



**HAL**  
open science

## **Askor, an R package for easy RNA-Seq data analysis.**

Susete Alves Carvalho, Kévin Gazengel, Anthony Bretaudeau, Stéphanie Robin, Stéphanie Daval, Fabrice Legeai

### ► **To cite this version:**

Susete Alves Carvalho, Kévin Gazengel, Anthony Bretaudeau, Stéphanie Robin, Stéphanie Daval, et al.. Askor, an R package for easy RNA-Seq data analysis.. Journées Ouvertes en Biologie, Informatique et Mathématiques (Jobim), Jun 2023, Tours, France. pp.263, 2018, JOBIM2023. 10.1080/02664763.2018.1454894 . hal-04720296

**HAL Id: hal-04720296**

**<https://hal.inrae.fr/hal-04720296v1>**

Submitted on 3 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Susete Alves Carvalho\*  
Kévin Gazengel\*  
Sylvain Masanelli  
Anthony Bretaudeau  
Stéphanie Robin  
Stéphanie Daval  
Fabrice Legeai  
\*Contributed equally

# AskOR

## an R tool for easy RNA-seq data analysis

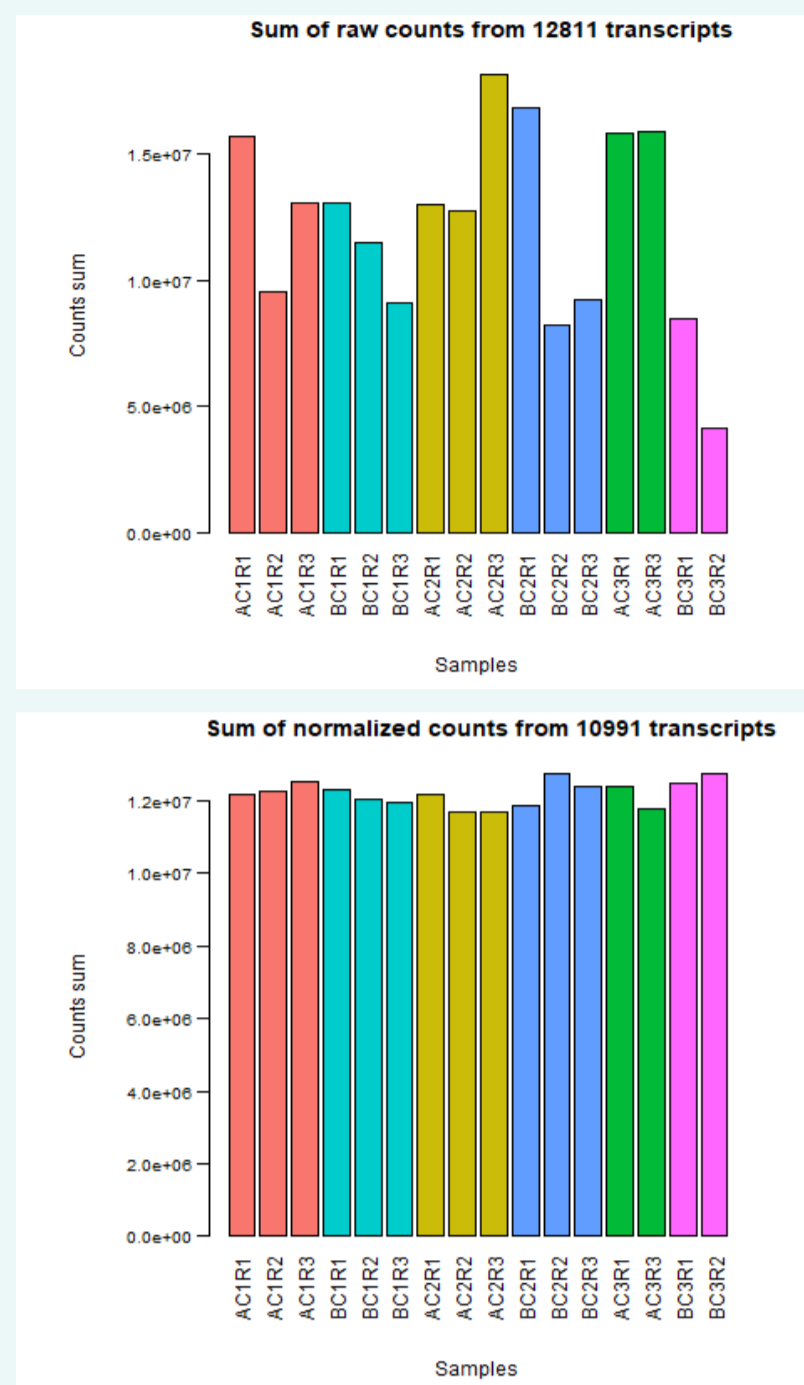
One contrast = comparison of two conditions or two groups of conditions

