

Genomeeting 2016

Análisis de datos RNA-Seq



**Anotación en datos de expresión diferencial para
RNA-seq**

Dra. Violeta Larios Serrato (vlarios@wintergenomics.com)

M.C. Liliana Andrés Hernández (landres@wintergenomics.com)



CONTENIDO

- ¿Que es anotación?
 - Biomart
 - Gene Ontology
 - Asignación de rutas metabólicas
- Anotación en Bioconductor
 - Anotación de los datos con la librería AnnotationDbi



CONTENIDO

- ¿Que es anotación?
 - Biomart
 - Gene Ontology
 - Asignación de rutas metabólicas
- Anotación en Bioconductor
 - Anotación de los datos con la librería AnnotationDbi

¿Que es anotación?

La anotación es el proceso de asociar o asignar información biológica a secuencias.

- Identificar elementos.
- Asociar información biológica a los elementos.
- Almacenar información biológica en bases de datos.

Anotación

Secuencia

ID de accesos	Nombre	Símbolo
GI, Genbank, Ensembl	Fosfocinasa	PK1

Rutas Metabólicas	GO	Literatura	Proteína	Posición
KEGG	GO:008998 inmune	Pubmed	P0235687	Cr2:30-178



CONTENIDO

- ¿Que es anotación?
 - **Biomart**
 - Gene Ontology
 - Asignación de rutas metabólicas
- Anotación en Bioconductor
 - Anotación de los datos con la librería AnnotationDbi



El proyecto BiomaRt es impulsado por la comunidad para proporcionar acceso a bases de datos para la investigación.

- Inició en el 2001
- Conectado a más de 40 bases de datos.
- Ayuda a resolver 2 situaciones:

Los nuevos datos sólo tienen sentido cuando se asocian con otros datos.
Las bases de datos se ajustan de forma lenta a nuevos conceptos o tecnologías.

Quantitative data: learning to share

Monya Baker

Nature Methods **9**, 39–41 (2012) | doi:10.1038/nmeth.1815

Published online 28 December 2011

<http://www.biomart.org/>

HOME

TOOLS

WEB SERVICE

COMMUNITY

PUBLICATIONS

NEWS

CREDITS

BioMart

is a community-driven project to provide unified access to distributed research data to facilitate the scientific discovery process.

The BioMart project provides free software and data services to the international scientific community in order to foster scientific collaboration and facilitate the scientific discovery process. The project adheres to the open source philosophy that promotes collaboration and code reuse.



