

Hands on Pathway Analysis

Bioinformatics Course UEB-VHIR
November 2021

Mireia Ferrer¹, Álex Sánchez^{1,2}, Esther Camacho¹, Angel Blanco^{1,2}, Berta Miró¹

¹ Unitat d'Estadística i Bioinformàtica (UEB) VHIR

² Departament de Genètica Microbiologia i Estadística, UB

TABLE OF CONTENTS

1. Resources for functional analysis of gene lists
2. A workflow in R using clusterProfileR
 - a) Preparation of the gene list
 - b) Overrepresentation/GSEA analysis
 - c) Visualization of results
3. Navigating through Reactome Pathway database
4. Other tools

What's the biology behind a list of genes ?



Resources for functional analysis

Resources for functional analysis

Resources of functional annotations

- 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:**

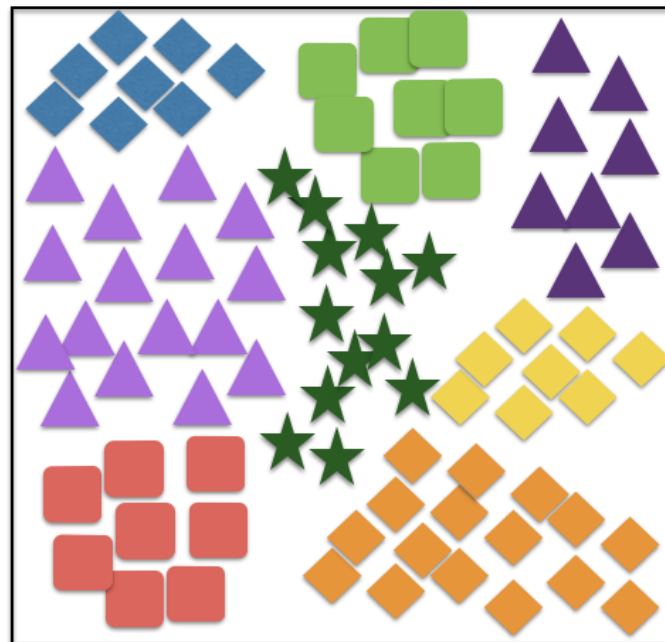
- [Gene Ontology](#) (GO)

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

- Pathways

- [KEGG](#)
 - [Reactome](#)
 - [WikiPathways](#)
 - [PantherDB](#)

All known genes in a species
(categorized into groups)



Resources of functional annotations

- 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](#)

- Disease-related

- DisGeNet
 - Disease Ontology

- Cell/Tissue markers

- Chromosomal location


- Regulatory motifs

- TRANSFAC
 - miRTarBase

Resources for functional analysis

Collections of gene sets

•MSigDB



MSigDB

Molecular Signatures Database

Molecular Signatures Database v7.2

Overview

The Molecular Signatures Database (MSigDB) is a collection of annotated gene sets for use with GSEA software. From this web site, you can

- ▶ **Search** for gene sets by keyword.
- ▶ **Browse** gene sets by name or collection.
- ▶ **Examine** a gene set and its annotations. See, for example, the [HALLMARK_APOPTOSIS](#) gene set page.
- ▶ **Download** gene sets.
- ▶ **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**

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

MSigDB database v7.2 updated September 2020. [Release notes](#).

Citing the MSigDB

To cite your use of the Molecular Signatures Database (MSigDB), a joint project of UC San Diego and Broad Institute, please reference Subramanian, Tamayo, et al. (2005) PNAS.

Collections

The MSigDB gene sets are divided into 9 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

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

C4

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

C5

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

C6

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

C7

immunologic signature gene sets defined directly from microarray gene expression data from immunologic studies.

C8

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


Web tools

- g:Profiler - <http://biit.cs.ut.ee/gprofiler/index.cgi>
- Reactome - <https://reactome.org/>
- DAVID - <http://david.abcc.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://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/>



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

[Home](#) » [BiocViews](#)

All Packages

http://bioconductor.org/packages/release/BiocViews.html#___Pathways

Bioconductor version 3.12 (Release)

Autocomplete [biocViews](#) search:

[ResearchField](#) (902)
[StatisticalMethod](#) (727)
[Technology](#) (1251)
[WorkflowStep](#) (1081)
 [Alignment](#) (80)
 [Annotation](#) (124)
 [BatchEffect](#) (52)
 [ExperimentalDesign](#) (22)
 [GenomeBrowsers](#) (2)
 [MultipleComparison](#) (163)
 [Normalization](#) (128)
 [Pathways](#) (179)
 [BioCarta](#) (7)
 [GO](#) (72)
 [KEGG](#) (51)
 [Reactome](#) (19)

Packages found under GO:

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

Show entries Search table:

Package	Maintainer	Title	Rank
annotate	Bioconductor Package Maintainer	Annotation for microarrays	17
GOSemSim	Guangchuang Yu	GO-terms Semantic Similarity Measures	36
clusterProfiler	Guangchuang Yu	statistical analysis and visualization of functional profiles for genes and gene clusters	38
enrichplot	Guangchuang Yu	Visualization of Functional Enrichment Result	43
GSEABase	Bioconductor Package Maintainer	Gene set enrichment data structures and methods	56
interactiveDisplayBase	Shawn Balcome	Base package for enabling powerful shiny web displays of Bioconductor objects	59

