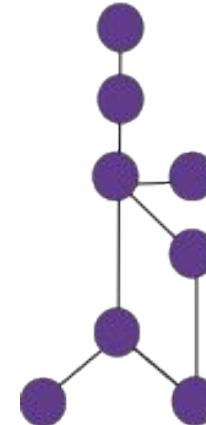
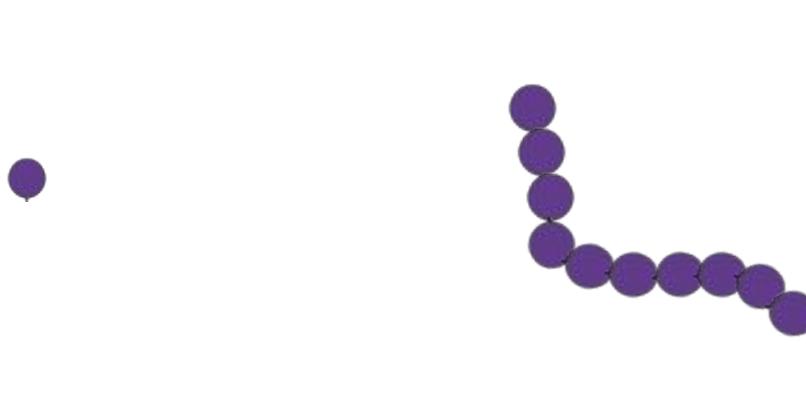


Functional Analysis

The biology behind the genes

What is Functional analysis?

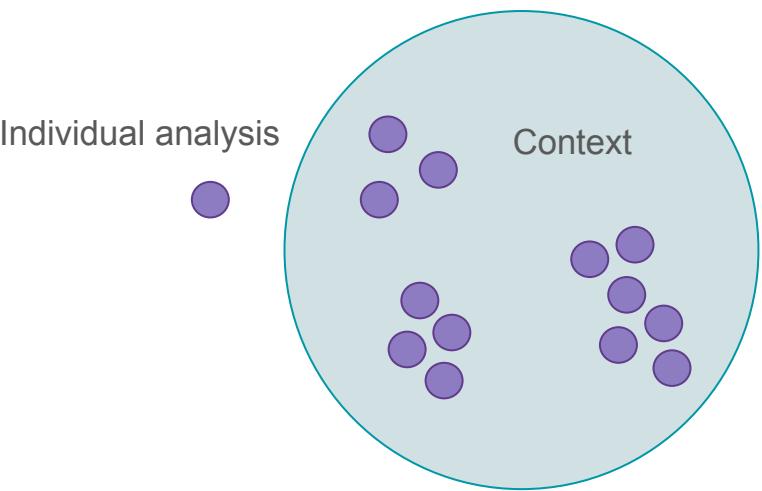


Biology behind genes

Looking into genes function:

Individually

- Focus in one main gene
- Out of context



Biology behind genes

Looking into genes function:

Individually

- Focus in one main gene
- Out of context

Lists of genes

Look



SEMANTICS

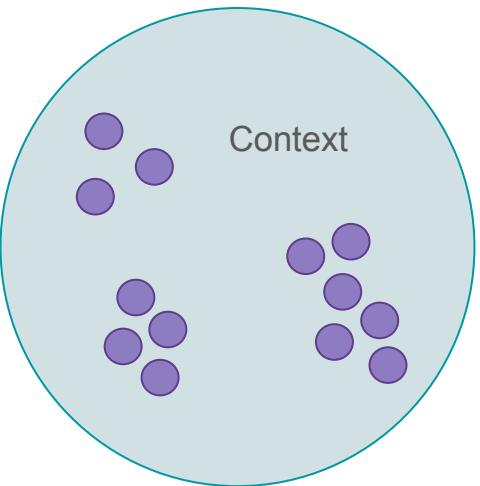
Look into



Look forward



Individual analysis



Biology behind genes

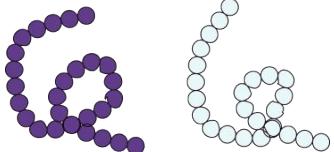
Looking into genes function:

Individually

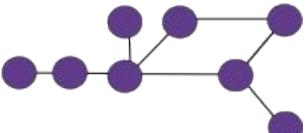
- Focus in one main gene
- Out of context

Lists of genes

- Prioritize context
- Understand biological relationship



Common functions/structure



Network related



Common transcription factors



Similar motifs



Biology behind genes

Looking into genes function:

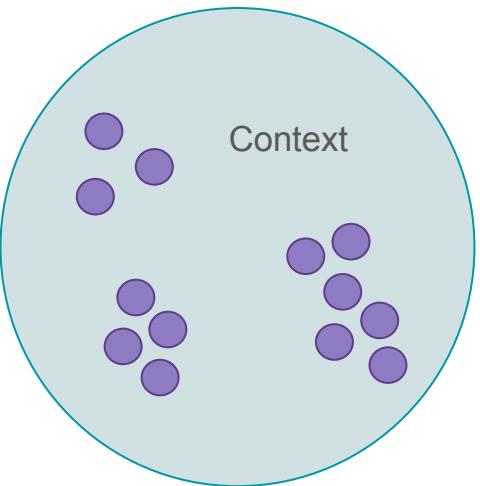
Individually

- Focus in one main gene
- Out of context

Lists of genes

- Prioritize context
- Understand biological relationship

Define what is their relation
within the context = condition



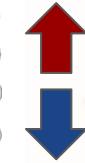
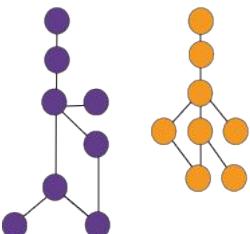
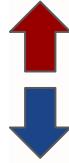
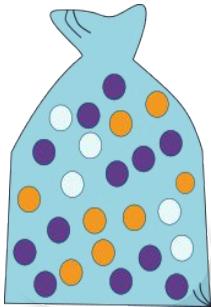
Biology behind genes

Different methods:

- Enrichment analysis
 - Gene Ontology (GO)
 - Gene-set (GSEA)
- Topology
 - Pathway
- Gene Expression
 - Differential Expression

Overrepresentation analysis

- Annotation
- Gene
 - Protein





Quality and context

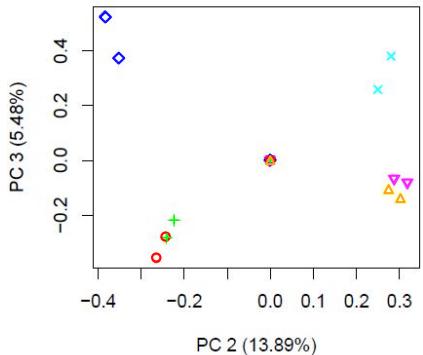
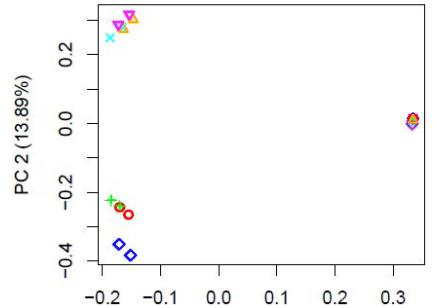
Data introduction



Quality control

- The assembling ...
 - PCA: Correlation between samples
 - Metadata: Information about samples
 - etc...

Hints of the context

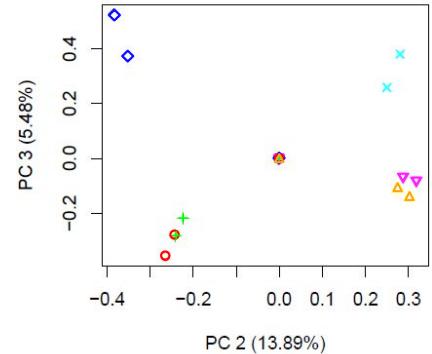
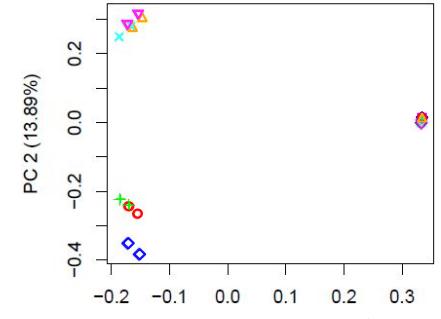


Quality control

Hints of the context

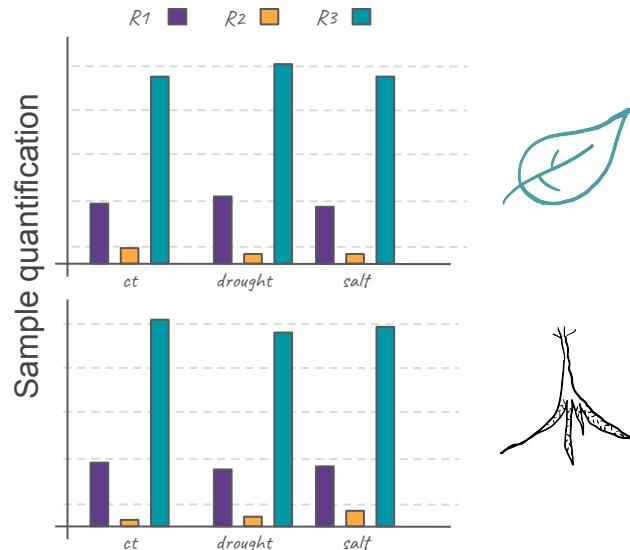
- The assembling ...
 - PCA: Correlation between samples
 - Metadata: Information about samples
 - etc...

Tissue	Treatment	Organism	AGE days	Cultivar	dev_stage	treatment
Leaf	Control	Prosopis juliflora	15	DLEG910003	Leaf	No stress
Root	Control	Prosopis juliflora	15	DLEG910003	Root	No stress
Leaf	Drought	Prosopis juliflora	15	DLEG910003	Leaf	15% PEG
Root	Drought	Prosopis juliflora	15	DLEG910003	Root	15% PEG
Leaf	Salt	Prosopis juliflora	15	DLEG910003	Leaf	150mM NaCl
Root	Salt	Prosopis juliflora	15	DLEG910003	Root	150mM NaCl

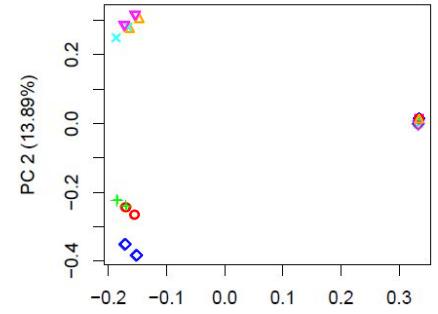


Quality control

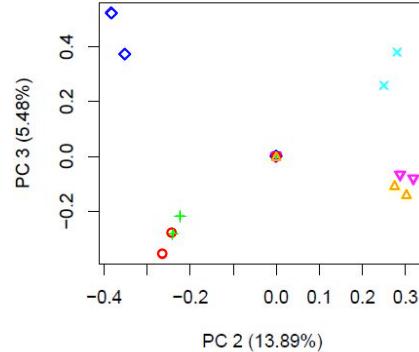
- The assembling ...
 - PCA: Correlation between samples
 - Metadata: Information about samples
 - etc...



Hints of the context



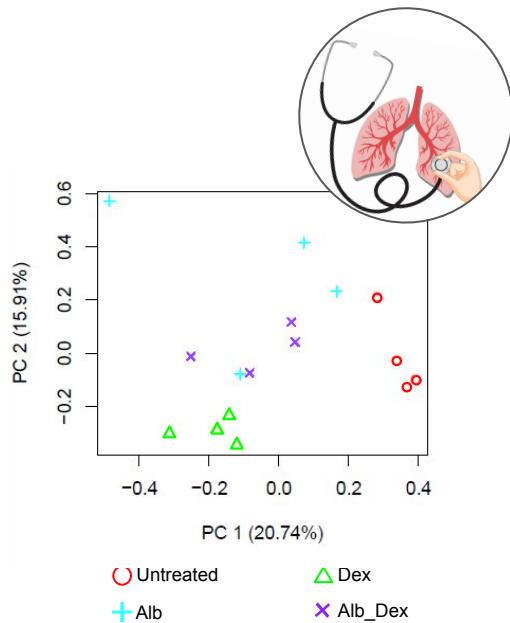
Leaf_CT Root_CT
Leaf_Drouth Root_Drouth
Leaf_Salt Root_Salt



Quality control → another example

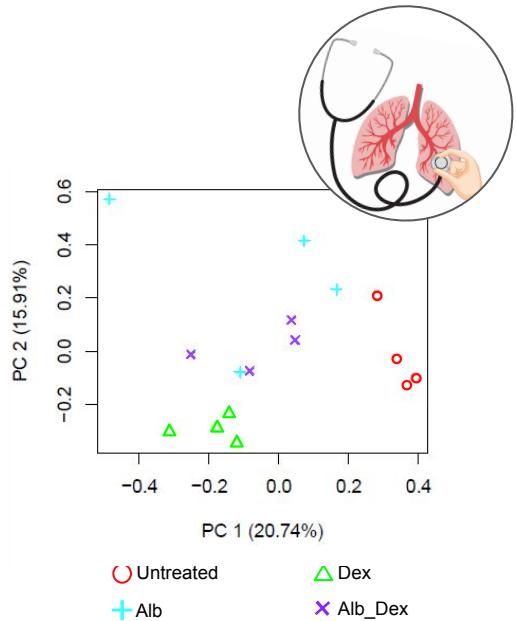
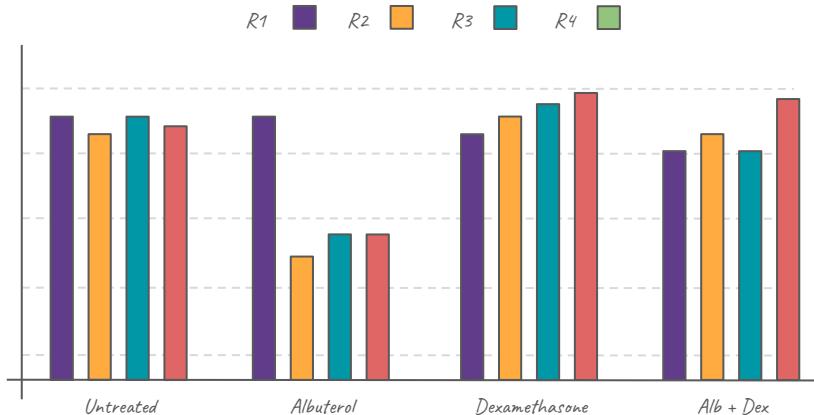
- The assembling ...
 - PCA: Correlation between samples
 - Metadata: Information about samples
 - etc...

Tissue	Treatment	Organism	Acession n	Condition	treatment
Lungs	Untreated	Homo sapiens	E-GEOD-52778	Asthma	No stress
Lungs	Albuterol	Homo sapiens	E-GEOD-52778	Asthma	1µM for 18h
Lungs	Dexamethasone	Homo sapiens	E-GEOD-52778	Asthma	1µM for 18h
Lungs	Alb + Dex	Homo sapiens	E-GEOD-52778	Asthma	1µM for 18h

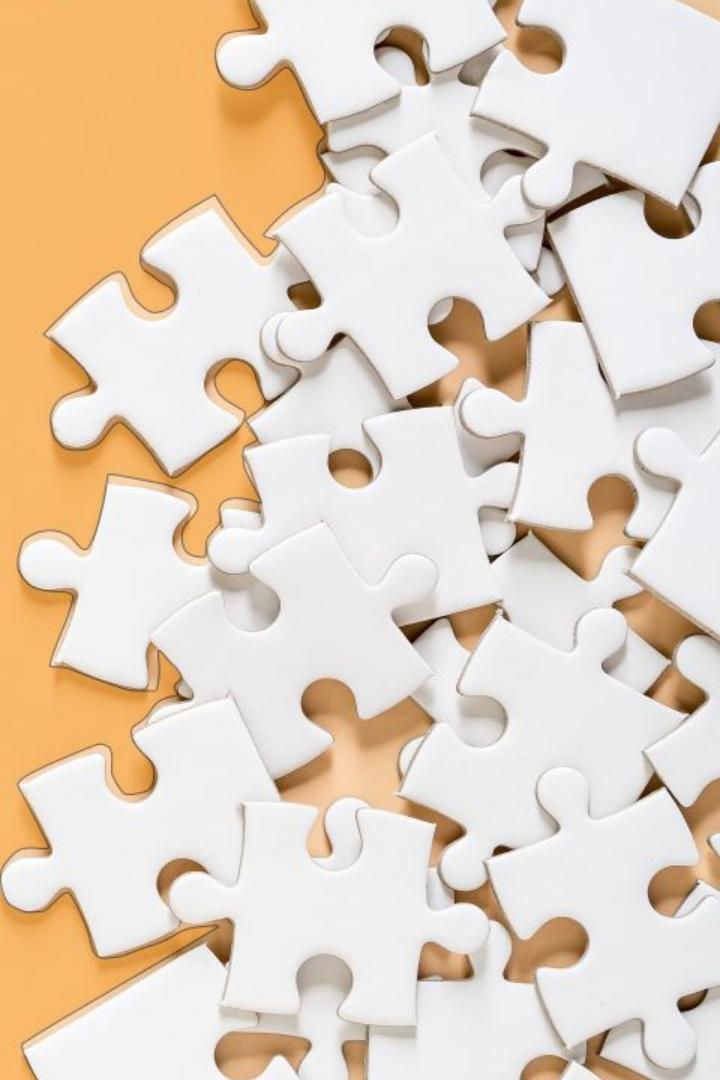


Quality control → another example

- The assembling ...
 - PCA: Correlation between samples
 - Metadata: Information about samples
 - etc...



