

## MATRiX app: Mining and functional Analysis of TRanscriptomics data

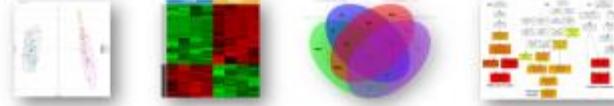
Une application dédiée à l'exploration de résultats d'analyses biostatistiques de données transcriptomiques



## Problématique

Des chercheurs complètes à explorer et ré-analyser pour des biologistes

- ④ Critères arbitraires pour la sélection de listes gènes
- ④ 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

<http://genotoul.fr> 

## Utilisation

How to import ?

Put click on the Import button to load the data.

- After you have imported, you will have to select the file in the menu "File" → "Import" or directly by clicking on the "Import" button. Then click on the "Import" button.
- You need three distinct files. These files are respectively named `experiments.xls`, `matrix.xls` and `experiments.xls`.
- Matrix : The experimental design is in a 2 column table that associates samples to three successive biological conditions.
- Experiments : The table of the normalized expression values (log2, 0, 1,...) with genes in rows and samples in columns. The first column must contains the unique gene identifier per transcript, probe, ...).
- Metadata : The table of the experimental conditions (treatment, probe and PBO) next to a fixed column with 1.
- The first line consists to select all the data of each and then confirm the selection by clicking on the same button.
- A green message will then appear to confirm the data loading with a summary table.

<http://genotoul.fr> 

## Quelle solution d'hébergement?

<http://genotoul.fr> 

## Démo

<https://matrix.toulouse.inrae.fr>

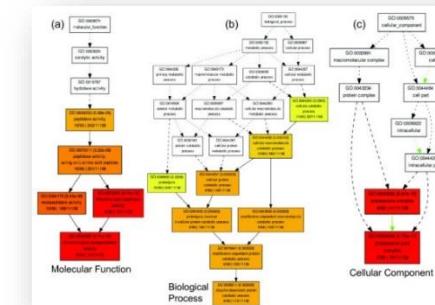
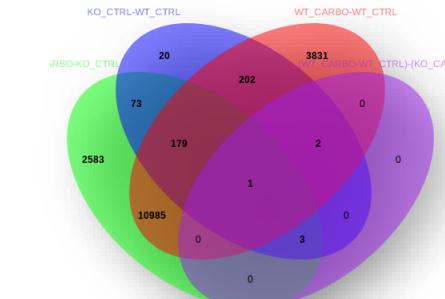
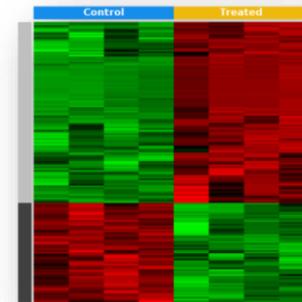
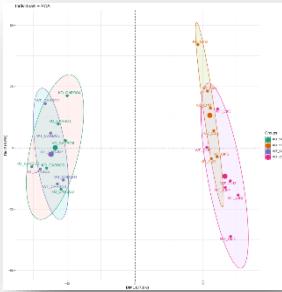
<http://genotoul.fr> 



# Problématique

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

- ④ Critères arbitraires pour la sélection de listee gènes
- ④ 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

