

# Functional Analysis & Visualizations (based on RNA-seq)

Laboratory Course Bioinformatics 2025

Dr. techn. Veronika Schusterbauer  
Sunday, November 23, 2025

# Functional Analysis

# Functional Profiling

**Summarize results from DGE (up to thousands of deregulated genes) into biologically meaningful functions or pathways.**

Scenario:

*E. coli* exposed to oxidative stress (e.g., hydrogen peroxide).

Functional profiling result:

- Upregulated GO terms: "*response to oxidative stress*", "*iron–sulfur cluster binding*", "*antioxidant activity*"
- Pathways enriched: *Glutathione metabolism*, *Superoxide dismutase pathway*

Example

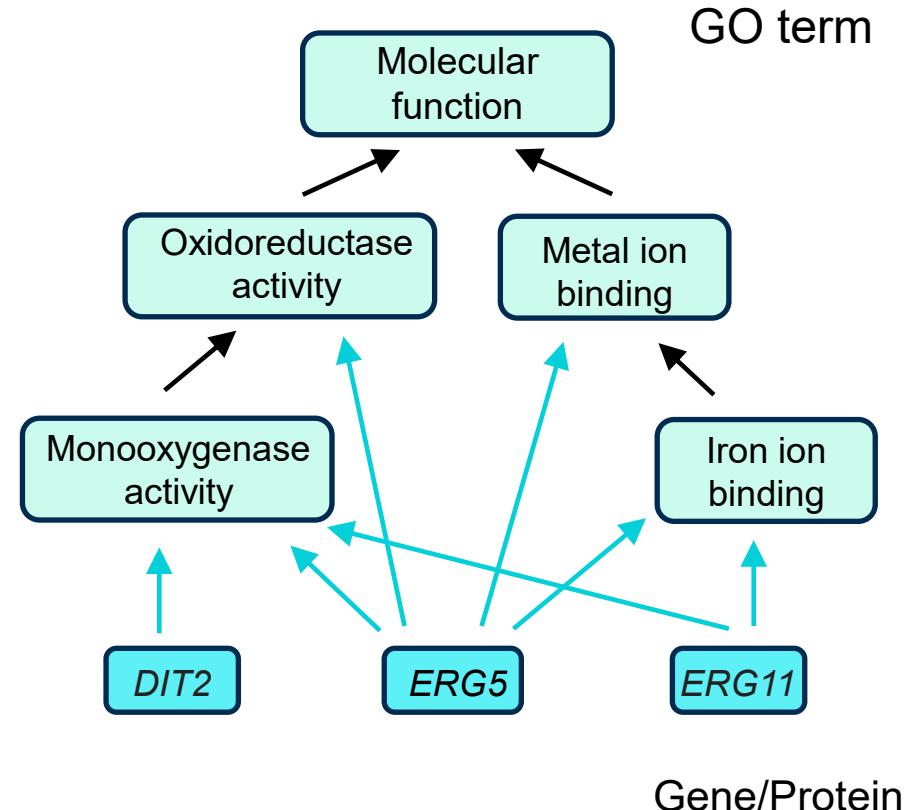
# Functional Profiling

## Input:

- Table/List of (changes in) gene expression  
(Or similar things: e.g. Protein levels)
- Functional Annotations of genes/proteins
  - Connections of genes/proteins to functional groups or pathways

# Functional Annotations

- Link genes/proteins to functions/pathways (e.g GO terms)
- Many to many relationship
  - One function can be connected to multiple genes
  - One gene can be annotated with multiple functions



# Functional Annotation Example

Keytypes

„Gene“	EntrezID	...	GO terms
sbcB	944740		<a href="#">GO:0000175</a> , <a href="#">GO:0003676</a> , <a href="#">GO:0008310</a> , <a href="#">GO:0004529</a> , <a href="#">GO:0006281</a> , <a href="#">GO:0003677</a>
puuA	944741		<a href="#">GO:0016879</a> , <a href="#">GO:0003824</a> , <a href="#">GO:0004356</a>
frlD	944742		<a href="#">GO:0016301</a> , <a href="#">GO:0016740</a> , <a href="#">GO:0016301</a> , <a href="#">GO:0016310</a> , <a href="#">GO:0016772</a>
gsiB	944743		<a href="#">GO:0043190</a> , <a href="#">GO:0055085</a> , <a href="#">GO:0042597</a> , <a href="#">GO:0030288</a>
puuE	944744		<a href="#">GO:0008483</a> , <a href="#">GO:0030170</a> , <a href="#">GO:0009448</a> , <a href="#">GO:0003867</a>

