

# An R pipeline using the “*targets*” package for Multi-Omics Integrative Analyses

**Ferran Briansó<sup>1</sup> and Alex Sánchez-Pla<sup>2</sup>**

Statistics and Bioinformatics Research Group,  
Department of Genetics, Microbiology & Statistics, Universitat de Barcelona.  
Av. Diagonal 645, 08028, Barcelona, Catalonia, Spain.

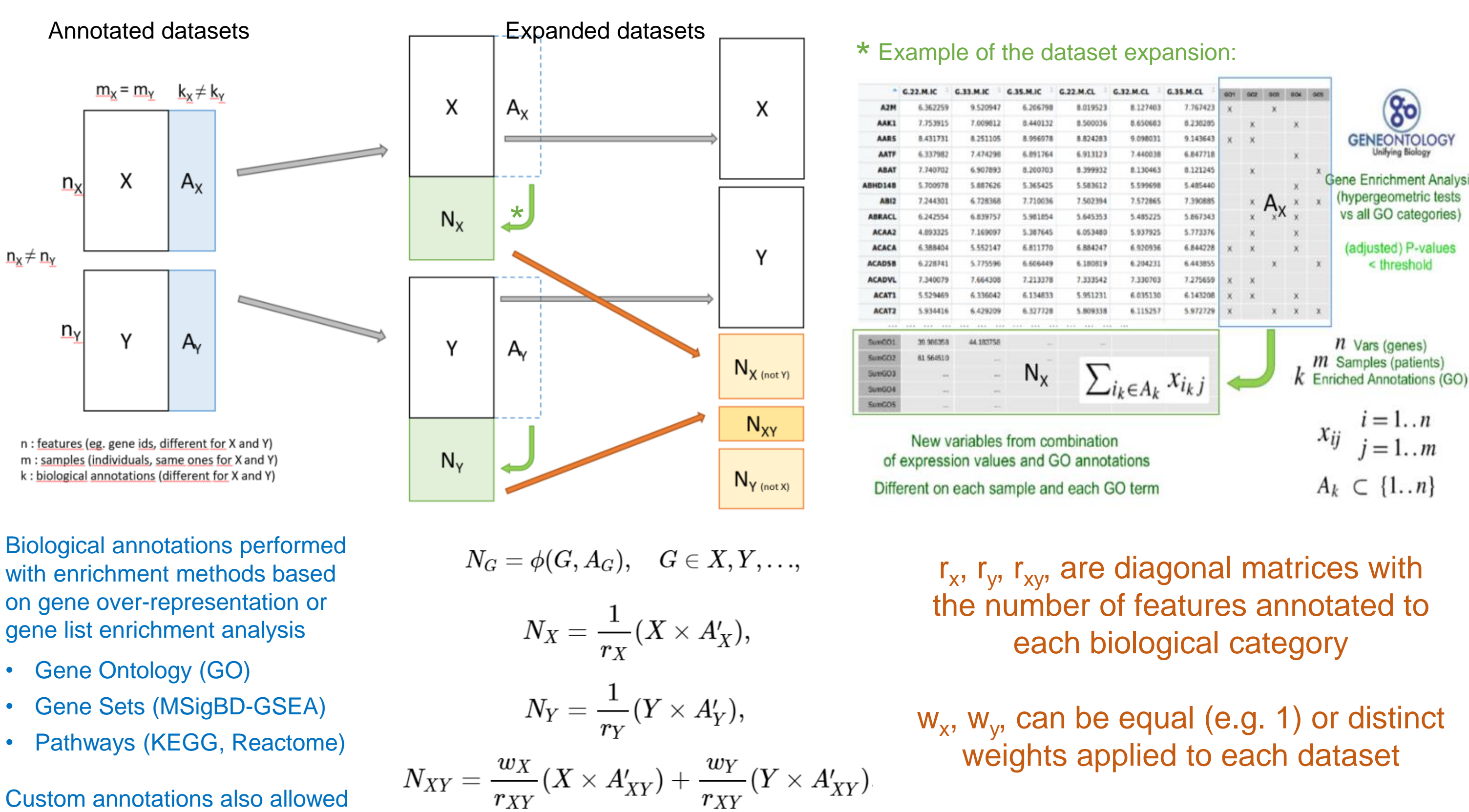
<sup>1</sup>[ferran.brianso@gmail.com](mailto:ferran.brianso@gmail.com), <sup>2</sup>[asanchez@ub.edu](mailto:asanchez@ub.edu)