### Descriptive analysis

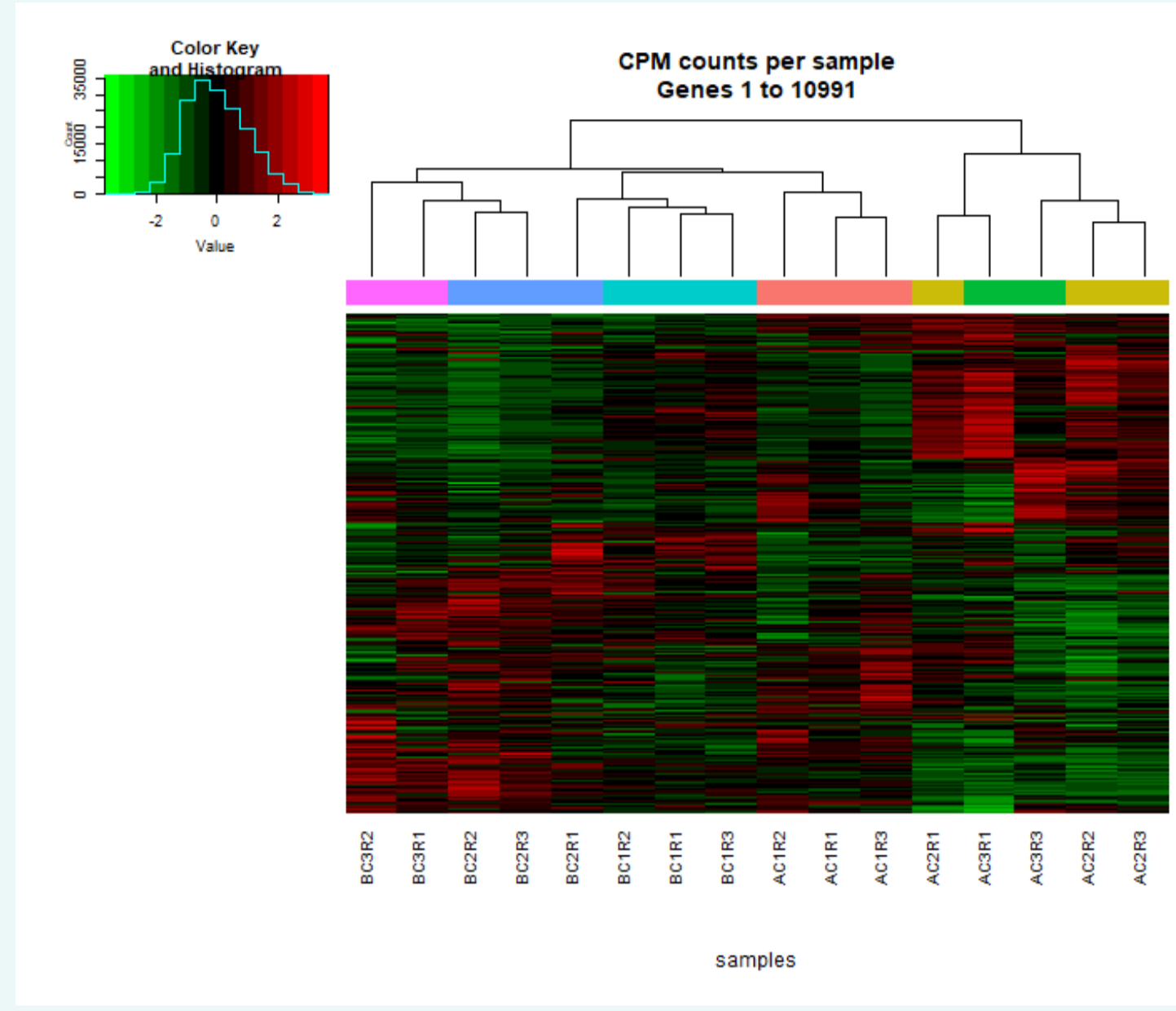
#### Input files :

- RNA-seq count table
- File of experimental design
- File describing contrasts to analyse
- Gene description (optionnal)
- GO-term annotations (optionnal)
- KEGG annotations (optionnal)