BROWSE
DATA



ID
CONVERSION



SEQUENCE
RETRIEVAL



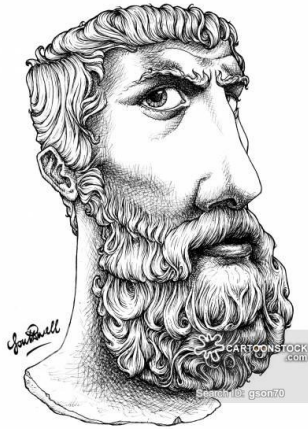
ENRICHMENT
ANALYSIS



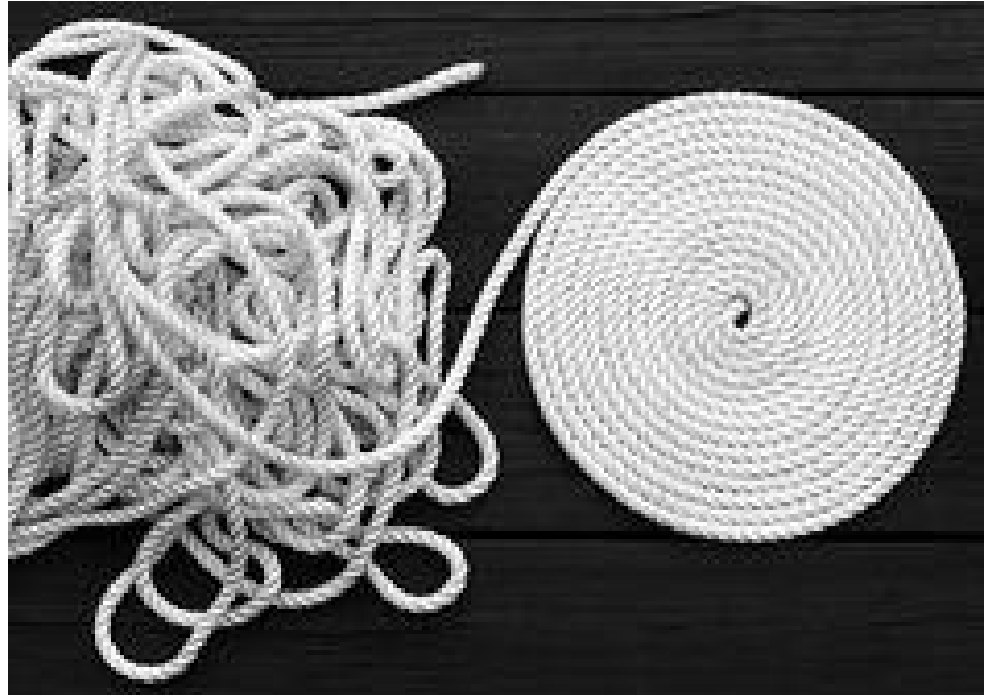
CONTENIDO

- ¿Que es anotación?
 - Biomart
 - Gene Ontology
 - Asignación de rutas metabólicas
- Anotación en Bioconductor
 - Anotación de los datos con la librería AnnotationDbi

Ontología



**Categorías
Clasificar**





Enrichment analysis

Your gene IDs here...

biological process ▾

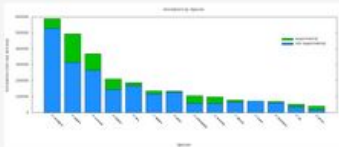
Homo sapiens ▾

Submit

[Advanced options / Help](#)

Powered by [PANTHER](#)

Statistics



Gene Ontology Consortium

Search GO data

Search for terms and gene products...

Search

Ontology

[Filter classes](#)

[Download ontology](#)

Gene Ontology: the framework for the model of biology. The GO defines concepts/classes used to describe gene function, and relationships between these concepts. It classifies functions along three aspects:

molecular function

molecular activities of gene products

cellular component

where gene products are active

Annotations

[Download annotations](#) (standard files)

[Filter and download](#) (customizable files <100k lines)

GO annotations: the model of biology. Annotations are statements describing the functions of specific genes, using concepts in the Gene Ontology. The simplest and most common annotation links one gene to one function, e.g. FZD4 + Wnt signaling pathway. Each statement is based on a specified piece of evidence. [more](#)

40, 000 conceptos

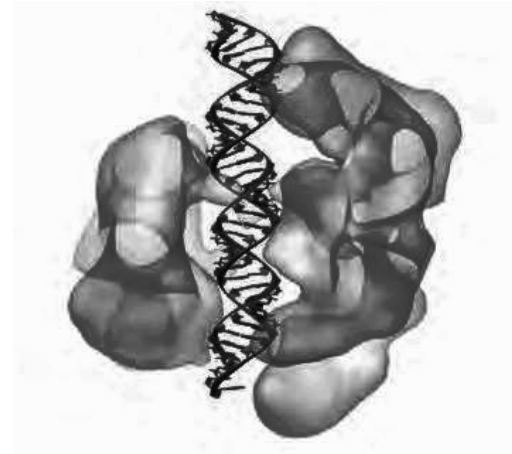
100,000 articulos

National Human Genome
Research Institute

Gene Ontology

Función Molecular: describe actividades que ocurren a nivel molecular que generalmente corresponden a genes individuales.

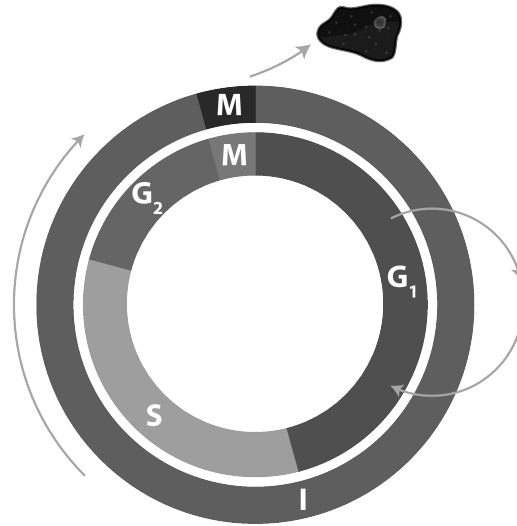
Actividad hidrolasa
Unión a nucleotidos
Transportador
Actividad enzimática



Gene Ontology

Procesos Biológicos: describe procesos acompañados por uno o más ensambles de funciones moleculares.

