Functional analysis of gene lists
A practical exploration of resources

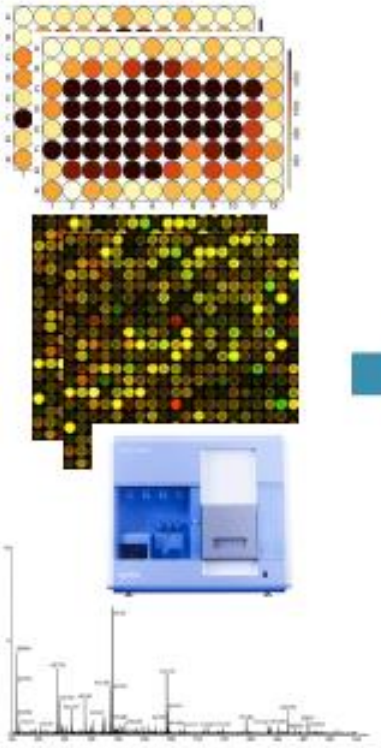
Bioinformatics Course UEB-VHIR
November 2020

Ricardo Gonzalo[1], Mireia Ferrer[1], Álex Sánchez[1,2]
Berta Miró[1], Angel Blanco[1,2]

1 Unitat d'Estadística i Bioinformàtica (UEB) VHIR
2 Departament de Genètica Microbiologia i Estadística, UB

Functional Analysis of gene lists

What's the biology behind a list of genes ?



From individual genes

From groups of genes

Functional Analysis of gene lists

TABLE OF CONTENTS

Resources for functional analysis

Resources for functional analysis

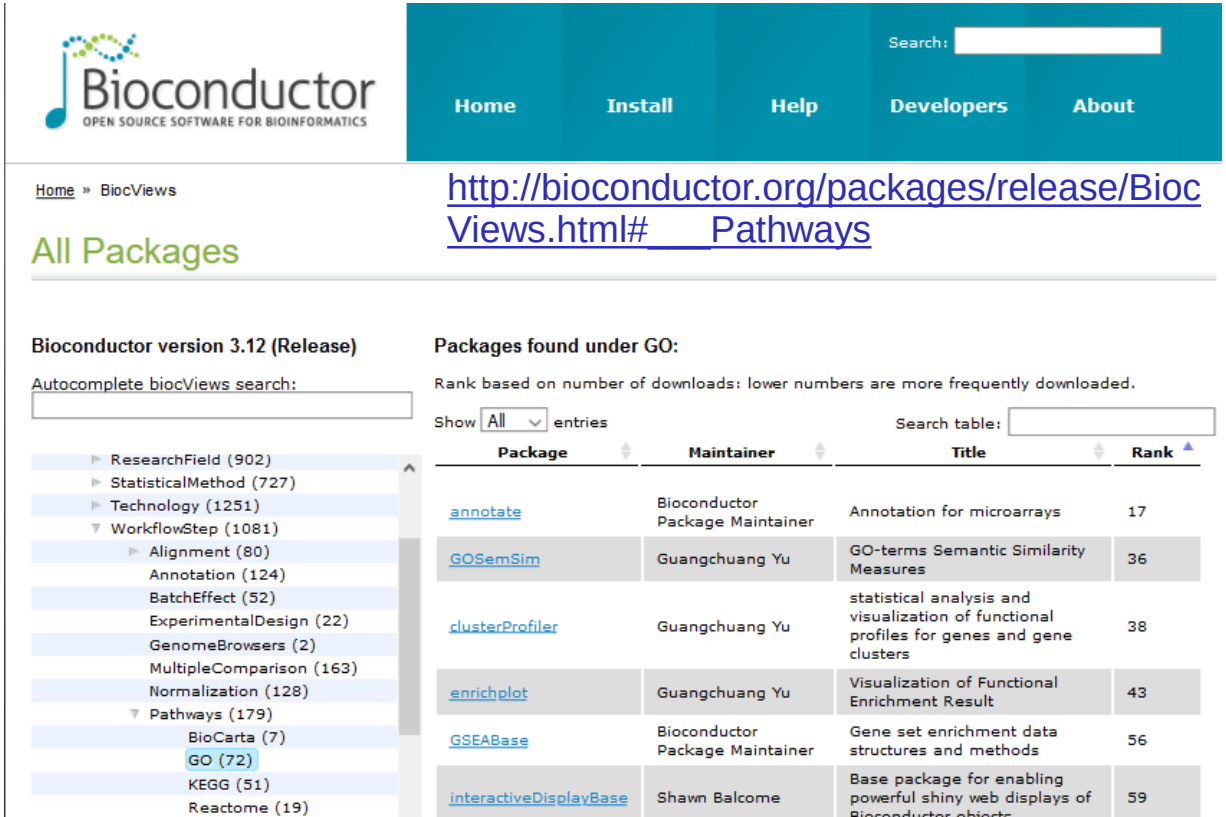
Web tools

- 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/



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, visualisation, post-processing

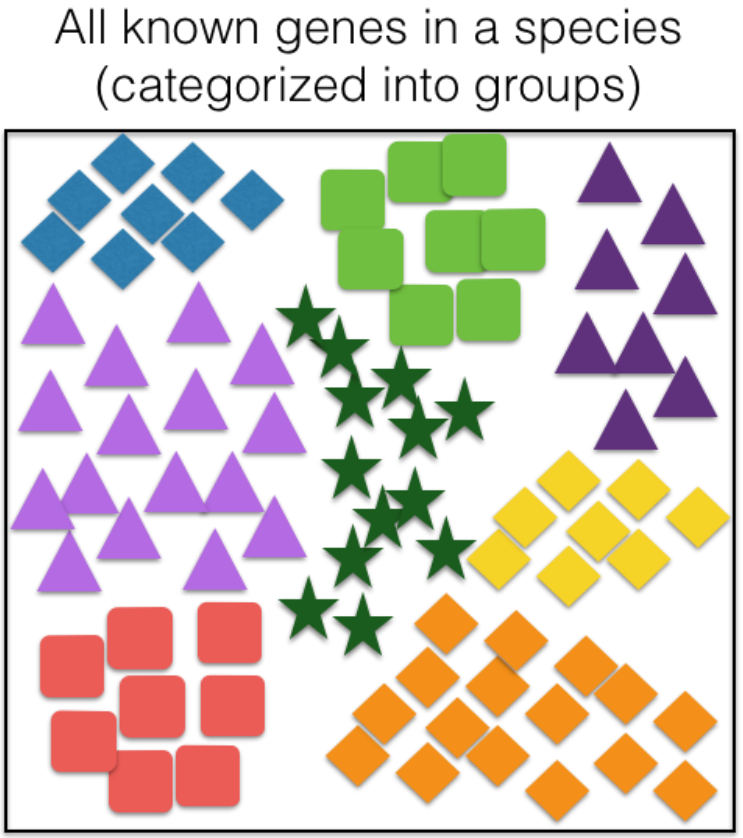
Resources for functional analysis

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

- The Gene Ontology
- Pathway databases
- Disease-related DB
- Cell/tissue markers
- Chromosome locations
- Transcription factor DB
- Micro-RNA / targets DB



Resources for functional analysis

Resources of functional annotations

- Gene Ontology (GO)
  - Cellular Components / Biological Processes / Molecular Function
- Pathways Databases
  - KEGG, Reactome, WikiPathways PantherDB
- Disease-related
  - DisGeNet, Disease Ontology
- Cell/Tissue markers databases
  - CellMarker, Cellfinder
- Chromosomal location
- Transcription Factors binding sites DBs, TRANSFAC
- MicroRNA-Target interactions DB

Resources for functional analysis

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,

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.

Browsing the Gene Ontology

Functional Analysis of gene lists

Functional Analysis of gene lists

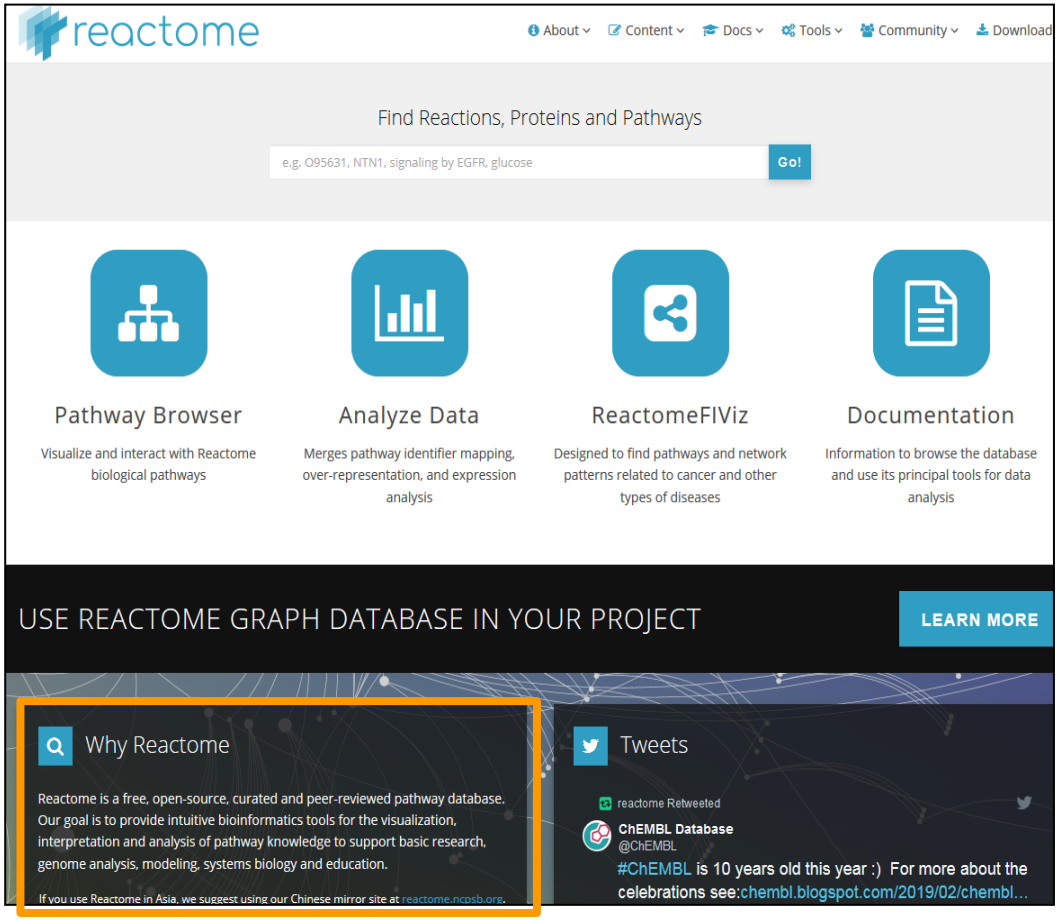
Queries that can be done with the GO

- 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 Reactome Pathways database

5. Introduction to Reactome and Pathway visualization

https://reactome.org/

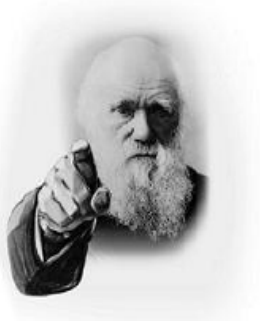


Tutorials: http://www.reactome.org/userguide/Usersguide.html
https://www.ebi.ac.uk/training/online/

5. Introduction to Reactome and Pathway visualization

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

UniProt ChEBI PubMed Ensembl the Gene Ontology

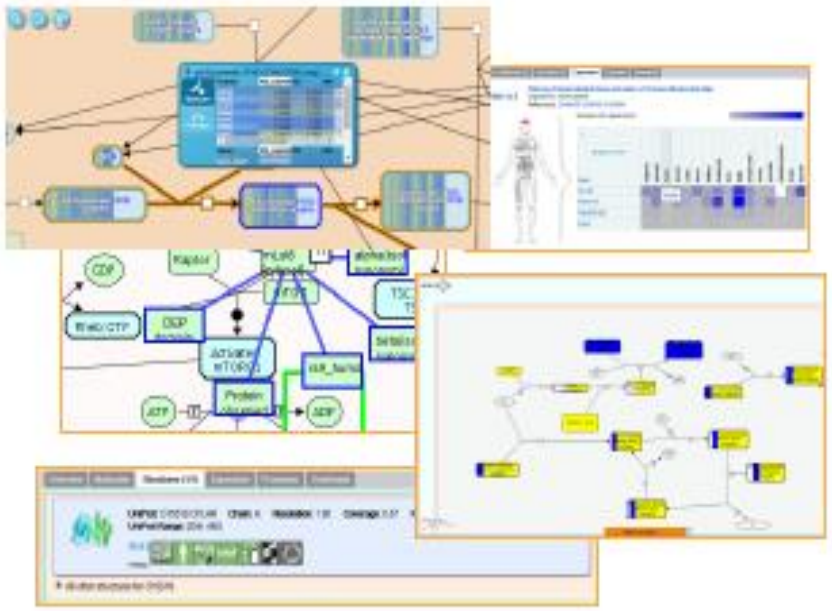
5. Introduction to Reactome and Pathway visualization

