MATRiX app: Mining and functional Analysis of TRanscriptomics data

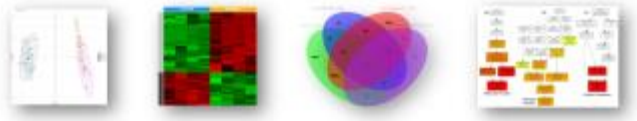
Une application dédiée à l'exploration de résultats d'analyses biostatistiques de données transcriptomiques

http://get.genotoul.fr
get@genotoul.fr
@get_genotoul

Problématique

Des données complexes à explorer et ré-analyser pour les biologistes

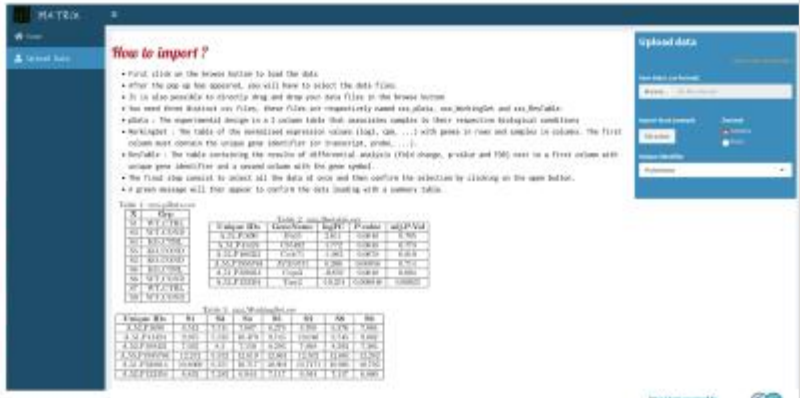
- Critères arbitraires pour la sélection de listee gènes
- Compétences spécifiques pour la génération de graphiques/ analyses complémentaires



➔ Développer une solution pour que les biologistes puissent manipuler leurs jeux de données de manière autonome

3

Utilisation



7

Quelle solution d'hébergement?

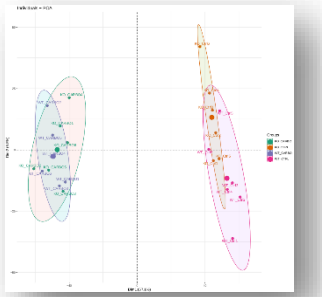
Démo

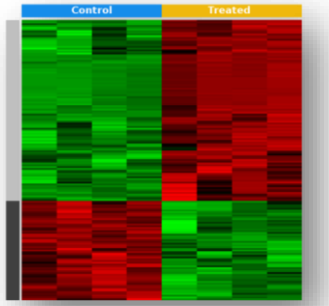
https://matrix.toulouse.inrae.fr

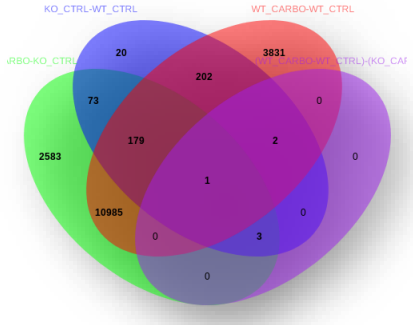
Problématique

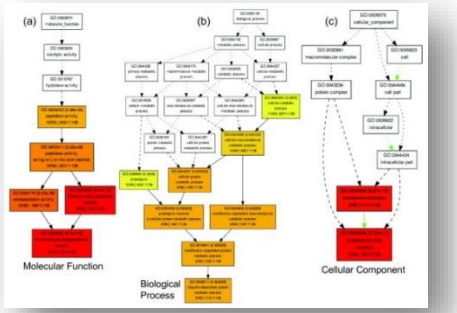
Des données complexes à explorer et ré-analyser pour les biologistes

- Critères arbitraires pour la sélection de listee gènes
- Compétences spécifiques pour la génération de graphiques/ analyses complémentaires









➔ Développer une solution pour que les biologistes puissent manipuler leurs jeux de données de manière autonome

Objectifs

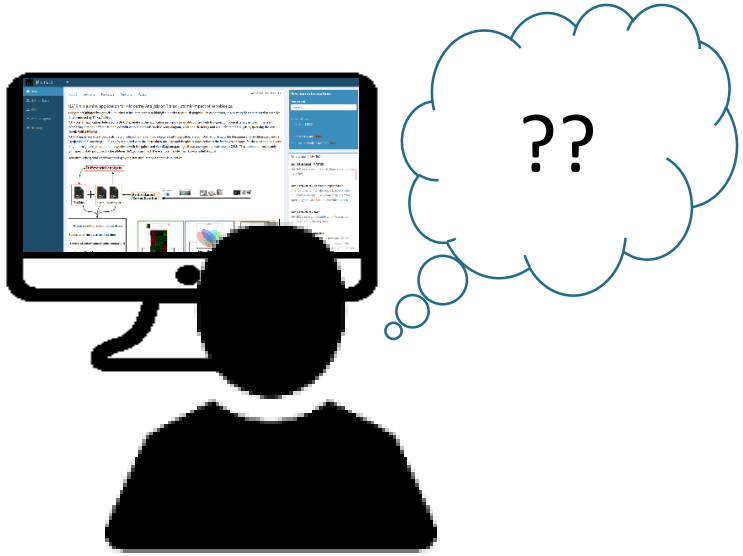
Stage M2 Bioinfo puis CDD Franck Soubès: développement et déploiement de l'application MATRiX

✓**Import datasets** (microarray, RNAseq) with statistical result table
- Upload from local
- Load from server

✓**Explore** and **Analyze** **data**
- ✓**PCA**
- ✓**Venn**
- ✓**Heatmap**
- ✓**Functional analysis**

