AskoR, an R package for easy RNA-Seq data analysis illustrated by the analysis of plant/pathogen/microbiote interactions.

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

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

INRAE

L'INSTITUT agro Rennes Angers

UNIVERSITÉ DE RENNES 1

BARIC

GOGEPP

AskoR, an R package for easy RNA-seq data analysis illustrated by the analysis of plant/pathogen/microbiote interactions[1]

Équipe Ecologie et Génétique des interactions **P**lantes-**M**icrobiotes-**B**ioagresseurs - UMR IGEPP

Plateforme GOGEPP - UMR IGEPP

CATI BARIC - INRAE, dpt SPE

Susete Alves Carvalho*
Kévin Gazengel*
Sylvin Masanelli
Anthony Bretaudeau
Stéphanie Robin
Stéphanie Daval
Fabrice Legeai

*Contributed equally

AskoR

- R tool
- Flexible: a large set of parameters for experimented users and default values for novices
- Including edgeR[2], topGO[3] and coseq[4,5] packages for differential expression analyses, GO-term enrichments, and co-expression clusters identification
- Ouputs compatible with AskOmics[6] tool for integration of heterogeneous data
- https://github.com/askomics/askoR

Descriptive

Before filtering low expressed genes
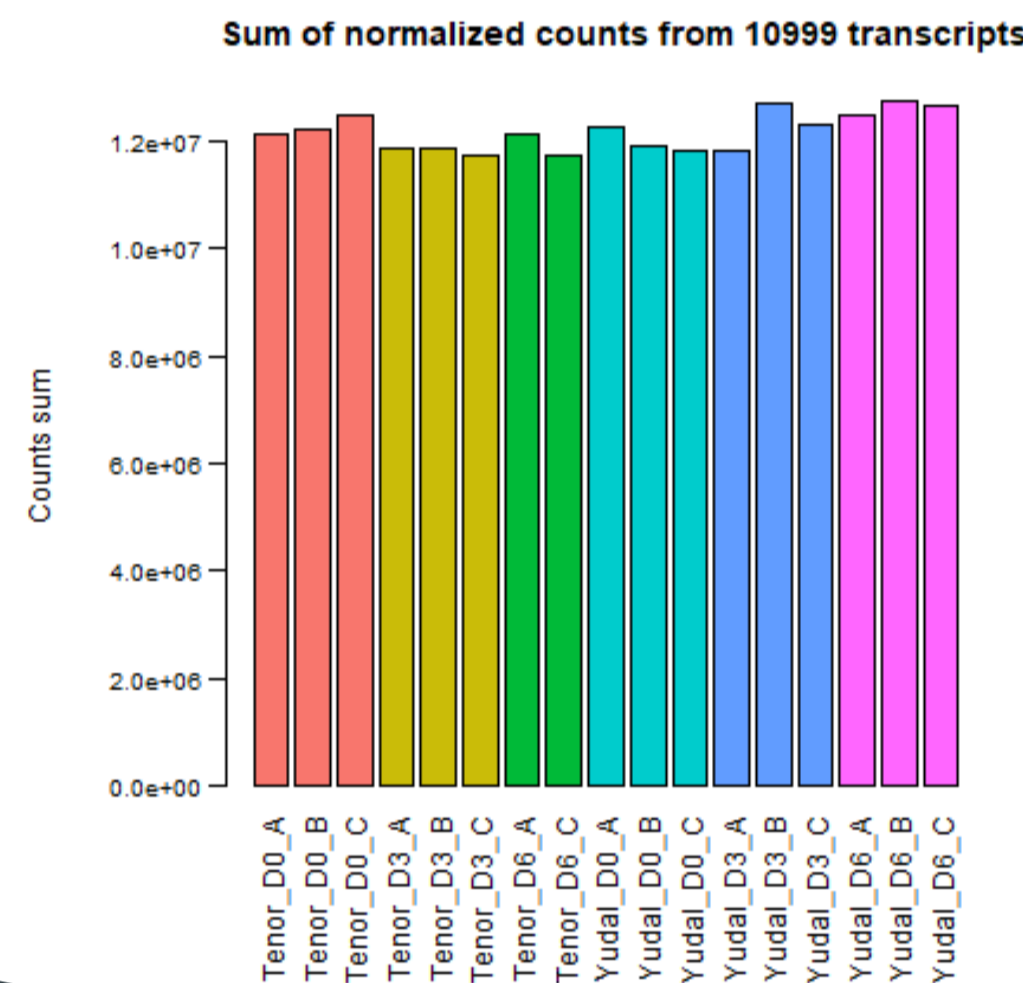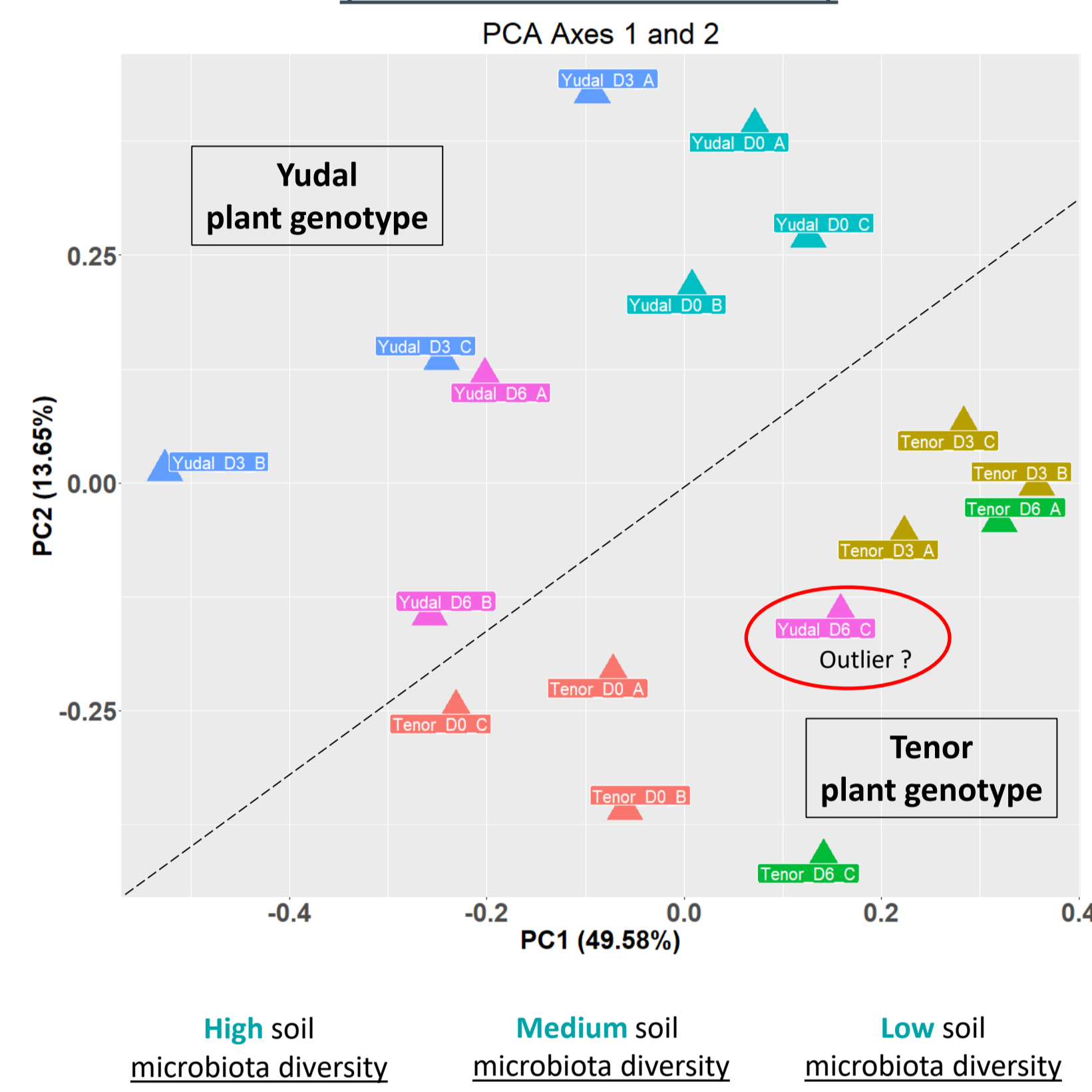Sum of raw counts from 12811 transcripts
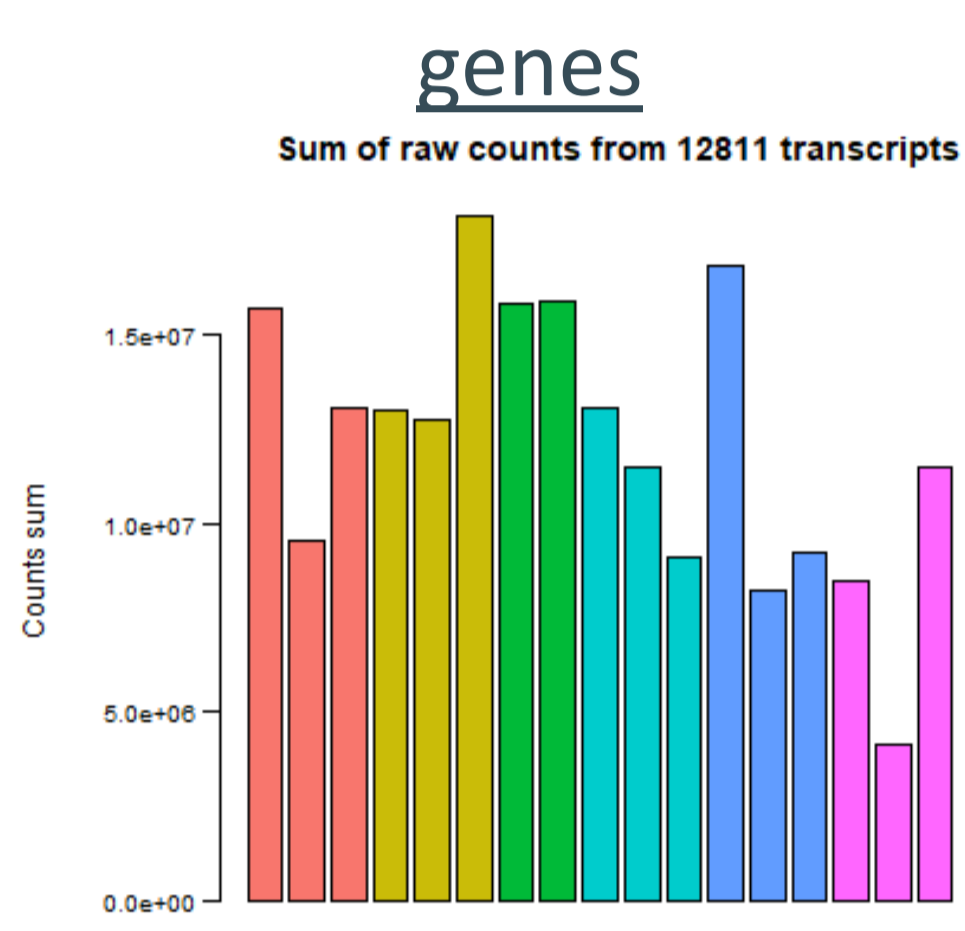
After filtering and normalization
Sum of normalized counts from 10999 transcripts

