

# BIOINFORMATICS COURSE

# Hands on Pathway Analysis

Bioinformatics Course UEB-VHIR  
December 2023

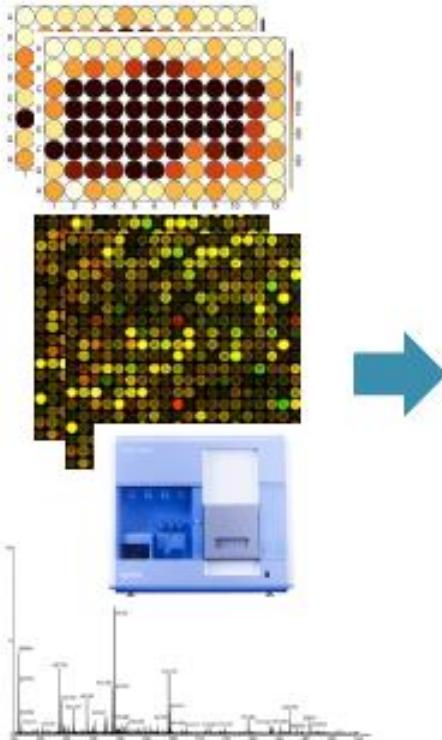
**Esther Camacho<sup>1</sup>, Mireia Ferrer<sup>1</sup>, Álex Sánchez<sup>1,2</sup>, Berta Miró<sup>1</sup>**

<sup>1</sup> Unitat d'Estadística i Bioinformàtica (UEB) VHIR

<sup>2</sup> Departament de Genètica Microbiologia i Estadística, UB

# Hands on Pathway Analysis

What's the biology behind a list of genes ?



Gene	Condition A	Condition B	Condition C	Condition D	Condition E	Condition F	Condition G	Condition H	Condition I	Condition J	Condition K	Condition L	Condition M	Condition N	Condition O	Condition P	Condition Q	Condition R	Condition S	Condition T	Condition U	Condition V	Condition W	Condition X	Condition Y	Condition Z	
Gene 1	High	Low	Medium	Very Low	Medium	High	Very Low																				
Gene 2	Medium	High	Very Low	Medium	High																						
Gene 3	Very Low	Medium	High	Very Low	Medium	High																					
Gene 4	Medium	Very Low	High	Medium	Very Low	High																					
Gene 5	High	Medium	Very Low	High	Medium	Very Low																					



From individual genes

From groups of genes

# Hands on Pathway Analysis

## TABLE OF CONTENTS

1. Resources for functional analysis of gene lists
2. Navigating through different resources
  - a) Browsing the GO
  - b) Pathway visualization in Reactome
3. Case of study: RNA-seq experiment
  - a) Preparation of the gene list
  - b) Functional annotation of individual genes
  - c) Overrepresentation analysis of top DEG (g:Profiler / Reactome analysis tool)
  - d) Visualization and downstream analyses
4. A workflow in R (clusterProfiler)

# **Resources for functional analysis**

# Resources for functional analysis

## Annotation datasets

- Functional annotations can be very diverse: molecular functions, pathways (genes that work together to carry out a biological process), interactions, gene regulation, involvement in disease...
- Collections of gene sets:

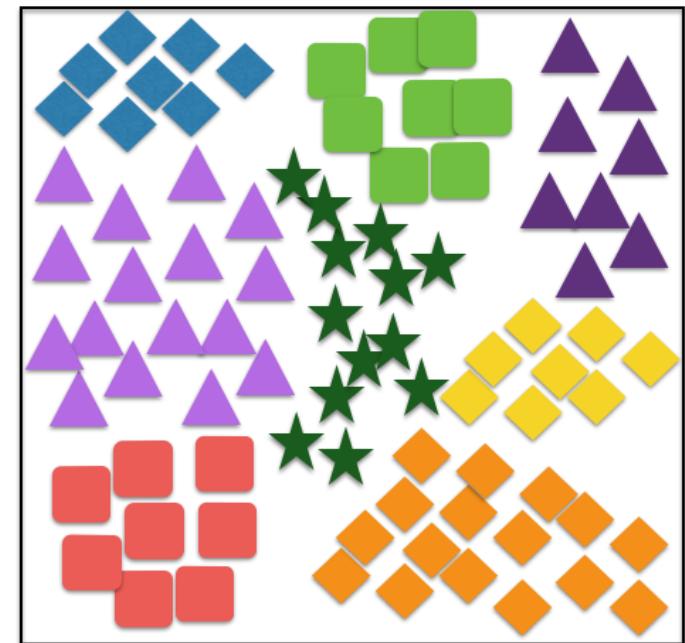
- Gene Ontology (GO)

- CC Cellular Components
- BP Biological Processes
- MF Molecular Function

- Pathways

- KEGG
- Reactome
- WikiPathways
- PantherDB

All known genes in a species  
(categorized into groups)



# Resources for functional analysis

## Annotation datasets

- Functional annotations can be very diverse: molecular functions, pathways (genes that work together to carry out a biological process), interactions, gene regulation, involvement in disease...

- Collections of gene sets:

- Gene Ontology (GO)

- CC Cellular Components
    - BP Biological Processes
    - MF Molecular Function

- Pathways (signaling, metabolic...)

- KEGG
    - Reactome
    - WikiPathways
    - PantherDB
    - BioCyc/HumanCyc

- Disease-related

- DisGeNet
    - Disease Ontology

- Cell/Tissue markers

- CellMarker

- Chromosomal location

- Regulatory motifs

- TRANSFAC (Transcr Fact)
    - miRTarBase (miRNA-target)

# Resources for functional analysis

## Annotation datasets

- [MSigDB](#)

MSigDB Home

- Human Collections
  - About
  - Browse
  - Search
  - Investigate
  - Gene Families
- Mouse Collections
  - About
  - Browse
  - Search
  - Investigate
- Help



**MSigDB**

Molecular Signatures  
Database

### Overview

The Molecular Signatures Database (MSigDB) is a resource of tens of thousands of annotated gene sets for use with GSEA software, divided into Human and Mouse collections. From this web site, you can

- ▶ **Examine** a gene set and its annotations. See, for example, the [HALLMARK\\_APOPTOSIS](#) human gene set page.
- ▶ **Browse** gene sets by name or collection.
- ▶ **Search** for gene sets by keyword.
- ▶ **Investigate** gene sets:
  - ▶ **Compute overlaps** between your gene set and gene sets in MSigDB.
  - ▶ **Categorize** members of a gene set by gene families.
  - ▶ **View the expression profile** of a gene set in a provided public expression compendia.
  - ▶ Investigate the gene set in the online **biological network repository NDEX**
  - ▶ **Download** gene sets.

### License Terms

GSEA and MSigDB are available for use under these license terms.

Please [register](#) to download the GSEA software and the MSigDB gene sets, and to use our web tools. After registering, you can log in at any time using your email address. Registration is free. Its only purpose is to help us track usage for reports to our funding agencies.

### Current Version

Human MSigDB v2023.2.Hs updated October 2023. [Release notes](#).

Mouse MSigDB v2023.2.Mm updated October 2023. [Release notes](#).

## Molecular Signatures Database

### Human Collections

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

**C5**

**C1** positional gene sets corresponding to human chromosome cytogenetic bands.

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

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

**C7** immunologic signature gene sets represent cell states and perturbations within the immune system.

**C3** regulatory target gene sets based on gene target predictions for microRNA seed sequences and predicted transcription factor binding sites.

**C8** cell type signature gene sets curated from cluster markers identified in single-cell sequencing studies of human tissue.

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

### Mouse Collections

**MH** mouse-ortholog hallmark gene sets are versions of gene sets in the MSigDB Hallmarks collection mapped to their mouse orthologs.

**M3** regulatory target gene sets based on gene target predictions for microRNA seed sequences and predicted transcription factor binding sites.

**M1** positional gene sets corresponding to mouse chromosome cytogenetic bands.

**M5** ontology gene sets consist of genes annotated by the same ontology term.

# Resources for functional analysis

## Web tools and software

- g:Profiler - <http://biit.cs.ut.ee/gprofiler/index.cgi>
- Reactome - <https://reactome.org/>
- DAVID - <https://david.ncifcrf.gov/tools.jsp>
- GeneMANIA - <http://www.genemania.org/>
- WebGestalt - <http://webgestalt.org/>
- AmiGO - <http://amigo.geneontology.org/amigo>
- QuickGO - <https://www.ebi.ac.uk/QuickGO/>
- ReviGO - <http://revigo.irb.hr/>
- GSEA - <http://software.broadinstitute.org/gsea/index.jsp>
- ClueGO - <http://apps.cytoscape.org/apps/cluego>
- PANTHER - <http://pantherdb.org/>
- GOrilla – <http://cbl-gorilla.cs.technion.ac.il/>

# Resources for functional analysis

## R packages

- clusterProfiler - <http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html>
- GAGE/Pathview - <http://www.bioconductor.org/packages/release/bioc/html/gage.html>
- SPIA - <https://www.bioconductor.org/packages/release/bioc/html/SPIA.html>
- WGCNA - <https://cran.r-project.org/web/packages/WGCNA/index.html>



Search:

[Home](#)   [Install](#)   [Help](#)   [Developers](#)   [About](#)

[Home](#) » BiocViews

### All Packages

#### Bioconductor version 3.18 (Release)

Autocomplete biocViews search:

▶ StatisticalMethod (831)
▶ Technology (1467)
WorkflowManagement (1)
▼ WorkflowStep (1219)
▶ Alignment (98)
Annotation (138)
BatchEffect (59)
ExperimentalDesign (27)
GenomeBrowsers (2)
MultipleComparison (179)
Normalization (159)
▼ Pathways (200)
BioCarta (6)
GO (73)

#### Packages found under Pathways:

Rank based on number of downloads: lower numbers are more frequently downloaded.

Show All ▾ entries	Package	Maintainer	Title	Rank
	<a href="#">KEGGREST</a>	Bioconductor Package Maintainer	Client-side REST access to the Kyoto Encyclopedia of Genes and Genomes (KEGG)	16
	<a href="#">annotate</a>	Bioconductor Package Maintainer	Annotation for microarrays	23
	<a href="#">edgeR</a>	Yunshun Chen, Gordon Smyth, Aaron Lun, Mark Robinson	Empirical Analysis of Digital Gene Expression Data in R	24
	<a href="#">enrichplot</a>	Guangchuang Yu	Visualization of Functional Enrichment Result	35
	<a href="#">clusterProfiler</a>	Guangchuang Yu	A universal enrichment tool for interpreting omics data	38

<http://bioconductor.org/packages/release/BiocViews.html#Pathways>

# Resources for functional analysis

## Considerations when choosing a tool

- Type of analysis performed
  - Annotation
  - Enrichment analysis (Overrepresentation vs GSEA)
  - Network-based analyses
  - Coexpression, ...
- Statistical test performed, algorithms used, multiple testing adjustment
- Source of information
  - Available genesets, databases
  - Updated? Curated?
  - Organisms supported
- Input data accepted
  - Genes, miRNA, SNPs, metabolomics
  - Flexibility in managing different types of Ids
- Output, visualization, postprocessing

# Hands on Pathway Analysis

## TABLE OF CONTENTS

1. Resources for functional analysis of gene lists
2. Navigating through different resources
  - a) Browsing the GO
  - b) Pathway visualization in Reactome
3. Case of study: RNA-seq experiment
  - a) Preparation of the gene list
  - b) Functional annotation of individual genes
  - c) Overrepresentation analysis of top DEG (g:Profiler / Reactome analysis tool)
  - d) Visualization and downstream analyses
4. A workflow in R (clusterProfiler)

# Browsing the Gene Ontology

# Browsing the Gene Ontology

## THE GENE ONTOLOGY RESOURCE

The mission of the GO Consortium is to develop a comprehensive, **computational model of biological systems**, ranging from the molecular to the organism level, across the multiplicity of species in the tree of life.

The Gene Ontology (GO) knowledgebase is the world's largest source of information on the functions of genes. This knowledge is both human-readable and machine-readable, and is a foundation for computational analysis of large-scale molecular biology and genetics experiments in biomedical research.

Search GO term or Gene Product in AmiGO ...



Any  Ontology  Gene Product



The network of biological classes describing the current best representation of the "universe" of biology. The molecular functions, cellular locations, and processes gene products may carry out.

- [GO Ontology Overview](#)
- [Browse in AmiGO](#)
- [Download](#)



### ANNOTATION

Statements, based on specific, traceable scientific evidence, asserting that a specific gene product is a real exemplar of a particular GO class.

- [GO Annotations Overview](#)
- [Browse in AmiGO](#)
- [Download](#)



GO Causal Activity Model (GO-CAM) provides a structured framework to link standard GO annotations into a more complete model of a biological system.

- [GO-CAM Overview](#)
- [Browse GO-CAMs](#)
- [Download](#)



Tools to curate, browse, search, visualize and download both the ontology and annotations. Bioinformatic Guides (Notebooks) and simple API access to integrate GO into your research.

- [GO Tools Overview](#)
- [GO APIs Guide](#)
- [GO Github](#)

Current release 2020-12-08: 44,117 GO terms | 7,963,579 annotations  
1,562,091 gene products | 4743 species (see statistics)

### GO Enrichment Analysis ?

Powered by PANTHER

Your gene IDs here...

biological process

Homo sapiens

Examples

Launch

Note: can use UniProt ID/AC, Gene Name, Gene Symbol, MGI ID

# Browsing the Gene Ontology

## Queries that can be done with the GO

- AmiGO / QuickGO
- Hierarchy
- Search for a term like apoptosis
- Search for the annotations of a given gene
- Do enrichment analysis of a gene list

# Browsing the Gene Ontology

- Functional annotation of individual genes with QuickGO  
<https://www.ebi.ac.uk/QuickGO/annotations>

The screenshot shows the QuickGO search interface. At the top, there is a navigation bar with links to Help, Contact, API, and Basket. The main title is "Gene Ontology and GO Annotations". Below the title, a search bar contains the query "LACTB". A blue button labeled "Search" is to the right of the search bar. The results are displayed in two main sections: "Terms" and "Gene products". The "Terms" section is currently empty. The "Gene products" section lists several entries, each with a protein ID and a description. A red oval highlights the link "Show all 2,112 results", which is followed by a large blue arrow pointing towards it. To the right of the "Gene products" section, there is a sidebar with a blue background and white text, which includes the text "area of interest" and "GO slim annotations".