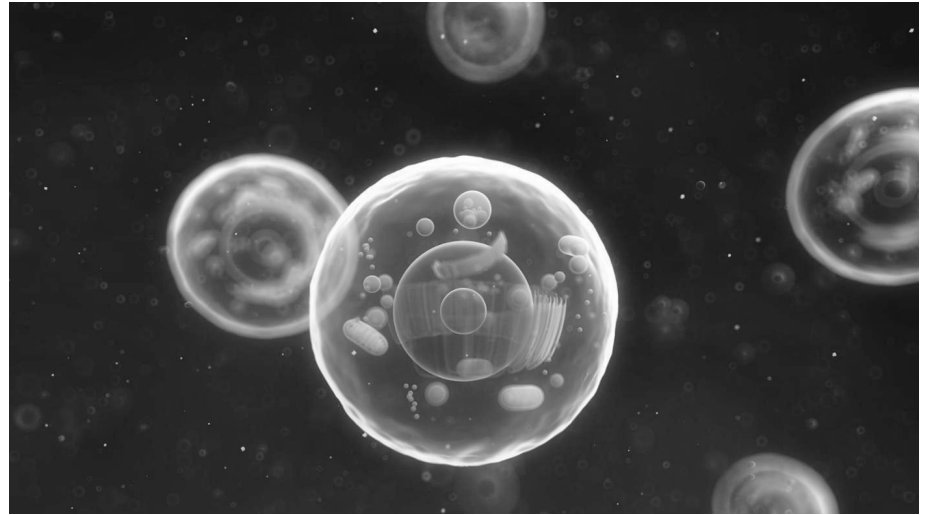
Biogenesis
Metabolismo de DNA o
RNA
Organización celular
Metabolismo de proteínas
Respuesta inmune



Gene Ontology

Componentes Celulares: relacionado a localización o complejos – estructuras subcelulares, localizaciones, complejos macromoleculares.

Membrana
Citosol
Nucleo



Ficha de un GO term

id: GO:0000016

name: lactase activity

namespace: molecular_function

def: "Catalysis of the reaction: lactose + H₂O = D-glucose + D-galactose."
[EC:3.2.1.108]

synonym: "lactase-phlorizin hydrolase activity" BROAD [EC:3.2.1.108]

synonym: "lactose galactohydrolase activity" EXACT [EC:3.2.1.108]



CONTENIDO

- ¿Que es anotación?
 - Biomart
 - Gene Ontology
 - Asignación de rutas metabólicas
- Anotación en Bioconductor
 - Anotación de los datos con la librería AnnotationDbi



