

RFLOMICS : Interactive web application for Omics-data analysis

(GNET team, IJPB, CATI Sysmics)

Delphine Charif, Nadia Bessoltane, Christine Paysant-Le-Roux

RFLOMICS : R package + Shiny application to perform omics-data analysis

Specifications

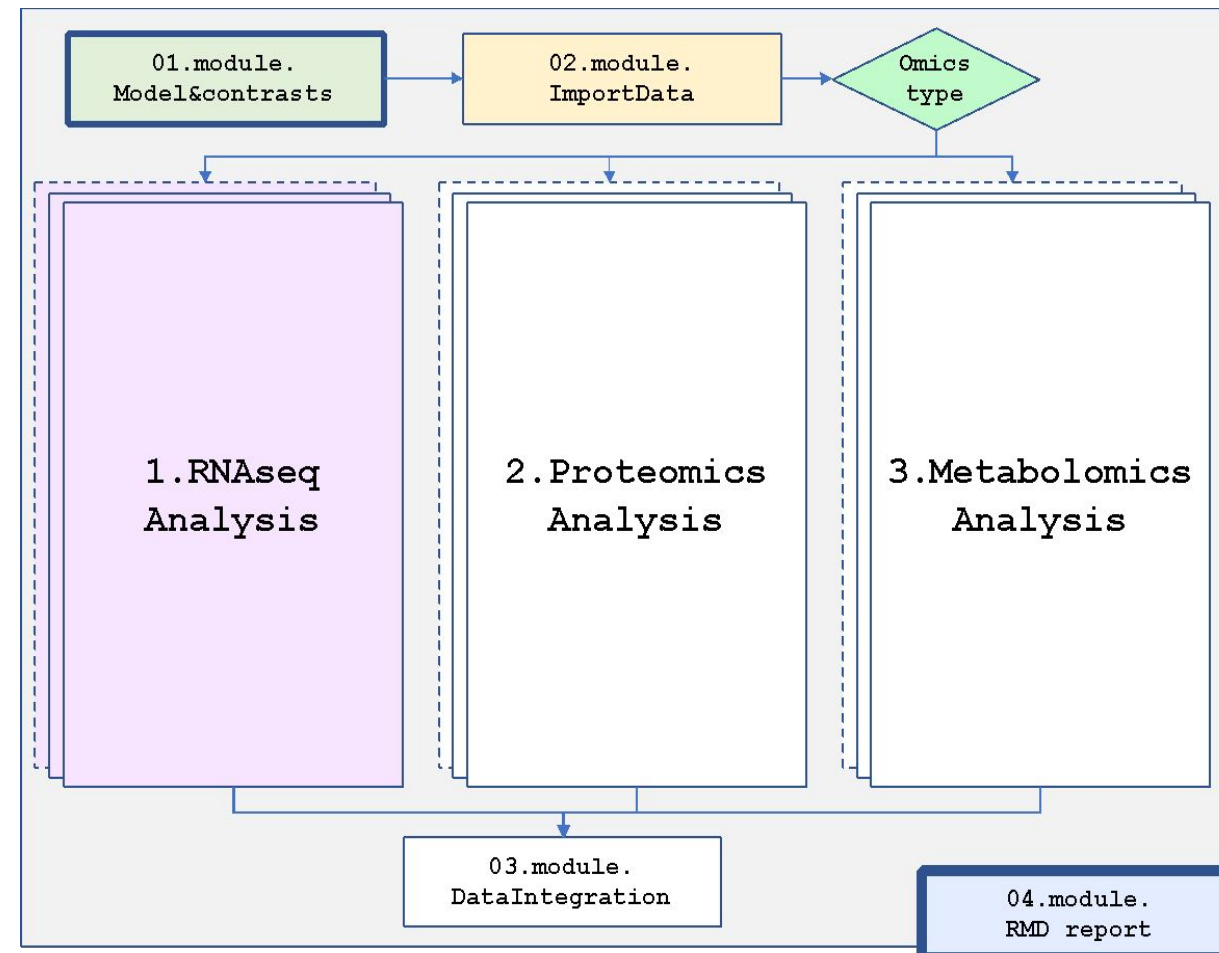
- Performs complete multi-omics project analysis,
- Support multi-factorial experimental design,
- Guarantees the relevance of the used methods,
- FAIR code
 - Reproducibility,
 - Share and capitalize our expertise,
 - ...
- Accessible via simple user-friendly interface

RFLOMICS : R package + Shiny application to perform omics-data analysis

- **Model & contrasts**
 - Multi-factorial statistics model
- **ImportData :**
 - Support 3 omics (same biological context)
 - Input type :
 - RNAseq -> read count (discrete)
 - proteomics/metabolomics -> abundance (continuous)
- **Data analysis**

(The relevance of the methods is guaranteed by experts)

 1. RNAseq : IPS2, IJPB
DiCoExpress (Ilana & Christine)
 2. Proteomics : CATI sysmics, pappso LCMS
 3. Metabolomics : CATI sysmics, IJPB (OV chimie)
- **Data integration** : CATI sysmics
- **RMD report**



Statistical translation of the context

- Project Ecoseed (L. Rajjou & G. Cueff)
 - Study of seed germination and vigor in *Arabidopsis thaliana*
 - 3 omics (RNAseq, proteomics and metabolomics)
- Statistical modeling : *Generalized Linear Model*

-> For each gene

- Interaction between imbibition and temperature

$$\log(\lambda_{i,t,r}) = \mu + \text{Replicat}_r + \text{Imbibition}_i + \text{Temperature}_t + \text{Imbibition}_i : \text{Temperature}_t$$

- No interaction

$$\log(\lambda_{i,t,r}) = \mu + \text{Replicat}_r + \text{Imbibition}_i + \text{Temperature}_t$$

		temperature		
		Elevated	Low	Medium
imb ibiti on	DS	3	3	3
	EI	3	3	3
	LI	3	3	3

DS : Dry Seed; EI : Erly Imb; LI : Late Imb.

Bio factor 1 : imbibition = {DS, EI, LI}
 Bio factor 2 : temperature = {Low, Medium, Elevate}
 Batch factor : replicat = {rep1, rep2, rep3}

* interactions between technical and biological factors are not considered

Statistical translation of the context

- Project Ecoseed (L. Rajjou & G. Cueff)
 - Study of seed germination and vigor in *Arabidopsis thaliana*
 - 3 omics (RNAseq, proteomics and metabolomics)

- Statistical modeling : **Generalized Linear Model**

-> For each gene

- Interaction between imbibition and temperature

$$\log(\lambda_{i,t,r}) = \mu + \text{Replicat}_r + \text{Imbibition}_i + \text{Temperature}_t + \text{Imbibition}_i : \text{Temperature}_t$$

- Statistical translation of the biological questions : contrasts

- 3 types of contrasts : sample (2 by 2), average (2nd factor), interaction effect.
- GLM_Contrasts function automatically writes a list of contrasts based on the model specified by the user (C. Paysant-Le-Roux)

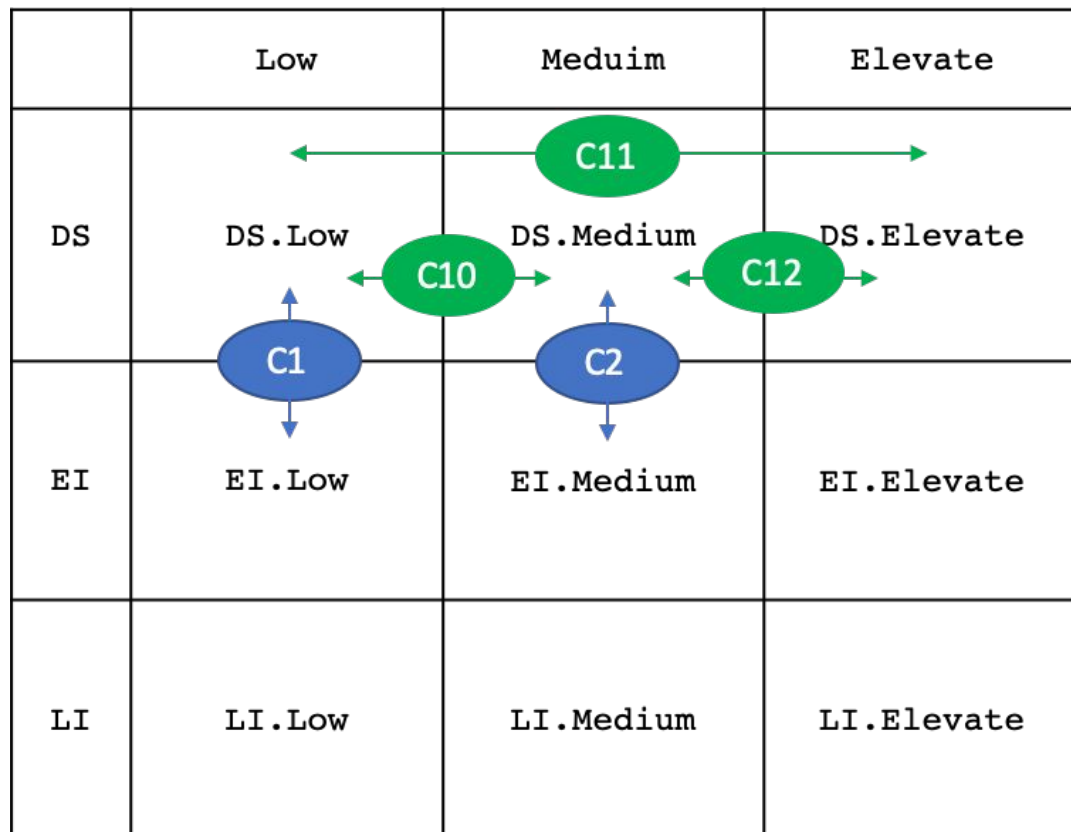
		temperature		
		Elevated	Low	Medium
imb ibiti on	DS	3	3	3
	EI	3	3	3
	LI	3	3	3

DS : Dry Seed; EI : Erly Imb; LI : Late Imb.

Statistical translation of the biological question

Simple contrasts : 2 by 2 comparisons

Ex. Which genes/proteins/metabolites are differentially expressed between DS and EI at Low temperature ?



Imbibition stage effect for each Temperature

C1: DS.Low - EI.Low

C2: DS.Medium - EI.Medium

...

