RFLOMICS : Interactive web application for Omics-data analysis

(IJPB, GNET team, CATI Sysmics)

(Delphine Charif, Nadia Bessoltane,
Christine Paysant-Le-Roux, Gwendal Cueff)

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

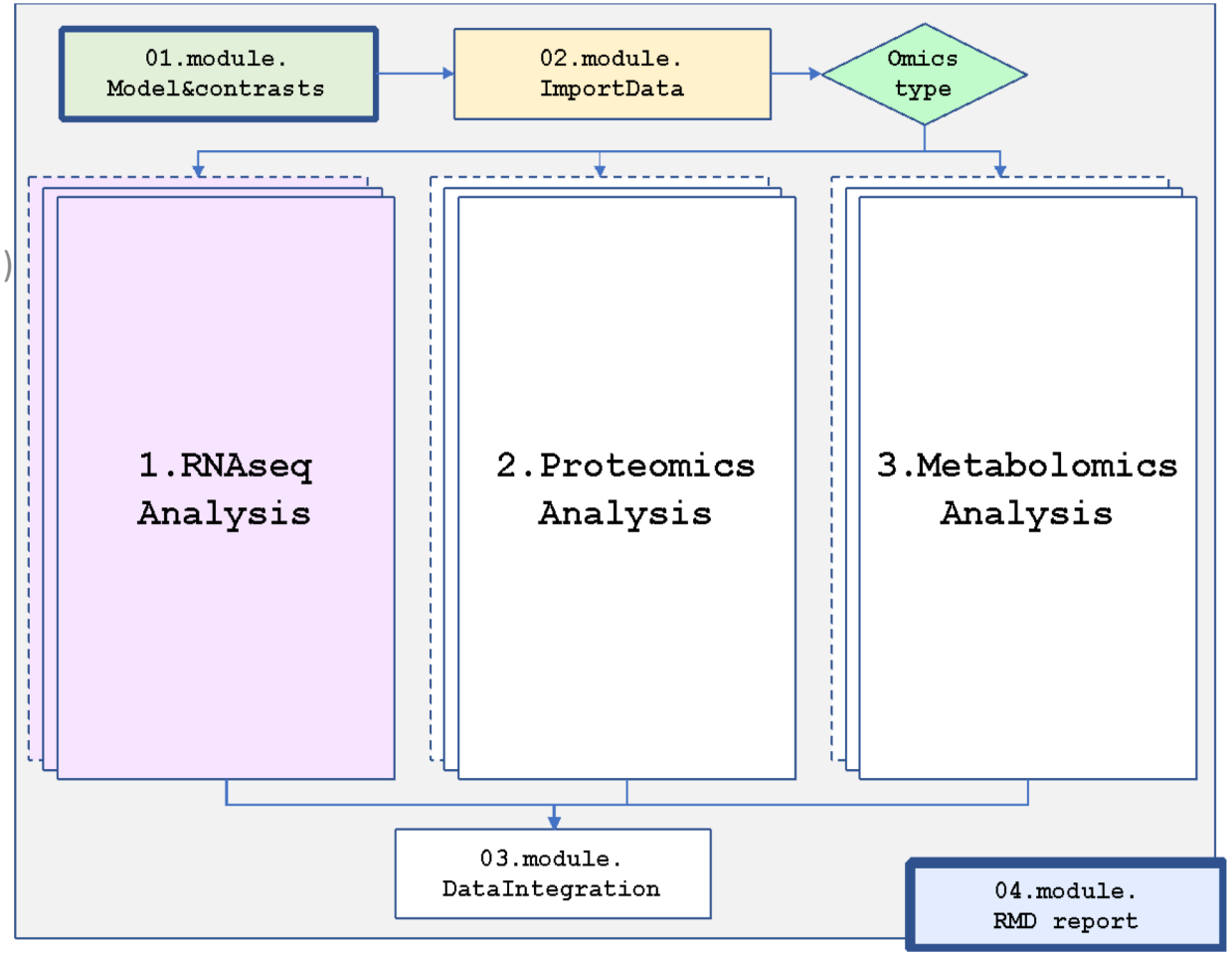
- **Model & contrasts**
  - Statistical translation of the context
  - Multi-factorial statistics model

- **ImportData :**
  - Support 3 types of omics (RNAseq, proteomics, metabolomics)
  - and multi dataset for each omics.
    → from same biological context