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

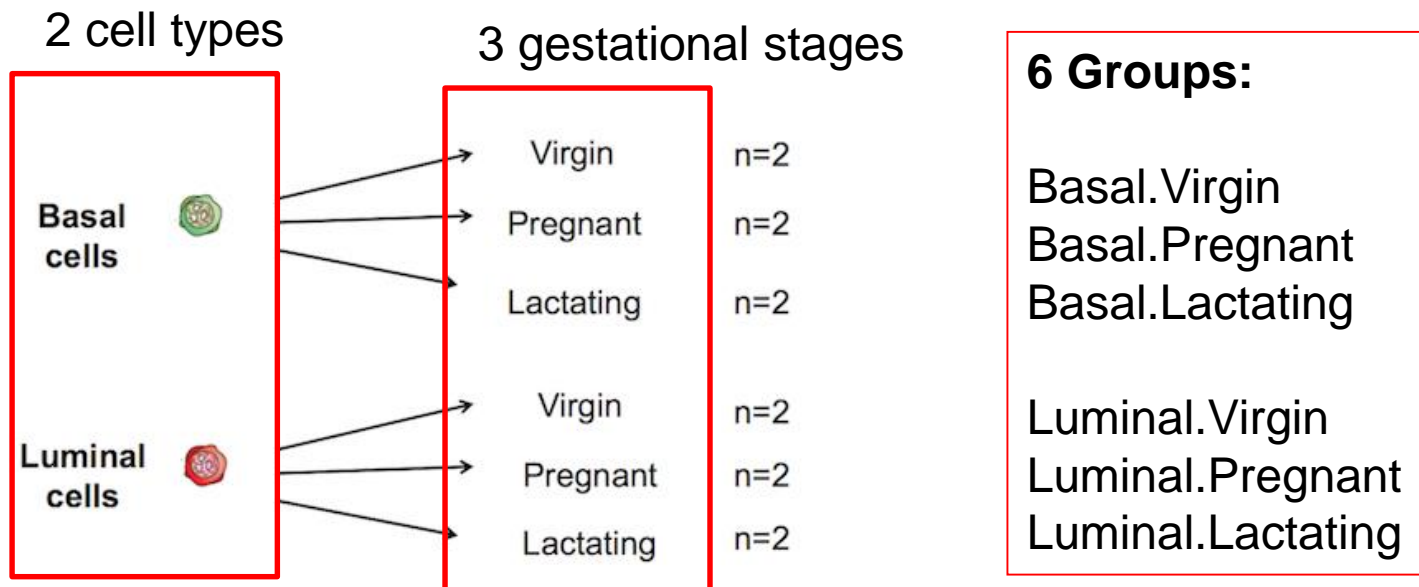
TABLE OF CONTENTS

1. Resources for functional analysis of gene lists
2. A workflow in R using clusterProfileR
 - a) Preparation of the gene list
 - b) Overrepresentation/GSEA analysis
 - c) Visualization of results
3. Navigating through Reactome Pathway database
4. Other tools

Hands On Pathway Analysis

Study overview

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



Fu *et al.* (2015) 'EGF-mediated induction of Mcl-1 at the switch to lactation is essential for alveolar cell survival' Nat Cell Biol

We analyzed the differential gene expression in the following comparisons:

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

Hands On Pathway Analysis

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?

A workflow in R

Follow file: *Practicum_EnrichmentAnalysis.html*

TABLE OF CONTENTS

1. Resources for functional analysis of gene lists
2. A workflow in R using clusterProfileR
 - a) Preparation of the gene list
 - b) Overrepresentation/GSEA analysis
 - c) Visualization of results
3. Navigating through Reactome Pathway database
4. Other tools

Hands On Pathway Analysis

Navigating through Reactome Pathway database

<https://reactome.org/>

The screenshot shows the Reactome website homepage. At the top, there is a navigation bar with links: About, Content, Docs, Tools, Community, and Download. Below this is a search bar with the text "Find Reactions, Proteins and Pathways" and a search button labeled "Go!". The search bar contains the example text "e.g. O95631, NTN1, signaling by EGFR, glucose". Below the search bar, there are four main sections, each with an icon and a title:

- Pathway Browser**: Visualize and interact with Reactome biological pathways. (Icon: A tree diagram)
- Analyze Data**: Merges pathway identifier mapping, over-representation, and expression analysis. (Icon: A bar chart)
- ReactomeFIViz**: Designed to find pathways and network patterns related to cancer and other types of diseases. (Icon: A network diagram)
- Documentation**: Information to browse the database and use its principal tools for data analysis. (Icon: A document)

Below these sections is a dark banner with the text "USE REACTOME GRAPH DATABASE IN YOUR PROJECT" and a "LEARN MORE" button. At the bottom, there are two sections: "Why Reactome" and "Tweets". The "Why Reactome" section is highlighted with an orange border and contains the following text:

Why Reactome

Reactome is a free, open-source, curated and peer-reviewed pathway database. Our goal is to provide intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge to support basic research, genome analysis, modeling, systems biology and education.

If you use Reactome in Asia, we suggest using our Chinese mirror site at reactome.ncpsb.org.

The "Tweets" section shows a tweet from @ChEMBL: "#ChEMBL is 10 years old this year :) For more about the celebrations see: chembl.blogspot.com/2019/02/chembl..."

Tutorials: <http://www.reactome.org/userguide/Usersguide.html>
<https://www.ebi.ac.uk/training/online/>

Hands On Pathway Analysis

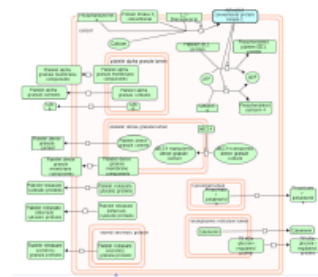
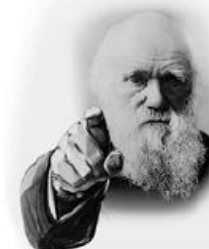
Navigating through Reactome Pathway database

Reactome is...

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

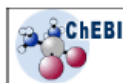
Authored by expert biologists,
maintained by Reactome editorial
staff (curators)

Mapped to cellular compartment



Extensively cross-referenced

Entrez Gene

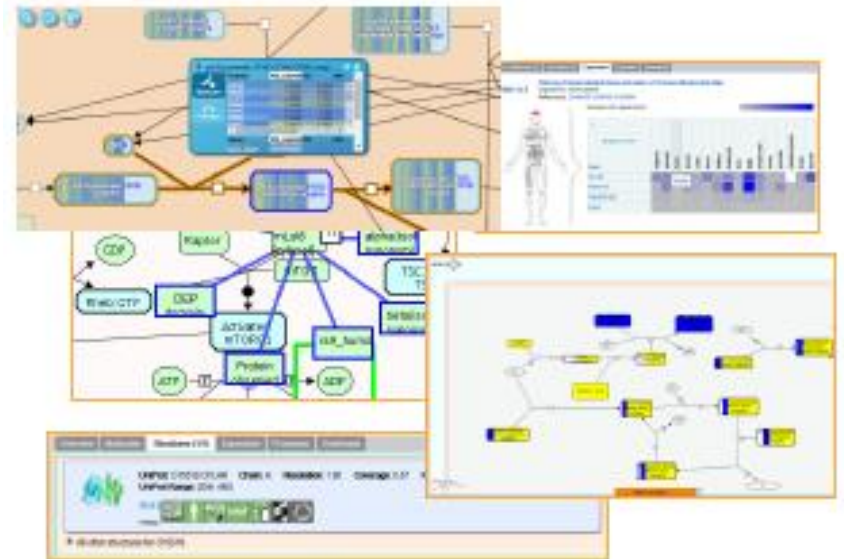


Hands On Pathway Analysis

Navigating through Reactome Pathway database

Reactome Tools

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



Hands On Pathway Analysis

Navigating through Reactome Pathway database

The Pathway Browser

The screenshot shows the Reactome Pathway Browser interface. The top navigation bar includes links for Home, Species, Analyse Data, Video Tour, and Layout. The main content area displays a complex pathway diagram with various biological processes represented by colored nodes and connecting lines. On the left, a 'Hierarchy Panel' lists various biological categories. At the bottom, a 'Detail Panel' provides information about the selected reaction. On the right, a 'Settings Sidebar' allows for customization. Various interactive tools are labeled with orange boxes and arrows, including 'Search Diagram', 'Open Diagram', 'Fit to Page', 'Thumbnail', 'Zoom/Move', 'Key', 'Export', and 'Illustrations'.