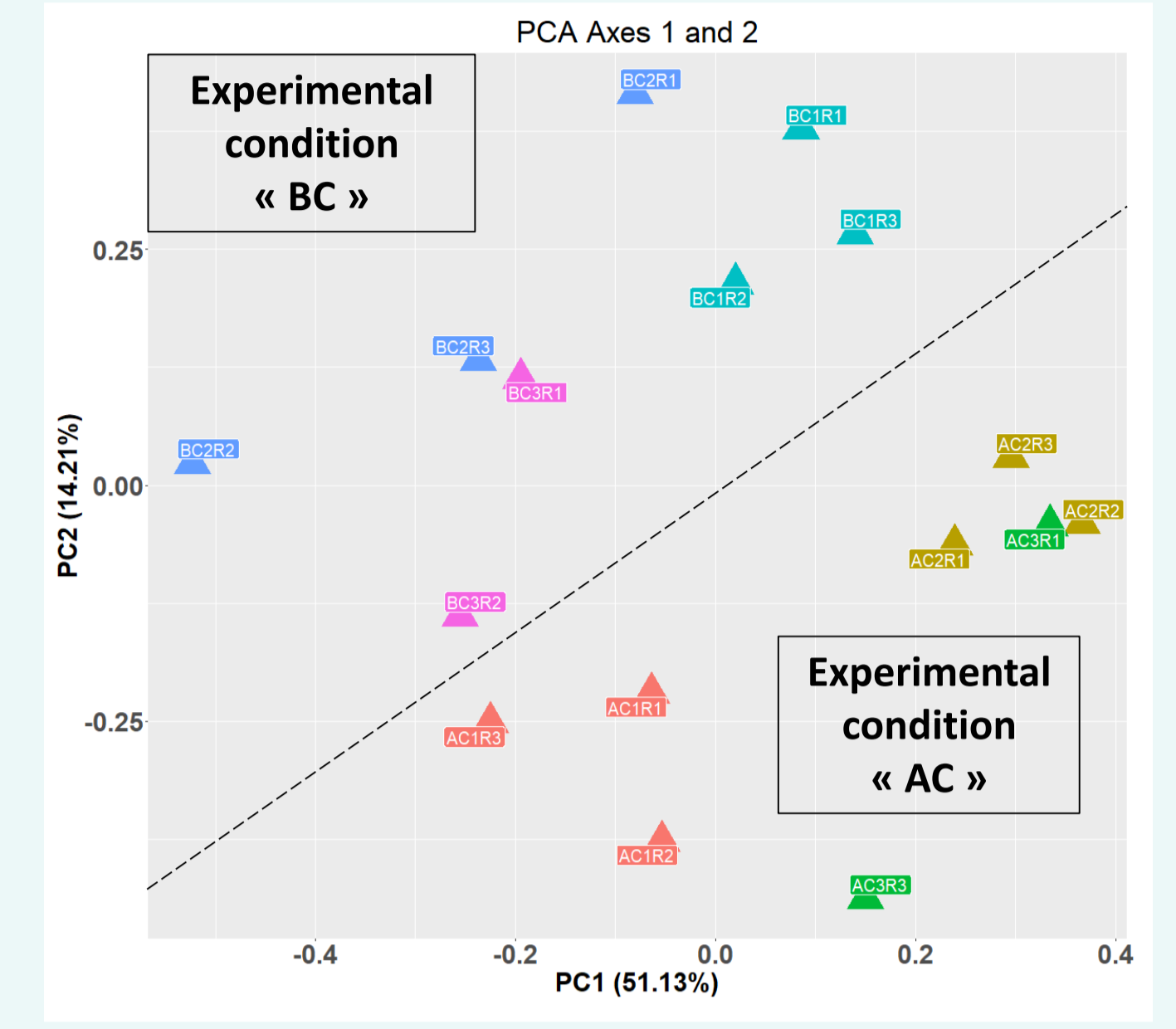
### RAW and NORMALIZED counts



### Heatmap of expression (on normalized counts)



### Principal Component Analysis (on normalized counts)

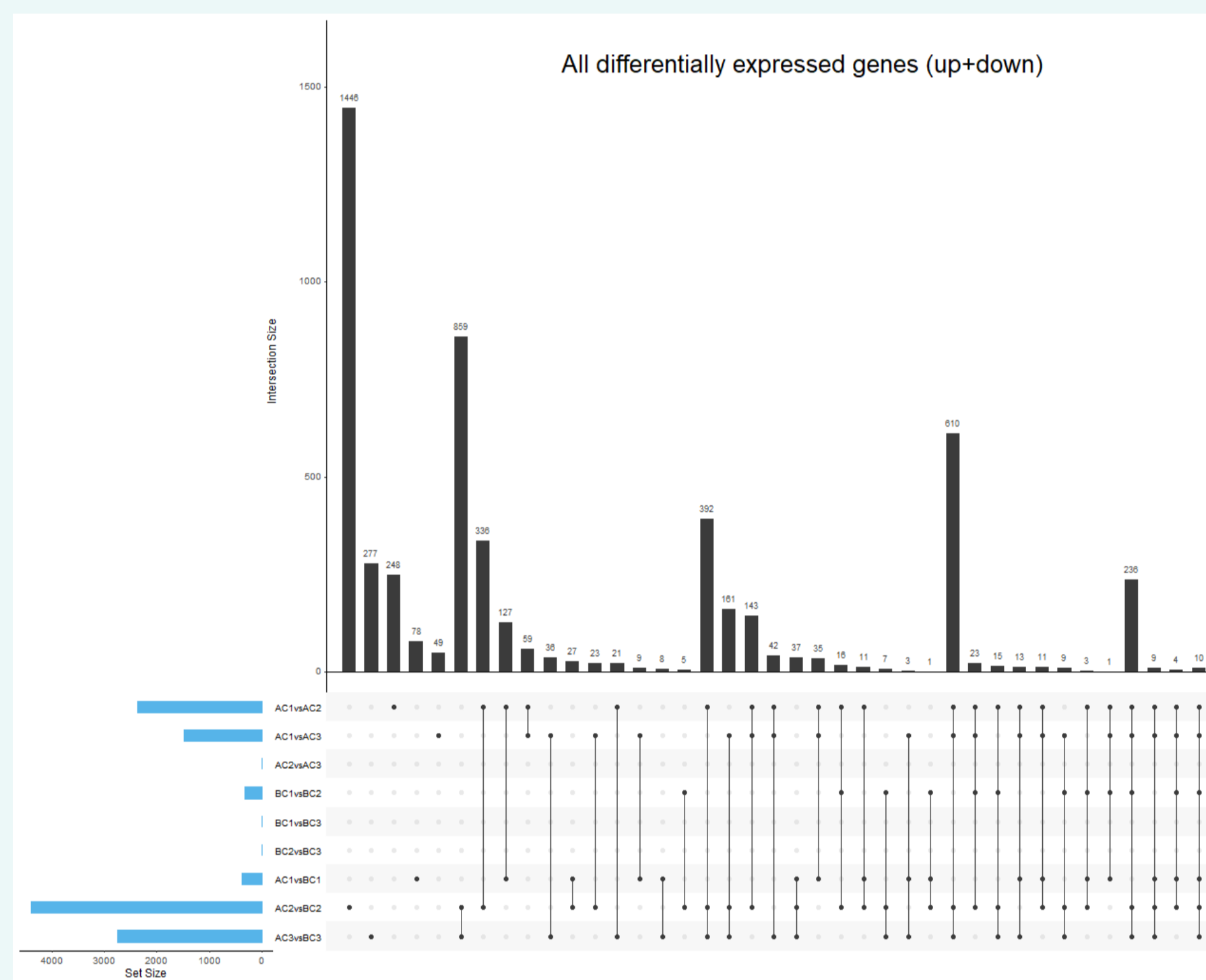


### Differential Expression (DE) analysis

#### Number