

XVIII Congreso de Biometría CEBMADRID

25, 26 y 27 mayo 2022, Madrid

# ***Integrative Analysis of Multi-Omics Data with Addition of Biological Knowledge***

Ferran Briansó & Alex Sánchez-Pla  
Dpt. Genetics, Microbiology and Statistics



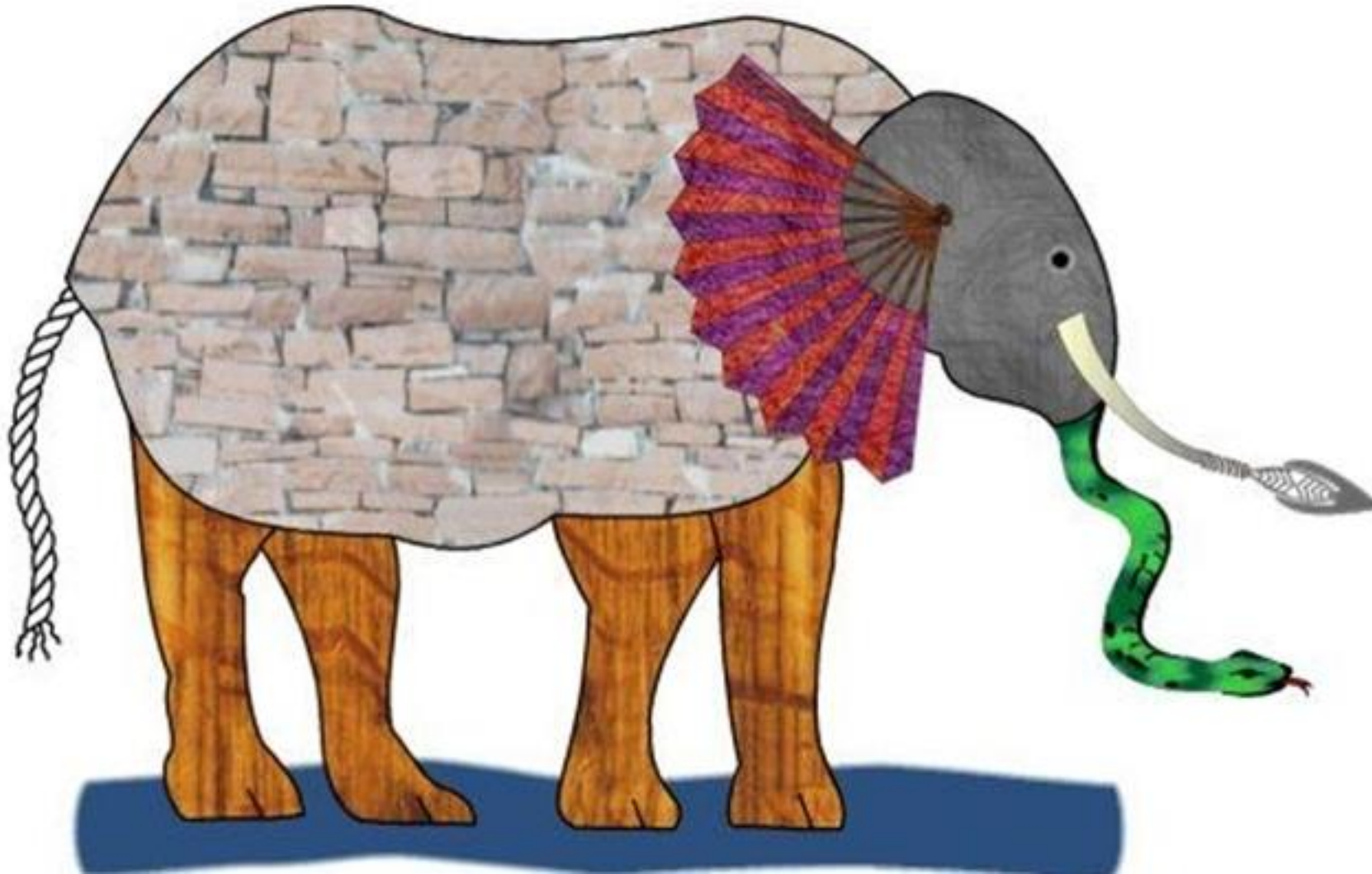
UNIVERSITAT DE  
BARCELONA



**Integrative analysis allows the combination of distinct omics data**

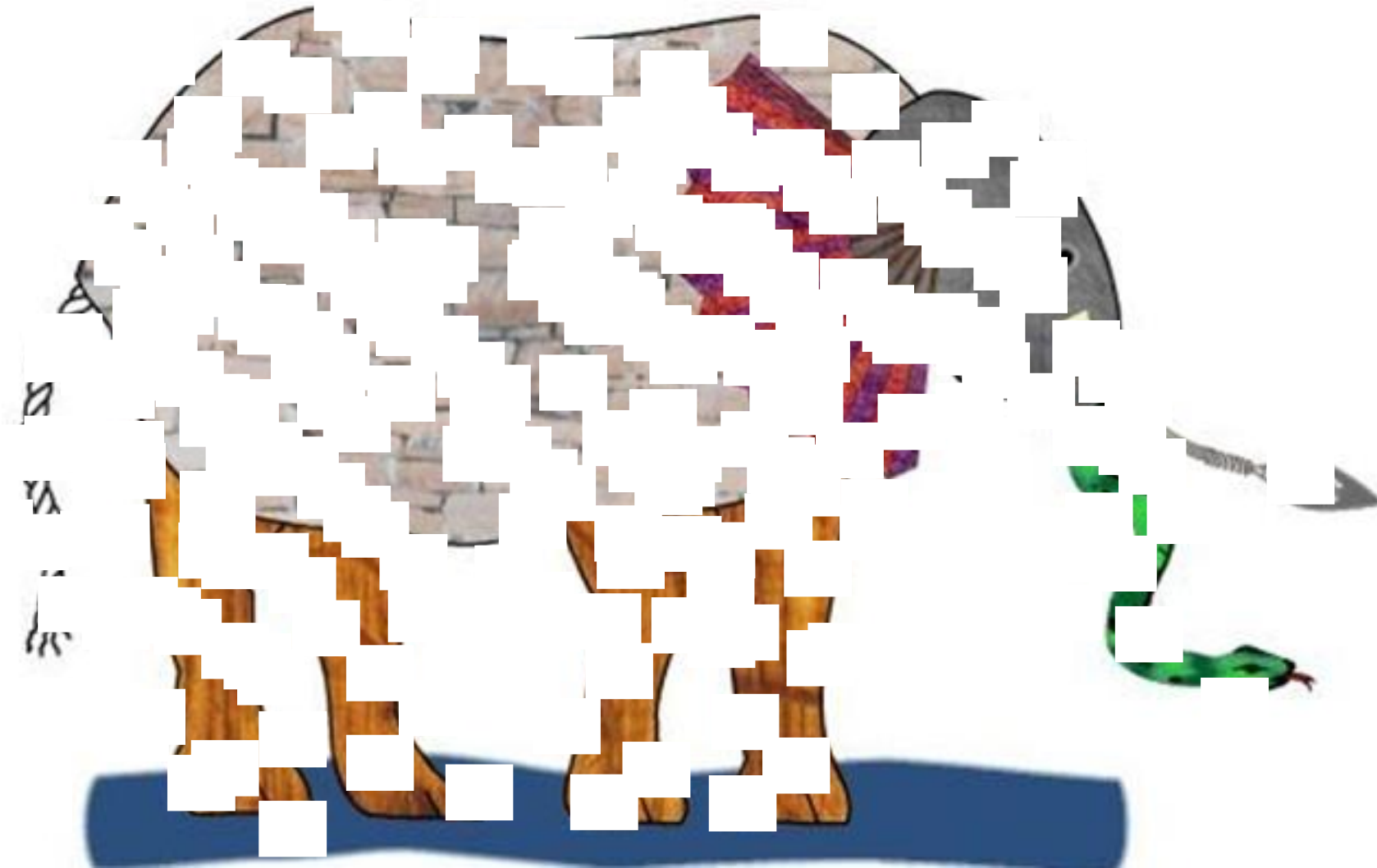