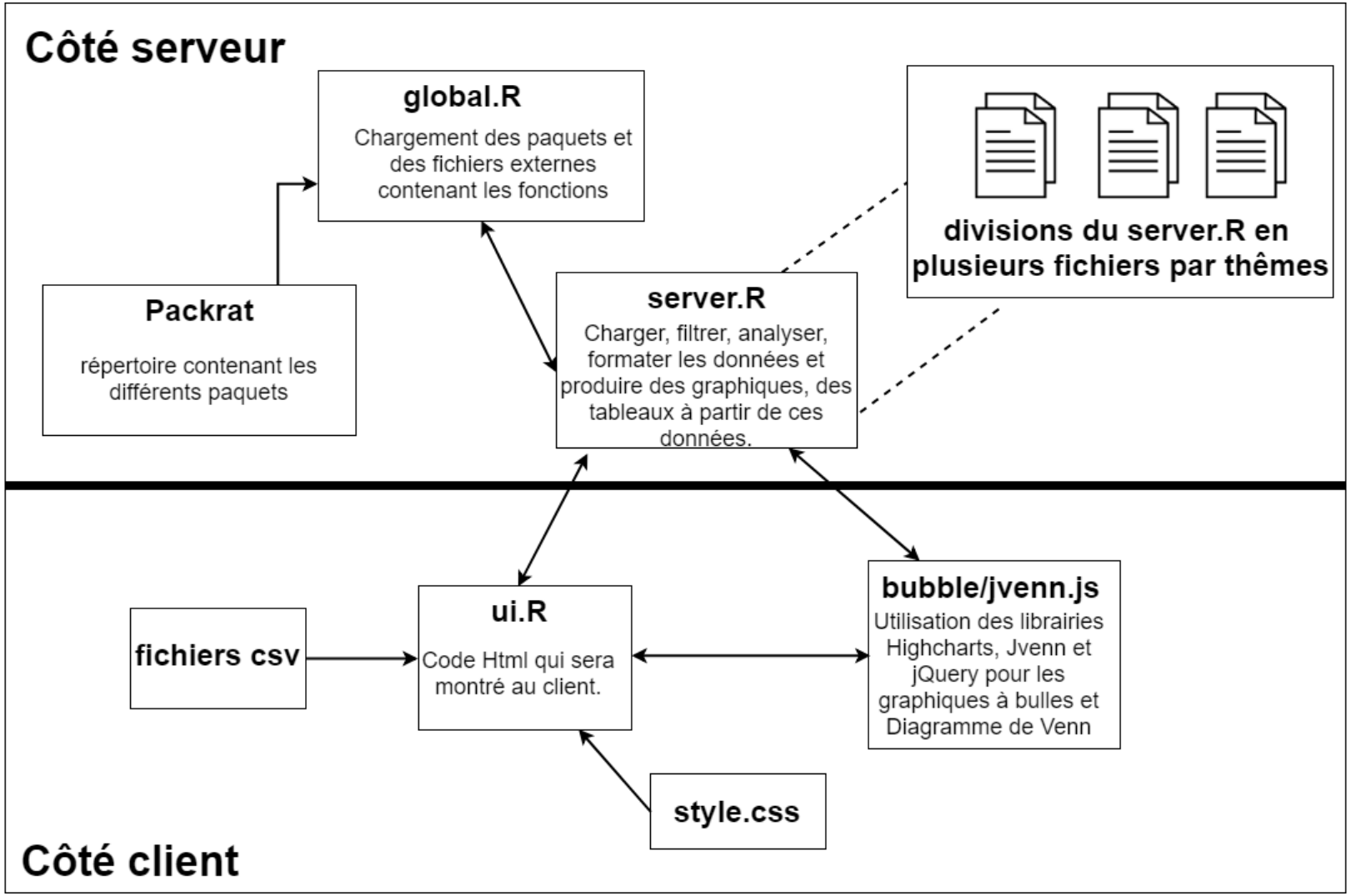
✓**Export publishable quality graphics**
- png
- eps, svg, pdf

✓Built into **a single web application**
- **Interactive** interface
- User friendly
- **Responsive**
- **Open** source and open access code (GPL-3)