Temperature effect at each imbibition stage

C10: DS.Low - DS.Medium

C11: DS.Medium - DS.Elevate

C12: DS.Low - DS.Elevate

...

18 a priori hypothesis can be formulated

Statistical translation of the biological question

Averaged contrasts

Ex. Which genes/proteins/metabolites are differentially expressed between the 2 imbibition stages DS, EI ?

	Low	Meduim	Elevate
DS	$\frac{DS.Low + DS.Medium + DS.Elevate}{3}$		
EI	$\frac{EI.Low + EI.Medium + EI.Elevate}{3}$		
LI	LI.Low	LI.Medium	LI.Elevate

Imbibition stage effect

C1: (DS - EI) in mean Temperature

C2: (EI - LI) in mean Temperature

...

Temperature effect

C1: (Low - Medium) in mean imbibition stage

C2: (Medium - Elevate) in mean imbibition stage

...

...

6 a priori hypothesis can be formulated

Statistical translation of the biological question

Interaction contrasts

Ex. Is the imbibition stage effect at Low temperature different from those at Medium temperature ?

	Low	Medium	Elevate
DS	DS.Low	DS.Medium	DS.Elevate
EI	EI.Low	EI.Medium	EI.Elevate
LI	LI.Low	LI.Medium	LI.Elevate

Interaction effect

$$C1.2: (DS.Low - EI.Low) - (DS.Medium - EI.Medium)$$

$$C2.3: (DS.Medium - EI.Medium) - (DS.Elevate - EI.Elevate)$$

...

Interaction effect

$$C1.2: (DS.Low - DS.Medium) - (EI.Low - EI.Medium)$$

$$C2.3: (DS.Medium - EI.Medium) - (DS.Elevate - EI.Elevate)$$

...

12 a priori hypothesis can be formulated

~/Documents/INRA/IJPB/PROJETS/IJPB-Bioinfo/FLOMICs/R-Flomics/rflomics/inst/RFLOMICSapp - Shiny

http://127.0.0.1:3540 | Open in Browser | Publish

RFLOMICS

- Presentation
- Experimental Design
- Import design
- Statistical model

Generate report

Select a model formulae

- Repeat + temperature + imbibition
- Repeat + temperature + imbibition + temperature:imbibition**
- Repeat + imbibition
- Repeat + temperature

Valid model choice

Contrast type : simple

- (temperatureLow - temperatureElevated) in imbibitionLI
- (temperatureMedium - temperatureElevated) in imbibitionLI
- (temperatureMedium - temperatureLow) in imbibitionLI
- (temperatureLow - temperatureElevated) in imbibitionEI
- (temperatureMedium - temperatureElevated) in imbibitionEI
- (temperatureMedium - temperatureLow) in imbibitionEI
- (temperatureLow - temperatureElevated) in imbibitionDS
- (temperatureMedium - temperatureElevated) in imbibitionDS
- (temperatureMedium - temperatureLow) in imbibitionDS
- (imbibitionEI - imbibitionLI) in temperatureElevated
- (imbibitionEI - imbibitionLI) in temperatureLow
- (imbibitionEI - imbibitionLI) in temperatureMedium
- (imbibitionDS - imbibitionLI) in temperatureElevated
- (imbibitionDS - imbibitionLI) in temperatureLow
- (imbibitionDS - imbibitionLI) in temperatureMedium
- (imbibitionDS - imbibitionEI) in temperatureElevated
- (imbibitionDS - imbibitionEI) in temperatureLow
- (imbibitionDS - imbibitionEI) in temperatureMedium

Contrast type : averaged

- (temperatureLow - temperatureElevated) in mean
- (temperatureMedium - temperatureElevated) in mean
- (temperatureMedium - temperatureLow) in mean
- (imbibitionEI - imbibitionLI) in mean
- (imbibitionDS - imbibitionLI) in mean
- (imbibitionDS - imbibitionEI) in mean

Contrast type : interaction

- (temperatureLow - temperatureElevated) in imbibitionEI - (temperatureLow - temperatureElevated) in imbibitionLI
- (temperatureMedium - temperatureElevated) in imbibitionEI - (temperatureMedium - temperatureElevated) in imbibitionLI
- (temperatureMedium - temperatureLow) in imbibitionEI - (temperatureMedium - temperatureLow) in imbibitionLI
- (temperatureLow - temperatureElevated) in imbibitionDS - (temperatureLow - temperatureElevated) in imbibitionLI
- (temperatureMedium - temperatureElevated) in imbibitionDS - (temperatureMedium - temperatureElevated) in imbibitionLI
- (temperatureMedium - temperatureLow) in imbibitionDS - (temperatureMedium - temperatureLow) in imbibitionLI
- (temperatureLow - temperatureElevated) in imbibitionDS - (temperatureLow - temperatureElevated) in imbibitionEI
- (temperatureMedium - temperatureElevated) in imbibitionDS - (temperatureMedium - temperatureElevated) in imbibitionEI