of DE/non-DE genes

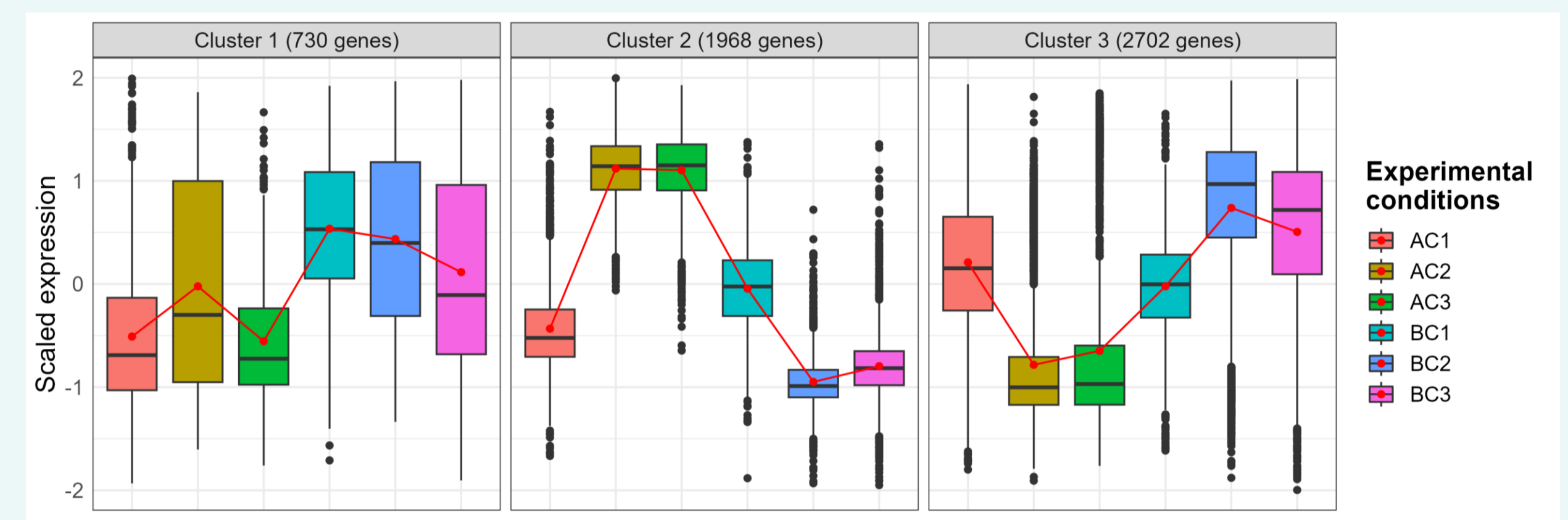
AC1<AC2	AC1=AC2	AC1>AC2	BC2=BC3
1459	8626	906	10991
AC1<AC3	AC1=AC3	AC1>AC3	AC1<BC1
963	9505	523	224
AC2=AC3			AC1=BC1
10991			10604
			AC1>BC1
			163
BC1<BC2	BC1=BC2	BC1>BC2	AC2<BC2
16	10661	314	2271
			AC2=BC2
			6597