Home Species Analyse Data Video Tour Layout

reactome Pathways for: Homo sapiens

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
- Hemostasis
- Immune
- Metabolism
- Metabolism
- Mitochondria
- Mitophagy
- Muscle contraction
- Nervous System
- Organelle biogenesis and maintenance
- Programmed Cell Death
- Reproduction
- Signal Transduction
- Transport of small molecules
- Vesicle-mediated transport

Open Diagram

Search Diagram

Fit to Page

Pathway Panel

Illustrations

Key

Export

Settings Sidebar

Thumbnail

Zoom/Move

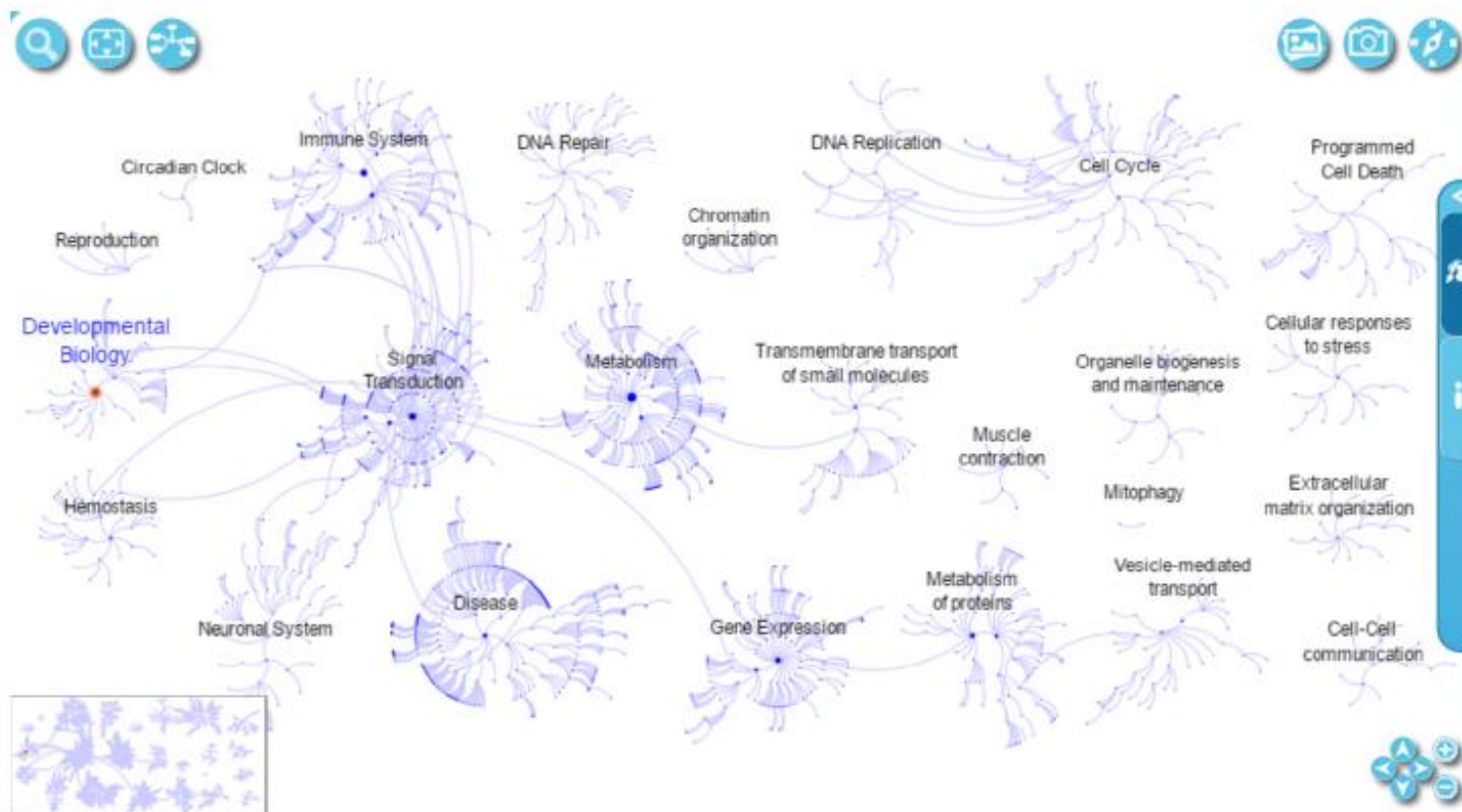
Detail Panel

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.

Hands On Pathway Analysis

Navigating through Reactome Pathway database

Pathway Overview



Hands On Pathway Analysis

Navigating through Reactome Pathway database

Navigating in the Pathway Browser

Home button

The screenshot displays the Reactome Pathway Browser interface. On the left, a sidebar lists various biological processes, with 'Signal Transduction' selected. Below it, a list of pathways is shown, including 'Signaling by Receptor Tyrosine Kinases', 'Signaling by TGF-beta family members', 'Signaling by GPCR', 'Signaling by NOTCH1', and 'Activated NOTCH1 Transmits Signal to the Nucleus'. The main area shows a network diagram of the 'Signal Transduction' pathway, with nodes representing molecules and edges representing reactions. A callout box points to a specific reaction in the diagram, stating 'Click reaction to open'. Another callout box points to the 'Details here' link in the reaction description, stating 'Details here'. A third callout box points to the 'Home button' in the top right corner, stating 'Click pathway to select, double-click to open diagram'.

reactome

Pathways for Mouse explore

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)
- Homeostasis
- Immune System
- Metabolism
- Metabolism of proteins
- Metabolism of RNA
- Mitophagy
- Muscle contraction
- Neuronal System
- Organelle 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 NOTCH1
- Pre-NOTCH1 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 DLL/AG ligands upon bind
 - Complex of NOTCH1 with its ligand is cleaved
 - NEXT1 is cleaved to produce NICD1
 - NICD1 translocates to nucleus
 - DTX binds NOTCH1
 - Conductin-1 (CNTN1) binds NOTCH1
 - Cleavage of CNTN1:NOTCH1 complex releases

Click pathway to select, double-click to open diagram

Click reaction to open

Details here

Description

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 by 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 pulses such as that of the vertebrate somitogenesis. This process is known as lateral inhibition: a molecular mechanism whereby individual cells direct their neighbors from doing the same. The NSP has been adopted by several other biological systems to divide the growing embryo into regular clocks called somitogenesis which are generated from a molecular oscillator in the presomitic mesoderm. The Notch receptor is synthesized in the rough endoplasmic reticulum as a single polypeptide chain, leading to a heterodimeric mature receptor consisting of non-covalently associated extracellular, transmembrane, and intracellular domains. Following ligand activation and further proteolytic cleavage, the intracellular domain of the receptor is released into the cytoplasm where it regulates gene expression.

Background literature references...

Hands On Pathway Analysis

Navigating through Reactome Pathway database

Navigating in the Pathway Browser

The screenshot shows the Reactome Pathway Browser interface. On the left is a hierarchical tree of biological processes. The main area displays a complex pathway diagram with nodes and connecting arrows. At the bottom, there is a detailed description of the selected pathway.

Selected: An orange box highlights the 'Signal Transduction' category in the left-hand navigation tree.

Highlights: An orange box highlights a specific node in the pathway diagram, with orange arrows pointing to it from the 'Selected' box.