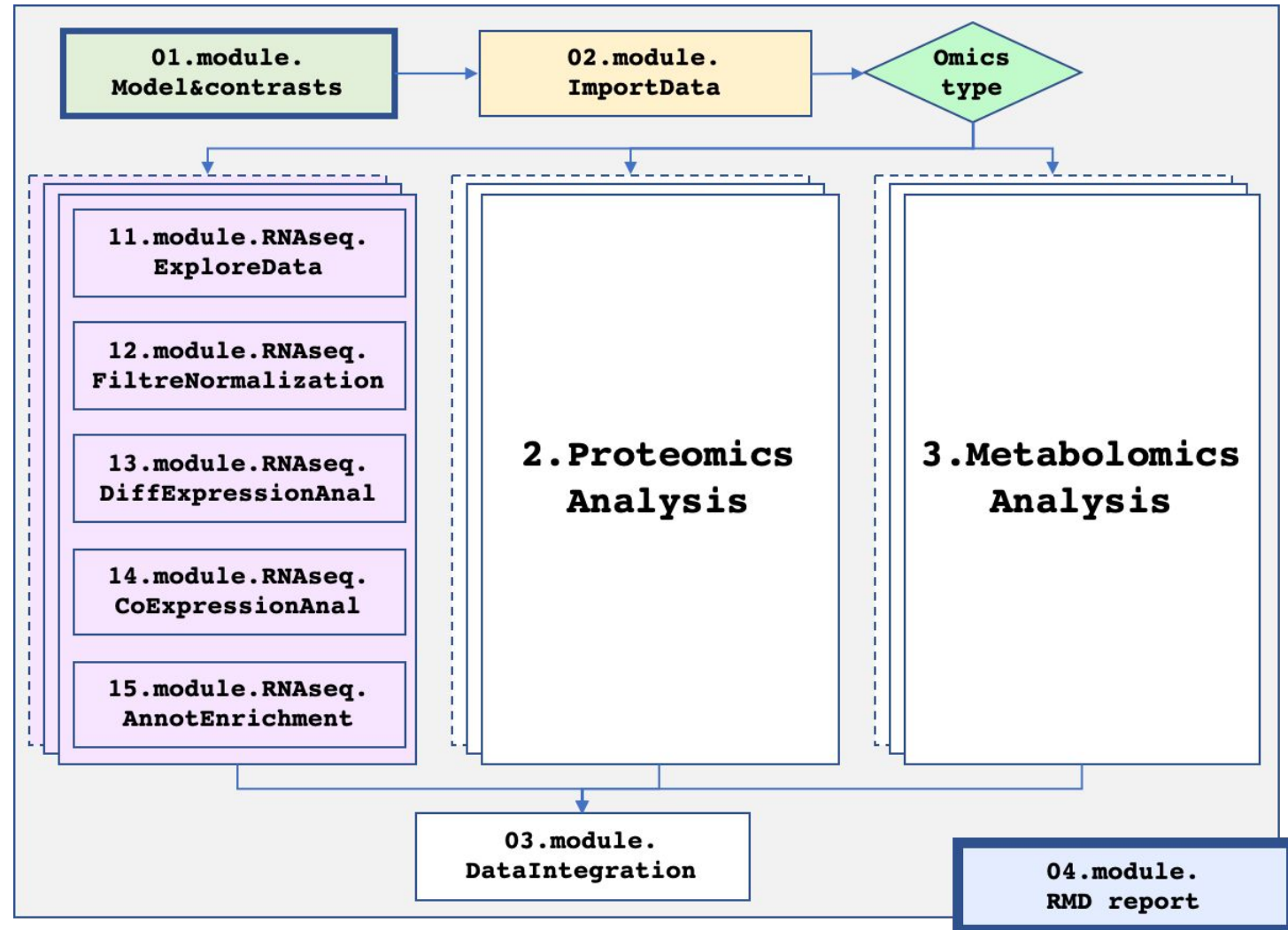
Workflow

*RNAseq Analysis**

***DiCoExpress**

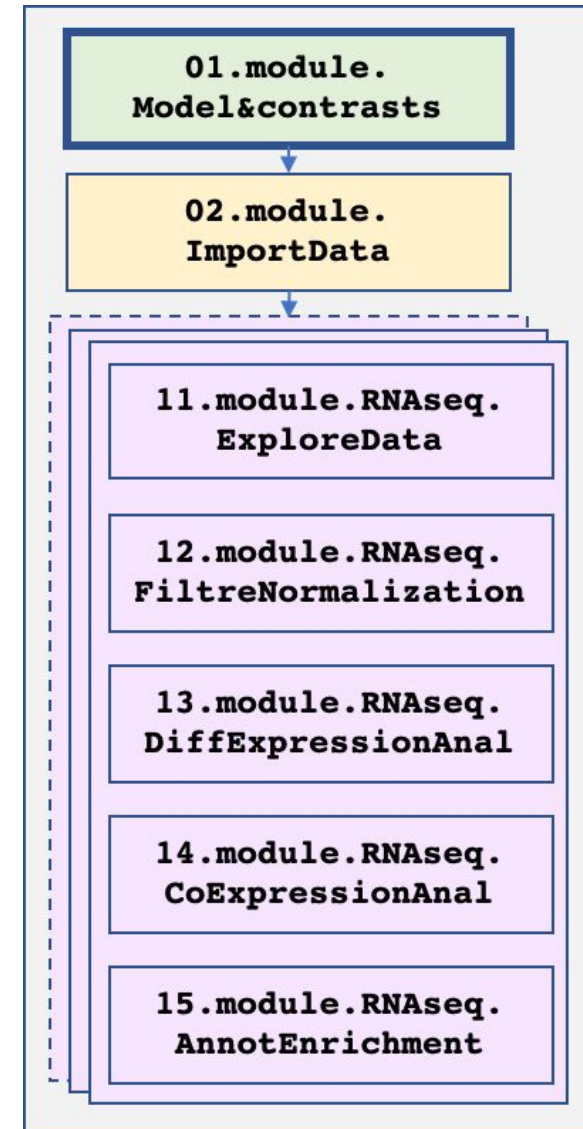
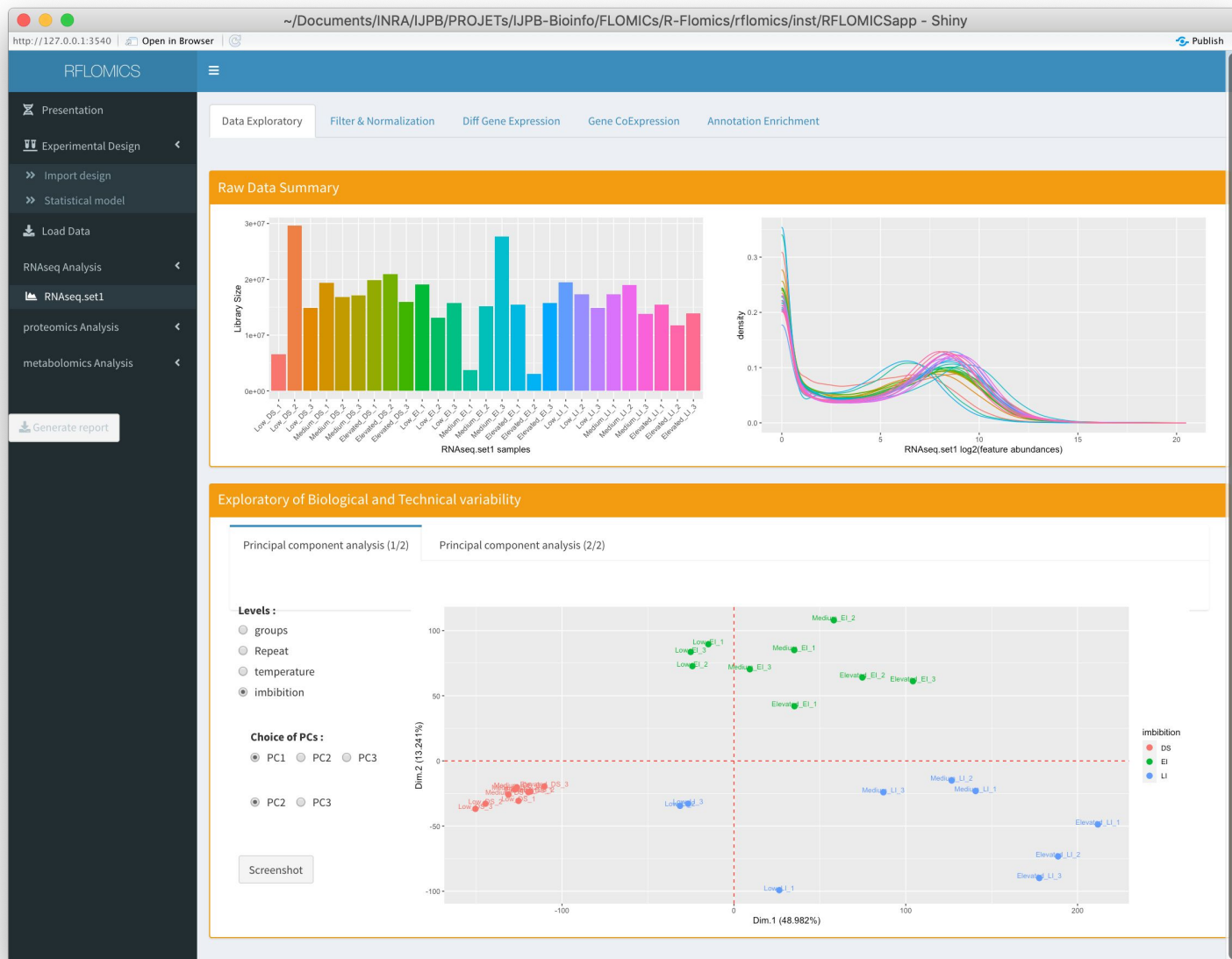
(I. Lambert and C. Paysant-Le Roux, 2020)
IPS2 (GeNet)

edgeR (McCarthy DJ et al, 2012, Nucleic Acids Research)
coseq (Rau A, et al, 2018, Bioinformatics)



RNAseq Analysis workflow

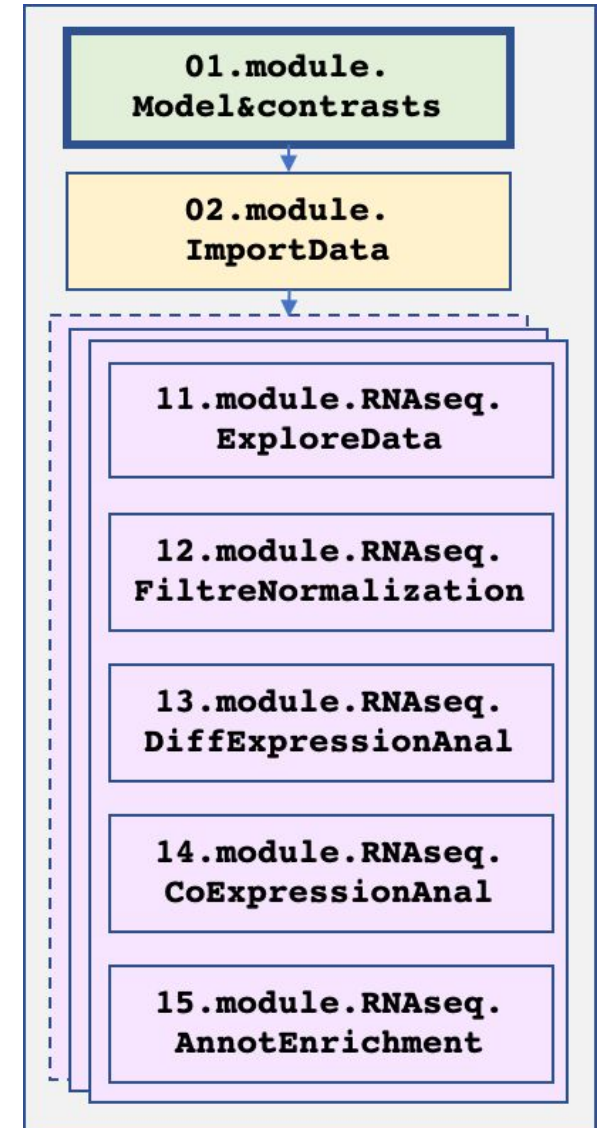
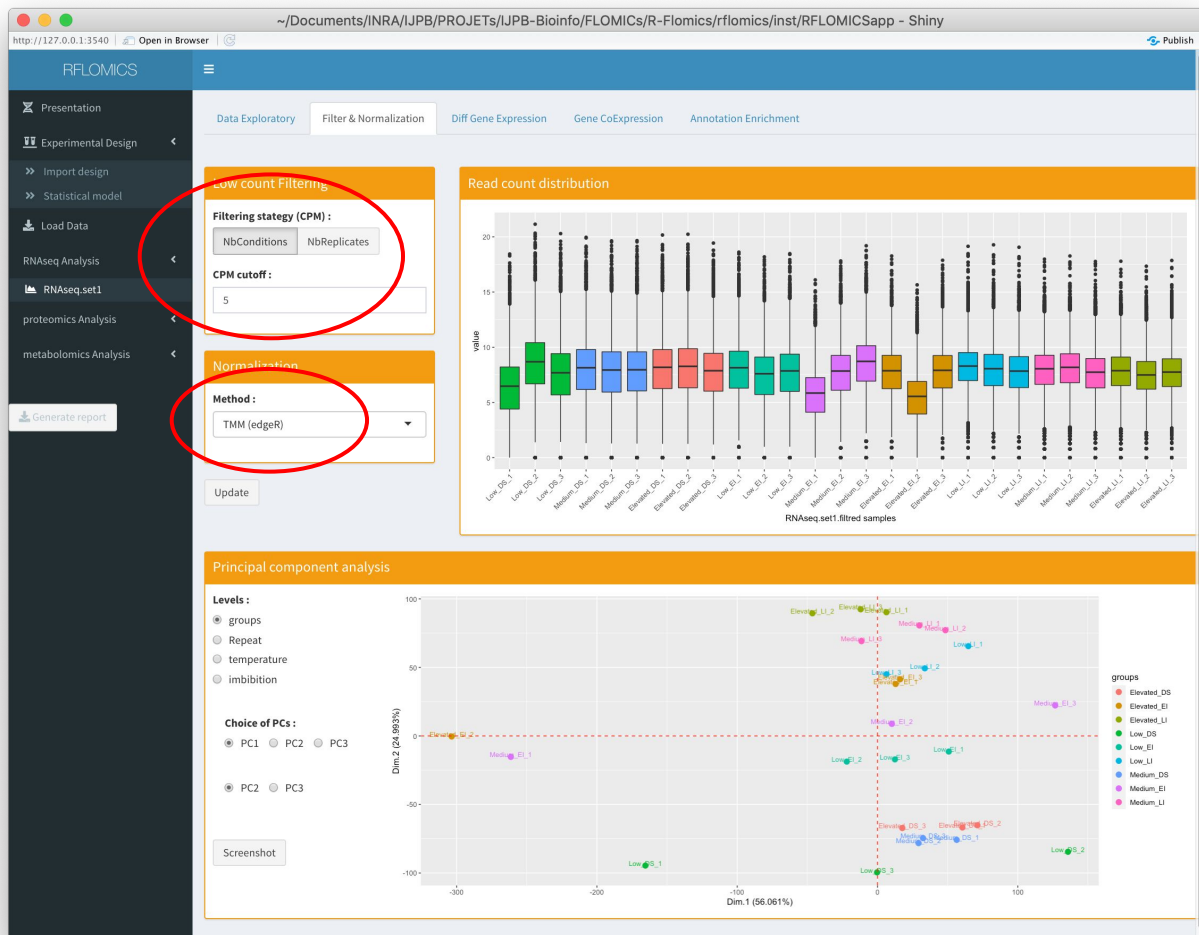
Quality control and data exploratory