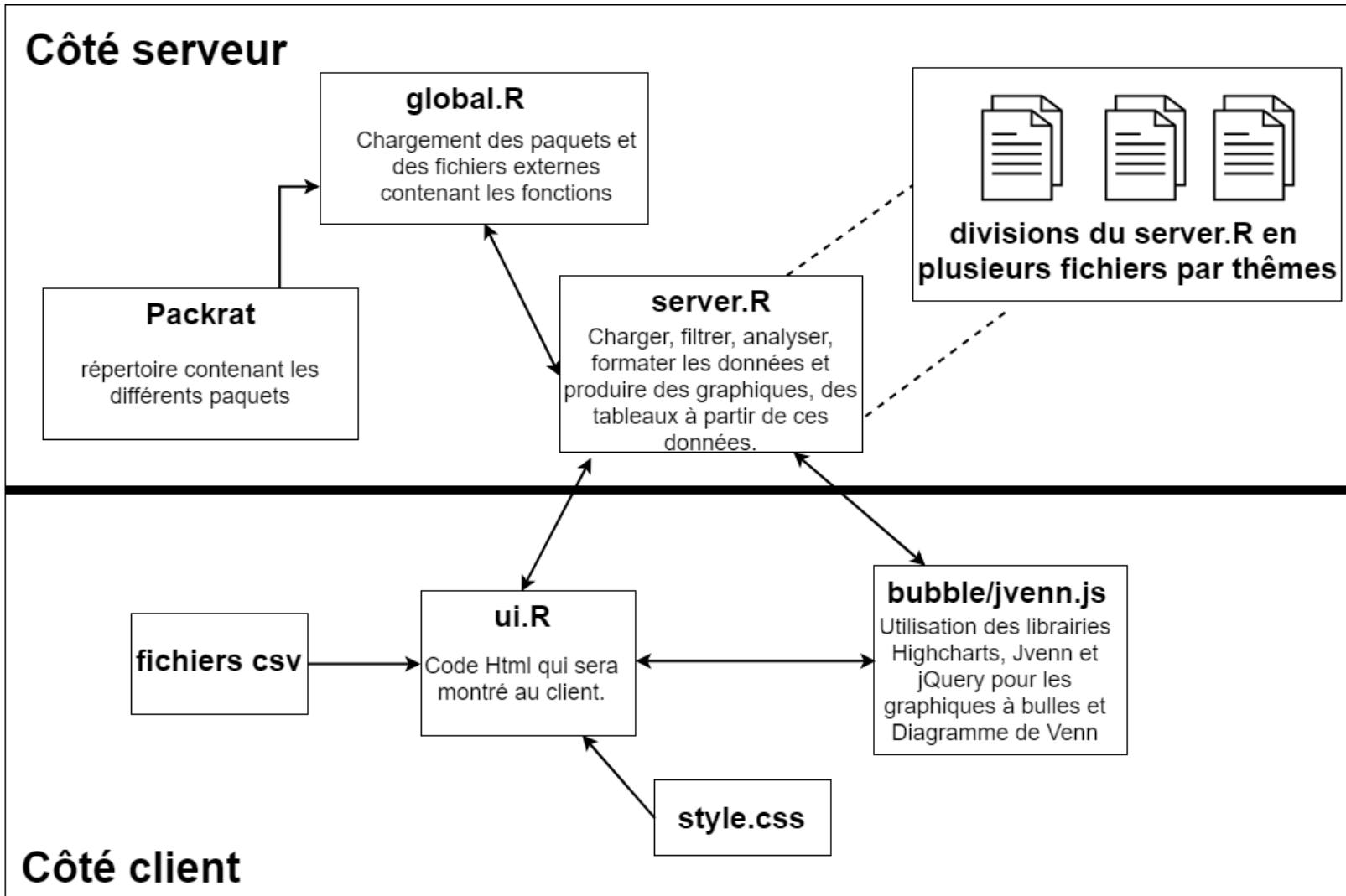
✓ **Built into a single web application**