RShiny : Principes

Architecture



Rshiny: Exemple

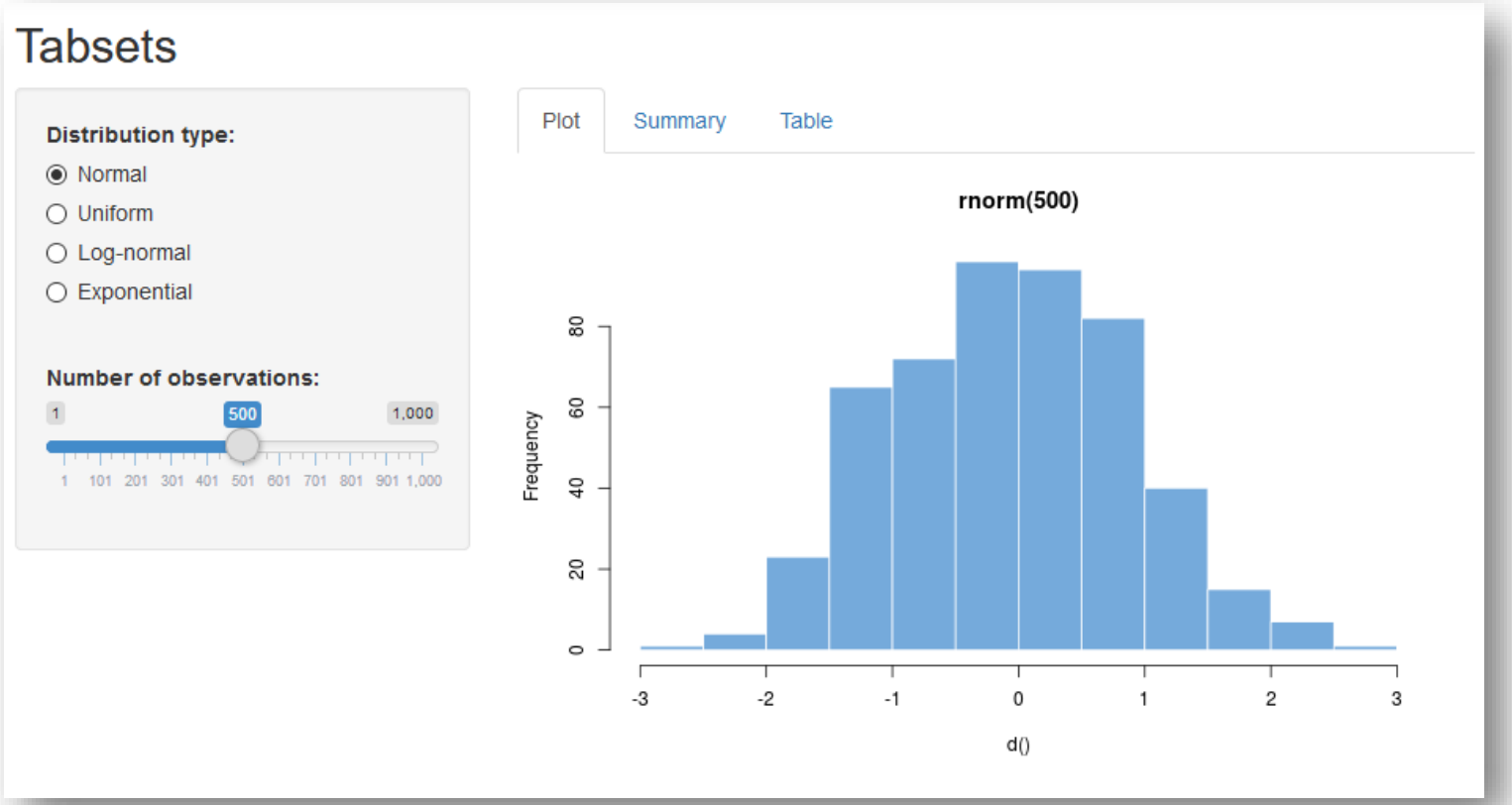
UI.R

```
ui <- fluidPage(
  titlePanel("Tabsets"),
  sidebarLayout(
    sidebarPanel(
      radioButtons("dist", "Distribution type:",
                   c("Normal" = "norm",
                     "Uniform" = "unif",
                     "Log-normal" = "lnorm",
                     "Exponential" = "exp")),
      sliderInput("n",
                  "Number of observations:",
                  value = 500,
                  min = 1,
                  max = 1000)
    ),
    mainPanel(
      tabsetPanel(type = "tabs",
                  tabPanel("Plot", plotOutput("plot")),
                  tabPanel("Summary", verbatimTextOutput("summary")),
                  tabPanel("Table", tableOutput("table"))
))))
```

Server.R

```
server <- function(input, output) {
  d <- reactive({
    dist <- switch(input$dist,
                   norm = rnorm,
                   unif = runif,
                   lnorm = rlnorm,
                   exp = rexp,
                   rnorm)
    dist(input$n)
  })
  output$plot <- renderPlot({
    dist <- input$dist
    n <- input$n
    hist(d(),
         main = paste("r", dist, "(", n, ")", sep = ""),
         col = "#75AADB", border = "white")
  })

  output$summary <- renderPrint({
    summary(d())
  })
  output$table <- renderTable({
    d()
  })
}
```



Utilisation

MATRiX

Home

Upload Data

How to import ?

- First click on the browse button to load the data
- After the pop up has appeared, you will have to select the data files.
- It is also possible to directly drag and drop your data files in the browse button
- You need three distinct csv files, these files are respectively named xxx_pData, xxx_WorkingSet and xxx_ResTable.
- pData : The experimental design in a 2 column table that associates samples to their respective biological conditions
- WorkingSet : The table of the normalised expression values (log2, cpm, ...) with genes in rows and samples in columns. The first column must contain the unique gene identifier (or transcript, probe, ...).
- ResTable : The table containing the results of differential analysis (fold change, p-value and FDR) next to a first column with unique gene identifier and a second column with the gene symbol.
- The final step consist to select all the data at once and then confirm the selection by clicking on the open button.
- A green message will then appear to confirm the data loading with a summary table.

Upload data

download sample data

User data (.csv format)