**Integrative analysis allows the combination of distinct omics data**



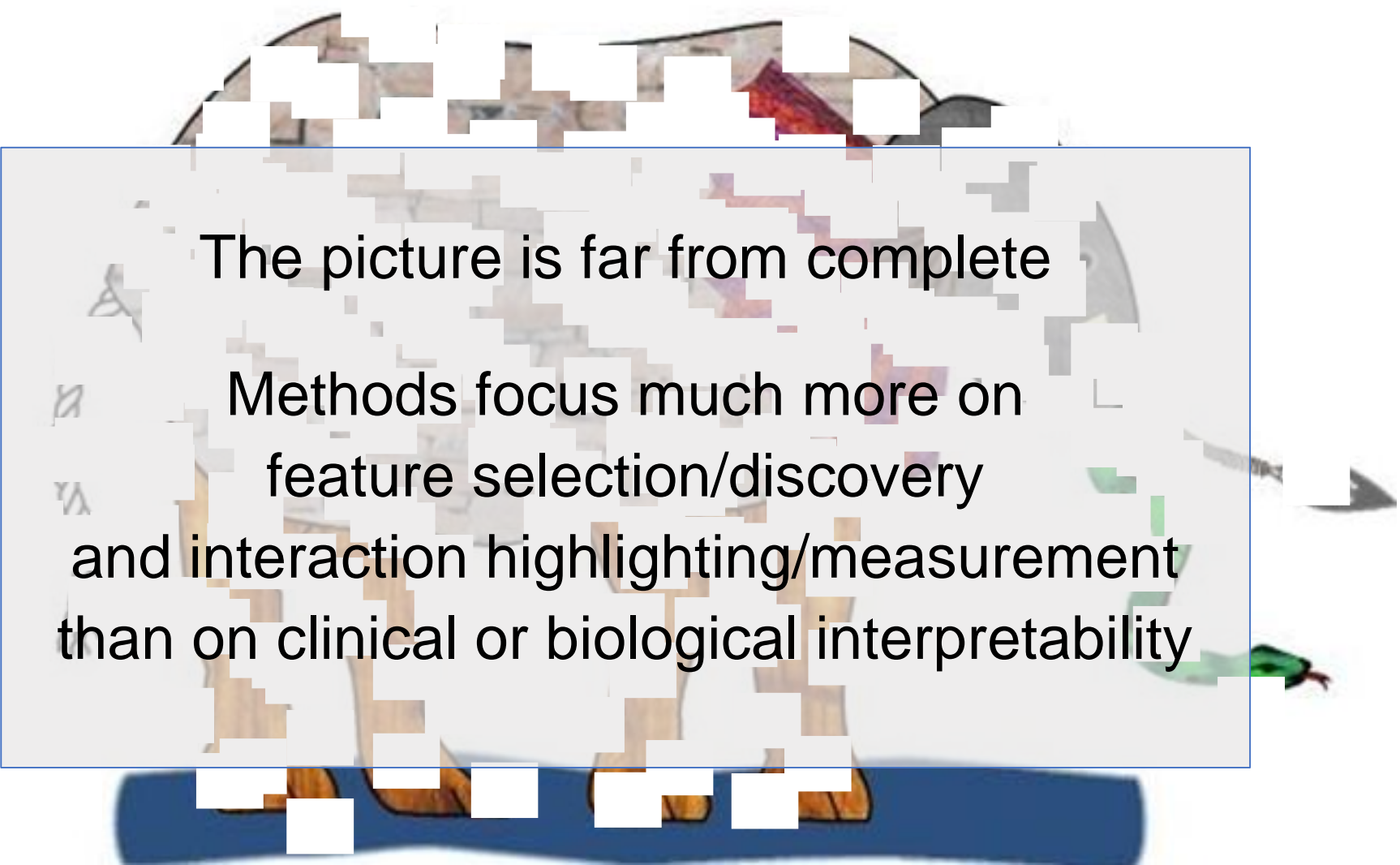
[https://en.wikipedia.org/wiki/Blind\\_men\\_and\\_an\\_elephant](https://en.wikipedia.org/wiki/Blind_men_and_an_elephant)

