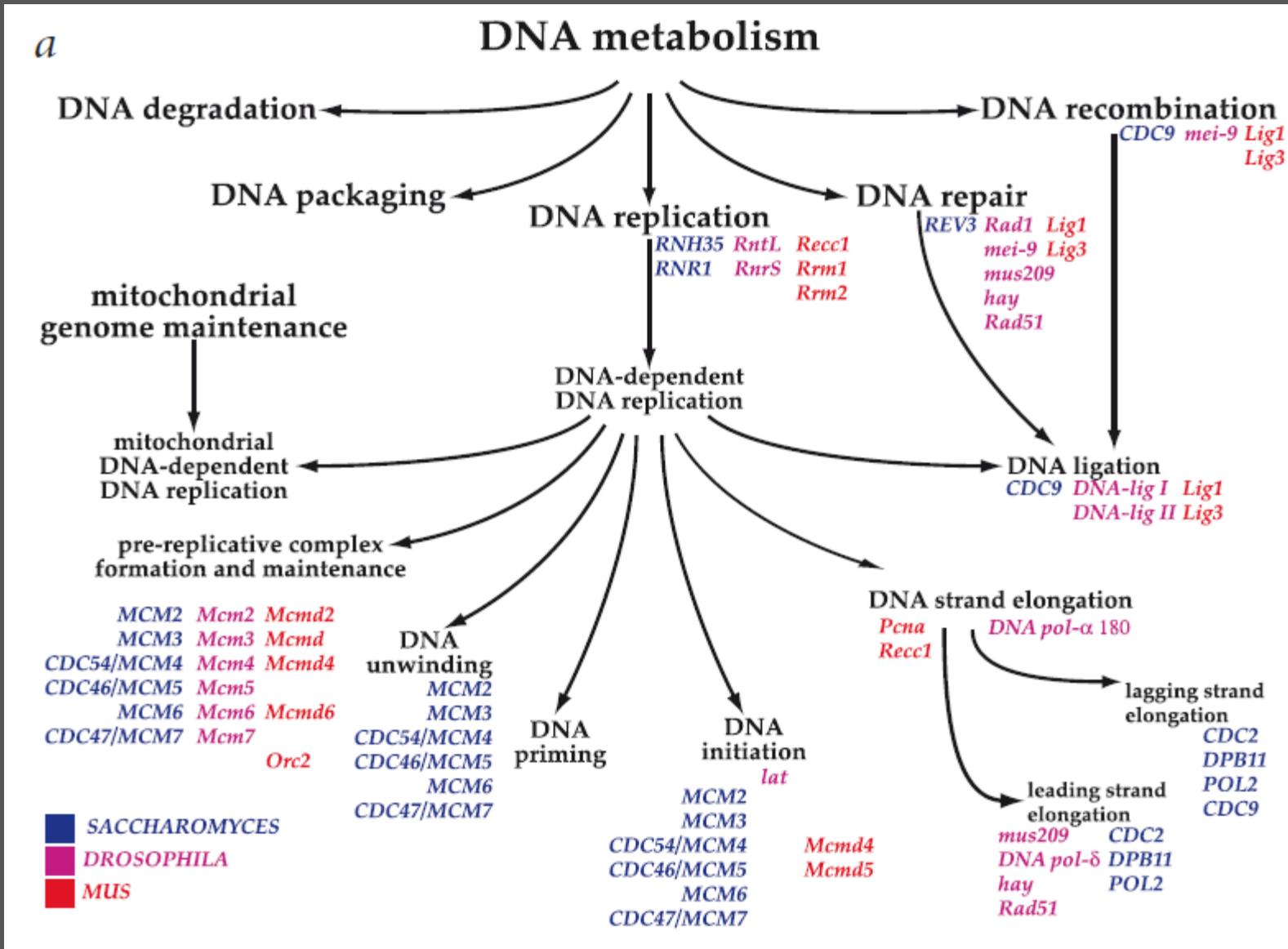
Describen las tareas que llevan a cabo los productos génicos. Ejemplos: *factor de transcripción*, *ADN helicasa*, etc.