What if it there is
no Ref. Genome?



Functional Annotation

Translating IDs according to the tool:

- Types of IDs

How it really starts:

- 1st level of annotation



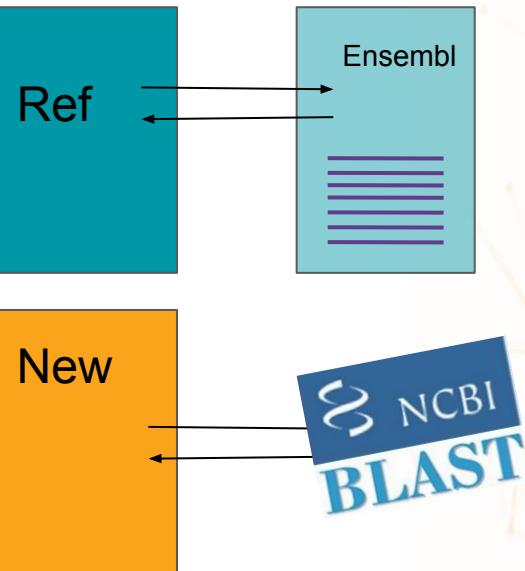
Functional Annotation

Translating IDs according to the tool:

- Types of IDs

How it really starts:

- 1st level of annotation
(for those who don't have a reference genome)



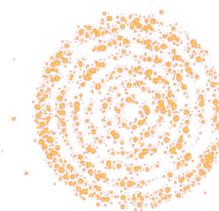


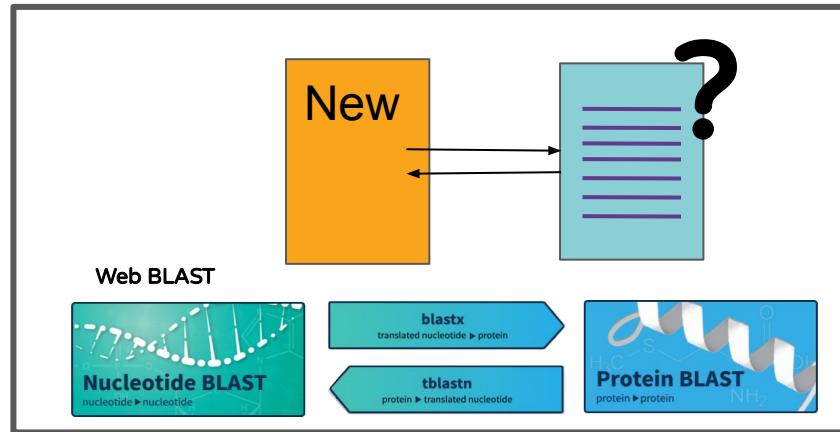
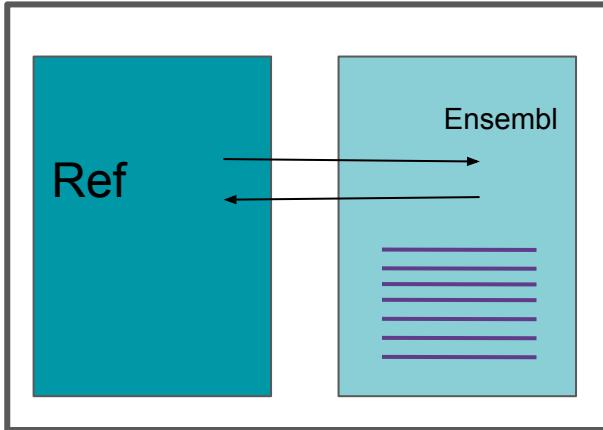
Activity: No ref Genome Organism

Get your fasta.

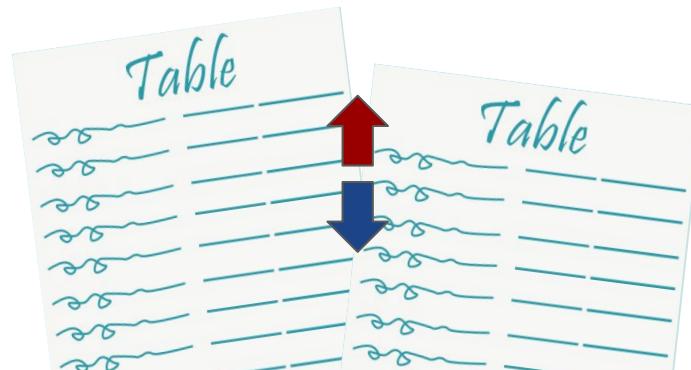
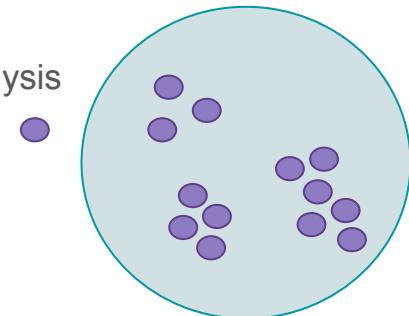
- You need to annotate (BLAST)
- Once you have your result filter
 - for 90-100% identity
 - check e-value (0.0 or close to)
- Translate to Uniprot ID and name

Web BLAST





Individual analysis



Functional Annotation

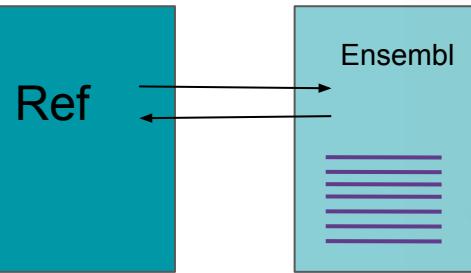
The basics - It starts



Functional Annotation

Translating IDs according to the tool:

- Types of IDs



Functional Annotation

Translating IDs according to the tool:

- Types of IDs



GenBank
RefSeq



How to get what you need:

- Uniprot ID mapping
- BioMart
- Programatically
 - Using R ([e-learning](https://www.uniprot.org/help/api_queries) or R course)
 - R and Python
 - https://www.uniprot.org/help/api_queries
 - https://www.uniprot.org/help/id_mapping



Functional Annotation

The screenshot shows the UniProt Functional Annotation interface. At the top, there is a navigation bar with links for BLAST, Align, Peptide search, ID mapping, SPARQL, UniProtKB (selected), Advanced, List, and Search.

Retrieve/ID mapping

Enter one or more IDs (100,000 max). You may also [load from a text file](#). Separate IDs by whitespace (space, tab, newline) or commas.

```
P31946 P62258 ALBU_HUMAN EFTU_ECOLI
```

From database: UniProtKB AC/ID ▾ To database: UniProtKB ▾

Name your ID Mapping job: "my job title"

Reset Map IDs

Functional Annotation

e!Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog

New Count Results

Dataset
Human genes (GRCh38.p14)

Filters
Gene stable ID(s) [e.g. ENSG00000000003]: [ID-list specified]

Attributes
Gene stable ID
Gene

New Count Results

URL XML Perl Help

Please restrict your query using criteria below
(If filter values are truncated in any lists, hover over the list item to see the full text)

REGION:

GENE:

Limit to genes (external references)...
 Input external references ID list [Max 500 advised]
 Limit to genes (microarray probes/probesets)...
 Input microarray probes/probesets ID list [Max 500 advised]

With BioGRID Interaction data, The General Repository for Interaction Datasets ... Only Excluded

Gene stable ID(s) [e.g. ENSG00000000003] No file chosen

ENS G00000003402
ENS G00000004799
ENS G00000006788
ENS G00000008256

With AFFY HC G110 probe ID(s) Only Excluded

AFFY HC G110 probe ID(s) [e.g. 737_at] No file chosen

Dataset
[None Selected]

Functional Annotation

e!Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog

New Count Results

Dataset
Human genes (GRCh38.p14)

Filters
Gene stable ID(s) [e.g. ENSG00000000003]: [ID-list specified]

Attribute
Gene stat
Gene stat
Transcript
Transcript
Protein st
Protein st
Gene des
GO term a
GO term i
UniProtKB
UniProtKB

Dataset
[None Selected]

Dataset
[None Selected]

Ensembl Genes 110

Human genes (GRCh38.p14)

New Count Results

URL XML Perl Help

Please select columns to be included in the output and hit 'Results' when ready

Missing non coding genes in your mart query output, please check the following [FAQ](#)

Features Variant (Germline)
 Structures Variant (Somatic)
 Homologues (Max select 6 orthologues) Sequences

GENE:

EXTERNAL:

GO
 GO term accession
 GO term name
 GO term definition

GO Slim GOA
 GO Slim GOA Accession(s)

External References (max 3)
 BioGRID Interaction data, The General Repository for Interaction Datasets ID
 CCDS ID
 ChEMBL ID
 DataBase of Aberrant 3' Splice Sites name
 DataBase of Aberrant 3' Splice Sites ID

NCBI gene (formerly Entrezgene) accession
 NCBI gene (formerly Entrezgene) ID
 PDB ID
 Reactome ID
 Reactome gene ID

Activity: Mapping IDs



Let's open the examples of data for
Plants and Humans.



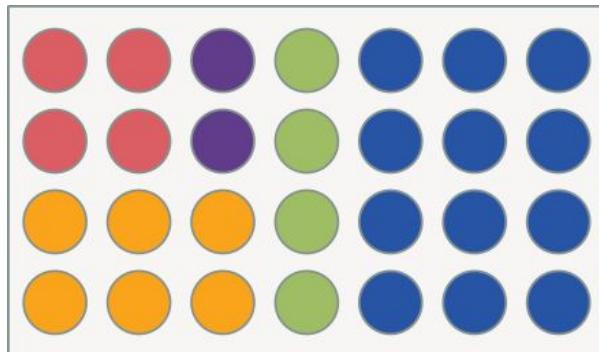
We want to get the UniProt ID and Ensemble ID of both datasets complementary to the list you are going to find. In one example you have Uniprot names and the other Ensemble IDs.

For that we will use the [id-mapping retriever](#) resource available in Uniprot, and [Biomart human id retriever](#) and [Biomart Plant id retriever](#) available in Ensemble.

Enrichment Analysis

Enrichment analysis

- Pull of genes
- Over-represented

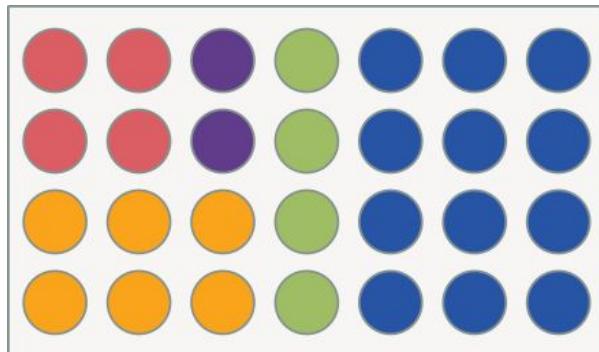


List of genes
Colors = functions

Image by janick

Enrichment analysis

- Pull of genes
- Over-represented
- Expression: Up or Down

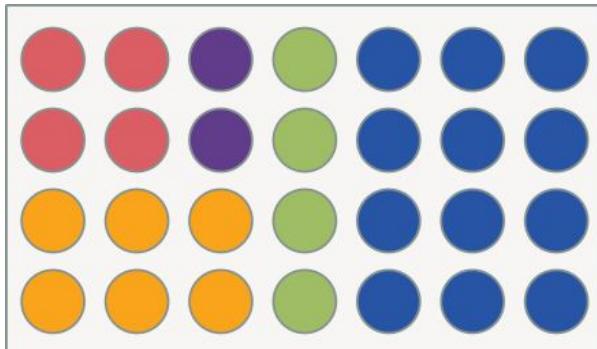


List of genes
Colors = functions

Image by janick

Enrichment analysis

- Pull of genes
- Over-represented
- Expression: Up or Down



List of genes
Colors = functions

Image by janick

Two types of enrichment:

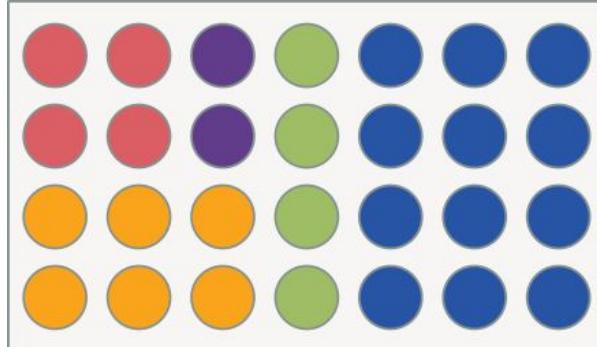
- Gene ontology enrichment
 - Define the related terms
- Gene-set enrichment
 - Select groups of genes

Enrichment analysis

Different points of view:

- Gene ontology enrichment:

All up against the Total - What terms are enriched?

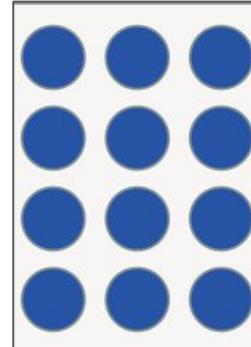


List of genes
Colors = functions

Image by janick

- Gene-set enrichment:

Ranked genes, which set is + represented among most differentially expressed?



Enrichment analysis

Different points of view:

- Gene ontology enrichment:

All up against the Total - What terms are enriched?



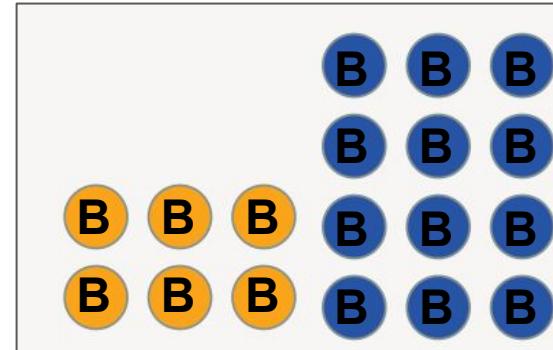
List of genes
Colors = functions

Image by janick

- Gene-ontology enrichment:

Which terms are more represented.
Consider the relation among terms

- Biological process
- Molecular function
- Cellular component



Enrichment analysis and Context

- Context and scientific answers:
 - Specific or Nonspecific
 - Whole plant *vs* leaves
 - Leaves during the day *vs* leaves in the sunset
- Whole human body *vs* lung sample
- Health lungs *vs* Cancer lungs



Combination of Image by katemangostar on Freepik and Freepik image



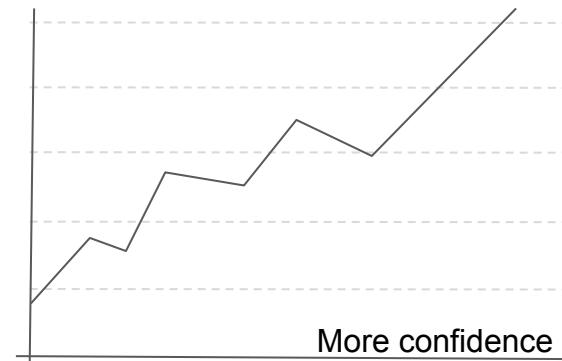
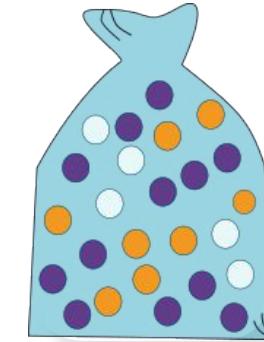
Image by brgfx on Freepik