LACTB

Search

Terms

Gene products

A0A2Y9NW76 Beta-lactamase-related domain-containing protein

A0A2Y9NZZ5 Beta-lactamase-related domain-containing protein

A0A341AMC2 Beta-lactamase-related domain-containing protein

A0A2K5LU97 Beta-lactamase-related domain-containing protein 5 annotations

A0A2B4RPP6 Beta-lactamase-related domain-containing protein 1 annotations

Show all 2,112 results

area of interest

GO slim annotations

# Browsing the Gene Ontology

- Functional annotation of individual genes with QuickGO  
<https://www.ebi.ac.uk/QuickGO/annotations>

The screenshot shows the QuickGO website interface. At the top, there is a navigation bar with links to EMBL-EBI, Services, Research, Training, About us, and EMBL-EBI. Below the navigation bar is a search bar with the text 'Search'. On the left side, there is a sidebar with filters for UniProtKB (selected), TrEMBL (3), Swiss-Prot (3), Type, MiRNA (8), Protein (6), Organism (selected as Homo Sapiens), and Proteome status (Reference Proteomes (Gene Centric, Canonical) (3)). The main content area displays the results for the gene product 'LACTB' with 14 results. The results table has columns for Database, ID, Name, Type, Taxon, and Annotations. The first row shows 'UniProtKB H0YNN5 Beta-lactamase-related domain-containing protein PROTEIN Homo sapiens'. The second row shows 'UniProtKB P83111 Serine beta-lactamase-like protein LACTB, mitochondrial PROTEIN Homo sapiens 22 annotations'. The third row shows 'UniProtKB Q53H82 Endoribonuclease LACTB2 PROTEIN Homo sapiens 15 annotations'. The fourth row shows 'UniProtKB A8MY62 Putative beta-lactamase-like 1 PROTEIN Homo sapiens'. The fifth row shows 'UniProtKB H0Y608 Beta-lactamase-related domain-containing protein PROTEIN Homo sapiens 1 annotations'. The sixth row shows 'RNACentral URS00007C010D\_9606 Homo sapiens (human) non-protein coding LACTB2-AS1:4 miRNA Homo sapiens'. A blue arrow points to the '22 annotations' button in the second row.

## Results for gene product "LACTB"

UniProtKB

14 results

TrEMBL (3)

Swiss-Prot (3)

Type

MiRNA (8)

Protein (6)

Organism

Homo Sapiens

Proteome status

Reference Proteomes  
(Gene Centric, Canonical) (3)

Database	ID	Name	Type	Taxon	Annotations
UniProtKB	H0YNN5	Beta-lactamase-related domain-containing protein	PROTEIN	Homo sapiens	
UniProtKB	P83111	Serine beta-lactamase-like protein LACTB, mitochondrial	PROTEIN	Homo sapiens	22 annotations
UniProtKB	Q53H82	Endoribonuclease LACTB2	PROTEIN	Homo sapiens	15 annotations
UniProtKB	A8MY62	Putative beta-lactamase-like 1	PROTEIN	Homo sapiens	
UniProtKB	H0Y608	Beta-lactamase-related domain-containing protein	PROTEIN	Homo sapiens	1 annotations
RNACentral	URS00007C010D_9606	Homo sapiens (human) non-protein coding LACTB2-AS1:4	miRNA	Homo sapiens	

# Browsing the Gene Ontology

- Functional annotation of individual genes with QuickGO



[Help](#) | [Contact](#) | [API](#) | [Basket](#)

[Search](#)

## GO annotations

[Taxon](#) | [Gene Products](#) | [GO terms](#) | [References](#) | [Aspect](#) | [Evidence](#) | [Extension](#) | [More](#) | [Clear all](#)

[Annotations](#)

[Statistics](#)

[Customise](#) | [Export](#) | 22 annotations

- Biological Process?
- Molecular function?
- Cellular localization?

Gene Product	Symbol	Qualifier	GO Term	Evidence	Reference	With / From	Taxon	Assigned By	Annotation Extension			
UniProtKB:P83111	LACTB	enables	GO:0008233	ECO:0000318	IBA	PANHER:PTN000532814 more...	9606 Homo sapiens	GO_Central				
UniProtKB:P83111	LACTB	involved_in	GO:0019216			regulation of lipid metabolic process	ECO:0000318	IBA	PANHER:PTN000532814 more...	9606 Homo sapiens	GO_Central	
UniProtKB:P83111	LACTB	is_active_in	GO:0005739			mitochondrion	ECO:0000318	IBA	PANHER:PTN000532814 more...	9606 Homo sapiens	GO_Central	
UniProtKB:P83111	LACTB	involved_in	GO:0006508			proteolysis	ECO:0000318	IBA	PANHER:PTN000532814 more...	9606 Homo sapiens	GO_Central	
UniProtKB:P83111	LACTB	enables	GO:0008233			peptidase activity	ECO:0000314	IDA	PANHER:PTN000532814 more...	9606 Homo sapiens	UniProt	
UniProtKB:P83111	LACTB	involved_in	GO:0019216			regulation of lipid metabolic process	ECO:0000314	IDA	PANHER:PTN000532814 more...	9606 Homo sapiens	UniProt	
UniProtKB:P83111	LACTB	located_in	GO:0005739			mitochondrion	ECO:0000314	IDA	PANHER:PTN000532814 more...	9606 Homo sapiens	UniProt	
UniProtKB:P83111	LACTB	involved_in	GO:0006508			proteolysis	ECO:0000314	IDA	PANHER:PTN000532814 more...	9606 Homo sapiens	UniProt	
UniProtKB:P83111	LACTB	enables	GO:0016787			hydrolase activity	ECO:0007322	IEA	GO_REF:0000043	UniProtKB-KW-KW-0378	9606 Homo sapiens	UniProt
UniProtKB:P83111	LACTB	located_in	GO:0005739			mitochondrion	ECO:0007322	IEA	GO_REF:0000043	UniProtKB-KW-KW-0496	9606 Homo sapiens	UniProt
UniProtKB:P83111	LACTB	enables	GO:0008233			peptidase activity	ECO:0007322	IEA	GO_REF:0000043	UniProtKB-KW-KW-0645	9606 Homo sapiens	UniProt
UniProtKB:P83111	LACTB	involved_in	GO:0006508			proteolysis	ECO:0007322	IEA	GO_REF:0000043	UniProtKB-KW-KW-0645	9606 Homo sapiens	UniProt
UniProtKB:P83111	LACTB	involved_in	GO:0006629			lipid metabolic process	ECO:0007322	IEA	GO_REF:0000043	UniProtKB-KW-KW-0443	9606 Homo sapiens	UniProt
UniProtKB:P83111	LACTB	enables	GO:0042802			identical protein binding	ECO:0000265	IEA	GO_REF:0000107	UniProtKB-D3ZFJ6 more...	9606 Homo sapiens	Ensembl
UniProtKB:P83111	LACTB	located_in	GO:0005739			mitochondrion	ECO:0007322	IEA	GO_REF:000044	UniProtKB-SubCell:SL-0173	9606 Homo sapiens	UniProt

# Hands on Pathway Analysis

## TABLE OF CONTENTS

1. Resources for functional analysis of gene lists
2. Navigating through different resources
  - a) Browsing the GO
  - b) Pathway visualization in Reactome
3. Case of study: RNA-seq experiment
  - a) Preparation of the gene list
  - b) Functional annotation of individual genes
  - c) Overrepresentation analysis of top DEG (g:Profiler / Reactome analysis tool)
  - d) Visualization and downstream analyses
4. A workflow in R (clusterProfiler)

# Browsing the Reactome Pathways database

# Browsing the Reactome Pathways

<https://reactome.org/>

The screenshot shows the Reactome homepage. At the top, there's a search bar with placeholder text "e.g. O95631, NTN1, signaling by EGFR, glucose" and a "Go!" button. Below the search bar are four main tool icons: "Pathway Browser" (with a server icon), "Analyze Data" (with a bar chart icon), "ReactomeFLViz" (with a network icon), and "Documentation" (with a document icon). A black banner below these tools says "USE REACTOME GRAPH DATABASE IN YOUR PROJECT" with a "LEARN MORE" button. The bottom section features a "Why Reactome" box with a magnifying glass icon, containing text about the database's purpose and a note about its Chinese mirror site. To the right of this box is a "Tweets" section with a Twitter icon, showing tweets from the @Reactome account.

Tutorials: <https://reactome.org/userguide>

[https://www.ebi.ac.uk/training/online/courses/reactome-tools-for-analysing-biological-pathways/#vf-tabs\\_section--overview](https://www.ebi.ac.uk/training/online/courses/reactome-tools-for-analysing-biological-pathways/#vf-tabs_section--overview)

# Browsing the Reactome Pathways

## Navigating through Reactome Pathway database

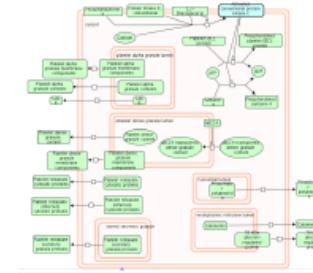
Reactome is...

Free, online, open-source curated database of pathways and reactions in human biology

Authored by expert biologists, maintained by Reactome editorial staff (curators)

Mapped to cellular compartment

Extensively cross-referenced



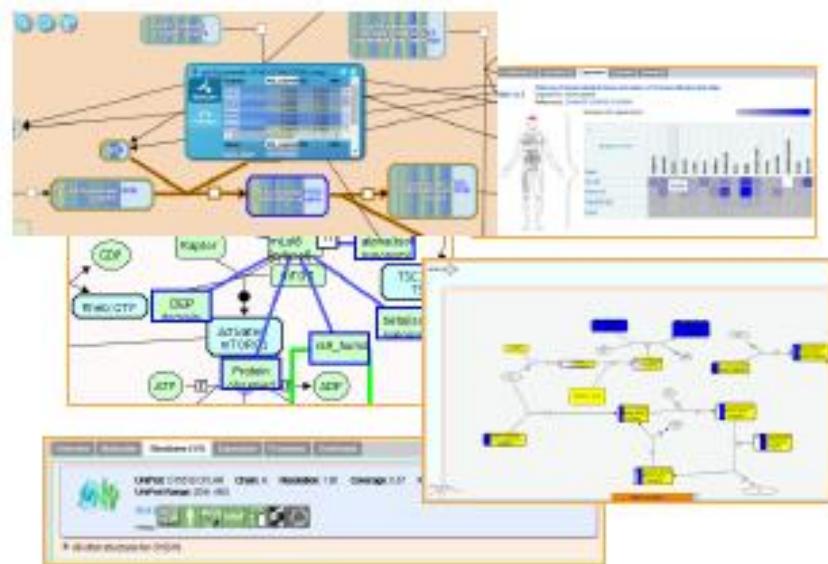
Entrez Gene



# Browsing the Reactome Pathways

## Reactome Tools

- Interactive Pathway Browser
- Analysis
  - Over-representation
  - Pathway topology
  - Expression overlay
- Molecular Interaction overlay
- Species Comparison



# Browsing the Reactome Pathways

## The Pathway Browser

Home
Species
Analyze Data
Video Tour
Layout

In-Diagram Search
Fit to Page
Open Diagram
Illustrations
Export
Settings
Zoom/Move

Thumbnail
Details Panel

Pathway Hierarchy Panel

Event Hierarchy:

- Cell Cycle
- Cell-Cell communication
- Cellular responses to external stimuli
- Chromatin organization
- Circadian Clock
- Developmental Biology
- Digestion and absorption
- Disease
- DNA Repair
- DNA Replication
- Extracellular matrix organization
- Metabolism of RNA
- Mitophagy
- Mitochondria
- Muscle contraction
- Neuronal System
- Hemostasis
- Signal Transduction
- Gene expression (transcription)
- Disease
- Metabolism of proteins
- Vesicle-mediated transport
- Cell-Cell communication
- Cellular responses to external stimuli
- Protein localization
- Reproduction
- Signal Transduction
- Transport of small molecules
- Vesicle-mediated transport

Pathway Viewport

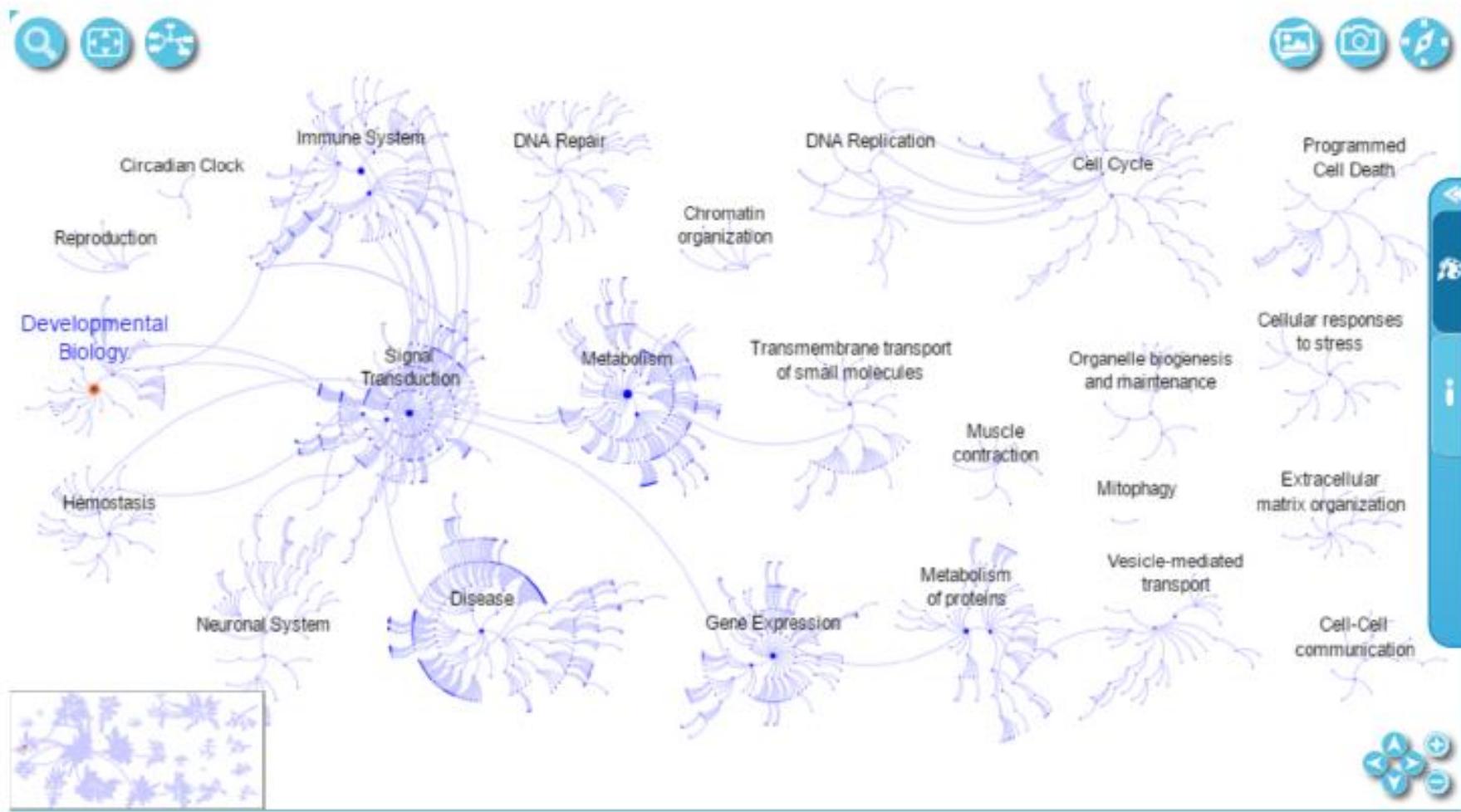
Thumbnail

Description    Molecules    Structures    Expression    Analysis    Downloads

Displays details when you select an item in the Pathway Browser. For example, when a reaction is selected, shows details including the input and output molecules, summary and references containing supporting evidence. When relevant, shows details of the catalyst, regulators, preceding and following events.