RNAseq Analysis workflow

Filtering and Normalization / QC

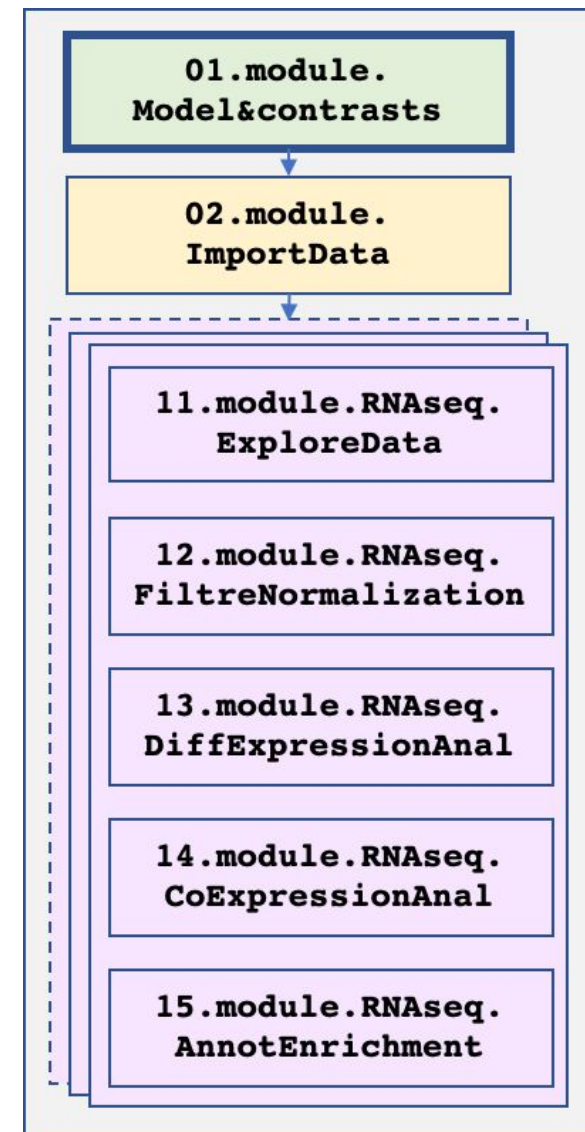
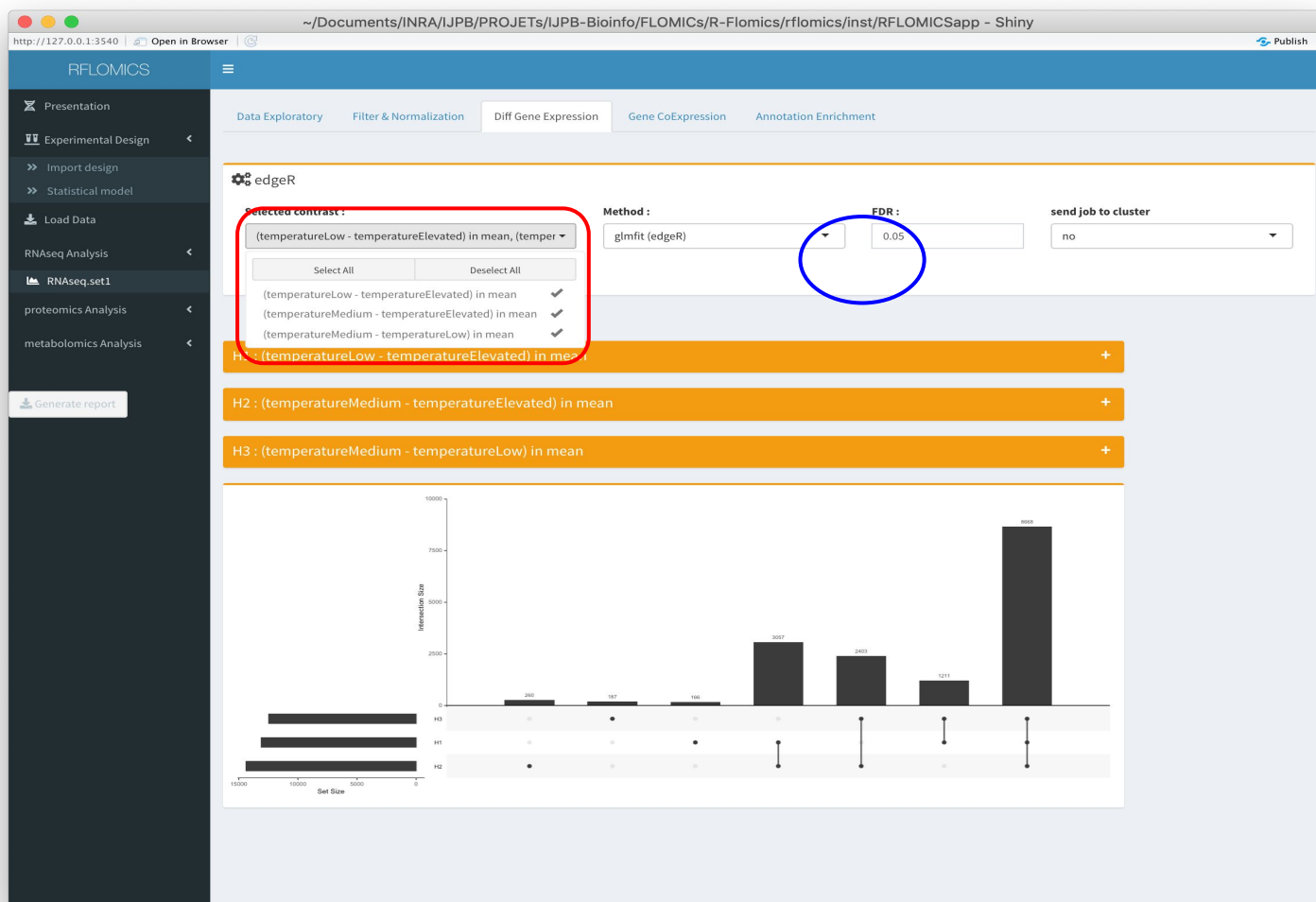
- Filtering : Counts per million (CPM) method
- Normalization : Trimmed Mean of M-values (TMM) from edgeR package



RNAseq Analysis workflow

Differential Expression Analysis

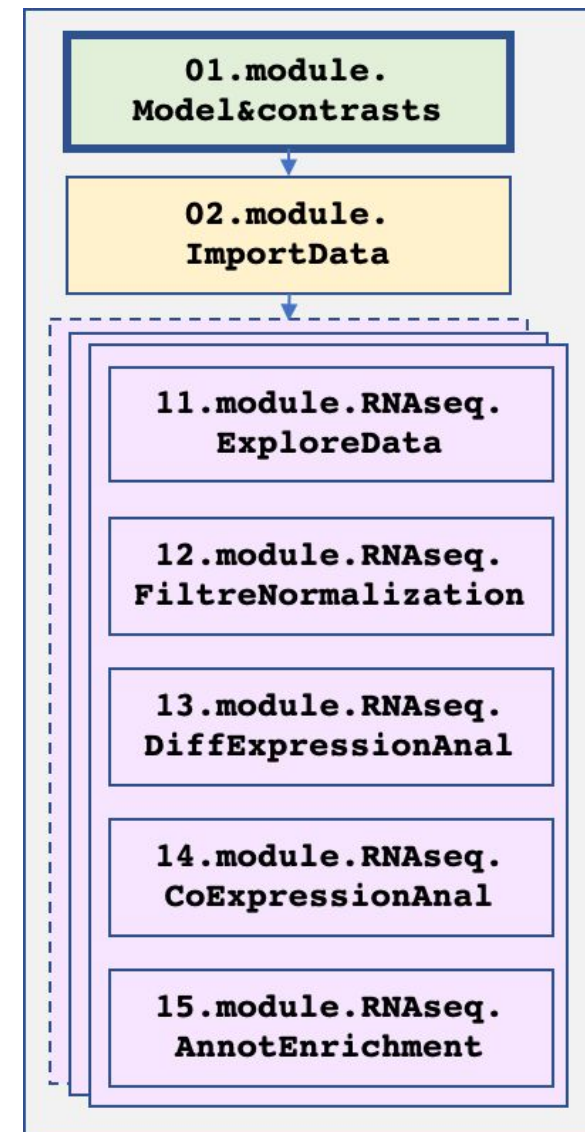
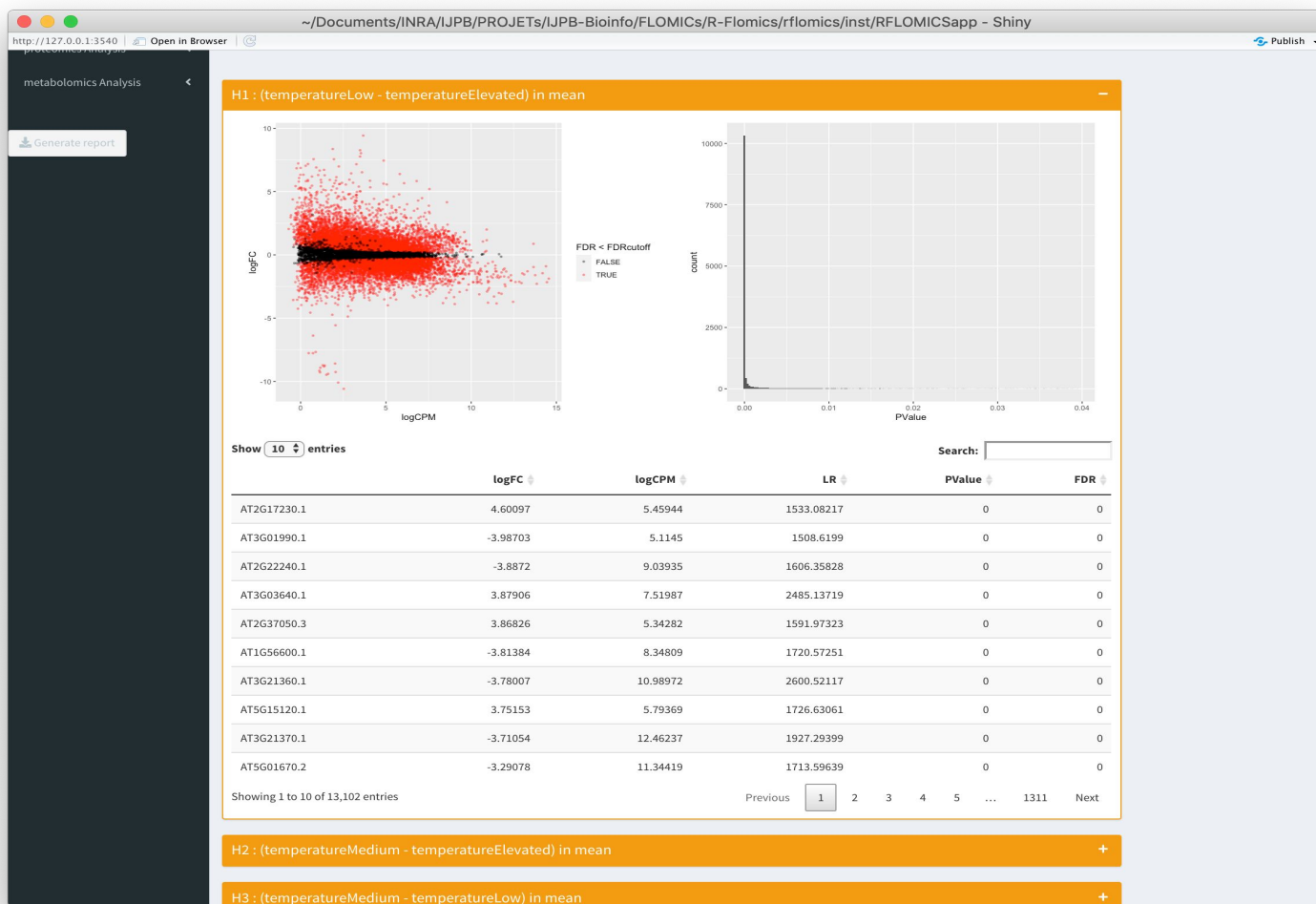
- Tests performed by edgeR : Negative Binomial models