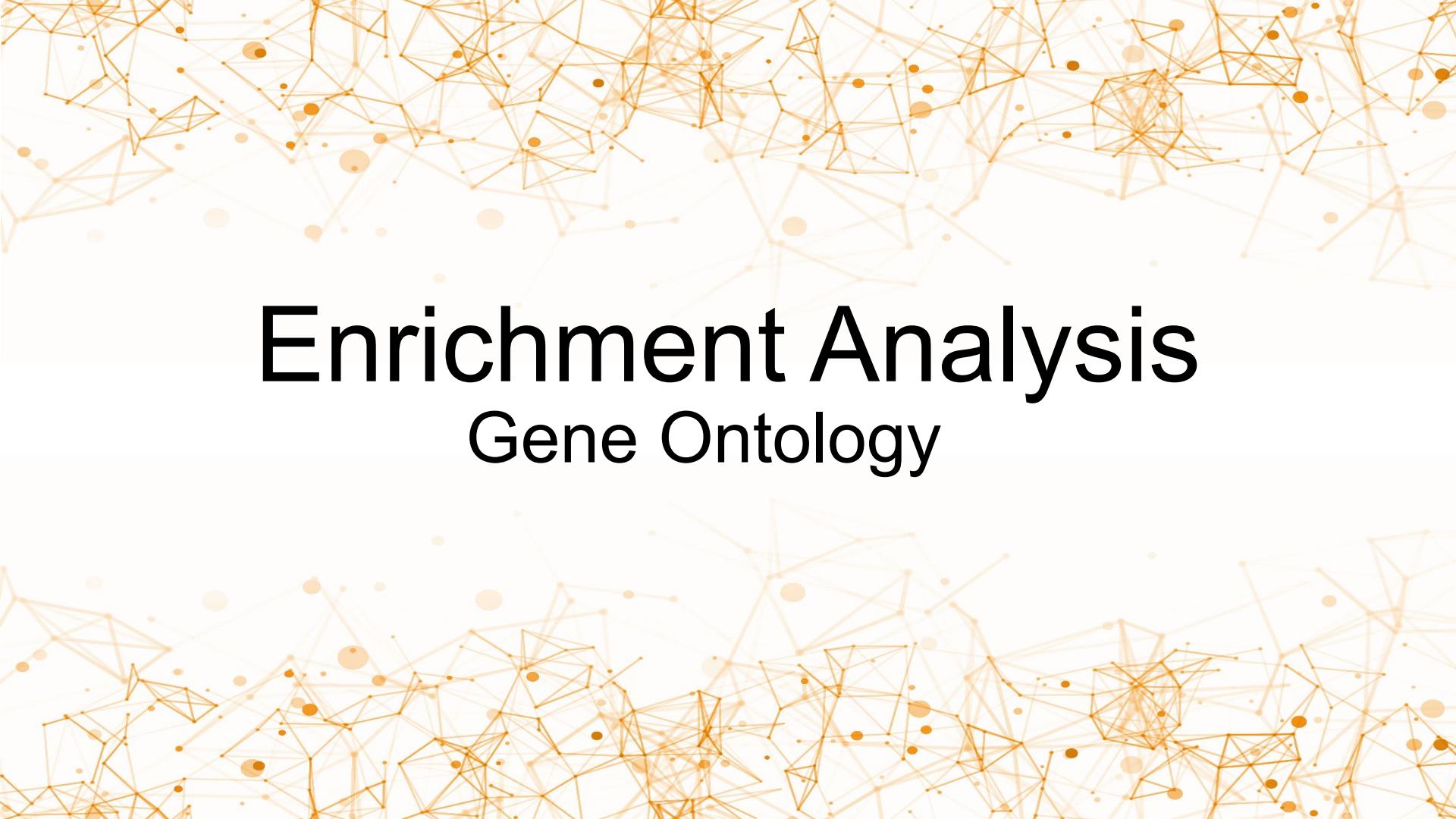
Quality control enrichment analysis

Corrections for false positive cases

You need to correct False positive rates

- Multiple testing
- Bonferroni:
 - More conservative
 - Less False Positive
 - Higher chance of False Negative
- FDR
 - Control proportion of False Positive
 - B&H
 - Find highest ranking



The background of the slide features a complex, abstract network structure composed of numerous small, semi-transparent orange dots connected by thin lines, creating a sense of depth and connectivity.

Enrichment Analysis

Gene Ontology

Enrichment analysis

GO enrichment

Ontology:

A set of **concepts** related to a domain, also used to define **relationships** among entities.

Enrichment analysis

GO enrichment

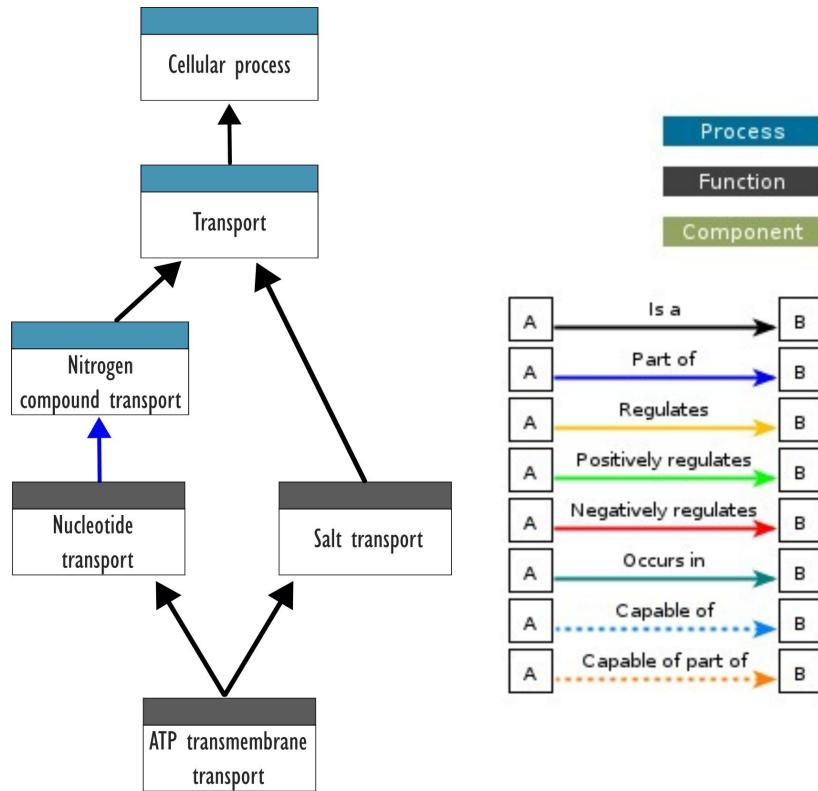
Ontology:

A set of concepts related to a domain, also used to define relationships among entities.

Gene Ontology:

Consortium created to provide the “*state of art*” about genes. The **gold** standard.

- Molecular function: Activities
- Biological process: Large process
- Cellular component: Location



adapted from QuickGO - <https://www.ebi.ac.uk/QuickGO>

Enrichment analysis

GO enrichment

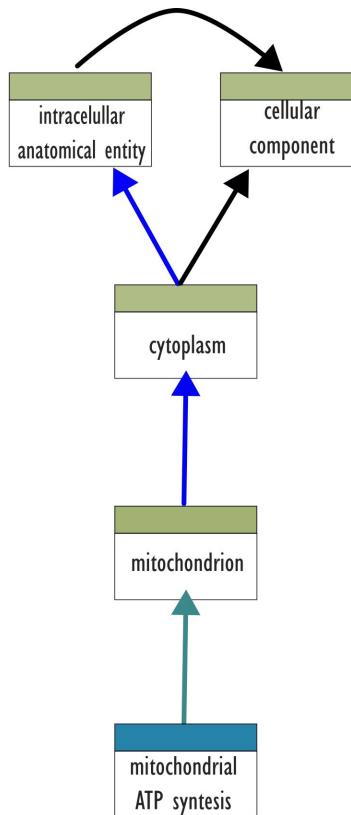
Ontology:

A set of concepts related to a domain, also used to define relationships among entities.

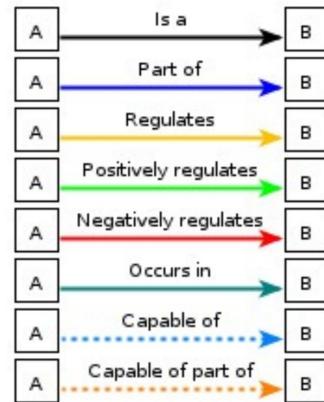
Gene Ontology:

Consortium created to provide the “*state of art*” about genes. The **gold** standard.

- Molecular function: Activities
- Biological process: Large process
- Cellular component: Location



Process
Function
Component



adapted from QuickGO - <https://www.ebi.ac.uk/QuickGO>

Enrichment analysis

GO enrichment

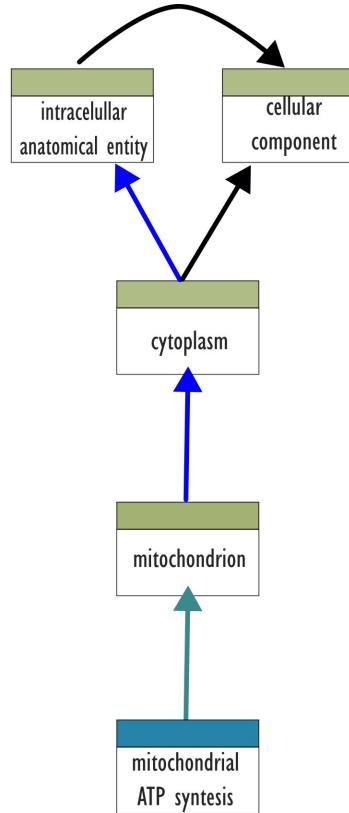
Ontology:

A set of **concepts** related to a domain, also used to define **relationships** among entities.

Gene Ontology:

Consortium created to provide the “*state of art*” about genes. The **gold standard**.

- Molecular function: Activities
- Biological process: Large process
- Cellular component: Location

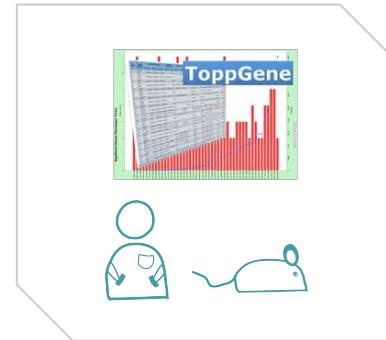


adapted from QuickGO - <https://www.ebi.ac.uk/QuickGO>

Taxon IDs:

Implicit, mostly Neutral

GO enrichment



GO enrichment



GO enrichment



About Ontology Annotations Downloads Help

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

Current release 2023-11-15: 42,769 GO terms | 7,648,957 annotations
1,534,763 gene products | 5,341 species (see statistics)

GO Enrichment Analysis

Powered by PANTHER

Your gene IDs here...

Q5VRD1
A0AOPOUZP7
Q9SDD6
A2ZRY8
Q5JNT2

biological process

Oryza sativa

Korarchaeum cryptophilum

Leishmania major

Leptospira interrogans

Listeria monocytogenes

Macaca mulatta

Medicago truncatula

Methanocaldococcus jannaschii

Methanosaerica acetylavorans

Monodelphis domestica

Monosiga brevicollis

Musa acuminate subsp. malaccensis

Mycobacterium tuberculosis

Neisseria meningitidis serogroup b

Nematostella vectensis

Neosartorya fumigata

Neurospora crassa

Nicotiana tabacum

Nitrosopumilus maritimus

Ornithodoros anatinus

Oryza sativa

[Home page](#)

The network of biological classes describing the current best representation of the "universe" of biology: the molecular functions, cellular locations, and processes gene

ANNOTATION

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

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

visualize and edit annotations. Notebooks) and the GO into

GO enrichment



[Home page](#)



About Ontology Annotations Downloads Help

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

- Biological process
- Cellular function
- Molecular component

Current release 2023-11-15: 42,769 GO terms | 7,648,957 annotations
1,534,763 gene products | 5,341 species ([see statistics](#))

GO Enrichment Analysis

Powered by PANTHER

Your gene IDs here...

biological process

Oryza sativa

Korarchaeum cryptofilum
Leishmania major
Leptospira interrogans
Listeria monocytogenes
Macaca mulatta
Medicago truncatula
Methanocaldococcus jannaschii
Methanosaerica acetylavorans
Monodelphis domestica
Monosiga brevicollis
Musa acuminata subsp. malaccensis
Mycobacterium tuberculosis
Neisseria meningitidis serogroup b
Nematostella vectensis
Neosartorya fumigata
Neurospora crassa
Nicotiana tabacum
Nitrosopumilus maritimus
Ornithorhynchus anatinus
Oryza sativa

Examples Launch

MOD IDs

Visualize and
Annotations.
(notebooks) and
the GO into

GO enrichment

[Documentation](#)Results [?](#)

Reference list upload_1

43659 out of 43659 93 out of 93

Unmapped IDs: 0 7

Multiple mapping information: 0 0

Export [Table](#) [XML with user input ids](#) [JSON with user input ids](#)Displaying only results for FDR P < 0.05, [click here to display all results](#)

	Oryza sativa (REF)	upload_1 (▼ Hierarchy NEW! ?)						
	#	# expected	Fold Enrichment	+/-	raw P value	FDR		
GO biological process complete	43659	5	.02	> 100	+	4.94E-11	5.28E-09	
adventitious root development	0	5	.22	22.35	+	3.99E-06	2.42E-04	
↳ root development	105	105						
↳ root system development	105	513						
↳ system development	513	824						
↳ anatomical structure development	824	910						
↳ developmental process	910	664						
↳ multicellular organism development	664	720						
↳ multicellular organismal process	720	246						
↳ plant organ development	246	5						
protein glutathionylation	5	8932						
↳ organic substance metabolic process	8932	9567						
↳ metabolic process	9567	13						
auxin export across the plasma membrane	13	61						
↳ export across plasma membrane	61	1340						
↳ transmembrane transport	1340	10813						
↳ cellular process	10813	139						
↳ export from cell	139	24						
↳ regulation of auxin mediated signaling pathway	24	4	.05	78.24	+	3.77E-07	3.13E-05	
↳ regulation of signal transduction	4	147	6	.31	19.16	+	9.83E-07	7.61E-05

Analysis Summary: Please report in publication [?](#)

Analysis Type: PANTHER Overrepresentation Test (Released 20231017)

Annotation Version and Release Date: GO Ontology database DOI: 10.5281/zenodo.7942786 Released 2023-01-05

Analyzed List: upload_1 (Homo sapiens) [Change](#)Reference List: Homo sapiens (all genes in database) [Change](#)Annotation Data Set: [GO biological process complete](#) [?](#)Test Type: Fisher's Exact BinomialCorrection: Calculate False Discovery Rate Use the Bonferroni correction for multiple testing [?](#) No correction

GO enrichment

[Documentation](#)

SELECT REFERENCE LIST

For a reference list, you may upload your own list (recommended) or choose from available whole genome lists.

Upload Reference List from flat file or Workspace

Select Organism: (Not applicable for Generic mapping file or Reference Proteome ids)

Homo sapiens
Mus musculus
Rattus norvegicus
Gallus gallus
Danio rerio

Upload list:

Please select list type...

Gene, Transcript, Protein and Alternate ID

PANTHER Generic Mapping File

ID's from Reference Proteome Genome

Organism for id list

VCF file Flanking region 20 Kb

Upload list: No file chosen

[supported IDs](#)

[Upload list](#)

If there are redundant IDs, only the first will be used in the analysis.

Please [login](#) to be able to select lists from your workspace.

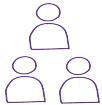
Use a Reference List that includes all genes from a whole genome

Default whole-genome lists:

- Ensembl: Ensembl gene identifier. Example: "ENSG00000126243"
- Ensembl_PRO: Ensembl protein identifier. Example: "ENSP00000337383"
- Ensembl_TRS: Ensembl transcript identifier. Example: ENST00000391828"
- Gene ID: EntrezGene IDs. examples include, "GenelD:10203", "10203" (for Entrez gene GenelD:10203)
- Gene symbol: for example, "CALCA"
- GI: NCBI GI numbers. Example: "16033597"
- HGNC: HUGO Gene Nomenclature ids. Example: "HGNC:16673"
- IPI: International Protein Index ids. Example: "IPI00740702"
- UniGene: NCBI UniGene ids. Examples: "Hs.654587", "At.36040"
- UniProtKB:UniProt accession. Example: "Q80536"
- UniProtKB-ID: UniProt ID. Example: "AGAP3_HUMAN"



Activity: Mapping IDs



Use the Example you've translated in the previous activity

Plants and Humans.



Now you want to have the GO terms annotated on Gene Ontology - Panther. Don't forget to user your own background.

HOW DO YOU DEFINE THE BACKGROUND?

- Context and scientific answers:
 - Specific or Nonspecific
 - Whole plant *vs* leaves
 - Leaves during the day *vs* leaves in the sunset



Combination of Image by katemangostar on Freepik and Freepik image

- Whole human body *vs* lung sample
- Health lungs *vs* Cancer lungs



Image by brgfx on Freepik

HOW DO YOU DEFINE THE BACKGROUND?