Browse... No file selected

Import local example

File select

Decimal

Comma

Point

Unique identifier

ProbeName

Table 1: xxx_pData.csv

X	Grp
S1	WT_CTRL
S3	WT_COND
S4	KO_CTRL
S5	KO_COND
S2	KO_COND
S8	KO_CTRL
S6	WT_COND
S7	WT_CTRL
S9	WT_COND

Table 2: xxx_Restable.csv

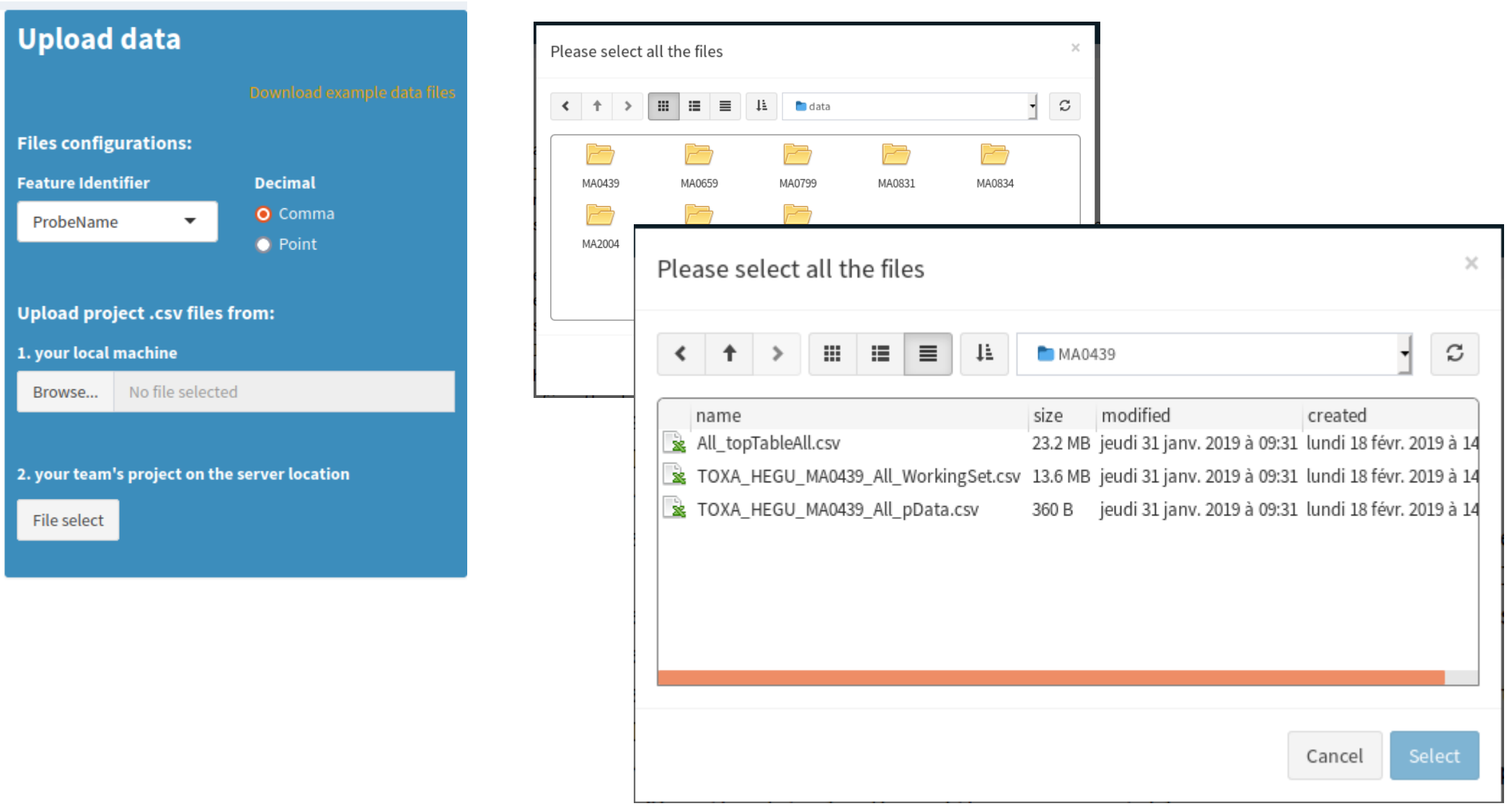
Unique IDs	GeneName	logFC	P.value	adj.P.Val
A_52_P1690	Dbil5	2,611	0,0048	0,785
A_51_P41424	C85492	1,772	0,0048	0,773
A_52_P108321	Ccdc71	-1,965	0,0078	0,819
A_55_P1985764	AV310571	0,208	0,00058	0,714
A_51_P328014	Cops3	-0,859	0,0048	0,604
A_52_P123354	Tusc2	-10,254	0,000048	0,00023

Table 3: xxx_WorkingSet.csv

Unique IDs	S1	S3	S4	S5	S2	S8	S6
A_52_P1690	6,512	7,511	7,007	6,276	6,760	6,276	7,064
A_51_P41424	9,975	5,525	10,479	9,745	10,040	9,745	9,662
A_52_P108321	7,035	8,2	7,150	8,293	7,698	8,293	7,384
A_55_P1985764	12,252	9,352	12,619	12,661	12,302	12,661	12,282
A_51_P328014	10,8000	6,251	10,717	10,901	10,7171	10,901	10,705
A_52_P123354	6,832	7,282	6,944	7,117	6,944	7,117	6,860