## Proceso Biológico

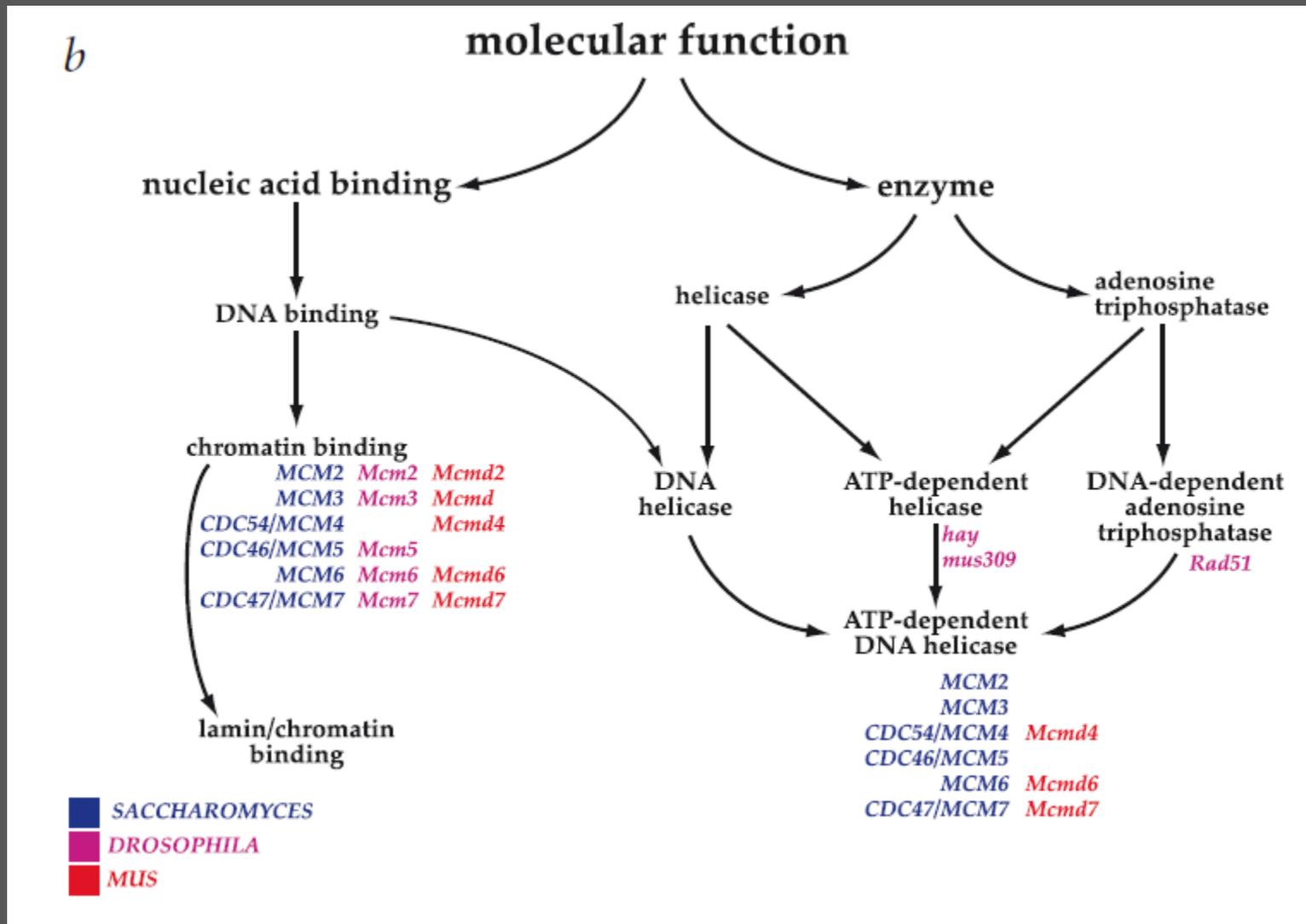
Se trata de los grandes procesos biológicos, como la *mitosis* o el *metabolismo de las purinas*, que son llevados a cabo por conjuntos ordenados de funciones moleculares.