**But interpretability is a weak point of most multi-omics approaches**



[https://en.wikipedia.org/wiki/Blind\\_men\\_and\\_an\\_elephant](https://en.wikipedia.org/wiki/Blind_men_and_an_elephant)

## But interpretability is a weak point of most multi-omics approaches

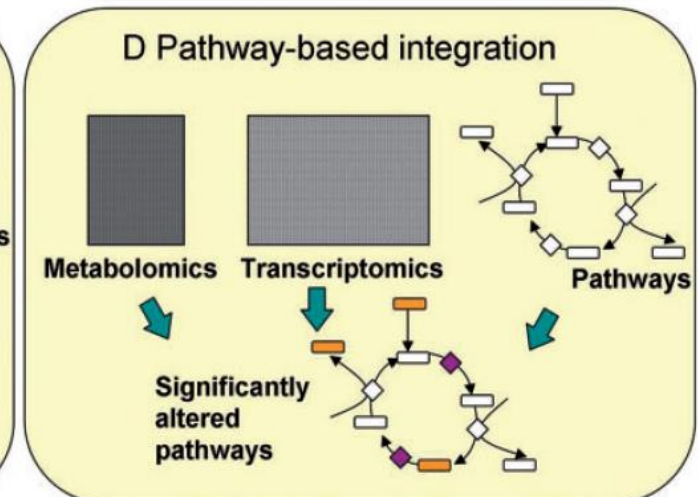
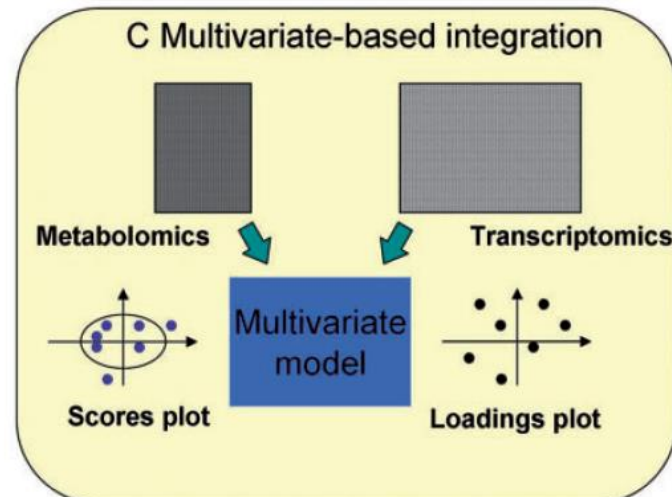
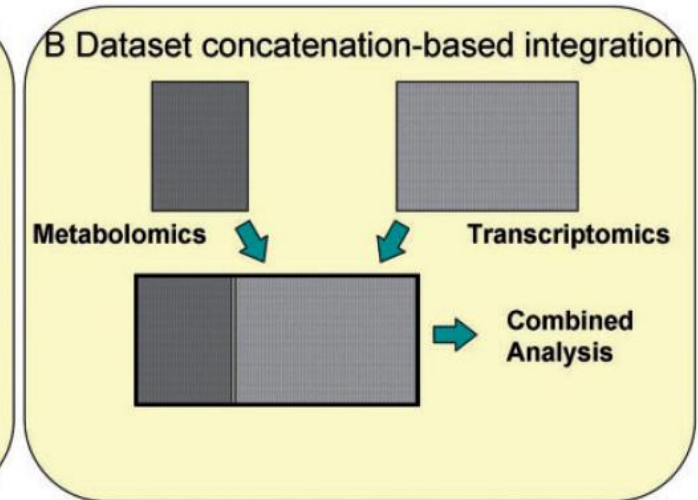
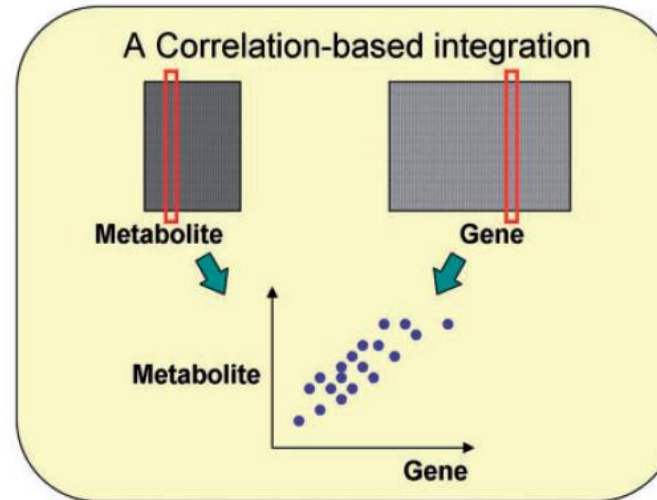
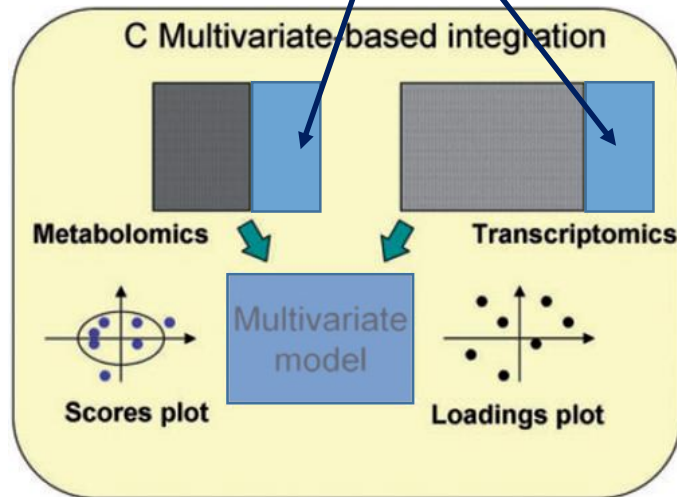