RNAseq Analysis workflow

Differential Expression Analysis

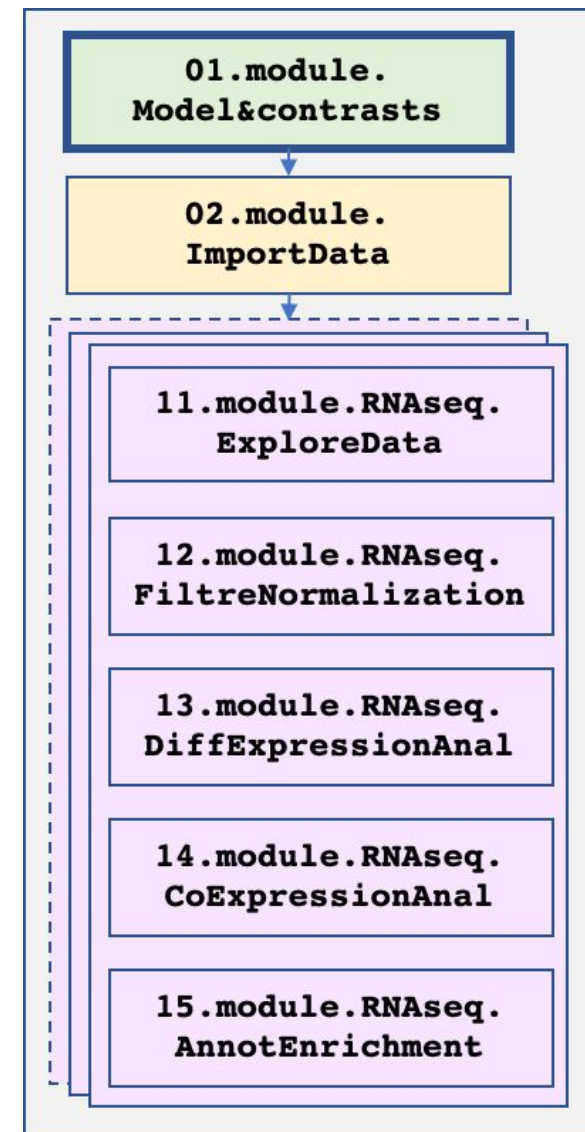
- Tests performed by edgeR : Negative Binomial models



RNAseq Analysis workflow

Differential Expression Analysis

- Tests performed by edgeR : Negative Binomial models

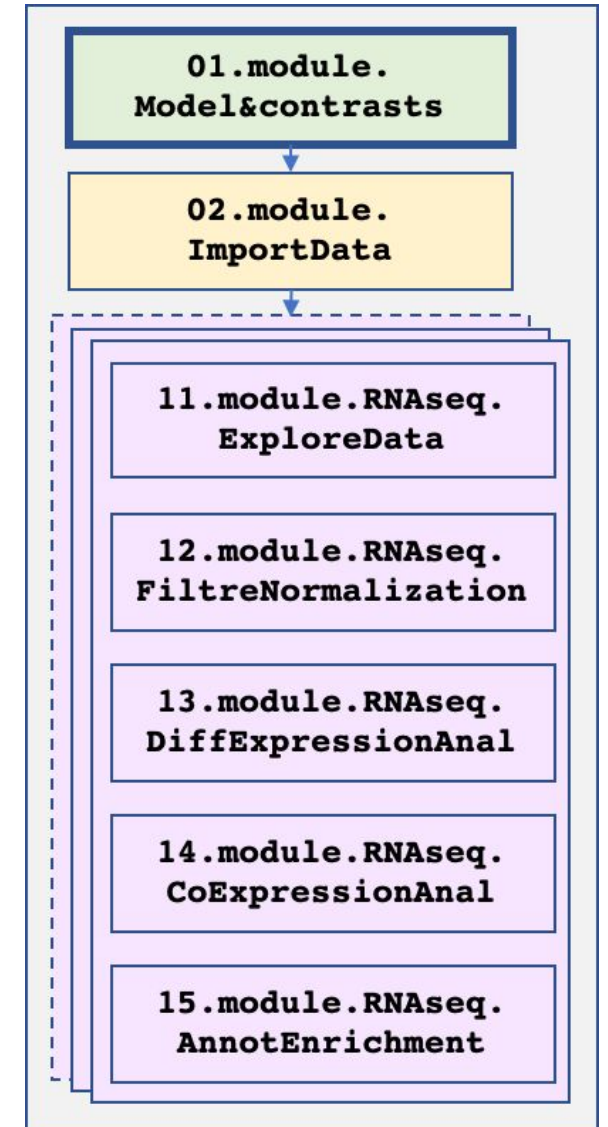
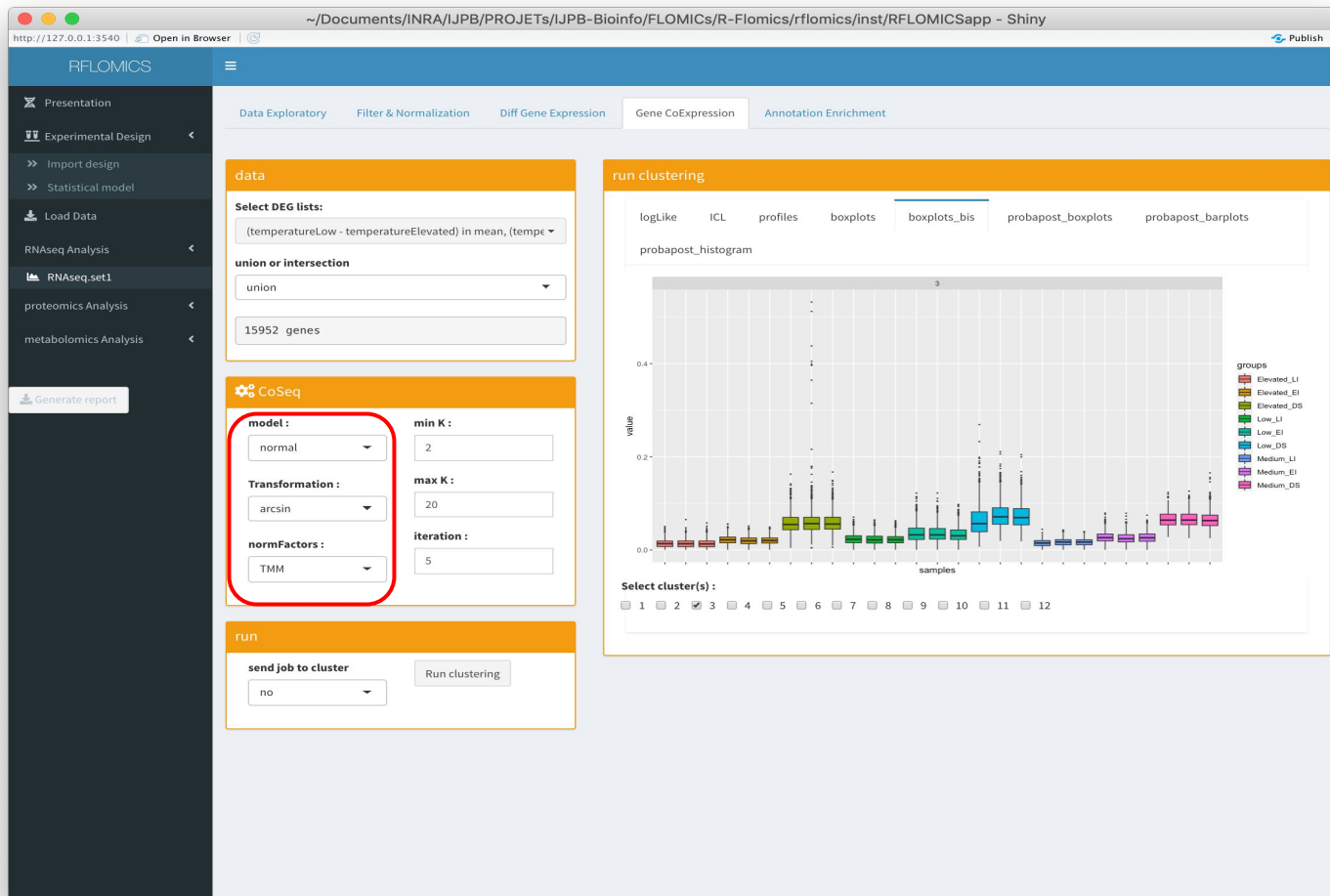