- **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()
  })
}

```

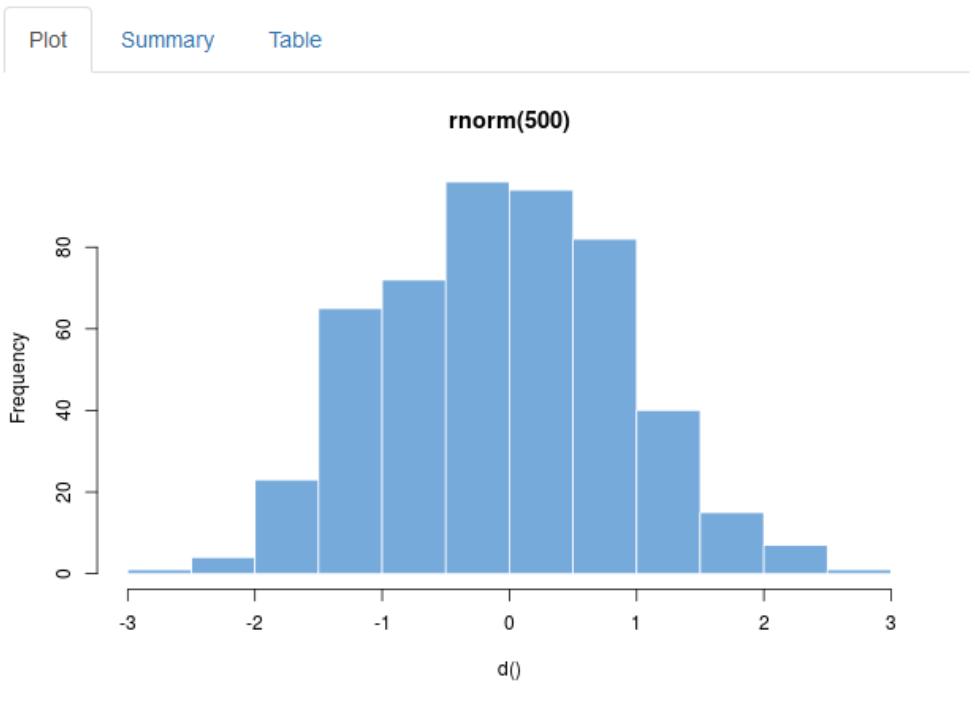
## Tabsets

**Distribution type:**

- Normal
- Uniform
- Log-normal
- Exponential

**Number of observations:**

1
500
1,000



# Utilisation

 MATRIX

- [!\[\]\(01fb5058363dcb3bfe1ee1159e9c248e\_img.jpg\) Home](#)
- [!\[\]\(54f0ad8b6afbf069171bcb3f2d838cc1\_img.jpg\) 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.

Table 1: xxx\_pData.csv

X	Grp
S1	WT_CTRL
S3	WT_COND
S4	KO_CTRL
S5	KO_COND
S2	KO_CTRL
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

### Upload data

[download sample data](#)

User data (.csv format)

Browse... No file selected

Import local example Decimal