Details don't update until selection: An orange box highlights the detailed description at the bottom of the page, which remains static until a specific pathway is selected.

Hover mouse: An orange box highlights the 'Signal Transduction' category in the left-hand navigation tree, indicating the action of hovering over it.

Pathway Diagram: The central area shows a complex network of biological processes, including 'Signal Transduction', 'Cell Cycle', and 'Metabolism'. The diagram uses various symbols to represent different types of molecules and interactions.

Navigation Tree: The left-hand side features a hierarchical tree of biological processes, including 'Signal Transduction', 'Cell Cycle', 'Metabolism', and 'Developmental Biology'. The 'Signal Transduction' category is currently selected.

Description Panel: The bottom panel provides a detailed description of the selected pathway, including a title, a brief summary, and a list of associated genes and proteins. The title is 'Activated NOTCH1 Transmits Signal to the Nucleus'.

Hands On Pathway Analysis

Navigating through Reactome Pathway database

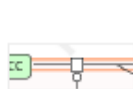
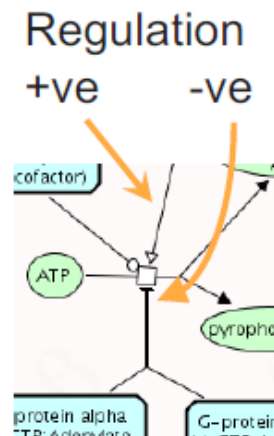
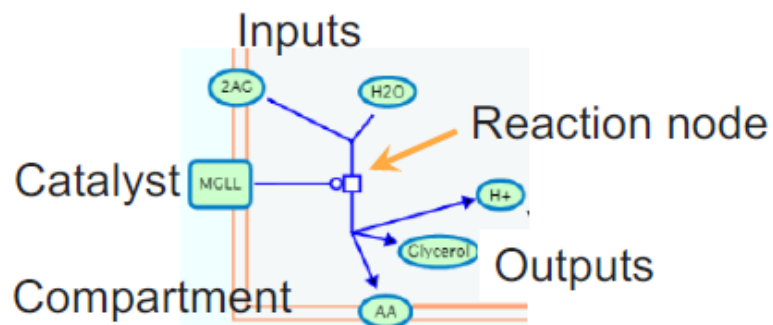
The Pathway Browser - Pathway Diagrams

Ovals are small molecules (or sets of)

Green boxes are proteins,

Blue are complexes,

Blue with double-boundary are sets



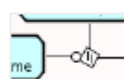
Transition



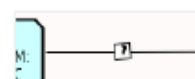
Binding



Dissociation



Omitted

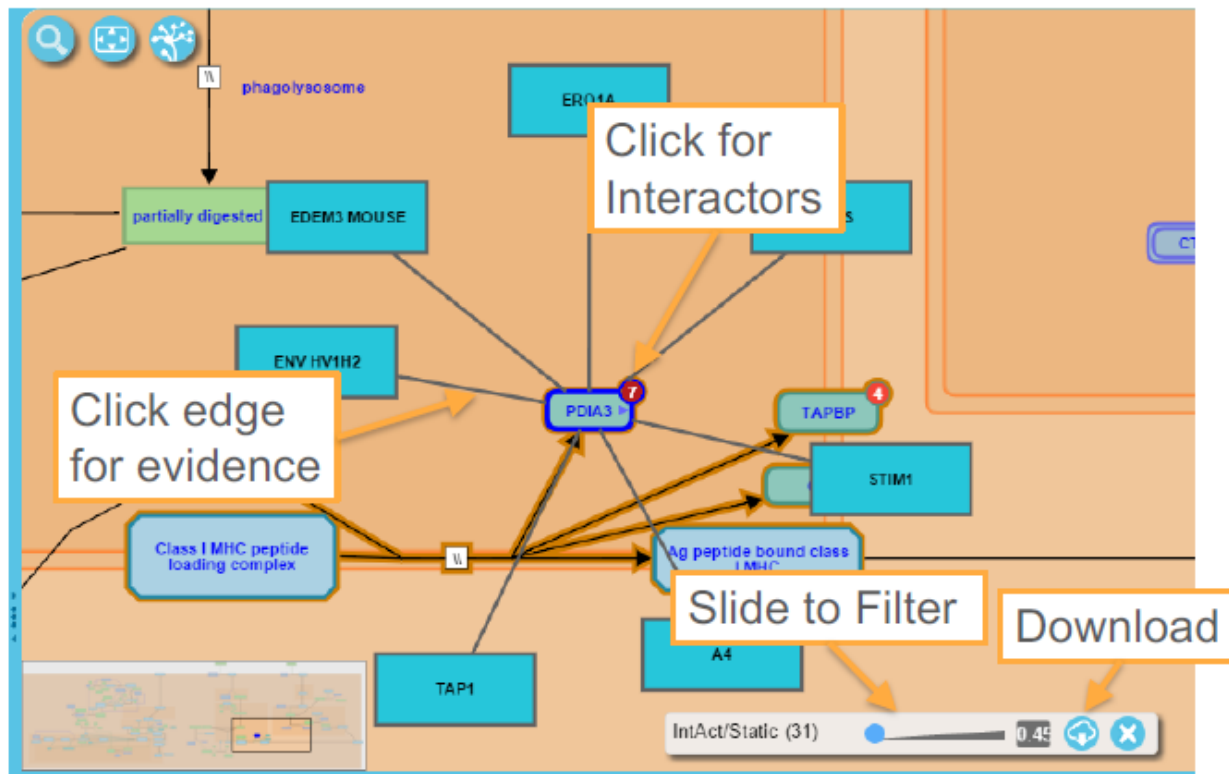


Uncertain

Hands On Pathway Analysis

Navigating through Reactome Pathway database

Interactors



Hands On Pathway Analysis

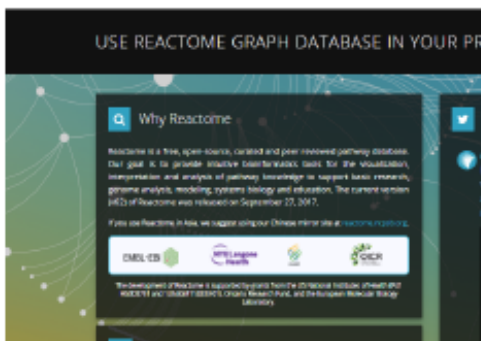
Navigating through Reactome Pathway database