# Browsing the Reactome Pathways

## Pathway Overview



# Browsing the Reactome Pathways

## Navigating in the Pathway Browser

Home button

The screenshot shows the Reactome Pathway Browser interface. On the left, there is a tree view of biological pathways under the heading "Event Hierarchy". A blue box highlights the "Signal Transduction" category. An orange arrow points from this category to a specific reaction in the main diagram. The main area displays a network diagram of "Signal Transduction" with various nodes and interactions. A blue box highlights a node in the center-right labeled "Notch Signaling Pathway". An orange arrow points from this node to a callout box containing the text "Click reaction to open". Another orange box highlights the "Details here" section at the bottom right of the callout. The bottom navigation bar includes tabs for Description, Molecules, Structures, Expression, Analytic, and Downloads.

Event Hierarchy:

- Cell Cycle
- Cell-Cell communication
- Cellular responses to external stimuli
- Chromatin organization
- Circadian Clock
- Developmental Biology
- Digestion and absorption
- Disease
- DNA Repair
- DNA Replication
- Extracellular matrix organization
- Gene expression (Transcription)
- Hainogenesis
- Immune System
- Metabolism
- Metabolism of proteins
- Metabolism of RNA
- Mitophagy
- Muscle contraction
- Neuronal System
- Organellar biogenesis and maintenance
- Programmed Cell Death
- Reproduction
- Signal Transduction**
- Signaling by Receptor Tyrosine Kinases
- Signaling by TGF-beta family members
- Signaling by GPCR
- Signaling by NOTCH**
- Pre-NOTCH Expression and Processing
- Signaling by NOTCH1
- Activated NOTCH1 Transmits Signal to the Nucleus
  - NOTCH1 binds DLL1
  - NOTCH1 binds DLL4
  - NOTCH1 binds JAG1
  - NOTCH1 binds JAG2
  - Ubiquitination of DLL4AG ligands upon bind
  - Complex of NOTCH1 with its ligand is cleaved
  - NEXT1 is cleaved to produce NICD1
  - NICD1 traffics to nucleus
  - DTX binds NOTCH1
  - Contactin-1 (CNTN1) binds NOTCH1
  - Cleavage of CNTN1-NOTCH1 complex rela

Pathways for: Homo sapiens

Analysis: Tour Layout

Click pathway to select, double-click to open diagram

Click reaction to open

Details here

The Notch Signaling Pathway (NSP) is a highly conserved pathway for cell-cell communication. NSP is involved in the regulation of cellular differentiation, proliferation, and specification. For example, it is utilized to continuously renewing adult tissues such as blood, skin, and gut epithelium not only to maintain stem cells in a proliferative, pluripotent, and undifferentiated state but also to direct the cellular progeny to adopt different developmental cell fates. Analogously, it is used during embryonic development to create fine-grained patterns of differentiated cells, notably during neurogenesis where the NSP controls patches such as that of the ventral midbrain and dorsal forebrain. This process is known as lateral inhibition: a molecular mechanism whereby individual cells inhibit their neighbors from doing the same. The NSP has been adopted by several other biological processes to divide the growing embryo into regular blocks called somites which are generated from a molecular oscillator in the pre-somatic mesoderm. The Notch receptor is synthesized in the rough endoplasmic reticulum as a single polypeptide chain, creating a heteromeric mature receptor consisting of non-covalently associated intracellular specific ligands. Following ligand activation and further proteolytic cleavage, the receptor is cleaved in the trans-golgi network to the cell surface ready to signal where it regulates gene expression.

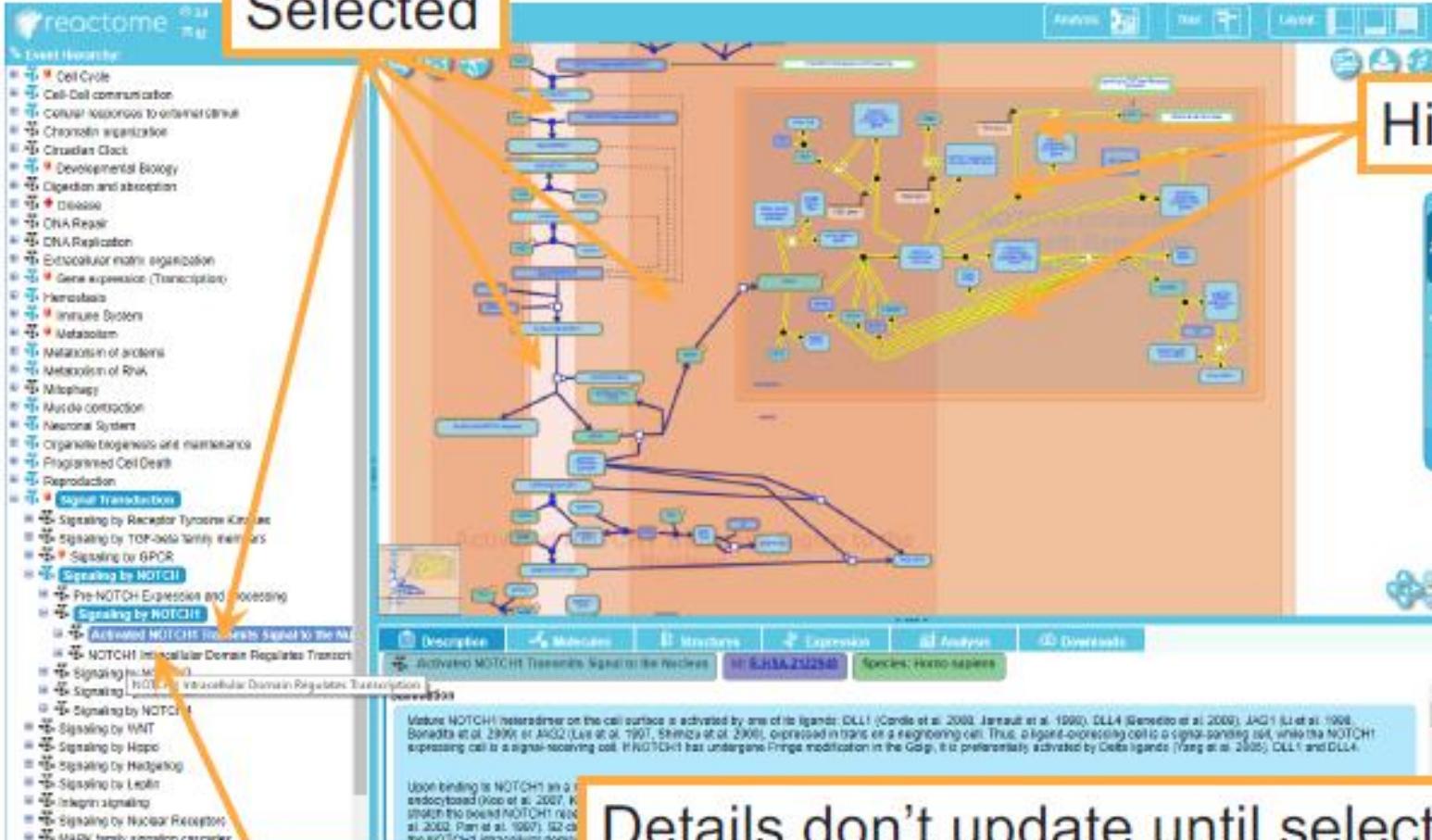
Background literature references...

# Browsing the Reactome Pathways

## Navigating in the Pathway Browser

**Selected**

- Selected
- Not Selected



**Highlights**

**Details don't update until selection**

**Hover mouse**

# Browsing the Reactome Pathways

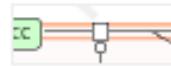
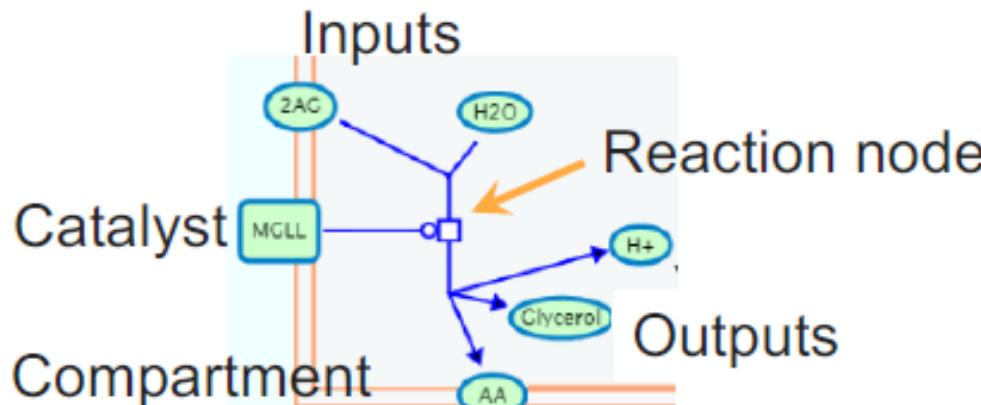
## The Pathway Browser - Pathway Diagrams

Ovals are small molecules (or sets of)