Reactome Tools

- Interactive Pathway Browser

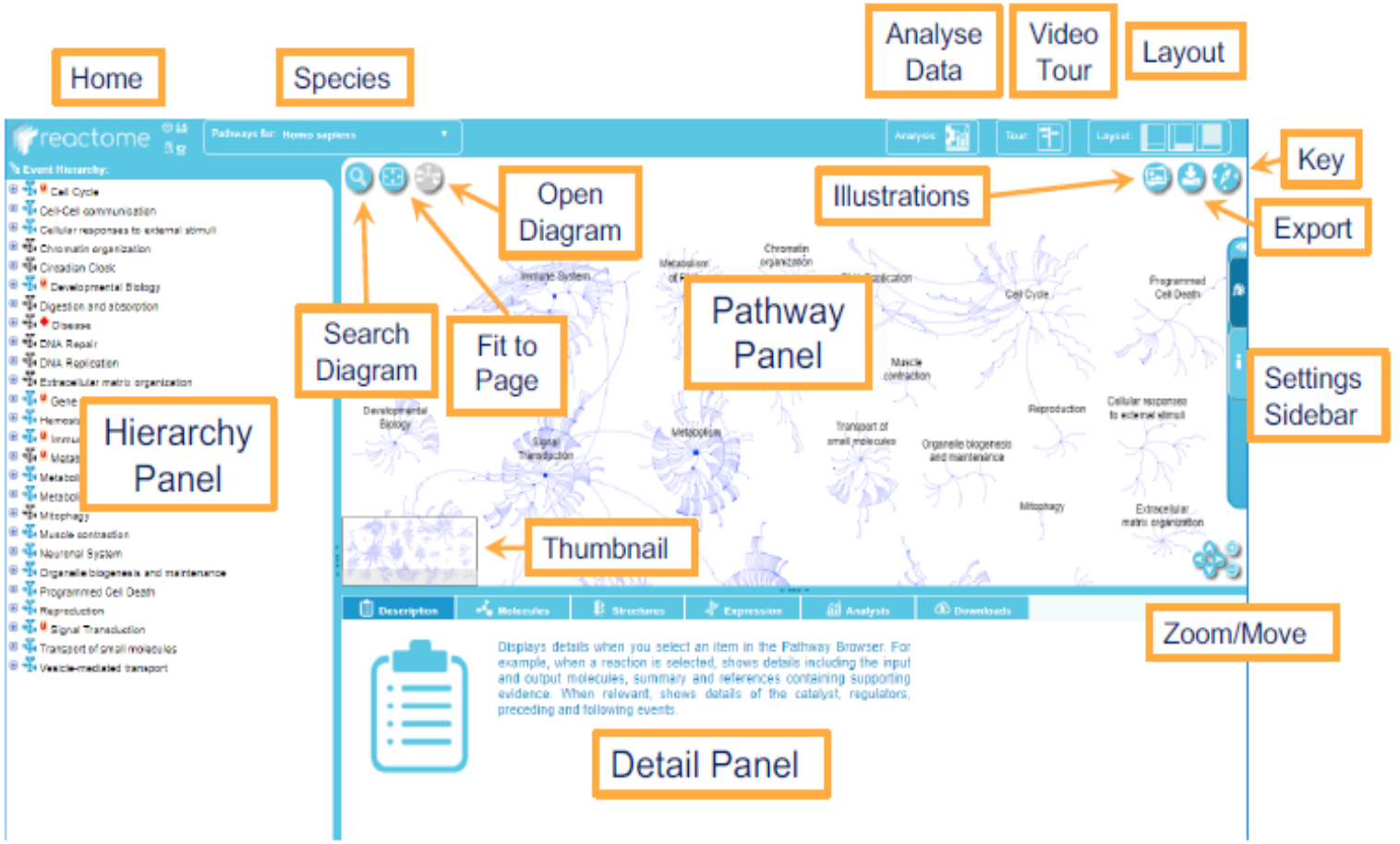
- Analysis
  - Over-representation
  - Pathway topology
  - Expression overlay

- Molecular Interaction overlay
- Species Comparison

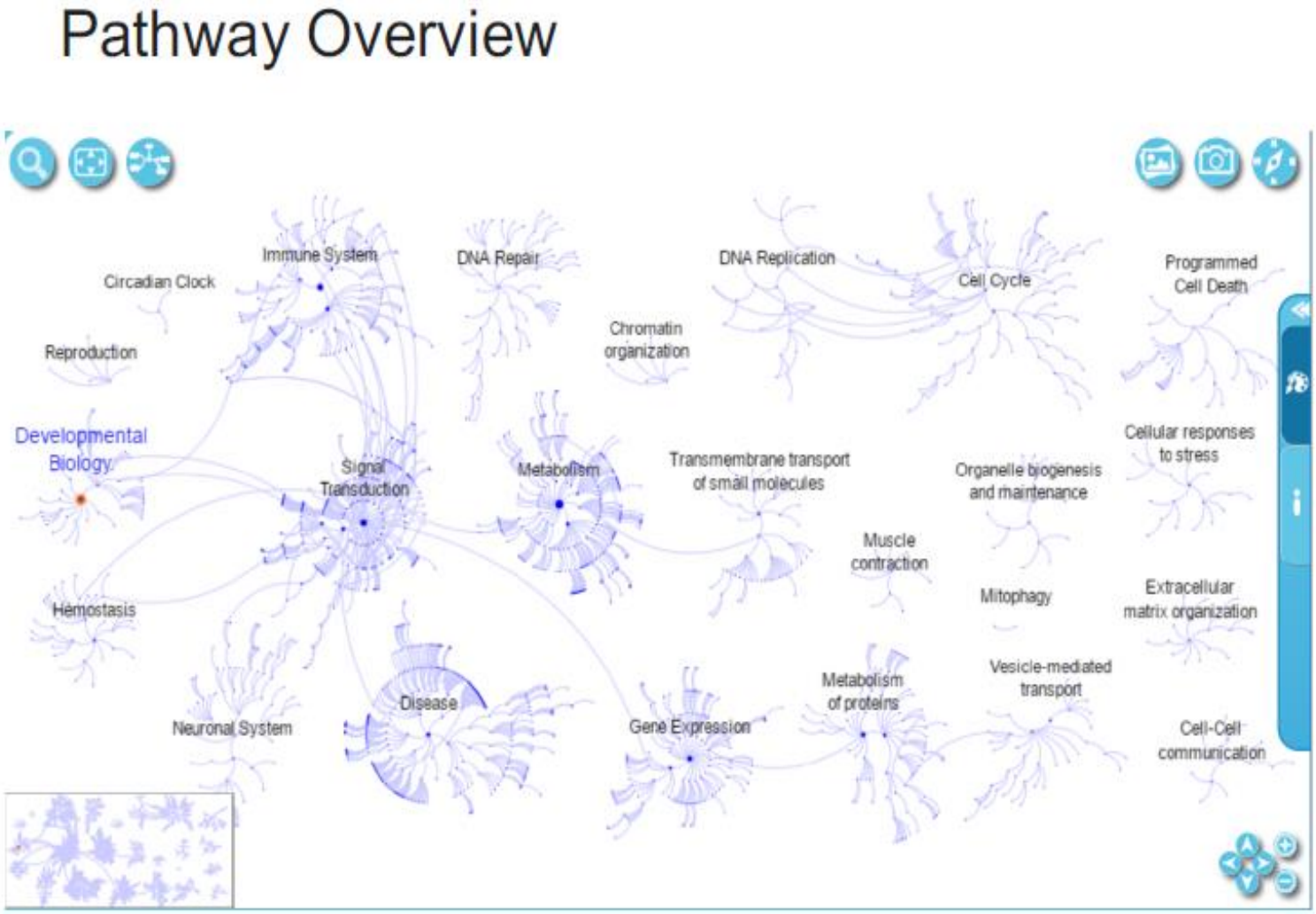


5. Introduction to Reactome and Pathway visualization

The Pathway Browser

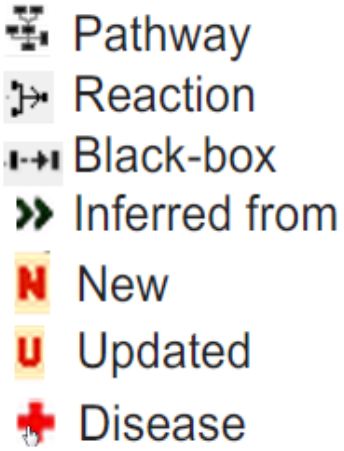


5. Introduction to Reactome and Pathway visualization



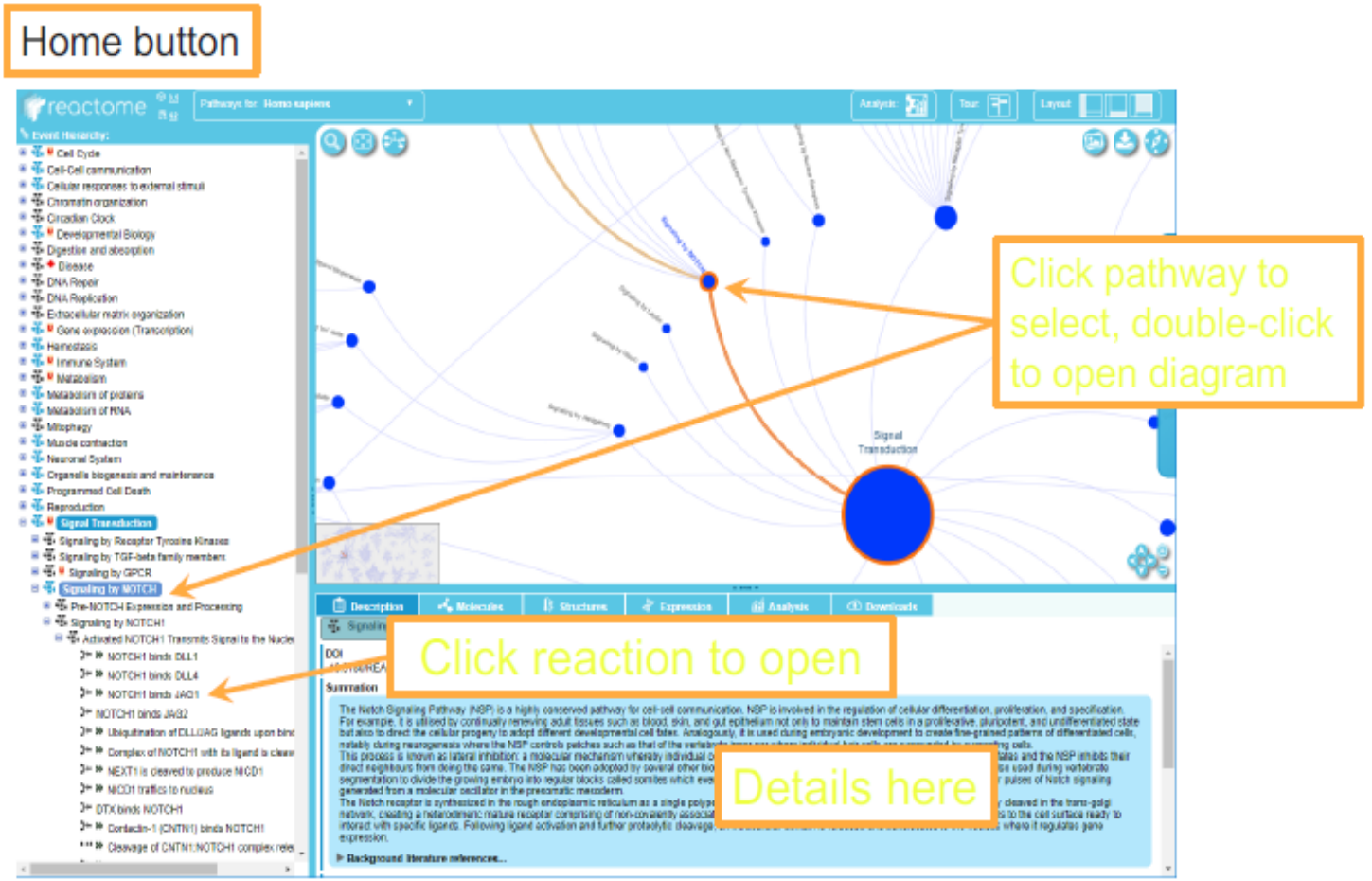
5. Introduction to Reactome and Pathway visualization