Show Illustration

The screenshot displays the Reactome Pathway database interface. On the left, a sidebar shows a hierarchical list of biological processes, with 'Assembly of the primary cilium' selected. The main area features a detailed illustration of the primary cilium assembly process, including the axoneme, transition zone, and basal body. The illustration is divided into two panels: 'CILIogenesis' and 'ASSEMBLY OF THE PRIMARY CILIUM INTRAFAGELLAR TRANSPORT AND CARGO TRAFFICKING'. The 'CILIogenesis' panel shows the formation of the cilium from the mother centriole. The 'ASSEMBLY OF THE PRIMARY CILIUM' panel shows the transport of IFT particles and cargo into the cilium. The illustration includes labels for various components such as the cilium membrane, periciliary membrane, plasma membrane, cilium vesicle, distal appendages, cilium sheath, mother centriole, transition zone, cilium pocket, transition fibres, basal body, and IFT particles (IFTB, IFTA, IFTC, IFTD, IFTF, IFTG, IFTH). It also shows the transport of cargo (e.g., RPL2, ARL13, UNC119, ARMC9, RFX2, RFX3, RFX4, RFX5, RFX6, RFX7, RFX8, RFX9, RFX10, RFX11, RFX12, RFX13, RFX14, RFX15, RFX16, RFX17, RFX18, RFX19, RFX20, RFX21, RFX22, RFX23, RFX24, RFX25, RFX26, RFX27, RFX28, RFX29, RFX30, RFX31, RFX32, RFX33, RFX34, RFX35, RFX36, RFX37, RFX38, RFX39, RFX40, RFX41, RFX42, RFX43, RFX44, RFX45, RFX46, RFX47, RFX48, RFX49, RFX50, RFX51, RFX52, RFX53, RFX54, RFX55, RFX56, RFX57, RFX58, RFX59, RFX60, RFX61, RFX62, RFX63, RFX64, RFX65, RFX66, RFX67, RFX68, RFX69, RFX70, RFX71, RFX72, RFX73, RFX74, RFX75, RFX76, RFX77, RFX78, RFX79, RFX80, RFX81, RFX82, RFX83, RFX84, RFX85, RFX86, RFX87, RFX88, RFX89, RFX90, RFX91, RFX92, RFX93, RFX94, RFX95, RFX96, RFX97, RFX98, RFX99, RFX100) and the role of various proteins (e.g., IFTB, IFTA, IFTC, IFTD, IFTF, IFTG, IFTH, RPL2, ARL13, UNC119, ARMC9, RFX2, RFX3, RFX4, RFX5, RFX6, RFX7, RFX8, RFX9, RFX10, RFX11, RFX12, RFX13, RFX14, RFX15, RFX16, RFX17, RFX18, RFX19, RFX20, RFX21, RFX22, RFX23, RFX24, RFX25, RFX26, RFX27, RFX28, RFX29, RFX30, RFX31, RFX32, RFX33, RFX34, RFX35, RFX36, RFX37, RFX38, RFX39, RFX40, RFX41, RFX42, RFX43, RFX44, RFX45, RFX46, RFX47, RFX48, RFX49, RFX50, RFX51, RFX52, RFX53, RFX54, RFX55, RFX56, RFX57, RFX58, RFX59, RFX60, RFX61, RFX62, RFX63, RFX64, RFX65, RFX66, RFX67, RFX68, RFX69, RFX70, RFX71, RFX72, RFX73, RFX74, RFX75, RFX76, RFX77, RFX78, RFX79, RFX80, RFX81, RFX82, RFX83, RFX84, RFX85, RFX86, RFX87, RFX88, RFX89, RFX90, RFX91, RFX92, RFX93, RFX94, RFX95, RFX96, RFX97, RFX98, RFX99, RFX100). The bottom of the interface shows a table with columns for Description, Molecules, Structures, Expression, Analysis, and Downloads. A note indicates that analysis results are shown when an analysis has been run.