## Componente Celular

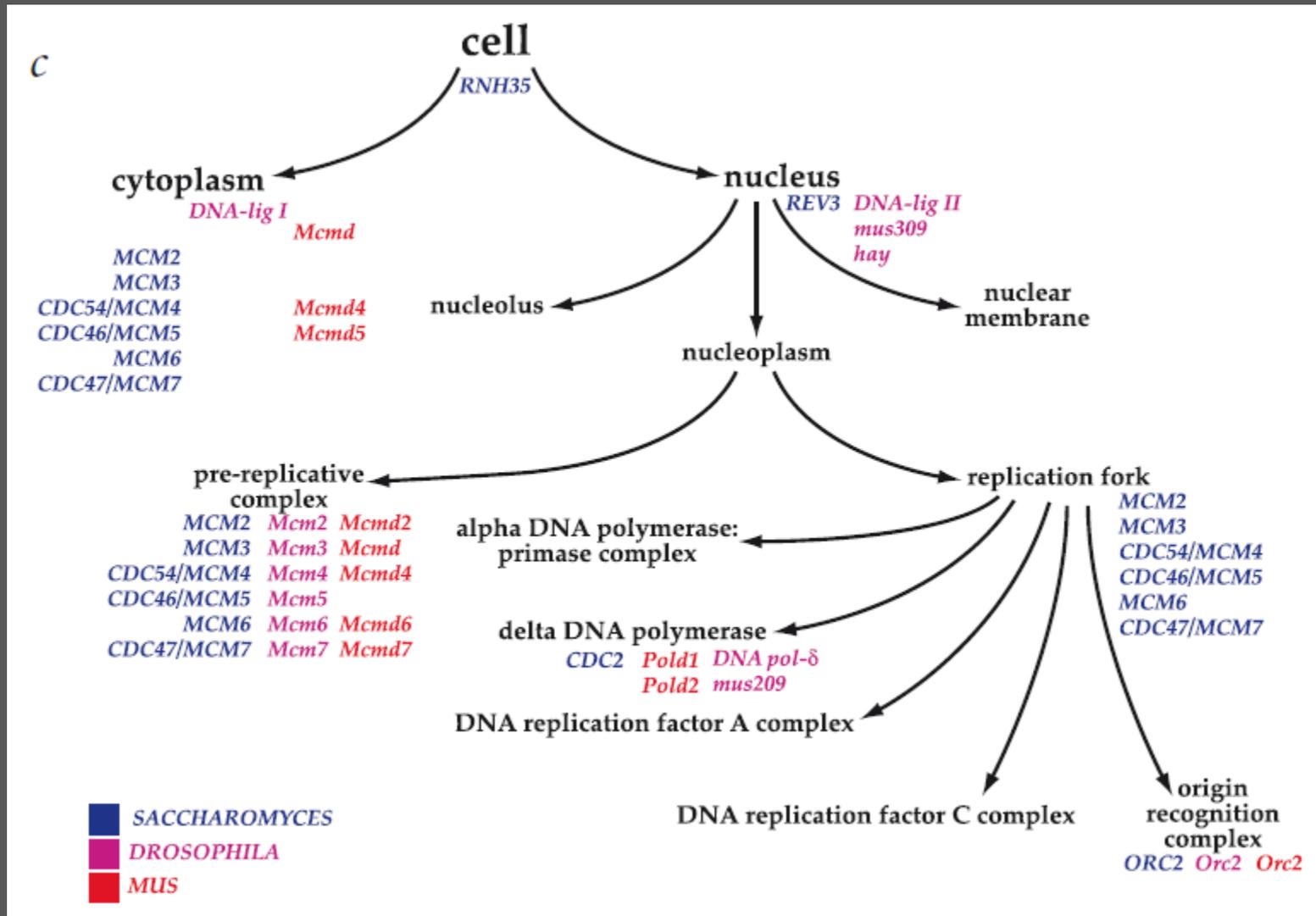
Estructuras subcelulares, localizaciones y complejos macromoleculares. Ejemplos: *núcleo*, *telómero*, *origen de replicación*, etc.



# Molecular function ontology



# Cellular component ontology



## ¿Qué es lo que hace pues Gene Ontology?

- Gene Ontology toma un producto génico (por ejemplo, una proteína) y devuelve su contexto celular
- Para cada una de las tres ontologías, los productos génicos se sitúan donde pertenecen, y se pueden tomar varios términos para encontrar los productos génicos asociados

## Términos GO

- Se trata de términos descriptivos que permiten asociar a cada producto génico un contexto molecular, celular y de proceso biológico
- Los términos son estandard en todas las bases de datos y se usan sinónimos para salvar las pequeñas diferencias ortográficas o gramaticales, o para describir funciones similares
- Los términos se van renovando a medida que avanza el conocimiento biológico

# Ejemplo de búsqueda

La búsqueda [“Q59J86”](#)

Produce:

- Name(s) “DNA polymerase”
- Type “protein”
- Species [“Gallus gallus \(chicken\)”](#)
- Synonyms “IPI00588123”
- Sequence
- References
- Term associations