Hierarchy Panel

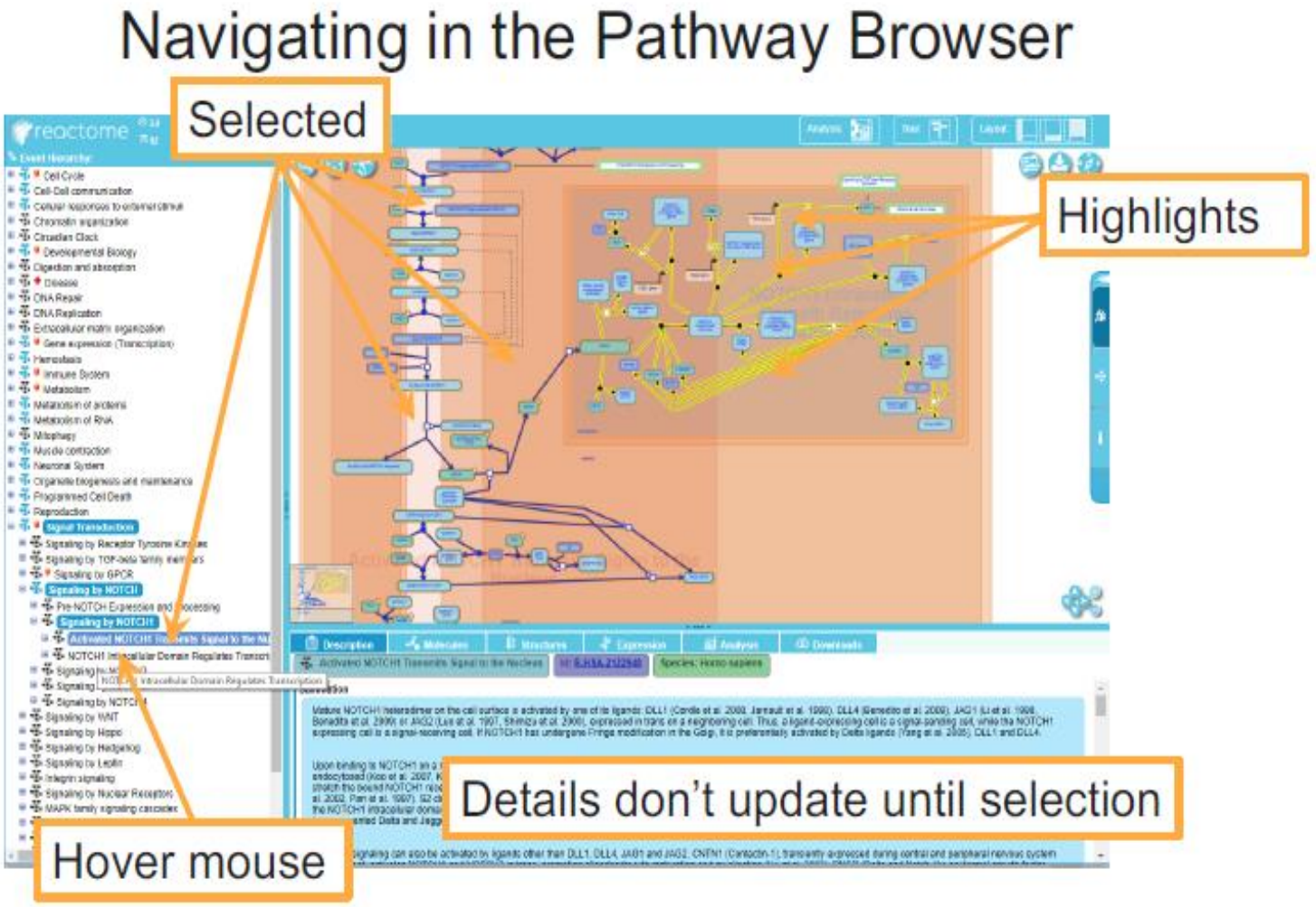


5. Introduction to Reactome and Pathway visualization

Navigating in the Pathway Browser



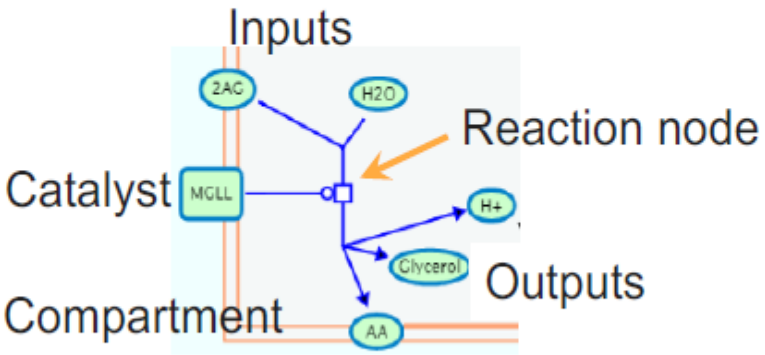
5. Introduction to Reactome and Pathway visualization

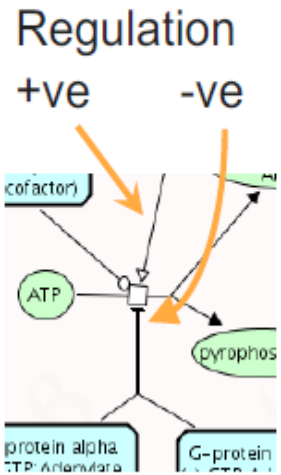


5. Introduction to Reactome and Pathway visualization

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

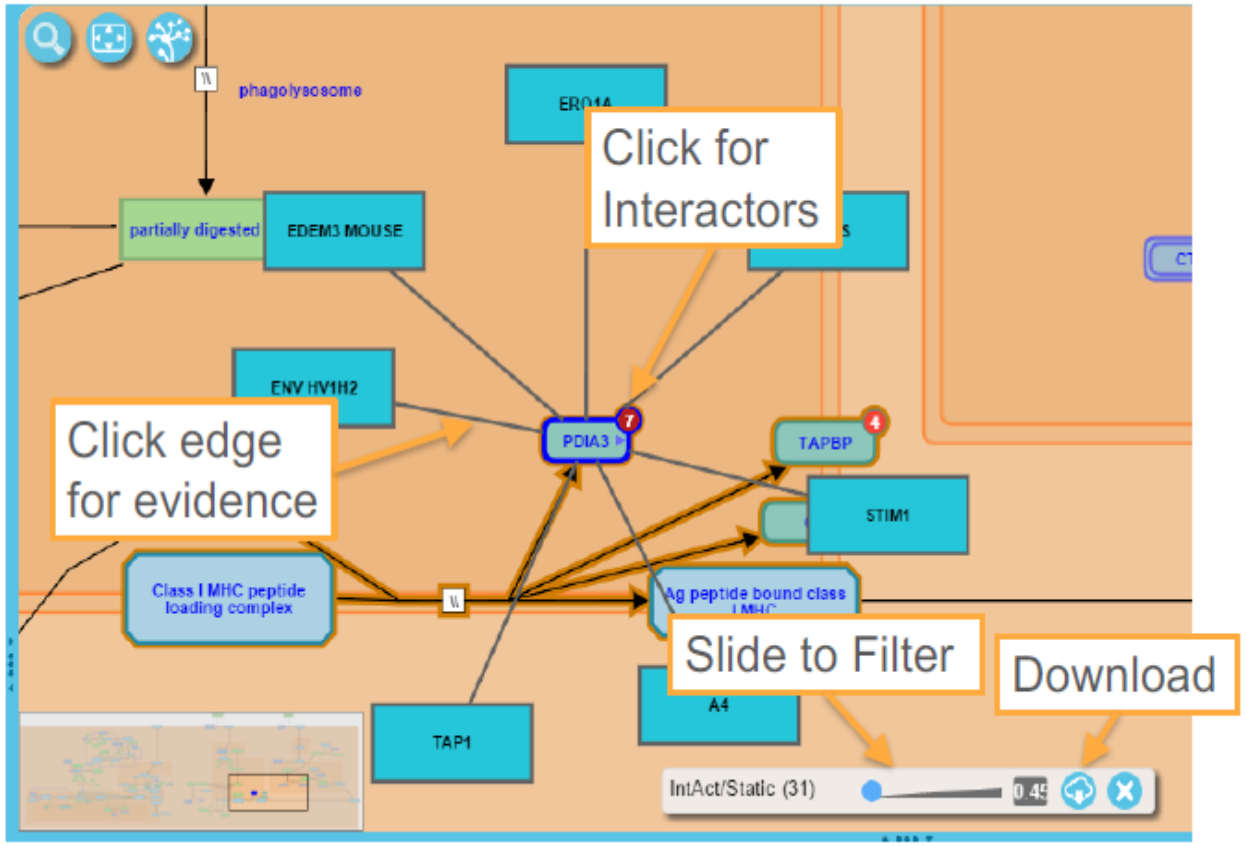




Transition Binding Dissociation Omitted Uncertain

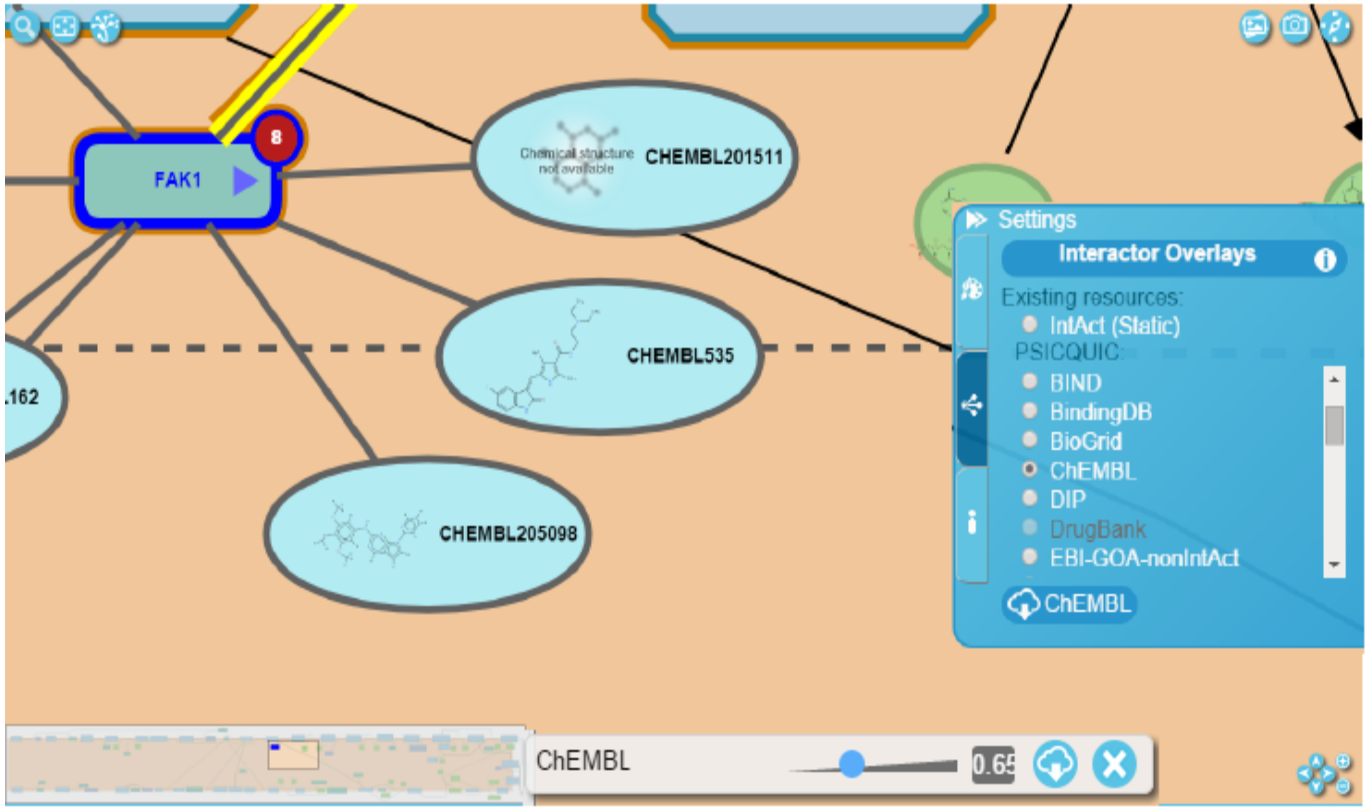
5. Introduction to Reactome and Pathway visualization