Hands On Pathway Analysis

Navigating through Reactome Pathway database

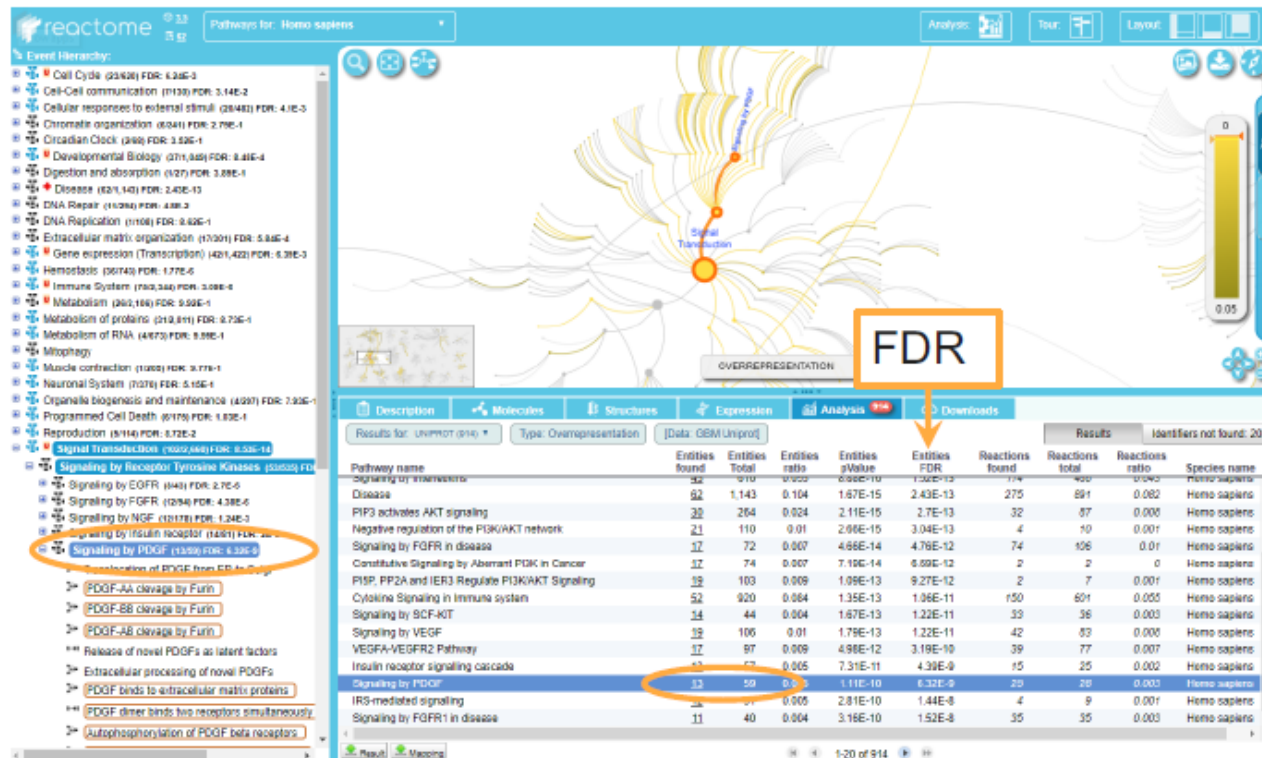


Familiarize with
Reactome using example
data

Hands On Pathway Analysis

Navigating through Reactome Pathway database

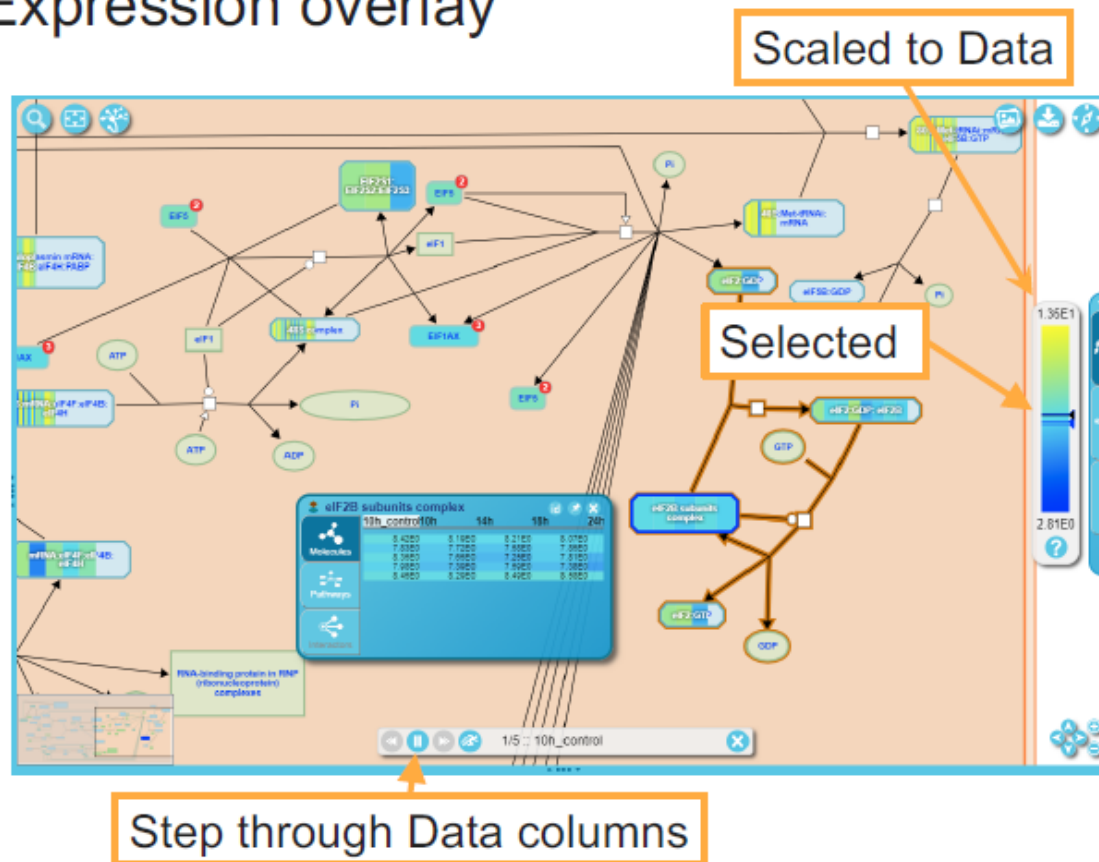
Analysis Result – Over-representation



Hands On Pathway Analysis

Navigating through Reactome Pathway database

Expression overlay



Hands On Pathway Analysis

Navigating through Reactome Pathway database

Exercise

- 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

Hands On Pathway Analysis

Navigating through Reactome Pathway database

reactome 3.7 78

Pathways for: **Homo sapiens**

Citation: Analysis: Tour: Layout:

Event Hierarchy:

- Autophagy (1/166) FDR: 9.82E-1
- Cell Cycle (44/734) FDR: 1.1E-5
 - Cell Cycle Checkpoints (23/280) FDR: 5.0E-6
 - Cell Cycle, Mitotic (40/596) FDR: 5.51E-6
 - Mitotic G1 phase and G1/S transition
 - S Phase (10/180) FDR: 1.69E-1
 - Mitotic G2-G2/M phases (13/214) FDR: 4.9E-6
 - G2 Phase (2/7) FDR: 1.65E-1
 - G2/M Transition (13/212) FDR: 4.9E-6
 - Cyclin A/B1/B2 associated even
 - Regulation of PLK1 Activity at C
 - Polo-like kinase mediated even
 - Centrosome maturation (2/83) F
 - The role of GTSE1 in G2/M pro
 - AURKA Activation by TPX2 (5/
 - FBXL7 down-regulates AURKA
 - Interaction between PHILDA1 a
 - M Phase (21/416) FDR: 3.64E-2
 - Regulation of mitotic cell cycle (9/92)
 - Chromosome Maintenance (8/138) FDR: 1.1E-5
 - Meiosis (2/92) FDR: 6.51E-1
 - Cell-Cell communication (6/133) FDR: 4.44E-6
 - Cellular responses to stimuli (13/970) FDR: 1.1E-5
 - Chromatin organization

Search for a term, e.g. pten ...

OVERREPRESENTATION Showing pValue

OVERREPRESENTATION analysis results for **TOTAL** Data submitted with no name

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	Hom
Cell Cycle	44	734	0.051	3.49E-8	1.1E-5	203	449	0.033	Hom
Resolution of Sister Chromatid Cohesion	17	134	0.009	4.7E-8	1.1E-5	7	8	0.001	Hom
Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal	14	94	0.007	1.09E-7	1.54E-5	4	4	0	Hom
Amplification of signal from the kinetochores	14	94	0.007	1.09E-7	1.54E-5	4	4	0	Hom
Mitotic Spindle Checkpoint	15	111	0.008	1.32E-7	1.54E-5	6	7	0.001	Hom
Cell Cycle Checkpoints	23	280	0.02	5.06E-7	5.06E-5	14	56	0.004	Hom
Interleukin-10 signaling	12	86	0.006	1.72E-6	1.51E-4	3	15	0.001	Hom
EML4 and NUDC in mitotic spindle formation	14	121	0.008	2.09E-6	1.63E-4	2	5	0	Hom
RHO GTPases Activate Formins	15	149	0.01	4.82E-6	3.37E-4	4	27	0.002	Hom

1-20 of 680

Hands On Pathway Analysis

TABLE OF CONTENTS

1. Resources for functional analysis of gene lists
2. A workflow in R using clusterProfileR
 - a) Preparation of the gene list
 - b) Overrepresentation/GSEA analysis
 - c) Visualization of results
3. Navigating through Reactome Pathway database
4. Other tools

Hands On Pathway Analysis

g:Profiler

- ✓ 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, ..)