- The whole genome?
- The hole list of genes of your experiment?
- Only the Up regulated genes? Or down?
- What do we do with the non-differentially expressed?



Combination of Image by katemangostar on Freepik and Freepik image



Image by brgfx on Freepik

HOW DO YOU DEFINE THE BACKGROUND?

- We use the filtered pull of genes in our experiment.
 - Filtered by??
 - Non-differentially expressed very low difference
 - Avoid bias (false positive)

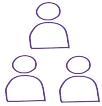


Combination of Image by katemangostar on Freepik and Freepik image



Image by brgfx on Freepik

Activity: Mapping IDs



Use the Example you've translated in the previous activity

Plants and Humans.



Now you want to have the GO terms annotated on Gene Ontology - Panther. Don't forget to user your own background.

GO enrichment



GO enrichment



GO enrichment



g:Profiler

g:Profiler

[News](#) [Archives](#) [Beta](#) [API](#) [R client](#) [FAQ](#) [Docs](#) [Contact](#) [Cite g:Profiler](#) [Services using g:P](#) [GMT Helper](#) [☰](#)**g:GOST**

Functional profiling

input can consist of mixed types
of gene IDs, SNP IDs,
chromosomal intervals or term
IDs[Query](#) [Upload query](#)

Input is whitespace-separated list of genes

Os01g0138800
Os01g0388101
Os01g0511600
Os01g0588900
Os01g0597800
Os01g0638000
Os01g0764800
Os02g0704300
Os02g0743400
Os01g0638000
Os01g0764800
Os02g0704300
Os02g0743400
Os01g0638000
Os01g0764800

[Run query](#)[random example](#)[mixed query example](#)**g:Convert**

ne ID conversion

g:Orth

Orthology search

g:SNPense

SNP id to gene name

Options

Organism:

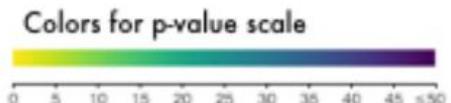
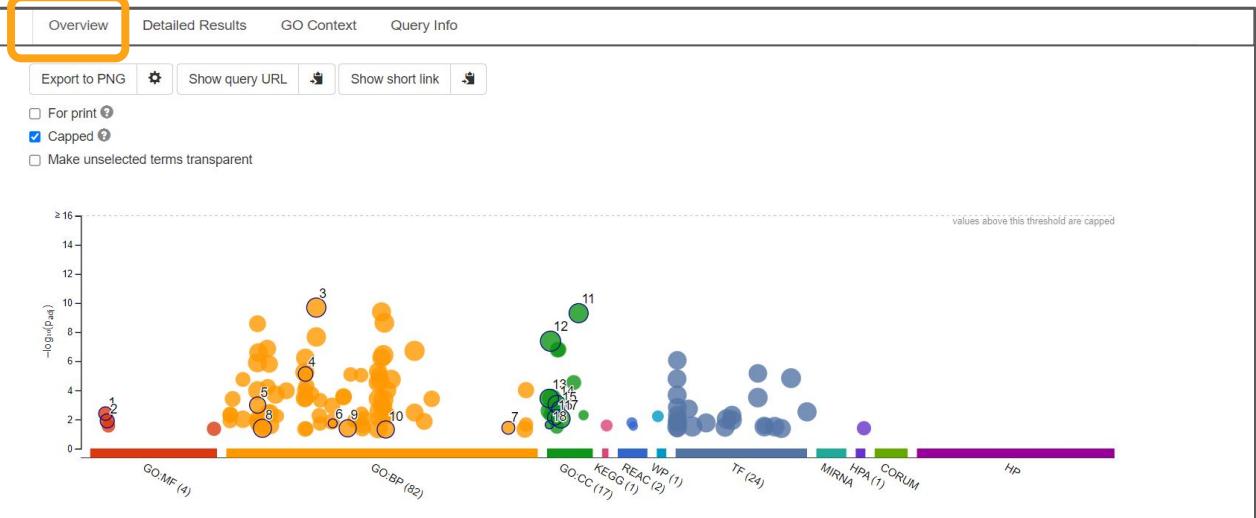
[Oryza sativa \(rice\)](#)

Organism:

[Homo sapiens \(Human\)](#)

- ▶ Favourites (14)
- ▶ Ensembl (205)
- ▶ Ensembl Genomes Fungi (67)
- ▶ Ensembl Genomes Metazoa (249)
- ▶ Ensembl Genomes Plants (147)
- ▶ Ensembl Genomes Protists (33)
- ▶ WormBase ParaSite (240)

GO enrichment



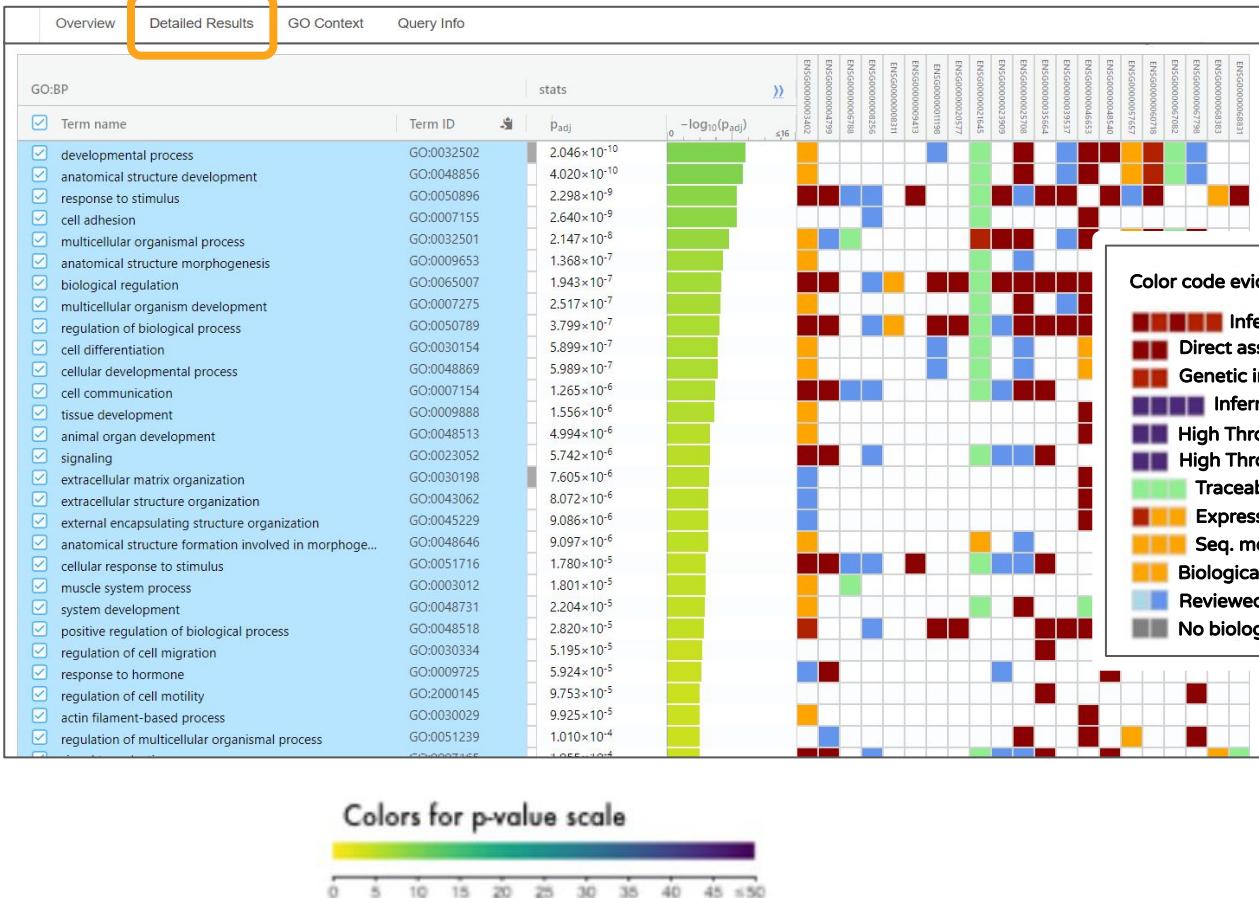
ID	Source	Term ID	Term Name	Padj (query_1)
1	GO:MF	GO:0005201	extracellular matrix structural constituent	3.966×10^{-3}
2	GO:MF	GO:0005539	glycosaminoglycan binding	1.264×10^{-3}
3	GO:BP	GO:0032502	developmental process	2.046×10^{-10}
4	GO:BP	GO:0030198	extracellular matrix organization	7.605×10^{-6}
5	GO:BP	GO:0007167	enzyme-linked receptor protein signaling pathway	1.035×10^{-3}
6	GO:BP	GO:0035360	positive regulation of peroxisome proliferator acti...	1.874×10^{-2}
7	GO:BP	GO:1904659	glucose transmembrane transport	3.759×10^{-2}
8	GO:BP	GO:0008283	cell population proliferation	4.087×10^{-2}
9	GO:BP	GO:0042592	homeostatic process	4.111×10^{-2}
10	GO:BP	GO:0051049	regulation of transport	4.924×10^{-3}
11	GO:CC	GO:0071944	cell periphery	4.966×10^{-10}
12	GO:CC	GO:0005737	cytoplasm	4.191×10^{-8}
13	GO:CC	GO:0005576	extracellular region	3.636×10^{-4}
14	GO:CC	GO:0030054	cell junction	8.896×10^{-4}
15	GO:CC	GO:0031252	cell leading edge	1.850×10^{-3}
16	GO:CC	GO:0015629	actin cytoskeleton	7.262×10^{-3}
17	GO:CC	GO:0031982	vesicle	9.019×10^{-3}
18	GO:CC	GO:0005587	collagen type IV trimer	2.347×10^{-3}

GO enrichment

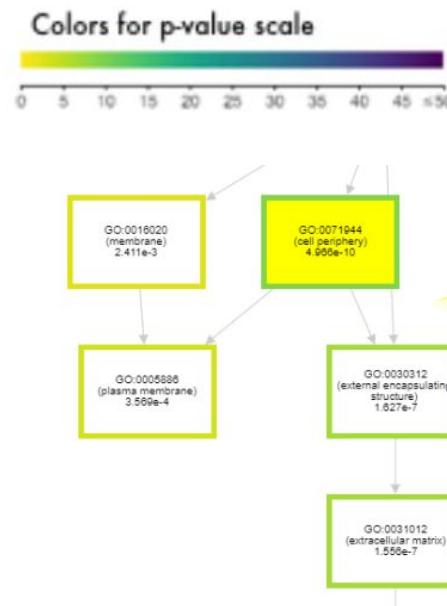
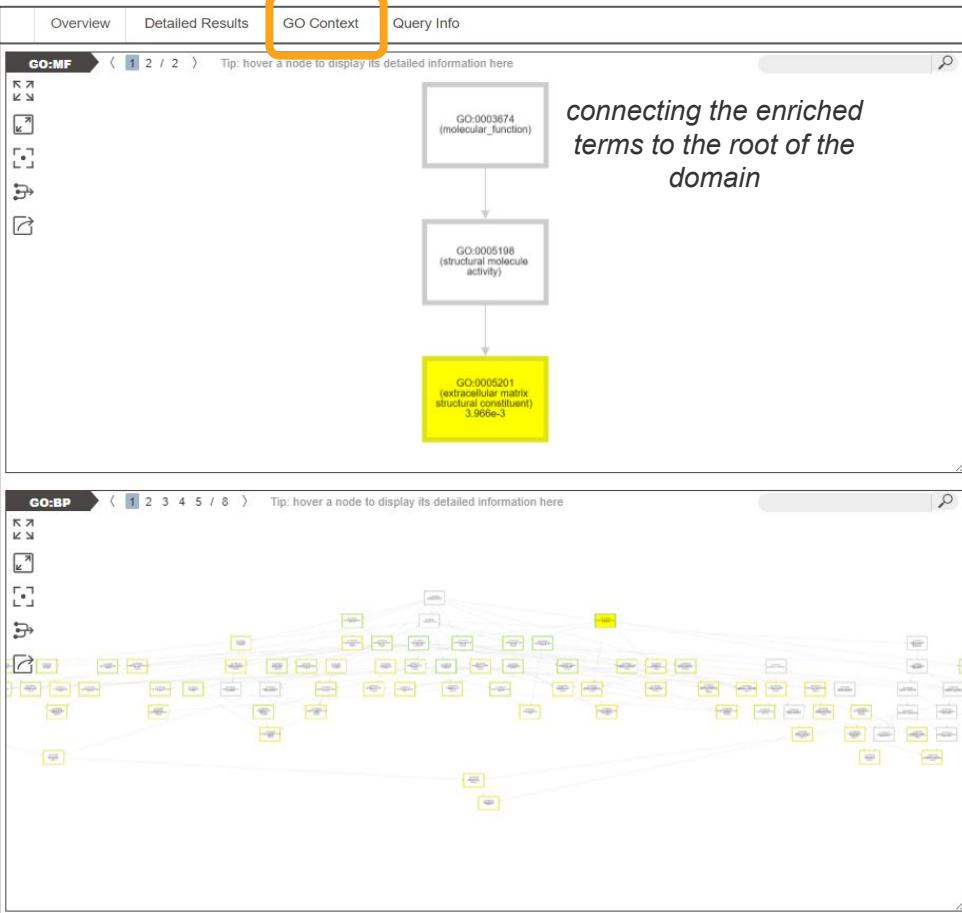


Documentation

g:Profiler



GO enrichment



g:Profiler

Activity: Mapping IDs



Use the Example you've translated in the previous activity
[Plants and Humans.](#)



Now you want to have the GO terms annotated on g:profiler.

GO enrichment



GO enrichment





GO enrichment



Basic parameters

Organism of Interest 

Homo sapiens

Method of Interest 

Over-Representation Analysis (ORA)

Functional Database 

Select a function database category



Select a function database name

Gene List

Select Gene ID Type 

Gene symbol

Upload Gene List 

 Click to upload

Reset

OR

Please enter gene ids...

Reference Gene List

Select Reference Set 

Select the reference set

Upload User Reference Set File and

Select ID type 

Select the ID type of reference set

 Click to upload

Reset



GO enrichment



Basic parameters

Organism of Interest ?

- Homo sapiens
- Arabidopsis thaliana
- Bos taurus
- Caenorhabditis elegans
- Canis lupus familiaris
- Danio rerio
- Sus scrofa
- Drosophila melanogaster
- Gallus gallus
- Homo sapiens**
- Mus musculus
- Rattus norvegicus
- Saccharomyces cerevisiae
- Others



Method of Interest ?

Functional Database ?



Gene List

Select Gene ID Type ?

Upload Gene List ?

OR

Reference Gene List

Select Reference Set ?

Select the reference set

Upload User Reference Set File and

Select ID type ?