KEGG: Kyoto Encyclopedia of Genes and Genomes

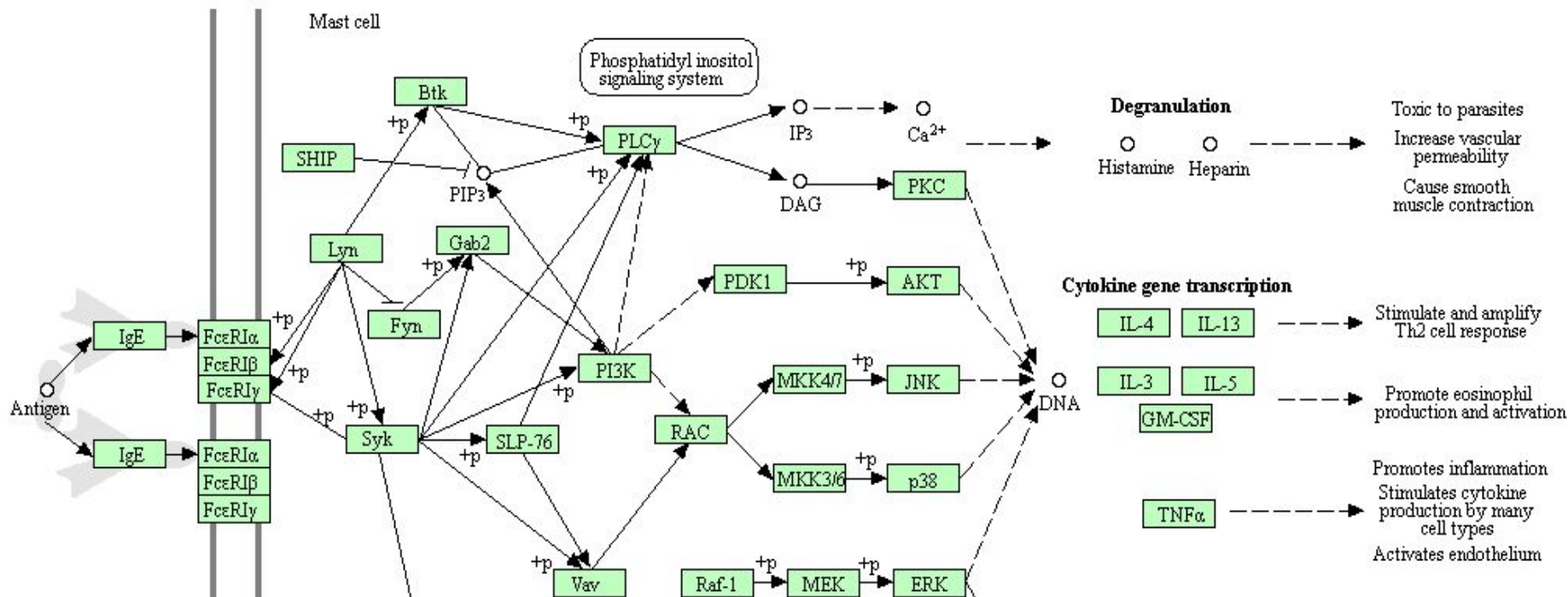
KEGG Database as of 2016/10/15

Systems information

KEGG PATHWAY	Pathway maps, reference (total)	498 (467,014)
KEGG BRITE	Functional hierarchies, reference (total)	231 (167,437)
KEGG MODULE	KEGG modules, reference (total)	764 (380,917)

Mapa de una ruta metabólica en KEGG

FcεRI SIGNALING PATHWAY





Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate

[[Module menu](#) | [Organism menu](#) | [Module entry](#) | [KO/Compound list](#)]

Entry M00001

Name Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate

Definition ([K00844](#),[K12407](#),[K00845](#),[K00886](#),[K08074](#),[K00918](#)) ([K01810](#),[K06859](#),[K13810](#),[K15916](#))
([K00850](#),[K16370](#),[K00918](#)) ([K01623](#),[K01624](#),[K11645](#),[K16305](#),[K16306](#)) [K01803](#) (([K00134](#),[K00150](#))
[K00927](#),[K11389](#)) ([K01834](#),[K15633](#),[K15634](#),[K15635](#)) [K01689](#) ([K00873](#),[K12406](#))

[Ortholog table](#)

[Taxonomy](#)

Type Pathway module

Pathway [map01200](#) [map00010](#)

Reaction [R01786](#),[R02189](#),[R09085](#) [C00267](#) -> [C00668](#)
[R02740](#) [C00668](#) -> [C05345](#)
[R04779](#),[R09084](#) [C05345](#) -> [C05378](#)
[R01070](#) [C05378](#) -> [C00111](#) + [C00118](#)
[R01015](#) [C00111](#) -> [C00118](#)
[R01061](#),[R01063](#) [C00118](#) -> [C00236](#)
[R01512](#) [C00236](#) -> [C00197](#)
[R07159](#) [C00118](#) -> [C00197](#)
[R01518](#) [C00197](#) -> [C00631](#)
[R00658](#) [C00631](#) -> [C00074](#)
[R00200](#) [C00074](#) -> [C00022](#)

Entry	K00844	K0
Name	HK	
Definition	hexokinase [EC:2.7.1.1]	
Pathway	ko00010 Glycolysis / Gluconeogenesis ko00051 Fructose and mannose metabolism ko00052 Galactose metabolism ko00500 Starch and sucrose metabolism ko00520 Amino sugar and nucleotide sugar metabolism ko00521 Streptomycin biosynthesis ko00524 Neomycin, kanamycin and gentamicin biosynthesis ko01200 Carbon metabolism ko04066 HIF-1 signaling pathway ko04910 Insulin signaling pathway ko04930 Type II diabetes mellitus ko04973 Carbohydrate digestion and absorption ko05230 Central carbon metabolism in cancer	
Module	M00001 Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate M00549 Nucleotide sugar biosynthesis, glucose => UDP-glucose	
Disease	H00664 Anemia due to disorders of glycolytic enzymes	
Brite	KEGG Orthology (K0) [BR: ko00001] Metabolism Overview 01200 Carbon metabolism	