Keys  
(or Identifier)

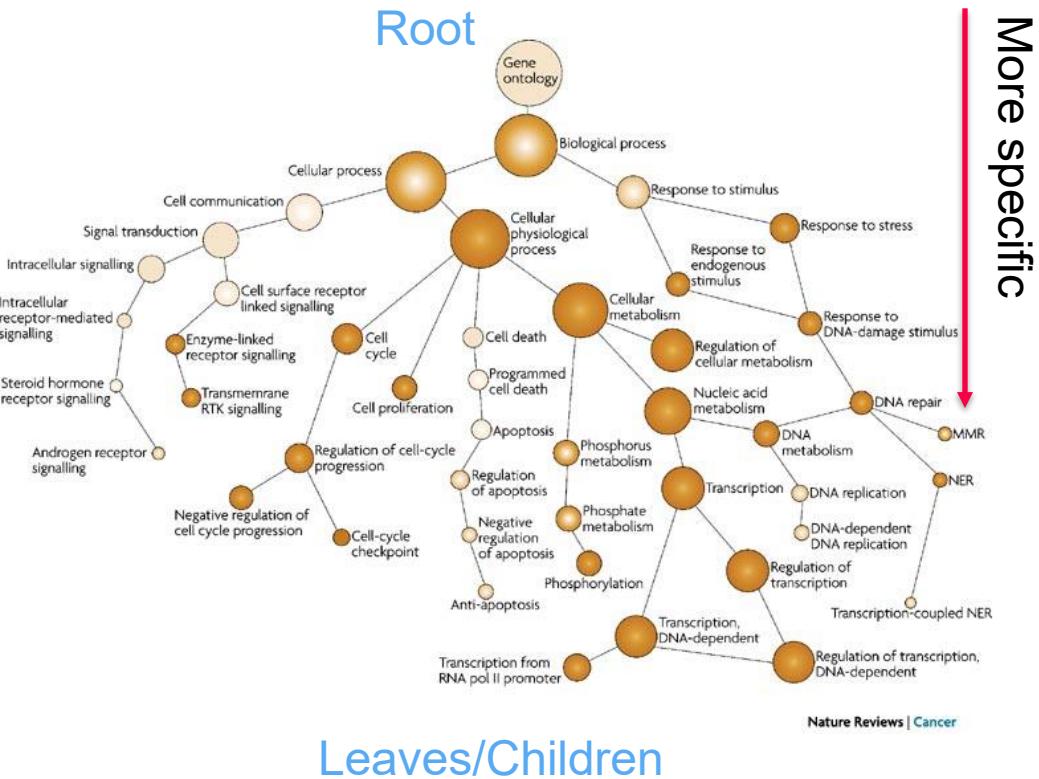
# Gene Ontology (GO)

An ontology is a standardized vocabulary for a scientific field

Three ontologies:

- **biological processes (BP)**
- cell components (CC)
- molecular functions (MF)

Gene Ontologies are organized as tree-shaped graphs



# Sources for GO-Terms (Databases):

EMBL-EBI QuickGO:

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

AMIGO:

<https://amigo.geneontology.org/amigo>

Directly in R as Annotation Data/OrgDB package:

<https://bioconductor.org/packages/3.20/data/annotation/>

Species specific databases:

*E. coli*: <https://biocyc.org/ECOLI/class-tree?object=Gene-Ontology-Terms>

*S. cerevisiae*: <https://www.yeastgenome.org/goTermFinder>

# Pathway databases

- Kyoto Encyclopedia of Genes and Genomes (KEGG)
  - Knowledge base of metabolism and other cellular processes
  - Maps of metabolism annotated with additional information
  - <https://www.genome.jp/kegg/pathway.html>
- Reactome Database
  - Open-source, open access, manually curated and peer-reviewed pathway database
  - <https://reactome.org/>

# Overrepresentation Analysis (ORA)

Are **genes annotated with pre-defined terms** (e.g. having a specific GO function) present more than would be expected (**over-represented**) in a **subset of your data (genes)**?

Common subsets: up-/ down regulated genes

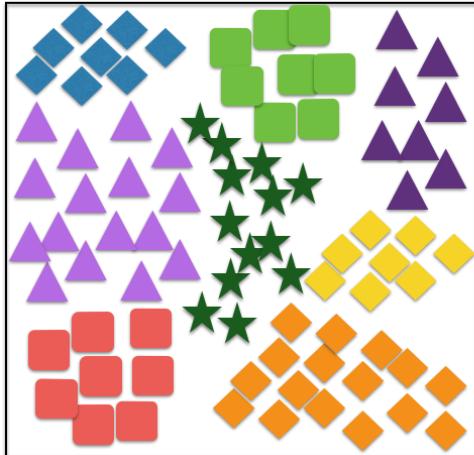
- Fruit fly transcriptome has about **10,000 genes**
- **260** genes are annotated with GO term “axon guidance”
- In our set of 500 upregulated genes **150** are annotated with “axon guidance”

is that significant?