			AC2>BC2
			2123
BC1=BC3			AC3<BC3
10991			1283
			AC3=BC3
			8246
			AC3>BC3
			1462

#### UpSet graphs



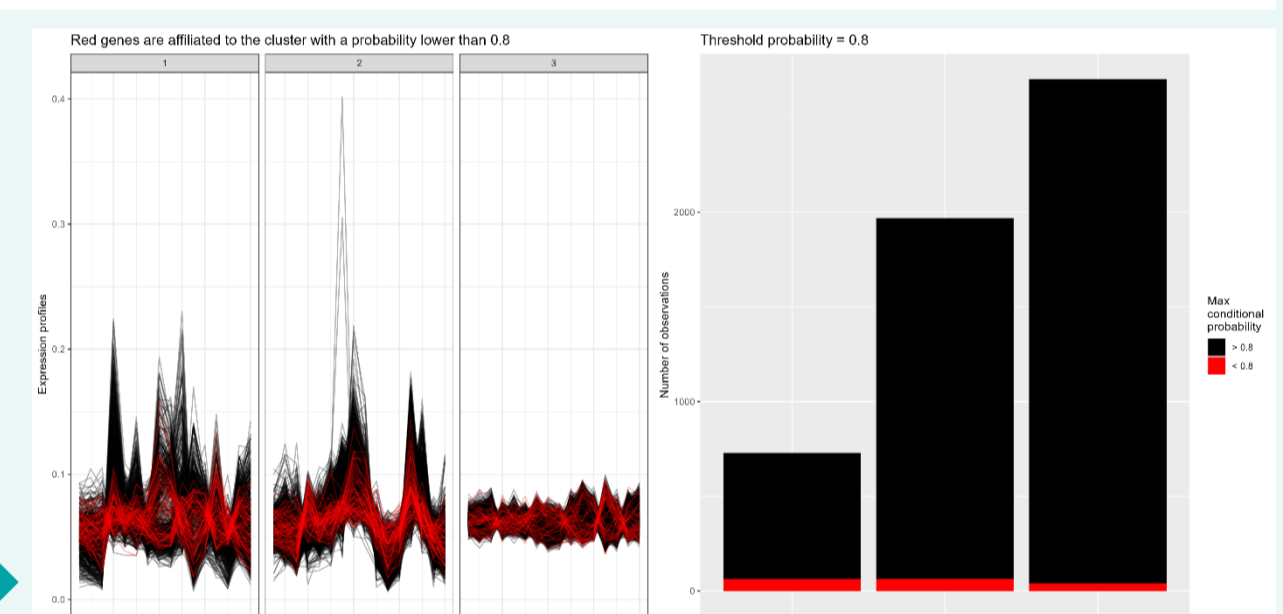
### Co-Expression clustering

- ⇒ On differentially expressed genes (DEG) in at least one contrast
- ⇒ On gene lists of interest