RNAseq Analysis workflow

Co-Expression analysis

Rau et al., 2015, *Bioinformatics*

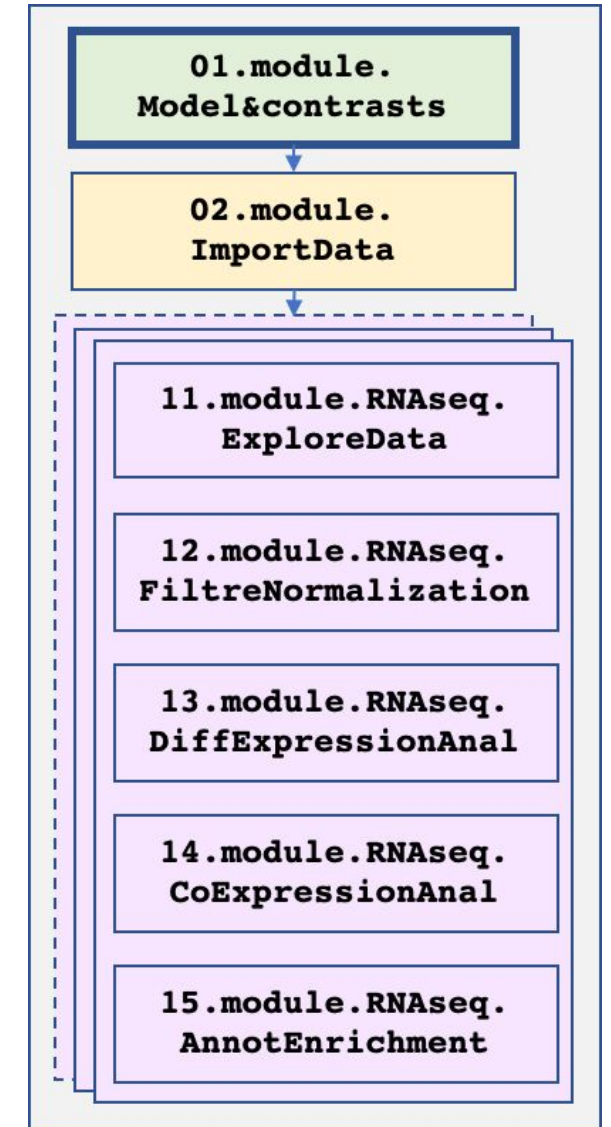
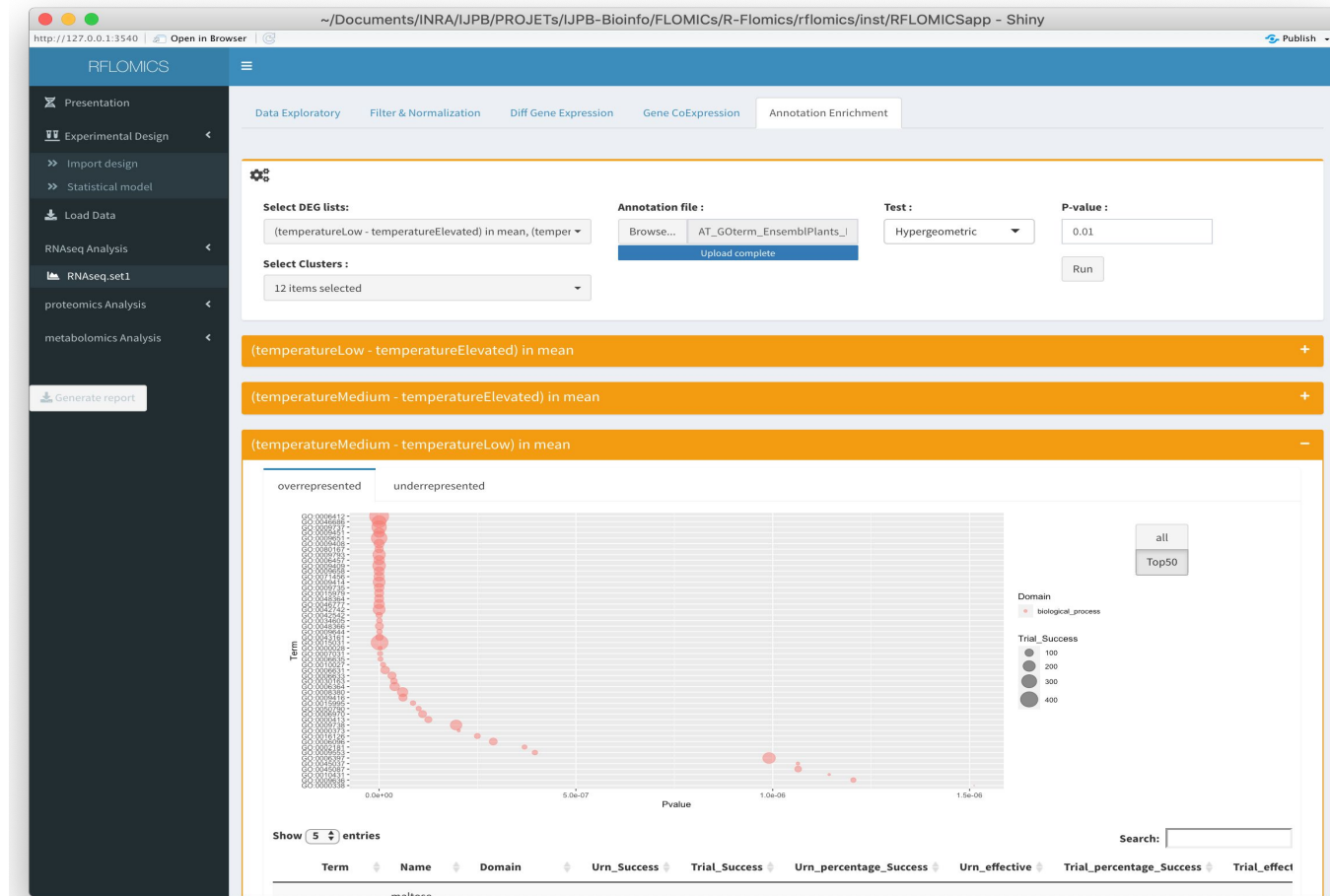
- coseq : Gaussian mixture model
- Input = validated DEG lists (union or intersection)
- Fixed parameter : model=normal; transformation=arcsin; normalization=TMM



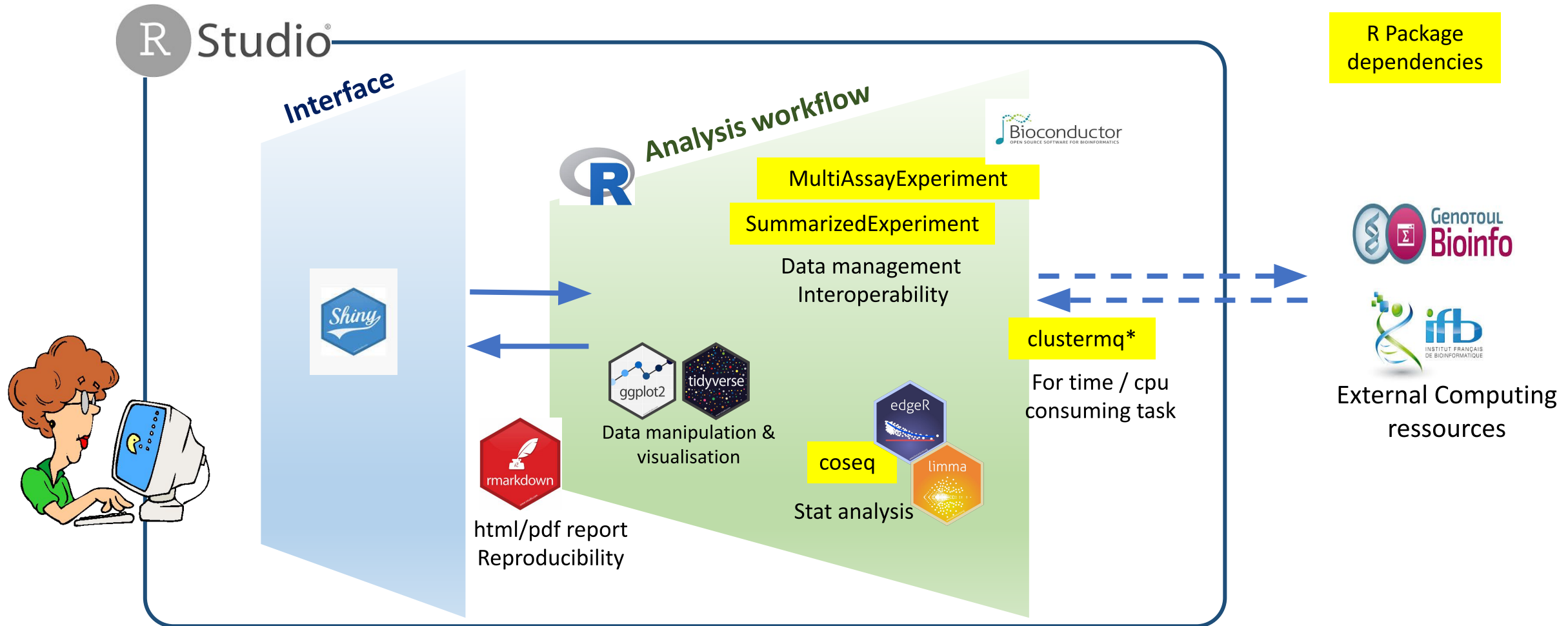
RNAseq Analysis workflow

Enrichment Annotation Analysis

- Test : Hypergeometric
- Input : validated DEG lists, cluster lists
- Reference : gene annotation (gene ID, annotationTerm, Domain).Ex. GO, KEGG...