Comma  Point

Unique identifier

ProbeName



# Utilisation

Chargement des données (données stockées sur le serveur)

### Upload data

Download example data files

**Files configurations:**

Feature Identifier      Decimal  
 Comma       Point

ProbeName

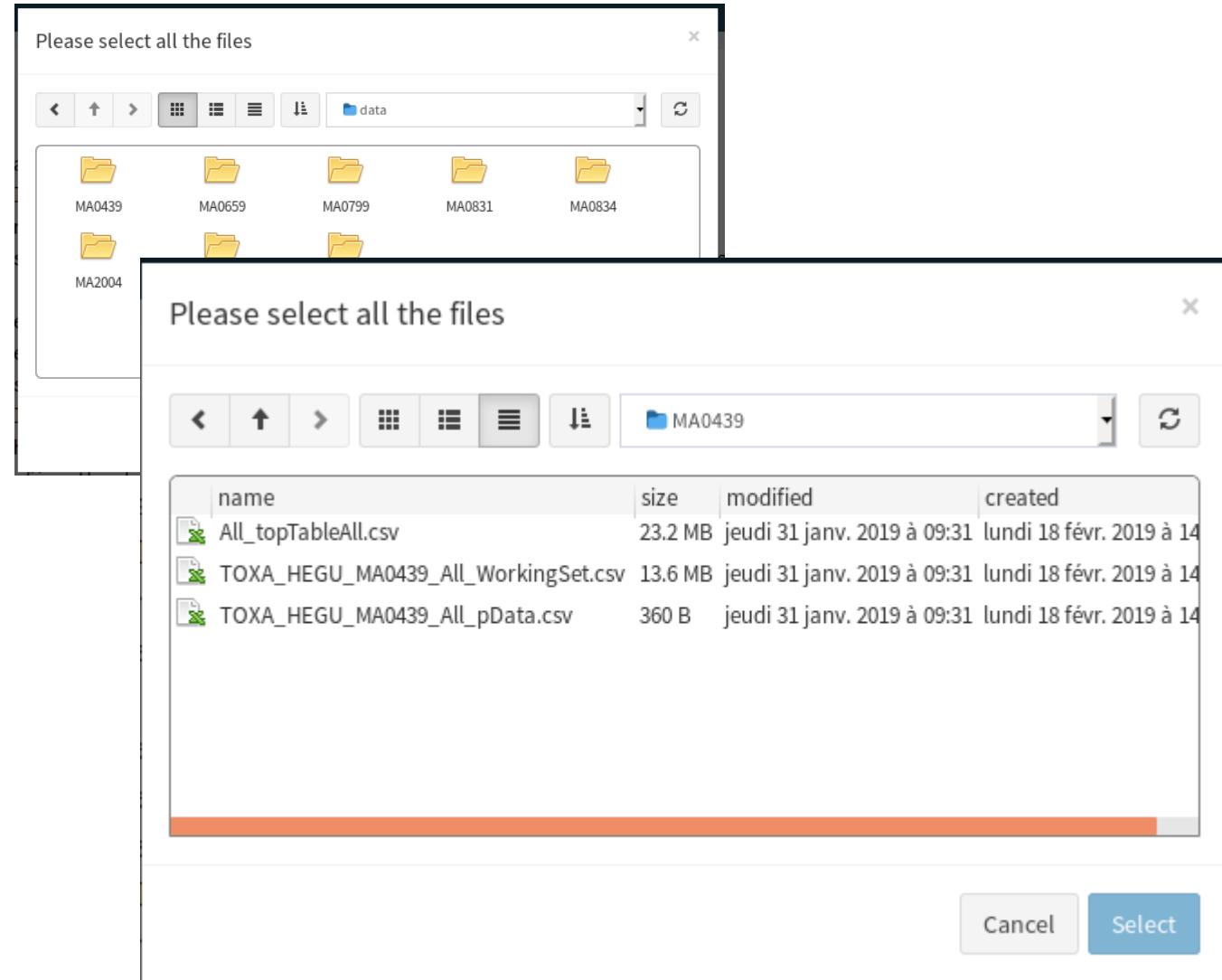
Upload project .csv files from:

1. your local machine

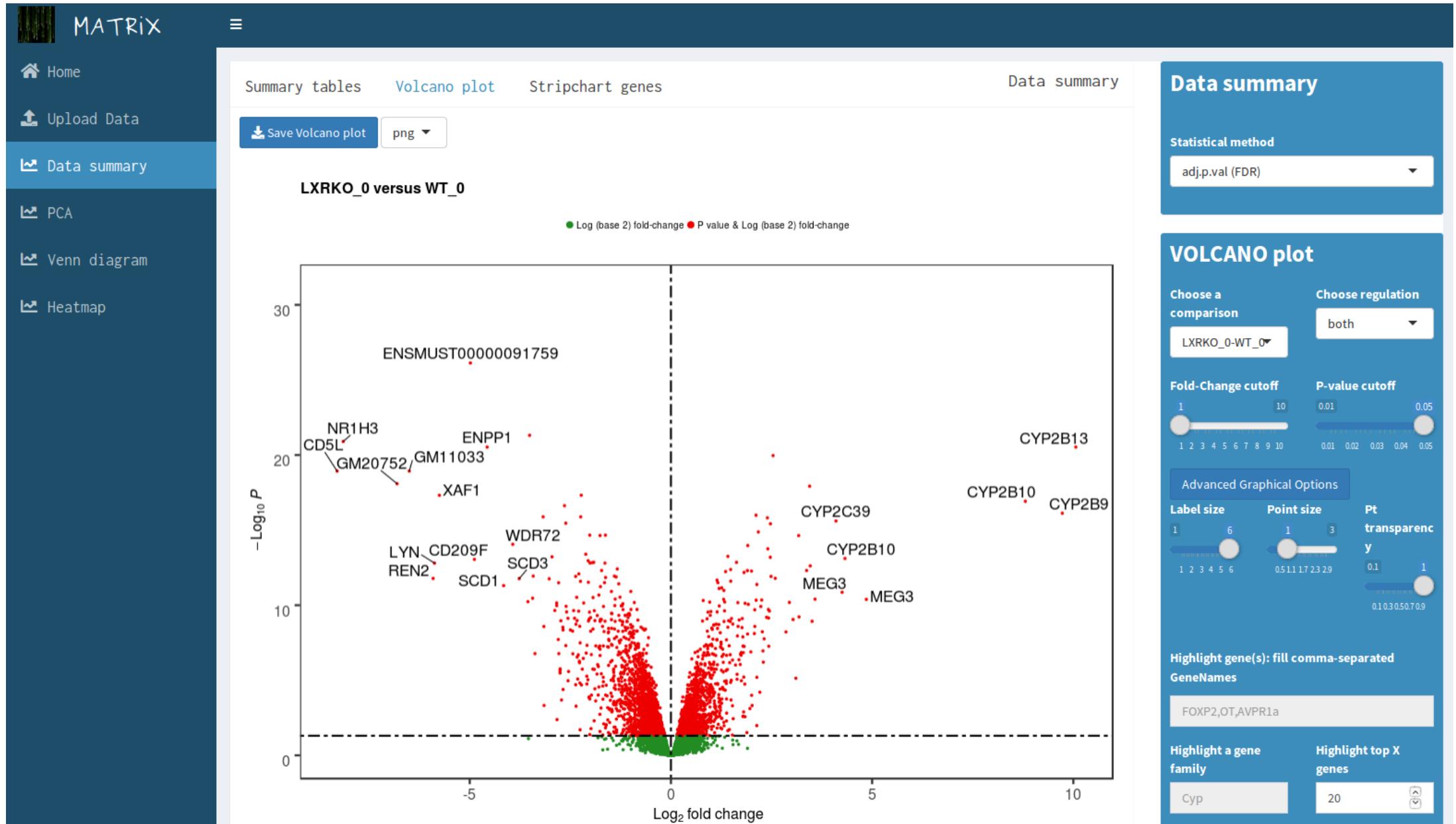
Browse... No file selected

2. your team's project on the server location

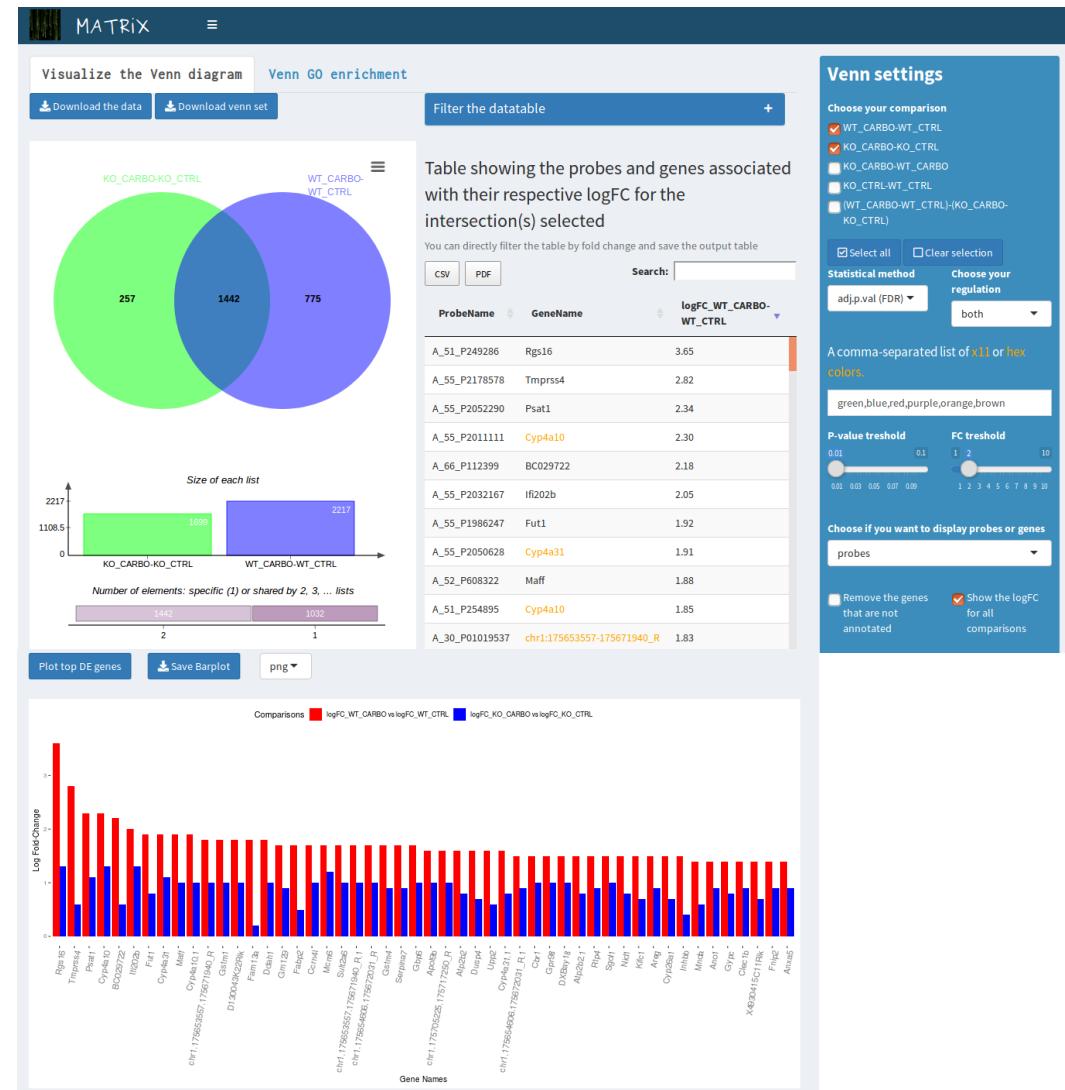
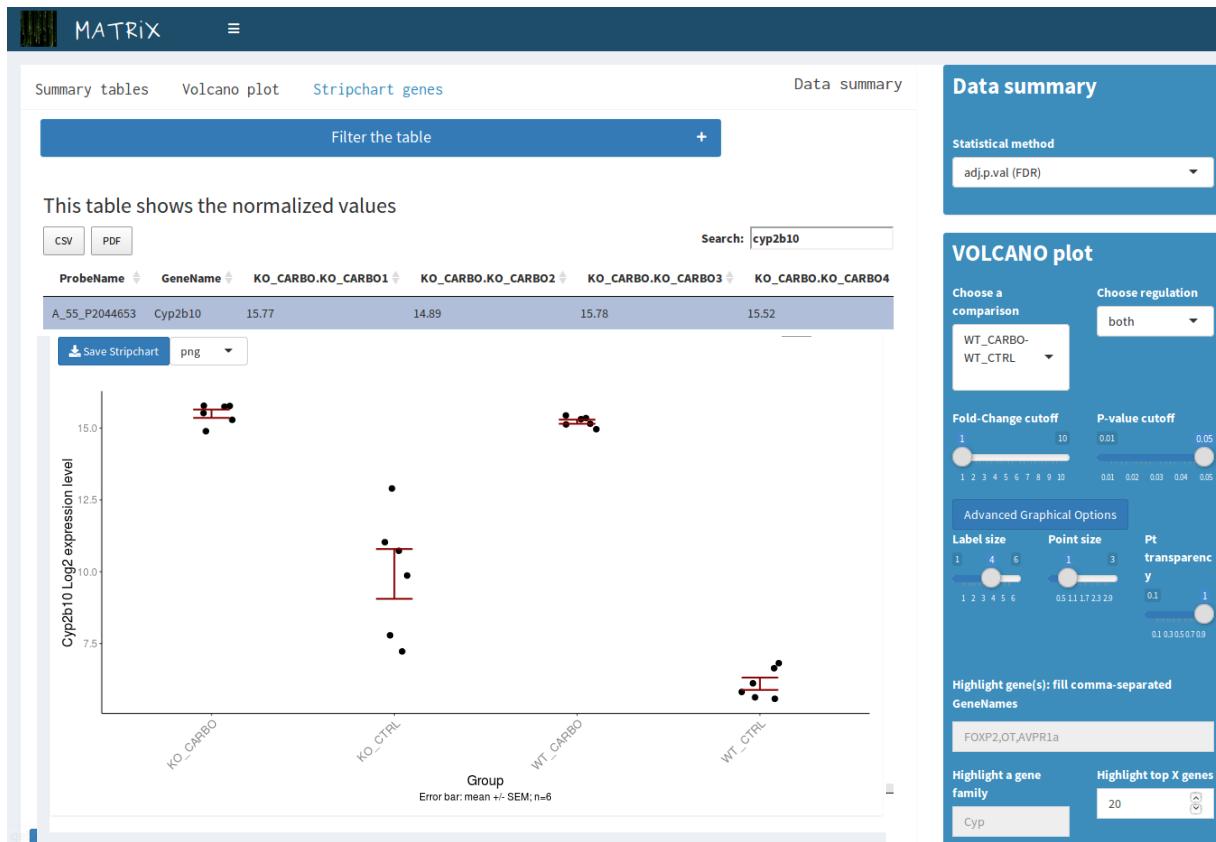
File select



## Explore Results

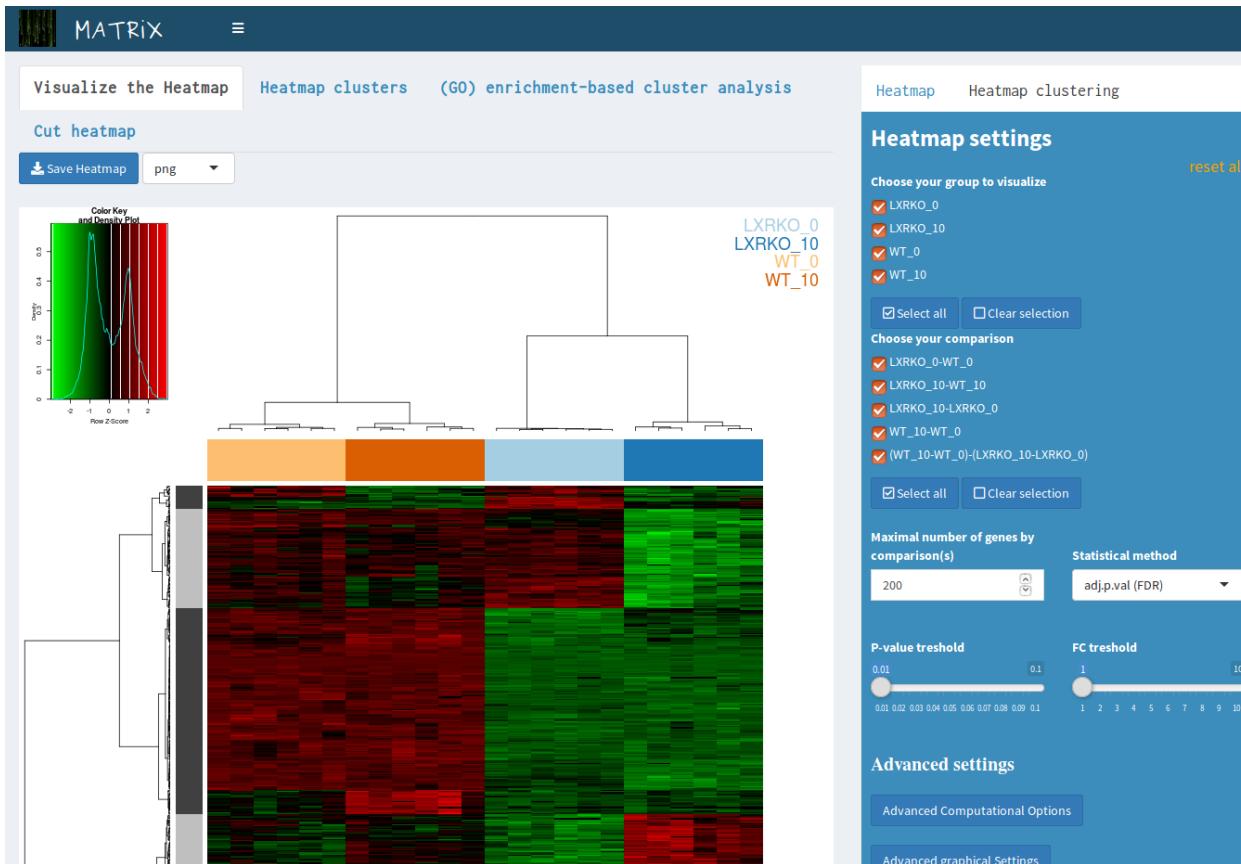


# Explore results



# Analyses fonctionnelles

- Heatmap clustering