Select the ID type of reference set



GO enrichment



Basic parameters

Organism of Interest

Homo sapiens

Method of Interest

Over-Representation Analysis (ORA)

Functional Database

Select a function database category

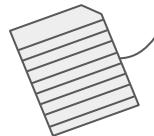


Select a function database name

Gene List

Select Gene ID Type

Upload Gene List



Gene symbol



NCBI Entrez gene

Gene name

Uniprot Swiss-Prot

Phosphosite sequence

Ensembl protein ID with phosphosite annotation

Refseq protein ID with phosphosite annotation

Uniprot protein ID with phosphosite annotation

The Genotype-Tissue Expression Project GTEx

affy Axiom BioBank1

affy Axiom PMRA

affy Axiom tx v1

affy GenomeWideSNP 5

affy GenomeWideSNP 6

affy Mapping10K Xba142

affy Mapping250K Nsp

affy Mapping250K Sty

affy Mapping50K Hind240

affy Mapping50K Xba240

affy OncoScan

NAT2

A1BG

A2M

A2MP1

NAT1

Reference Gene List

Select Reference Set

Upload User Reference Set File and

Select ID type

GO enrichment



Basic parameters

Organism of Interest

Homo sapiens

Method of Interest

Over-Representation Analysis (ORA)

Functional Database

My Axiom BioBank1

affy Axiom PMRA

affy Axiom tx v1

affy GenomeWideSNP 5

affy GenomeWideSNP 6

affy Mapping10K Xba142

affy Mapping250K Nsp

affy Mapping250K Sty

affy Mapping50K Hind240

affy Mapping50K Xba240

affy OncoScan

affy RosettaMerck Human RSTA

affy hc g110

affy hg focus

affy hg u133 plus 2

affy hg u133a

affy hg u133a 2

affy hg u133b

affy hg u95a

affy hg u95av2

Gene List

Select Gene ID Type

Upload Gene List

Reference Gene List

Select Reference Set

Select the reference set

Upload User Reference Set File and

Select ID type

Select the ID type of reference set

Click to upload

Reset

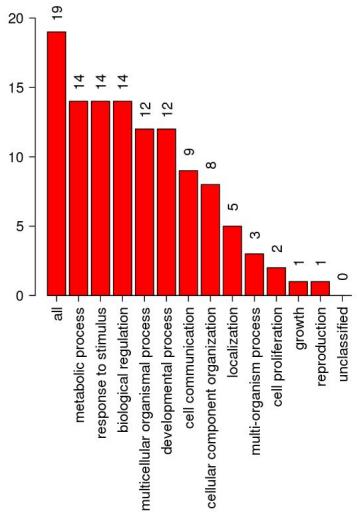
- My Axiom BioBank1
- affy Axiom PMRA
- affy Axiom tx v1
- affy GenomeWideSNP 5
- affy GenomeWideSNP 6
- affy Mapping10K Xba142
- affy Mapping250K Nsp
- affy Mapping250K Sty
- affy Mapping50K Hind240
- affy Mapping50K Xba240
- affy OncoScan
- affy RosettaMerck Human RSTA
- affy hc g110
- affy hg focus
- affy hg u133 plus 2
- affy hg u133a
- affy hg u133a 2
- affy hg u133b
- affy hg u95a
- affy hg u95av2



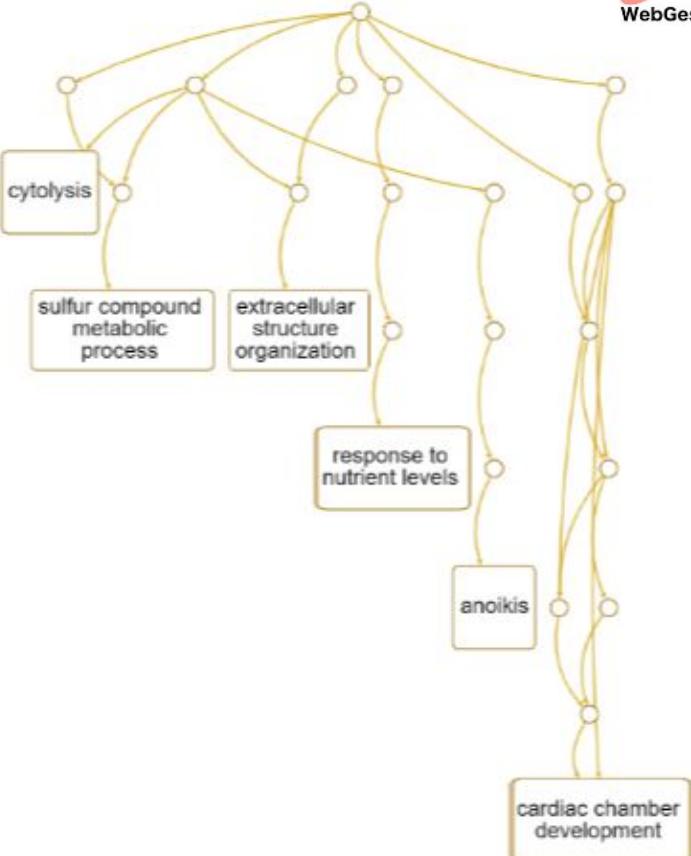
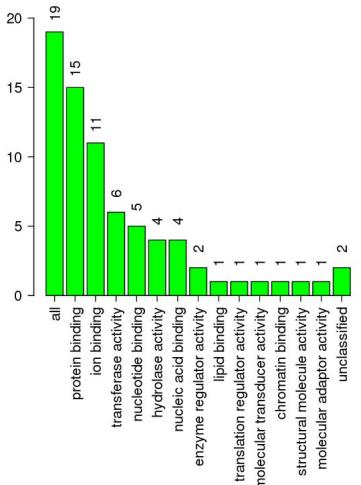
GO enrichment



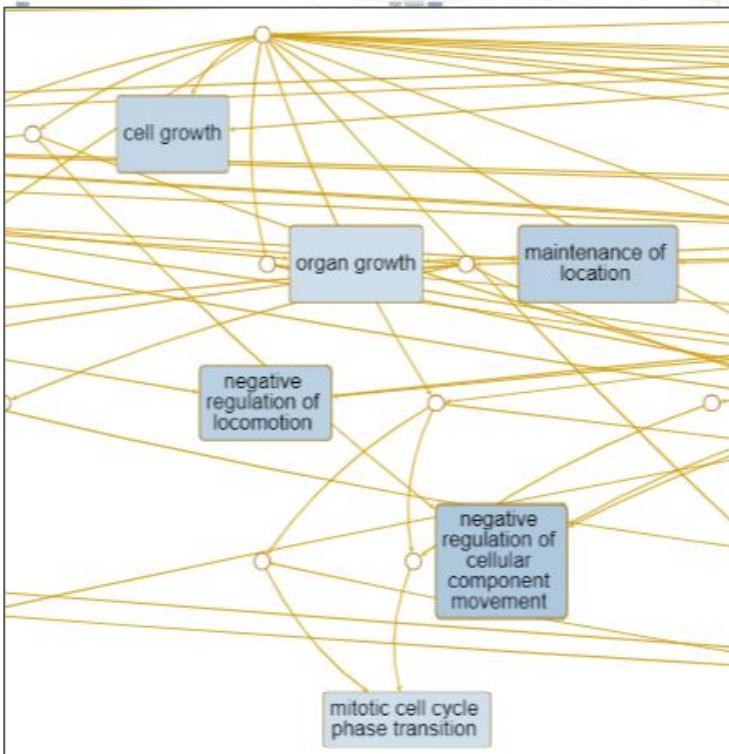
Bar chart of Biological Process categories



Bar chart of Molecular Function categories



GO enrichment



Enrichment Results

Redundancy reduction: All Affinity propagation Weighted set cover

[Table](#)[Bar chart](#)[Volcano plot](#)[DAG](#)

 FDR ≤ 0.05  FDR > 0.05

Activity: Mapping IDs



Use the Example you've translated in the previous activity
[Plants and Humans](#).



Now you want to have the GO terms annotated on WebGstalt.

GO enrichment



GO enrichment



GO enrichment plants



What is PLAZA?

PLAZA is an access point for plant comparative genomics, centralizing genomic data produced by different genome sequencing initiatives.

Available PLAZA instances

News

Dicots PLAZA 5.0 (2021) Primary instance!

Monocots PLAZA 5.0 (2021)

Latest news items

- Public release PLAZA 5.0 2021-05-13
- Preliminary release PLAZA Monocots 4.5 2019-07-10

View all news items

Go to Dicots PLAZA

Species included

Go to Monocots PLAZA

Species included

PLAZA

Home Tools Quick Selection

Analyze Data E2FAT1G01090 Gene All species

Introduction

PLAZA is an access point for plant comparative genomics methods and provides a central resource for plant sequence data and comparative analyses and data mining within the green plant lineage (*Virodip/antae*). Please use the search functions in the menu bar to find specific genes or genomic regions. For more details, please check the Documentation.

News

Dicots PLAZA 5.0 summary

- Integration of structural and functional data from 100 species.
- Includes 4,234,318 genes, of which 2,000,000 are protein coding. These 2,000,000 genes are clustered in 68,305 multi-species families (47.5% multi-specific).
- More information on each gene can be found in the data overview.

Analyze tools:

- Expansion plot
- Gene family finder
- Synteny plot
- Ks-graphs
- Skyline plot
- WGDotplot
- Colinear region finder
- Functional clusters
- WGMapping
- Interactive Phylogenetics Module
- BLAST
- Workbench

Arabidopsis lyrata
Arabidopsis thaliana
Capsella rubella
Cardamine hirsuta
Brassica carinata
Brassica oleracea
Brassica rapa
Brassica napus
Schrenkia parvula
Eutrema salsugineum
Aethionema arabicum
Tarenaya hassleriana
Carica papaya
Corchorus olitorius
Theobroma cacao
Gossypium hirsutum



GO enrichment plants



The screenshot shows the PLAZA workbench interface. At the top, there's a navigation bar with 'Data', 'Analyze', 'E2F,AT1G01090', 'Gene', 'All species', and a search bar. Below the navigation is a banner with the word 'Plaza' and a green grass background. The main area has a 'Workbench' tab and a 'Features' section. The 'Authentication' section contains a dropdown menu with 'PLAZA' selected. A large hand cursor is pointing at the 'PLAZA' option. To the right, there's a detailed view of the 'Authentication' form with fields for 'Login' and 'Password', and buttons for 'Login', 'Register', and 'Reset password'. A second hand cursor is pointing at the 'Login' button.

Workbench User Registration

Registration

To use the PLAZA workbench, a user account must first be setup. To register, please provide a valid e-mail address below. A mail containing the login and password to be used will be sent automatically.

E-mail address *

Title

University/Institute/Company name *

Access Type (Licensing information)

Academic Access Industry Access

Country

I have read and I agree with the [terms of service](#)

GO enrichment plants



PLAZA

Data Analyze E2FAT1G01090 Gene All species Dicots PLAZA 5.0

Workbench

User Information

Login: bruna.piereckmoura@vib.be
Authentication Type: PLAZA
Profile: Change profile settings
Password: Change password
API: Generate API key and view API history

Add New Experiment — 96 experiments remaining

Experiment name*: BE_PJ_Salt_dec23

Experiment description: Describe your experiment here:
Plant name: XXXXX
Type of experiment: Stress Salt
Tissue: Roots and Leaves
Conditions: 1 week time, X% concentration, etc ...

Create experiment

Workbench: **Pro. juliflora - JP**

Description: Prosopis juliflora cultivar:DLEG910003 Stress: Salt & Drought Assembly: Trinity
Last edited: 2023-11-24 15:19:15
Gene count: 1

Note: Only coding genes are clustered into gene families. See the documentation for more detailed information on gene families.
The indicated percentages represent the number of coding genes in non-orphan gene families divided by the total number of genes.

Actions

- Import using gene identifiers
- Import using BLAST
- Import using BED file
- Export genes, annotations and sequences
- Copy experiment
- Share experiment
- Change experiment settings
- Back to experiments overview

Species

Name	Count
Coffea canephora	1

Genes

Type	Count
Coding gene	1

Gene Families

Type	Count
Homologous gene family	1 (100%)
Orthologous gene family	1 (100%)

Toolbox

Compare

- ... exon-intron gene structures.
- ... gene duplication types.
- ... with other PLAZA workbench experiments.

Analyze

- ... the enrichment of
 - GO functional data
 - InterPro functional data
 - MapMan functional data

View

- ... associated gene families.
- ... the associated functional annotation.
 - GO terms
 - InterPro domains
 - MapMan terms
- ... the genome-wide organization.
- ... the orthologous genes
 - using the PLAZA integrative method.
 - using the PLAZA gene families.

GO enrichment plants



Workbench: **Pro. juliflora - JP** ⓘ

Experiment overview

Name	Pro. juliflora - JP
Last edit	2023-11-24 15:19:15
Gene count	1

Import genes

Gene identifiers (e.g. AT1G01270) can be entered in the text area below. Multiple genes can be separated by whitespaces, tabs and/or newlines.

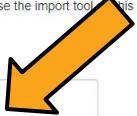
Non-standard identifiers will be converted to the primary gene identifiers, whenever possible.

If you use the *direct mapping* approach, then no secondary screen will be used to determine the species and genes.

If you do **NOT** use the *direct mapping* approach, then you can import only **500** identifiers at the same time.

If you wish to have more genes in your experiment, you can use the import tool on this web page multiple times with a different gene set.

Input

 Map directly to species


Workbench: **Pro. juliflora - JP** ⓘ

Description

Prosopis juliflora cultivar:DLEG910003 Stress: Salt & Drought Assembly: Trinity
- by Joao Pacifico 15 days plants

Last edited

2023-11-24 15:19:15

Gene count

1

Note: Only coding genes are clustered into gene families. See the documentation for more detailed information on gene families.
The indicated percentages represent the number of coding genes in non-orphan gene families divided by the total number of genes.

Actions

- Import using gene identifiers
- Import using BLAST
- Import using BED file
- Export genes, annotations and sequences
- Copy experiment
- Share experiment
- Change experiment settings
- Back to experiments overview



Species

Name	Count
Coffea canephora	1

Genes

Type	Count
Coding gene	1

Gene Families

Type	Count
Homologous gene family	1 (100%)
Orthologous gene family	1 (100%)

Toolbox

Compare

- ... exon-intron gene structures.
- ... gene duplication types.
- ... with other PLAZA workbench experiments.

Analyze

- ... the enrichment of
 - GO functional data
 - InterPro functional data
 - MapMan functional data

Toolbox

Compare

- ... exon-intron gene structures.
- ... gene duplication types.
- ... with other PLAZA workbench experiments.

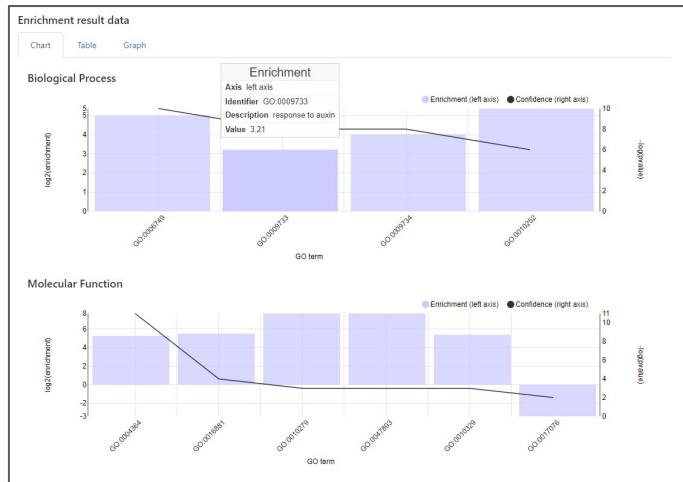
Analyze

- ... the enrichment of
 - GO functional data
 - InterPro functional data
 - MapMan functional data
- ... associated gene families.
- ... the associated functional annotation.
 - GO terms
 - InterPro domains
 - MapMan terms
- ... the genome-wide organization.
- ... the orthologous genes.
 - using the PLAZA integrative method.
 - using the PLAZA gene families.

View



GO enrichment plants

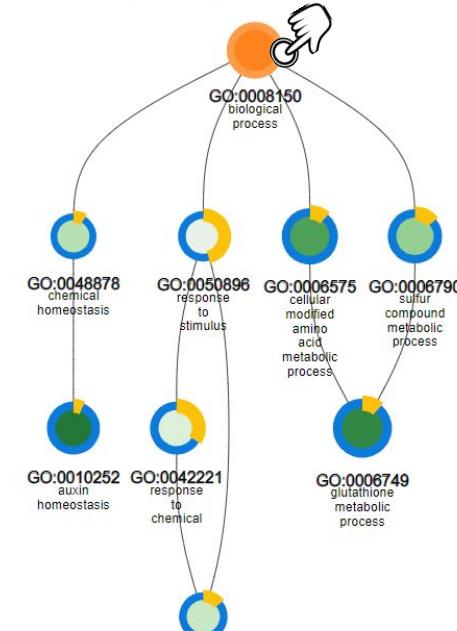


Enrichment result data

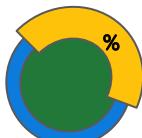
Chart

Table

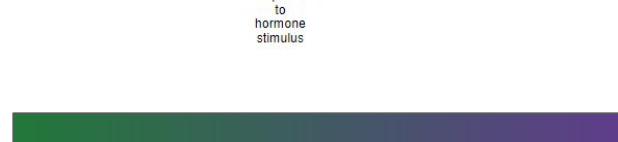
Graph



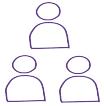
Overrepresented



Underrepresented



Activity: Mapping IDs



Now if you work with plant or if you are curious to try ...

Use the Example you've translated in the previous activity

[Plants](#) .

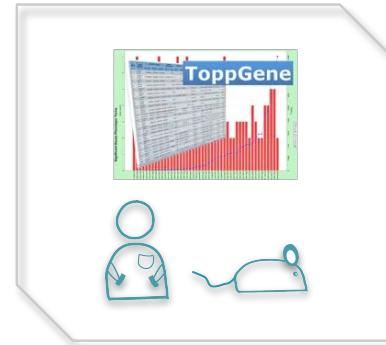


Now you want to have the GO terms annotated on PLAZA.

