



UNIVERSITÀ DEGLI STUDI  
DI SALERNO



Dario Righelli

Ph.D. Student

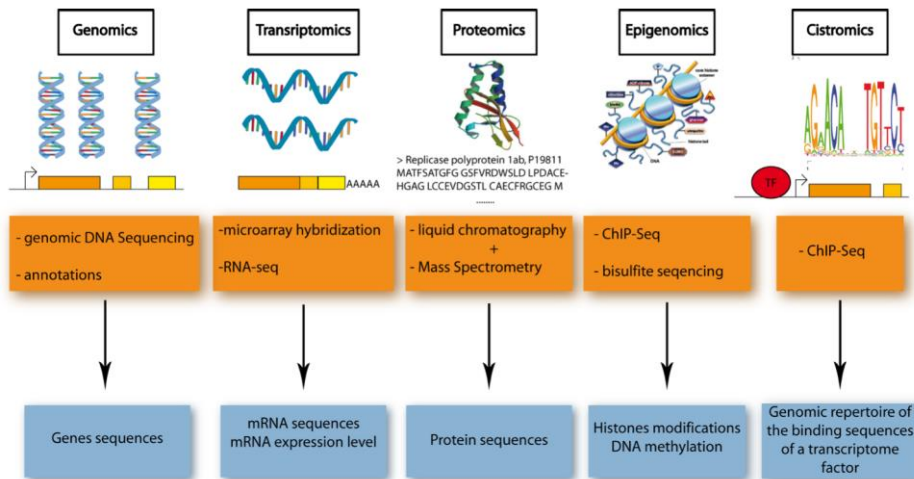
CHARME/EMBnet/NETTAB

25-10-2016

A Shiny based tool  
for NGS data  
Integration  
in Reproducible  
Spirit

Righelli D., Franzese M. , Angelini C.

# Background



- Sequencing techniques are developed to study cellular -omics
- Useful to inspect the cellular behaviour from many different point of view.

Image from: evodenomics, Coraline Petit, Master Biosciences, departement de Biologie, Ecole Normale Supérieure de Lyon. 01-10-2012

# Background

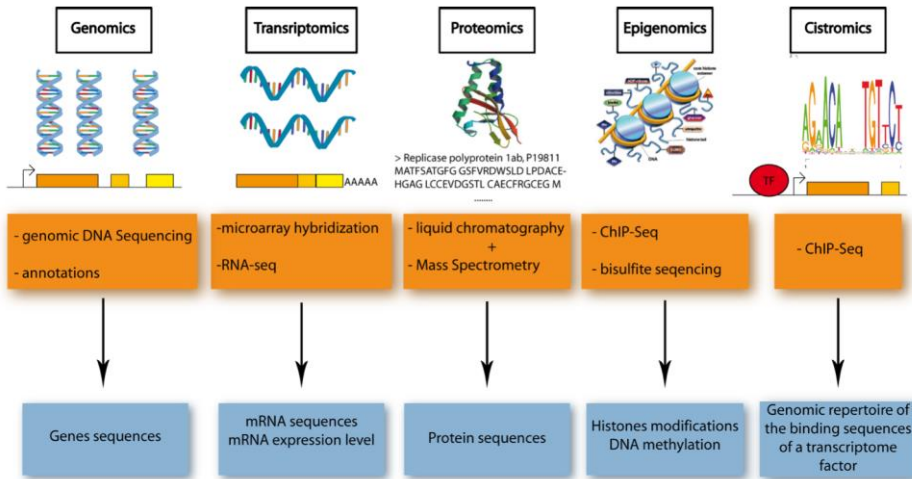


Image from: evodenomics, Coraline Petit, Master Biosciences, departement de Biologie, Ecole Normale Supérieure de Lyon. 01-10-2012

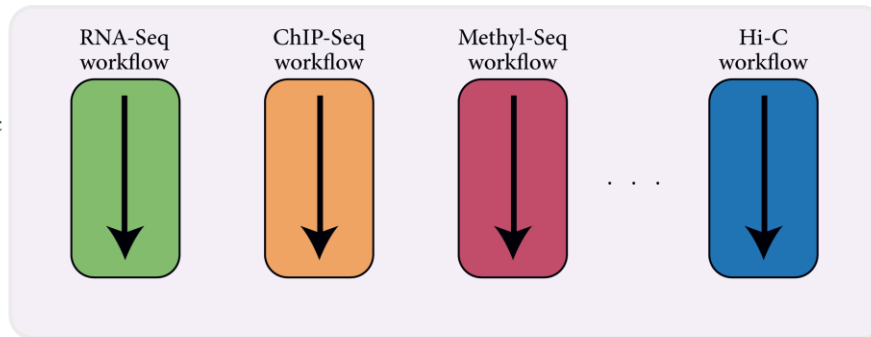
- Sequencing techniques are developed to study cellular -omics
- Useful to inspect the cellular behaviour from many different point of view.

- Single -seqs are like a camera view in a multi-view camera system.
- Single views need to be integrated to obtain the 3d model reconstruction.



# How Analysis works

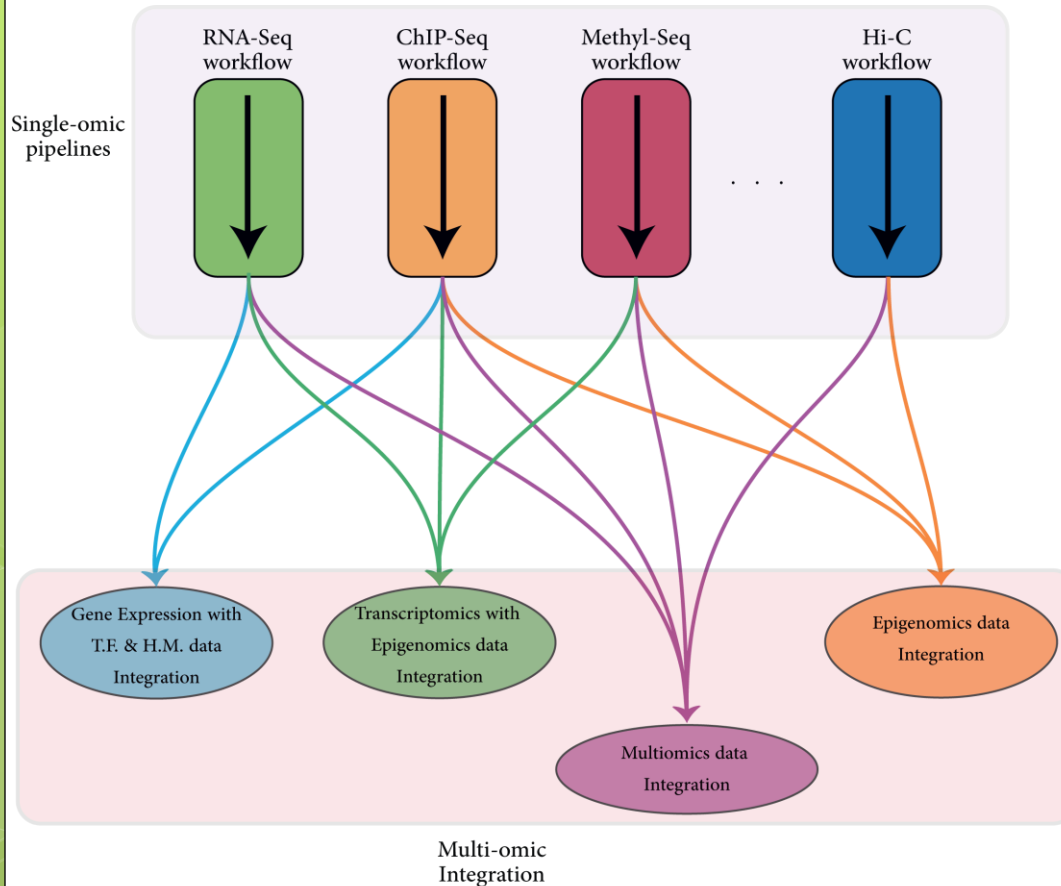
Single-omic  
pipelines



## Single-omic pipelines

- Complex analysis
- Several methods developed
- Analysis standards reached
- Several tools available
- Platform dependent