# Estructura del árbol GO

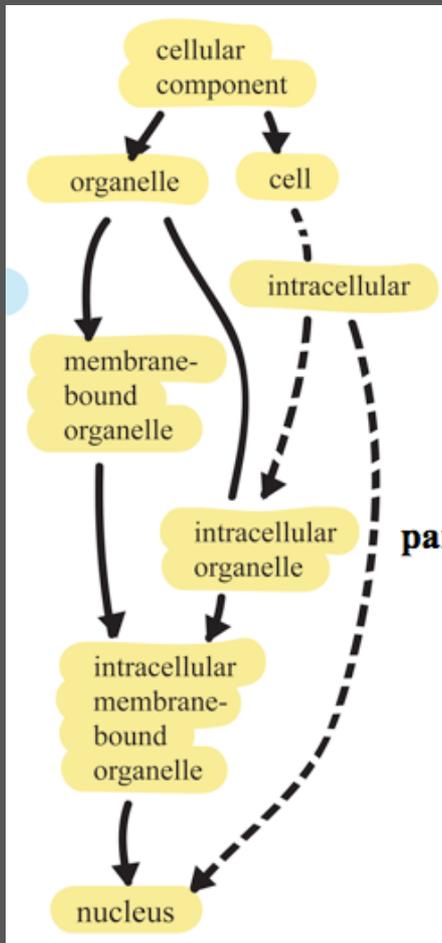
- [-] [GO:0003673 : Gene Ontology \(65883\)](#) 
  - [+] [GO:0008150 : biological process \(44405\)](#) 
    - [+] [GO:0007610 : behavior \(357\)](#)
    - [GO:0000004 : biological process unknown \(7877\)](#)
    - [-] [GO:0009987 : cellular process \(32672\)](#) 
      - [+] [GO:0007154 : cell communication \(5384\)](#)
      - [+] [GO:0008219 : cell death \(744\)](#)
      - [+] [GO:0030154 : cell differentiation \(464\)](#)
      - [+] [GO:0008151 : cell growth and/or maintenance \(28802\)](#)
      - [+] [GO:0006928 : cell motility \(911\)](#)
      - [+] [GO:0006944 : membrane fusion \(257\)](#)
    - [+] [GO:0016265 : death \(793\)](#)
    - [+] [GO:0007275 : development \(4615\)](#)
    - [+] [GO:0008371 : obsolete \(1581\)](#)
    - [+] [GO:0007582 : physiological processes \(31124\)](#)
    - [+] [GO:0016032 : viral life cycle \(115\)](#)
  - [+] [GO:0005575 : cellular component \(32869\)](#)
  - [+] [GO:0003674 : molecular function \(53910\)](#)

# Tipos de relaciones

- Is\_a [i]
- Part\_of [p]
- Regulates/ positively\_regulates / negatively\_regulates [r]

GO:0010467 : gene expression  
[r] GO:0010468 : regulation of gene expression  
---[i] GO:0045449 : regulation of transcription  
[p] GO:0006350 : transcription  
---[r] GO:0045449 : regulation of transcription

# Ejemplo



Las líneas sólidas son relaciones Is\_a

Las líneas de puntos son relaciones Part\_of

# Estructura de grafo

- Las ontologías se estructuran en forma de grafo acíclico dirigido (DAG), que son grafos que no contienen ciclos ni repeticiones.
- Son similares a las jerarquías pero difieren en que un término más especializado (hijo) puede relacionarse con más de un término más genérico (parental).



## Estructura de datos

- Las tres ontologías, las anotaciones y los productos génicos se almacenan en una base de datos relacional.
- GO es una base de datos en MySQL, actualizada permanentemente en línea.
- La base de datos se puede acceder mediante diferentes 'clientes' gráficos, como por ejemplo *AmiGO*.
- También puede accederse desde la línea de comando por diferentes programas, o puede descargarse completa.

http://amigo.geneontology.org/cgi-bin/amigo/search.cgi?session\_id=1830amigo1294836292

**the Gene Ontology** **AmiGO**

Search Browse BLAST More Tools Help

Search the Gene Ontology database

Q59J86

GO terms  genes or proteins  exact match

Enviar consulta

AmiGO Version: [1.7](#)  
[Try AmiGO Labs](#)

GO database release 2011-01-08  
[Cite this data](#) • [Terms of use](#) • [GO helpdesk](#)  
Copyright © 1999-2009 [the Gene Ontology](#)



Search Browse BLAST More Tools Help

Search GO   terms  genes or proteins  exact match

# Q59J86

Gene product information ↓ Peptide sequence ↓ Sequence information ↓ 33 term associations →