GO enrichment



GO enrichment



GO enrichment humans



ToppGene Suite

A one-stop portal for gene list enrichment analysis and candidate gene prioritization based on functional annotations and protein interactions network

ToppGene:

- Home
- API
- Links
- Database details
- Supplementary
- Help
- Publications
- Terms of Use
- Contacts

Other Projects:

- ToppCluster
- Lungmap Heatmaps

TopFun
Functional enrichment

↑

- **ToppFun:** Transcriptome, ontology, phenotype, proteome, and pharmacome annotations based gene list functional enrichment analysis
Detect functional enrichment of your gene list based on Transcriptome, Proteome, Regulome (TFBS and miRNA), Ontologies (GO, Pathway), Phenotype (human disease and mouse phenotype), Pharmacome (Drug-Gene associations), literature co-citation, and other features.
- **ToppGene:** Candidate gene prioritization
Prioritize or rank candidate genes based on functional similarity to training gene list.
- **ToppNet:** Relative importance of candidate genes in networks
Prioritize or rank candidate genes based on topological features in protein-protein interaction network.
- **ToppGenet:** Prioritization of neighboring genes in protein-protein interaction network
Identify and prioritize the neighboring genes of the seeds in protein-protein interaction network based on functional features in protein-protein interaction network (ToppNet).

Select your gene identifier type, paste your sets below or select example set, then submit.

Symbol Type: HGNC Symbol

Example gene sets: HGNC Symbol

Enrichment Gene Set: HGNC Symbol and Synonyms

Background Symbol Type: HGNC Symbol

Optional Background Set: (the example training and test set of genes)

If empty, the full gene set of each category will be used.

Ensembl ID

RefSeq

Uniprot

ENSG00000003402

ENSG00000004799

ENSG00000006788

ENSG00000008256

ENSG00000008311

ENSG00000009413

Clear

Submit

GO enrichment humans



ToppGene Suite

A one-stop portal for gene list enrichment analysis and candidate gene prioritization based on functional annotations and protein interactions network

Input Gene List (380 / 403)		Background Gene List (16350 / 20405)	
Entered		Human Symbol	Gene ID
ENSG00000103198 CRISPLD2	(cysteine rich secretory protein LOC1 domain containing 2)	59189	
ENSG00000105835 NAMPT	(nucleoside phosphorylase/nampt)	83716	
ENSG00000106032 CPEO1	(coiledin like and POFV-like eukaryotic domain containing 1)	10935	
ENSG00000106123 EPB46	(EPB receptor BB)	79874	
ENSG00000106611 HEB	(Ets homolog, mTORC1 binding)	2051	
ENSG00000106617 PRKG2	(protein kinase AMP-activated non-catalytic subunit gamma 2)	6099	
ENSG00000107104 KANK1	(KAN motif and ankyrin repeat domains 1)	51422	
ENSG00000107968 MAP3K8	(mitogen-activated protein kinase kinase kinase 8)	23184	
ENSG00000108387 SEPTIN4	(septin 4)	59	
ENSG00000108604 SMARD2	(SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 2)	6609	
ENSG00000108924 HLF	(HLF transcription factor, PAR BZIP family member)	3131	
ENSG00000108950 FAM20A	(FAM20A gelatin associated secretory pathway pseudokinase)	54757	
ENSG00000109860 MMD	(monocyte to macrophage differentiation associated)	23531	
ENSG00000109905 BTB16	(BTB finger and BTB domain containing 16)	7704	
ENSG00000110915 IP55	(IP55 biogenesis of lysosomal organelles complex 2 subunit 2)	11234	
ENSG00000112216 GRN	(natural killer cell expressed, developmentally down-regulated 9)	4739	
ENSG00000112216 GRN	(natural killer cell expressed, developmentally down-regulated 9)	81491	
ENSG00000112216 GRN	(natural killer cell expressed, developmentally down-regulated 9)	730	
ENSG00000112396 CT	(complement C7)	730	
ENSG00000114270 COL7A1	(collagen type VII alpha 1 chain)	1294	
ENSG00000115419 Q1 S	(Q1 S (alpha 1 chainless))	2744	
Genes Not Found			
Entered	Status		
ENSG00000123262	Not Found		
ENSG00000141150	Not Found		
ENSG00000152268	Not Found		
ENSG00000162840	Not Found		
ENSG00000174680	Not Found		
ENSG00000180672	Not Found		
ENSG00000213763	Not Found		
ENSG00000226121	Not Found		
ENSG00000233117	Not Found		
ENSG00000233929	Not Found		
ENSG00000237697	Not Found		
ENSG00000247311	Not Found		
ENSG00000250978	Not Found		
ENSG00000260396	Not Found		
ENSG00000260841	Not Found		
ENSG00000261468	Not Found		
ENSG00000261490	Not Found		
ENSG00000261685	Not Found		

Calculations

pValue Method: Probability density function

Feature	p-Value cutoff	Gene Limits
<input checked="" type="checkbox"/> All	FDR	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> GO: Molecular Function	Bonferroni	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> GO: Biological Process	FDR B&Y	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> GO: Cellular Component	FDR	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> Human Phenotype	None	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> Mouse Phenotype	FDR	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> Domain	FDR	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> Pathway	FDR	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> Pubmed	FDR	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> Interaction	FDR	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> Cytoband	FDR	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> Transcription Factor Binding Site	FDR	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> Gene Family	FDR	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> Coexpression	FDR	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> Coexpression Atlas	FDR	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> ToppCell Atlas	FDR	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> Computational	FDR	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> MicroRNA	FDR	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> Drug	FDR	1 ≤ n ≤ 2000
<input checked="" type="checkbox"/> Disease	FDR	1 ≤ n ≤ 2000

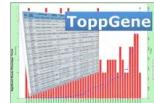
Group results within a category by source.



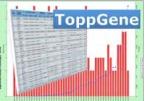
Home

Modify Query

Start



GO enrichment humans


[Go To Start Page](#)

Input Parameters [Show Detail]

Training Results [Expand All] [Download All] [Sparse Matrix] Display pValues and Scores as Scientific (4 significant digits) Table row limit 50

1: GO: Molecular Function [Display Chart] 38 input genes in category / 1000 annotations before applied cutoff / 14053 genes in category

ID	Name	Source	pValue	FDR B&H	FDR B&Y	Bonferroni	Genes from Input	Genes in Annotation
1	GO:0005539 glycosaminoglycan binding		1.822E-5	1.583E-2	1.185E-1	1.822E-2	16	183
2	GO:0005201 extracellular matrix structural constituent		3.166E-5	1.583E-2	1.185E-1	3.166E-2	14	151

2: GO: Biological Process [Display Chart] 31 input genes in category / 5073 annotations before applied cutoff / 13934 genes in category

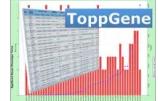
ID	Name	Source	pValue	FDR B&H	FDR B&Y	Bonferroni	Genes from Input	Genes in Annotation
1	GO:0009725 response to hormone		9.019E-10	4.575E-6	4.168E-5	4.575E-6	52	813
2	GO:0007155 cell adhesion		1.202E-8	3.049E-5	2.778E-4	6.099E-5	68	1254
3	GO:0048846 anatomical structure formation involved in morphogenesis		8.770E-8	1.483E-4	1.351E-3	4.449E-4	69	1407
4	GO:0030198 extracellular matrix organization		1.763E-7	1.795E-4	1.635E-3	8.943E-4	28	316
5	GO:0043062 extracellular structure organization		1.876E-7	1.795E-4	1.635E-3	9.517E-4	28	317

[Show 12 more annotations](#)

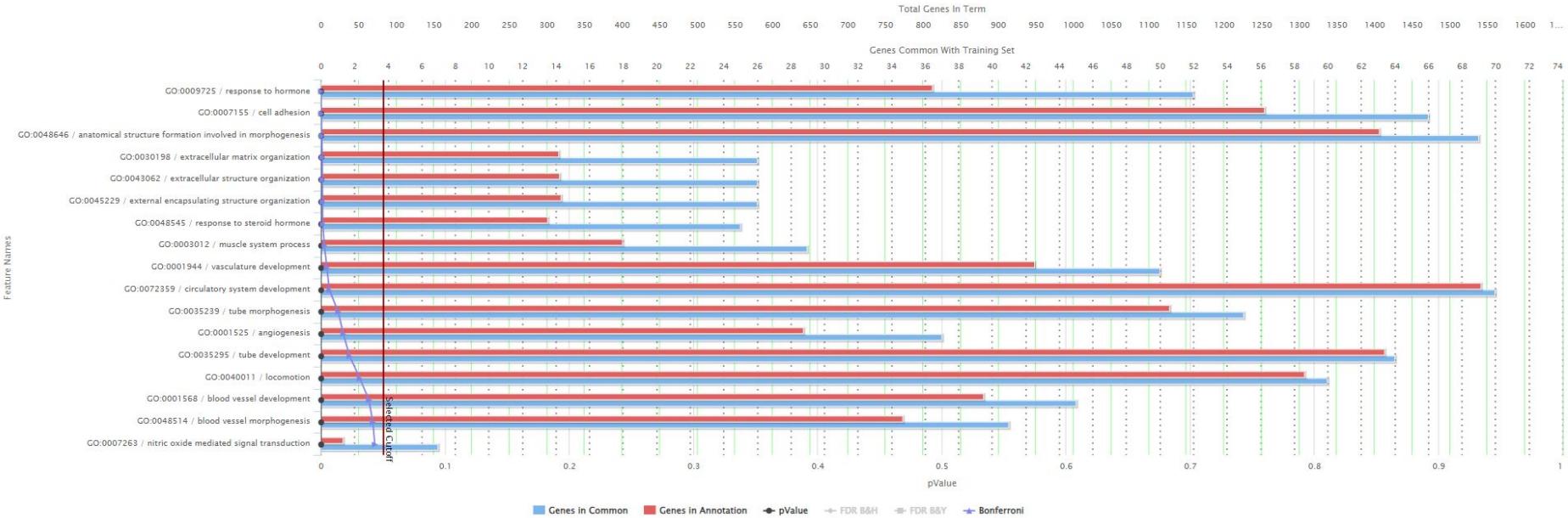
3: GO: Cellular Component [Display Chart] 12 input genes in category / 499 annotations before applied cutoff / 14083 genes in category

ID	Name	Source	pValue	FDR B&H	FDR B&Y	Bonferroni	Genes from Input	Genes in Annotation
1	GO:0030312 external encapsulating structure		4.314E-8	1.078E-5	7.309E-5	2.153E-5	34	466
2	GO:0031012 extracellular matrix		4.314E-8	1.078E-5	7.309E-5	2.153E-5	34	466
3	GO:0062023 collagen-containing extracellular matrix		2.305E-6	3.984E-4	2.705E-3	1.195E-3	27	385

GO enrichment humans



Significant Terms For: GO: Biological Process



Activity: Mapping IDs



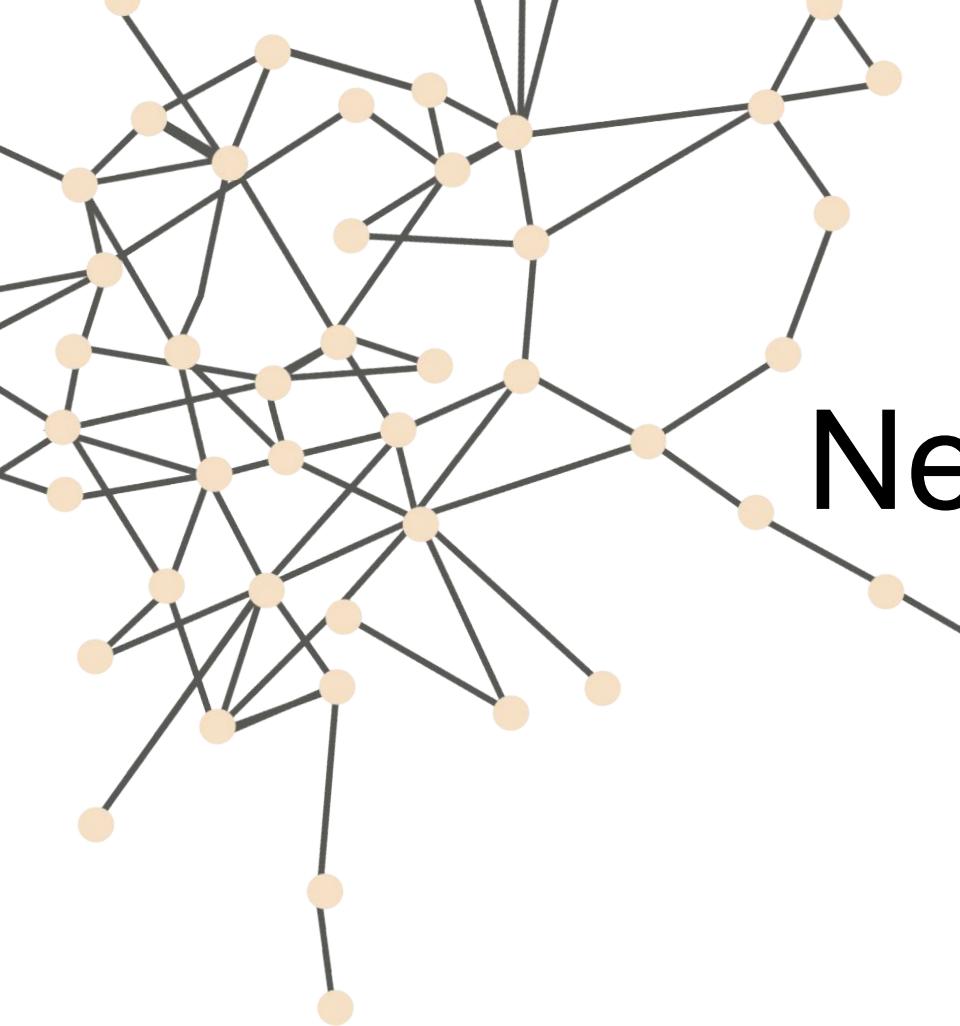
Now if you work with Humans or mouse or if you are curious to try ...

Use the Example you've translated in the previous activity

[Humans.](#)



Now you want to have the GO terms annotated on ToppGene.

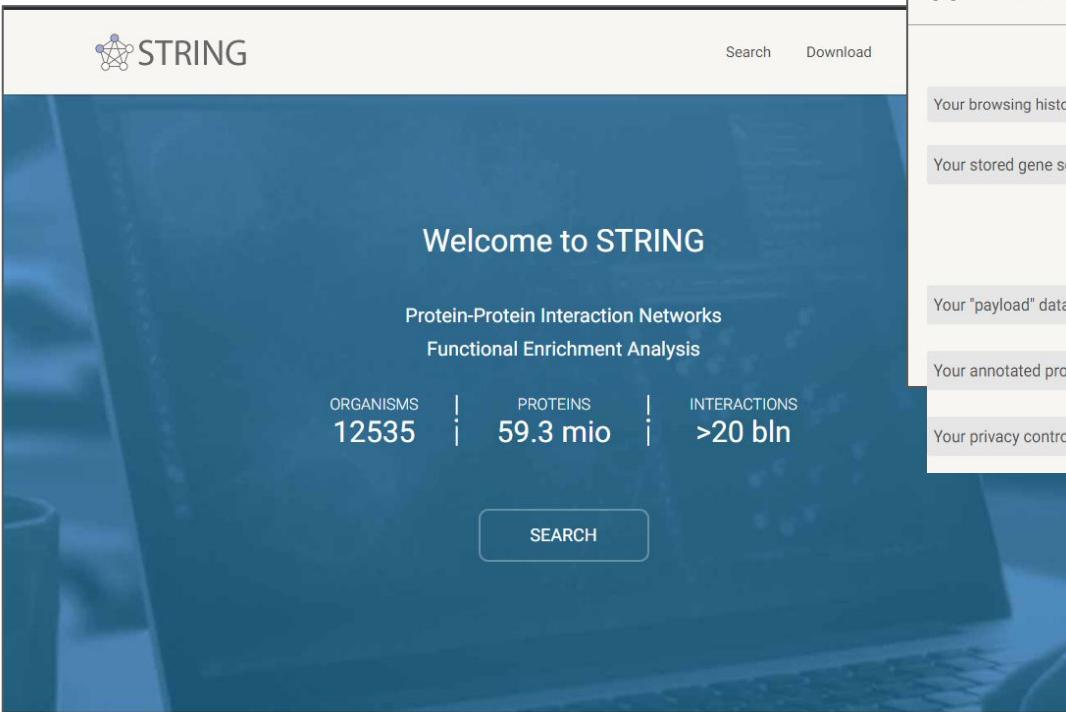


Network Analysis

Pathways and what they say

Network analysis

String



STRING

Welcome to STRING

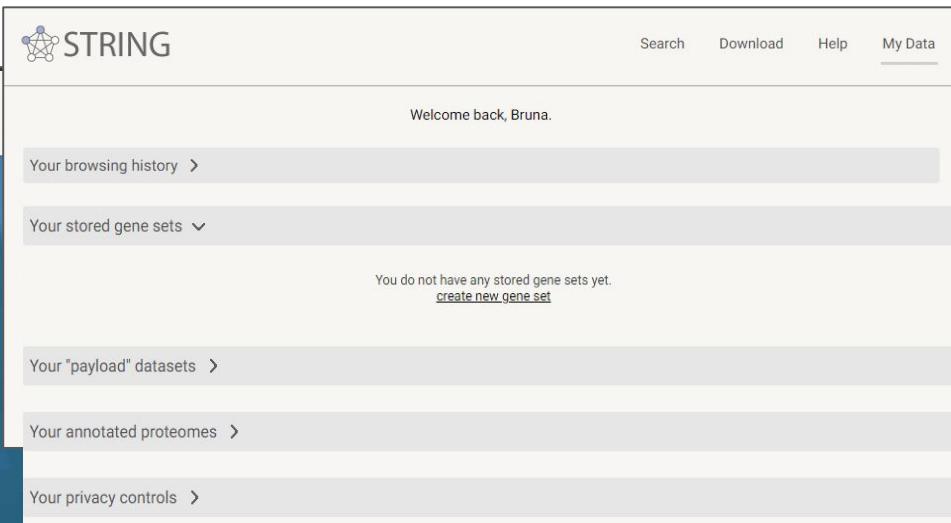
Protein-Protein Interaction Networks
Functional Enrichment Analysis

ORGANISMS | PROTEINS | INTERACTIONS

12535 | 59.3 mio | >20 bln

SEARCH

The screenshot shows the main landing page of the STRING website. It features a dark blue background with a faint image of a laboratory setup. At the top right are links for "Search", "Download", and "Help". Below that is a navigation bar with the STRING logo and a search bar. The central area displays key statistics: 12535 organisms, 59.3 million proteins, and over 20 billion interactions. A large "SEARCH" button is prominently displayed at the bottom.