# How Analysis works



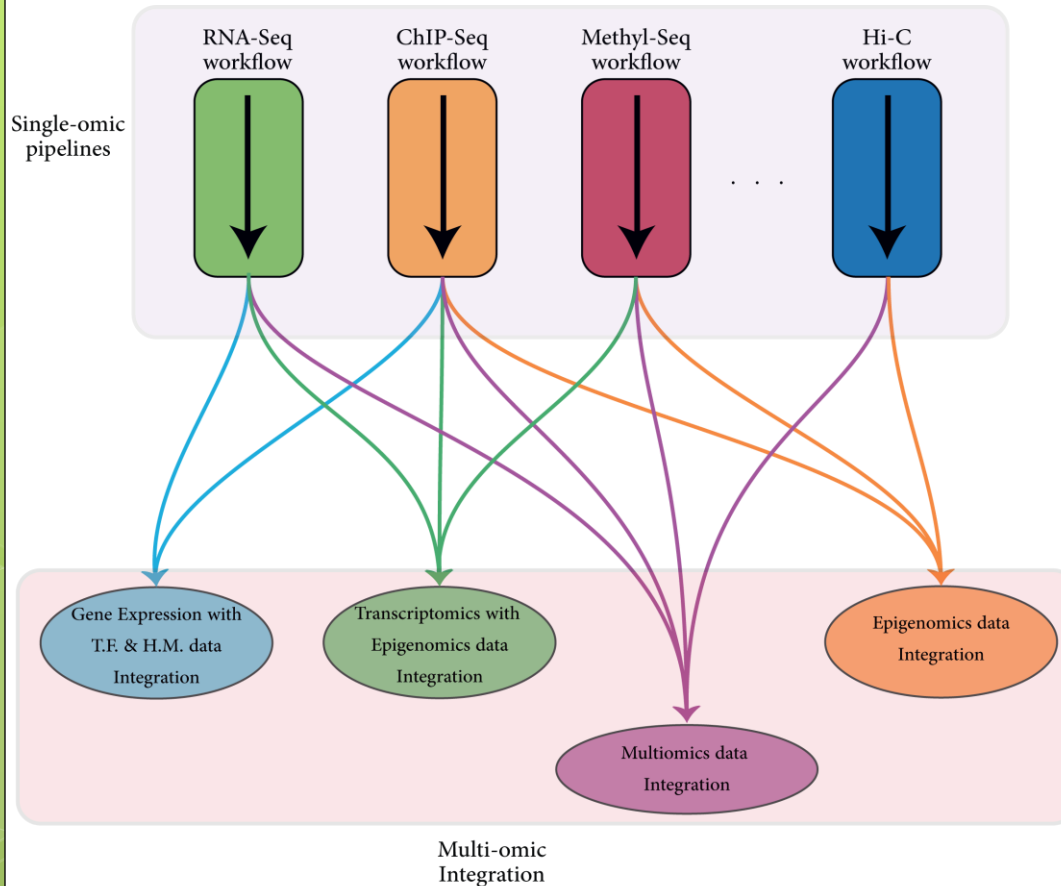
## Single-omic pipelines

- Complex analysis
- Several methods developed
- Analysis standards reached
- Several tools available
- Platform dependent

## Multi-omic Integration

- Emerging novel methods
- No past examples as guide
- No standards reached, yet
- Few implemented tools

# How Analysis works



## Single-omic pipelines

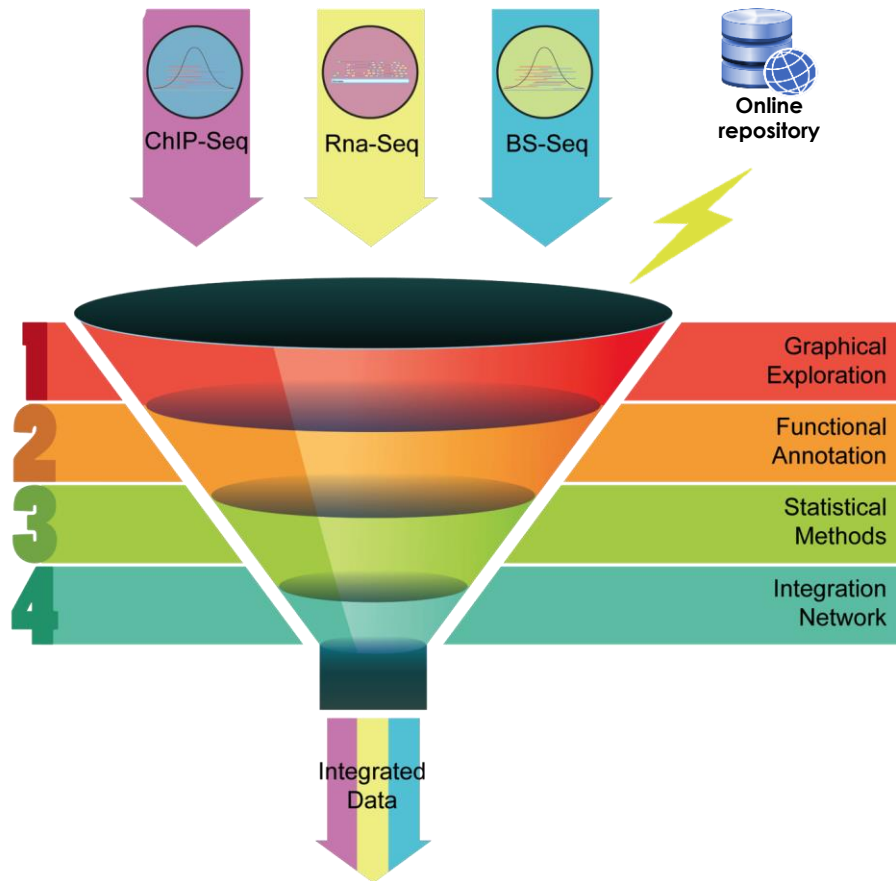
- Complex analysis
- Several methods developed
- Analysis standards reached
- Several tools available
- Platform dependent

## Multi-omic Integration

- Emerging novel methods
- No past examples as guide
- No standards reached, yet
- Few implemented tools

Focus on Next Generation Sequencing data

# How to Integrate



## Data

- Experimental data
- Online repositories
  - Consortium data
  - Annotation data
  - ...

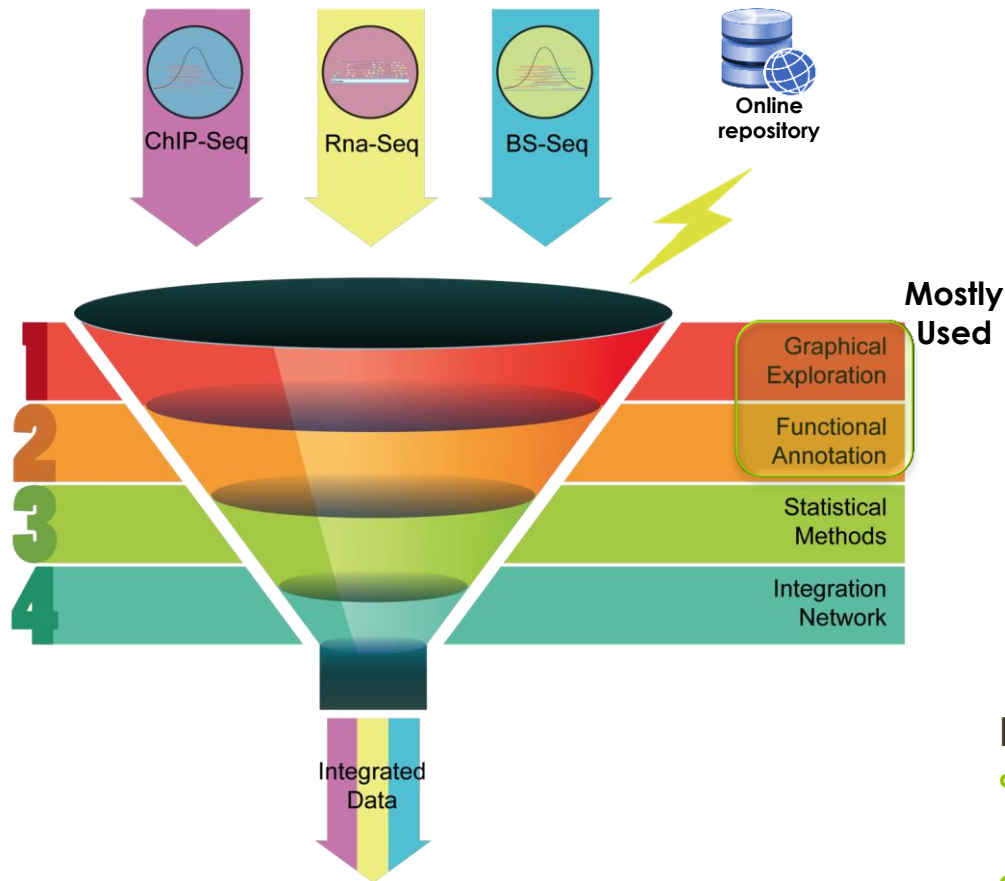
## Several integration levels

- Graphical level
- Pathway/GO Analysis
- Statistical methods
- Regulation Networks

## References

- MENG, Chen, et al. Dimension reduction techniques for the integrative analysis of multi-omics data. Briefings in bioinformatics, 2016.
- TUNCBAG, Nurcan, et al. Network-Based Interpretation of Diverse High-Throughput Datasets through the Omics Integrator Software Package. PLoS Comput Biol,
- ...