- Functional Analysis

## Functional Analysis

Send genes from a cluster to a web service

Choose a Cluster: 1      Submit to: Enrichr

## Enrichr

Login | Register

Description: HeatmapCluster\_1 (49 genes)

**GO Biological Process** 2023

- Fatty Acid Elongation (GO:0030497)
- Very Long-Chain Fatty Acid Biosynthetic Process
- Long-Chain fatty-acyl-CoA Biosynthetic Process
- Long-Chain fatty-acyl-CoA Metabolic Process
- Sphingolipid Metabolic Process (GO:000666)

**GO Cellular Component** 2023

- Insulin-Responsive Compartment (GO:0032122)
- Vesicle Membrane (GO:0012506)
- Membrane Attack Complex (GO:0005579)
- Vesicle (GO:0031982)
- CatSper Complex (GO:0036128)

**GO Molecular Function** 2023

- Cytidylyltransferase Activity (GO:0070567)
- Malar Dehydrogenase Activity (GO:0016615)
- Telethonin Binding (GO:0031433)
- Insulin-Like Growth Factor II Binding (GO:000666)
- Fatty Acid Elongase Activity (GO:0009922)

**MGI Mammalian Phenotype Level 4 2021**

- abnormal circulating glucose level MP:0000
- macrovacular hepatic steatosis MP:003111
- pancreatic islet hyperplasia MP:0005491
- increased insulin secretion MP:0003058
- decreased cardiac output MP:0003393

**Human Phenotype Ontology**

- Abnormality of salivation (HP:0100755)
- Abnormality of liposaccharide metabolism (HP:0005491)
- Abnormality of glycolipid metabolism (HP:0003058)
