

Ontology... from scratch !

Claire Toffano-Nioche & Audrey Onfroy

Introduction

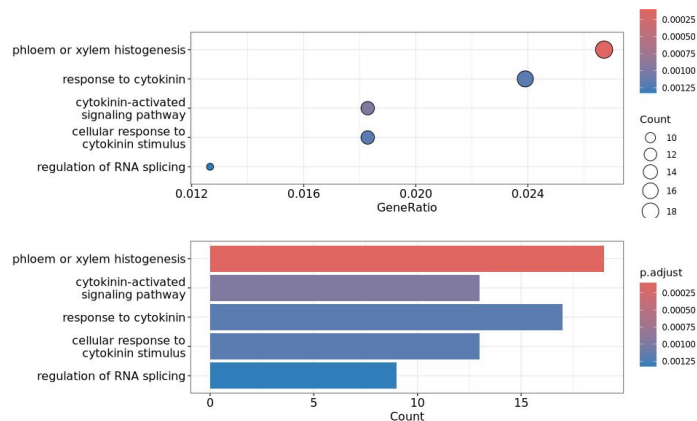
Introduction... is the conclusion of the previous session

Over-Representation Analysis (ORA)

Input: database + a custom gene set

- up OR down-regulated genes (filtered !)
- gene identifiers only

(possible) **Output:**

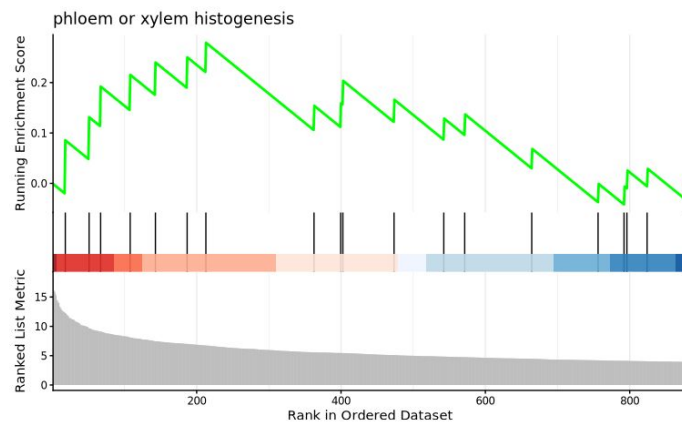


Gene Set Enrichment Analysis (GSEA)

Input: database + a custom **ranked** gene set

- up AND down-regulated genes (all !)
- gene identifiers and gene weights

(possible) **Output:**



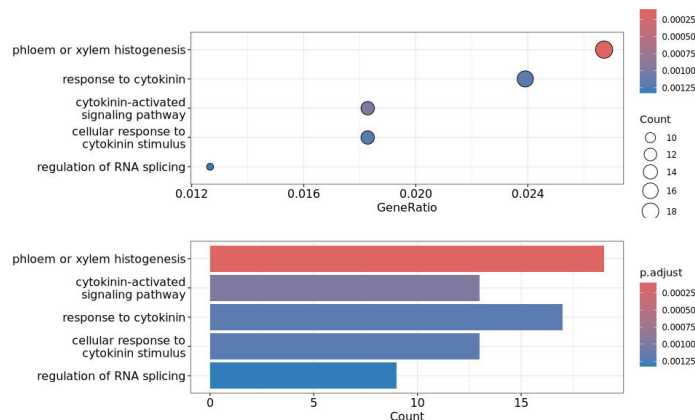
Introduction... is the conclusion of the previous session

Over-Representation Analysis (ORA)

Input: database + a custom gene set

- up OR down-regulated genes (filtered !)
- gene identifiers only

(possible) Output:

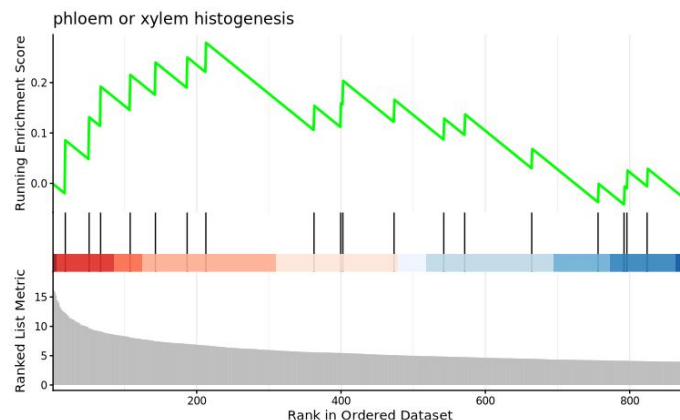


Gene Set Enrichment Analysis (GSEA)

Input: database + a custom **ranked** gene set

- up AND down-regulated genes (all !)
- gene identifiers and gene weights

(possible) Output:



EB3I n1 in 2025...