# How to Integrate



## Data

- Experimental data
- Online repositories
  - Consortium data
  - Annotation data
  - ...

## Several integration levels

- Graphical level
- Pathway/GO Analysis
- Statistical methods
- Regulation Networks

## References

- MENG, Chen, et al. Dimension reduction techniques for the integrative analysis of multi-omics data. Briefings in bioinformatics, 2016.
- TUNCBAG, Nurcan, et al. Network-Based Interpretation of Diverse High-Throughput Datasets through the Omics Integrator Software Package. PLoS Comput Biol,
- ...



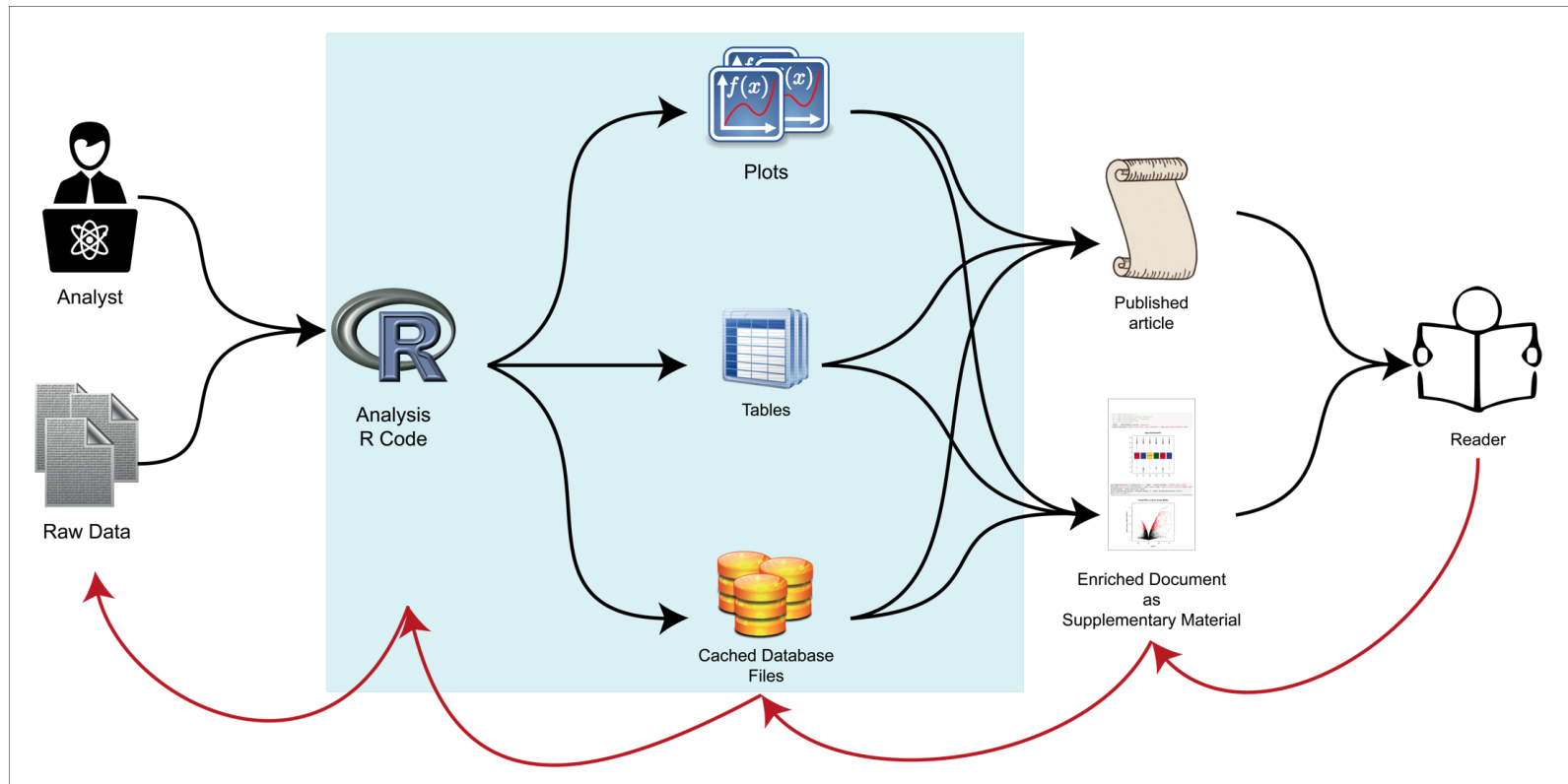
# Reproducible Background

## Handling problems

- Trace complex analysis
- Trace used methods
- Trace input/output data

## Third party reproducibility

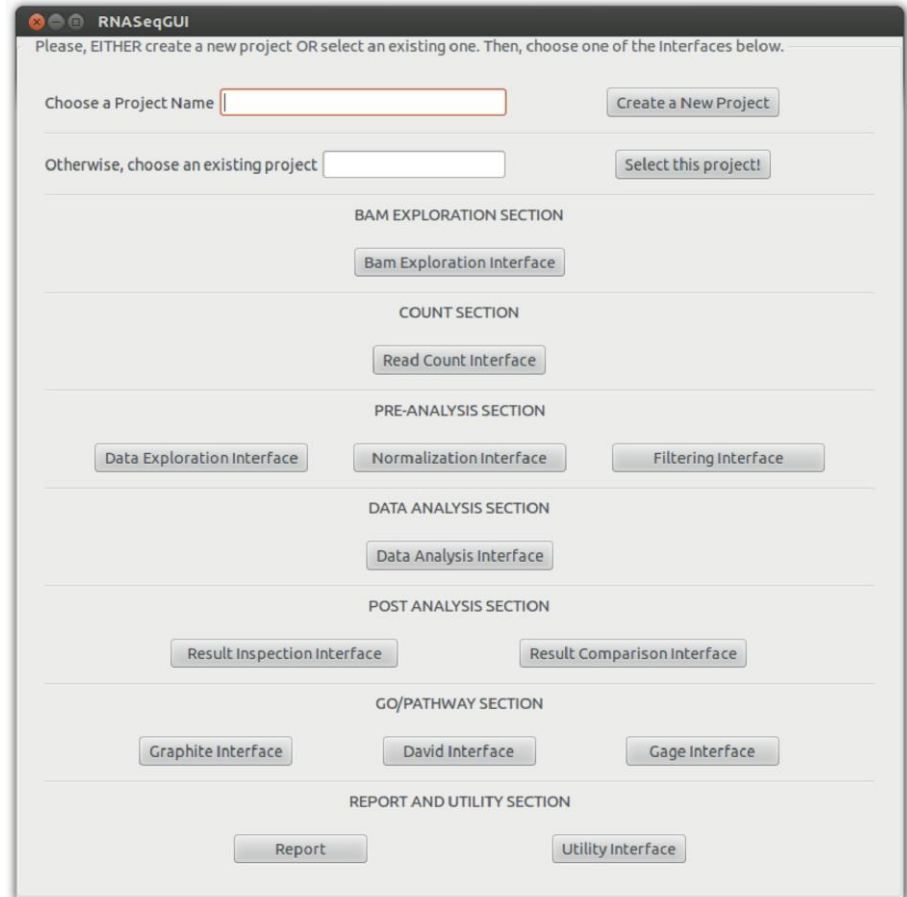
- Markdown language
- Data Caching
- Analysis Report



# How it's Implemented

## RNASeqGUI

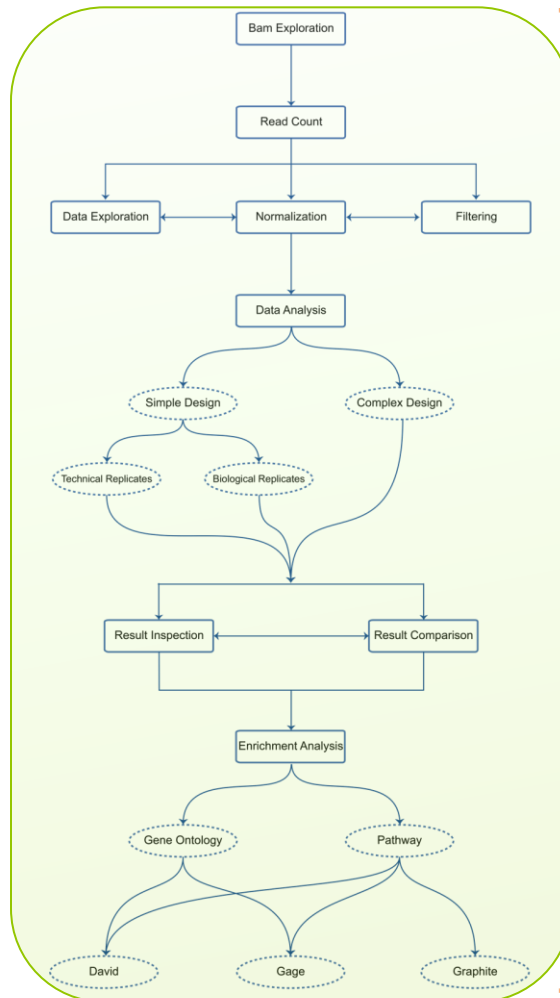
- Analyse RNA-Seq data
  - Preprocessing/normalization
  - DE methods
  - Pathway/GO analysis
  - Data inspection plots
- Reproducible Research
  - R-Markdown
  - Data Caching
    - filehash package
  - HTML Report
    - knitr package