Interactors

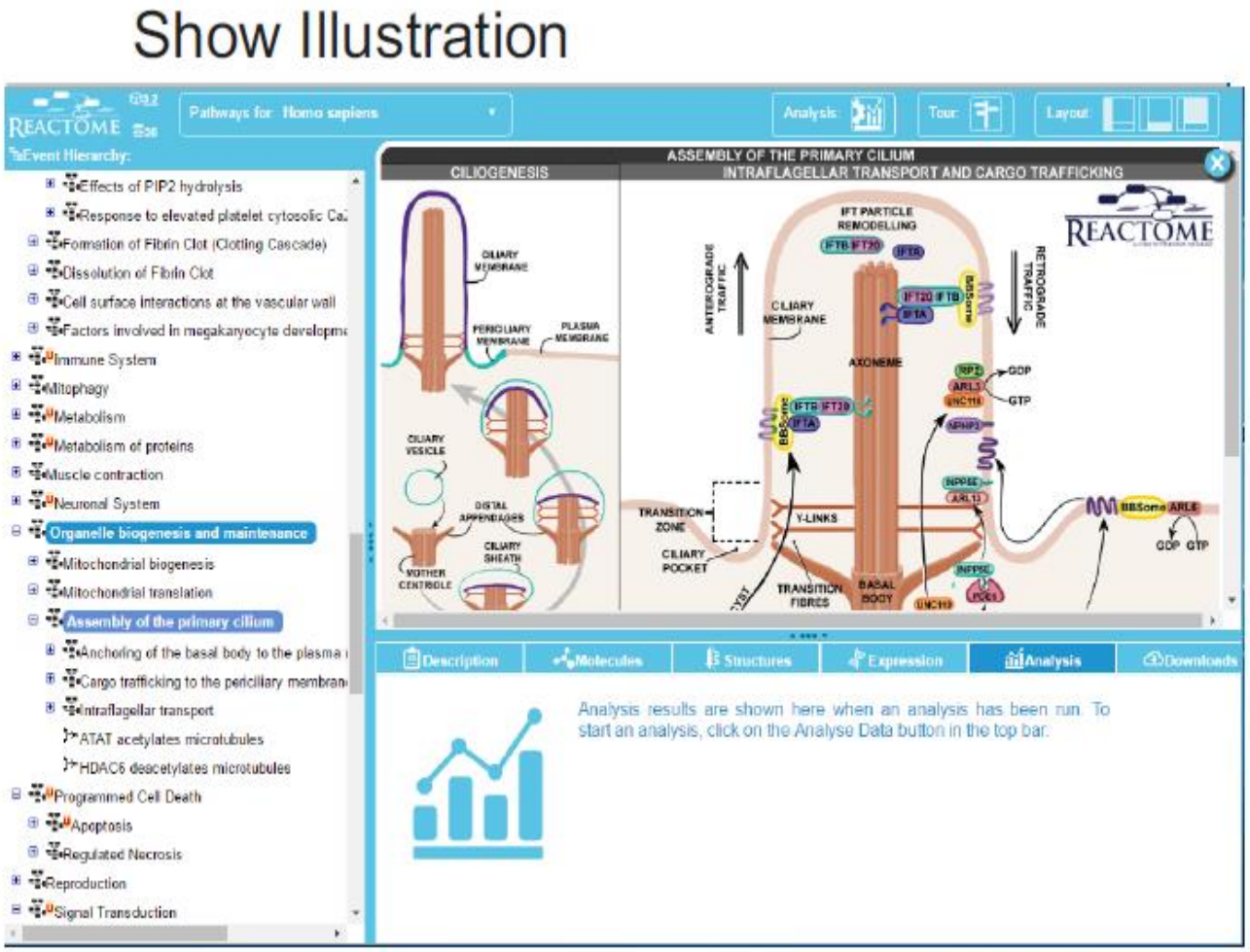


5. Introduction to Reactome and Pathway visualization

Molecular Interaction Overlay – Set source



5. Introduction to Reactome and Pathway visualization



5. Introduction to Reactome and Pathway visualization

Searching the Reactome

Simple text search

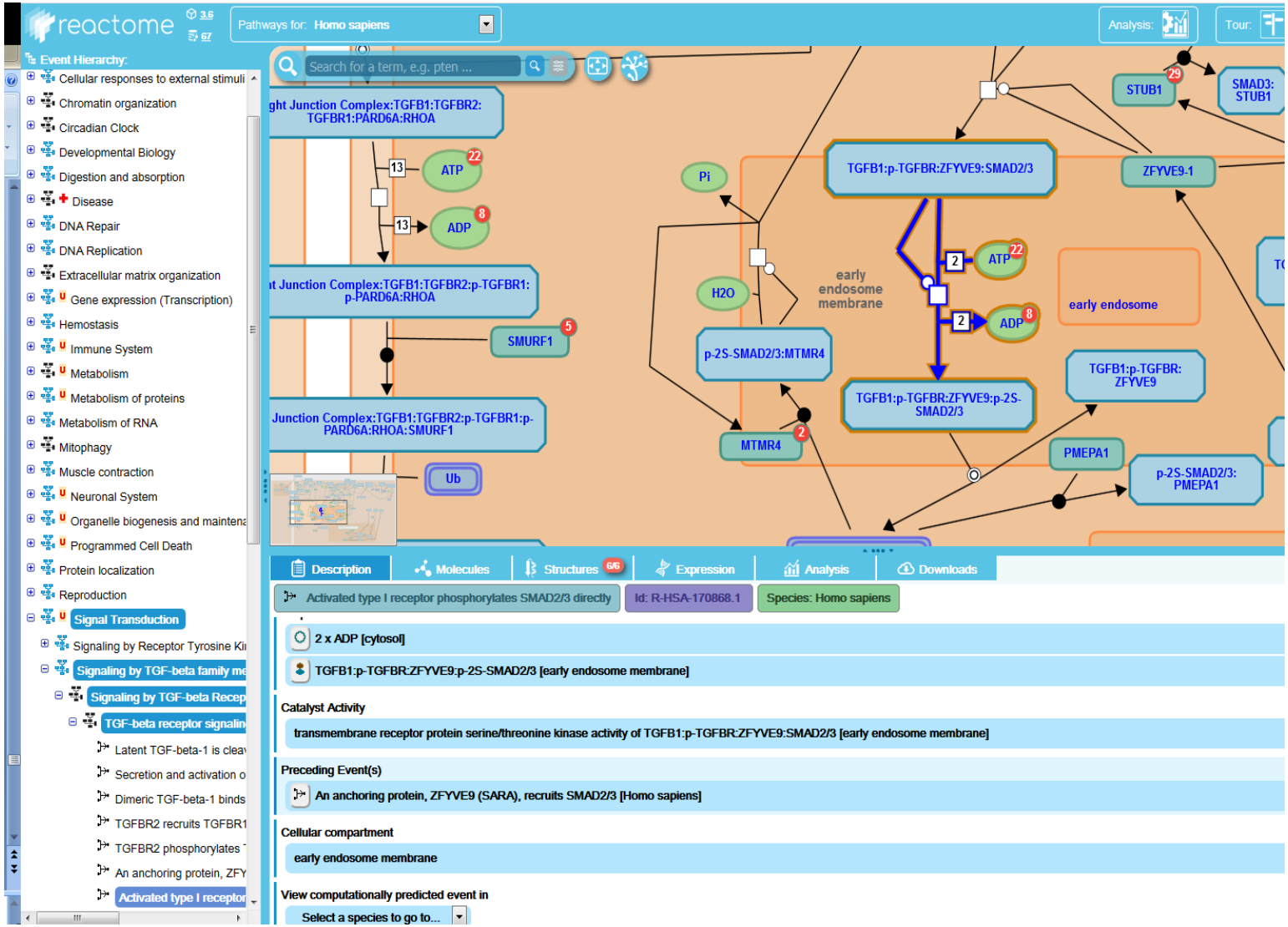
The simple text search tool is located top right of the Home page. Several other Reactome pages include this search tool. To search Reactome type a word, phrase or identifier in the search box:

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

5. Introduction to Reactome and Pathway visualization

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?


reactome
Pathways for: Homo sapiens
Analysis:
Tour:
Event Hierarchy:
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)
Hemostasis
Immune System
Metabolism
Metabolism of proteins
Metabolism of RNA
Mitophagy
Muscle contraction
Neuronal System
Organelle biogenesis and maintena
Programmed Cell Death
Protein localization
Reproduction
Signal Transduction
Signaling by Receptor Tyrosine Ki
Signaling by TGF-beta family me
Signaling by TGF-beta Recep
TGF-beta receptor signalin
Latent TGF-beta-1 is cleav
Secretion and activation o
Dimeric TGF-beta-1 binds
TGFBR2 recruits TGFBR1
TGFBR2 phosphorylates
An anchoring protein, ZFY
Activated type I receptor
Search for a term, e.g. pten ...
ght Junction Complex:TGFB1:TGFBR2:
TGFBR1:PARD6A:RHOA
ATP
ADP
t Junction Complex:TGFB1:TGFBR2:p-TGFBR1:
p-PARD6A:RHOA
SMURF1
Junction Complex:TGFB1:TGFBR2:p-TGFBR1:p-
PARD6A:RHOA:SMURF1
Ub
Pi
H2O
early
endosome
membrane
TGFB1:p-TGFBR:ZFYVE9:SMAD2/3
ZFYVE9-1
STUB1
SMAD3:
STUB1
early endosome
p-2S-SMAD2/3:MTMR4
MTMR4
TGFB1:p-TGFBR:ZFYVE9:p-2S-
SMAD2/3
TGFB1:p-TGFBR:
ZFYVE9
PMEPA1
p-2S-SMAD2/3:
PMEPA1
Description
Molecules
Structures
Expression
Analysis
Downloads
Activated type I receptor phosphorylates SMAD2/3 directly
Id: R-HSA-170868.1
Species: Homo sapiens
2 x ADP [cytosol]
TGFB1:p-TGFBR:ZFYVE9:p-2S-SMAD2/3 [early endosome membrane]
Catalyst Activity
transmembrane receptor protein serine/threonine kinase activity of TGFB1:p-TGFBR:ZFYVE9:SMAD2/3 [early endosome membrane]
Preceding Event(s)
An anchoring protein, ZFYVE9 (SARA), recruits SMAD2/3 [Homo sapiens]
Cellular compartment
early endosome membrane
View computationally predicted event in
Select a species to go to...

Case Study

PLoS One. 2014 Jun 13;9(6):e99625. doi: 10.1371/journal.pone.0099625. eCollection 2014.

RNA-Seq transcriptome profiling identifies CRISPLD2 as a glucocorticoid responsive gene that modulates cytokine function in airway smooth muscle cells.

Himes BE[1], Jiang X[2], Wagner P[2], Hu R[2], Wang Q[2], Klanderman B[3], Whitaker RM[4], Duan Q[4], Lasky-Su J[4], Nikolos C[5], Jester W[5], Johnson M[5], Panettieri RA Jr[5], Tantisira KG[4], Weiss ST[6], Lu Q[2].

Introduction to Functional Analysis

The differential expression analysis performed on RNA-seq data (airway dataset) returned some genes differentially expressed in DEX-**Treated vs Untreated 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.**

```
res <- results(dds, contrast=c("dex","trt","untrt"))
res
```

```
log2 fold change (MLE): dex trt vs untrt
Wald test p-value: dex trt vs untrt
DataFrame with 31604 rows and 6 columns
                   baseMean log2FoldChange     lfcSE      stat      pvalue       padj
                  <numeric>      <numeric> <numeric> <numeric>   <numeric>  <numeric>
ENSG00000000003.14 739.940717     -0.3611537  0.106869 -3.379419 0.000726392 0.00531137
ENSG00000000419.12 511.735722      0.2063147  0.128665  1.603509 0.108822318 0.29318870
ENSG00000000457.13 314.194855      0.0378308  0.158633  0.238479 0.811509461 0.92255697
ENSG00000000460.16  79.793622     -0.1152590  0.314991 -0.365912 0.714430444 0.87298038
ENSG00000000938.12   0.307267     -1.3691185  3.503764 -0.390757 0.695977205         NA
...                       ...            ...       ...       ...         ...        ...
ENSG00000285979.1   38.353886      0.3423657  0.359511  0.952310    0.340940   0.600750
ENSG00000285987.1    1.562508      0.7064145  1.547295  0.456548    0.647996         NA
ENSG00000285990.1    0.642315      0.3647333  3.433276  0.106235    0.915396         NA
ENSG00000285991.1   11.276284     -0.1165515  0.748601 -0.155692    0.876275   0.952921
ENSG00000285994.1    3.651041     -0.0960094  1.068660 -0.089841    0.928414         NA
```

Introduction to Functional Analysis

Exercise 2

• 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**

Tip: You can use excel (with caution) or **R** to filter your TopTable

• Under these criteria, how many genes are up-regulated? And how many down-regulated?

• What is the **Entrez ID** of your top **most up/down-regulated gene**? And its **UniProt ID** (SwissProt)?

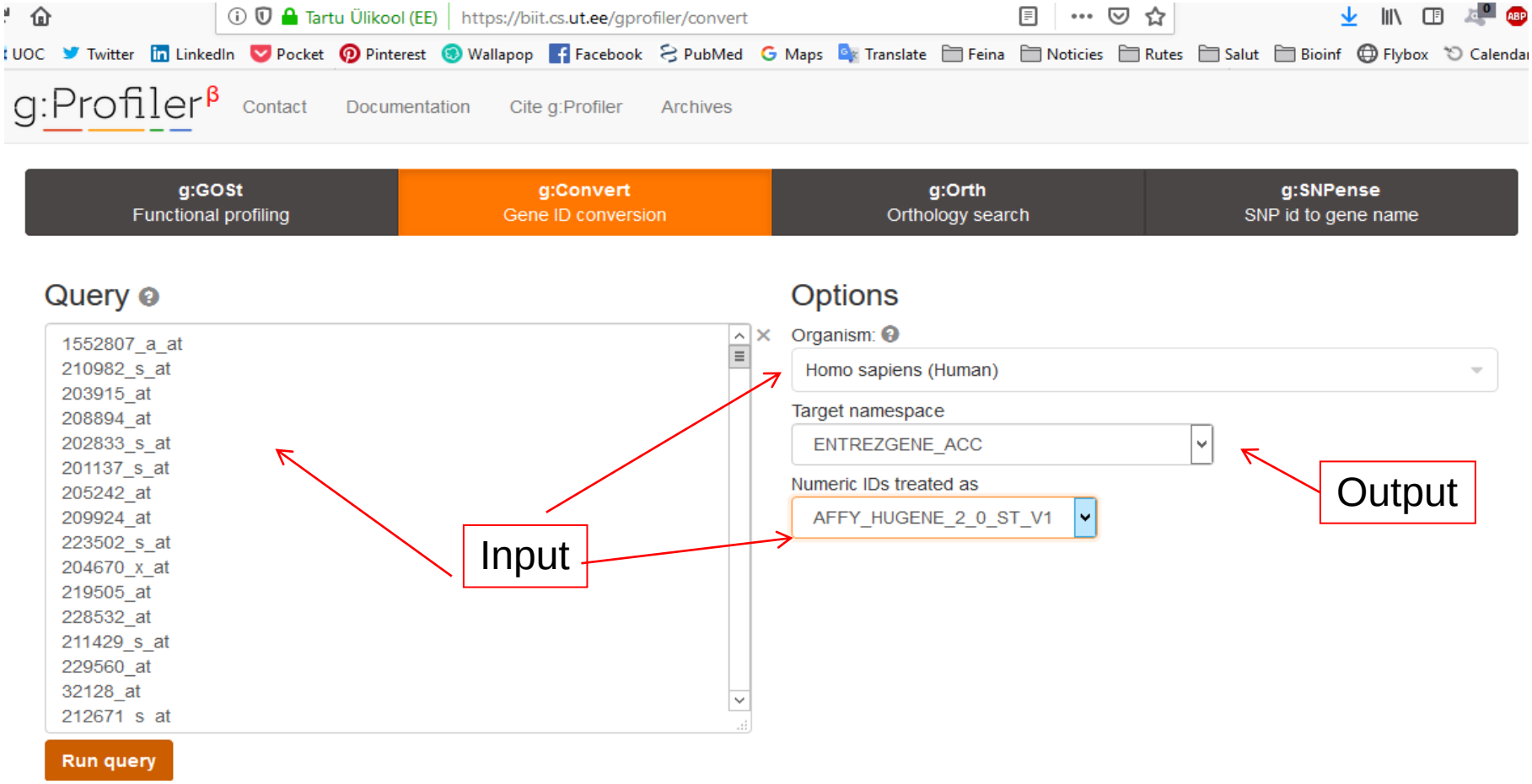
Tip: You can use g:Convert tool from g:Profiler for ID conversion