Example

# Overrepresentation Analysis (ORA)

All known genes in a species  
(categorized into groups)



DEGs

## Measures:

- Statistical significance
  - P-value
  - adjusted p-value (or q-value)
- GeneRatio:
  - # colored tags in subset (DEGs) / # colored tags in total
  - Yellow: 3/8
  - Fly example: 150/260

# Visualizations

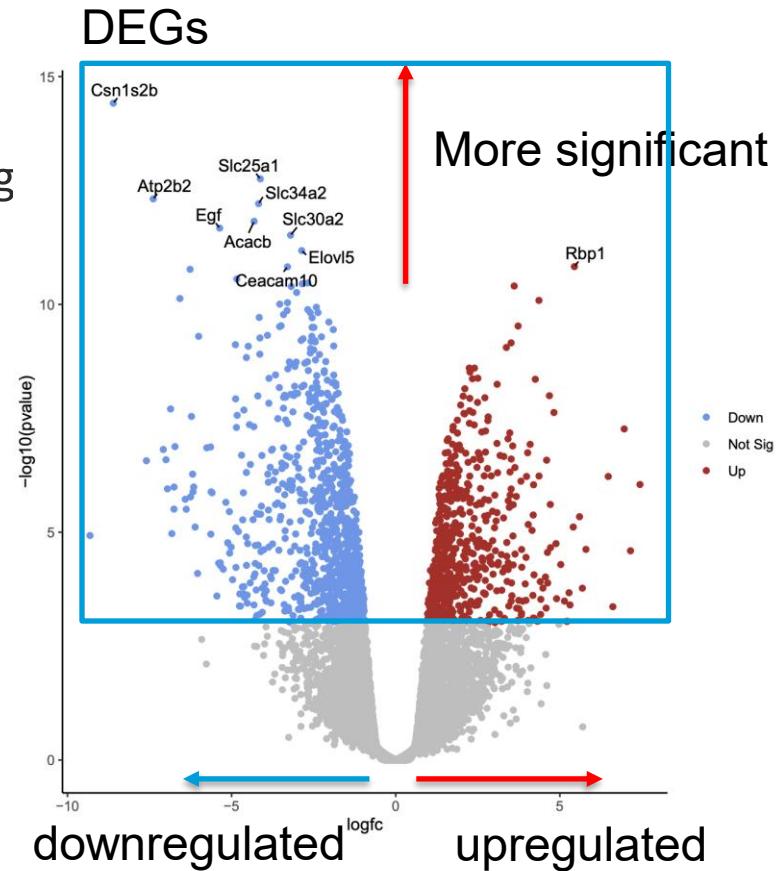
# Volcano Plot

**Statistical significance** (adjusted p-value, on log scale)  
vs **magnitude of change** (log fold change)

Quick identification of most up/downregulated genes

See trends like:

- Levels of expression/log fold change (LFC)
- More up/down regulation
- Levels of significance



# DotPlot

## Overview of most overrepresented functions

By default sorted by GeneRatio

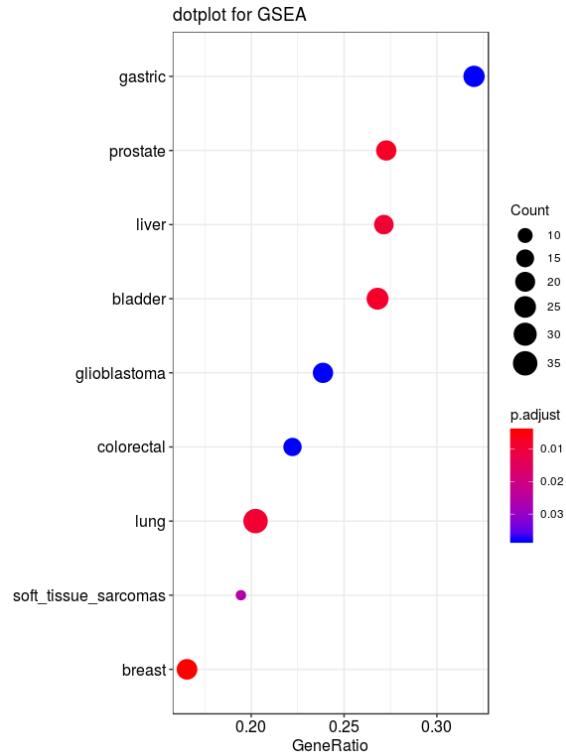
- Possible alternatives:
  - Adjusted p-value
  - Count