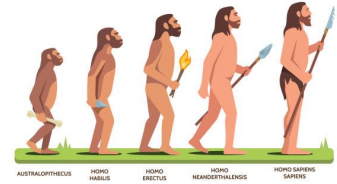
Rattus norvegicus



Solanum lycopersicum



Diplodus vulgaris



Homo sapiens



Oryzias latipes



Arabidopsis thaliana



Xenopus laevis



Danio rerio



Loripes orbiculatus



Mus musculus

How to search for an existing database ?

Several solutions:

- centralized databases:
 - KEGG <https://www.genome.jp/kegg/>
 - Reactome <https://reactome.org/>
 - WikiPathways <https://www.wikipathways.org/>
 - GeneOntology geneontology.org
 - ...
- species-specific websites:
 - *Zea mays* maizegdb.org
 - *Daucus carota* carrotomics.org
 - various plants <https://bioinformatics.psb.ugent.be/plaza/>
- some research articles propose original genome assembly and annotation
- heck https://bioconductor.org/packages/release/BiocViews.html#_Organism



Zea mays



Daucus carota

Missing database ? Build your own !

Make your own annotation

How to build one's own annotation ?

Several ideas:

- transfer the annotation, using orthologs, from a closely related well-annotated organism
- use eggNOG-mapper <http://eggnog-mapper.embl.de/>
- use FANTASIA <https://github.com/CBBIO/FANTASIA>
- use the AnnotationForge R package



10.1101/2024.02.14.580341
10.1101/2024.02.28.582465

How to build one's own annotation ?

Several ideas:

- transfer the annotation, using orthologs, from a closely related well-annotated organism
- use eggNOG-mapper <http://eggnog-mapper.embl.de/>
- use FANTASIA <https://github.com/CBBIO/FANTASIA>
- use the AnnotationForge R package



10.1101/2024.02.14.580341
10.1101/2024.02.28.582465

How to use the AnnotationForge R package ?

Check the documentation (for `makeOrgPackageFromNCBI` or `makeOrgPackage` functions):

<https://bioconductor.org/packages/release/bioc/html/AnnotationForge.html>

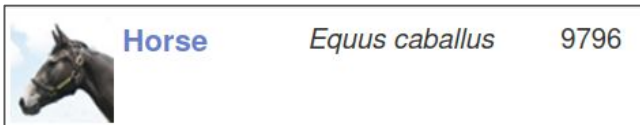
Install the AnnotationForge R package:

```
if (!require("BiocManager", quietly = TRUE))  
  install.packages("BiocManager")  
  
BiocManager::install("AnnotationForge")
```

And a required dependency which is not automatically installed:

```
BiocManager::install("biomaRt")
```

Input parameter settings



The most important parameter is **tax_id**:

```
out_dir = "/meaningful/output/directory/"
genus = "Equus"           # can be whatever you want
species = "caballus"      # can be whatever you want
tax_id = "9796"           # must exist
```



Vertebrates:

<https://www.ensembl.org/info/about/species.html>

Metazoa:

<https://metazoa.ensembl.org/info/about/species.html>

Fungi:

<https://fungi.ensembl.org/info/about/species.html>

Plants:

<https://plants.ensembl.org/info/about/species.html>

Bacteria and Archeae:

<https://bacteria.ensembl.org/info/about/species.html>

Protists:

<https://protists.ensembl.org/info/about/species.html>

Find the taxon ID

Species List

metazoa.ensembl.org/info/about/species.html

EnsemblMetazoa

BLAST | BioMart | Tools | Downloads | Help & Docs | Blog

Search Ensembl Metazoa...

Using this website | Data access | API & software | About us

In this section

- Species List
- Release Cycle
- Scientific Publications
- Workshops
 - Geek for a Week
- Acknowledgements
- Collaborations
 - AllBio
 - Culicoides sonorensis
 - Infravec2
 - VectorBase
 - WormBase
 - WormBase Parasite
- Contact Us
 - Mailing Lists
- Legal
 - Browser Genome Release Agree
 - Disclaimer
 - Ensembl Image Reuse
 - Ensembl name and logo policy
 - Software Licence
- Code of Conduct
- Job Vacancies
- Projects using Ensembl
- Scientific Advisory Board