Functional annotation of individual genes

Id mapping
Searching for a gene in different databases

2. Preparation of the Gene list

ID mapping with g:Profiler



2. Preparation of the Gene list

ID mapping with g:Profiler

Run query

Export to CSV | Show query URL | Copy URL to clipboard

Export results as *gProfiler_IDconvert_GSE7586_sel.csv*

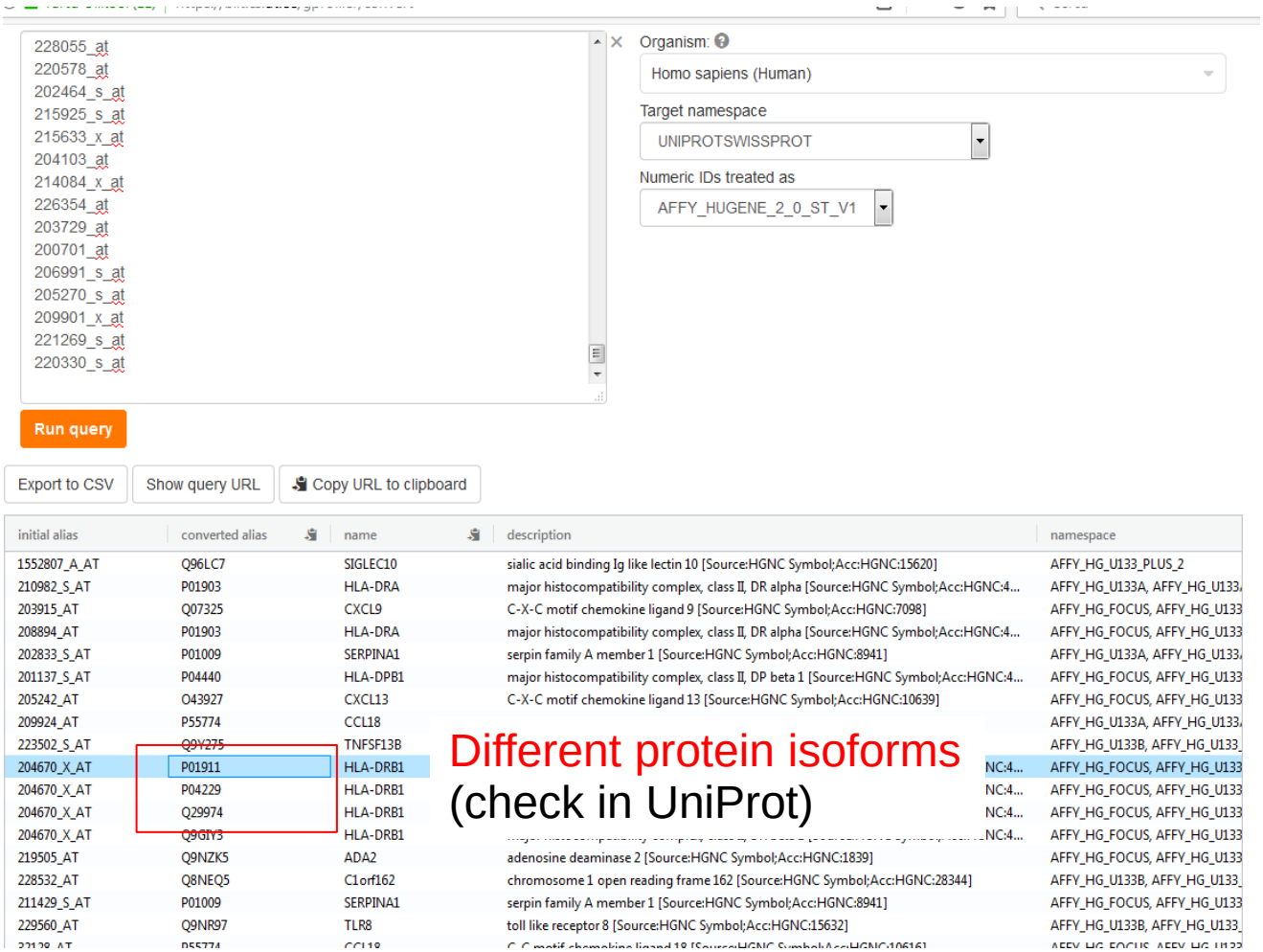
initial alias	converted alias	name		namespace
1552807_A_AT	89790	SIGLEC10		AFFY_HG_U133_PLUS_2
210982_S_AT	3122	HLA-DRA	...c:HGNC:4...	AFFY_HG_U133A, AFFY_HG_U133
203915_AT	4283	CXCL9		AFFY_HG_FOCUS, AFFY_HG_U133
208894_AT	3122	HLA-DRA	major histocompatibility complex, class II, DR alpha [Source:HGNC Symbol;Acc:HGNC:4...	AFFY_HG_FOCUS, AFFY_HG_U133
202833_S_AT	5265	SERPINA1	serpin family A member 1 [Source:HGNC Symbol;Acc:HGNC:8941]	AFFY_HG_U133A, AFFY_HG_U133
201137_S_AT	3115	HLA-DPB1	...class II, DP beta 1 [Source:HGNC Symbol;Acc:HGNC:4...	AFFY_HG_FOCUS, AFFY_HG_U133
205242_AT	10563	CXCL13	...Source:HGNC Symbol;Acc:HGNC:10639]	AFFY_HG_FOCUS, AFFY_HG_U133
209924_AT	6362	CCL18	...ource:HGNC Symbol;Acc:HGNC:10616]	AFFY_HG_U133A, AFFY_HG_U133
223502_S_AT	10673	TNFSF13B	TNF superfamily member 13b [Source:HGNC Symbol;Acc:HGNC:11929]	AFFY_HG_U133B, AFFY_HG_U133_
204670_X_AT	105369230	HLA-DRB1	major histocompatibility complex, class II, DR beta 1 [Source:HGNC Symbol;Acc:HGNC:4...	AFFY_HG_FOCUS, AFFY_HG_U133
204670_X_AT	3123	HLA-DRB1	major histocompatibility complex, class II, DR beta 1 [Source:HGNC Symbol;Acc:HGNC:4...	AFFY_HG_FOCUS, AFFY_HG_U133
219505_AT	51816	ADA2	adenosine deaminase 2 [Source:HGNC Symbol;Acc:HGNC:1839]	AFFY_HG_FOCUS, AFFY_HG_U133
228532_AT	128346	C1orf162	chromosome 1 open reading frame 162 [Source:HGNC Symbol;Acc:HGNC:28344]	AFFY_HG_U133B, AFFY_HG_U133_
211429_S_AT	5265	SERPINA1	serpin family A member 1 [Source:HGNC Symbol;Acc:HGNC:8941]	AFFY_HG_FOCUS, AFFY_HG_U133
209969_S_AT	6772	STAT1	signal transducer and activator of transcription 1 [Source:HGNC Symbol;Acc:HGNC:11362]	AFFY_HG_U133A, AFFY_HG_U133
209619_AT	972	CD74	CD74 molecule [Source:HGNC Symbol;Acc:HGNC:1697]	AFFY_HG_FOCUS, AFFY_HG_U133
206881_S_AT	None	None	None	
207111_AT	2015	ADGRE1	adhesion G prote...ymbol;Acc:HGNC:3336]	AFFY_HG_FOCUS, AFFY_HG_U133
212788_X_AT	2512	FTL	ferritin light chain [Source:HGNC Symbol;Acc:HGNC:3999]	AFFY_HG_U133A, AFFY_HG_U133
204118_AT	962	CD48	CD48 molecule [Source:HGNC Symbol;Acc:HGNC:1683]	AFFY_HG_FOCUS, AFFY_HG_U133
202902_S_AT	1520	CTSS	cathepsin S [Source:HGNC Symbol;Acc:HGNC:2545]	AFFY_HG_FOCUS, AFFY_HG_U133
229390_AT	441168	CALHM6	calcium homeostasis modulator family member 6 [Source:HGNC Symbol;Acc:HGNC:333...	AFFY_HG_U133B, AFFY_HG_U133_
204588_S_AT	9056	SLC7A7	solute carrier family 7 member 7 [Source:HGNC Symbol;Acc:HGNC:11065]	AFFY_HG_FOCUS, AFFY_HG_U133

Some ambiguities

Some missing

2. Preparation of the Gene list

ID mapping with g:Profiler



3. Functional annotation of individual genes

What is known of gene X?

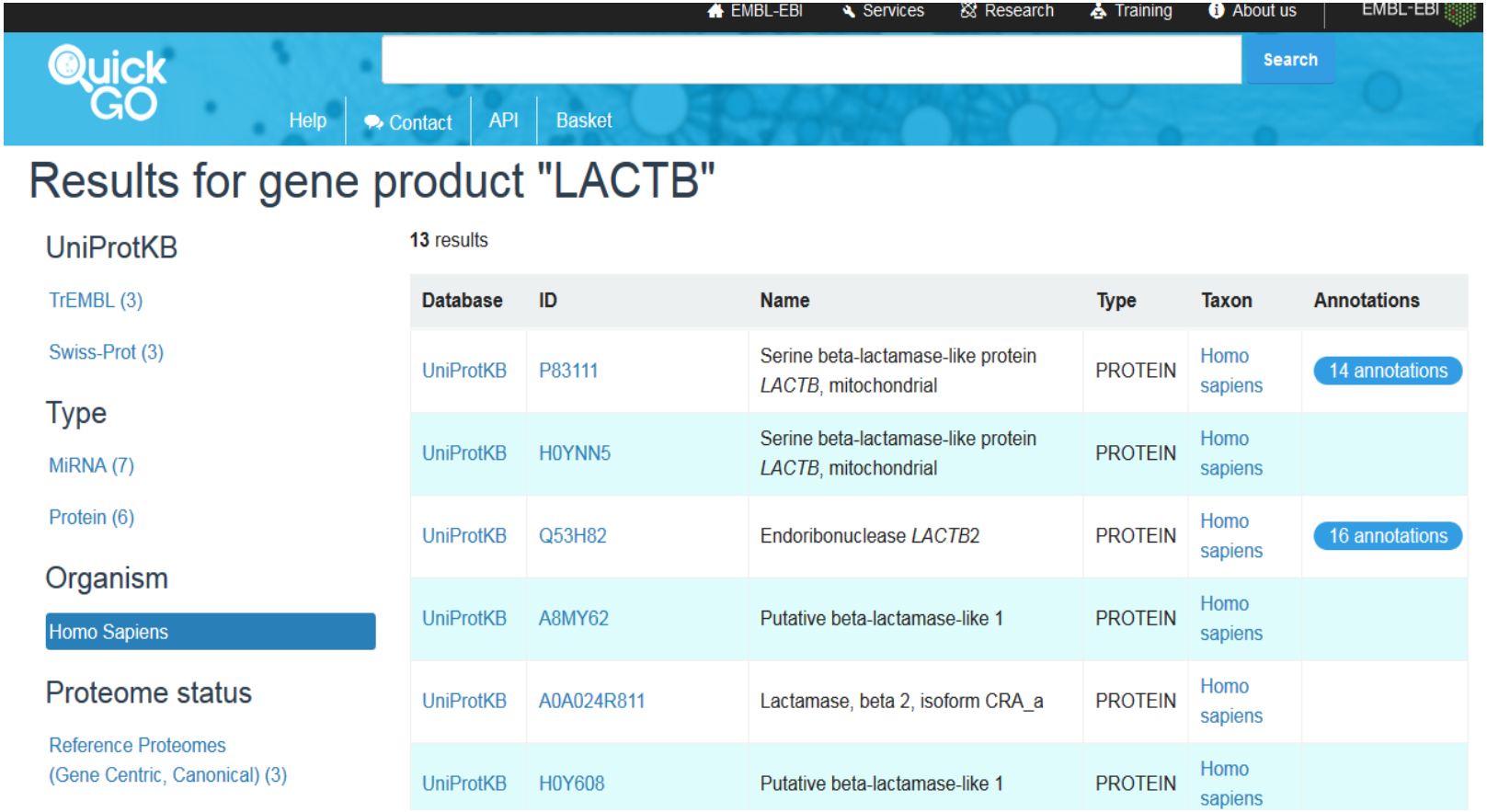
➔Get information for one of the genes in some database (Pubmed, UniProt, NCBI-Gene…)

➔ 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/

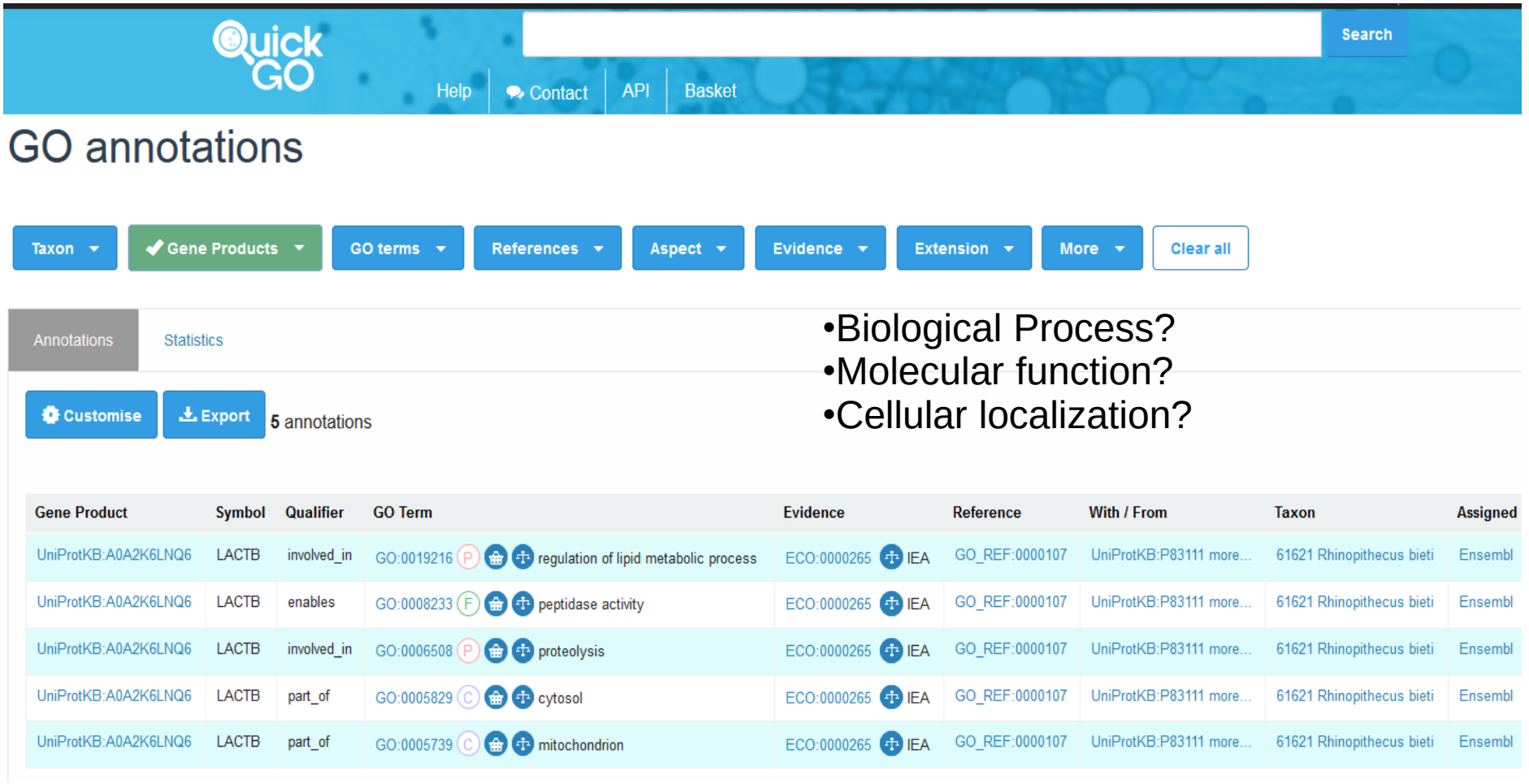
3. Functional annotation of individual genes