- Abnormality of glycosphingolipid metabolism (HP:0003393)
- Recurrent gram-negative bacterial infection: (HP:0005491)

**Jensen TISSUES**

- Venom duct
- Primary root
- Abdominal adipose tissue
- Cladode
- Uroepithelial cell line

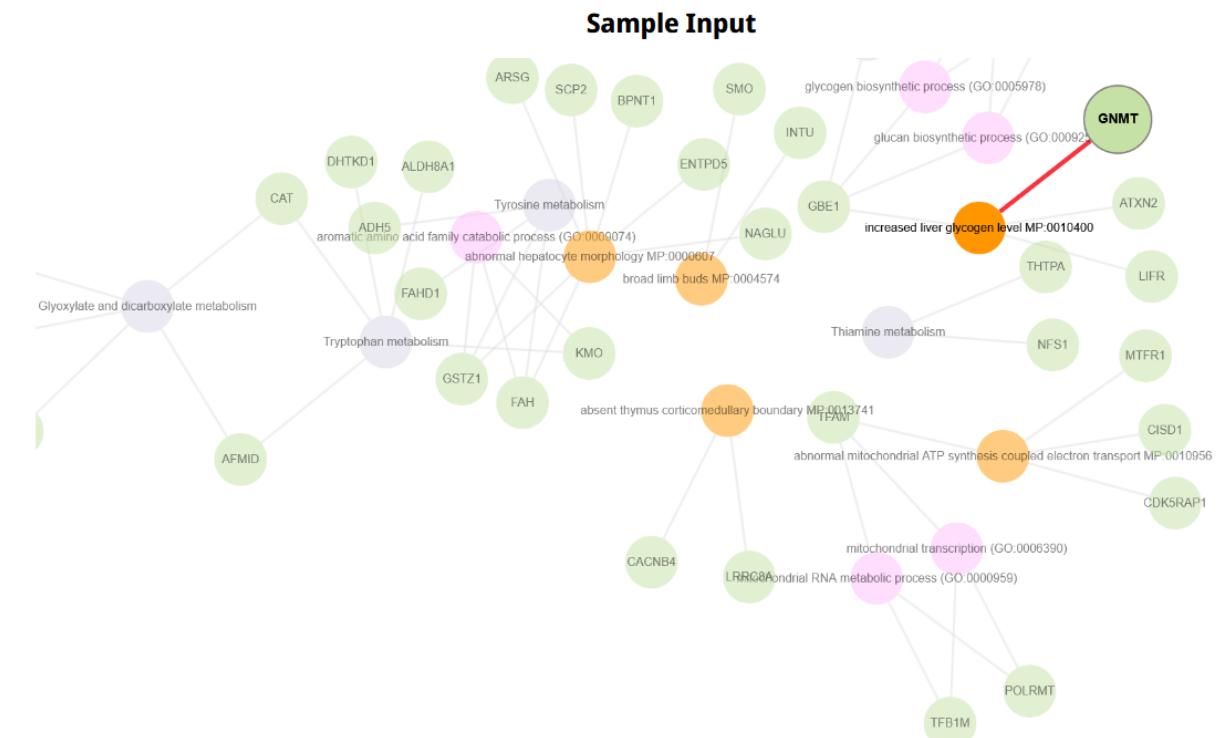
<https://maayanlab.cloud/enrich>

<http://get.genotoul.fr>



# Evolutions prévues

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

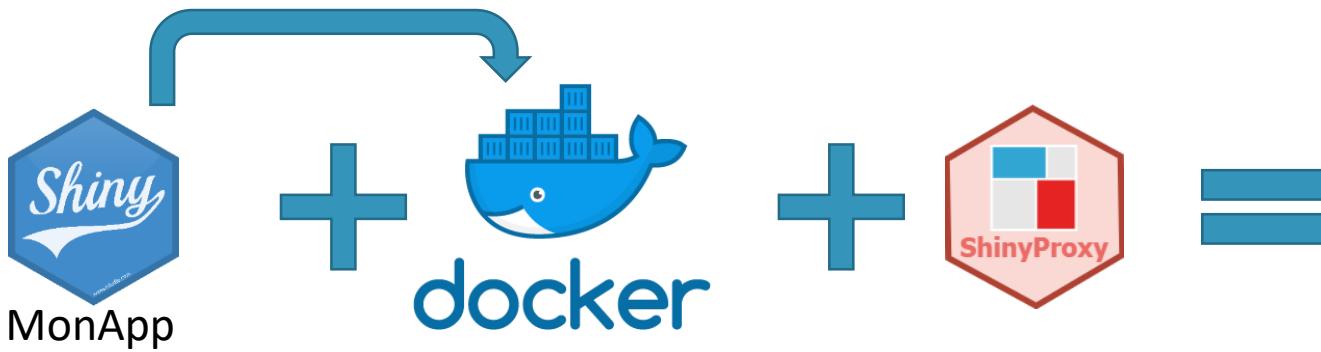


# Quelle solution d'hébergement?



# Hebergement 1: shinyproxy/ VM genotoul

<https://www.shinyproxy.io/>



[https://forgemia.inra.fr/ylliipi/MA\\_Trix\\_App](https://forgemia.inra.fr/ylliipi/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
```