## Information

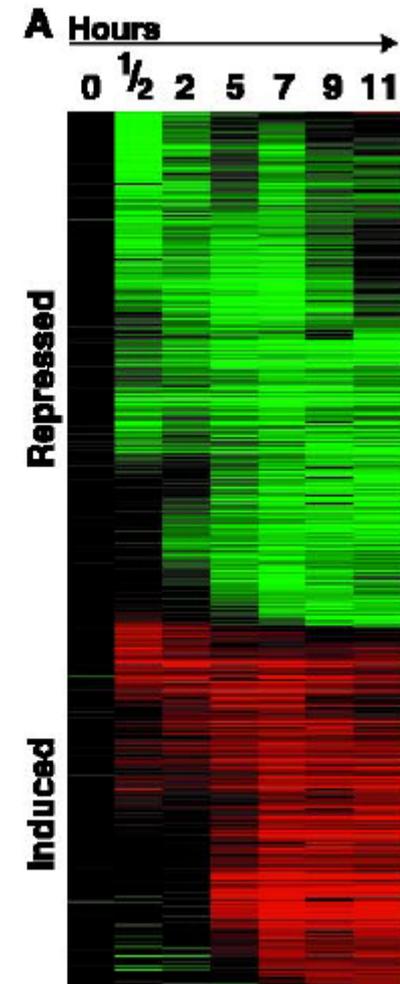
<b>Symbol</b>	Q59J86
<b>Name(s)</b>	DNA polymerase
<b>Type</b>	<b>protein</b>
<b>Species</b>	<u><i>Gallus gallus</i></u> (chicken)
<b>Synonyms</b>	IPI00588123 Q59J86_CHICK
<b>Database</b>	UniProtKB, <u>UniProtKB:Q59J86</u>
<b>Sequence</b>	<u>View sequence</u> ; <a href="#">use as BLAST query sequence</a>
<b>Ref Genome</b>	<u>Homology</u> under POLA ( <span>R. norvegicus</span> <span>M. musculus</span> <span>G. gallus</span> <span>S. cerevisiae</span> <span>S. pombe</span> <span>D. discoideum</span> <span>C. elegans</span> <span>H. sapiens</span> <span>D. rerio</span> )

# The Transcriptional Program of Sporulation in Budding Yeast

S. Chu,\* J. DeRisi,\* M. Eisen, J. Mulholland, D. Botstein,  
P. O. Brown,† I. Herskowitz†

SCIENCE VOL 282 23 OCTOBER 1998

Fig. 4. The global pattern of gene expression during sporulation. (A) The genes that showed the most significant induction or repression during sporulation (20) (a total of 1116) were ordered so that genes with similar expression patterns were grouped together and so that the order roughly reflected the time of first induction or repression. Changes in mRNA concentration for each gene, relative to those in cells immediately prior to transfer to sporulation medium, are shown with a color to represent the quantitative changes as described in Fig. 3. Time after transfer to sporulation medium is shown on the horizontal axis. (B) expression profiles for each set were calculated.



# BABELOMICS<sup>4</sup>

gene expression and functional profiling analysis suite

Upload data

Processing

Expression

Genomic

Functional analysis

Utilities

*anonymous* working on project *default* 0 Kb of 1.00 Gb (0.00%)

## Functional analysis

- Single enrichment analysis
  - Fatigo

Provides significant over-representation of functional annotations by single enrichment analysis

<http://babelomics.bioinfo.cipf.es/>



## Define your comparison

- Id list vs Id list
- Id List vs Rest of genome
- Id List vs Rest of ids contained in your annotations (complementary list)

## Select your data

List 1 :

no data selected.

List 2 :

no data selected. This data format is:

Or go to Upload Data form: [Upload \[idlist\]](#)

## Options

Fisher exact test

Remove duplicates?

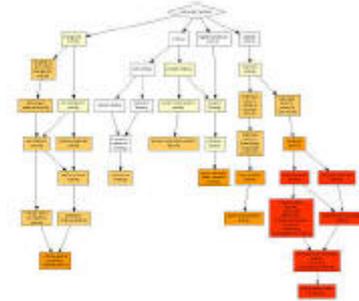
## Databases

Organism

- GO biological process [\[options\]](#)
- GO molecular function [\[options\]](#)
- GO cellular component [\[options\]](#)
- GOSlim GOA [\[options\]](#)
- Interpro [\[options\]](#)
- KEGG pathways [\[options\]](#)
- Reactome [\[options\]](#)
- Biocarta [\[options\]](#)



# GORILLA



*Gene Ontology enRIchment anaLysis and visuaLizAtion tool*

<http://cbl-gorilla.cs.technion.ac.il/>