CONTENIDO

- ¿Que es anotación?
 - Biomart
 - Gene Ontology
 - Asignación de rutas metabólicas
- Anotación en Bioconductor
 - Anotación de los datos con la librería AnnotationDbi



- Fuente de paquetes de anotación
 - Centradas en genes
 - Centras en el genoma
- Paqueterías curadas
- Creación de paquetes de anotación



Gene centric *AnnotationDbi* packages include:

- Organism level: e.g. *org.Mm.eg.db*.
- Platform level: e.g. *hgu133plus2.db*, *hgu133plus2.probes*, *hgu133plus2.cdf*.
- Homology level: e.g. *hom.Dm.inp.db*.
- System-biology level: *GO.db*

Genome centric *GenomicFeatures* packages include

- Transcriptome level: e.g. *TxDb.Hsapiens.UCSC.hg19.knownGene*, *EnsDb.Hsapiens.v75*.
- Generic genome features: Can generate via *GenomicFeatures*

Data Mining con BiomaRt

BiomaRt es una paquetería de R que provee una interfaz con colecciones de bases de datos implementados en BioMart (<http://www.biomart.org>).

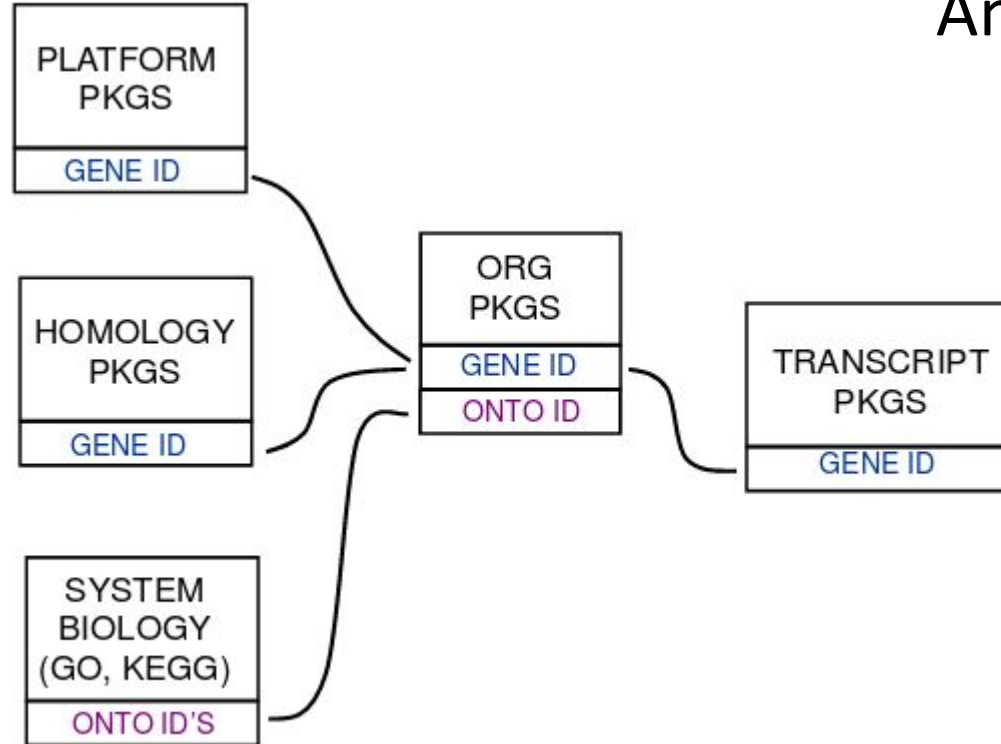




CONTENIDO

- ¿Que es anotación?
 - Biomart
 - Gene Ontology
 - Asignación de rutas metabólicas
- Anotación en Bioconductor
 - Anotación de los datos con la librería AnnotationDbi