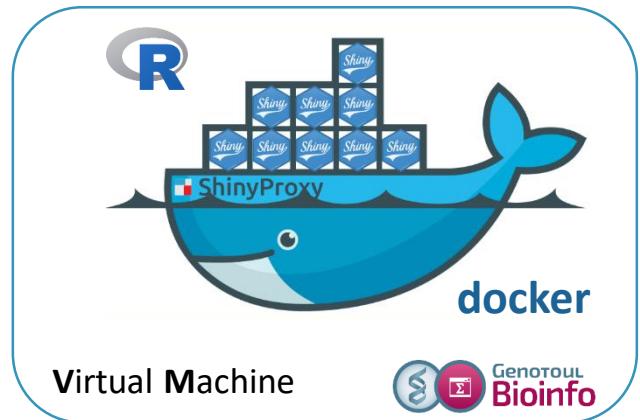
```
$ java -jar shinyproxy-3.0.1.jar
```

### Pro:

- configurable

### Cons:

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



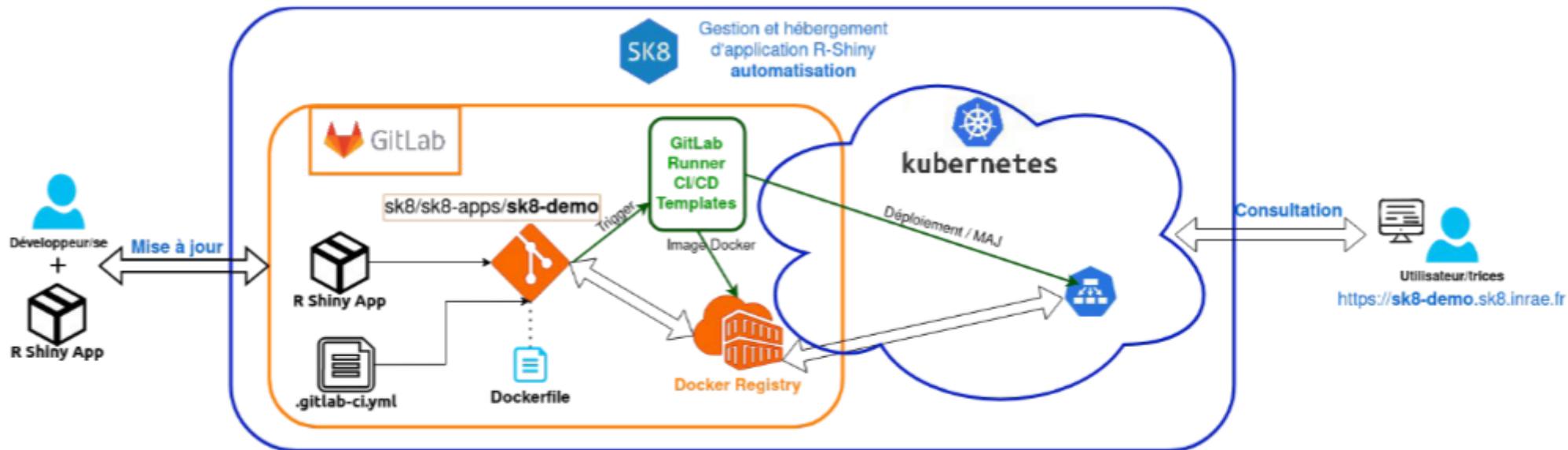
<https://matrix.toulouse.inrae.fr:8080>

# 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
- Performance

### Cons:

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



# Démo

<https://matrix.toulouse.inrae.fr>