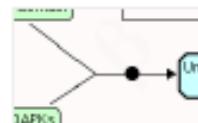
Green boxes are proteins,

Blue are complexes,

Blue with double-boundary are sets



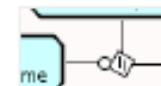
Transition



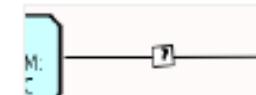
Binding



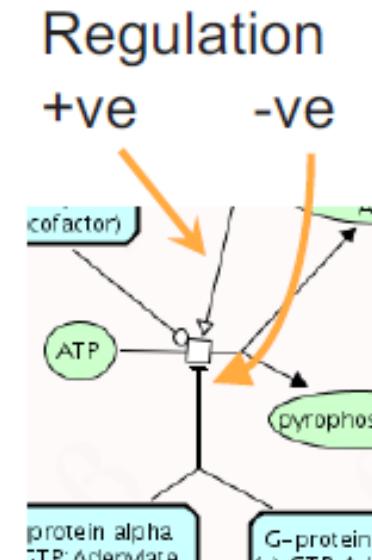
Dissociation



Omitted

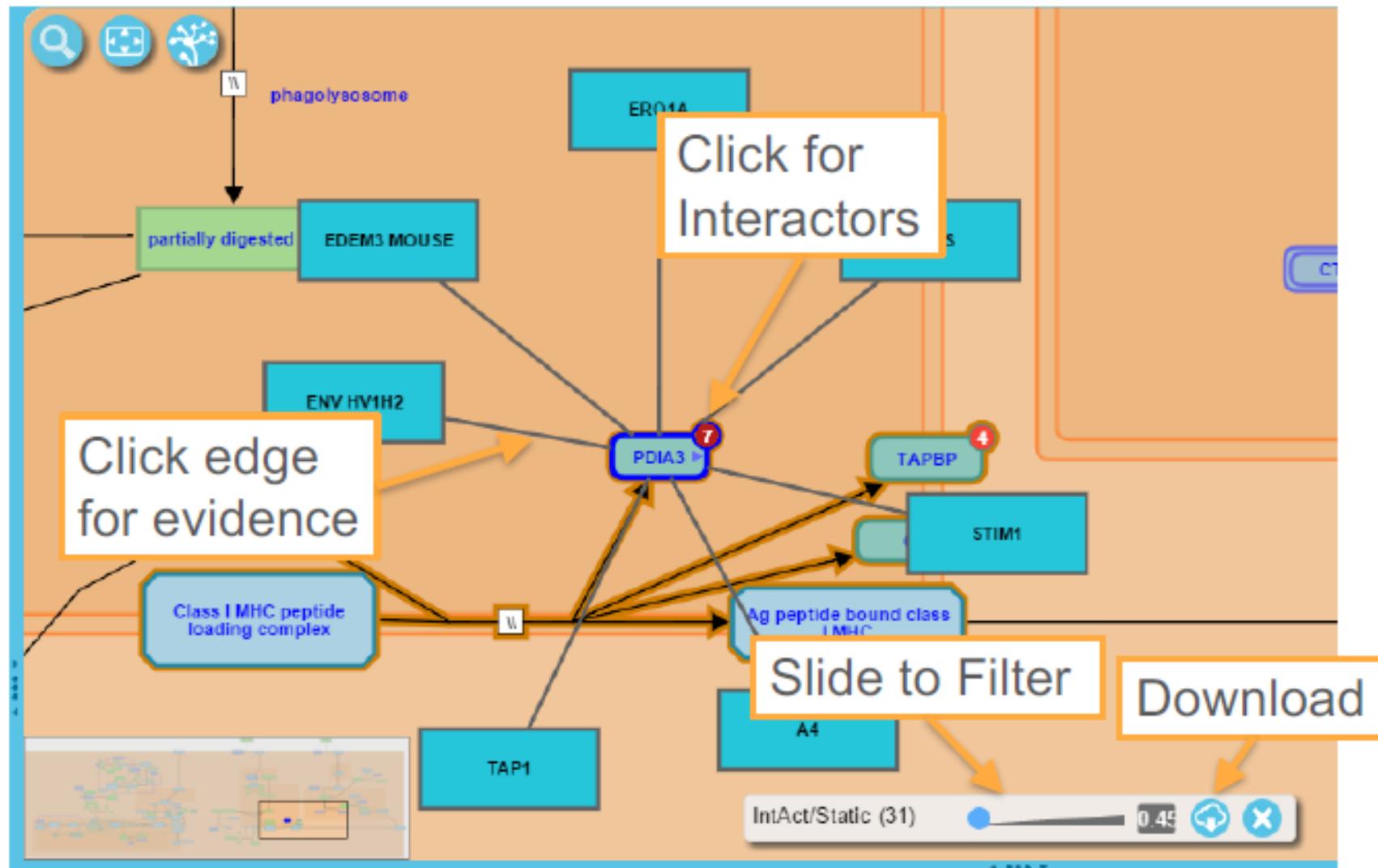


Uncertain



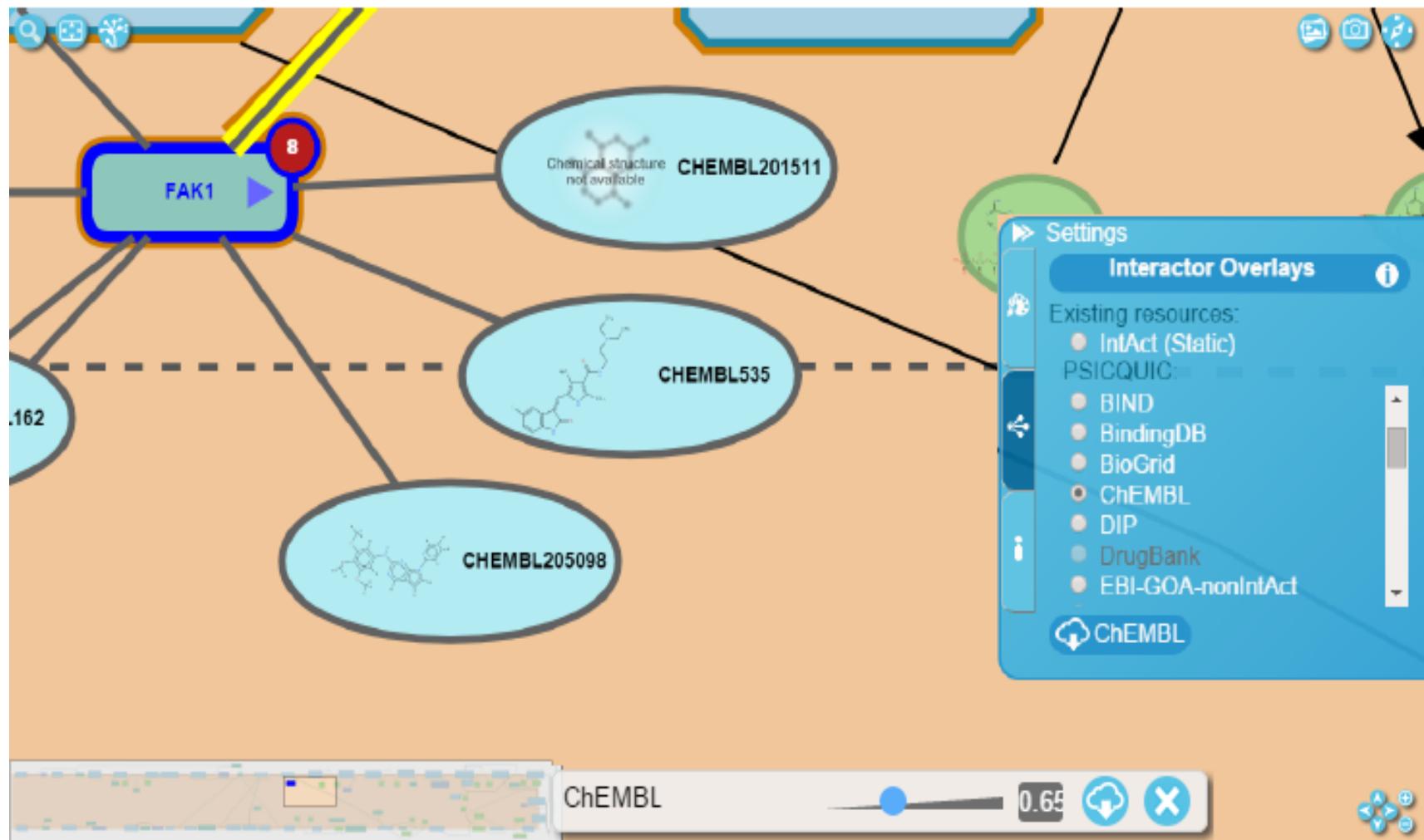
# Browsing the Reactome Pathways

## Interactors



# Browsing the Reactome Pathways

## Molecular Interaction Overlay – Set source



# Browsing the Reactome Pathways

## Show Illustration

REACTOME Pathways for: Homo sapiens.

Event Hierarchy:

- Effects of PIP2 hydrolysis
- Response to elevated platelet cytosolic Ca<sup>2+</sup>
- Formation of Fibrin Clot (Clotting Cascade)
- Dissolution of Fibrin Clot
- Cell surface interactions at the vascular wall
- Factors involved in megakaryocyte development
- Immune System
- Mitophagy
- Metabolism
- Metabolism of proteins
- Muscle contraction
- Neuronal System
- Organelle biogenesis and maintenance
- Mitochondrial biogenesis
- Mitochondrial translation
- Assembly of the primary cilium
- Anchoring of the basal body to the plasma membrane
- Cargo trafficking to the periciliary membrane
- Intraflagellar transport
- ATAT acetylates microtubules
- HDAC6 deacetylates microtubules
- Programmed Cell Death
- Apoptosis
- Regulated Necrosis
- Reproduction
- Signal Transduction

**CILIOGENESIS**

**ASSEMBLY OF THE PRIMARY CILIJUM: INTRAFLAGELLAR TRANSPORT AND CARGO TRAFFICKING**

Description Molecules Structures Expression Analysis Downloads

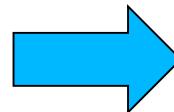
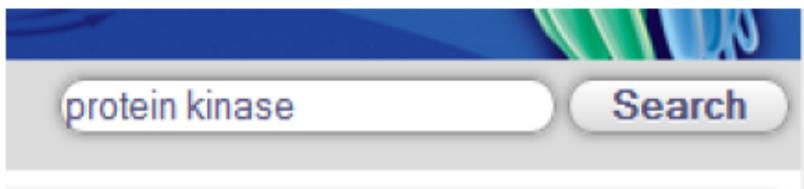
Analysis results are shown here when an analysis has been run. To start an analysis, click on the Analyse Data button in the top bar.

# Browsing the Reactome Pathways

## Searching the Reactome

### Simple text search

The simple text search tool is located top right of the Home page. Several other Reactome pages include this search tool. To search Reactome type a word, phrase or identifier in the search box:



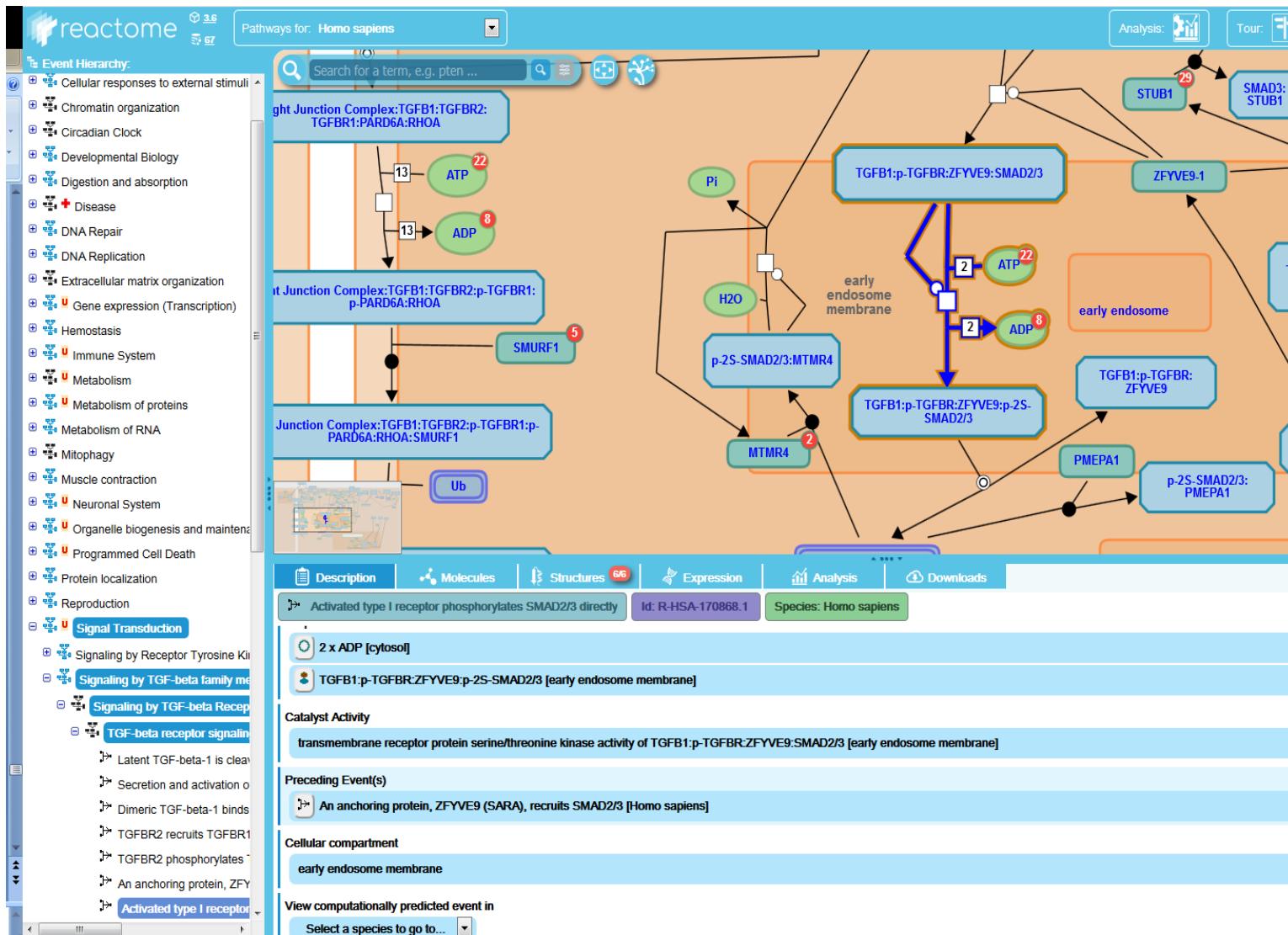
Different types of results (Pathway, Reactions, Prot...)

# Browsing the Reactome Pathways

## Searching the Reactome

1. Find the reaction Activated type I receptor phosphorylates SMAD2/3 directly. What pathway does it belong to?
2. In which cellular compartment does this reaction take place?
3. What is the GO molecular function associated with the catalyst?
4. What references verify this reaction?
5. Is this reaction predicted to occur in *Canis familiaris*? In *Saccharomyces cerevisiae*?
6. Is this event likely to occur in liver?
7. Are 3D structures available for TGF-beta1?

# Browsing the Reactome Pathways



# Hands on Pathway Analysis

## TABLE OF CONTENTS

1. Resources for functional analysis of gene lists
2. Navigating through different resources
  - a) Browsing the GO
  - b) Pathway visualization in Reactome
3. Case of study: RNA-seq experiment
  - a) Preparation of the gene list
  - b) Functional annotation of individual genes
  - c) Overrepresentation analysis of top DEG (g:Profiler / Reactome analysis tool)
  - d) Visualization and downstream analyses
4. A workflow in R (clusterProfiler)

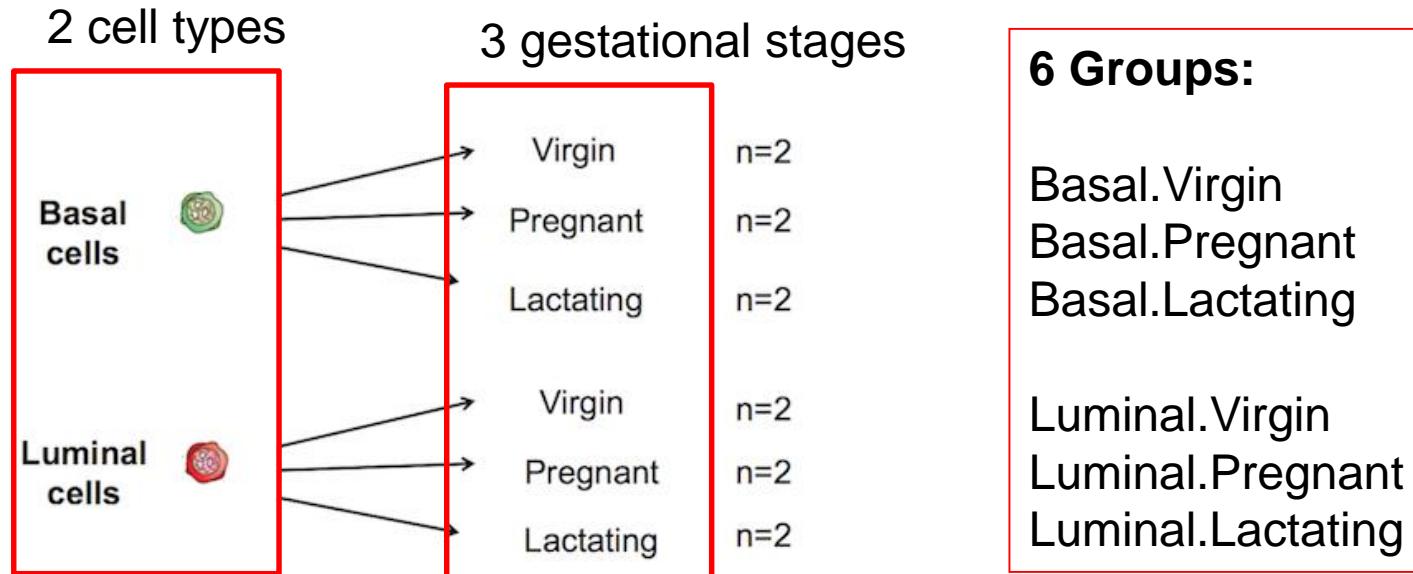


# Case of Study

# Case of study

## Study overview

Dataset: RNA-seq data of mouse mammary gland ([GSE60450](#))



We analyzed the differential gene expression in the following comparisons:

- Basal.Pregnant vs. Basal.Lactating
- Luminal.Pregnant vs. Luminal.Lactating

# Case of study

## Study overview

**Objective:** Our experiment in RNA-seq returned some genes differentially expressed in **basal.pregnant vs basal.lactate cells**. We want now to extract some biological meaning from these lists, such as **which biological processes or pathways the differentially expressed genes are implicated in**.

- What type(s) of test/study can you perform to get this information?
- What is the input data in such tests?
- Do you know some databases where you can find information on biological processes/pathways?

# Case of study

## Preparation of the gene list

- Prepare a Gene List with the “top” up-regulated genes and one with the “top” down-regulated genes according to the following statistical criteria:
  - **Up-regulated genes:** genes with an **adjusted p-value < 0.05 AND logFC > 2**
  - **Down-regulated genes:** genes with an **adjusted p-value < 0.05 AND logFC < -2**
- Under these criteria, how many genes are up-regulated? And how many down-regulated?
- What is the **Ensembl ID** of your top **most up/down-regulated gene?** And its **Gene Symbol / UniProt ID** (SwissProt)?