Differential expression (DE)

Principal Componant Analysis
(on normalized counts)

PCA Axes 1 and 2

Yudal plant genotype

Outlier ?

Tenor plant genotype

High soil microbiota diversity
Tenor_D0   Yudal_D0

Medium soil microbiota diversity
Tenor_D3   Yudal_D3

Low soil microbiota diversity
Tenor_D6   Yudal_D6

Quantitative

UpSet Graph for all contrasts
(alternative to Venn Diagram for more than 4 lists)

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

Intersection Size

Yudal_D0vsYudal_D3
Yudal_D3vsYudal_D6
Yudal_D0vsYudal_D6
Tenor_D0vsTenor_D3
Tenor_D3vsTenor_D6
Tenor_D0vsTenor_D6

Set Size
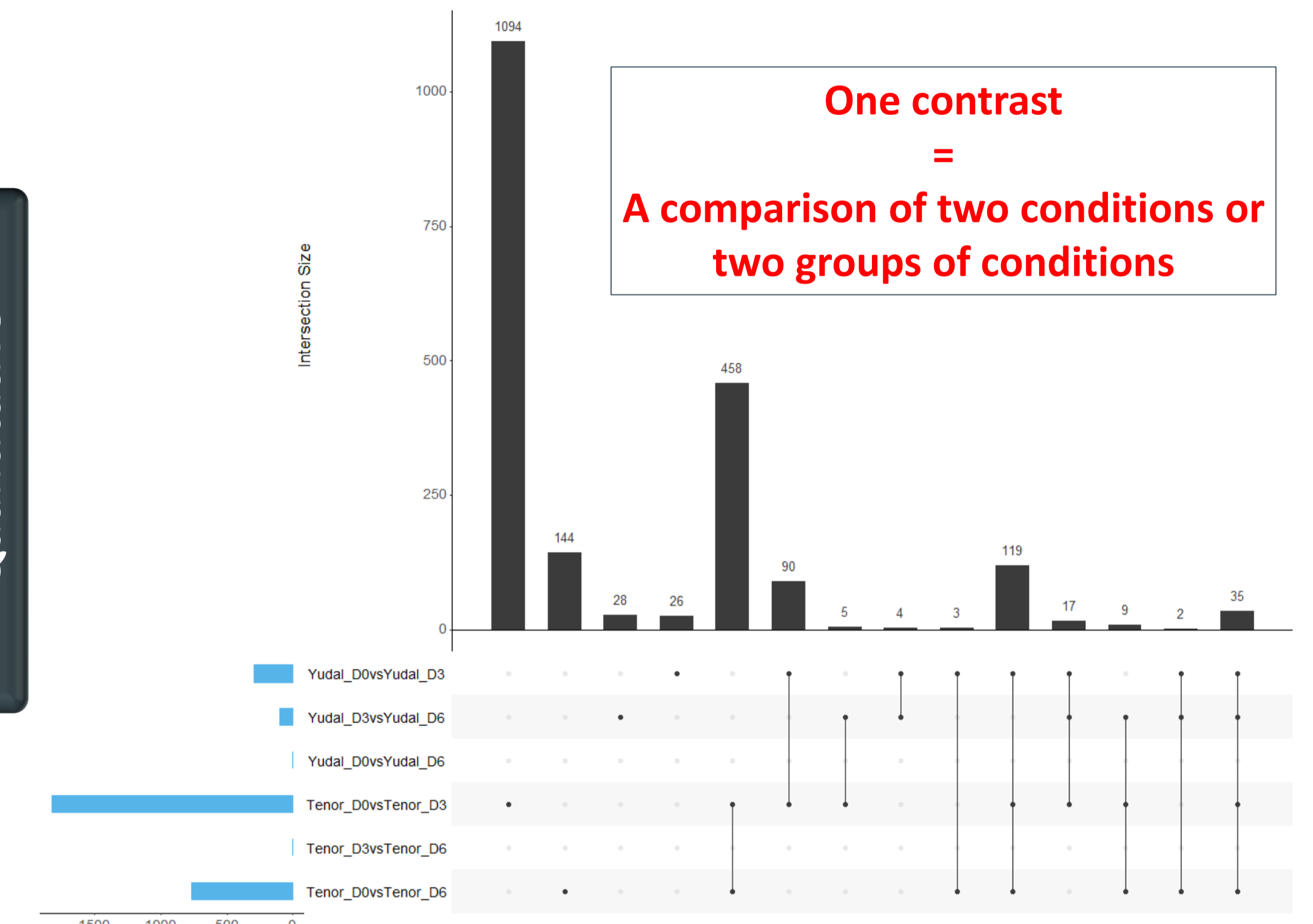
DE tables   AskOmics tables   MA and Vulcano plots

GO-term enrichment

⇒ **On contrasts**
⇒ **On gene lists of interest**

Summary tables

Separated graphs for each GO category

Lists of genes for each enriched GO-term

GO Enrichment for contrast Tenor_D0 < Tenor_D3
(956 annotated genes among 1360 genes)

BP
chitin metabolic process
phosphate ion transmembrane transport
aminophospholipid transport
chitin biosynthetic process
G protein-coupled glutamate receptor signaling pathway
cilium assembly
carbohydrate metabolic process

CC
cilium
myosin complex
integral component of plasma membrane
BBSome
motile cilium
intraciliary transport particle B
ATP-binding cassette (ABC) transporter complex

MF
phosphate ion binding
phospholipid binding
lipid binding
G protein-coupled GABA receptor activity
chitin synthase activity
hydrolase activity, hydrolyzing O-glycosyl compounds
chitin binding

GO categories
● Molecular Function
● Biological Process
● Cellular Component

Number of genes
● 10
● 20
● 30
● 40

Ratio Significant/Expected

➢ **Identify biological pathways :**
   e.g. chitin pathway of the pathogen (related to pathogenicity) reduced with high diversity of microorganisms.