## Motivation and Main Idea

One common approach to multi-omics integration is using Dimension Reduction techniques, which are also helpful for data visualization<sup>1,2</sup>. There is however some space for improvement: One point that may be improved is the difficulty in interpreting results from the biological point of view. Another is the lack of pipelines covering the whole integration process, which may be simpler to use than usual bioinformatic pipelines<sup>3,4</sup>, such as those based on Galaxy<sup>5</sup> and NextFlow<sup>6</sup>, which require a certain level of technical knowledge to be configured. Our proposal addresses these two points, by

1. Adding **biological annotations** (GO<sup>7</sup>, GSEA<sup>8</sup>...) to the original datasets (X, Y) **before** the joint analysis, thus creating the **expanded datasets** (N<sub>X</sub>, N<sub>Y</sub>...) with new annotation-derived features,
2. Performing the analysis of the original and expanded datasets with contrasted **Multi-Block Dimension Reduction methods** (such as RGCCA<sup>9</sup>, MFA<sup>10</sup>),
3. Implementing a **pipeline with the *targets* package**<sup>11</sup> that allows data scientists and researchers to work entirely within R, in a reproducible and easy-to-use workflow.

## Expanding Data with Biological Annotations



## The *targets* Package

Main `_targets.R` file, with definition of `tar_target()` objects

```
#####
#### To run this targets script:
## 1- load 'targets' and 'tarchetypes' libs
library(targets)
library(tarchetypes)

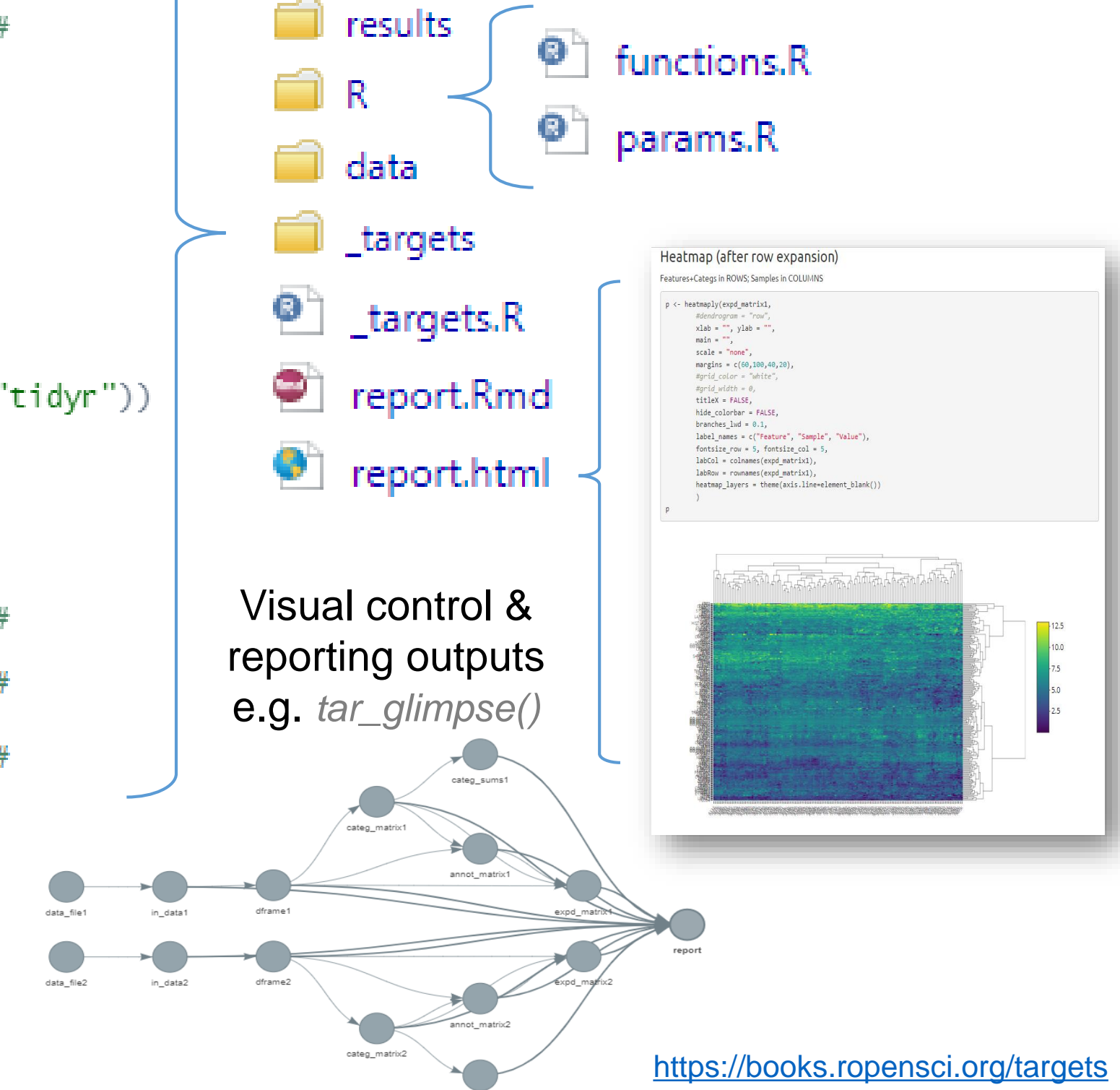
## 2- load functions, params and general options
source("R/functions.R")
source("R/params.R")
options(tidyverse.quiet = TRUE)
tar_option_set(packages = c("tidyverse", "dplyr", "readr", "tidyr"))

## 3- run the following commands IN CONSOLE
#tar_make() # to run the script
#tar_glimpse() # to see a simple view of the workflow
#tar_visnetwork() # to see more details of the workflow
#####

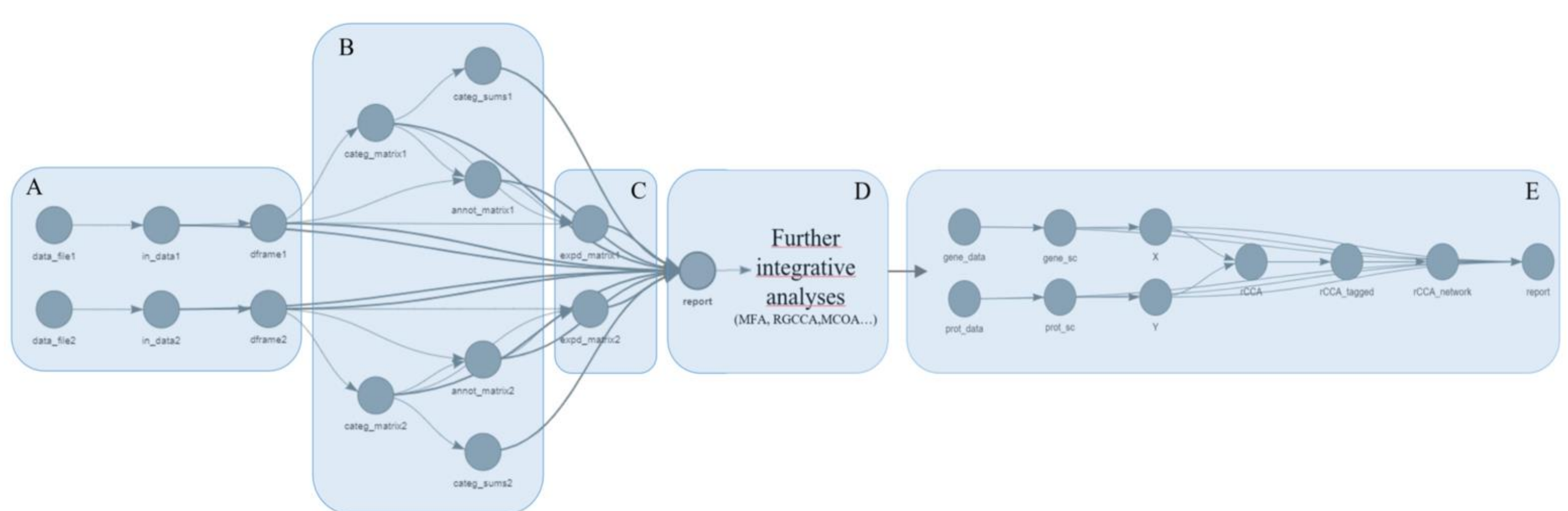
##### DO NOT RUN ANY OF THE REMAINING LINES
#####

list(
  ## 01- Lectura de dades d'entrada
  tar_target(
    data_file1,
    p.infile1,
    format = "file"
  ),
  ...
)
```