- **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)

	temperature			
imbibition		Elevated	Low	Medium
	DS	3	3	3
	EI	3	3	3
	LI	3	3	3

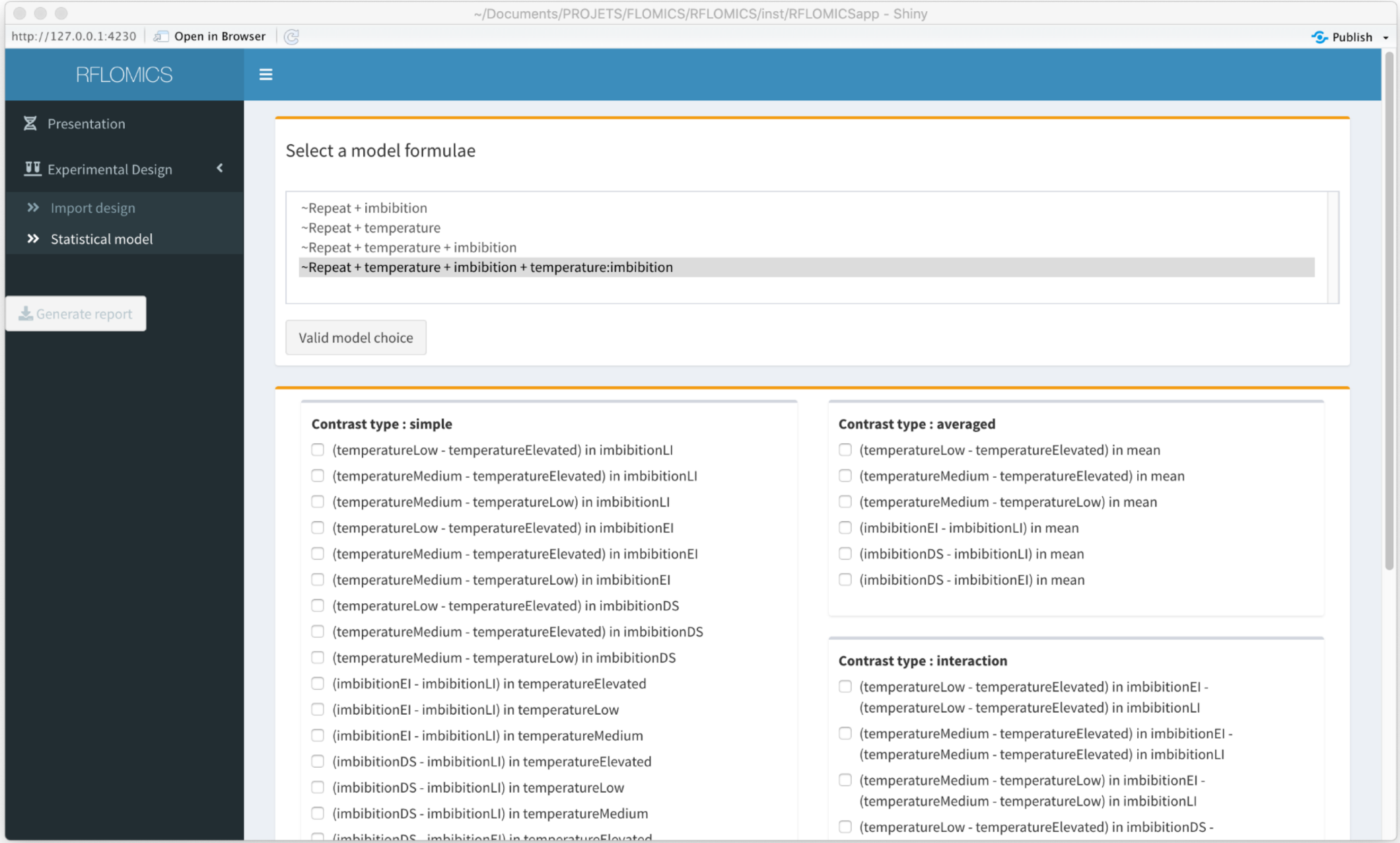
DS : Dry Seed; EI : Erly Imb; LI : Late Imb.

- Statistical modeling : ***Generalized Linear Model/Linear Model***

  -> for each gene

$$\log(\lambda_{i,t,r}) = \mu + Replicat_r + Imbibition_i + Temperature_t + Imbibition_i : Temperature_t$$

- Statistical translation of the biological questions : contrasts
  - 3 types of contrasts : sample (2 by 2), average (2nd factor), interaction effect.
  - The list of contrasts is generated automatically based on the model specified by the user (C. Paysant-Le-Roux)


~/Documents/PROJETS/FLOMICS/RFLOMICS/inst/RFLOMICSapp - Shiny
http://127.0.0.1:4230
Open in Browser
Publish
RFLOMICS
Presentation
Experimental Design
Import design
Statistical model
Generate report
Select a model formulae
~Repeat + imbibition
~Repeat + temperature
~Repeat + temperature + imbibition
~Repeat + temperature + imbibition + temperature:imbibition
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
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 -