Tip: You can use g:Convert tool from g:Profiler for ID conversion

# Case of study

## Preparation of the gene list

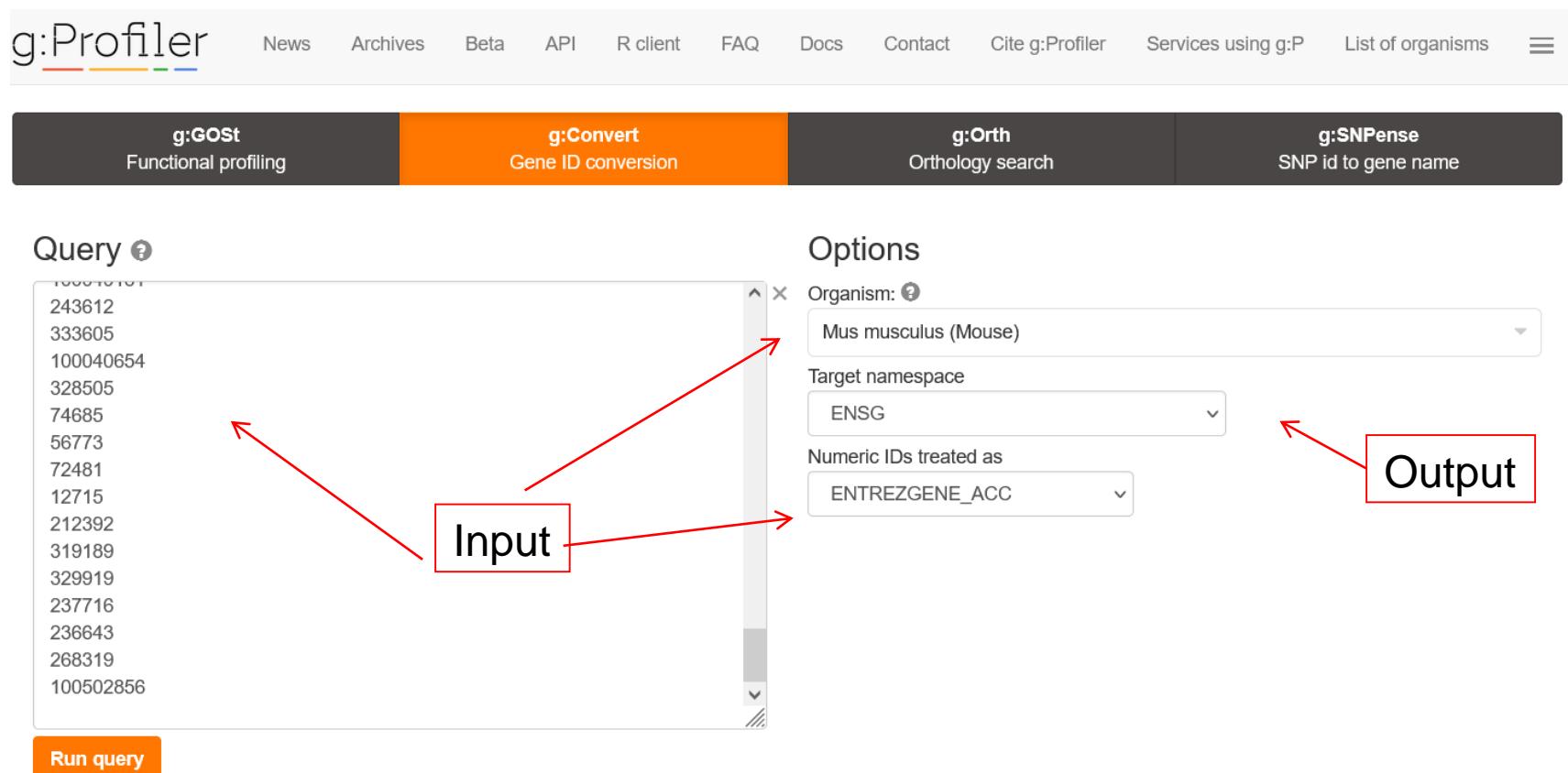
- Prepare a Gene List with the “top” up-regulated genes and one with the “top” down-regulated genes according to the following statistical criteria:
  - **Up-regulated genes:** genes with an **adjusted p-value < 0.05 AND logFC > 2**
  - **Down-regulated genes:** genes with an **adjusted p-value < 0.05 AND logFC < -2**
- Under these criteria, how many genes are up-regulated? And how many down-regulated? **294 up, 101 down**
- What is the **Ensembl ID** of your top **most up/down-regulated gene?** And its **Gene Symbol / UniProt ID** (SwissProt)?

Tip: You can use g:Convert tool from g:Profiler for ID conversion

# Case of study

## Preparation of the gene list

### ID mapping with g:Profiler – g:Convert



**Query** ⓘ

243612  
333605  
100040654  
328505  
74685  
56773  
72481  
12715  
212392  
319189  
329919  
237716  
236643  
268319  
100502856

**Run query**

**Input**

**Options**

Organism: ⓘ  
Mus musculus (Mouse)

Target namespace  
ENSG

Numeric IDs treated as  
ENTREZGENE\_ACC

**Output**

**g:Convert** enables to convert between various gene, protein, microarray probe and numerous other types of namespaces. We provide at least 40 types of IDs for more than 60 species. The 98 different namespaces supported for human include Ensembl, Refseq,

Illumina, Entrezgene and Uniprot identifiers. All namespaces are obtained through matching them via Ensembl gene identifiers as a reference.

# Case of study

## Preparation of the gene list

### ID mapping with g:Profiler

[Export to CSV](#) [Show query URL](#)  [Show short link](#) 

initial alias	converted alias	icon	name	icon	description	namespace
381290	ENSMUSG00000026463		Atp2b4		ATPase, Ca++ transporting, plasma membrane 4 [Source:MGI Symbol;Acc:MGI:88111]	ENTREZGENE_ACC
226101	ENSMUSG00000048612		Myof		myoferlin [Source:MGI Symbol;Acc:MGI:1919192]	ENTREZGENE_ACC
16012	ENSMUSG00000023046		Igfbp6		insulin-like growth factor binding protein 6 [Source:MGI Symbol;Acc:MGI:96441]	ENTREZGENE_ACC
16669	ENSMUSG00000020911		Krt19		keratin 19 [Source:MGI Symbol;Acc:MGI:96693]	ENTREZGENE_ACC
231991	ENSMUSG00000053007		Creb5		cAMP responsive element binding protein 5 [Source:MGI Symbol;Acc:MGI:2443973]	ENTREZGENE_ACC
211577	None		None		None	
270150	ENSMUSG00000070306		Ccdc153		coiled-coil domain containing 153 [Source:MGI Symbol;Acc:MGI:2448587]	ENTREZGENE_ACC
21953	ENSMUSG00000031097		Tnni2		troponin I, skeletal, fast 2 [Source:MGI Symbol;Acc:MGI:105070]	ENTREZGENE_ACC
12992	ENSMUSG00000061388		Csn1s2b		casein alpha s2-like B [Source:MGI Symbol;Acc:MGI:105312]	ENTREZGENE_ACC
170761	ENSMUSG00000032105		Pdzd3		PDZ domain containing 3 [Source:MGI Symbol;Acc:MGI:2429554]	ENTREZGENE_ACC
67111	ENSMUSG00000029413		Naaa		N-acylethanolamine acid amidase [Source:MGI Symbol;Acc:MGI:1914361]	ENTREZGENE_ACC
232016	ENSMUSG00000037973		Itprid1		ITPR interacting domain containing 1 [Source:MGI Symbol;Acc:MGI:2685304]	ENTREZGENE_ACC
76123	ENSMUSG00000027883		Gpsm2		G-protein signalling modulator 2 (AGS3-like, C. elegans) [Source:MGI Symbol;Acc:MGI:1...	ENTREZGENE_ACC
18019	ENSMUSG00000027544		Nfatc2		nuclear factor of activated T cells, cytoplasmic, calcineurin dependent 2 [Source:MGI Sy...	ENTREZGENE_ACC
67971	ENSMUSG00000014846		Tppp3		tubulin polymerization-promoting protein family member 3 [Source:MGI Symbol;Acc:M...]	ENTREZGENE_ACC
74134	ENSMUSG00000040703		Cyp2s1		cytochrome P450, family 2, subfamily s, polypeptide 1 [Source:MGI Symbol;Acc:MGI:192...	ENTREZGENE_ACC
12833	ENSMUSG0000001119		Col6a1		collagen, type VI, alpha 1 [Source:MGI Symbol;Acc:MGI:88459]	ENTREZGENE_ACC
66234	ENSMUSG00000031604		Msmo1		methylsterol monooxygenase 1 [Source:MGI Symbol;Acc:MGI:1913484]	ENTREZGENE_ACC
12834	ENSMUSG00000020241		Col6a2		collagen, type VI, alpha 2 [Source:MGI Symbol;Acc:MGI:88460]	ENTREZGENE_ACC
21827	ENSMUSG00000028047		Thbs3		thrombospondin 3 [Source:MGI Symbol;Acc:MGI:98739]	ENTREZGENE_ACC
235505	ENSMUSG00000046186		Cd109		CD109 antigen [Source:MGI Symbol;Acc:MGI:2445221]	ENTREZGENE_ACC
59031	ENSMUSG00000036599		Chst12		carbohydrate sulfotransferase 12 [Source:MGI Symbol;Acc:MGI:1929064]	ENTREZGENE_ACC
57266	ENSMUSG00000021508		Cxcl14		chemokine (C-X-C motif) ligand 14 [Source:MGI Symbol;Acc:MGI:1888514]	ENTREZGENE_ACC
242122	ENSMUSG00000051076		Vtcn1		V-set domain containing T cell activation inhibitor 1 [Source:MGI Symbol;Acc:MGI:30396...	ENTREZGENE_ACC
320311	ENSMUSG00000047496		Rnf152		ring finger protein 152 [Source:MGI Symbol;Acc:MGI:2443787]	ENTREZGENE_ACC
107769	ENSMUSG00000038623		Tm6sf1		transmembrane 6 superfamily member 1 [Source:MGI Symbol;Acc:MGI:1933209]	ENTREZGENE_ACC
329738	ENSMUSG00000049565		Aknad1		AKNA domain containing 1 [Source:MGI Symbol;Acc:MGI:3584453]	ENTREZGENE_ACC
12722	ENSMUSG00000040212		Fmp2		epithelial membrane protein 2 [Source:MGI Symbol;Acc:MGI:1009720]	ENTREZGENE_ACC

# Case of study

## Functional annotation of individual genes

### What is known of gene X?

- Get information for one of the genes in some database (Pubmed, UniProt, NCBI-Gene...)
- Look for GO terms (CC, MF, BP) associated to your favorite gene:

- Tip: Search by gene/protein symbol in QuickGO:

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

# Case of study

## Functional annotation of individual genes

*Note that doing this for every one of your genes may take you some long time...*

PubMed ▾ GNAQ

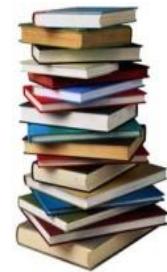
Create RSS Create alert Advanced

Format: Summary ▾ Sort by: Most Recent ▾ Per page: 20 ▾ Send to ▾

See 271 articles about **GNAQ** gene function  
See also: [GNAQ G protein subunit alpha\\_q](#) in the Gene database  
[gnaq](#) in *Homo sapiens* *Mus musculus* *Rattus norvegicus* [All 160 Gene records](#)  
See also: [38 tests](#) for **GNAQ** in the Genetic Testing Registry

**Search results**  
Items: 1 to 20 of 370

<< First < Prev Page  of 19 Next > Last >>



# Case of study

## Functional annotation of individual genes

- Genes usually don't act on their own
- Extracting information at the “group” level may be more meaningful and facilitate interpretation of the processes involved in the condition of study



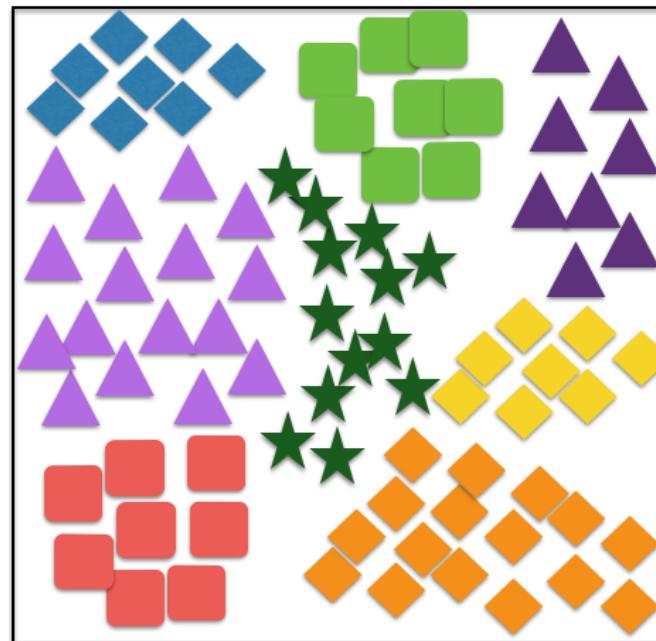
# Overrepresentation analysis

# Case of study

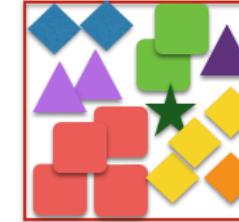
## Overrepresentation analysis

- Perform a statistical test (eg. Fisher exact test) to determine whether any gene sets (GO terms/ pathways) are “surprisingly” enriched in your gene list

All known genes in a species  
(categorized into groups)



Gene list



# Case of study

## Overrepresentation analysis

### What do we need?

Our list of genes



Selected genes from topTable

The universe from which comes our list of genes



Remaining genes in microarray/genome

Gene sets of annotations



From annotation database  
(eg. *.gmt* file /*.gaf* file for GO)

<http://geneontology.org/docs/download-go-annotations/>

Structure of relationships between annotation terms



From annotation database  
(eg. *.obo* file for GO)

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

# Overrepresentation analysis

## **g:Profiler – g:GOST**

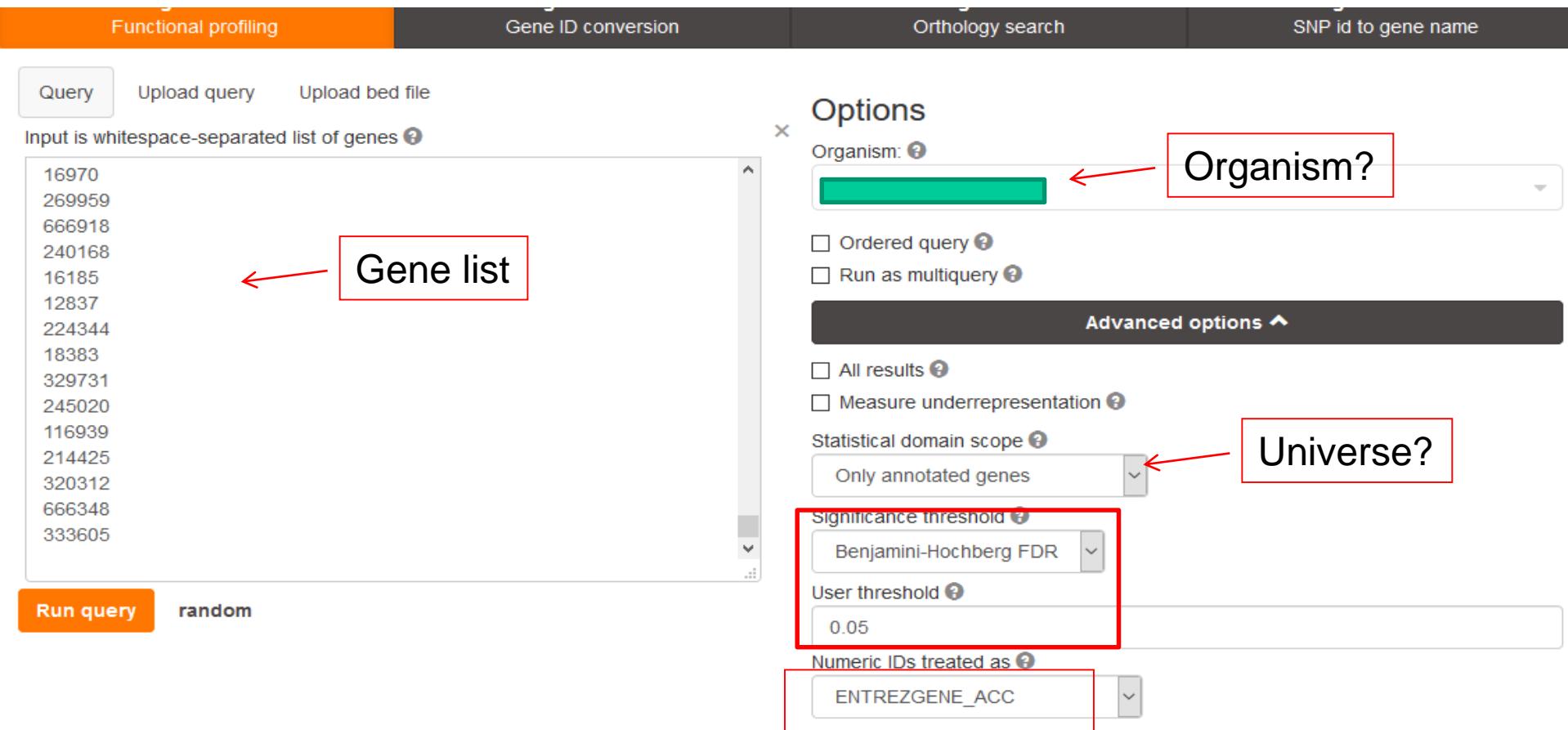
- ✓ Allows to perform enrichment analysis across different annotation databases
- ✓ Accepts different types of IDs as input
- ✓ P-value adjustments for multiple comparisons
- ✓ Source of evidence (experimental, computational, ...)
- ✓ Also for ranked lists

The screenshot shows the 'Data sources' section of the g:Profiler interface. At the top, there are buttons for 'select all', 'clear all', and 'Show data versions'. Below this, there are several sections of checked checkboxes:

- Gene Ontology**: GO molecular function, GO cellular component, GO biological process, No electronic GO annotations (with a question mark icon).
- biological pathways**: KEGG, Reactome, WikiPathways.
- regulatory motifs in DNA**: TRANSFAC, miRTarBase.
- protein databases**: Human Protein Atlas, CORUM.
- Human phenotype ontology**: HP.