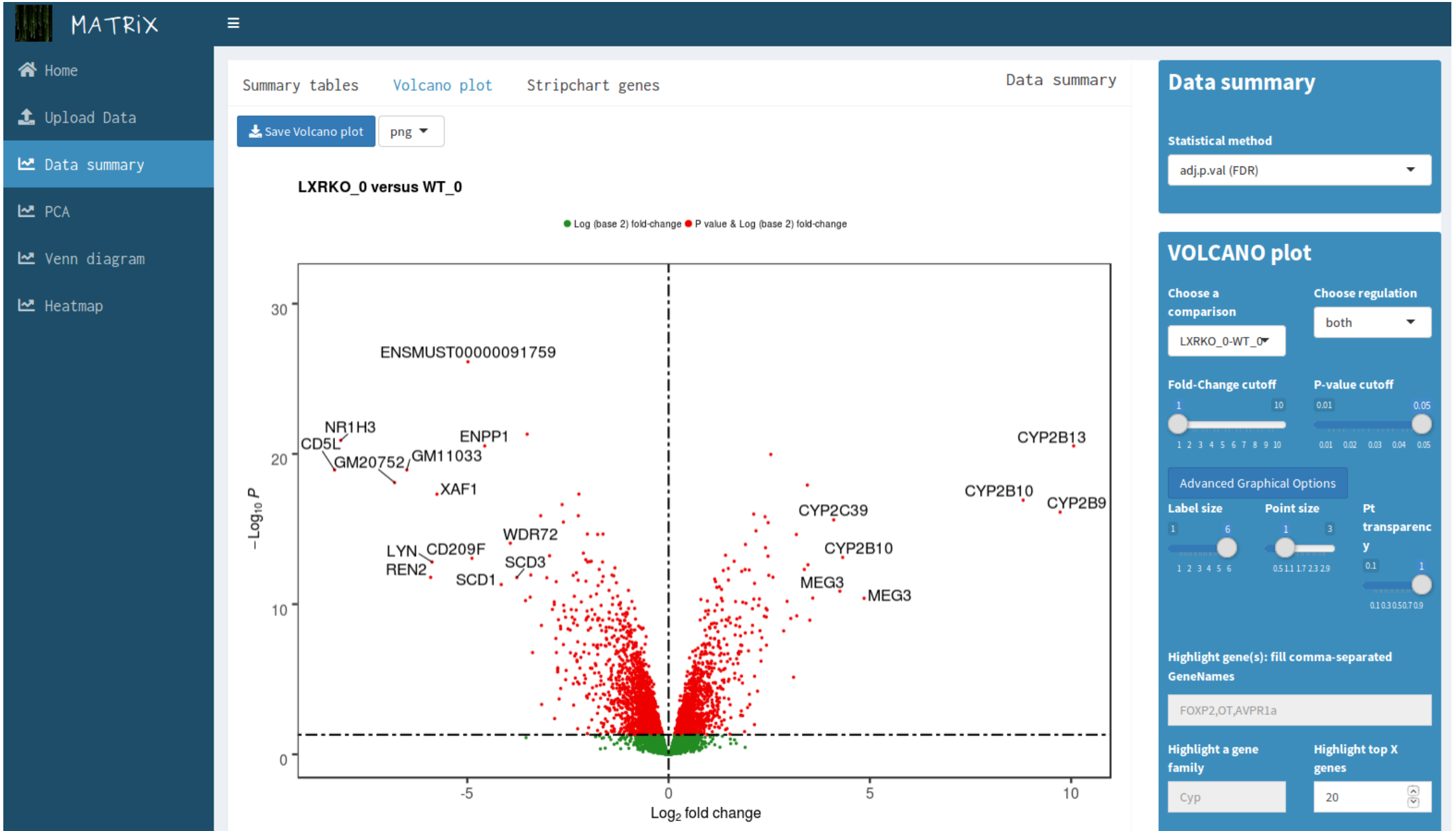
Utilisation

Chargement des données (données stockées sur le serveur)