All Species

Show All entries

Show columns

Name	Classification	Taxon ID	Assembly	Accession	Variation database	Regulation database	Whole genome alignments	Other alignments	In peptide compara	In pan-taxonomic compara
 Acanthaster planci (Crown-of-thorns starfish)	Echinodermata	133434	OKI_Apl_1.0	GCA_001949145.1	-	-	-	✓	✓	-
 Acromyrmex echinator (Panamanian leaf-cutter ant)	Hymenoptera	103372	Aech_3.9	GCA_000204515.1	-	-	-	✓	✓	-
 Acropora millepora (Stony coral, JS-1)	Cnidaria	45264	Amil_v2.1	GCA_013753865.1	-	-	-	✓	✓	-
 Actinia equina (Beadlet anemone, AE1)	Cnidaria	6106	equina_smartden.arrow4.noredun	GCA_011057435.1	-	-	-	✓	✓	-
 Actinia tenebrosa (Australian red waratah sea anemone)	Cnidaria	6105	ASM960242v1	GCA_009602425.1	-	-	-	✓	-	-
 Acyrtosiphon pisum (Pea aphid, AL4f)	Hemiptera	7029	pea_aphid_22Mar2018_4r6ur_v2	GCA_005508785.2	-	-	-	✓	✓	-
 Adelges cooleyi (Spruce gall adelgid, 19-005CV)	Hemiptera	133065	UGA_ACOO_1.1	GCA_023614345.1	-	-	-	✓	✓	-
 Adineta vaga (Rotifer)	Rotifera	104782	AMS_PRJEB1171_v1	GCA_000513175.1	-	-	-	✓	✓	-
 Aedes aegypti (Yellow fever mosquito, LVP_AGWG)	Diptera	7159	AaegL5	GCA_002204515.1	✓	-	✓	✓	✓	✓
 Aedes albopictus - (Asian tiger mosquito, Foshan)	Diptera	7160	GCA035046485v1	GCA_035046485.1	-	-	-	✓	-	-

Create your own org.Ecaballus.eg.db package

The function extracts all the information from NCBI and builds a (local) package:

```
AnnotationForge::makeOrgPackageFromNCBI(  
  version = "0.0.1",  
  author = "Some One <so@someplace.org>",  
  maintainer = "Some One <so@someplace.org>",  
  outputDir = out_dir,  
  NCBIFilesDir = out_dir,  
  tax_id = tax_id,  
  genus = genus,  
  species = species,  
  ensemblVersion = 115,          # better from traceability  
  # Default additional parameter  
  databaseOnly = FALSE,  
  useDeprecatedStyle = FALSE,  
  rebuildCache = TRUE,  
  verbose = TRUE)
```

Overview of the output directory

Ensembl database

Important size ! but
downloaded only once,
then updated quickly...

Custom package

```
[3.7G] gene2accession.gz
[1.2G] gene2go.gz
[224M] gene2pubmed.gz
[1.9G] gene2refseq.gz
[1.3G] gene_info.gz
[ 72G] NCBI.sqlite
[184M] org.Ecaballus.eg.db
├── [ 473] DESCRIPTION
├── [184M] inst
│   ├── [184M] extdata
│   └── [184M] org.Ecaballus.eg.sqlite
├── [7.9K] man
│   ├── [ 705] org.Ecaballus.egBASE.Rd
│   ├── [2.4K] org.Ecaballus.eg_dbconn.Rd
│   └── [ 845] org.Ecaballus.egORGANISM.Rd
├── [ 382] NAMESPACE
├── [5.3K] R
│   └── [1.3K] zzz.R
```

Install and use the newly made package

First, build the package name so the script stays pseudo-automatic:

```
package_name = paste0("org.", substr(genus, 1, 1), species, ".eg.db")
```

org. Equus caballus .eg.db

└──┘

org.Ecaballus.eg.db

Then, install the package:

```
install.packages(paste0(out_dir, "/", package_name), repos = NULL)
```

The package can be loaded and used:

```
library(package_name, character.only = TRUE)
```

Supplementary Slide

The only code

```
# Input parameter
out_dir = "/meaningful/output/directory/"
genus = "Equus"      # can be whatever you want
species = "caballus" # can be whatever you want
tax_id = "9796"      # must exist

# Create the package from NCBI
AnnotationForge::makeOrgPackageFromNCBI(
  version = "0.0.1",
  author = "Some One <so@someplace.org>",
  maintainer = "Some One <so@someplace.org>",
  outputDir = out_dir,
  NCBIFilesDir = out_dir,
  tax_id = tax_id,
  genus = genus,
  species = species,
  ensemblVersion = 115,      # better from traceability
  # Default additional parameter
  databaseOnly = FALSE,
  useDeprecatedStyle = FALSE,
  rebuildCache = TRUE,
  verbose = TRUE)

# Retrieve package name
package_name = paste0("org.", substr(genus, 1, 1), species, ".eg.db")

# Install the package
install.packages(paste0(out_dir, "/", package_name), repos = NULL)

# Load the package
library(package_name, character.only = TRUE)
```