# Overrepresentation analysis

## g:Profiler – g:GOSt



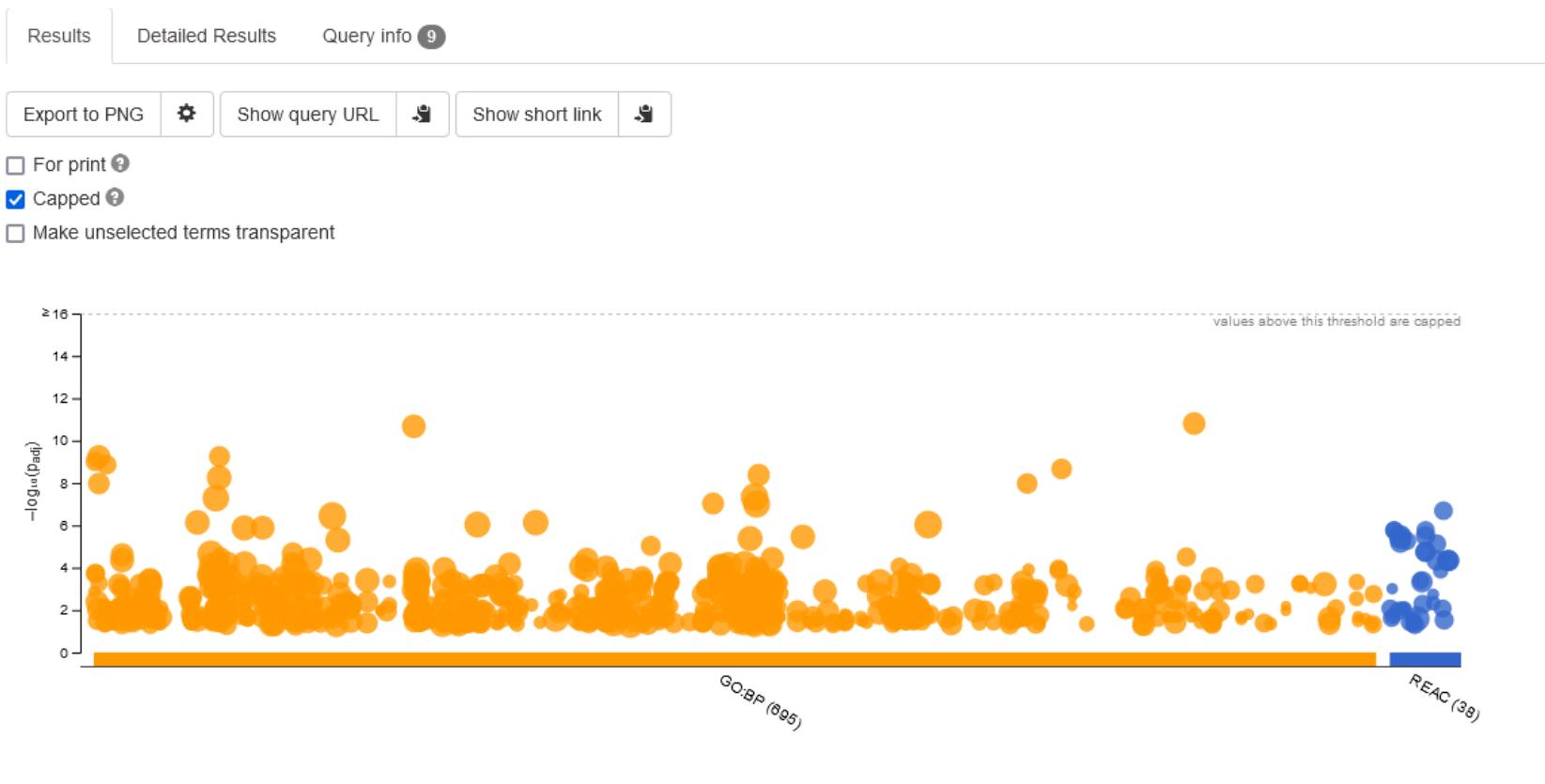
The screenshot shows the g:Profiler interface with the following components:

- Top Navigation:** Functional profiling, Gene ID conversion, Orthology search, SNP id to gene name.
- Left Panel:** A "Query" tab is selected. It includes "Upload query" and "Upload bed file" buttons. A text input field says "Input is whitespace-separated list of genes". Below it is a list of gene IDs:
  - 16970
  - 269959
  - 666918
  - 240168
  - 16185
  - 12837
  - 224344
  - 18383
  - 329731
  - 245020
  - 116939
  - 214425
  - 320312
  - 666348
  - 333605
- Bottom Left Buttons:** Run query (orange), random.
- Right Panel (Options):** An "Options" panel is open.
  - Organism?**: A dropdown menu labeled "Organism" with a red box around it and an arrow pointing to it from the left.
  - Universe?**: A dropdown menu labeled "Statistical domain scope" with a red box around it and an arrow pointing to it from the right.
  - Significance threshold**: A dropdown menu set to "Benjamini-Hochberg FDR" with a red box around it.
  - User threshold**: A text input field set to "0.05" with a red box around it.
  - Numeric IDs treated as**: A dropdown menu set to "ENTREZGENE\_ACC" with a red box around it.

# Overrepresentation analysis

## g:Profiler – g:GOSt

### g:Profiler graphical output



# Overrepresentation analysis

## g:Profiler – g:GOST

### g:Profiler graphical output

#### g:Profiler specific vocabulary:

Query (Q): genes in my gene list

Term (T): tested gene-set (pathway)

Q&T (common genes): number of genes that overlap between my gene list and the tested gene-sets

(Q&T) / T: overlap normalized by the gene-set size

p-value: FDR (corrected for multiple hypothesis testing)



#### > Gene-sets (pathways/processes) that were found enriched

source	term name	term ID	n. of terms	n. of query genes	n. of common genes	p-value	size of the overlap	significance	> Genes in input list
BP	negative regulation of biological process	GO:0048519	4340	391	127	2.29e-03			
BP	negative regulation of cellular process	GO:0048523	3985	391	119	2.17e-03			
BP	single-organism process	GO:0044699	13478	391	305	1.49e-02			
BP	death	GO:0016265	1945	391	69	3.41e-03			
BP	developmental process	GO:0032502	5560	391	150	1.82e-02			
BP	single-organism developmental process	GO:0044767	5476	391	148	2.00e-02			
BP	multicellular organismal development	GO:0007275	4696	391	132	1.16e-02			
BP	anatomical structure development	GO:0048656	4925	391	138	7.09e-03			
BP	system development	GO:0048731	4114	391	122	2.11e-03			
BP	tissue development	GO:0009888	1763	391	63	8.57e-03			
BP	organ development	GO:0048513	2992	391	95	3.79e-03			
BP	single-organism cellular process	GO:0044763	12236	391	282	2.32e-02			
BP	cell death	GO:0002519	1941	391	69	3.17e-03			
BP	programmed cell death	GO:0012591	1850	391	68	1.36e-03			
BP	apoptotic process	GO:0006915	1830	391	68	7.76e-04			
BP	regulation of cell death	GO:0010941	1467	391	61	7.24e-05			
BP	regulation of programmed cell death	GO:0043067	1395	391	60	2.83e-05			
BP	regulation of apoptotic process	GO:0042981	1383	391	60	2.06e-05			

# Overrepresentation analysis

## GORilla

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

- ✓ Direct visualization of enriched terms in hierarchy
- ✓ Needs to specify the universe
- ✓ Also for ranked lists
- ✓ Only for GO

# Overrepresentation analysis

## GORILLA

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

### Step 1: Choose organism

Mus musculus

### Step 2: Choose running mode

- Single ranked list of genes     Two unranked lists of genes (target and background lists)

### Step 3: Paste a ranked list of gene/protein names

Names should be separated by an <ENTER>. The preferred format is gene symbol. Other supported formats are: gene and protein RefSeq, Uniprot, Unigene and Ensembl.

Target set:

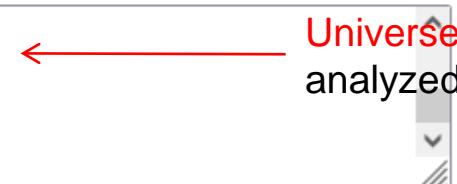
Myix2  
F630040K05Rik  
Cers3  
Gm10658  
Fer114  
Chac1  
Stra61



Or upload a file:  No se ha seleccionado ningún archivo.

Background set:

Gdf11  
Rcor1  
Smarcb1  
Zfp551  
Kbtbd6  
Psen2



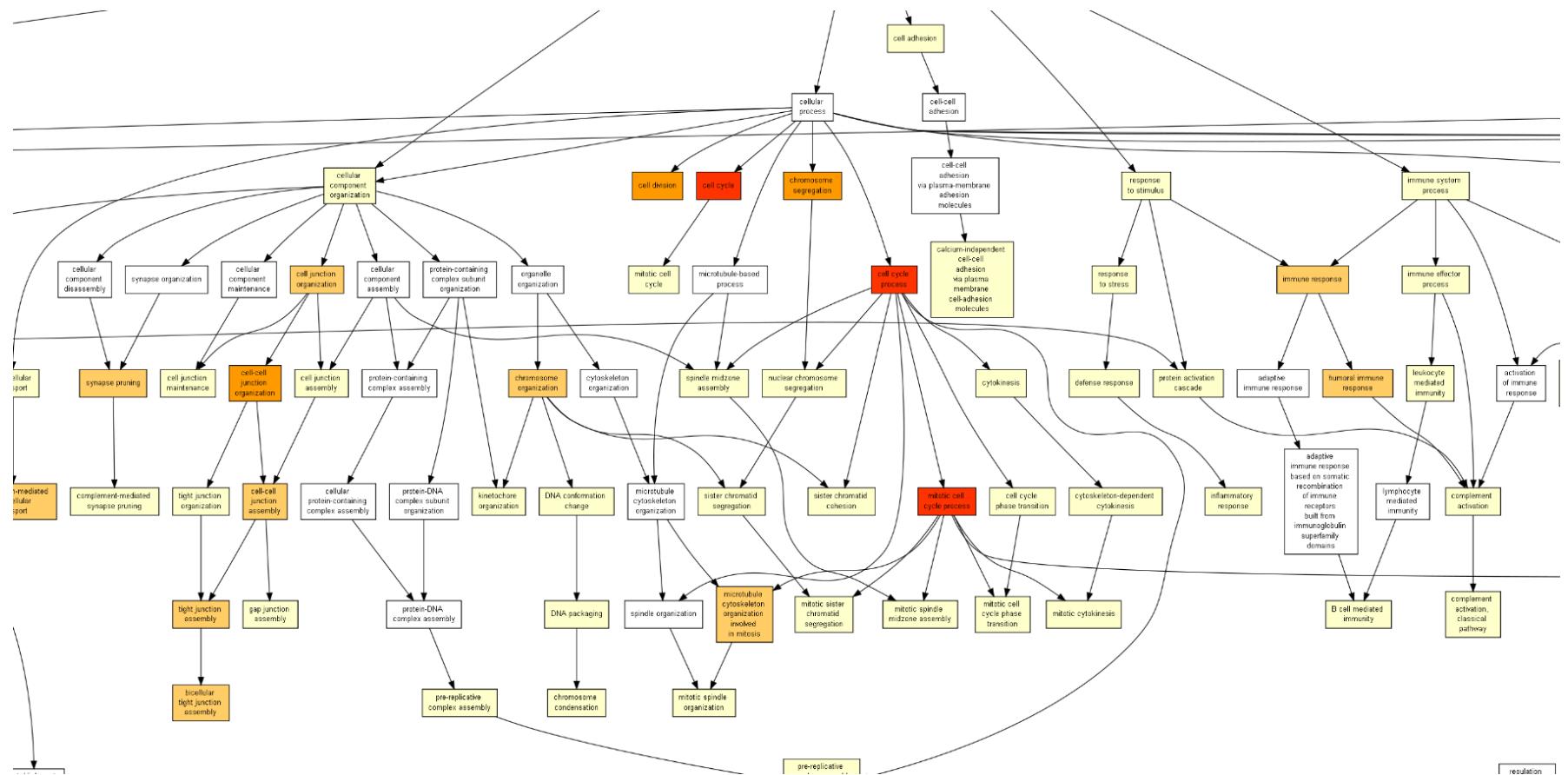
Or upload a file:  No se ha seleccionado ningún archivo.

### Step 4: Choose an ontology

- Process     Function     Component     All

# Overrepresentation analysis

## GORILLA



# Overrepresentation analysis

## Reactome Pathways analysis tool

### Analysis

The screenshot shows the Reactome homepage. At the top, there's a search bar with placeholder text "e.g. cdk1, mitogen, signaltransit, glucose". Below the search bar are four main icons: "Pathway Browser" (blue icon with a tree), "Analyze Data" (orange icon with a bar chart, circled in orange), "ReactomeFIViz" (blue icon with a network graph), and "Documentation" (blue icon with a document). The "Analyze Data" button is highlighted with a large orange circle. The background features a dark banner with the text "USE REACTOME GRAPH DATABASE IN YOUR PROJECT" and a "Why Reactome" section.