RFLOMICS : R package + Shiny application



*M Schubert. clustermq enables efficient parallelisation of genomic analyses. *Bioinformatics* (2019)

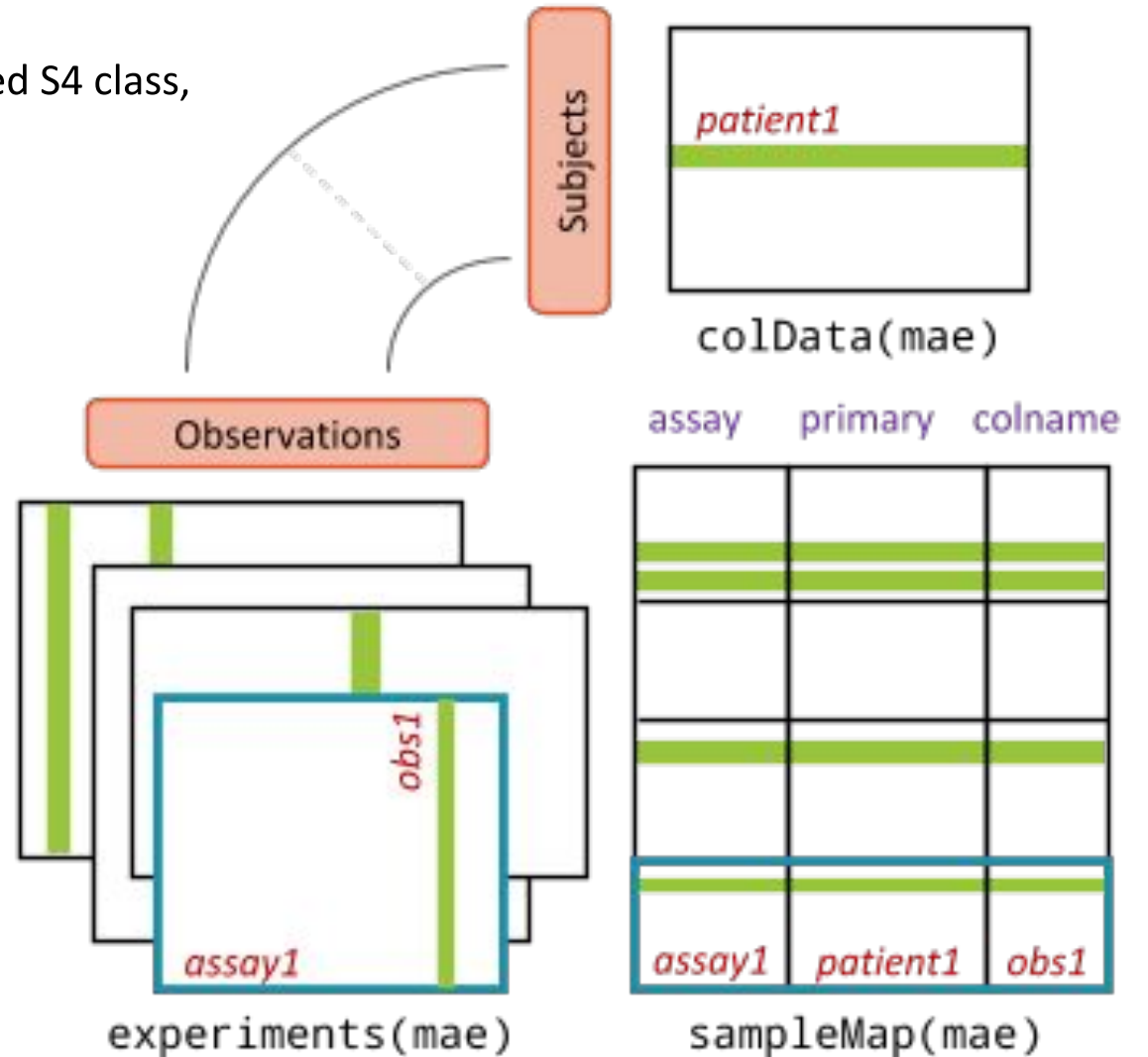
RFLOMICS : R package *workflow*

- R package (Rstudio, roxygen, ...)
- Object-oriented programming
 - For all analysis, one object are created from MultiAssayExperiment S4 class
 - All functionalities are implemented as generic methods for this class.

MultiAssayExperiment

multi-omics/multi-data

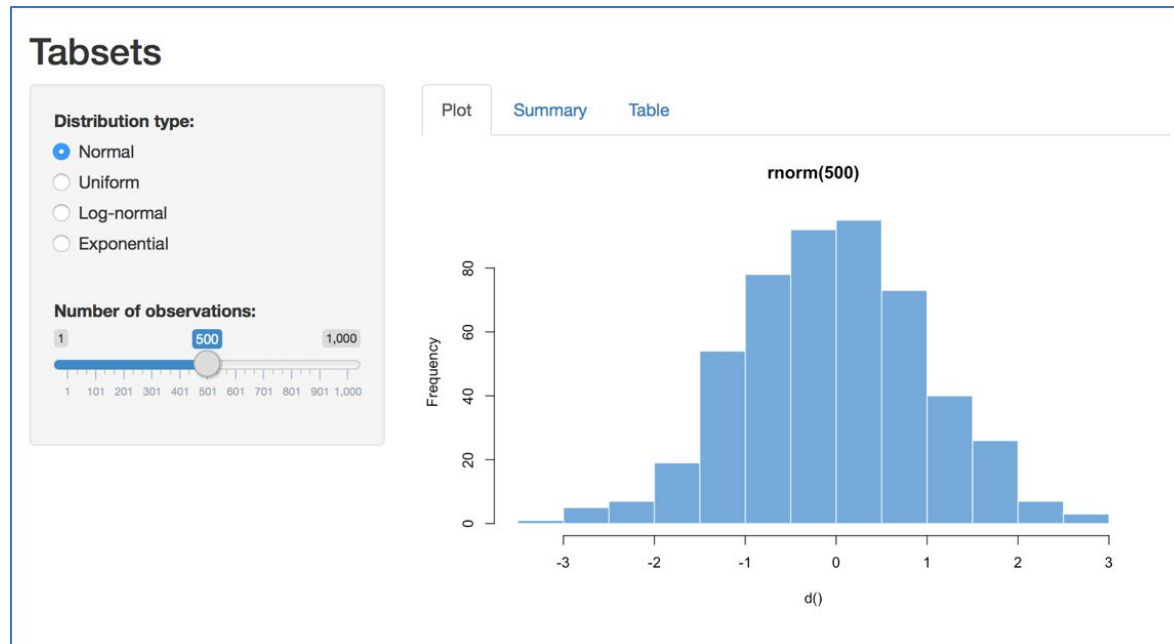
- MultiAssayExperiment introduces a Bioconductor object-oriented S4 class, defining a general data structure for representing multi-omics experiments.
- **colData (data.frame)**
 - experimental plan / sample sheet
- **experiments (list)**
 - omics data (summarized experiments)
- **sampleMap (data.frame)**
 - correspondence sample names - platform
- **metadata (list)**
 - project names
 - model formulae
 - selected contrasts
 - ...



RFLOMICS : Shiny application

Interface

- Shiny is an R package, developed by RStudio, that makes it easy to build interactive web applications straight from R



APP TEMPLATE

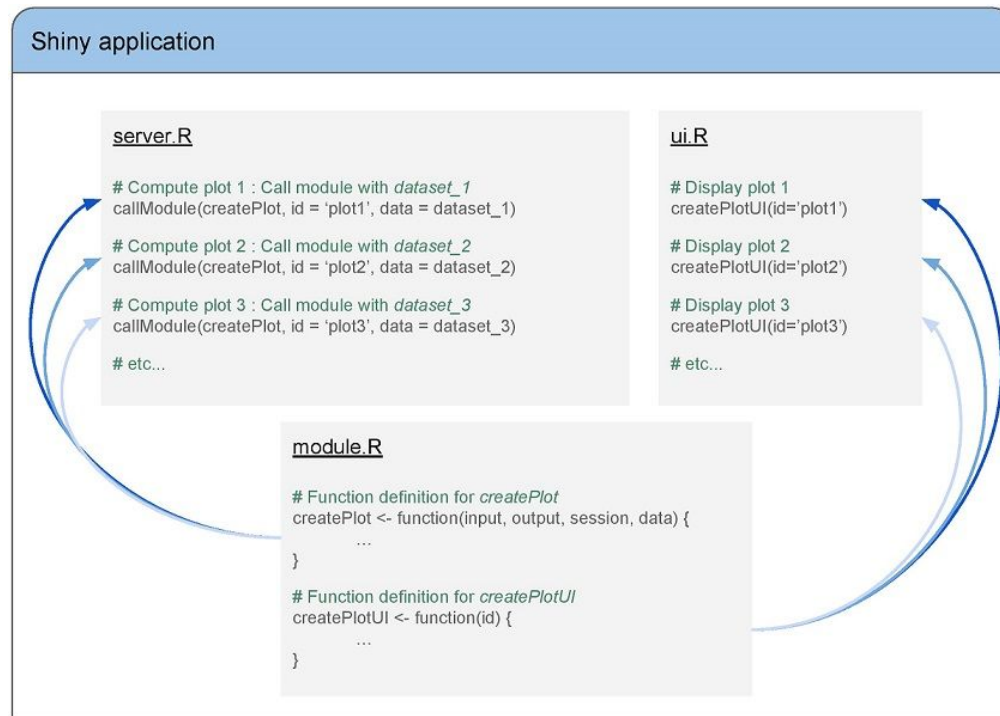
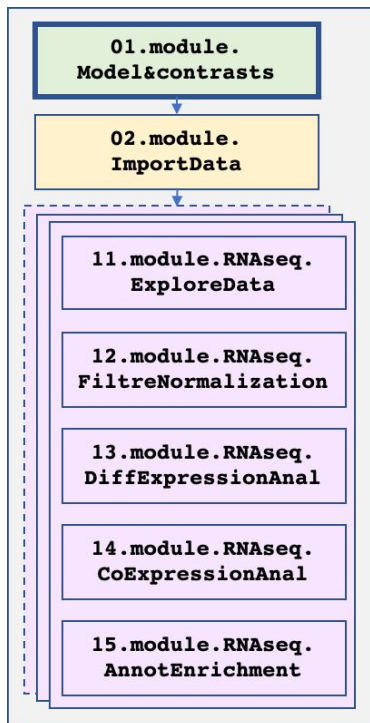
Begin writing a new app with this template. Preview the app by running the code at the R command line.



```
library(shiny)
ui <- fluidPage()
server <- function(input, output){}
shinyApp(ui = ui, server = server)
```