Data sources ^

[Show data versions](#)

Gene Ontology

☒ GO molecular function

☒ GO cellular component

☒ GO biological process

☐ No electronic GO annotations ?

biological pathways

☒ KEGG

☒ Reactome

☒ WikiPathways

regulatory motifs in DNA

☒ TRANSFAC

☒ miRTarBase

protein databases

☒ Human Protein Atlas

☒ CORUM

Human phenotype ontology

☒ HP

Hands On Pathway Analysis

g:Profiler

g:Profiler graphical output

q: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

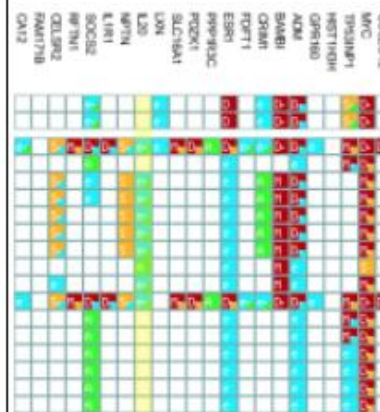
source	term name	term ID
BP	negative regulation of biological process	G0:0048819
BP	negative regulation of cellular process	G0:0048823
BP	single-organism process	G0:0044699
BP	death	G0:0016205
BP	developmental process	G0:0032502
BP	single-organism developmental process	G0:0044767
BP	multicellular organismal development	G0:0007275
BP	anatomical structure development	G0:0048896
BP	system development	G0:0048731
BP	tissue development	G0:0009888
BP	organ development	G0:0048813
BP	single-organism cellular process	G0:0044763
BP	cell death	G0:0008219
BP	programmed cell death	G0:0012501
BP	apoptotic process	G0:0006915
BP	regulation of cell death	G0:0010941
BP	regulation of programmed cell death	G0:0043067
BP	regulation of apoptotic process	G0:0042981

size of the
overlap

signi-
fican-
ce

> Genes in input list

n. of terra genes	n. of quarry genes	n. of common genes	p-value
4340	391	127	2.29e-03
3985	391	119	2.17e-03
13478	391	305	1.49e-02
19495	391	69	3.41e-03
9560	391	150	1.82e-02
5476	391	148	2.00e-02
4696	391	132	1.16e-02
4905	391	138	7.09e-03
4114	391	122	2.11e-03
1763	391	63	8.57e-03
2992	391	95	3.79e-03
12236	391	262	2.32e-02
1941	391	69	3.17e-03
1870	391	68	1.16e-03
1830	391	68	7.76e-04
1467	391	61	7.24e-05
1395	391	60	2.83e-05
1383	391	60	2.96e-05



Hands On Pathway Analysis

Functional profiling

Gene ID conversion

Orthology search

SNP id to gene name

Query

Upload query

Upload bed file

Input is whitespace-separated list of genes ?

16970
269959
666918
240168
16185
12837
224344
18383
329731
245020
116939
214425
320312
666348
333605

Gene list

ENTREZGENE_ACC

Run query

random

Data sources ^

select all

clear all

Show data versions

Gene Ontology

- ☒ GO molecular function
- ☒ GO cellular component
- ☒ GO biological process
- ☐ No electronic GO annotations ?

biological pathways

- ☐ KEGG
- ☒ Reactome
- ☐ WikiPathways

Options

Organism: ?

Organism?

☐ Ordered query ?

☐ Run as multiquery ?

Advanced options ^

☐ All results ?

☐ Measure underrepresentation ?

Statistical domain scope ?

Only annotated genes

Universe?

Significance threshold ?

Benjamini-Hochberg FDR

User threshold ?

0.05

Numeric IDs treated as ?

ENTREZGENE_ACC

Hands On Pathway Analysis

170770
14115
14585
12740
14299
12266
14621

Run query

random example

Data sources ▼

Bring your data (Custom GMT) ▼

Results

Detailed Results

Query info 9

Export to PNG



Show query URL



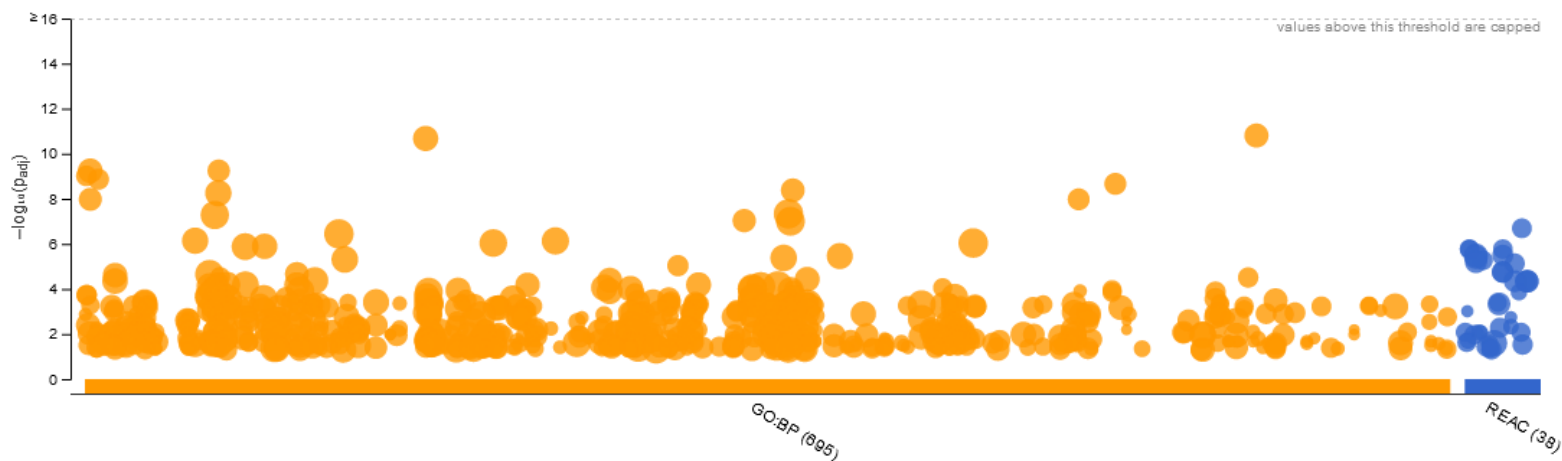
Show short link



☐ For print ?

☒ Capped ?

☐ Make unselected terms transparent



Hands On Pathway 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
- ✗ low flexibility in gene ids

Hands On Pathway 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:

Myxk
F630040K05Rik
Cers3
Gm10658
Fer114
Chac1
Stra61

← Gene list

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

Background set:

Gdf11
Rcor1
Smarchb1
Zfp551
Kbtbd6
Psen2

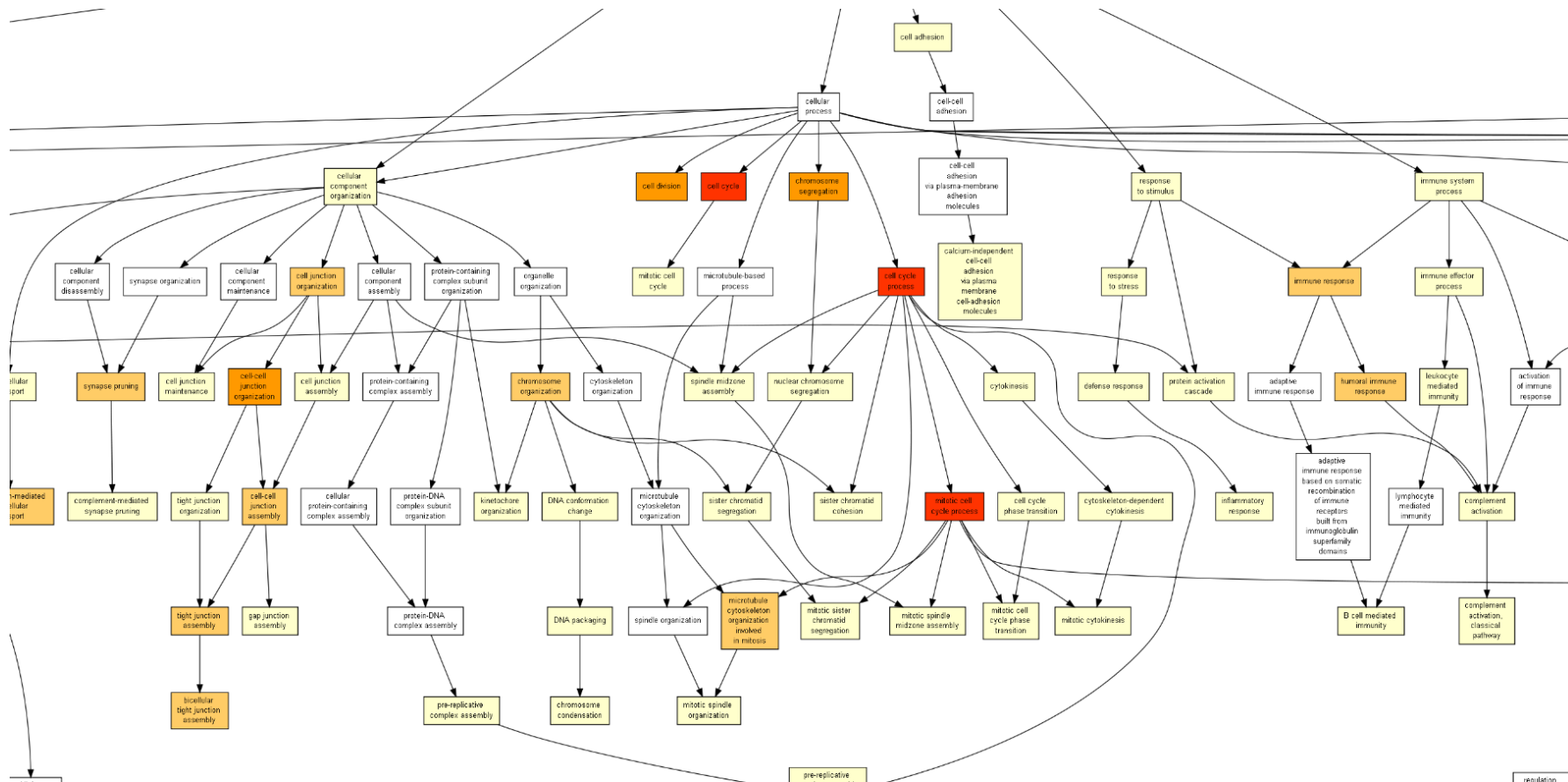
← Universe (take it from file
topTab.csv)

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

Step 4: Choose an ontology

- ☒ Process ☐ Function ☐ Component ☐ All

Hands On Pathway Analysis



Hands On Pathway Analysis

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



WebGestalt

WEB-based GENE SeT AnaLysis Toolkit

Translating gene lists into biological insights...

[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 ⓘ

OR

100038566
74562
69065
74152

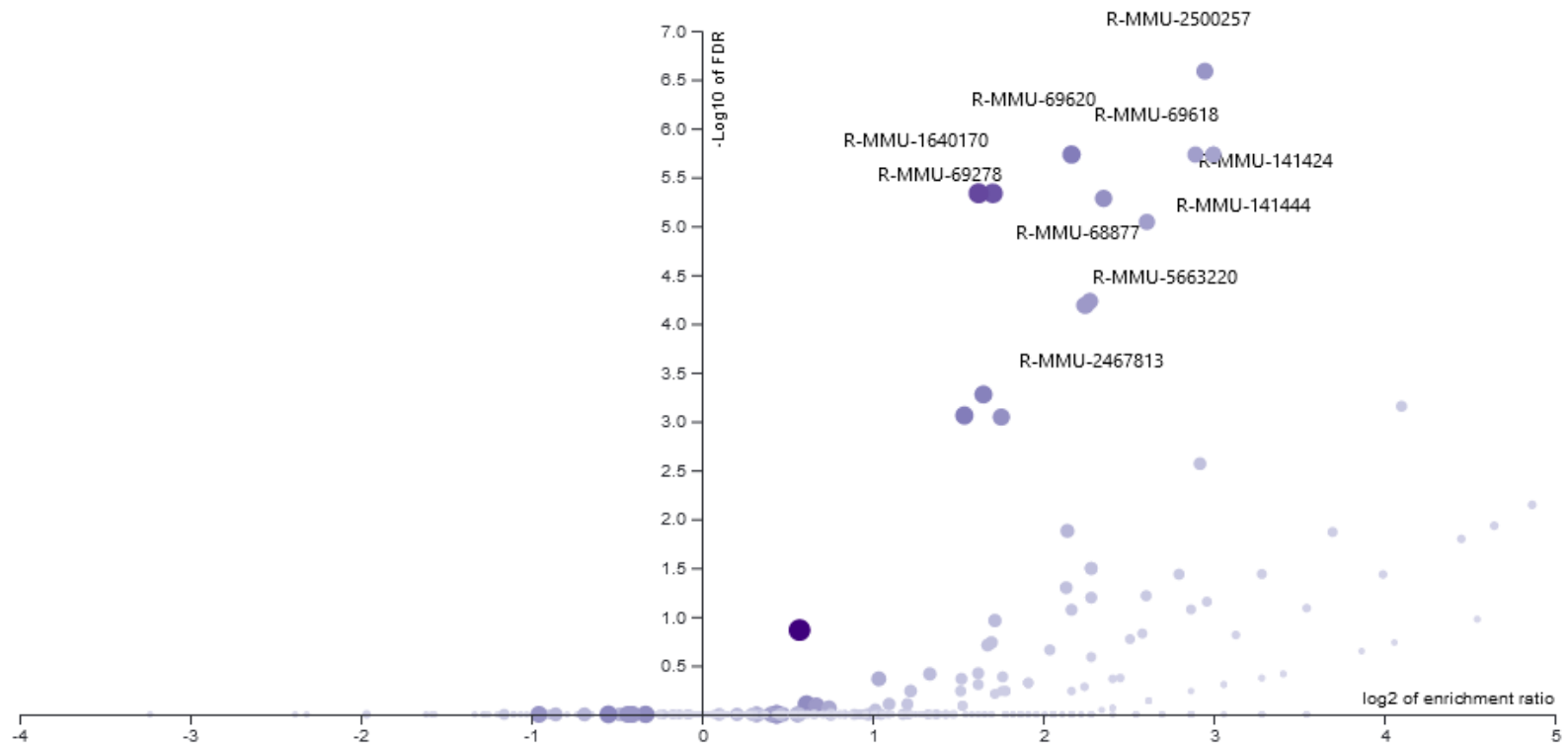
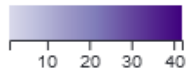
Reference Gene List

Select Reference Set ⓘ genome ▼

Upload User Reference Set File and Select ID type ⓘ

Select the ID type of reference set ▼

Hands On Pathway Analysis



Label:

geneSet



PNG



Download

Draw Link



Pan & Zoom