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

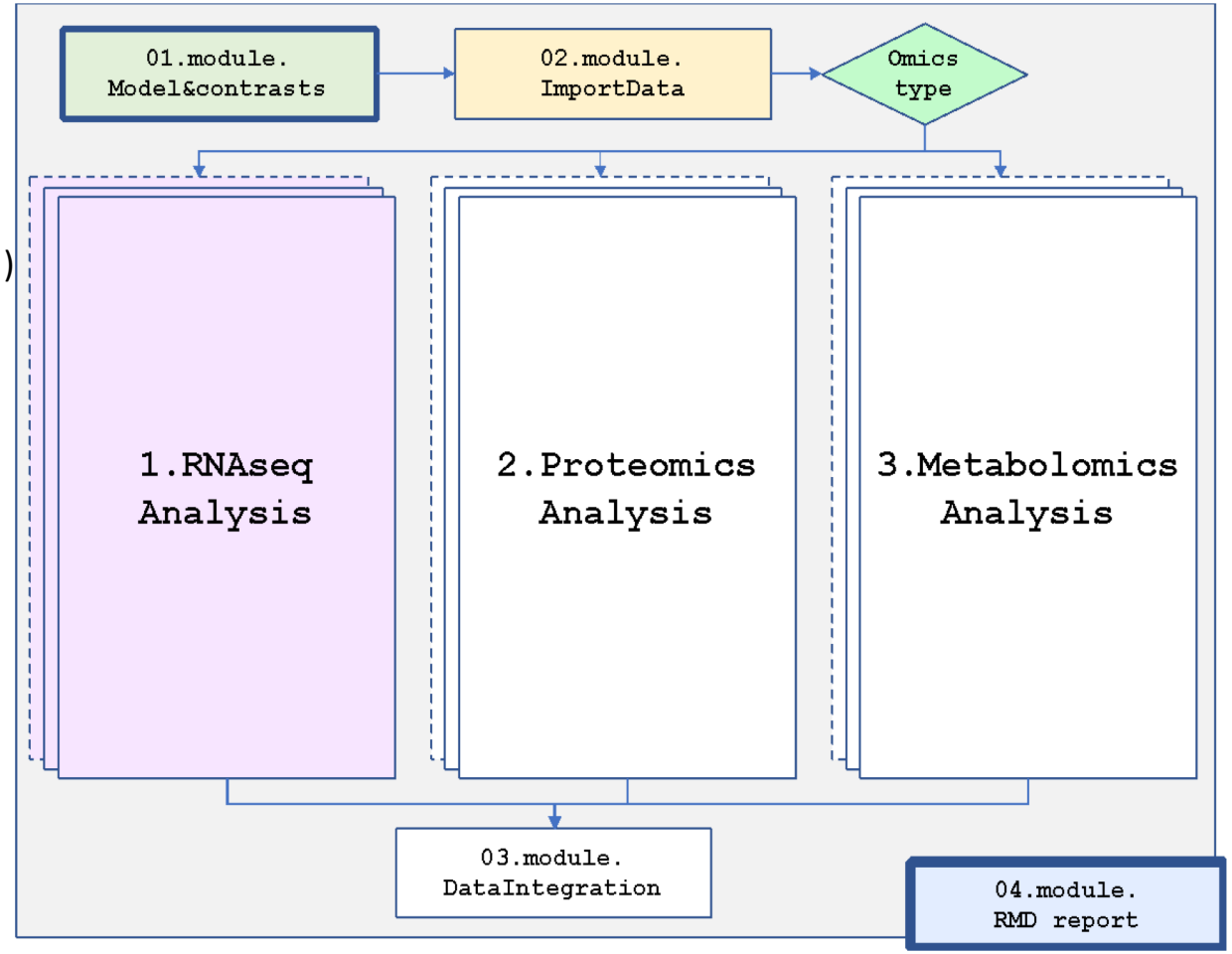
- Model & contrasts
  - Statistical translation of the context
  - Multi-factorial statistics model

- **ImportData :**
  - Support 3 types of omics (RNAseq, proteomics, metabolomics)
  - and multi dataset for each omics.
    → from same biological context

- **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**



Workflow : RNAseq Analysis

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

ExploreData:

- QC

FilterNormalization:

- CPM, TMM

DiffExpressionAnalysis:

edgeR (McCarthy DJ et al, 2012, NAR)