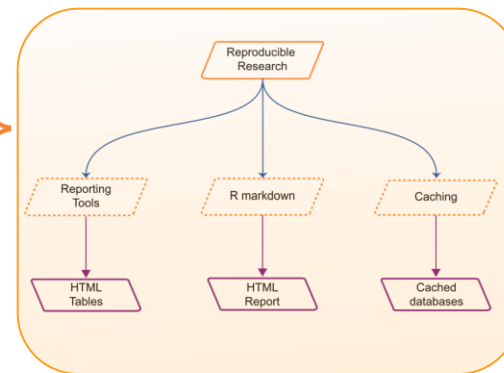
- Russo F., Angelini C. - RNASeqGUI: A GUI for analyzing RNA-seq data. Bioinformatics. 2014. 30(17): 2514-2516.
- Russo F., Righelli D. and Angelini C. - Advancements in RNASeqGUI towards a Reproducible Analysis of RNA-Seq Experiments. BioMed Research International, 2016, 7972351.

## How it's Implemented

### User Interactive Layer

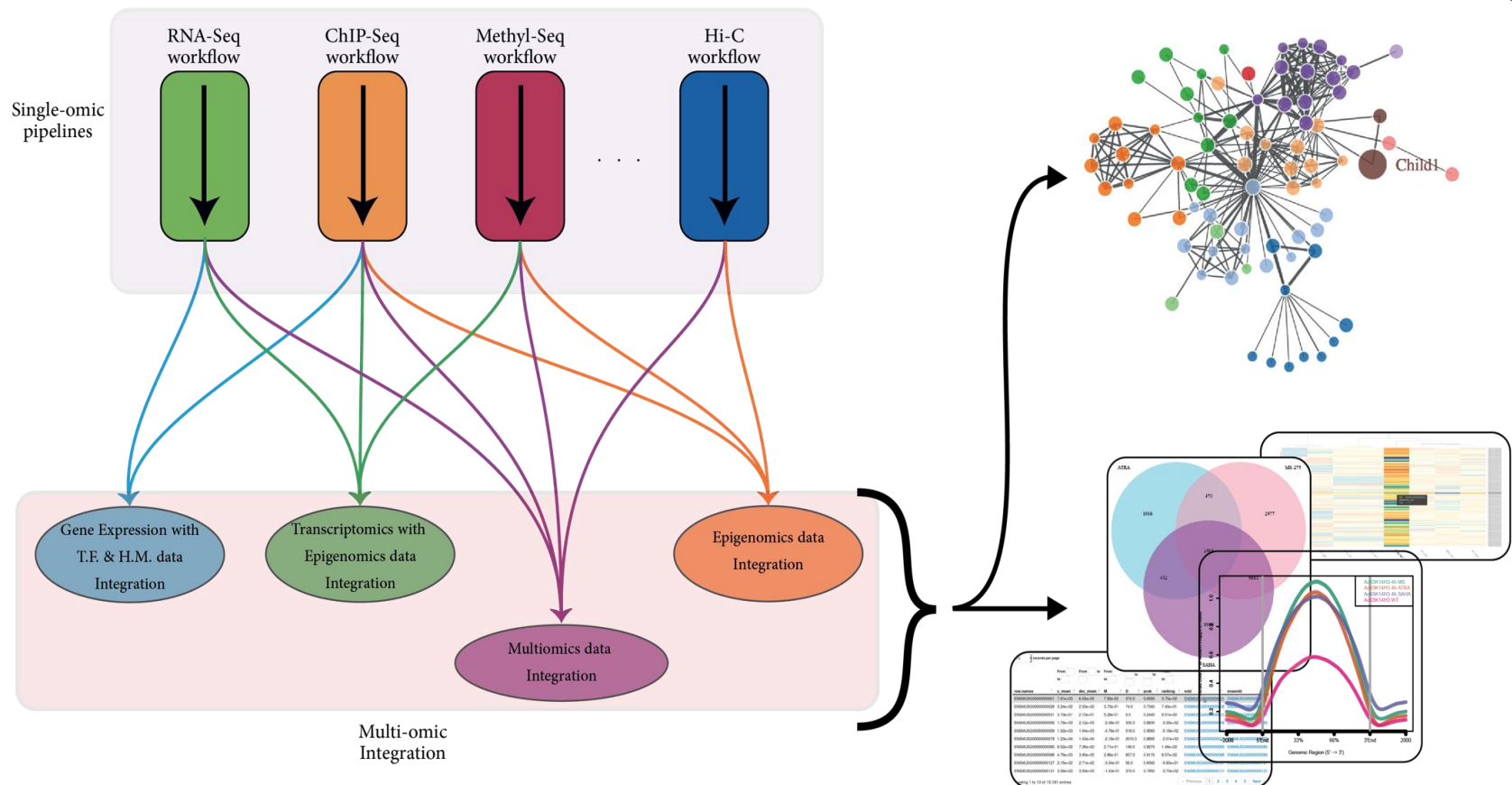


### RR Hidden Layer



# How to Improve

## Integration of High-throughput Omics data



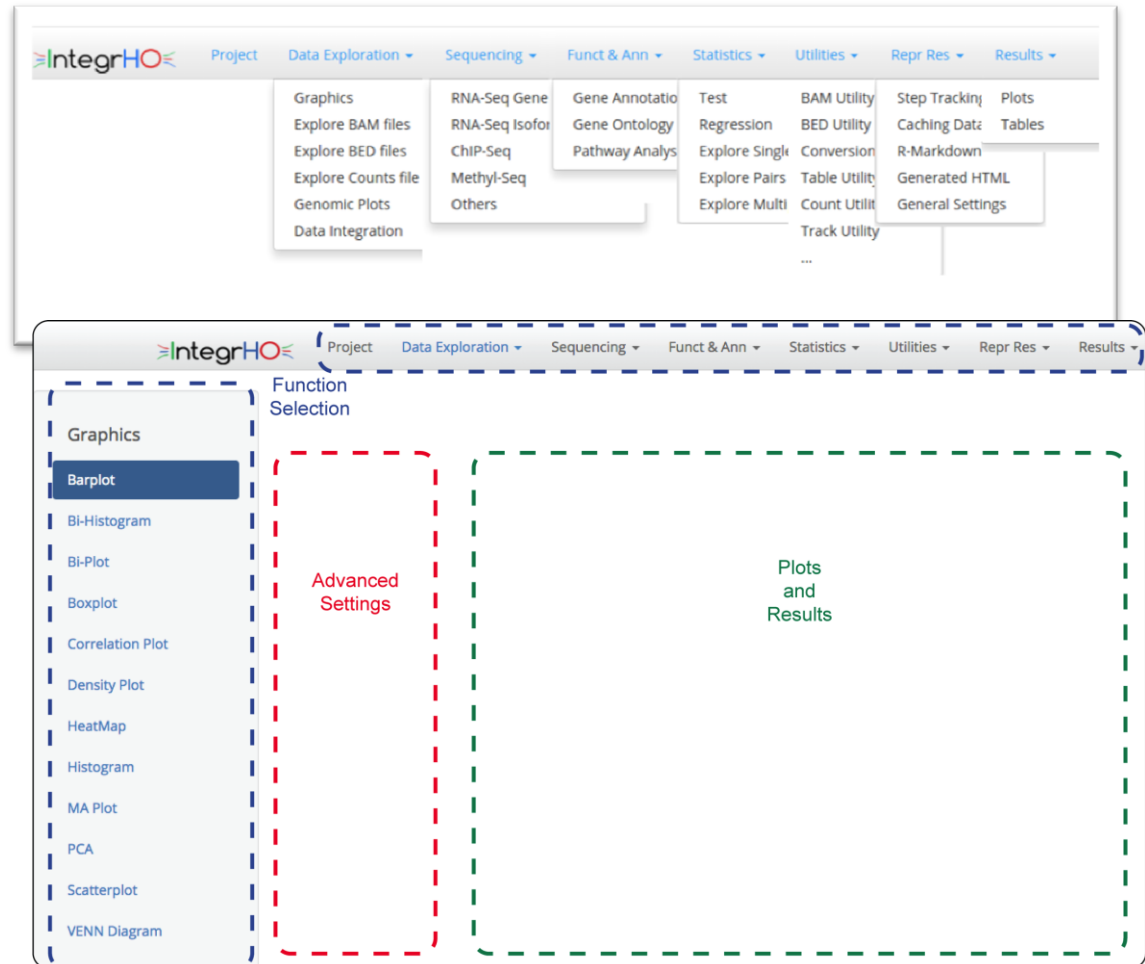
# How it's Made

## Graphical User Interface

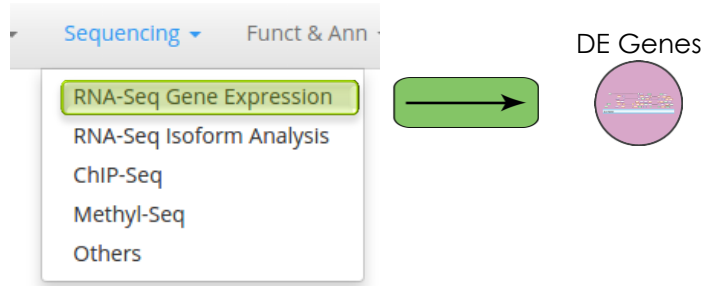
- HTML based (shiny)
- Stand-alone
- No pipeline oriented
- Interactive plots (plotly)
- Multi-omic analysis
- Multi-omic integration
- Reproducible Research
- open-source

## On-going implementation

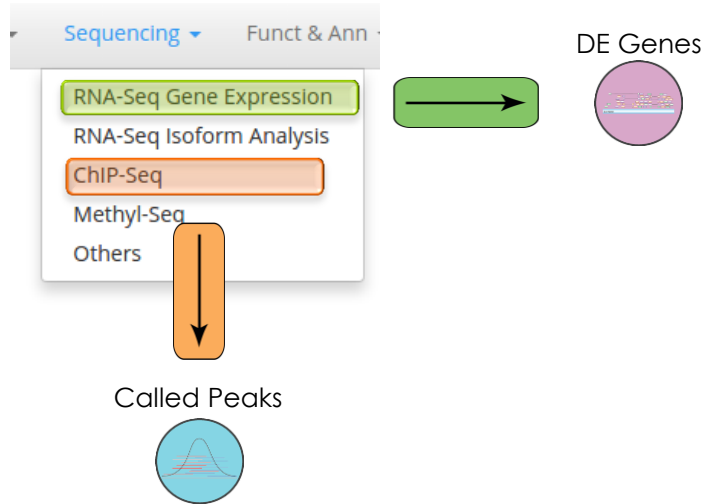
- RNA-Seq data Analysis
- ChIP-Seq data Analysis
- BS-Seq data Analysis
- Live Markdown editing