STRING

Welcome back, Bruna.

Your browsing history >

Your stored gene sets ▾

You do not have any stored gene sets yet.
[create new gene set](#)

Your "payload" datasets >

Your annotated proteomes >

Your privacy controls >

This screenshot shows the user profile section of the STRING website. It includes a welcome message for the user "Bruna". The sidebar contains links for browsing history, stored gene sets (which are currently empty), payload datasets, annotated proteomes, and privacy controls.

Create a New Gene Set

1) name your new set ...

name for future reference:

my gene set no. 1

2) identify your organism ...

Homo sapiens

3) provide identifiers ...

list of gene/protein names:

(one identifier per line or CSV; examples: #1 #2 #3)

... or upload a file:
 Choose File No file chosen

4) set additional options ...

usage:

 enable usage as a statistical background set

5) launch ...

< BACK

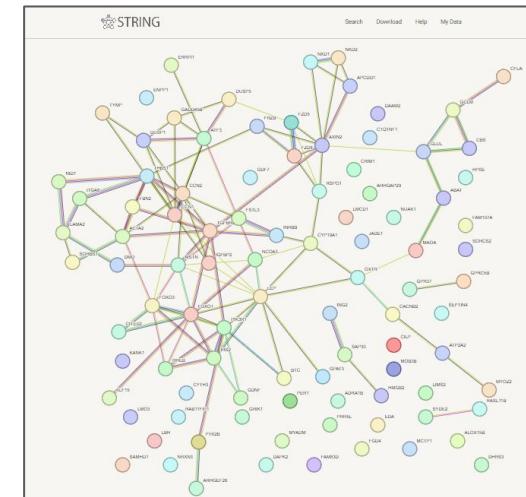
CONTINUE ➔

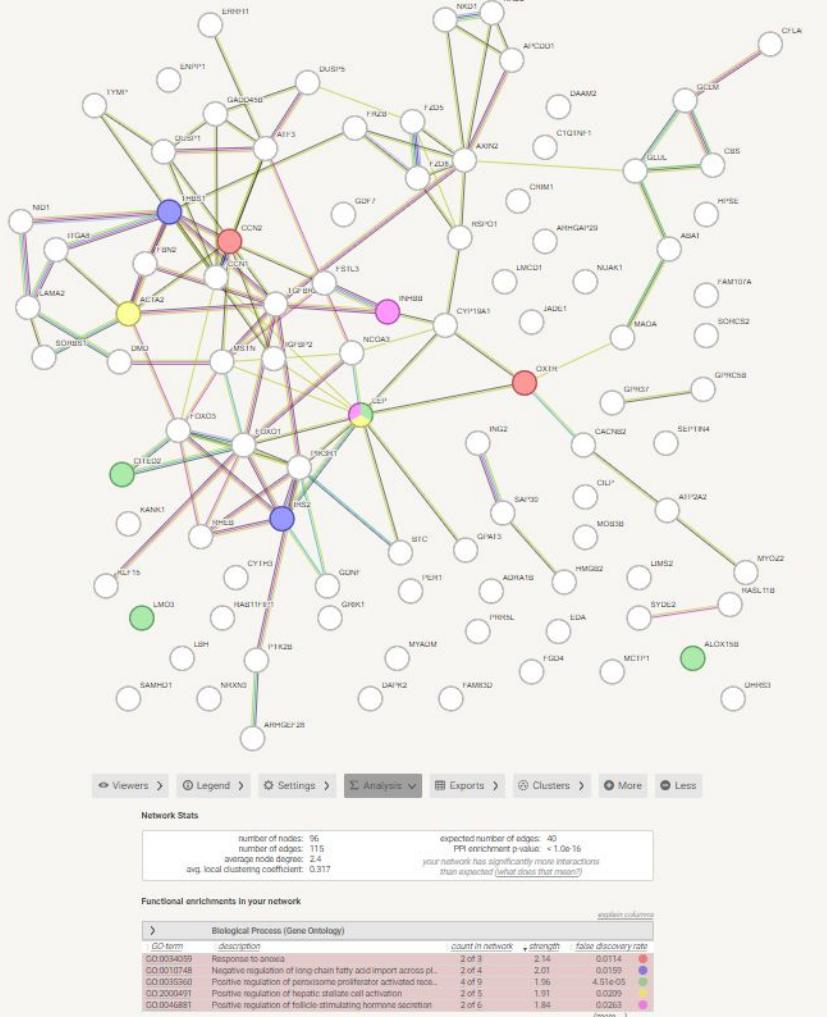
Stored Geneset

reference:	"exercise"
organism:	Homo sapiens
distinct input items:	96
distinct genes mapped:	96
enable statistics use:	yes
last updated:	2023-12-02, 21:26:32

VIEW AS INTERACTION NETWORK

input rank	input term	mapped gene	annotation
96	ENPP1	ENPP1	Ectonucleotide pyrophosphatase/phosphodiesterase family member 1, secreted form; Nucleotid...
95	SAMHD1	SAMHD1	Deoxyribonucleoside triphosphate triphosphohydrolase SAMHD1; Protein that acts both as a host res...
94	NRXN3	NRXN3	Neurexin-3; Neuronal surface protein that may be involved in cell recognition and cell adhesio...
93	DAPK2	DAPK2	Death-associated protein kinase 2; Calcium/calmodulin-dependent serine/threonine kinase involv...
92	FAMB3D	FAMB3D	Protein FAMB3D; Probable proto-oncogene that regulates cell proliferation, growth, migration and...
91	DHRS3	DHRS3	Short-chain dehydrogenase/reductase 3; Catalyzes the reduction of all-trans-retinol to all-trans-re...
90	KANK1	KANK1	KN motif and ankyrin repeat domain-containing protein 1; Involved in the control of cytoskeleton f...
89	CITED2	CITED2	Cbp/p300-interacting transactivator 2; Transcriptional coactivator of the p300/CBP-mediated tran...
88	LM03	LM03	LIM domain only 3.
87	ARHGEF28	ARHGEF28	Rho guanine nucleotide exchange factor 28; Functions as a RHOA-specific guanine nucleotide ex...
86	ATP2A2	ATP2A2	FYVE, RhoGEF and PH domain-containing protein 4; Activates CDC42, a member of the Ras-like fa...
85	FGD4	FGD4	Phosphatidylinositol 3-kinase regulatory subunit alpha; Binds to activated (phosphorylated) protei...
84	PIK3R1	PIK3R1	Fibrillin-2 C-terminal peptide; [Fibrillin-2]; Fibrillins are structural components of 10-12 nm extracel...
83	FBX2	FBX2	Multiple C2 and transmembrane domain-containing protein 1; Calcium sensor which is essential fo...
82	MCTP1	MCTP1	VPS10 domain-containing receptor SCRS2; The heterodimer formed by NGFR and SCRS2 functi...
81	SORCS2	SORCS2	Actin-associated protein FAM107A; Stress-inducible actin-binding protein that plays a role in syna...
80	FAM107A	FAM107A	Glial cell line-derived neurotrophic factor; Neurotrophic factor that enhances survival and mor...
79	CDNE	CDNE	





Viewers **Legend** **Settings** **Analysis** **Exports** **Clusters** **More** **Less**

 **Network** currently showing Summary view: shows current interactions. Nodes can be moved; popups provide information on nodes & edges.

 **Neighborhood** Groups of genes that are frequently observed in each other's genomic neighborhood.

 **Experiments** Co-purification, co-crystallization, Yeast2Hybrid, Genetic Interactions, etc ... as imported from primary sources.

 **Databases** Known metabolic pathways, protein complexes, signal transduction pathways, etc ... from curated databases.

 **Textmining** Automated, unsupervised textmining - searching for proteins that are frequently mentioned together.

 **Coexpression** Proteins whose genes are observed to be correlated in expression, across a large number of experiments.

Basic Settings

Network type:

- full STRING network (the edges indicate both functional and physical protein associations)
- physical subnetwork (the edges indicate that the proteins are part of a physical complex)

UPDATE

meaning of network edges:

- evidence (line color indicates the type of interaction evidence)
- confidence (line thickness indicates the strength of data support)

active interaction sources:

- Textmining
- Experiments
- Databases
- Co-expression
- Neighborhood
- Gene Fusion
- Co-occurrence

max number of interactors to show:

1st shell: - none / query proteins only - 2nd shell: - none - **Advanced Settings**

network display mode:

- static png (network is a simple bitmap image; not interactive)
- interactive svg (network is a scalable vector graphic [SVG]; interactive)

network display options:

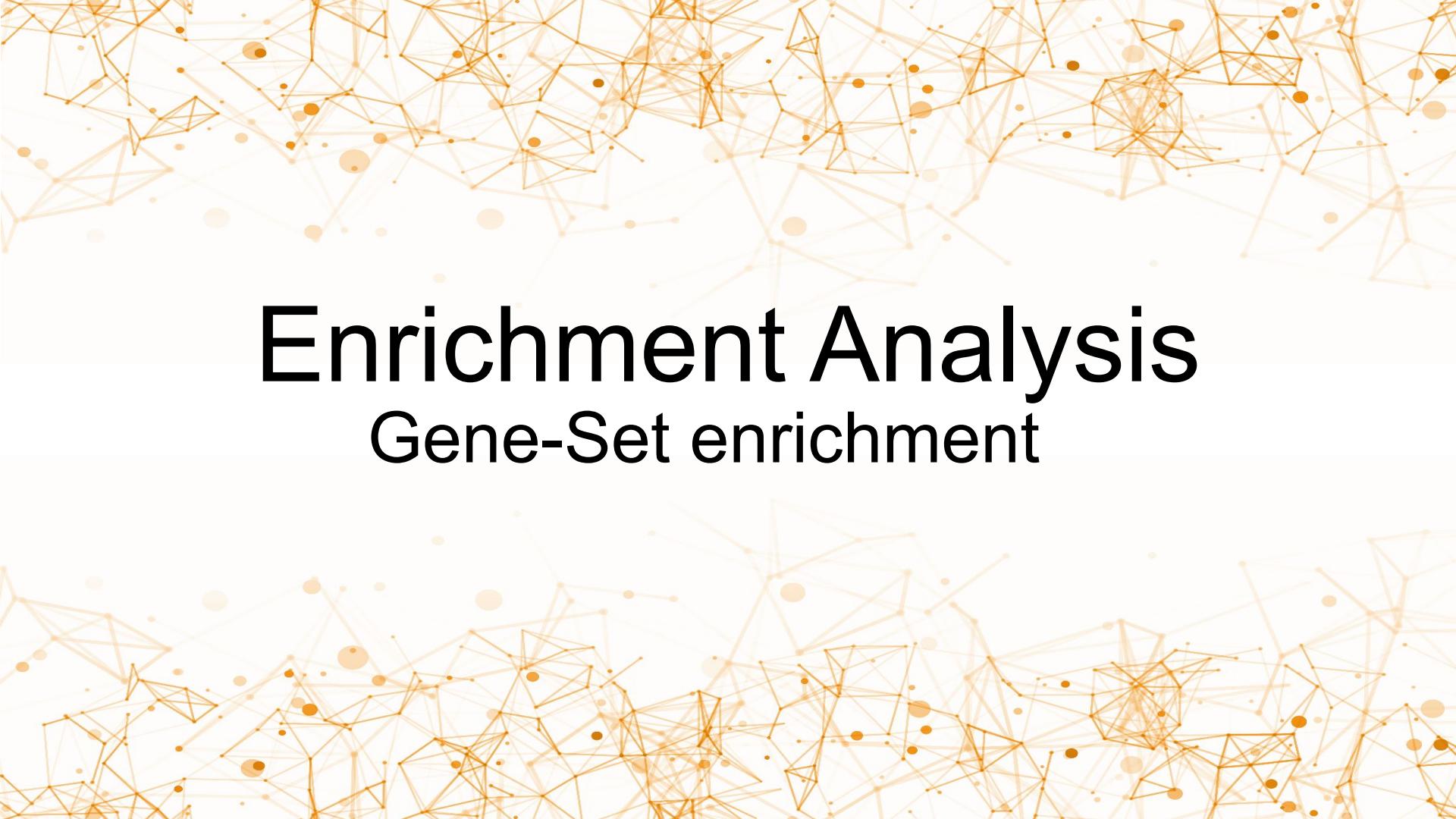
- Enable node coloring mode
- enable 3D bubble design
- disable structure previews inside network bubbles
- center protein names on nodes
- show your query protein names
- hide disconnected nodes in the network
- hide protein names
- protein name font size

Activity: Network analysis plant



Now try to use the up-regulated genes of *Homo sapiens* or *Oryza sativa* visualize the topology. Filter based on confidence and highlight some functional GO annotation.



The background of the slide features a complex, abstract network structure composed of numerous small, semi-transparent orange dots connected by thin, light-orange lines. This pattern creates a sense of depth and connectivity, resembling a molecular or neural network.