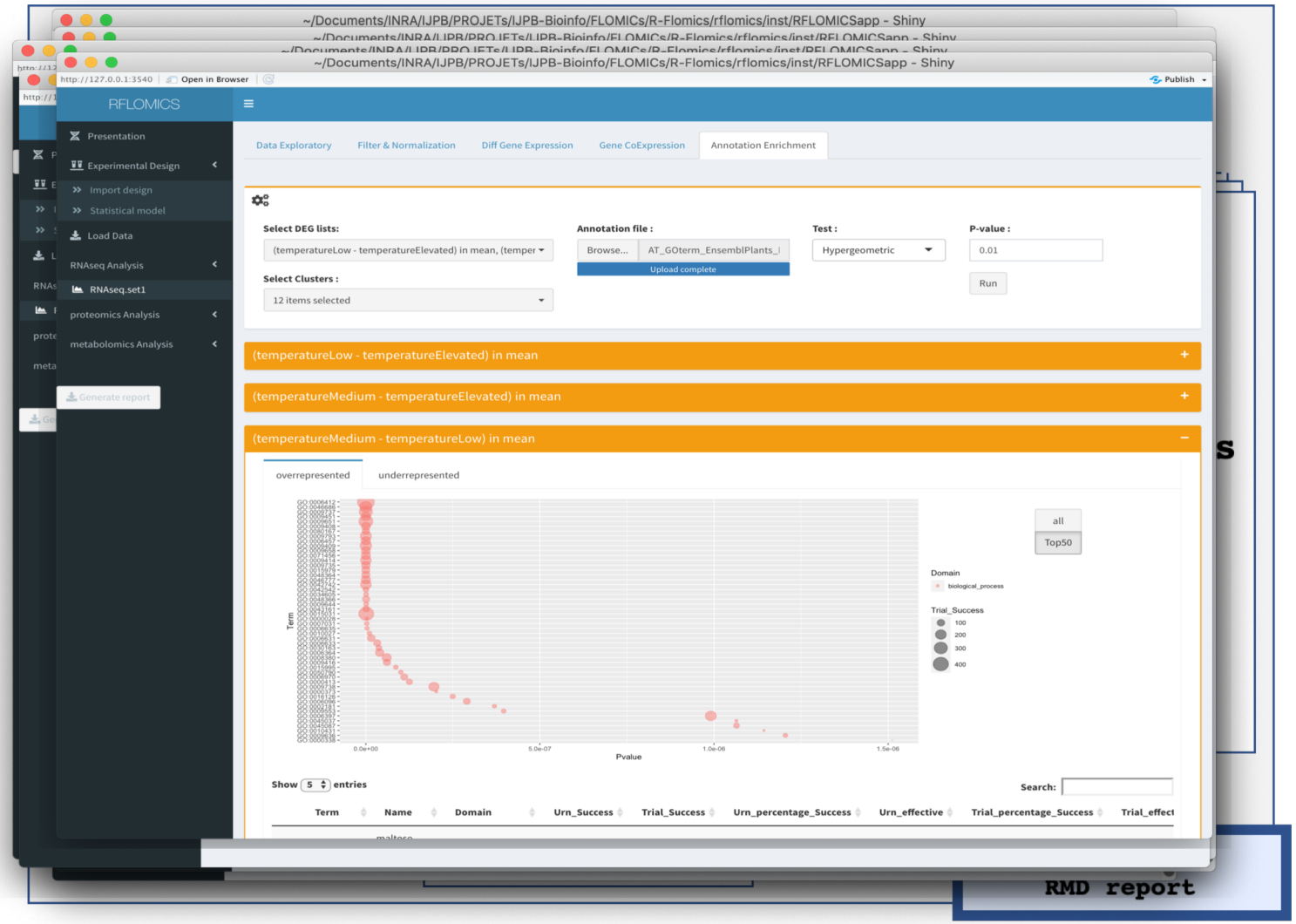
Co-Expression Analysis:

coseq (Rau A, et al, 2018, Bioinformatics)

- Gaussian mixture model
- Fixed parameters :
  - Transformation=’arcsin’
  - Normalization=’TMM’

Annotation & Enrichment:

- Ref : GO, KEGG…
- Test hypergeometric



Workflow : Metabolomics & Proteomics* Analysis

* Gwendal internship at Pasteur Institute, Papsso, ProteoSign (Efstathiou G, et al, 2017, NAR)

ExploreData & transformation:

- QC
- log2 intensity transformation

→ start with pre-processed matrices

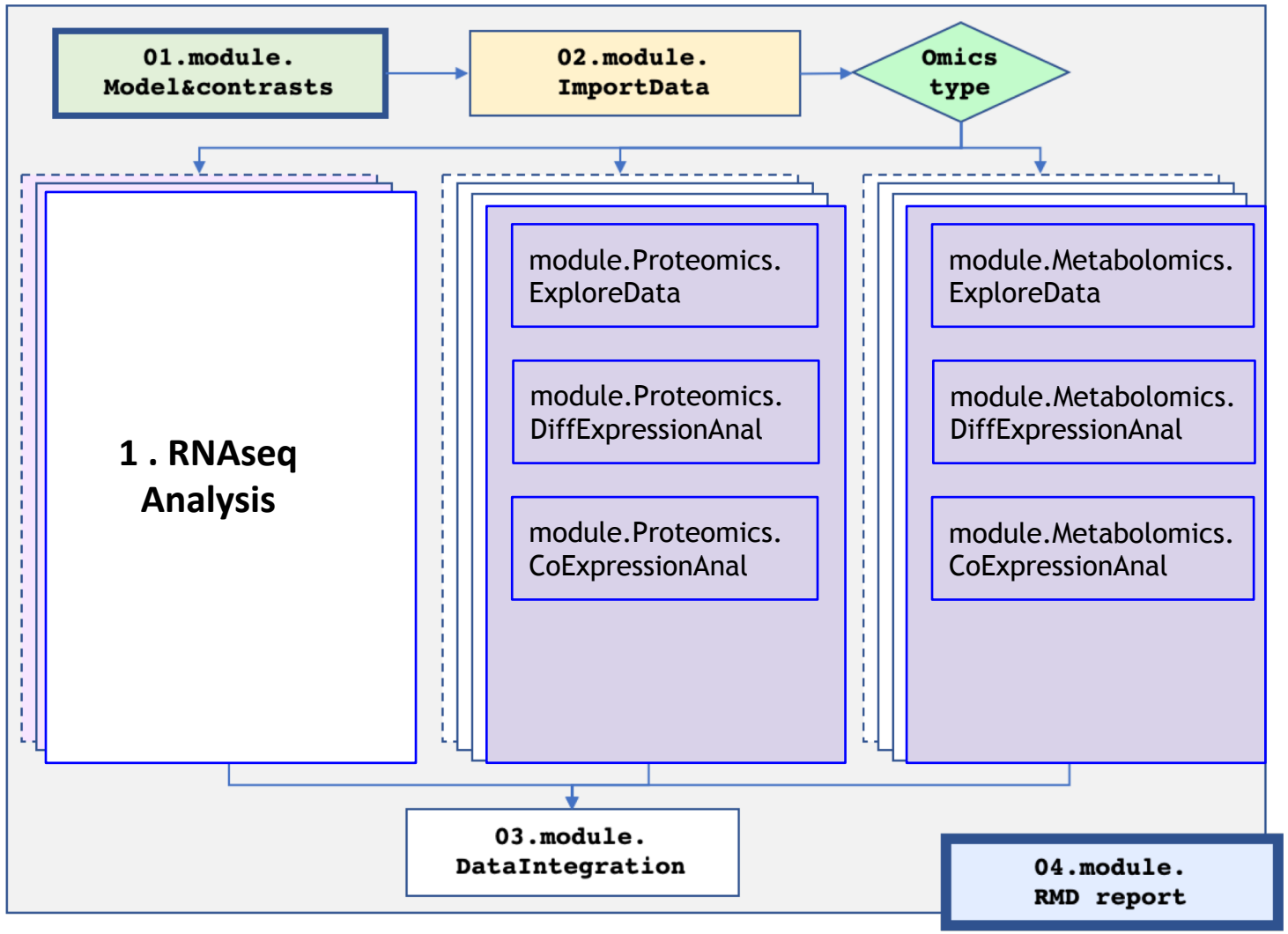
DiffExpressionAnalysis:

limma (Ritchie ME, et al, 2015, NAR)