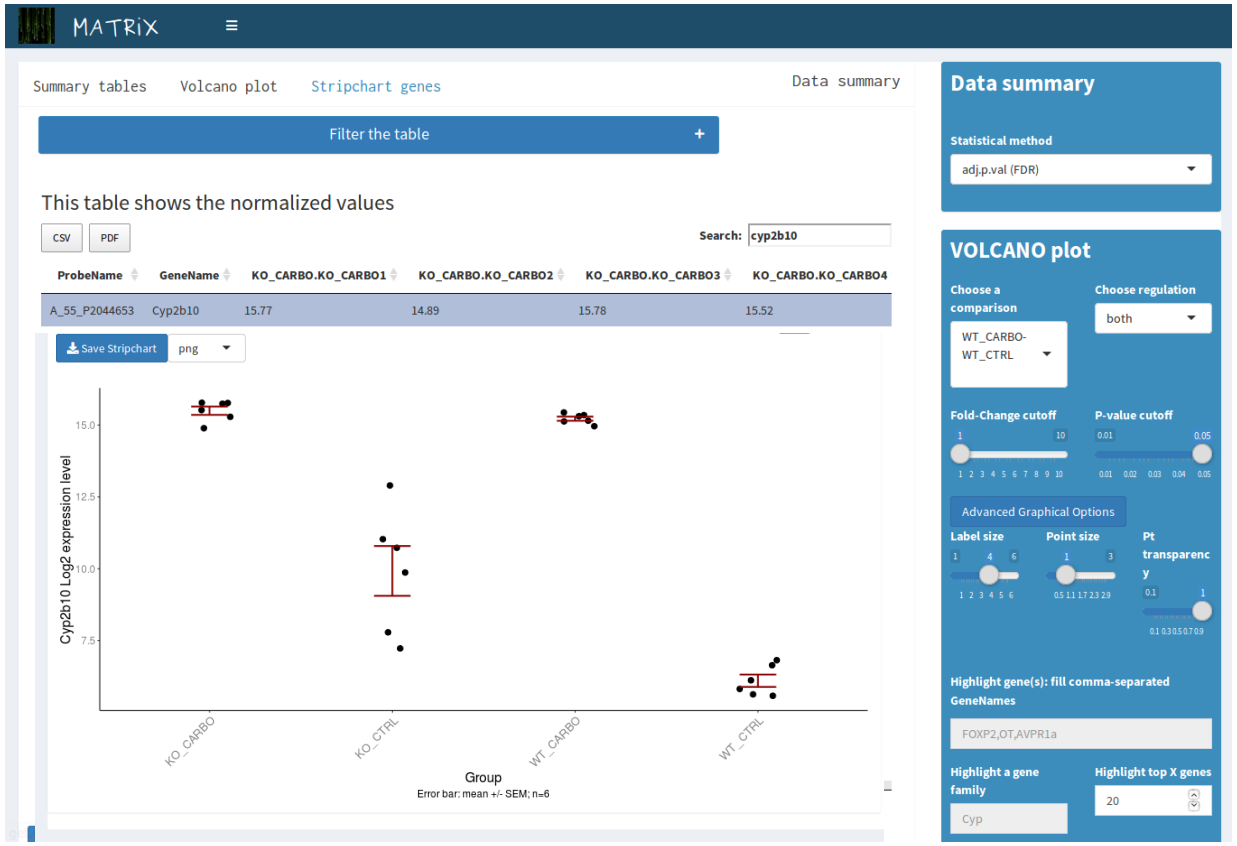
Genotoul GeT

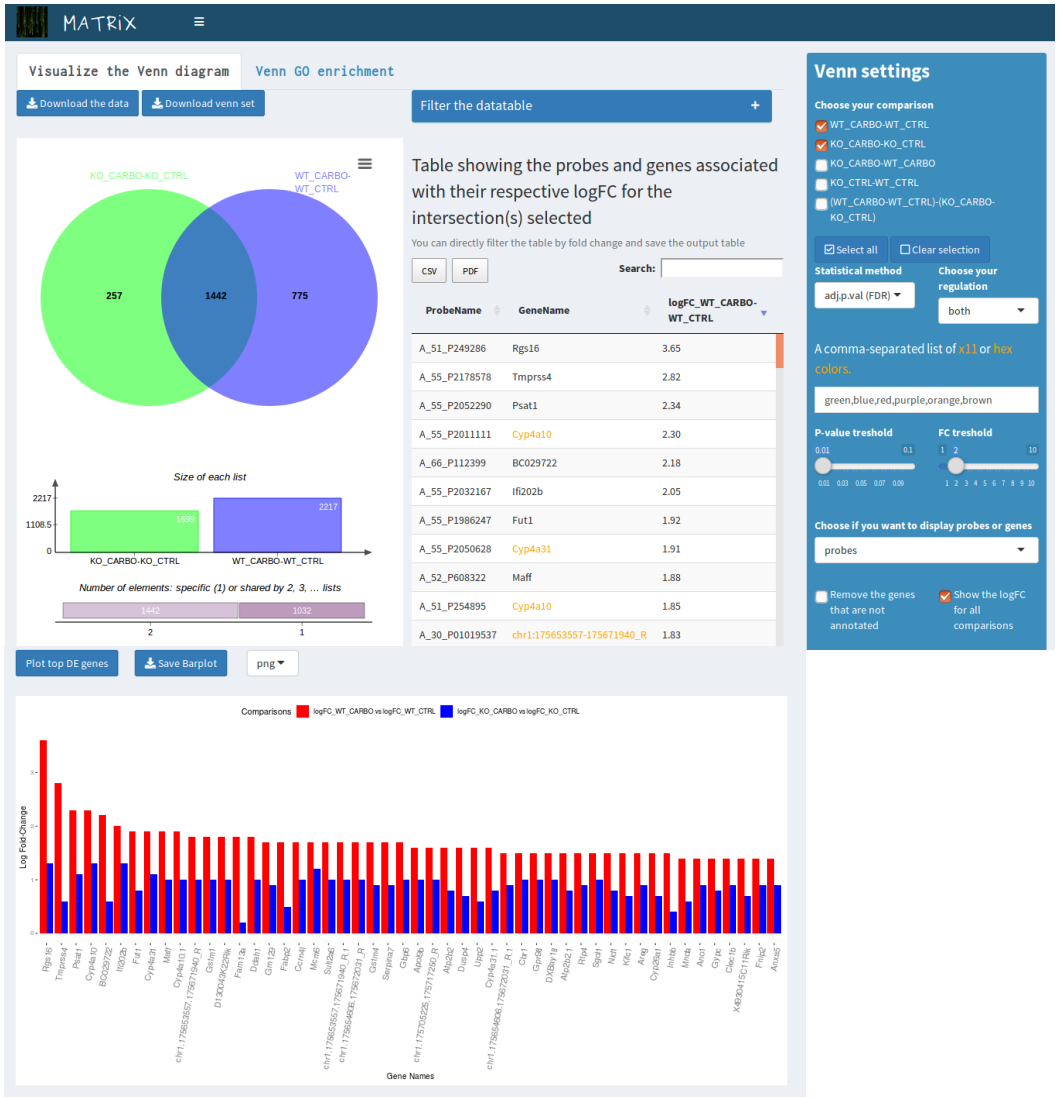
Explore Results



Genotoul GeT

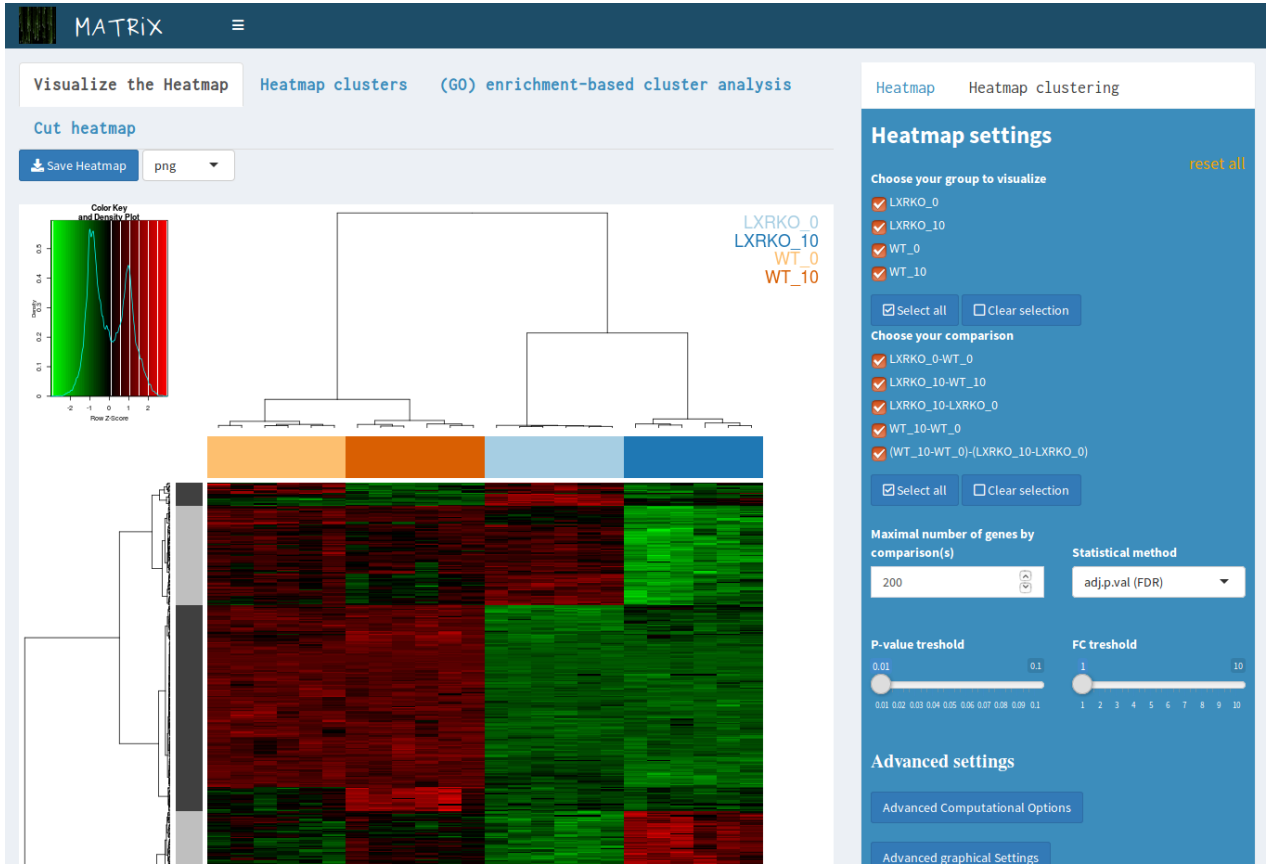
Explore results



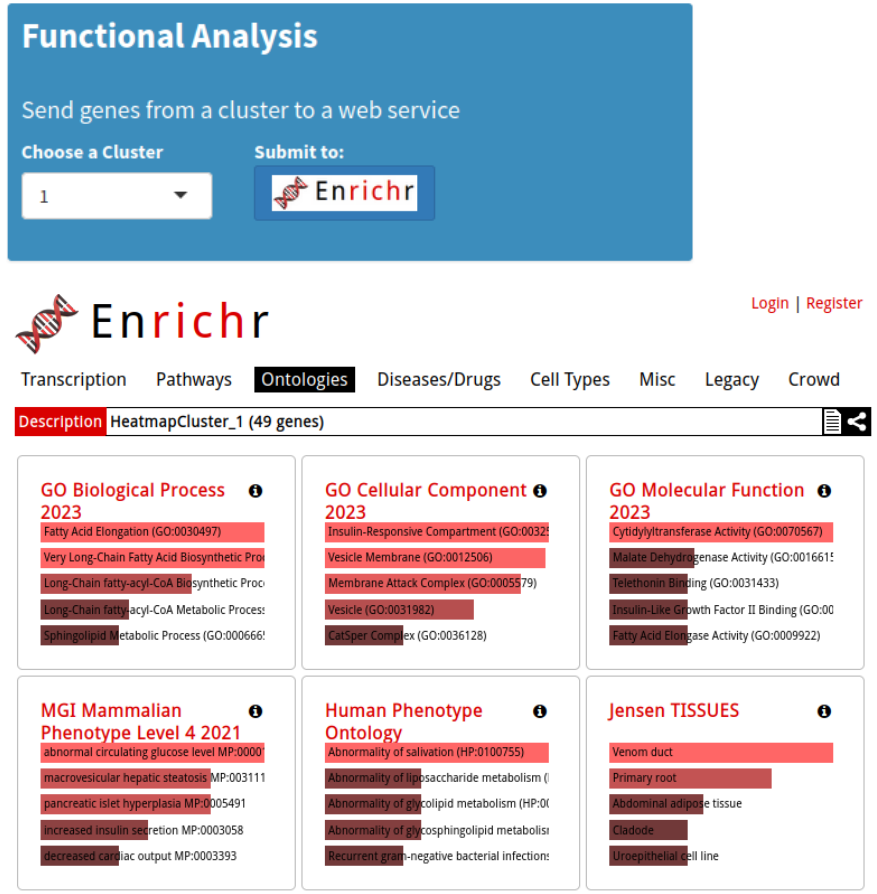


Analyses fonctionnelles

- **Heatmap** clustering



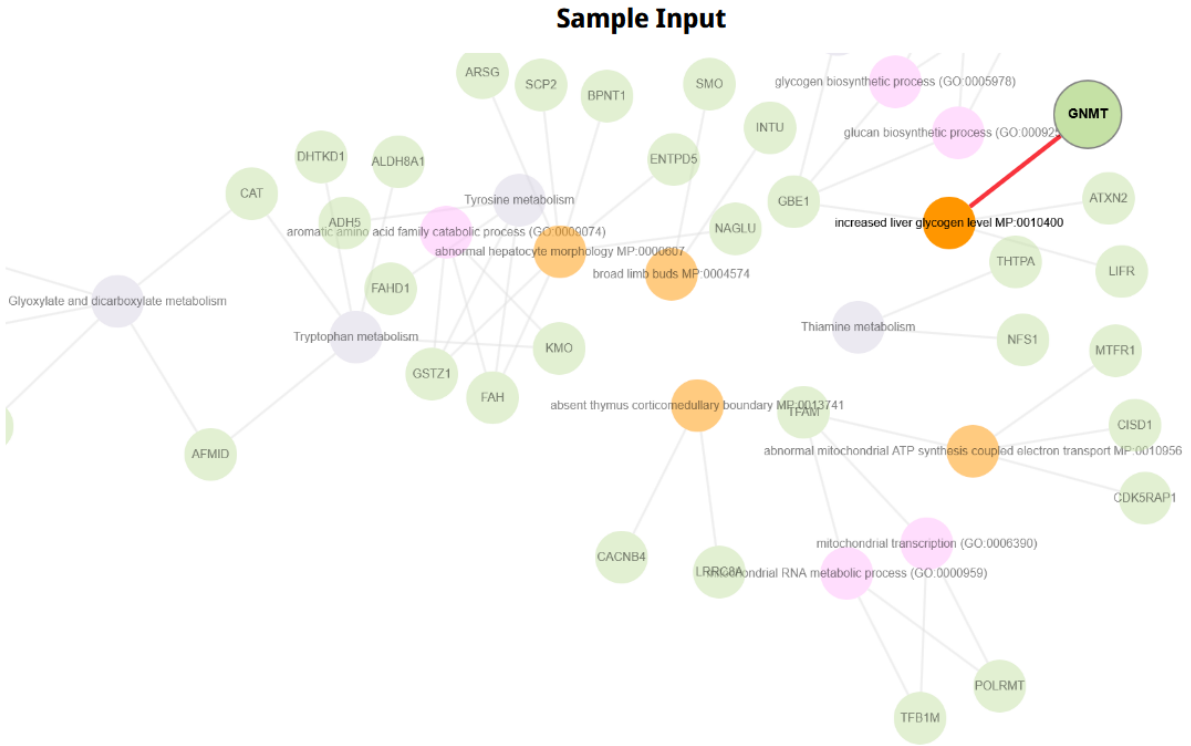
- Functional **Analysis**



https://maayanlab.cloud/enrich

Evolutions prévues

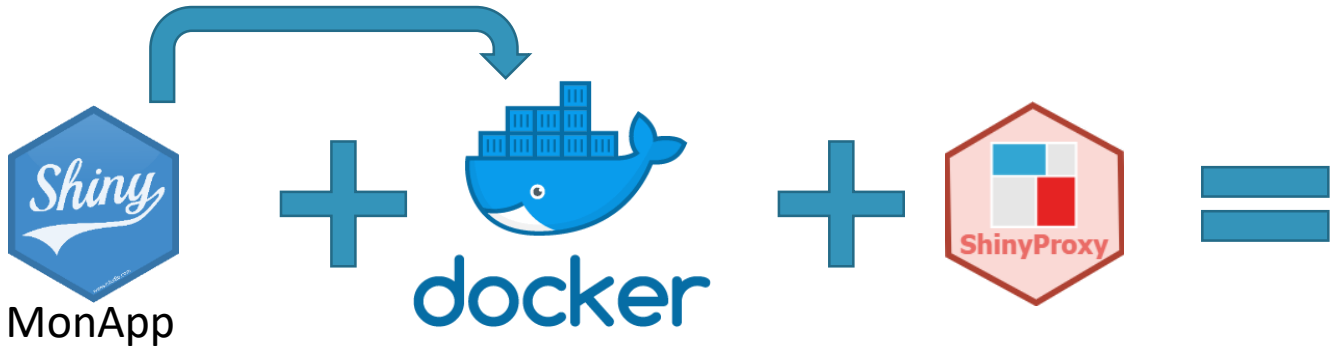
- **Correction « bugs »**
  - Table sorting
  - Scroll lag
- **Améliorations**
  - Send venn list to enricher…
  - Implémenter API string.DB
  - enrichr-KG