- Per-cluster analysis
- Summary tables

Cluster robustness

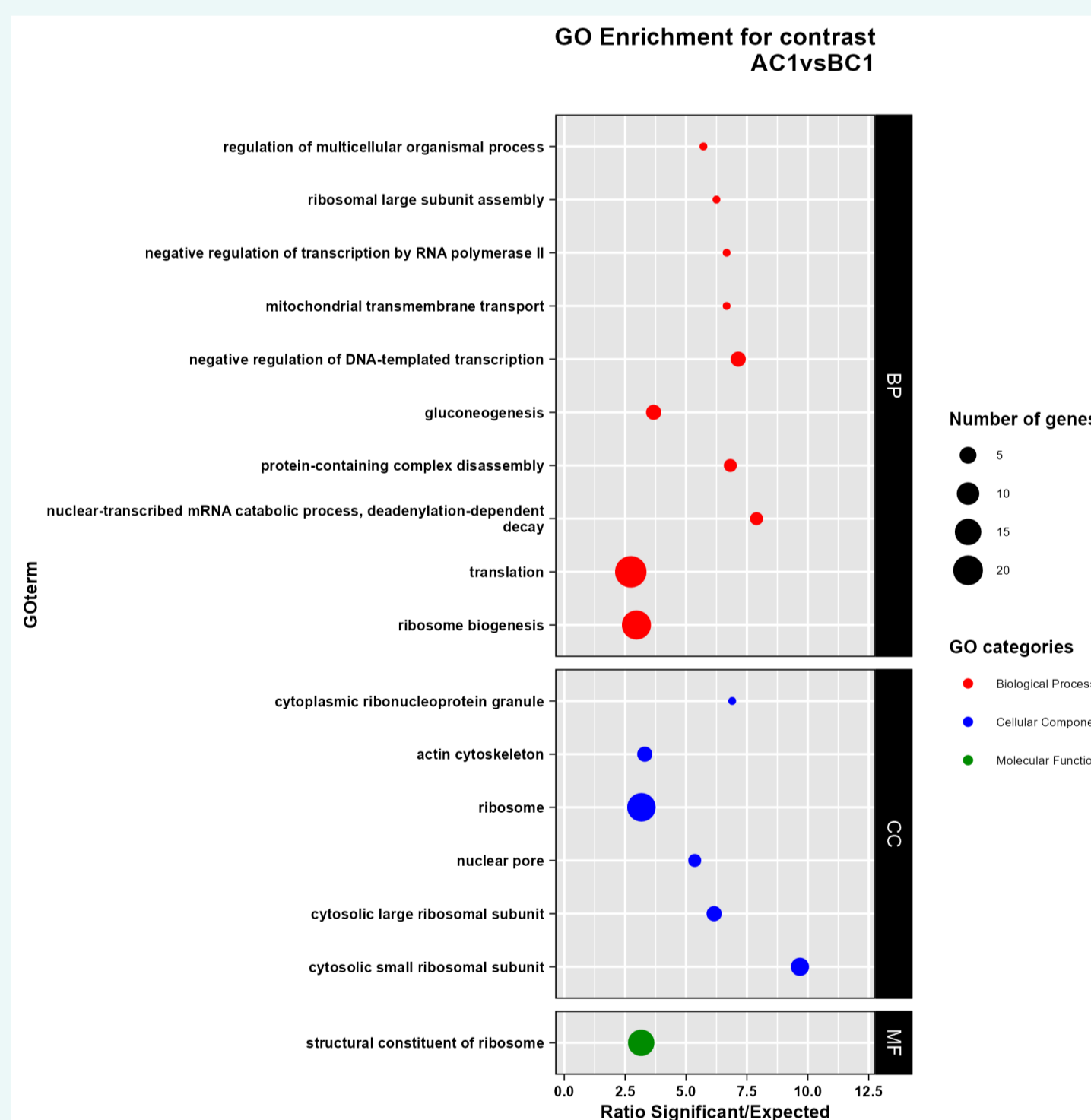


### Further analyses