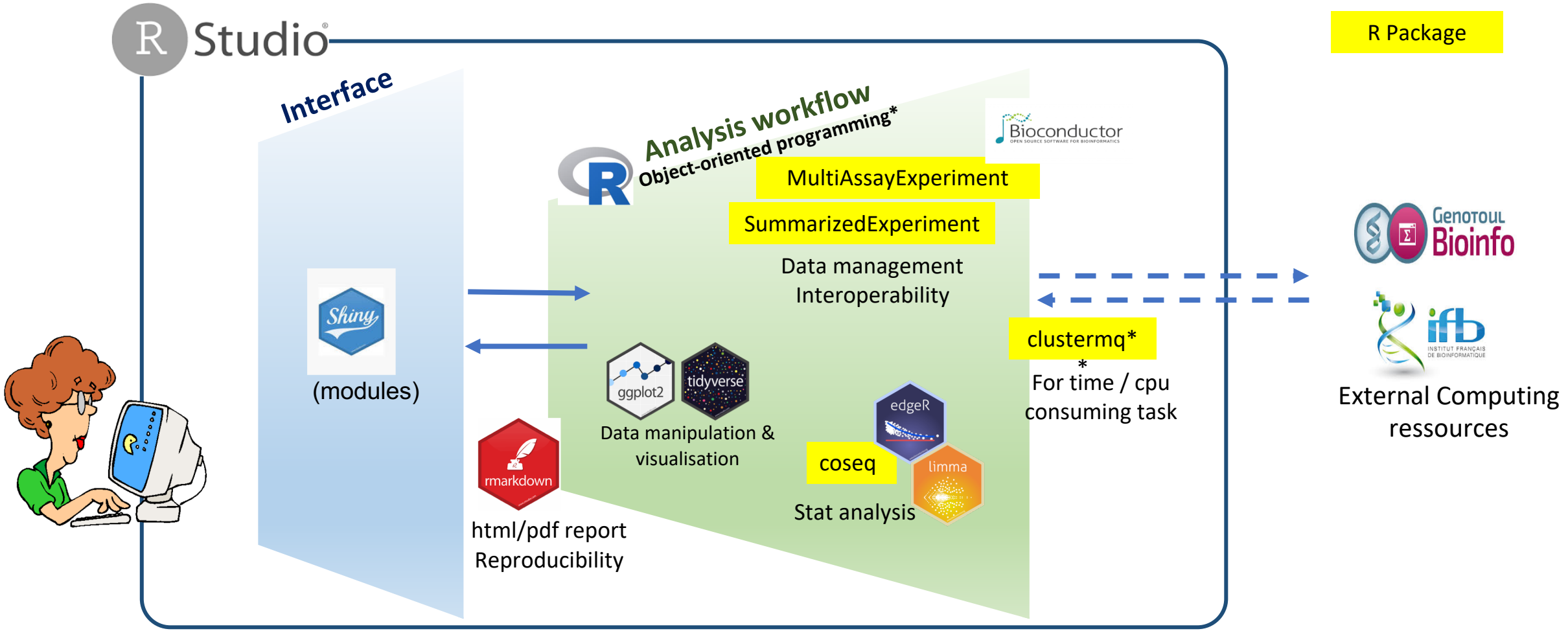
Co-Expression Analysis:

coseq (Rau A, et al, 2018, Bioinformatics)

- Gaussian mixture model
- Fixed parameters :
  - Data scaled by proteins/metabolites
  - Transformation=none
  - Normalization=none



RFLOMICS : R package + Shiny application



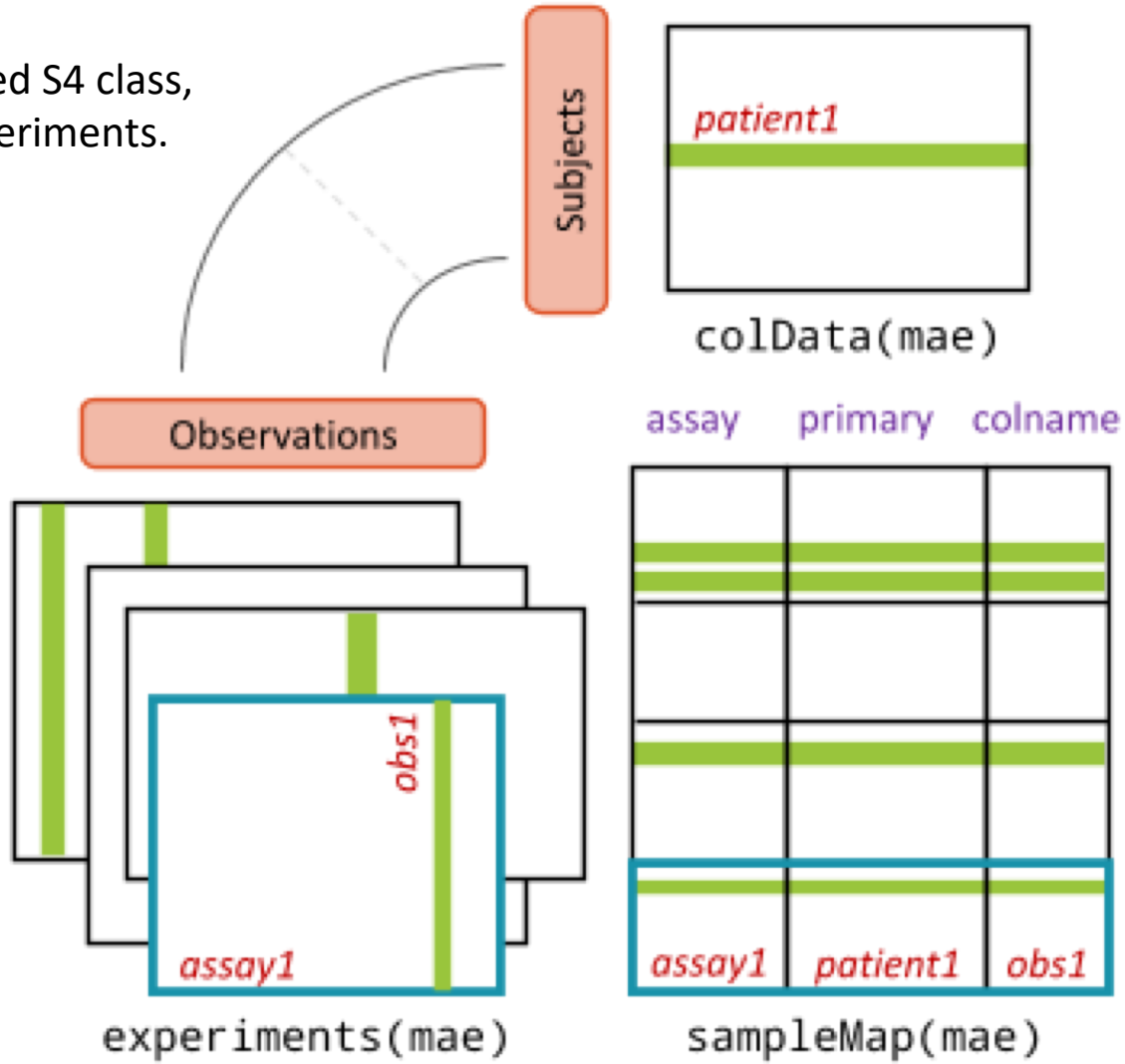
*For all analysis, one object are created from MultiAssayExperiment S4 class, All functionalities are implemented as generic methods for this class.