Quelle solution d’hébergement?

Hebergement 1: shinyproxy/ VM genotoul

https://www.shinyproxy.io/



https://forgemia.inra.fr/ylippi/MA_Trix_App

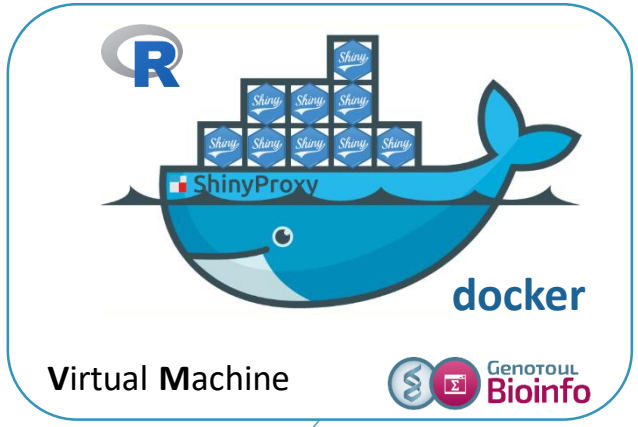
Dockerfile
Ubuntu 20.04 with r-base
Dependencies packages
COPY MonApp /root/monapp
EXPOSE 3838
Shiny::runApp("/root/monapp")

$ sudo docker build -t **MyDockApp** .

application.yml

```
specs:
  - id: monapp
    display-name: Mon application
    container-cmd: ["R", "-e",
"shiny::runApp('/root/monapp')"]
    container-image: MyDockApp
```

$ java -jar shinyproxy-3.0.1.jar



https://matrix.toulouse.inrae.fr:8080

Pro:
- configurable