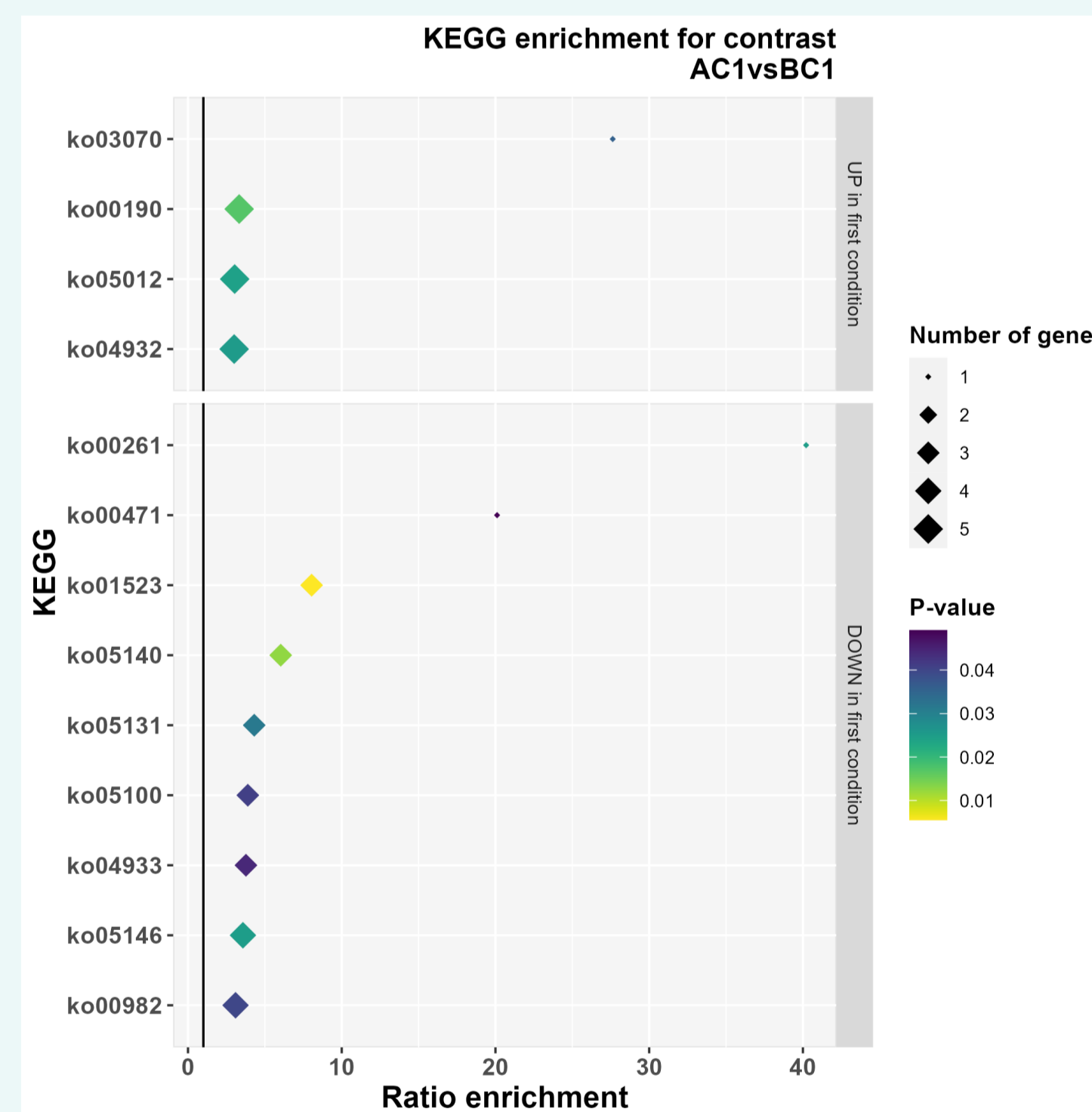
- ⇒ On contrasts
- ⇒ On co-expression clusters
- ⇒ On gene lists of interest