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

MultiAssayExperiment (MAE)
multi-omics/multi-data

- MultiAssayExperiment introduces a Bioconductor object-oriented S4 class, defining a general data structure for managing multi-omics experiments.

  - **colData (data.frame)**
    - experimental plan / sample sheet
  - **experiments (list)**
    - omics data (**summarized experiments**)
      - raw data & results
      - used methods & chosen parameters
  - **sampleMap (data.frame)**
    - correspondence sample names - platforme
  - **metadata (list)**
    - project names
    - model formulae
    - selected contrasts
    - ...



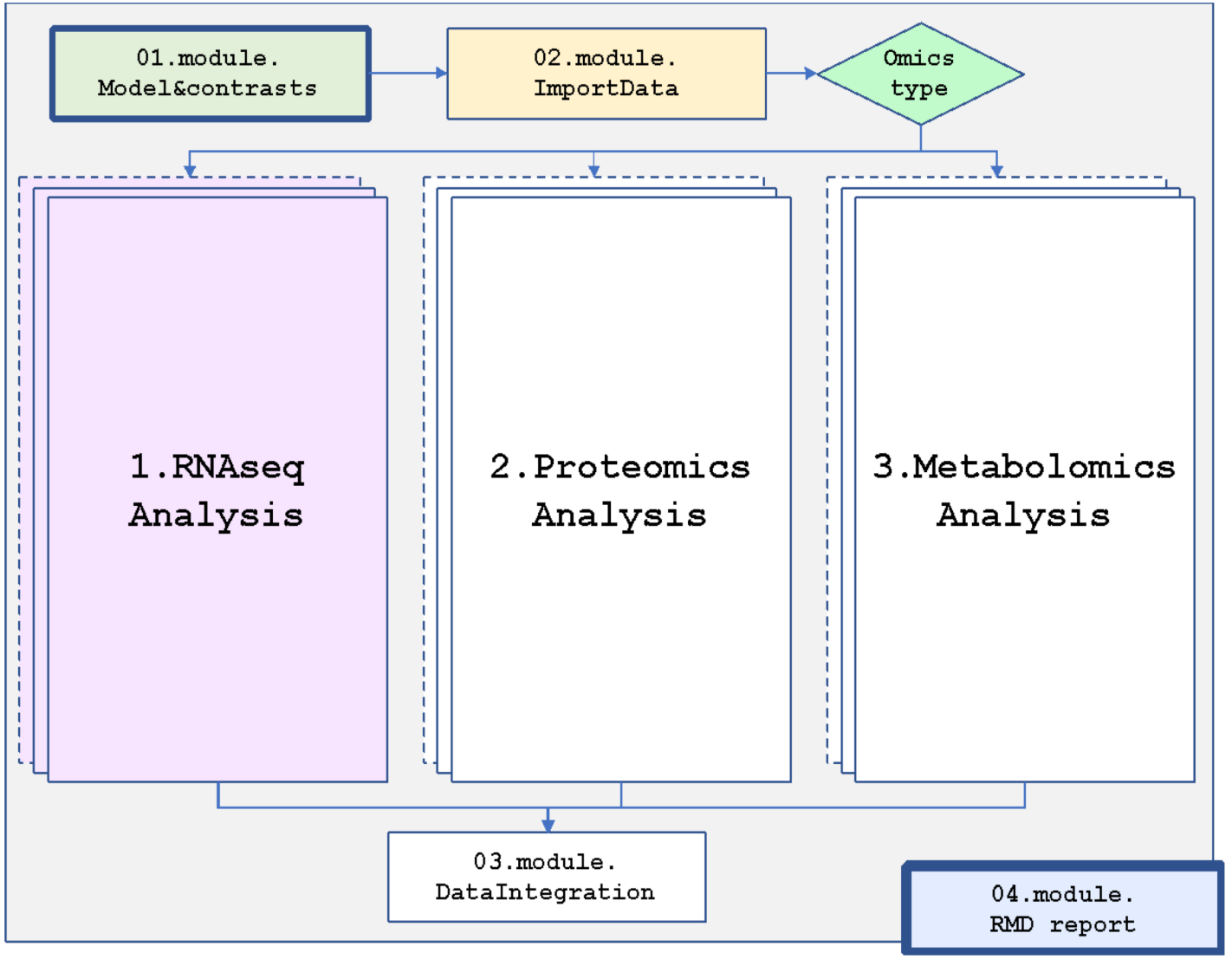
The MultiAssayExperiment object schematic (Ramos M. et all, 2017 nature)