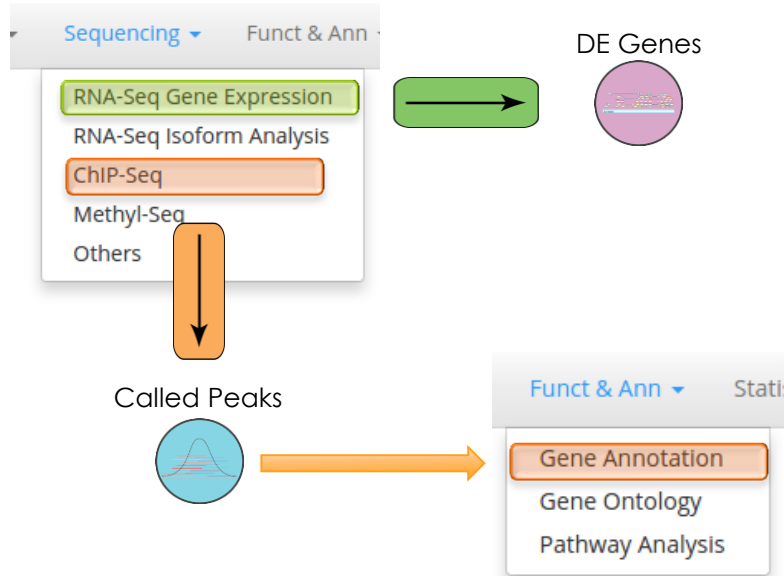
# How to Use



# How to Use

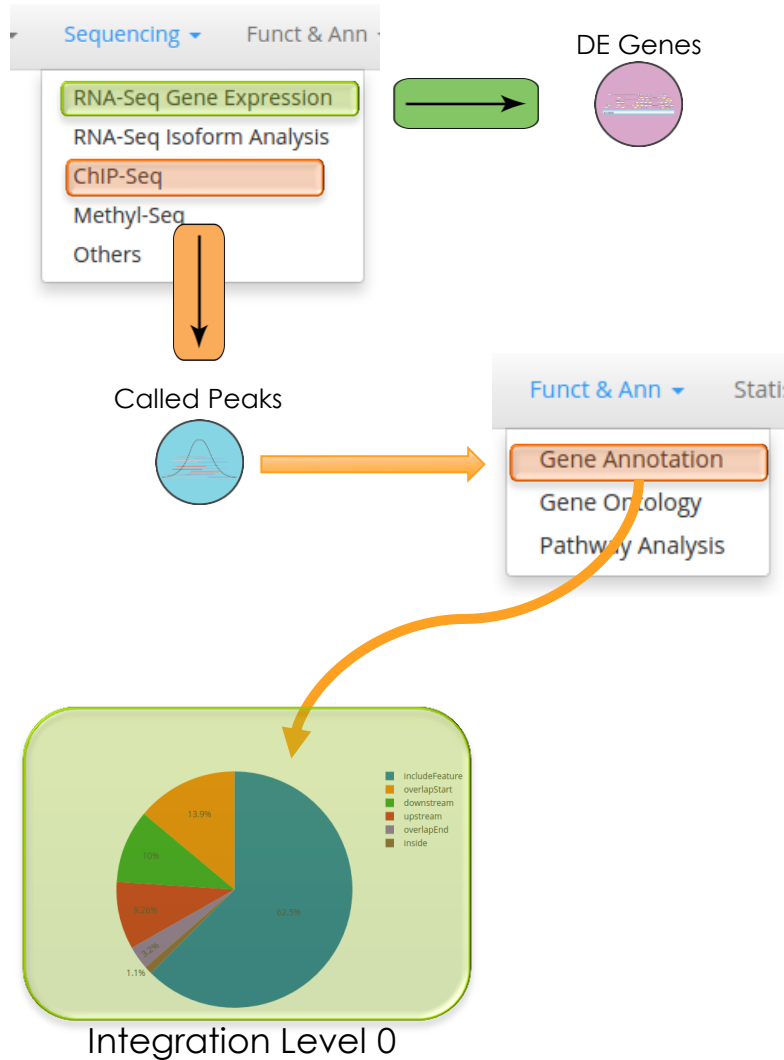


# How to Use

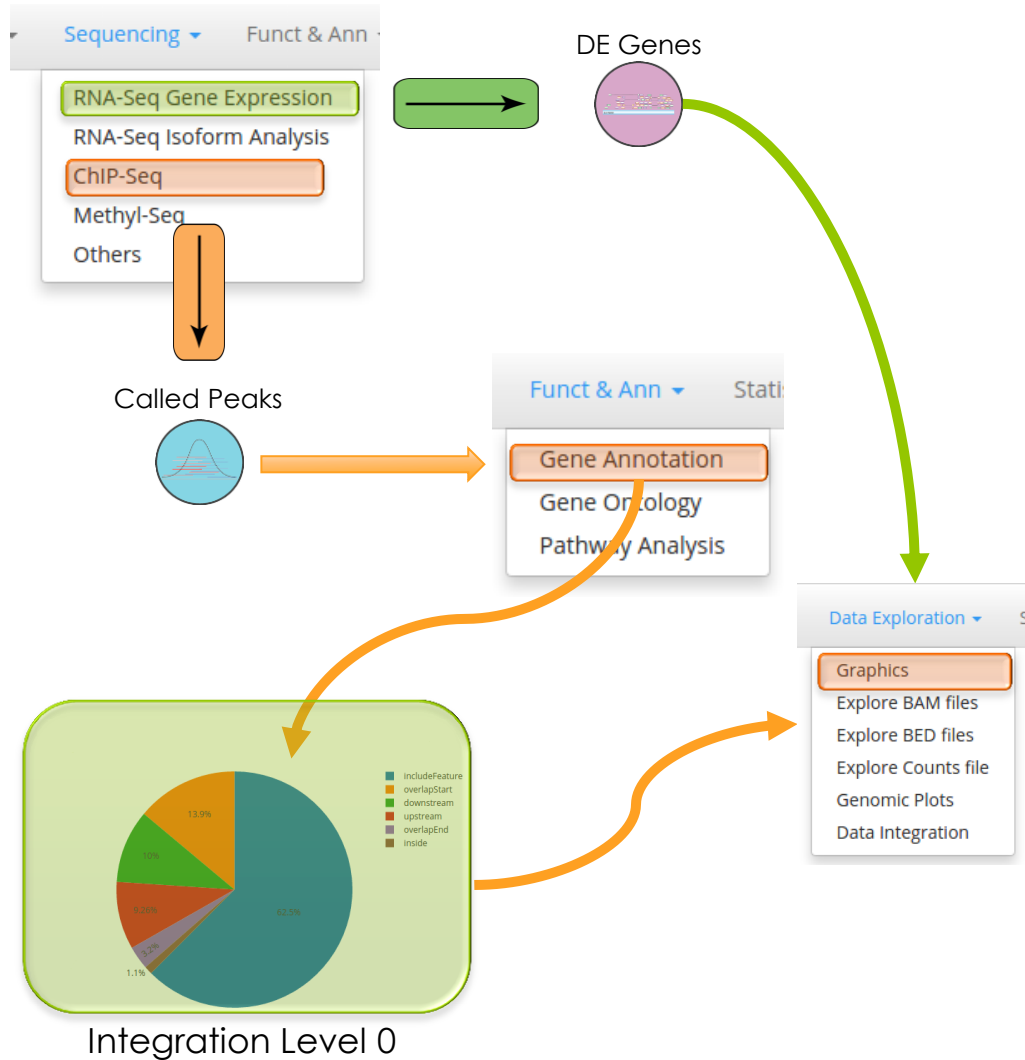




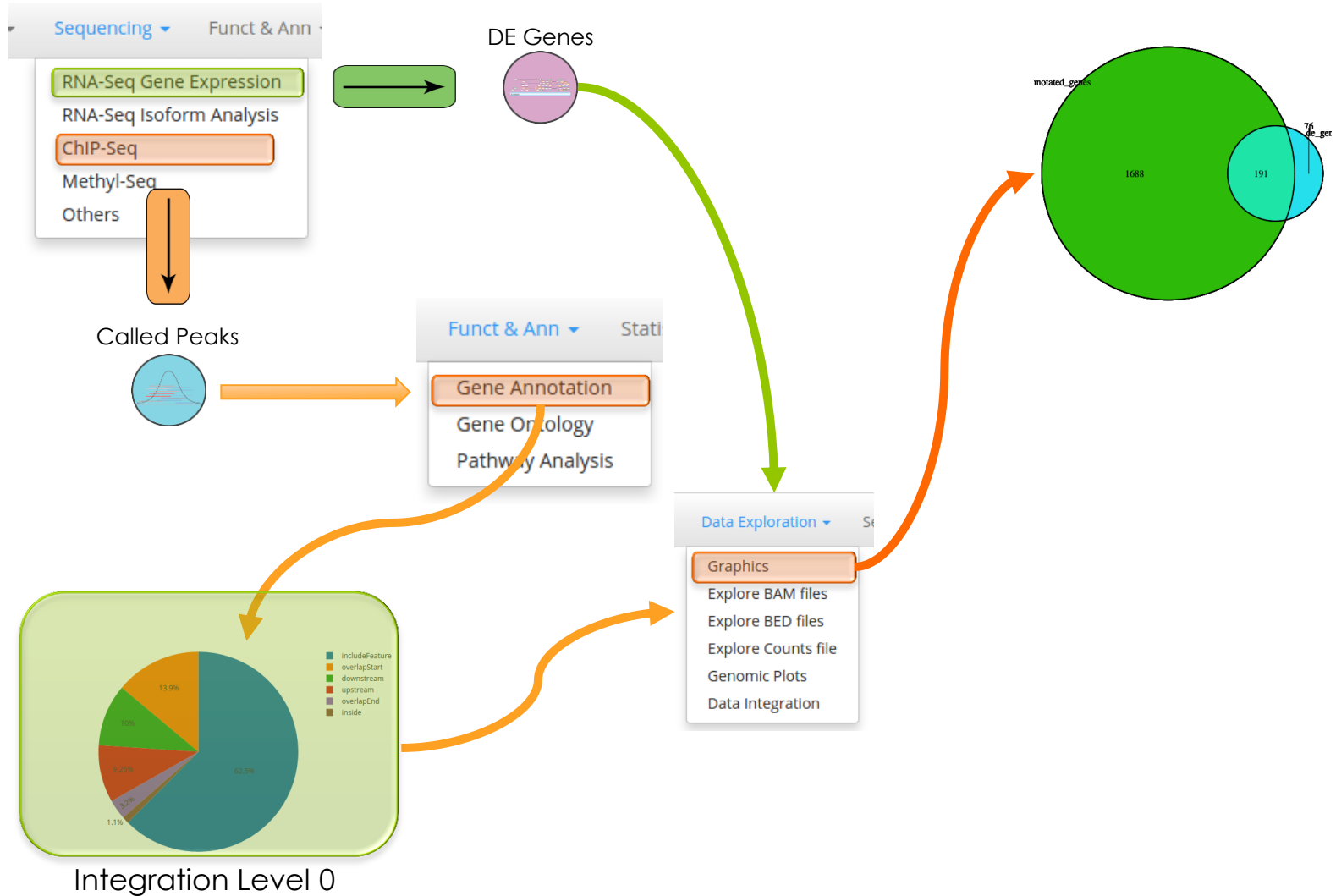
# How to Use



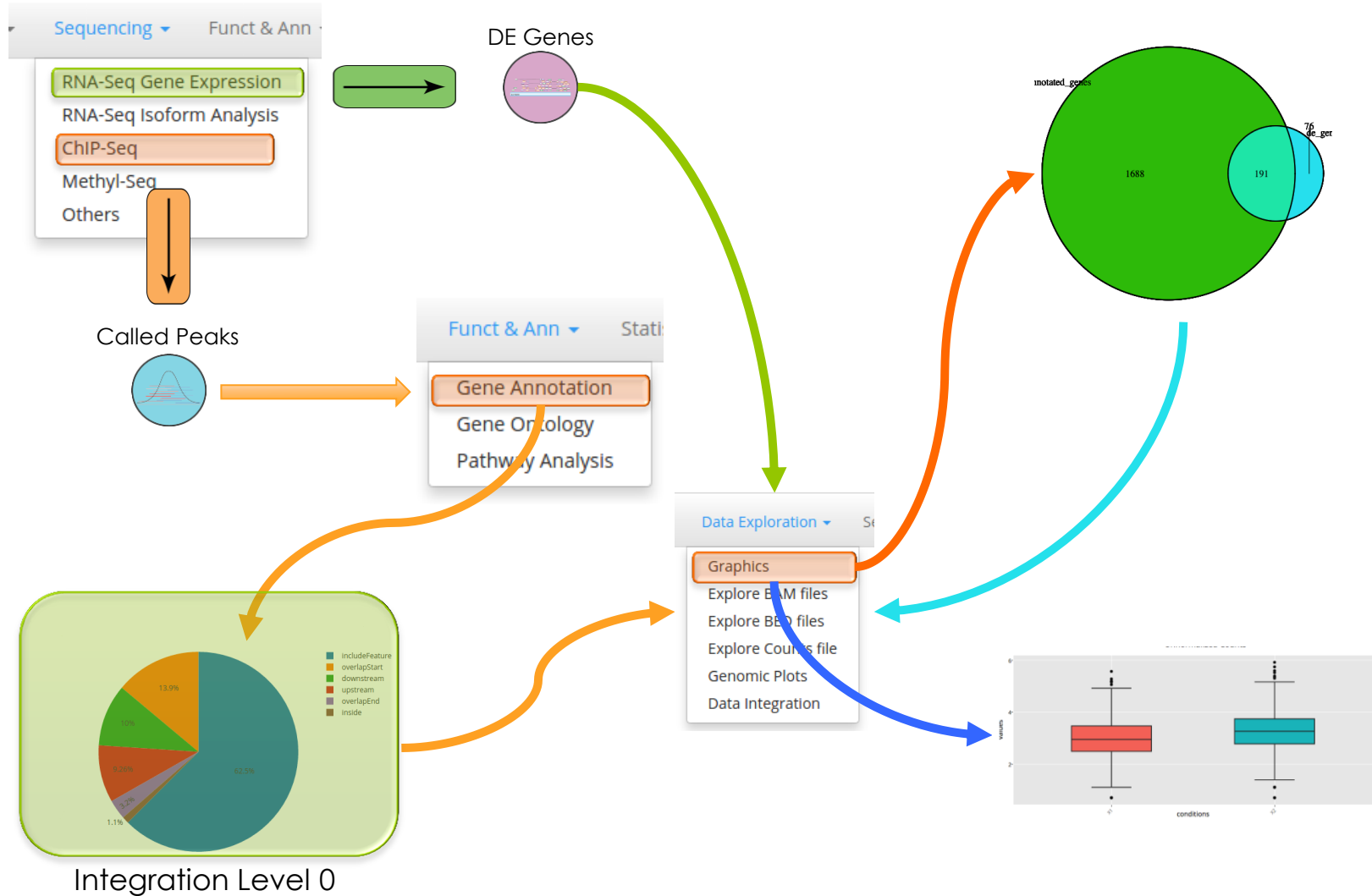
# How to Use



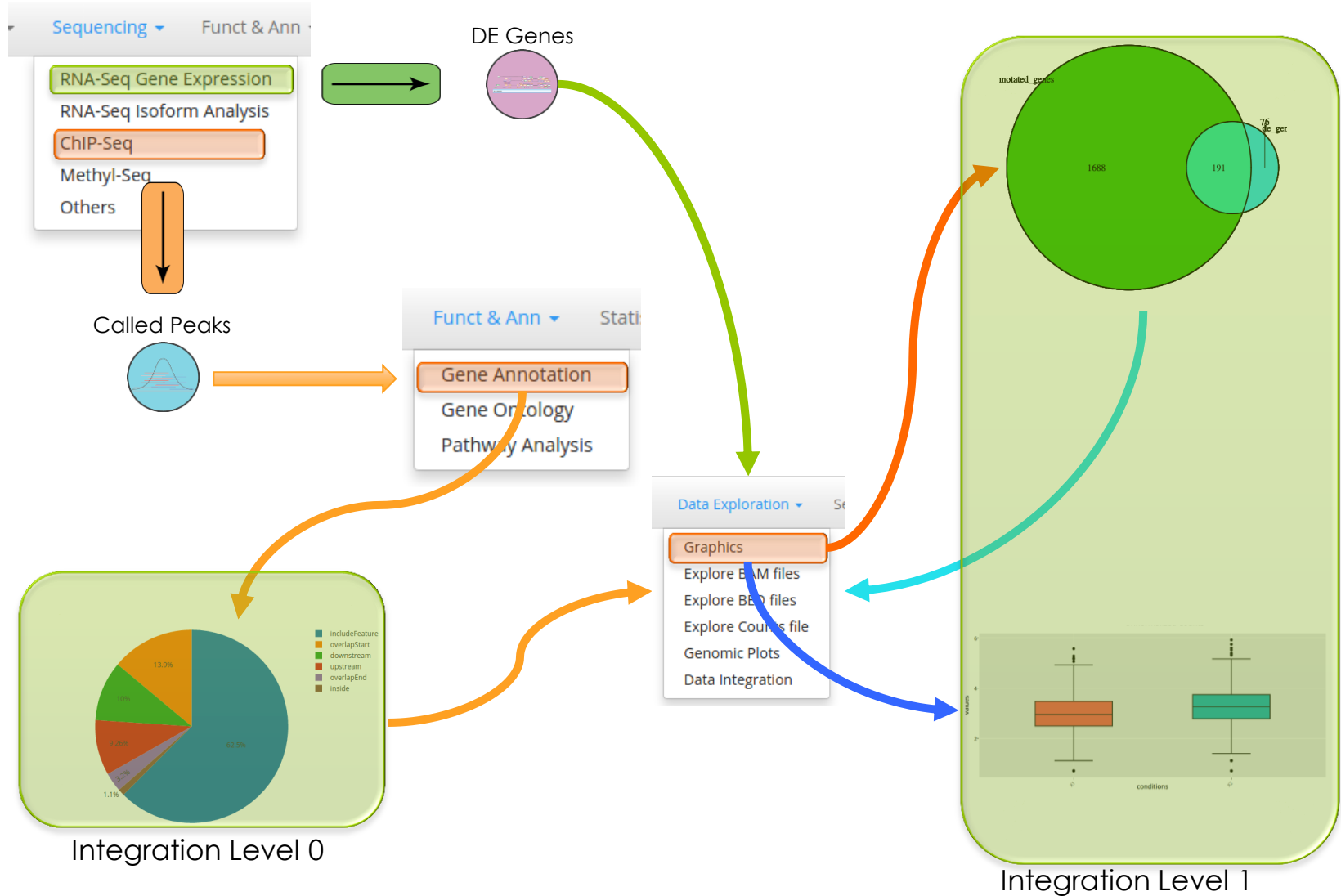
# How to Use



# How to Use

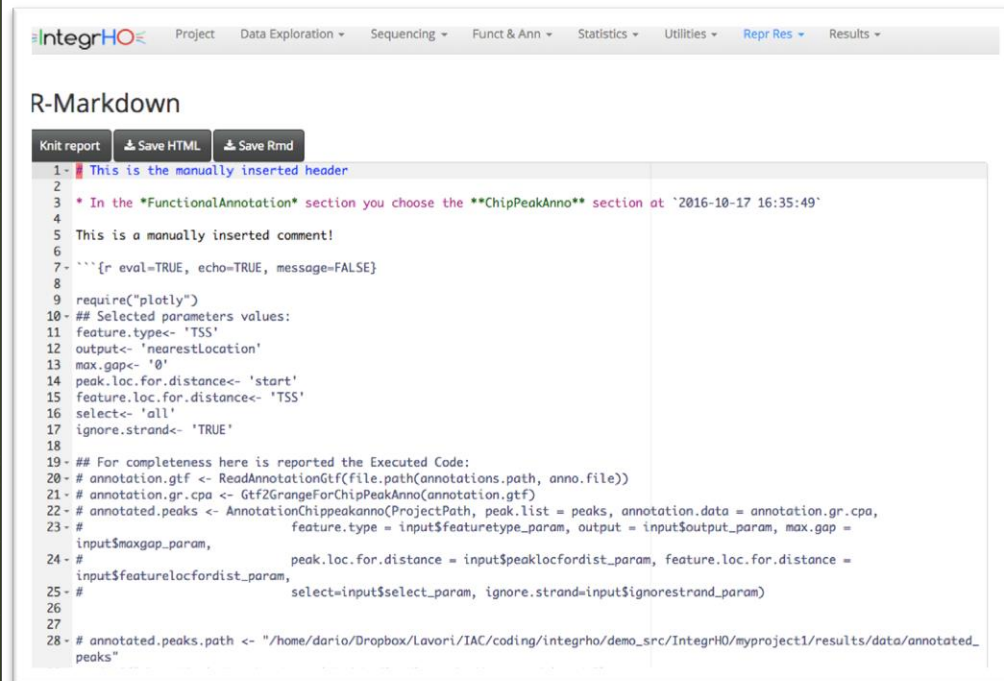


# How to Use



# How it Reproduces

- Dedicated interface
  - Fast R-markdown editing
  - Fast report compilation
- Packages
  - shinyAce
  - rmarkdown/knitr



```

1- This is the manually inserted header
2
3- * In the *FunctionalAnnotation* section you choose the **ChipPeakAnno** section at `2016-10-17 16:35:49`