## Easy working dir structure



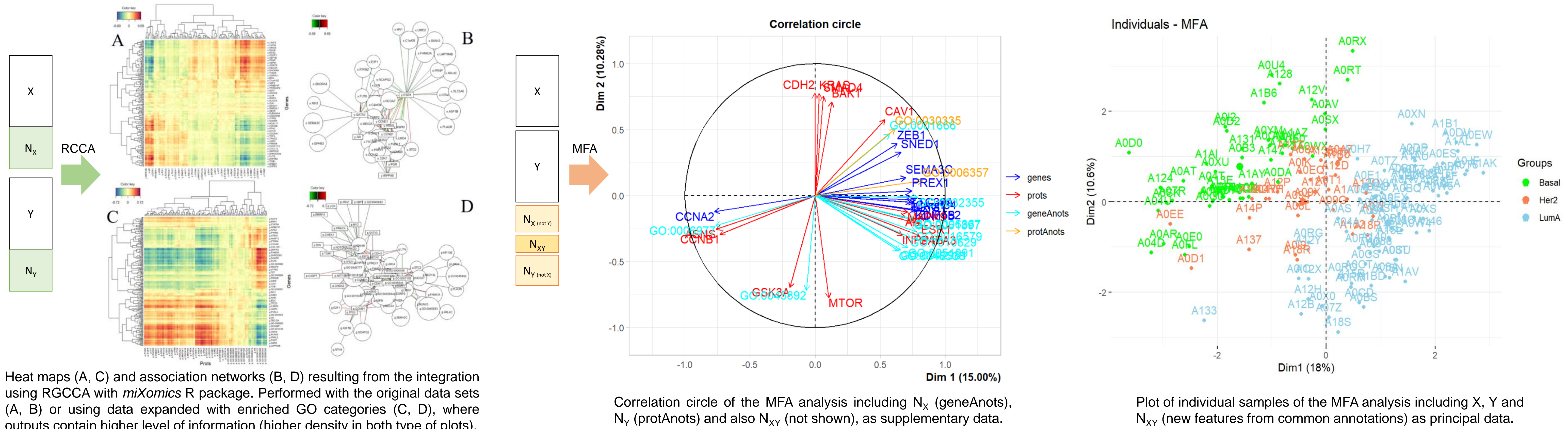
## The Analysis Pipeline



- A) Two 'omics-derived input data sets are loaded and converted to R data frames.
- B) Annotations are created, or loaded if custom annotation file provided, as additional objects, one for each given input matrix, and used to expand these original data, to end up with
- C) A pair of data frames containing the starting values plus the average expression/abundance values of the features related to each annotation as new variables. Then, a first R markdown report is rendered to show steps and main results of the annotate-and-expand process, and
- D) This output is used for further integrative analyses (MFA, or RGCCA in the depicted example: E).

## Integrative Analyses with Dimension Reduction Techniques

Integrative multi-omics analysis, combining protein and gene expression from 150 breast cancer samples (data from TCGA, available in *miXomics*<sup>12</sup>), expanded using GO categories, and applying distinct dimension reduction techniques (Regularized Generalized Canonical Correlation Analysis and Multiple Factor Analysis) implemented with packages *miXomics* or *FactoMineR*<sup>13</sup> & *factoextra*<sup>14</sup>, respectively.



## References

- 1) Cavill C. et al. Briefings in Bioinformatics, 2016
- 2) Meng C. et al. Briefings in Bioinformatics, 2016
- 3) Leipzig J. Briefings in Bioinformatics, 2017
- 4) Wratten L. et al. Nature Methods, 2021
- 5) Galaxy Project: <https://usegalaxy.org/>
- 6) NextFlow: [www.nextflow.io](http://www.nextflow.io)
- 7) GO Consortium, Nucleic Acids Res., 2017
- 8) Subramanian, A. et al. PNAS, 2005
- 9) Lé Cao, K-A. et al. Bioinformatics, 2009
- 10) de Tayrac, M. BMC Genomics, 2009.
- 11) targets: <https://books.ropensci.org/targets/>
- 12) miXomics [www.mixomics.org](http://www.mixomics.org)
- 13) FactoMineR: <http://factominer.free.fr/>
- 14) factoextra: <https://rpkgs.datanovia.com/factoextra>

## Acknowledgements



**Ricardo Gonzalo**, former member of the  
Statistics & Bioinformatics Unit, VHIR

Work financially supported by the Spanish Ministerio de Ciencia e Innovacion, grant PID2019-104830RB-I00, and by the Departament d'Economia i Coneixement de la Generalitat de Catalunya, grant 2017SGR622 (GRBIO).