- Summary tables
- Lists of genes for each enriched term

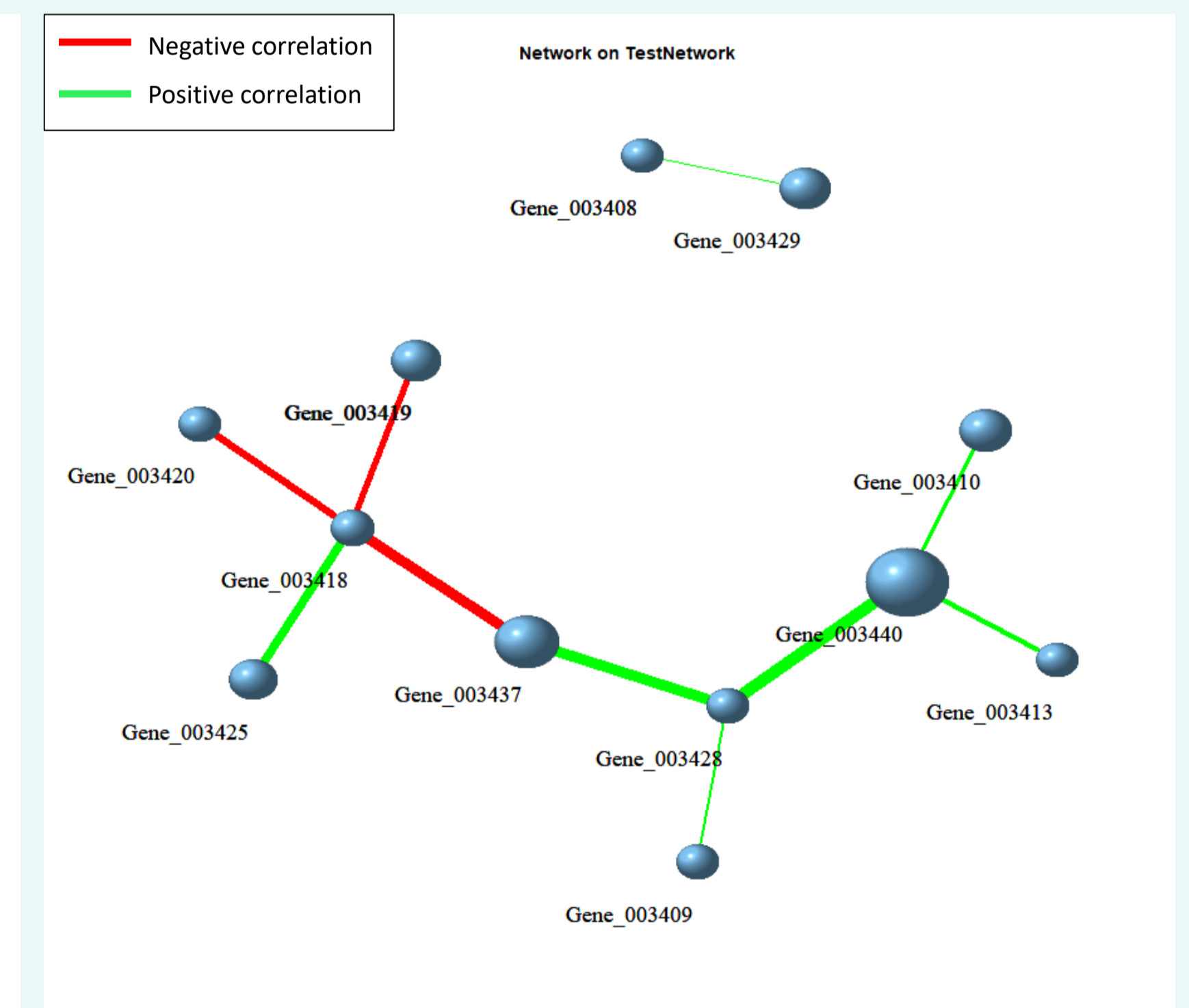
#### GO-term enrichment



#### NEW KEGG enrichment



#### NEW Gene Network analysis



### Take home message

#### AskOR facilitates transcriptomics data analyses

- Automatic and quick analysis generating outputs organized in folders
- Easy to handle in R environment (<https://github.com/askomics/askOR>)

- Large set of parameters and outputs (tables and graphs)
- Preconfigured RStudio instance available on Galaxy (<https://usegalaxy.fr/>)

- Use of edgeR<sup>1</sup>, topGO<sup>2</sup>, coseq<sup>3,4</sup>, and igraph<sup>5</sup> packages for differential expression analyses, GO-term and KEGG enrichments, co-expression clusters identification, and gene network graphs
- Outputs compatible with AskOmics<sup>6</sup> tool for integration of heterogeneous data

- Robinson MD. *et al.* edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140. doi:10.1093/bioinformatics/btp616, 2010.
- Adrian A. and Rahnenfuhrer J. topGO: Enrichment Analysis for Gene Ontology. R package version 2.46.0, 2021.
- Rau A. and Maugis-Rabusseau C. Transformation and model choice for co-expression analysis of RNA-seq data. *Briefings in Bioinformatics*, 19(3), 425-436, 2018.
- Godichon-Baggioni A. *et al.* Clustering transformed compositional data using K-means, with applications in gene expression and bicycle sharing system data. *Journal of Applied Statistics*, doi:10.1080/02664763.2018.1454894, 2018.
- Csardi G. and Nepusz T. The igraph software package for complex network research. *Inter-Journal, Complex Systems*, 1695, 2006
- Garnier X. *et al.* AskOmics: a user-friendly interface to Semantic Web technologies for integrating local datasets with reference resources. *JOBIM, Nantes, France*. pp.1. hal-02401750, 2019.