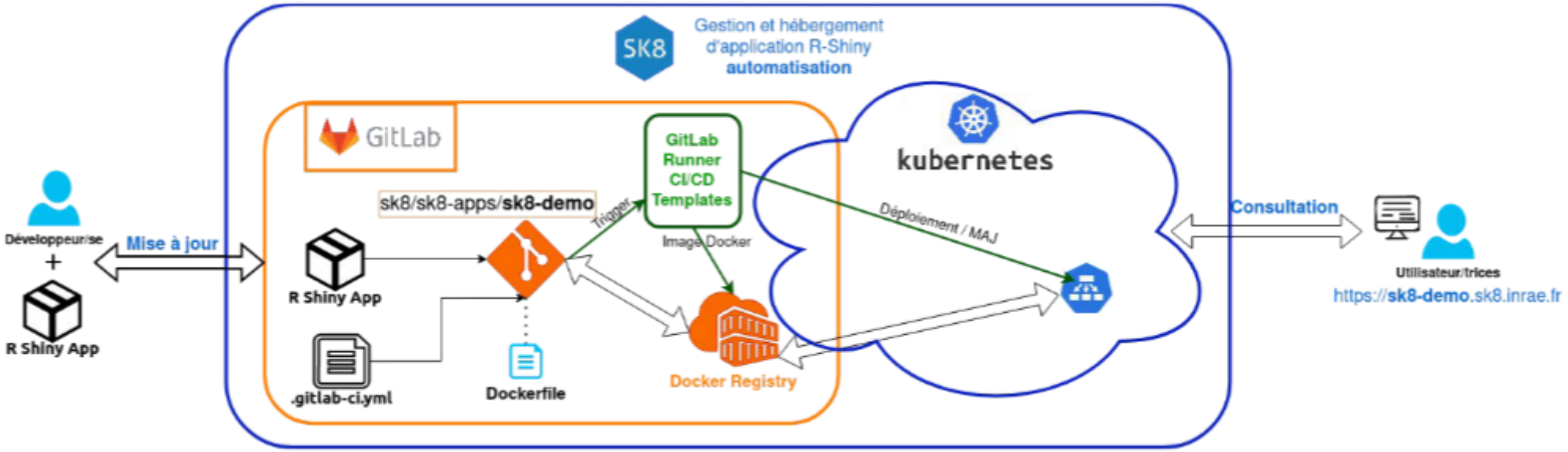
Cons:
- Configure and maintain host server (VM)
- Pay VM location service

Hebergement 2: kubernetes/ SK8 INRAE

https://sk8.inrae.fr

https://sk8.inrae.fr/

Hebergement et déploiement à INRAE



Version 1 instance/ multi-users https://matrixapp.sk8.inrae.fr/

Version multi-instances / multi-users https://shiny.sk8.inrae.fr/app/get-trix-matrixapp (shinyProxy)

Pro:

- Déploiement automatisé via gitlab CI/CD
- Cluster kubernetes
- Peformance

Cons:

- Réservé développeurs INRAE
- Financement infra à identifier pour solution d'hébergement pérenne

Démo

https://matrix.toulouse.inrae.fr