•Functional annotation of individual genes with QuickGO



3. Functional annotation of individual genes

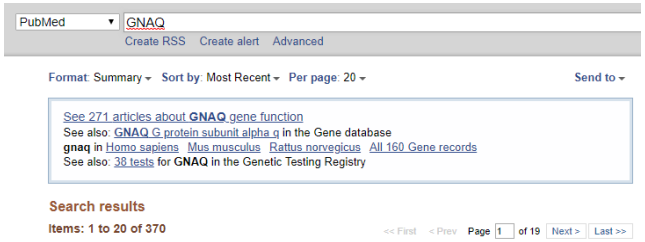
•Functional annotation of individual genes with QuickGO



Gene Product	Symbol	Qualifier	GO Term	Evidence	Reference	With / From	Taxon	Assigned
UniProtKB:A0A2K6LNQ6	LACTB	involved_in	GO:0019216 (P) regulation of lipid metabolic process	ECO:0000265 IEA	GO_REF:0000107	UniProtKB:P83111 more...	61621 Rhinopithecus bieti	Ensembl
UniProtKB:A0A2K6LNQ6	LACTB	enables	GO:0008233 (F) peptidase activity	ECO:0000265 IEA	GO_REF:0000107	UniProtKB:P83111 more...	61621 Rhinopithecus bieti	Ensembl
UniProtKB:A0A2K6LNQ6	LACTB	involved_in	GO:0006508 (P) proteolysis	ECO:0000265 IEA	GO_REF:0000107	UniProtKB:P83111 more...	61621 Rhinopithecus bieti	Ensembl
UniProtKB:A0A2K6LNQ6	LACTB	part_of	GO:0005829 (C) cytosol	ECO:0000265 IEA	GO_REF:0000107	UniProtKB:P83111 more...	61621 Rhinopithecus bieti	Ensembl
UniProtKB:A0A2K6LNQ6	LACTB	part_of	GO:0005739 (C) mitochondrion	ECO:0000265 IEA	GO_REF:0000107	UniProtKB:P83111 more...	61621 Rhinopithecus bieti	Ensembl

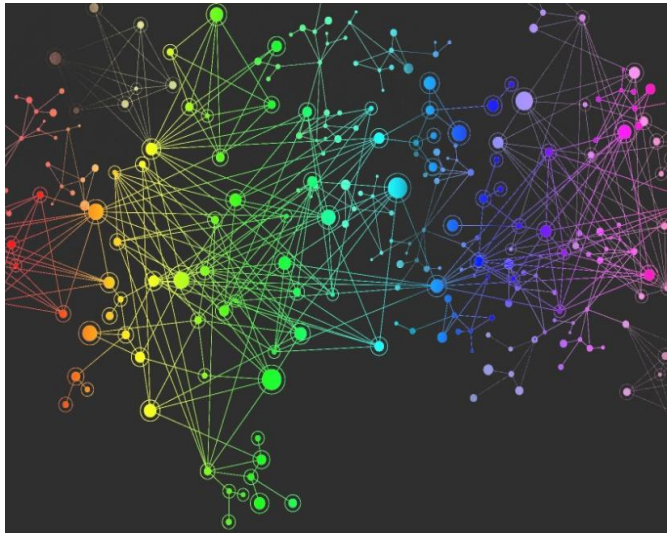
3. Functional annotation of individual genes

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



4. Enrichment analysis of selected genes

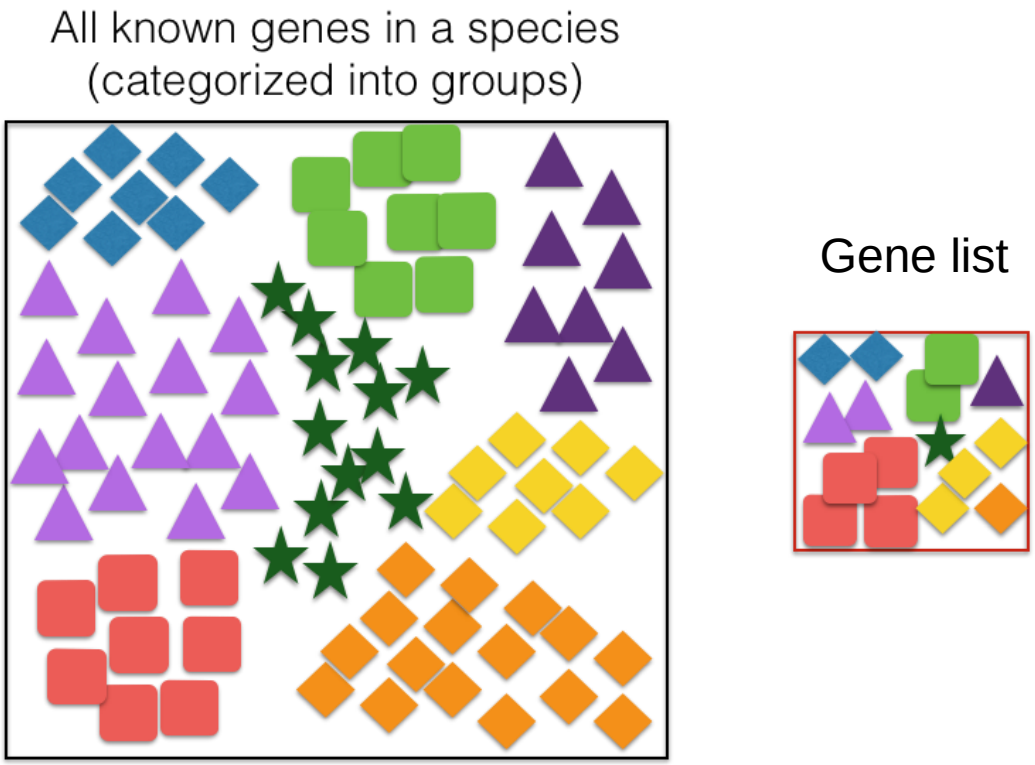
- Genes usually don’t act on their own
- Extracting information at the “group” level may be more meaningful and facilitate interpretation of the processes involved in the condition of study



Overrepresentation analysis

G:profiler
Gorilla
Reactome

• you can determine the probability of having the observed proportion of genes associated with a specific category in your gene list based on the proportion of genes associated with the same category in the background set (gene categorizations for the appropriate organism).



4. Enrichment analysis of selected genes

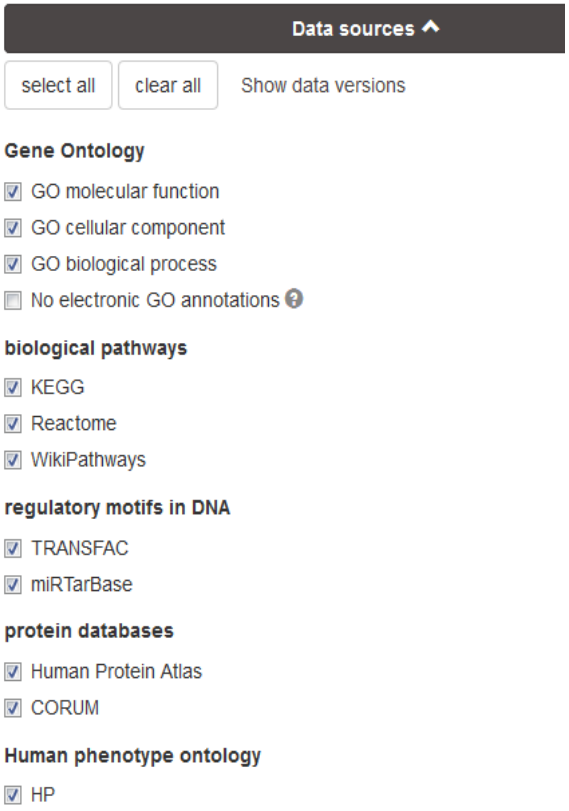
What do we need?

- ❑ 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. *.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/

4. Enrichment analysis of selected genes

g:Profiler

- ✓ Allows to perform enrichment analysis across different annotation databases
- ✓ Accepts different types of IDs as input
- ✓ Adjustment for multiple comparisons
- ✓ Source of evidence (experimental, computational, …



4. Enrichment analysis of selected genes

g:Profiler

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)

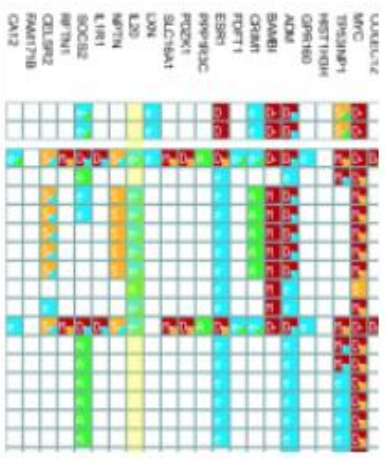
> Gene-sets (pathways/processes) that were found enriched

size of the overlap

signi-fican-ce

> Genes in input list

source	term name	term ID	n. of term genes	n. of query genes	n. of common genes	p-value
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:0048856	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	3.35e-03
BP	single-organism cellular process	GO:0044763	12236	391	282	2.32e-02
BP	cell death	GO:0008219	1941	391	69	3.17e-03
BP	programmed cell death	GO:0012501	1850	391	68	1.16e-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



4. Enrichment analysis of selected genes

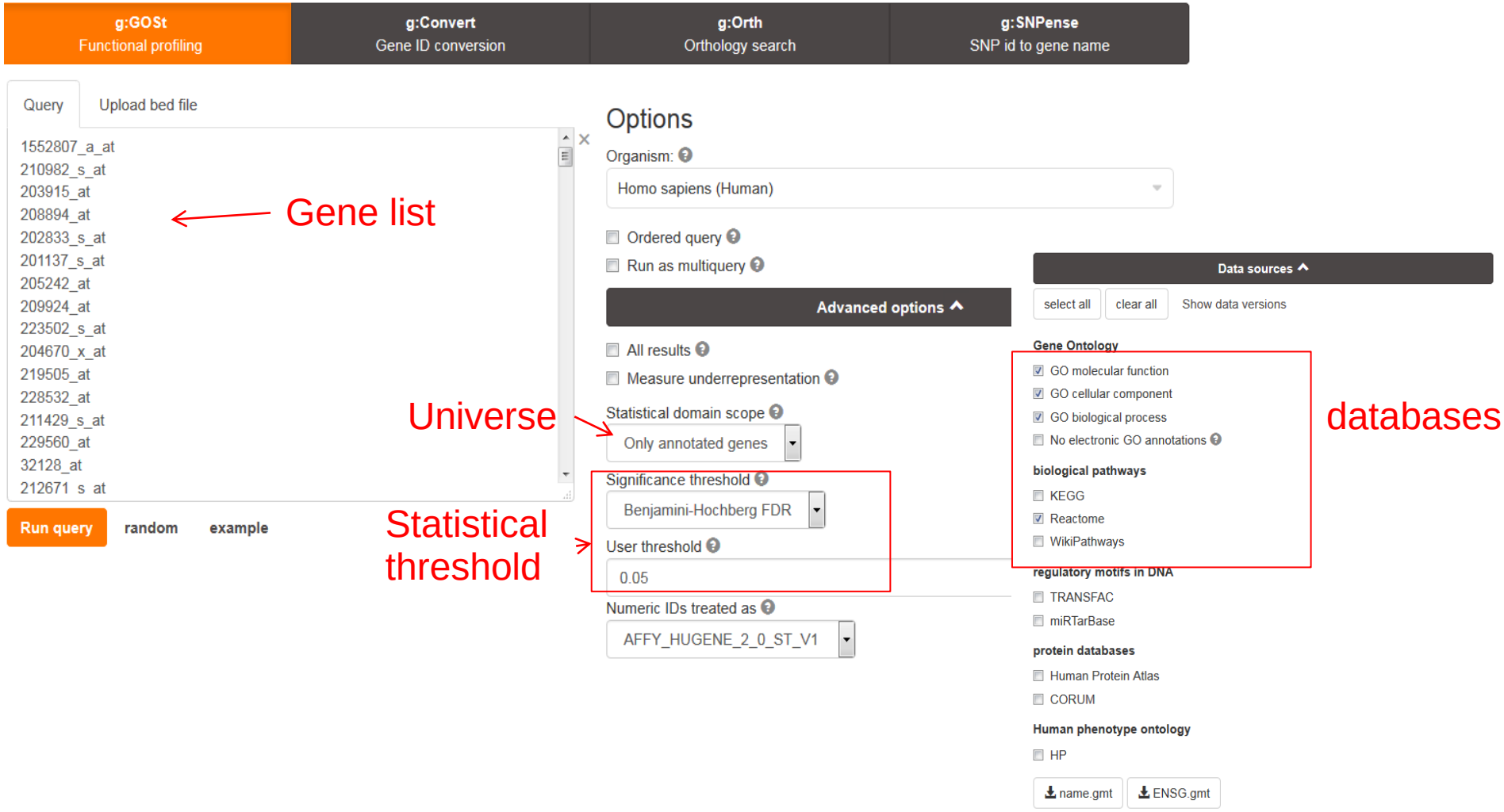
g:Profiler

What GO or Reactome terms are enriched in your gene list?