4
5- This is a manually inserted comment!
6
7- ```{r eval=TRUE, echo=TRUE, message=FALSE}
8
9- require("plotly")
10- ## Selected parameters values:
11- feature.type<- 'TSS'
12- output<- 'nearestLocation'
13- max_gap<- '0'
14- peak.loc.for.distance<- 'start'
15- feature.loc.for.distance<- 'TSS'
16- select<- 'all'
17- ignore.strand<- 'TRUE'
18
19- ## For completeness here is reported the Executed Code:
20- # annotation.gtf <- ReadAnnotationGtf(file.path(annotations.path, anno.file))
21- # annotation.gr.cpa <- Gtf2GrangeForChipPeakAnno(annotation.gtf)
22- # annotated.peaks <- AnnotationChippeakanno(ProjectPath, peak.list = peaks, annotation.data = annotation.gr.cpa,
23- #   feature.type = input$featuretype_param, output = input$output_param, max_gap =
24- #   input$maxgap_param,
25- #   peak.loc.for.distance = input$peaklocfordist_param, feature.loc.for.distance =
26- #   input$featurelocfordist_param,
27- #   select=input$select_param, ignore.strand=input$ignorestrand_param)
28- # annotated.peaks.path <- "/home/dario/Dropbox/Lavori/IAC/coding/integrho/demo_src/IntegrHO/myproject1/results/data/annotated_
   peaks"
  
```

# How it Reproduces

- Dedicated interface
  - Fast R-markdown editing
  - Fast report compilation
- Packages
  - shinyAce
  - rmarkdown/knitr