Toward a FAIR-code package

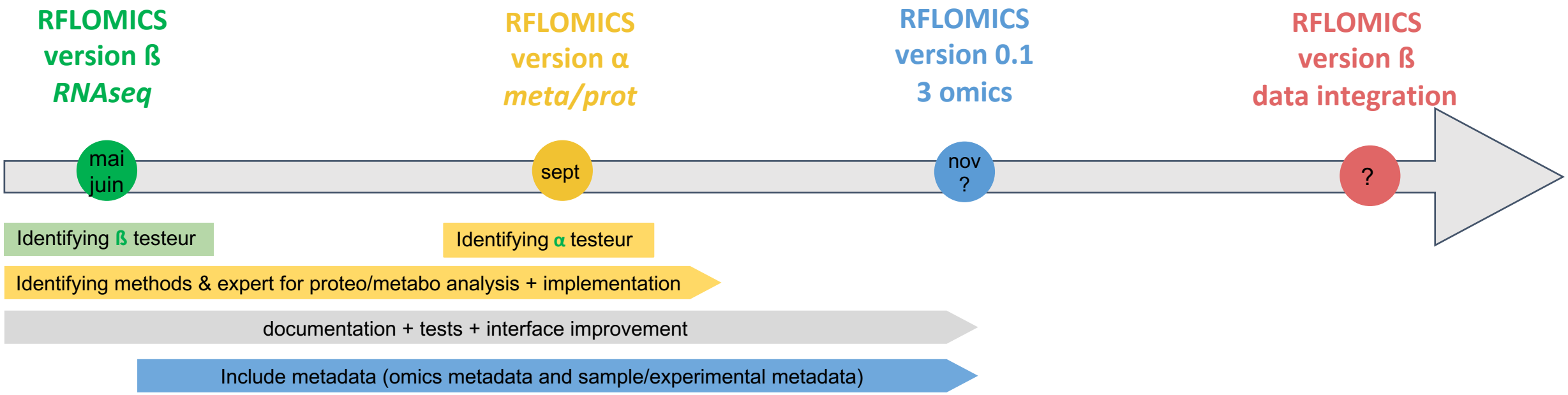
Technical choice		Findable/ Accessible	Interoperable	Reproducible (reusable)	Flexible code (new functionality)
R	**OOP**		generics functions		x
	devtools, roxygen	package, documentation	package		
	MAE/SE		supported by multiple applications/packages	raw data, results, methods, parameters...	all biological data type
	clustermq		most bioinfo clusters		
R shiny	**Shiny module**		x		manage complexity, simultaneous analysis
	Bookmarking (URL)		share		
	Repeatable			make a random function repeatable	
	Rmarkdown		x	report	Rmarkdown child
	gitlab (forgemia)	code source			
Deployment	**Docker**	x	x	x	
	Shiny Server (INRA)	x	x	x	

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



Identifying data projects with biological question

Identifying methods and expert

Implementation

Extending RFLOMICS to multi-omics data integration methods


Merci!