- Tip: Use g:GOSt tool in g:Profiler for functional enrichment analysis

4. Enrichment analysis of selected genes

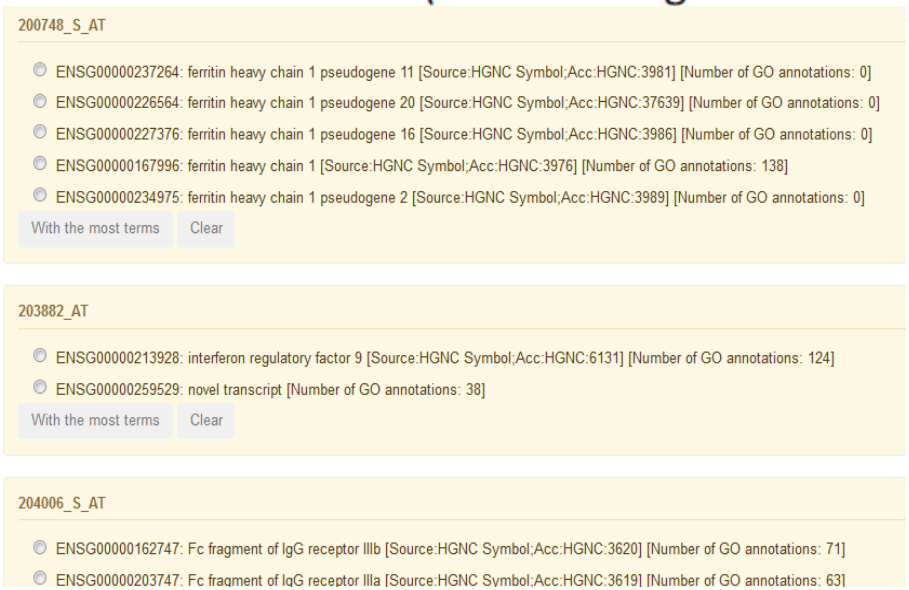
g:Profiler



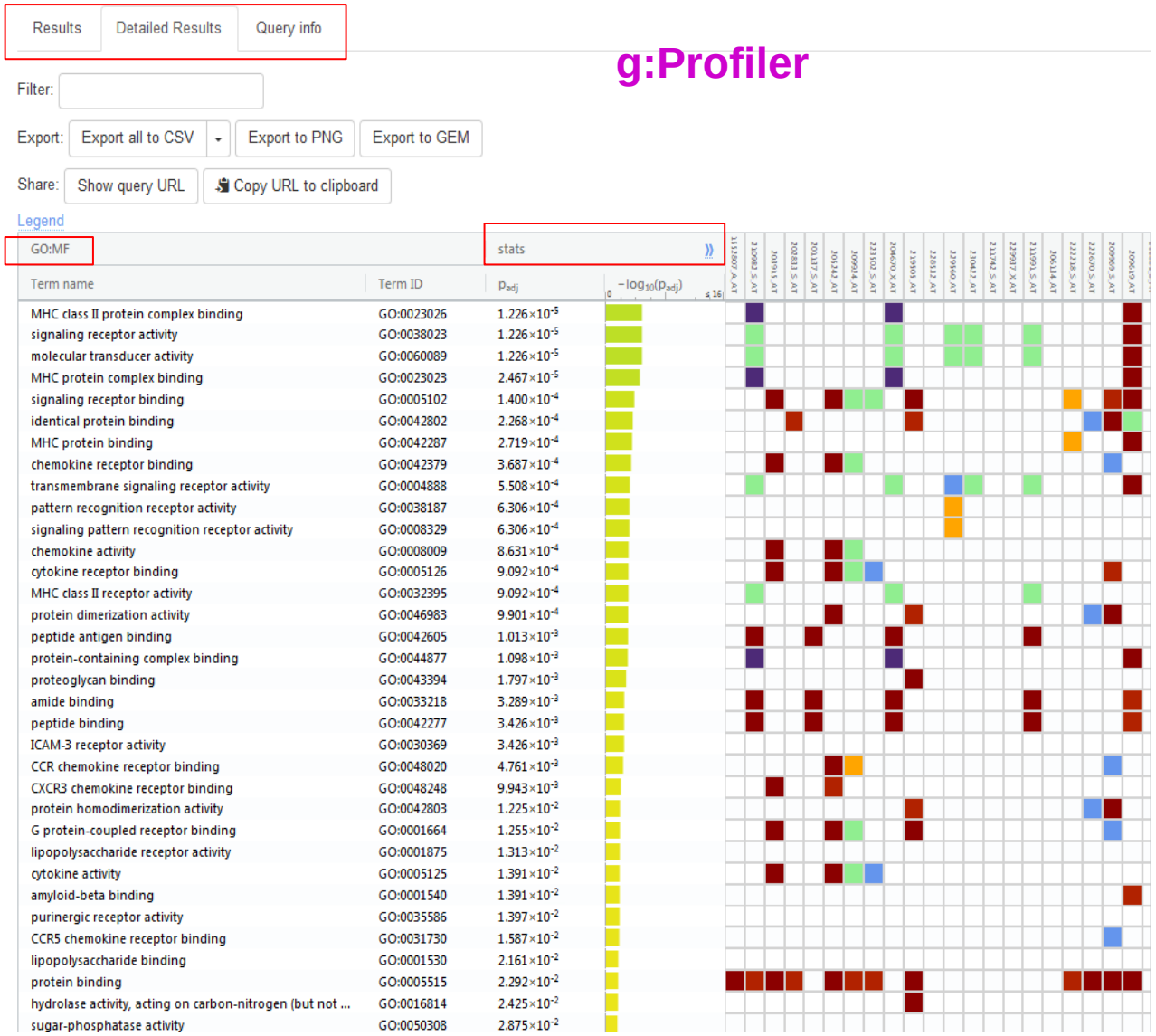
4. Enrichment analysis of selected genes

g:Profiler

- When mapping ambiguous symbols, pay attention to
  - Pseudogenes, fusion genes, read-throughs
  - Apparently unrelated symbols, from e.g. matching aliases
  - Families of very similar proteins (histones)
  - Number of annotated GO terms (Ensembl alignment errors, etc)



4. Enrichment analysis of selected genes



4. Enrichment analysis of selected genes

GOrilla

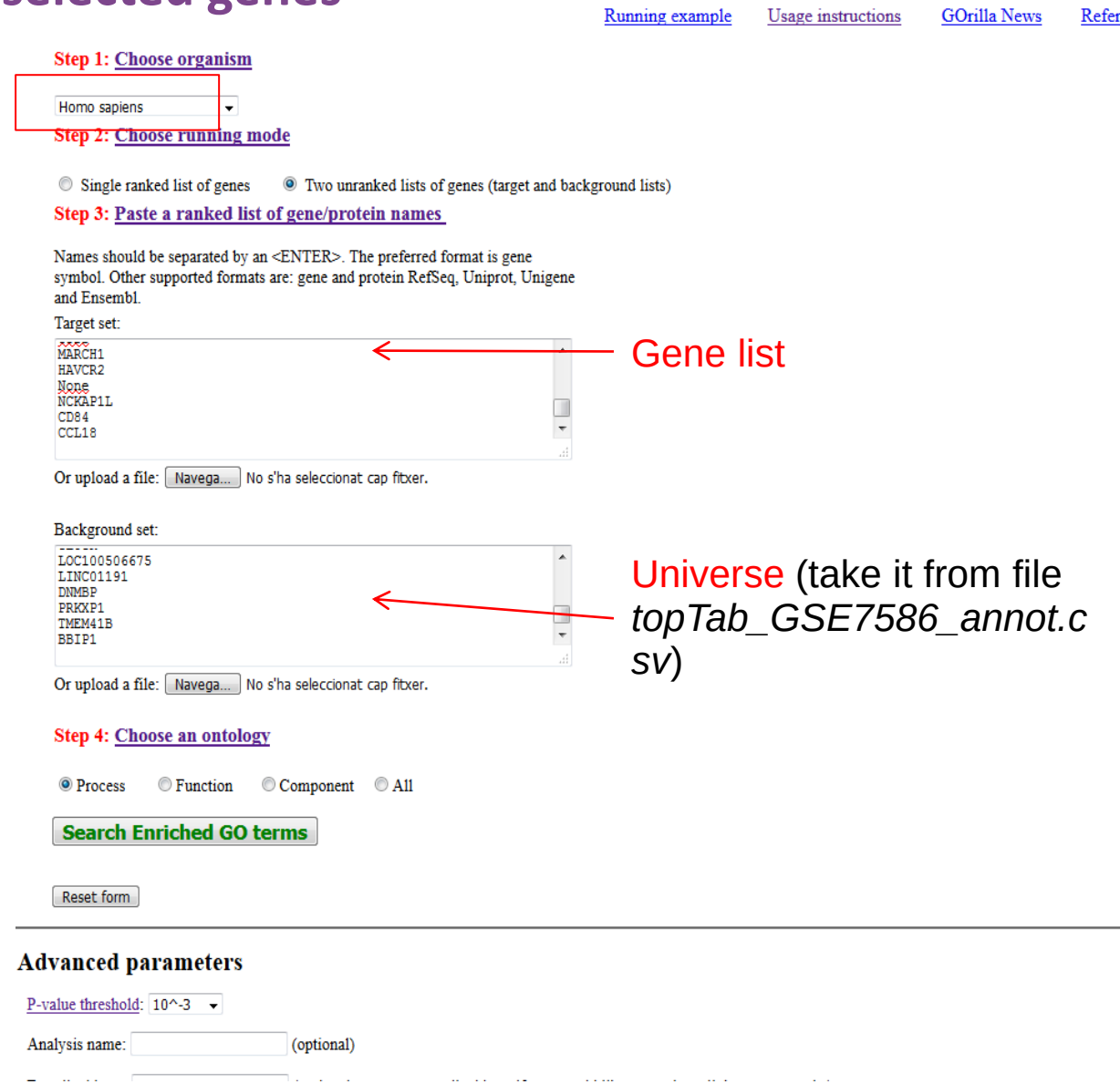
http://cbl-gorilla.cs.technion.ac.il/

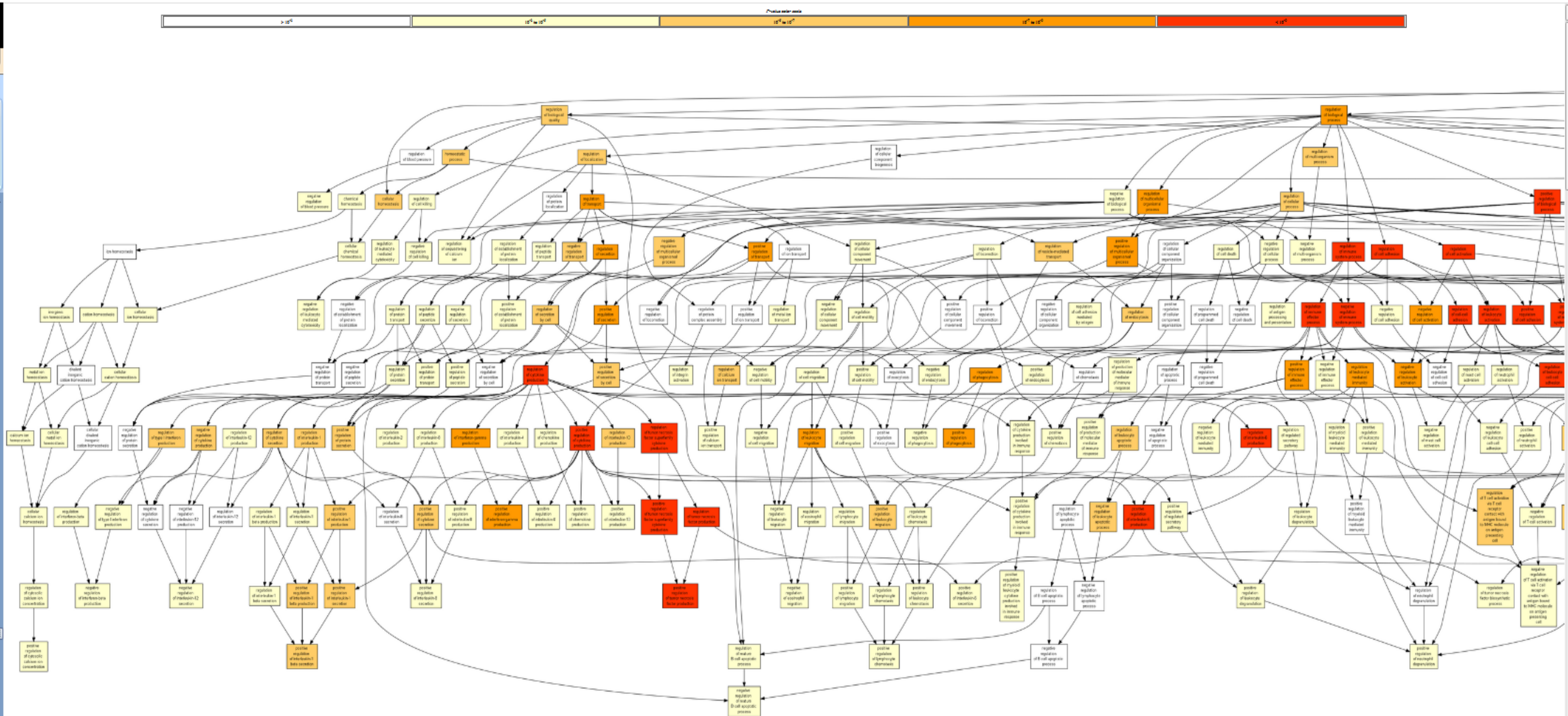
- ✓ Direct visualization of enriched terms in hierarchy
- ✓ Needs to specify the universe
- ✗ Only for GO
- ✗ Gene/Protein names as input