Enrichment Analysis

Gene-Set enrichment

Gene Set enrichment

[Bulk RNAseq func. analysis & codes](#)



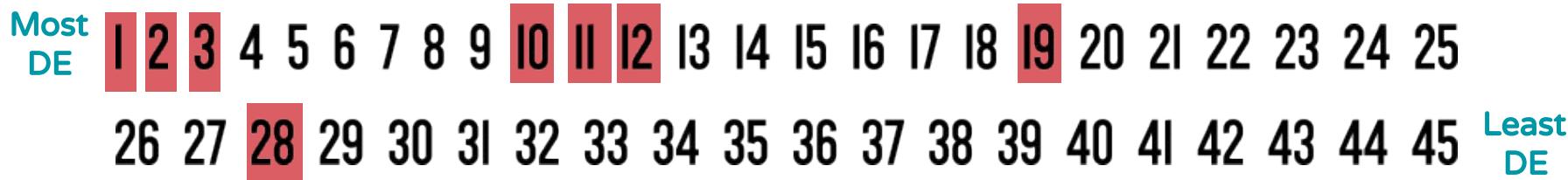
[Single cell funct. analysis & codes](#)



- We choose a set of genes based on:
 - All genes tested for DE
 - Pathway annotation
- We evaluate the set based on Diff expression
 - Sort p-value
 - Sort logFC
 - Sort normalized counts



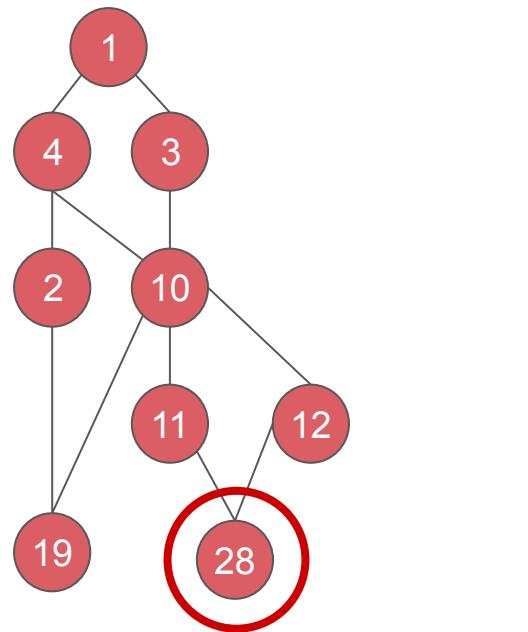
Gene Set enrichment



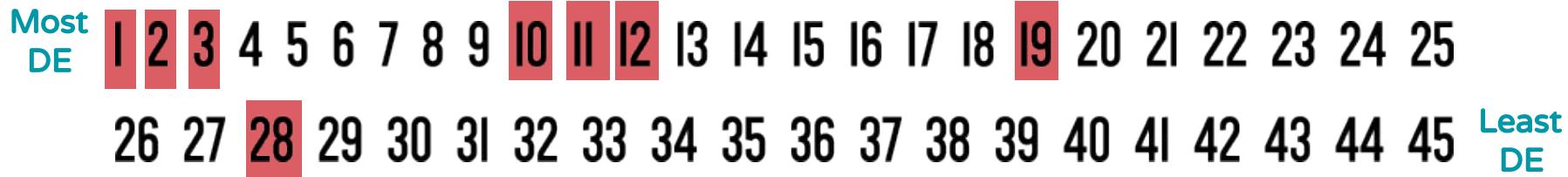
+DE 1

rank

- 45 - DE



Gene Set enrichment



Many are among most DE

Most probably pathway is affected

Most
DE

I 2 3 4 5 6 7 8 9 10 II 12 13 14 15 16 17 18 19 20 21 22 23 24 25

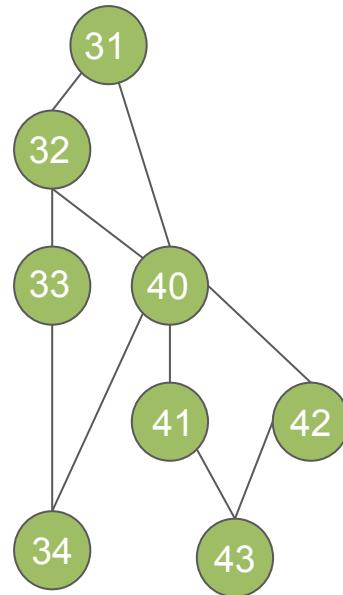
26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 Least
DE



+DE 1

rank

45 - DE



Most
DE

I 2 3 4 5 6 7 8 9 10 II 12 13 14 15 16 17 18 19 20 21 22 23 24 25

26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 Least
DE



All are not much DE

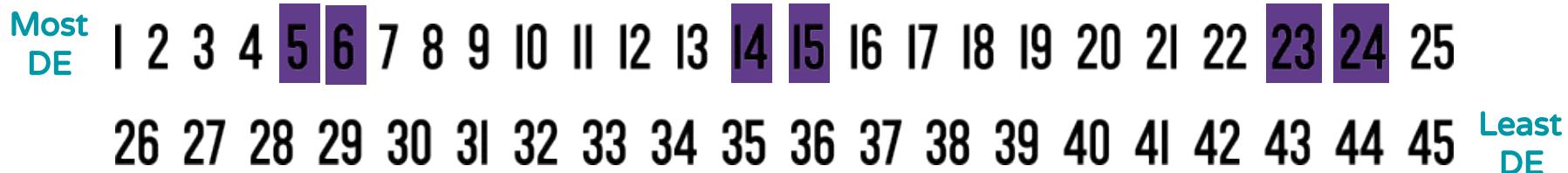
Most probably pathway is NOT affected

+DE 1

rank

45 - DE

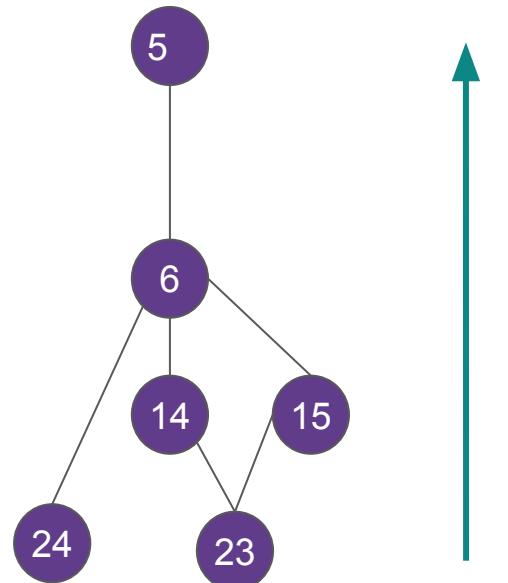
Gene Set enrichment



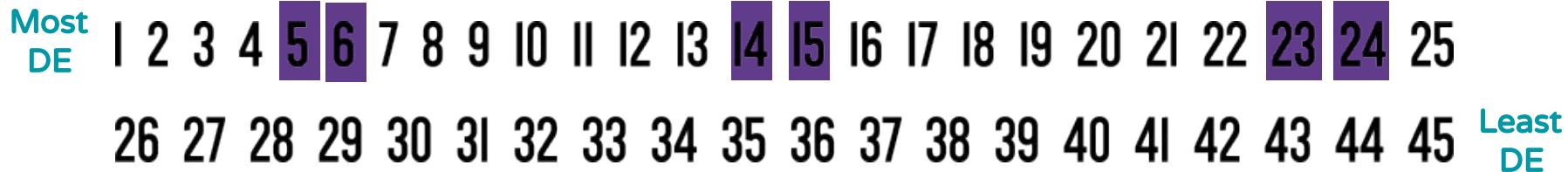
+DE 1

rank

45 - DE



Gene Set enrichment



All are among most DE

Most probably pathway is affected

+DE 1

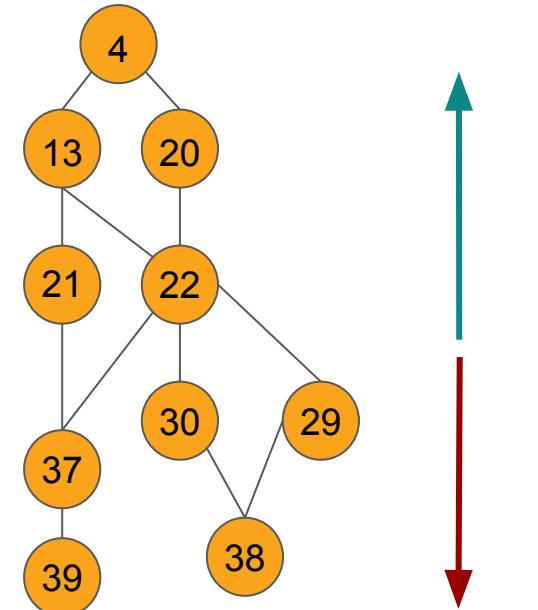
rank

45 - DE

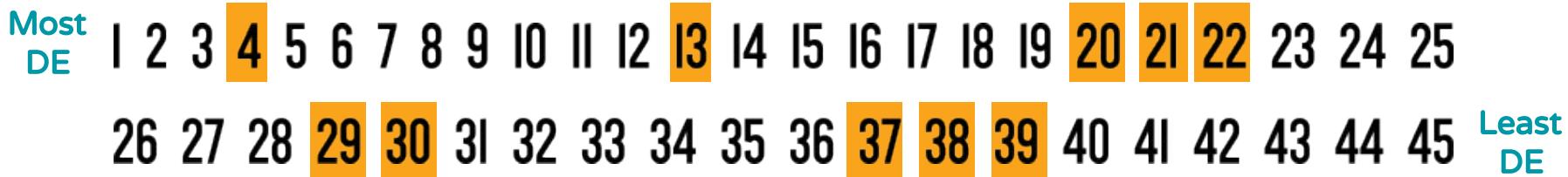
Gene Set enrichment



+DE 1 rank 45 - DE



Gene Set enrichment



Genes in are spread,
inconsistent answer

Most probably pathway is NOT affected

Gene Set enrichment

1

rank

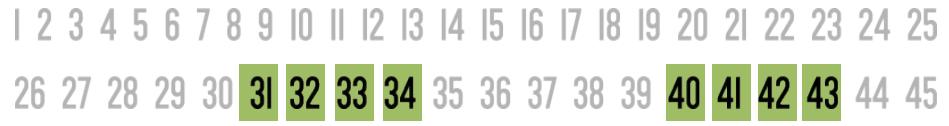
45

Most DiffExpressed**Least DiffExpressed**

Genes are spread in the rank.
Some are very DE others are not.
Not affected



Genes are mostly in top of the rank.
Most are very DE.
It is affected



Genes are spread in the rank.
Some are DE others are not.
Not affected



Genes are spread in the top.
All are very DE.
It is affected

Gene Set enrichment Analysis

Manual 4 interpretations

File name: file.rnk

It is a RANKED list format (tab delimited)

Number must be represented with points not commas

No headers



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](#) | [User Forum](#) | [GOView](#) | [WebGestaltR](#) | [WebGestalt 2017](#)

Basic parameters

Organism of Interest (dropdown menu)

Method of Interest (dropdown menu)
Over-Representation Analysis (ORA)
Gene Set Enrichment Analysis(GSEA)
Network Topology-based Analysis (NTA)

Functional Database (dropdown menu)

ID	uniprotID	T test P value
ENSG000000000003	O43657	-3.96E-01
ENSG000000000003	AOA087WYV6	-3.96E-01
ENSG000000000003	AOA087WZU5	-3.96E-01
ENSG000000000419	O60762	2.16E-01
ENSG000000000419	AOA804HIB3	2.16E-01
ENSG000000000419	AOA804HIK9	2.16E-01
ENSG000000000419	AOA804HJ93	2.16E-01
ENSG000000000419	H0Y368	2.16E-01
ENSG000000000419	Q5QPJ9	2.16E-01
ENSG000000000419	Q5QPK2	2.16E-01
ENSG000000000457	Q8IZE3	1.42E-02
ENSG000000000457	X6RHX1	1.42E-02
ENSG000000000460	Q9NSG2	-1.74E-02
ENSG000000000460	AOA1B0GTJ9	-1.74E-02
ENSG000000000460	AOA1B0GUP7	-1.74E-02
FNSG000000000460	Δ0Δ1B0GV14	-1.74E-02

Gene Set enrichment Analysis

File name: file.rnk

It is a RANKED list format (tab delimited)
Number must be represented with points not commas
No headers

Gene List

Select Gene ID Type ② Gene symbol

Upload Gene List ② Click to upload Reset

OR

A4GALT	-0.838394204314119
AAAS	2.7864702280181
AACS	0.828274229038452
AADAT	-0.025119056689585
AAED1	0.05011116125720201



ID	uniprotID	T test	P value
ENSG000000000003	O43657	-3.96E-01	
ENSG000000000003	A0A087WYV6	-3.96E-01	
ENSG000000000003	A0A087WZU5	-3.96E-01	
ENSG000000000419	O60762	2.16E-01	
ENSG000000000419	A0A804HIB3	2.16E-01	
ENSG000000000419	A0A804HIK9	2.16E-01	
ENSG000000000419	A0A804HJ93	2.16E-01	
ENSG000000000419	H0Y368	2.16E-01	
ENSG000000000419	Q5QPJ9	2.16E-01	
ENSG000000000419	Q5QPK2	2.16E-01	
ENSG000000000457	Q8IZE3	1.42E-02	
ENSG000000000457	X6RHX1	1.42E-02	
ENSG000000000460	Q9NSG2	-1.74E-02	
ENSG000000000460	A0A1B0GTJ9	-1.74E-02	
ENSG000000000460	A0A1B0GUP7	-1.74E-02	
ENSG000000000460	A0A1B0GV14	-1.74E-02	

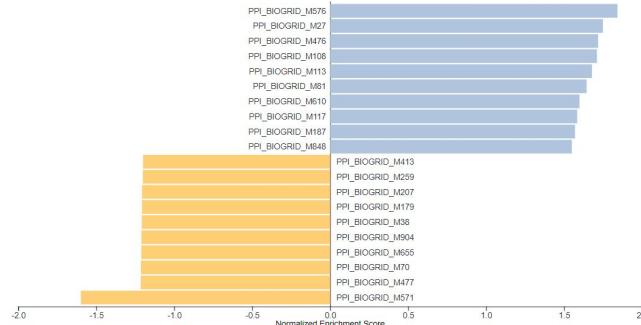
Supplementary [video](#) about graph interpretation

Enrichment Results

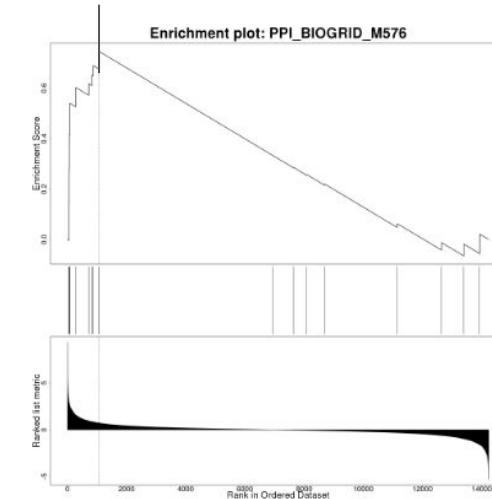
Redundancy reduction: All Affinity propagation Weighted set cover

[Table](#) [Bar chart](#) [Volcano plot](#) [DAG](#)

FDR ≤ 0.05 FDR > 0.05



Enrichment Score

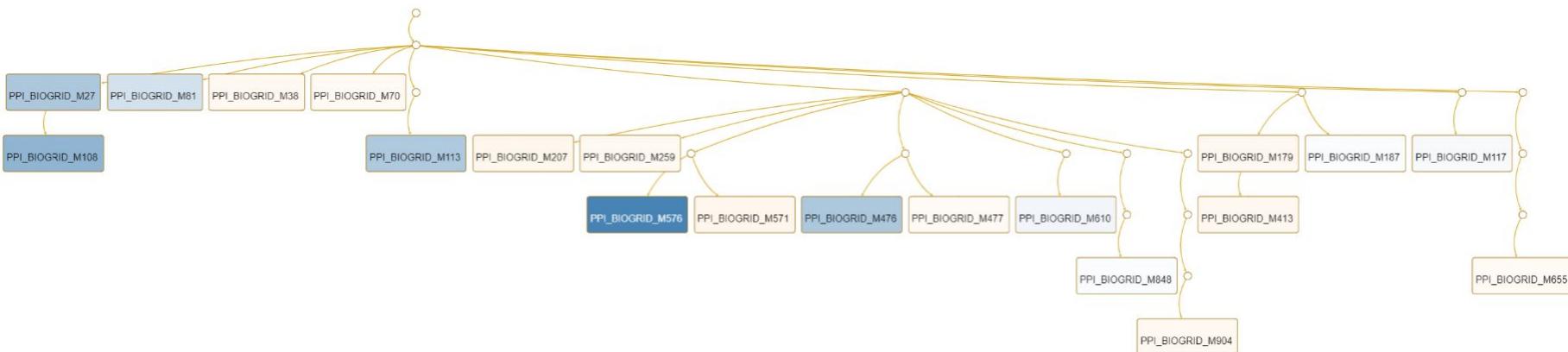


User ID	↓ Score	Gene Symbol	Gene Name	Entrez Gene ID
Q15327	3.03	ANKRD1	ankyrin repeat domain 1	27063
Q969Q1	2.92	TRIM63	tripartite motif containing 63	84676
Q9NPC6	2.8	MYOZ2	myozenin 2	51778
O76041	2.59	NEBL	nebulette	10529
Q9NZU5	1.59	LMCD1	LIM and cysteine rich domains 1	29995
P35609	0.951	ACTN2	actinin alpha 2	88
O15273	0.869	TCAP	titin-cap	8557
Q9NP98	0.844	MYOZ1	myozenin 1	58529
Q59H18	0.728	TNNI3K	TNNI3 interacting kinase	51086
O75112	0.726	LDB3	LIM domain binding 3	11155

Table Bar chart Volcano plot DAG

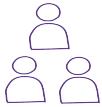
network_PPI_BIOGRID ▾

Compact mode: On





Activity: Gene-set enrichment



Get your experiment Expression and check for Gene enrichment using WebGestalt.

Or use the examples for

[Plant](#) and [Human](#).





Activity:

Use the list of genes in the **EXAMPLE** dataset and make the **Enrichment analysis** (GO annotation) using **at least 3** of the introduced tools. (Create a folder for your project so you can **save the results** of the 3 analysis.)

COMPARE them, are the results consistent?

(Run individually and discuss observations with your college (2 to 3 people))

Get a set of interesting proteins and check the topology in **STRING**.

At last, check for gene-set enrichment using **WebGestalt**.



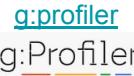
[ToppGene](#)



[WebGestalt](#)



[Plaza](#)



[g:Profiler](#)



[GeneOntology](#)
GENEONTOLOGY
Unifying Biology



[String](#)
STRING