Co-expression clusters

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

Cluster 1 (1099 genes)   Cluster 2 (558 genes)   Cluster 3 (377 genes)

Scaled expression

Experimental conditions
Tenor_D0
Tenor_D3
Tenor_D6
Yudal_D0
Yudal_D3
Yudal_D6
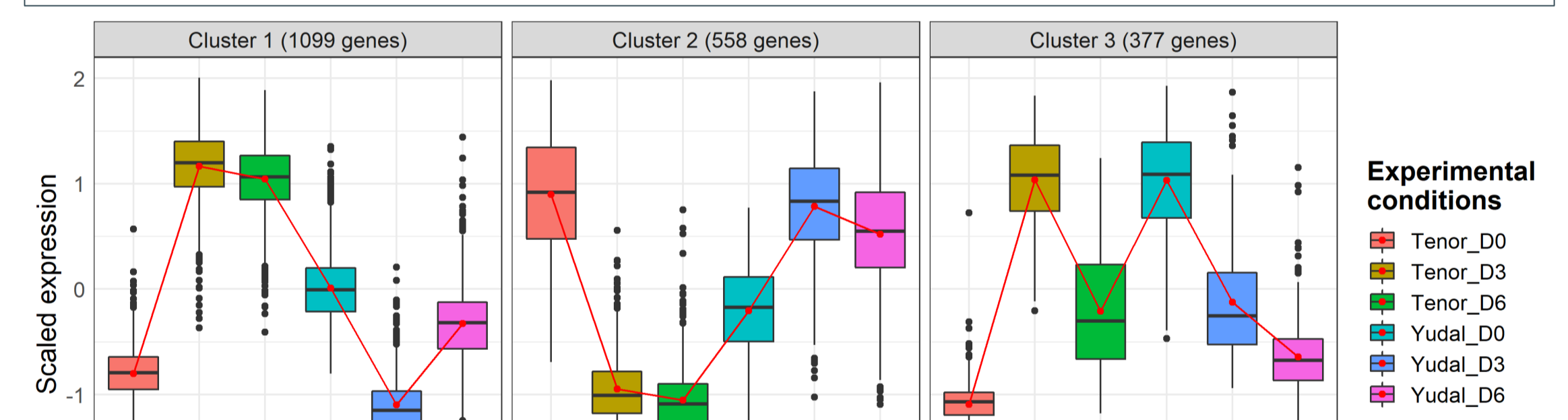
Cluster robustness

Summary tables

Per-cluster analysis

Heatmap

➢ **Define expression profiles between conditions**
➢ **Perform GO-term enrichment on each cluster**

Take home message

➢ **AskoR facilitates transcriptomics data analyses**
   → Automatic and quick analysis generating outputs organized in folders
   → Large set of parameters and outputs (tables and graphs)
   → Easy to handle in R environment (**https://github.com/askomics/askoR**)
   → Preconfigured RStudio instance available on Galaxy (**https://usegalaxy.fr/**)
   → Adaptable to many research questions from RNA-seq data but possible extrapolation to other count tables (SmallRNA-seq, metagenomics, …)

➢ **Perspectives**
   → Publish AskoR on CRAN
   → Integrate KEGG enrichment analysis
   → Integrate gene network analysis
   → Integrate specific outputs for metagenomics

1. Daval S. *et al*. Soil microbiota influences clubroot disease by modulating *Plasmodiophora brassicae* and *Brassica napus* transcriptomes. Microbial biotechnology, *13*(5), 1648–1672. https://doi.org/10.1111/1751-7915.13634, 2020.
2. 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.
3. Adrian A. and Rahnenfuhrer J. topGO: Enrichment Analysis for Gene Ontology. R package version 2.46.0, 2021.
4. 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.
5. 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.
6. 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.

IGEPP