4. Enrichment analysis of selected genes

GOrilla

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





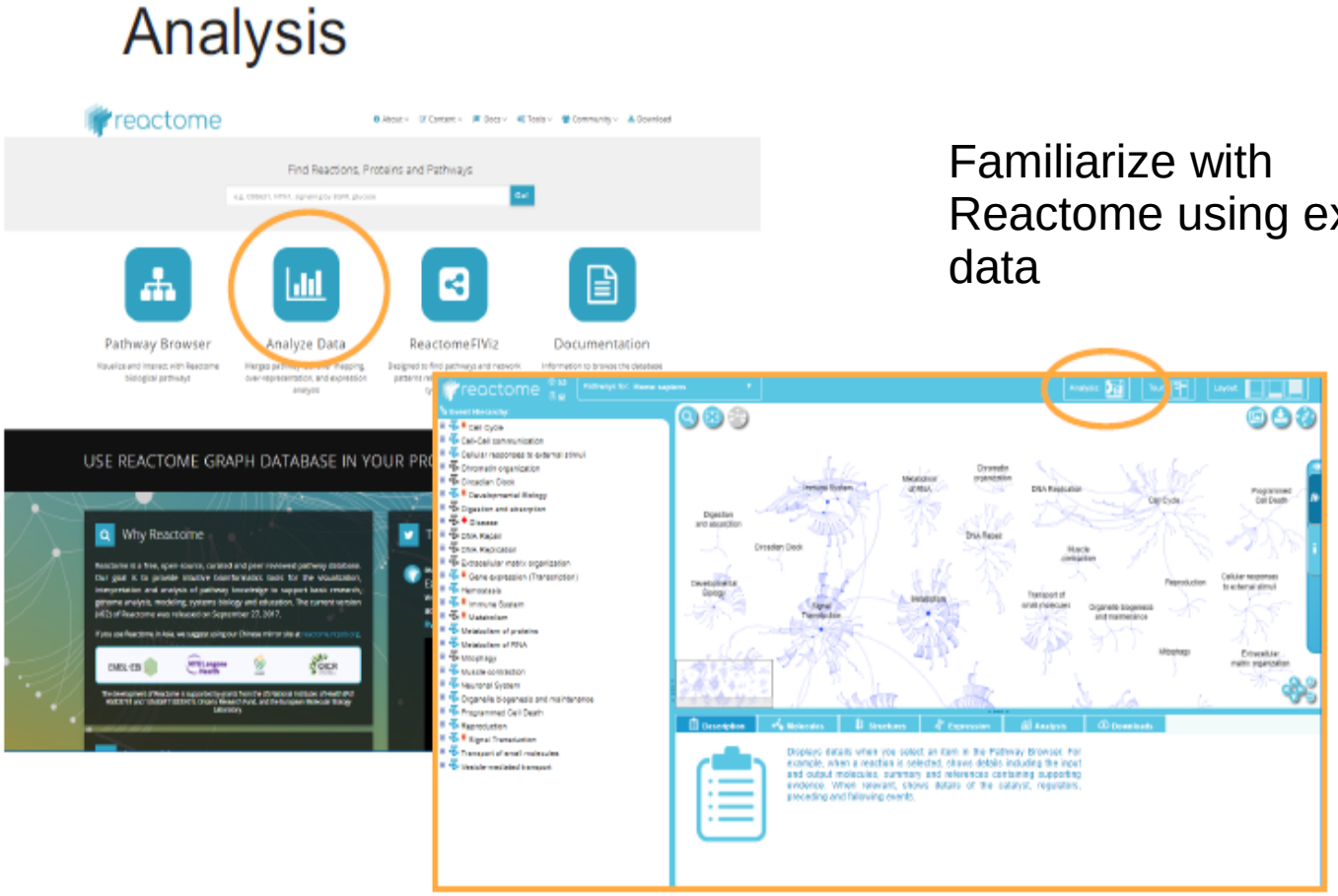
GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
[illegible]	immune system process	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	regulation of immune response	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	regulation of immune system process	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	positive regulation of immune system process	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	immune response	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	leukocyte activation	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	immune effector process	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	positive regulation of immune response	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	cell activation	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	defense response	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	myeloid leukocyte activation	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	leukocyte activation involved in immune response	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	cell activation involved in immune response	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	regulation of response to stimulus	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	granulocyte activation	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	myeloid cell activation involved in immune response	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	neutrophil activation	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	leukocyte degranulation	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	neutrophil activation involved in immune response	[illegible]	[illegible]	[illegible]	[+] Show genes
[illegible]	activation of immune response	[illegible]	[illegible]	[illegible]	[+] Show genes

6. Reactome Pathway Enrichment analysis of selected genes

Try with your dataset (list of genes): *selected_logFC_reactome.csv*

- What pathways are enriched below FDR < 0.05?
- Which has the major number of matching entities? What is its hierarchy?
- Explore “Diseases” section: are there similar pathways found in other infection
- …?

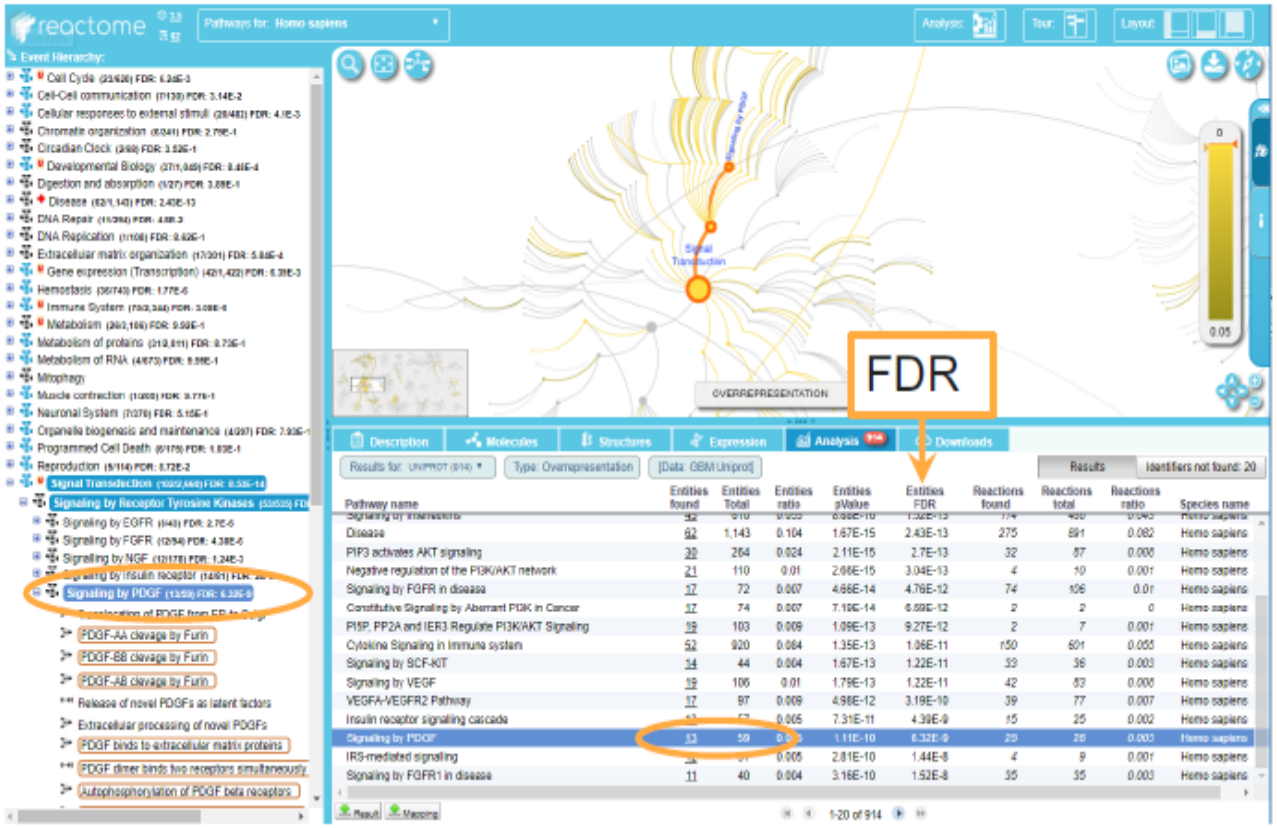
5. Introduction to Reactome and Pathway visualization



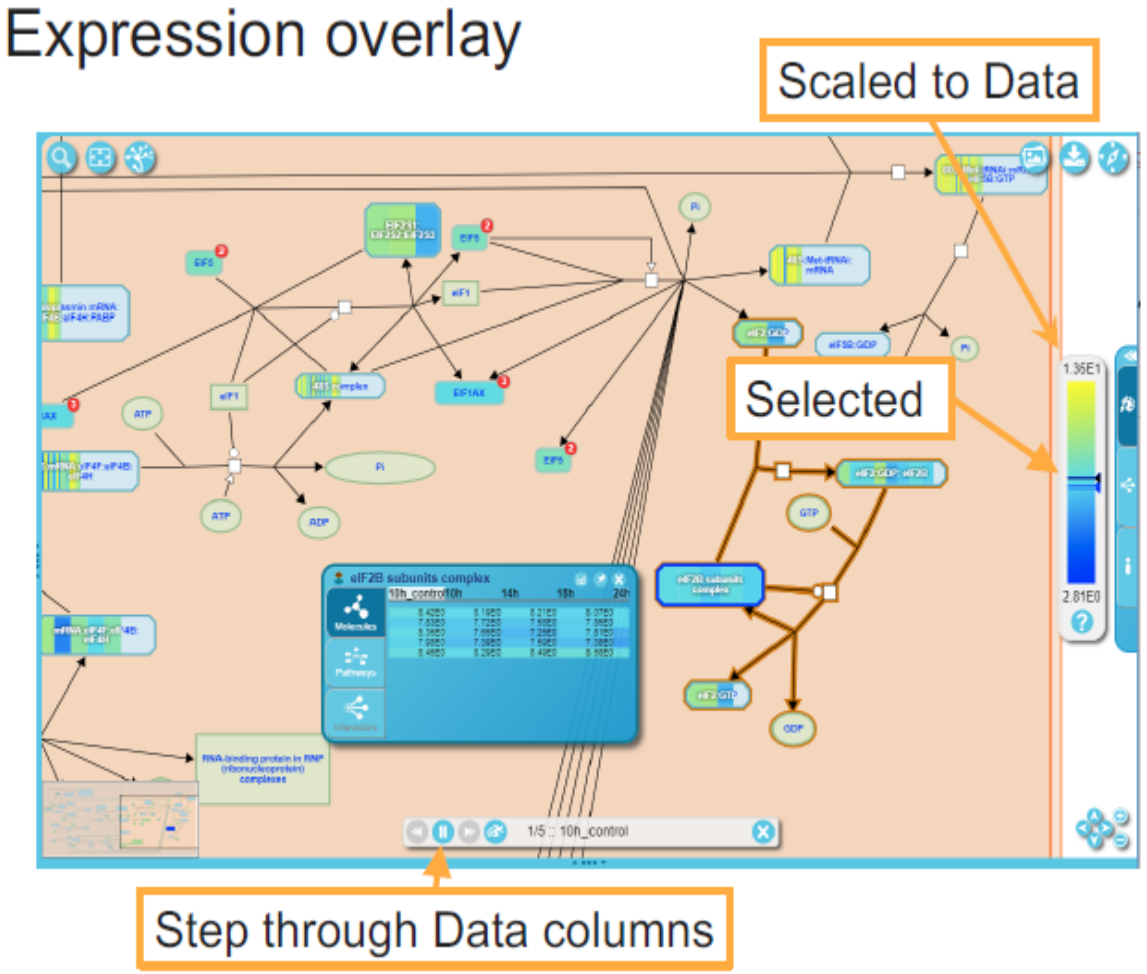
Familiarize with Reactome using example data

5. Introduction to Reactome and Pathway visualization

Analysis Result – Over-representation



5. Introduction to Reactome and Pathway visualization



Visualization and downstream analysis

REVIGO
EnrichmentMaps
Pathview

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 of pathways of interest

EnrichmentMaps

A workflow in R

clusterProfileR

Workflow gsea with clusterprofiler:

https://learn.gencore.bio.nyu.edu/rna-seq-analysis/gene-set-enrichment-analysis/

https://hbctraining.github.io/DGE_workshop_salmon/lessons/functional_analysis_2019.html

https://hbctraining.github.io/DGE_workshop/lessons/functional_analysis_other_methods.html

http://biit.cs.ut.ee/gprofiler/page/r