AnnotationDbi

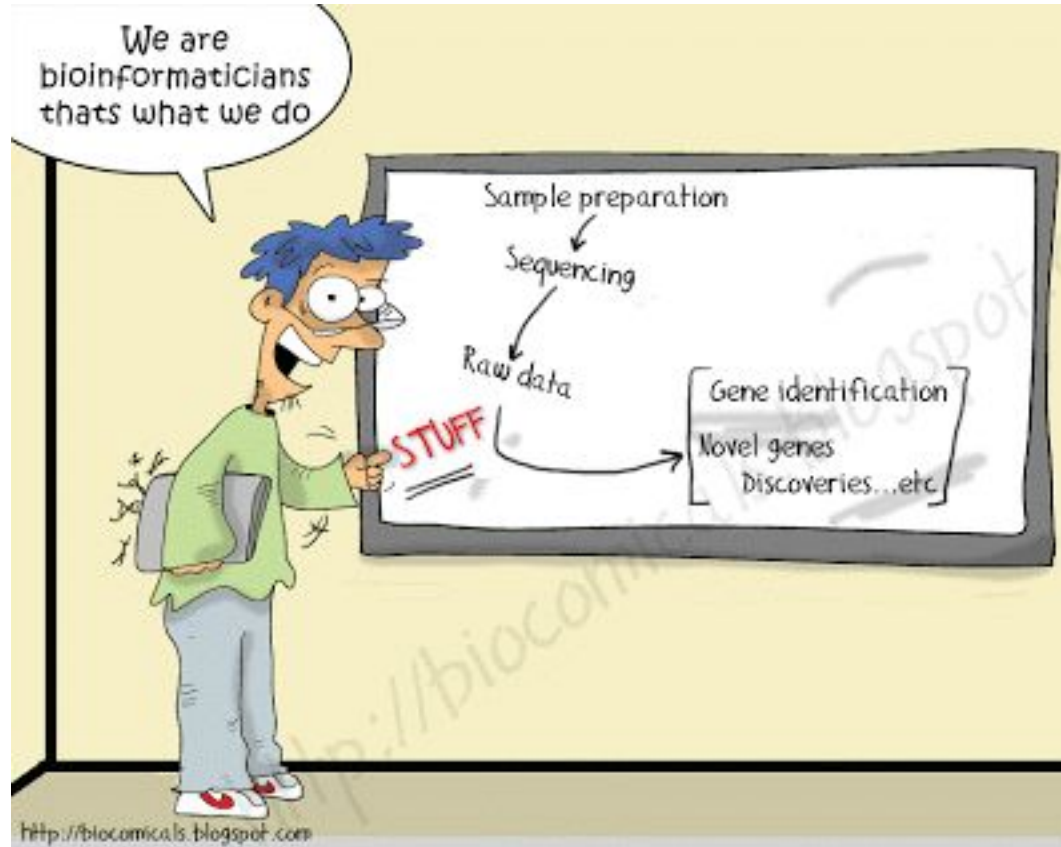


Enlace para la descarga del archivo de anotación en
Bioconductor

<http://bioconductor.org/packages/release/bioc/html/AnnotationDbi.html>



¡Último paso, hagámoslo!



Ruta del script de anotación, abrir en R

/home/alumno/genomeeting-2016/R-scripts

```
#Debemos cargar las paqueterias con la funcion library

library("AnnotationDbi")
library("org.Hs.eg.db")

#subir tabla de resultados
setwd("/Users/Lily/Desktop/EDTEST/")
deseq_Results <- read.table("deseq_Results.tsv", header = TRUE, as.is=TRUE,row.names=1, s

head(deseq_Results)
# Nombres de las columnas del archivo de anotacion
#siempre hay que checar si esta en minusculas o mayusculas
columns(org.Hs.eg.db)

#Anotacion
deseq_Results$symbol <- mapIds(org.Hs.eg.db, keys=row.names(deseq_Results),column="SYMBOL
deseq_Results$entrez <- mapIds(org.Hs.eg.db,keys=row.names(deseq_Results), column="GENE#

# ejecutamos head() para observar si se agrego la anotacion
head(deseq_Results)

#guardando los mejores 100 resultados anotados y con un pvalue ajustado
Ordenados <- deseq_Results[order(deseq_Results$padj),]
```

GRACIAS



WINTER GENOMICS

> **Turning data into forefront knowledge**



CONTACTO

Manizales No.906, Colonia Lindavista.
México, Distrito Federal.

Tel/fax: (52) (55) 5119-0240
(52) (55) 5119-5624

E-mail: contacto@wintergenomics.com

www.wintergenomics.com