RFLOMICS : Shiny application *Interface*

- Shiny is an R package, developed by RStudio, that makes it easy to build interactive web applications straight from R
- Shiny modules :



APP TEMPLATE

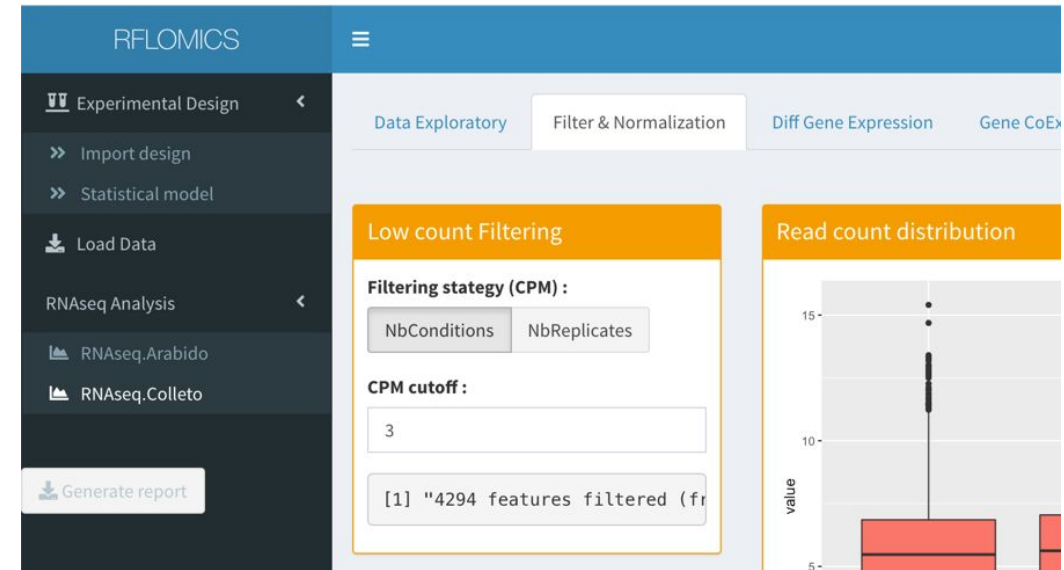
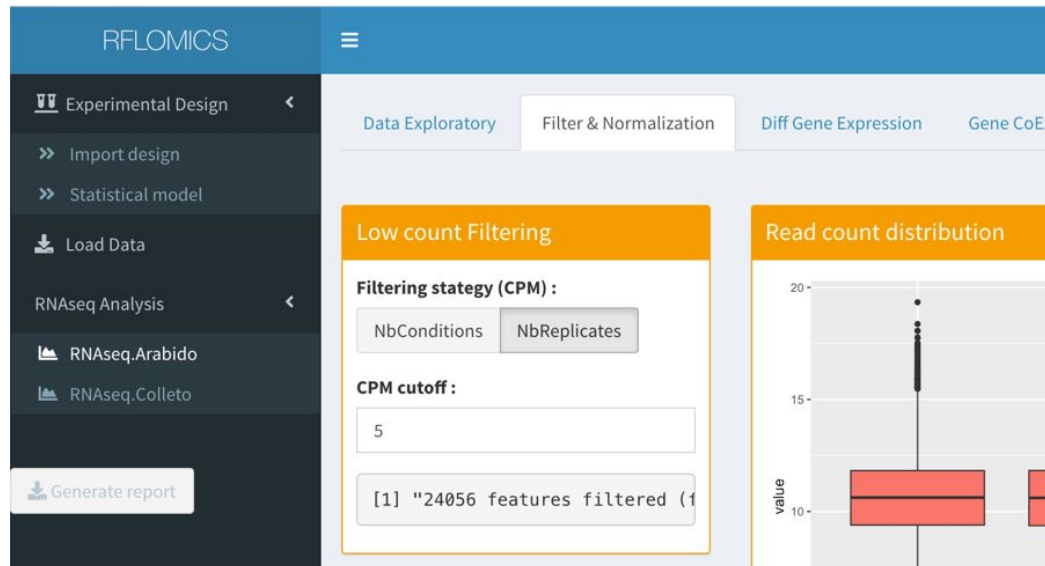
Begin writing a new app with this template. Preview the app by running the code at the R command line.



```
library(shiny)
ui <- fluidPage()
server <- function(input, output){}
shinyApp(ui = ui, server = server)
```

RFLOMICS : Shiny application *Interface*

- Shiny is an **R package**, developed by RStudio, that makes it easy to build interactive web applications straight from R
- Shiny modules
 - interoperability
 - simultaneous analysis



RFLOMICS : Shiny application

Interface

- Shiny is an **R package, developed by RStudio**, that makes it easy to build interactive web applications straight from R
- Shiny modules
 - interoperability
 - simultaneous analysis
 - facilitates development
- Bookmarking / repeatable
 - share
 - reproducibility
- Deployment
 - locally -> docker
 - Shiny server (INRAE)

APP TEMPLATE

Begin writing a new app with this template. Preview the app by running the code at the R command line.



```
library(shiny)
ui <- fluidPage()
server <- function(input, output){}
shinyApp(ui = ui, server = server)
```

RFLOMICS : Shiny application

RMD report

Using chunk option child for Markdown

<https://yihui.org/knitr/demo/child/>

Using chunk option `child` for Markdown (.Rmd) documents

The chunk option `child` supports the inclusion of child documents. For example here is how to combine a main document with a child document. We create a child document named `knitr-child.Rmd`:

```
# Details from the child document
```

```
Hi, there. I'm a child with a plot and as many details as necessary.
```

```
```{r test-child}  
plot(trees)
```
```

and we create a main document named `main.Rmd`

```
---  
title: "Main report"  
output: html_document  
---
```

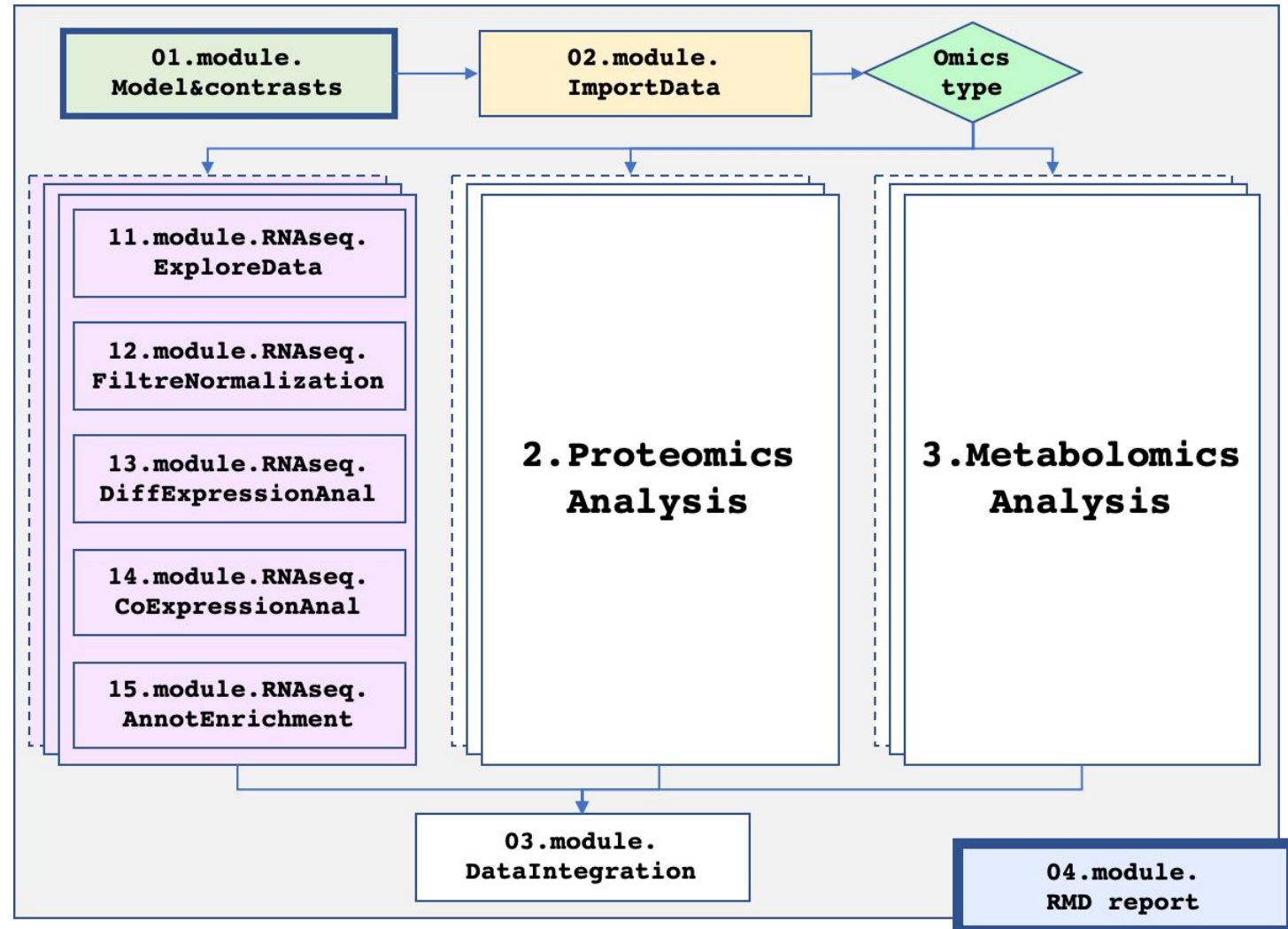
```
# Summary
```

```
```{r test-main, child = 'knitr-child.Rmd'}  
```
```

```
# Conclusion
```

RFLOMICS Outputs

- RMD report
 - Flexibility
 - Reproducibility
 - used methods & parameters
 - cmd line
 - results summary
- .Rdata = MAE object
 - Experimental Design
 - For each omics/dataset
 - raw data
 - results
 - used methods
 - chosen parameters
- Result tables



Perspectives

A lot of work...

Improvement (tests, documents..., docker, Bookmarking)

RNAseq analysis (others functionalities...)

Proteomics/Metabolomics data analysis...

Any contribution is welcome...

Beta-testers are welcome...

Incoming features

Extending RFLOMICS to multi-omics data integration methods



Merci!