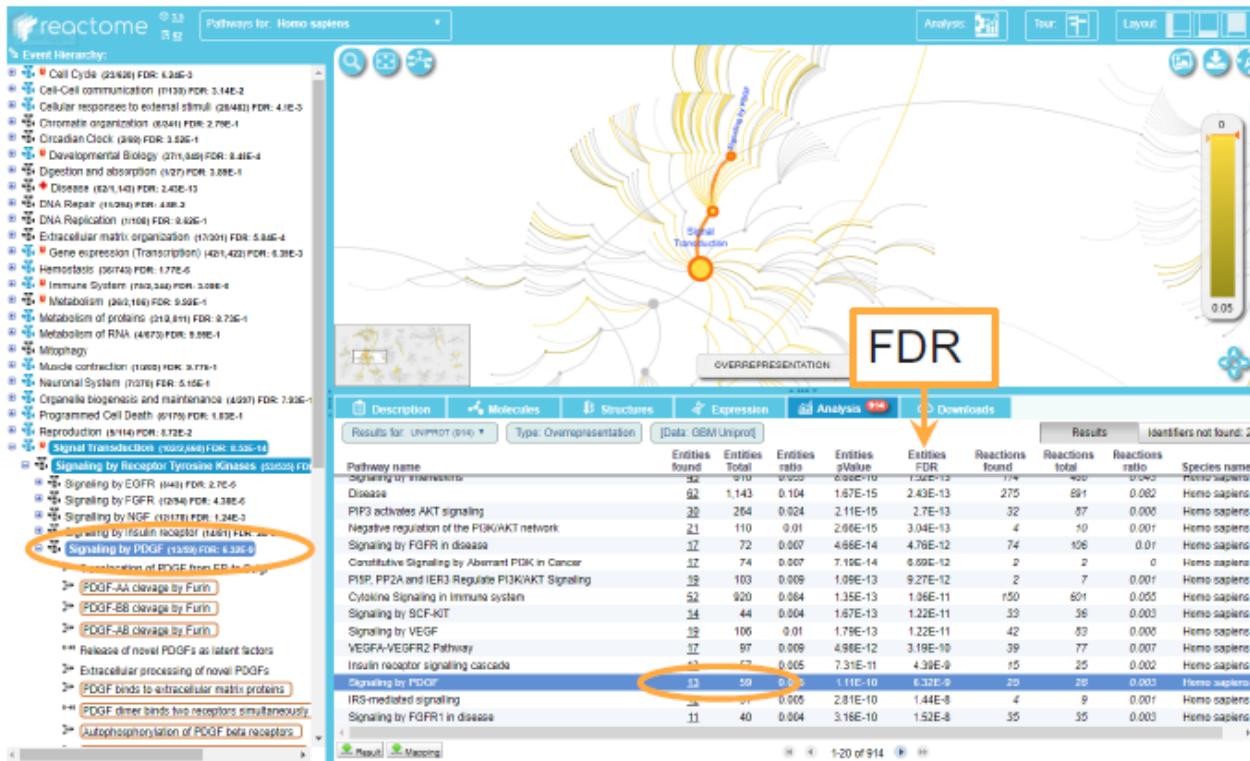
Familiarize with  
Reactome using example  
data

The screenshot shows the Reactome Pathway Browser interface. On the left, a sidebar lists biological pathways under categories like "Metabolic", "Signal Transduction", and "Cell Cycle". The main area displays a complex network of blue lines representing interactions between various cellular components. At the bottom, there are several tabs: "Description", "Molecules", "Reactions", "Enzymes", "Kinases", and "Downloads". The "Analyze" tab is highlighted with a large orange circle. A callout box at the bottom right provides a brief description of what the "Analyze" tab does: "Displays details when you select an item in the Pathway Browser. For example, when a reaction is selected, shows details including the input and output molecules, summary and references containing supporting evidence. When relevant, shows details of the catalyst, regulators, preceding and following events."

# Overrepresentation analysis

## Reactome Pathways analysis tool

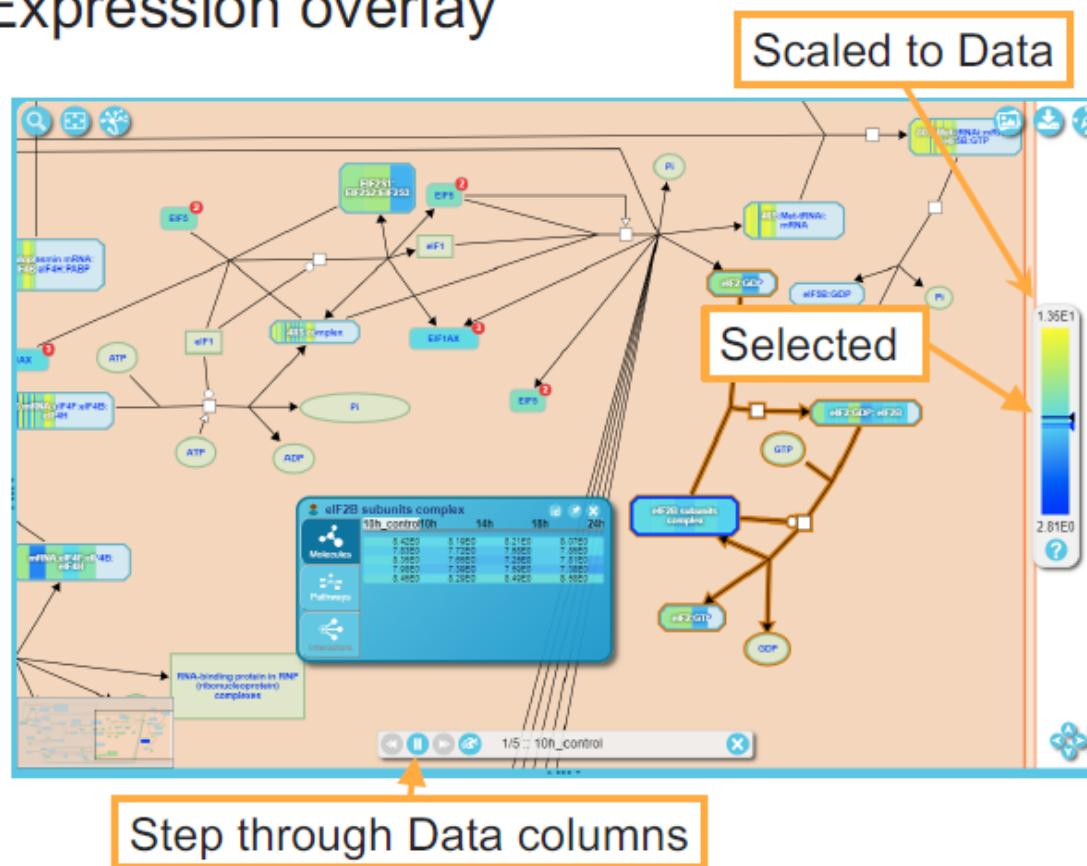
### Analysis Result – Over-representation



# Overrepresentation analysis

## Reactome Pathways analysis tool

### Expression overlay



# Overrepresentation analysis

## Reactome Pathways analysis tool

Try with your list of genes:

- Perform an overrepresentation analysis of the top up-regulated genes in B.PregVSLac comparison using Reactome Analysis web-based tool.
  - Examine the results and different outputs/tabs
  - What is the parent pathway of most of enriched terms?
  - Download a complete report of the results

# Overrepresentation analysis

## Reactome Pathways analysis tool

The Reactome Pathways analysis tool interface:

- Event Hierarchy:** A tree view showing biological pathways. Key nodes include Autophagy, Cell Cycle, Cell Cycle Checkpoints, Cell Cycle, Mitotic, Mitotic G1 phase and G1/S transition, S Phase, Mitotic G2-G2/M phases, G2 Phase, G2/M Transition, Cyclin A/B1/B2 associated events, Regulation of PLK1 Activity at G2/M, Polo-like kinase mediated events, Centrosome maturation, The role of GTSE1 in G2/M progression, AURKA Activation by TPX2, FBXL7 down-regulates AURKA, Interaction between PHLDA1 and AURKA, M Phase, Regulation of mitotic cell cycle, Chromosome Maintenance, Meiosis, Cell-Cell communication, Cellular responses to stimuli, and Chromatin organization.
- Search Bar:** "Search for a term, e.g. pten ..."
- Top Navigation:** Pathways for: Homo sapiens, Citation, Analysis, Tour, Layout.
- Pathway Network:** A complex network diagram showing interactions between various biological processes like DNA Repair, Cell Cycle, Mitotic Cycle, Programmed Cell Death, etc.
- Color Scale:** A vertical bar indicating the Overrepresentation score, ranging from 0 (yellow) to 0.05 (blue).
- Analysis Tab:** Shows OVERREPRESENTATION results for TOTAL Data submitted with no name. The table includes columns: Pathway name, Entities found, Entities Total, Entities ratio, Entities pValue, Entities FDR, Reactions found, Reactions total, Reactions ratio, and Species.
- Downloads:** Buttons for Results (110), Not found, and Downloads.

Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species
Cell Cycle, Mitotic	40	596	0.042	7.8E-9	5.51E-6	177	350	0.026	Homo sapiens
Cell Cycle	44	734	0.051	3.49E-8	1.1E-5	203	449	0.033	Homo sapiens
Resolution of Sister Chromatid Cohesion	17	134	0.009	4.7E-8	1.1E-5	7	8	0.001	Homo sapiens
Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal	14	94	0.007	1.09E-7	1.54E-5	4	4	0	Homo sapiens
Amplification of signal from the kinetochores	14	94	0.007	1.09E-7	1.54E-5	4	4	0	Homo sapiens
Mitotic Spindle Checkpoint	15	111	0.008	1.32E-7	1.54E-5	6	7	0.001	Homo sapiens
Cell Cycle Checkpoints	23	280	0.02	5.06E-7	5.06E-5	14	56	0.004	Homo sapiens
Interleukin-10 signaling	12	86	0.006	1.72E-6	1.51E-4	3	15	0.001	Homo sapiens
EML4 and NUDC in mitotic spindle formation	14	121	0.008	2.09E-6	1.63E-4	2	5	0	Homo sapiens
RHO GTPases Activate Formins	15	149	0.01	4.82E-6	3.37E-4	4	27	0.002	Homo sapiens

Page: 1-20 of 680

# Overrepresentation analysis

## Other tools

DAVID: <https://david.ncifcrf.gov/home.jsp>

The screenshot shows the DAVID Bioinformatics Database Analysis Wizard interface. At the top, there is a navigation bar with links for Home, Start Analysis (circled in red), Shortcut to DAVID Tools, Technical Center, Downloads & APIs, Term of Service, About DAVID, and About LHRI. Below the navigation bar, there are tabs for Upload, List (selected), and Background. On the left, the Gene List Manager section allows users to select species, with 'Homo sapiens(390)' currently selected. It also includes a 'Select Species' dropdown and a 'List Manager' section containing 'demolist2'. Below this are buttons for 'Select List to:' with options like 'Use', 'Rename', 'Remove', 'Combine', and 'Show Gene List'. The main right-hand area is titled 'Analysis Wizard' and displays the message 'Step 1. Successfully submitted gene list' with a checked checkbox. It shows 'Current Gene List: demolist2' and 'Current Background: Homo sapiens'. Below this, it says 'Step 2. Analyze above gene list with one of DAVID tools' with a downward arrow, followed by a list of tools: Functional Annotation Tool (with sub-options: Functional Annotation Clustering, Functional Annotation Chart, Functional Annotation Table), Gene Functional Classification Tool, Gene ID Conversion Tool, and Gene Name Batch Viewer. There are also links to 'Tell us how you like the tool' and 'Contact us for questions'.

# Overrepresentation analysis

## Other tools

**DAVID:** <https://david.ncifcrf.gov/home.jsp>


**DAVID Bioinformatics Resources**  
 Laboratory of Human Retrovirology and  
 Immunoinformatics (LHRI)



**Functional Annotation Clustering**

Current Gene List: demolist2  
 Current Background: Homo sapiens  
 372 DAVID IDs

Options    Classification Stringency: Medium

Kappa Similarity	Similarity Term Overlap: 3	Similarity Threshold: 0.50	
Classification	Initial Group Membership: 3	Final Group Membership: 3	Multiple Linkage Threshold: 0.50
Enrichment Thresholds	EASE: 1.0		
Display	<input type="checkbox"/> Fold Change <input type="checkbox"/> Bonferroni <input checked="" type="checkbox"/> Benjamini <input type="checkbox"/> FDR <input type="checkbox"/> LT,PH,PT		

**91 Cluster(s)**

Annotation Cluster 1	Enrichment Score: 3.22	G	Count	P_Value	Benjamini
<input type="checkbox"/> GOTERM_BP_DIRECT	<a href="#">negative regulation of transcription from RNA polymerase II promoter</a>	RT	42	3.9E-7	1.0E-3
<input type="checkbox"/> GOTERM_BP_DIRECT	<a href="#">positive regulation of transcription from RNA polymerase II promoter</a>	RT	46	3.0E-6	2.7E-3
<input type="checkbox"/> GOTERM_MF_DIRECT	<a href="#">transcriptional repressor activity, RNA polymerase II transcription regulatory region</a>	RT	20	2.0E-5	3.3E-3
<input type="checkbox"/> GOTERM_MF_DIRECT	<a href="#">sequence-specific binding</a>	RT	26	6.2E-5	8.2E-3
<input type="checkbox"/> GOTERM_CC_DIRECT	<a href="#">chromatin</a>	RT	37	1.1E-4	7.0E-3
<input type="checkbox"/> GOTERM_MF_DIRECT	<a href="#">sequence-specific double-stranded DNA binding</a>	RT	25	1.7E-4	1.9E-2
<input type="checkbox"/> GOTERM_MF_DIRECT	<a href="#">DNA binding</a>	RT	44	2.1E-4	1.9E-2
<input type="checkbox"/> GOTERM_MF_DIRECT	<a href="#">transcriptional activator activity, RNA polymerase II transcription regulatory region</a>	RT	22	3.0E-4	1.9E-2
<input type="checkbox"/> GOTERM_MF_DIRECT	<a href="#">sequence-specific binding</a>	RT	17	5.7E-4	3.1E-2
<input type="checkbox"/> GOTERM_MF_DIRECT	<a href="#">RNA polymerase II core promoter proximal region sequence-specific DNA binding</a>	RT	39	1.5E-3	5.2E-2
<input type="checkbox"/> GOTERM_BP_DIRECT	<a href="#">regulation of transcription from RNA polymerase II promoter</a>	RT	49	1.8E-3	1.1E-1

 [Download File](#)

# Overrepresentation analysis

## Other tools

PantherDB: <http://www.pantherdb.org/>



The mission of the PANTHER knowledgebase is to support biomedical and other research by providing comprehensive information about the evolution of protein-coding gene families, particularly protein phylogeny, function and genetic variation impacting that function. [Learn more](#)

PANTHER18.0 Released. [Click](#) for more details.

search keyword  All  Home About Data Version Tools API/Services Publications Workspace Downloads FAQ/Help/Tutorial Login Register Contact us

Current Release: PANTHER 18.0 | 15,693 family phylogenetic trees | 143 species | [News](#) [Whole genome function views](#)

Gene List Analysis      Browse      Sequence Search      Genetic Variant Impact      Keyword Search

Please refer to our article in [Nature Protocols](#) for detailed instructions on how to use this page.