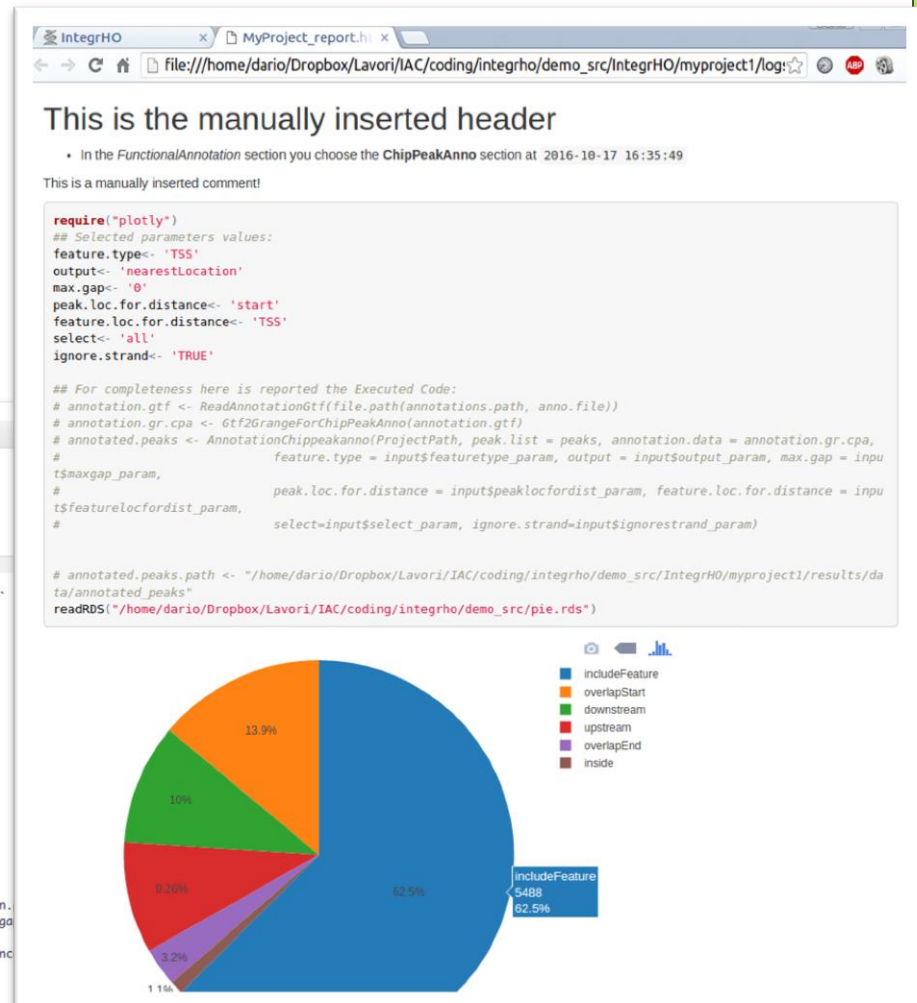
IntegrHO Project Data Exploration Sequencing Funct & Ann Statistics Utilities Repr Res

## R-Markdown

Knit report Save HTML Save Rmd

```

1- This is the manually inserted header
2
3- * In the *FunctionalAnnotation* section you choose the **ChipPeakAnno** section at `2016-10-17 16:35:49`
4
5- This is a manually inserted comment!
6
7- ```{r eval=TRUE, echo=TRUE, message=FALSE}
8
9- require("plotly")
10- ## Selected parameters values:
11- feature.type<- 'TSS'
12- output<- 'nearestLocation'
13- max_gap<- '0'
14- peak.loc.for.distance<- 'start'
15- feature.loc.for.distance<- 'TSS'
16- select<- 'all'
17- ignore.strand<- 'TRUE'
18
19- ## For completeness here is reported the Executed Code:
20- # annotation.gtf <- ReadAnnotationGtf(file.path(annotations.path, anno.file))
21- # annotation.gr.cpa <- Gtf2GrangeForChipPeakAnno(annotation.gtf)
22- # annotated.peaks <- AnnotationChippeakanno(ProjectPath, peak.list = peaks, annotation.data = annotation.
23- # feature.type = input$featuretype_param, output = input$output_param, max.gap = input$
24- # input$maxgap_param, peak.loc.for.distance = input$peaklocfordist_param, feature.loc.for.distance = input$
25- # input$featurelocfordist_param, select=input$select_param, ignore.strand=input$ignorestrand_param)
26
27
28- # annotated.peaks.path <- "/home/dario/Dropbox/Lavori/IAC/coding/integrho/demo_src/IntegrHO/myproject1/results/data/annotated_
29- peaks"
```



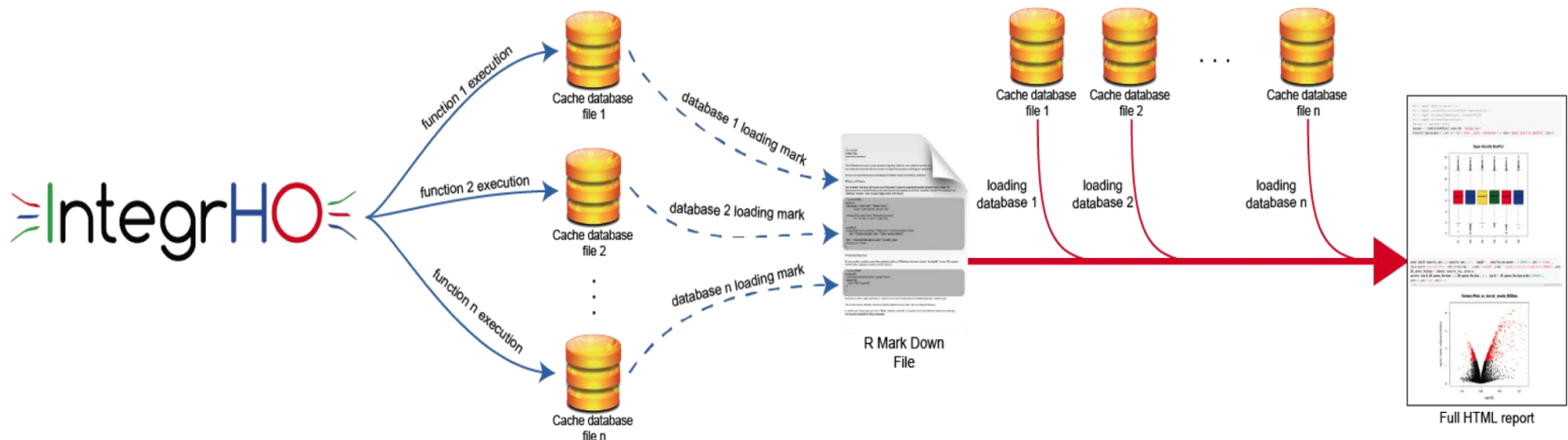
# How it Caches

## Motivations

- Helpful to reproduce analysis
- Stores input/output objects
- Useful to speed-up time demanding steps
- Useful to share data

## Implementation

- Used to speed-up report compilation
- Implemented with R base data management system
  - No external packages required





# How it's Improved

## side mini projects

### **Rmarkdown wrapper functions**

- Speed up code reporting
- Standard markdown based
  - <https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf>
- *rmarkdown* package
- No Markdown knowledge necessary
- If caching is reported
  - Some functions depends on caching wrapper class

### **Caching wrapper class**

- Based on *RData* files
- No external dependencies
- R base implementation

### **Online Accession**

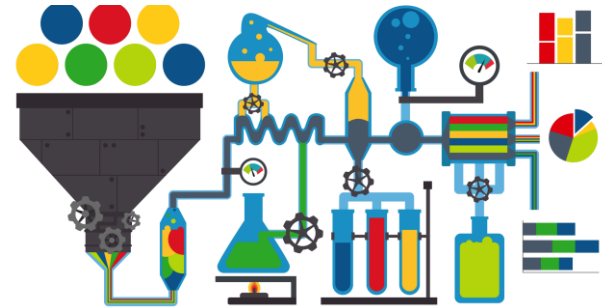
- <https://github.com/drighelli/RR>

(on-going development)

# What's next

## Integration methods

- Graphical methods
- Statistical methods
- Integration networks



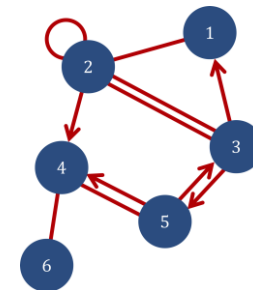
## User helping functionalities

- Preselected analysis workflow
- Live help
- Live results inspection



## Reproducible Research

- Live browsing of caching file
- Analysis step tracking
  - Graph construction





Thank you  
for Your  
Attention!



# How to Use