Dot size:

Count = # DE genes within group

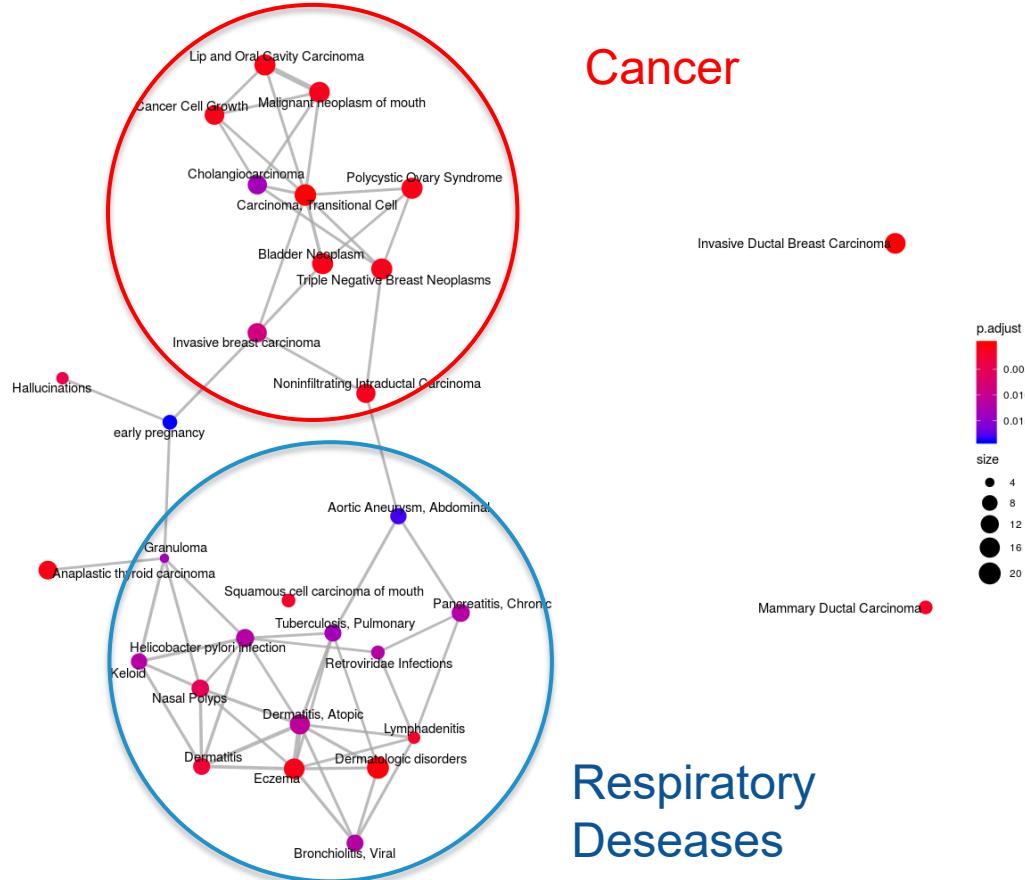
Dot color:

Adjusted p-value



# EMapPlot

- Connects GO terms based on their similarity
- Similarity calculated based on GO structure
- Makes **groups/clusters** of similar GO terms visible



# Report Submission

Hand in via TeachCenter:

- MOL923UF Laboratory Course Bioinformatics  
(<https://tc.tugraz.at/main/course/view.php?id=4192>)
- Choose one person per group to hand in all 3 reports
- All 3 Assignments are found in the Report Section
  - Here you can also find Report Instructions
- Deadline for all 3 reports is **14<sup>th</sup> of December** (Groups 1+2), or **4<sup>th</sup> of January** (Groups 3+4)

# Report Submission for RNA-seq

- Answers to questions in the Lab instructions (Lab09 & Lab10)
- Answer the questions in **full sentences**.
- Use the **numbering** that is provided in the lab instructions.
- Add **relevant plots and tables** and describe how you produced these results.
- **Plots and tables need to have a capture** describing what is presented.
  - E.g.: Figure 1: Dotplot of the ORA analysis, showing the 10 most significantly overrepresented biological processes within downregulated genes

---

## Other announcements:

We are looking for Student assistants for next year..

... also I encourage everybody to evaluate the course...

... the more specific your comments the better ...