**Help Tips**  
Steps:  
► 1. Select list and list type to analyze  
► 2. Select Organism  
► 3. Select operation  
[Using enhancer data](#)

1. Enter ids and or select file for batch upload. Else enter ids or select file or list from workspace for comparing to a reference list.  
Enter IDs:  Supported IDs  
Upload IDs:  Ninguno archivo selec.  
Please [login](#) to be able to select lists from your workspace.  
Select List Type:  ID List  
 Previously exported text search results  
 Workspace list  
 PANTHER Generic Mapping  
 ID's from Reference Proteome Genome  
Organism for id list    
 VCF File Flanking region   Search Enhancer Data

2. Select organism.

# Overrepresentation analysis

## Other tools



### WEB-based GEne SeT AnaLysis Toolkit

WebGestalt

Translating gene lists into biological insights...

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

[ORA Sample Run](#) | [GSEA Sample Run](#) | [NTA Sample Run](#) | [Phosphosite Sample Run \(New in 2019!\)](#) | [External Examples](#) | [Manual \(PDF, Web\)](#) | [Citation](#) |  
[GOView](#) | [WebGestaltR](#) | [WebGestalt 2017](#)

**Basic parameters**

**Organism of Interest** ⓘ Mus musculus

**Method of Interest** ⓘ Over-Representation Analysis (ORA)

**Functional Database** ⓘ pathway

+ Reactome

**Gene List**

**Select Gene ID Type** ⓘ NCBI Entrez gene

**Upload Gene List** ⓘ Click to upload Reset

OR

100038566  
74562  
69065  
74152

**Reference Gene List**

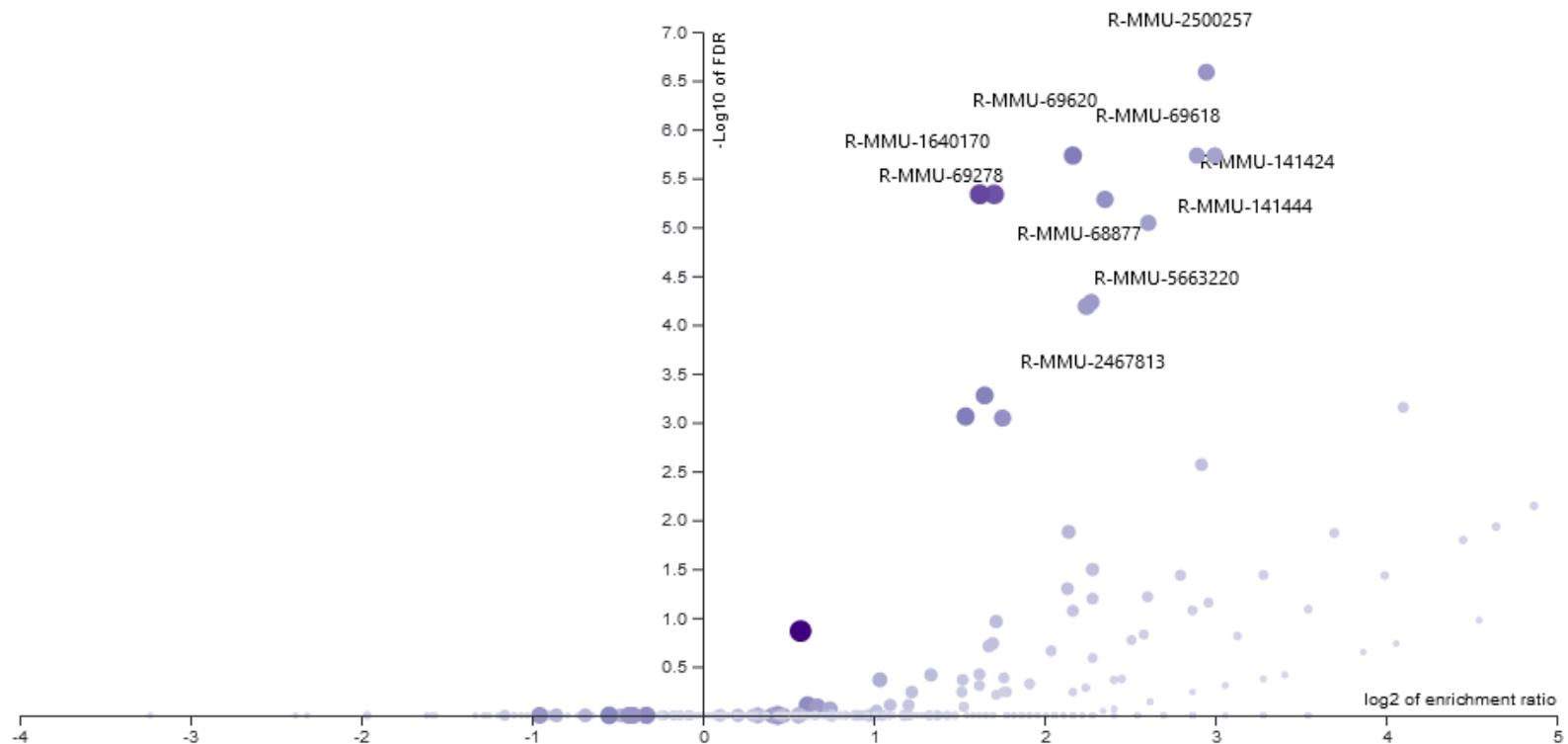
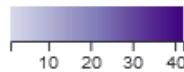
**Select Reference Set** ⓘ genome

**Upload User Reference Set** ⓘ Select the ID type of reference set

**File and Select ID type** ⓘ Click to upload Reset

# Overrepresentation analysis

## Other tools



Label: geneSet



PNG



Download

Draw Link

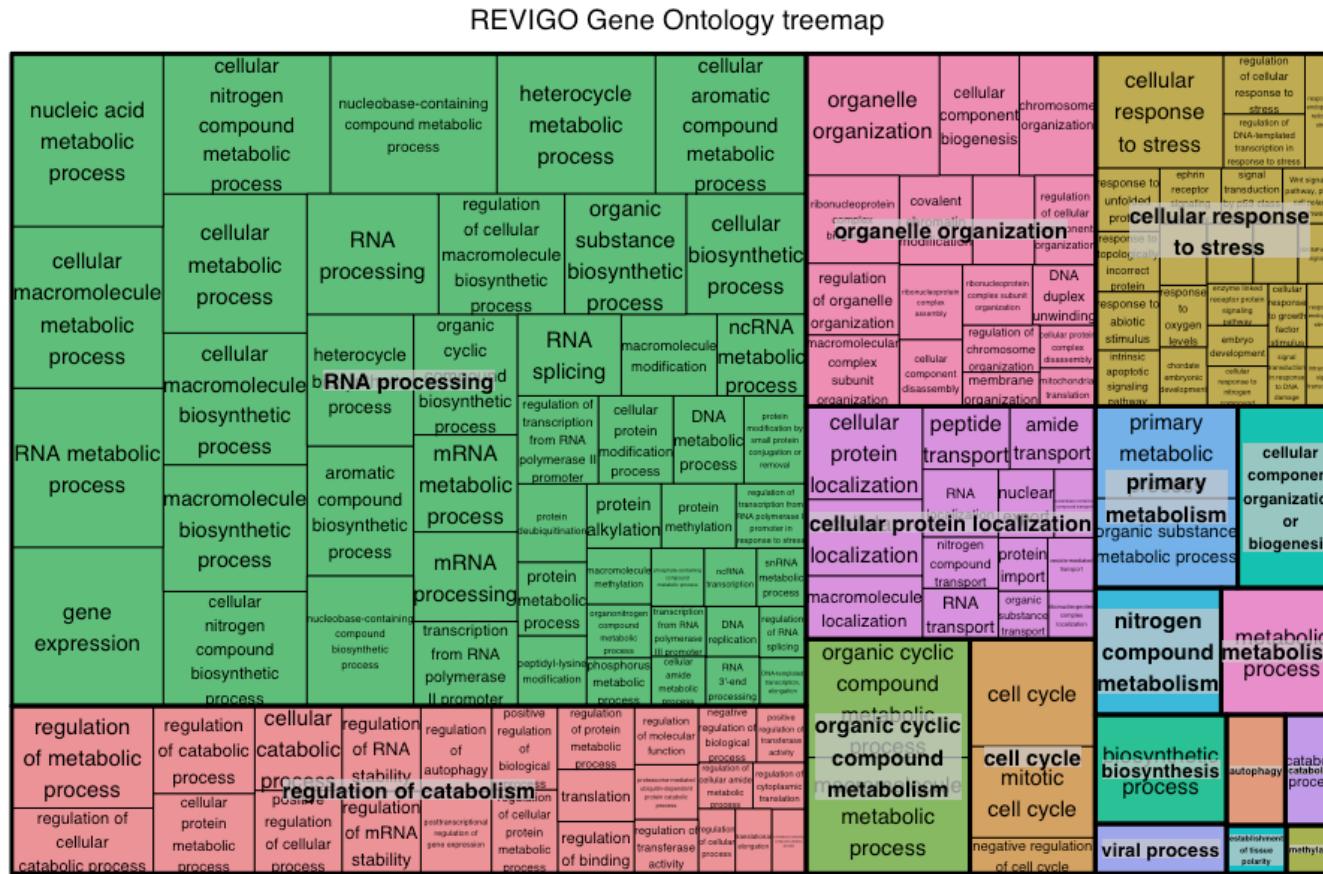


Pan & Zoom

# **Visualization and downstream analysis**

# Visualization and downstream analysis

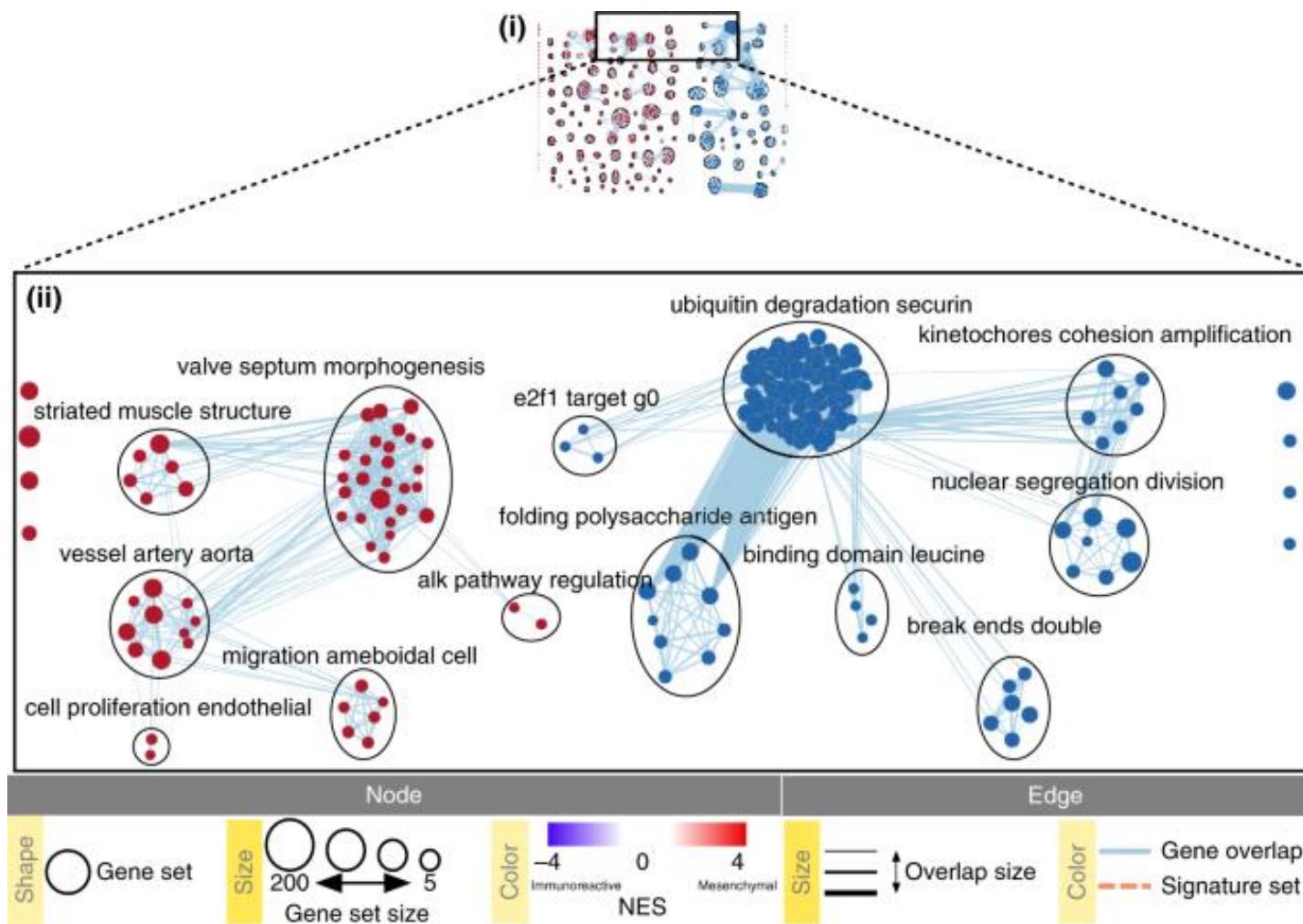
**REVIGO** is a web-based tool that can take our list of GO terms, collapse redundant terms by semantic similarity, and summarize them graphically.



# Visualization and downstream analysis

## EnrichmentMap Pipeline (Cytoscape)

- Nature Protocols: <https://www.nature.com/articles/s41596-018-0103-9>



# Hands on Pathway Analysis

## TABLE OF CONTENTS

1. Resources for functional analysis of gene lists
2. Navigating through different resources
  - a) Browsing the GO
  - b) Pathway visualization in Reactome
3. Case of study: RNA-seq experiment
  - a) Preparation of the gene list
  - b) Functional annotation of individual genes
  - c) Overrepresentation analysis of top DEG (g:Profiler / Reactome analysis tool)
  - d) Visualization and downstream analyses
4. A workflow in R (**clusterProfiler**)

# Hands On Pathway Analysis

## A workflow in R

Follow file: *Practicum\_EnrichmentAnalysis.Rmd / html*

# References

- [https://uebvhir.github.io/Pathway\\_Analysis-Guidelines/](https://uebvhir.github.io/Pathway_Analysis-Guidelines/)
  - Galaxy tutorial:  
<https://training.galaxyproject.org/training-material/topics/transcriptomics/tutorials/rna-seq-genes-to-pathways/tutorial.html>
  - Training materials:  
[https://hbctraining.github.io/DGE\\_workshop\\_salmon/lessons/functional\\_analysis\\_2019.html](https://hbctraining.github.io/DGE_workshop_salmon/lessons/functional_analysis_2019.html)
- [https://www.ebi.ac.uk/training/online/courses/reactome-tools-for-analysing-biological-pathways/#vf-tabs\\_section--overview](https://www.ebi.ac.uk/training/online/courses/reactome-tools-for-analysing-biological-pathways/#vf-tabs_section--overview)