The picture is far from complete  
Methods focus much more on  
feature selection/discovery  
and interaction highlighting/measurement  
than on clinical or biological interpretability



# Some approaches for multi-omics data integration

Addition of Biological Annotations  
(GO, GeneSets, pathways...)  
**before the integrative steps**



# Our proposal

**1. Pre-process omics datasets in order to include biological information before the joint analysis**

## → Expanded datasets

Applying standard biological significance analysis methods  
(e.g. Gene over-representation; Gene Sets; Pathways; Networks... + Custom annotations)

#	#	s1	s3	s4	s5	s6	s7	
#1	g	9.823383	9.981502	10.911122	9.249729	11.459667	8.369157	9.560181
#2	g	9.027626	8.301891	8.509765	9.393560	8.936765	8	
#3	g	9.750415	9.249729	9.249729	9.358899	9.208426	9	
#4	g	12.074250	9.340945	9.826612	9.893645	10.266810	#1	0
#5	g	7.799255	9.574843	9.401317	10.238429	10.641396	9	0
#6	g	54.181059	9.249729	9.249729	9.249729	10.238429	#2	1
#7	g	8.985552	9.943802	10.046111	6.552113	11.458644	#4	1
#8	g	9.456599	10.460557	10.687289	9.395681	10.655145	#5	0
#9	g	8.774890	9.968802	9.937178	9.192533	9.648703	#6	1
#10	g	8.907701	9.249729	9.249729	9.249729	12.210678	#7	1
#11	g	8.973982	8.036014	10.687557	9.099134	10.862929	#1	0
#12	g	7.951261	12.231027	11.448632	8.127683	11.126290	#5	0
#13	g	7.543131	9.924511	11.811728	10.828903	7.744870	#10	0
#14	g	10.052646	9.249729	9.768817	9.302912	10.074701	#2	0
#15	g	11.119820	9.750415	11.246757	8.349641	9.730128	#11	0
#16	g	8.931945	10.295648	11.925690	9.282700	10.125574	#1	0
#17	g	9.697817						
#18	g	10.133594						
#19	g	10.658743						
#20	g	10.042697						
#1	g	1.560181	6.354711	0	1	0	0	1
#2	g	9.341271	10.030435	1	0	1	0	0
#3	g	10.722181	11.039361	1	0	1	0	0
#4	g	10.945357	9.068662	0	0	0	1	0
#5	g	10.978791	9.069396	1	0	0	0	1
#6	g	11.005452	9.997342	1	0	1	1	1
#7	g	11.308446	9.069396	1	0	0	0	0
#8	g	9.390803	9.727783	1	0	1	0	1
#9	g	11.373304	10.757272	1	1	1	0	0
#10	g	10.000457	11.275904	0	1	0	0	1
#11	g	10.799194	11.289403	1	1	1	0	0
#12	g	9.373252	12.705092	1	1	0	0	1
#13	g	9.625648	8.803307	0	0	0	0	0
#14	g	10.552339	8.632491	1	1	0	0	1
#15	g	9.757147	10.776161	0	0	0	0	1
#16	g	9.613065	10.925900	1	1	0	0	0
#17	g	8.888190	10.280053	0	1	0	0	1
#18	g	11.158111	11.109993	1	0	1	0	0
#19	g	10.642895	10.191623	0	0	0	0	0
#20	g	9.734463	9.151632	1	0	0	1	0
#1	111.109328	112.485958	111	NA	NA	NA	NA	NA
#2	121.288059	123.553432	121	NA	NA	NA	NA	NA
#3	115.672151	112.602470	NA	11	NA	NA	NA	NA
#4	132.611944	137.078157	NA	13	NA	NA	NA	NA
#5	72.404721	70.615121	NA	72	NA	NA	NA	NA
#6	73.907199	69.405752	NA	NA	NA	7	NA	NA
#7	83.193258	80.509283	NA	NA	NA	NA	8	NA
#8	78.016520	79.934604	NA	NA	NA	NA	NA	8
#9	80.100083	91.304700	NA	NA	NA	NA	NA	NA
#10	103.283346	103.038291	103	NA	NA	NA	NA	NA

# Our proposal

## 1. Pre-process omics datasets in order to include biological information before the joint analysis

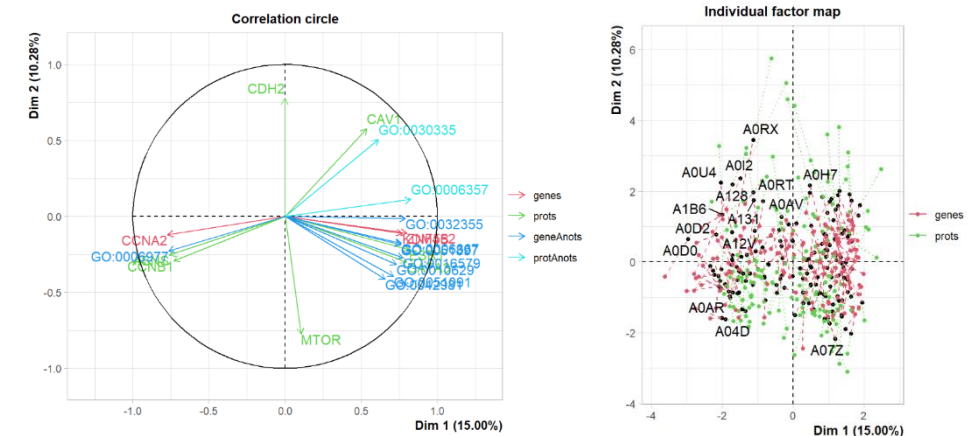
→ Expanded datasets

Applying standard biological significance analysis methods  
(e.g. Gene over-representation; Gene Sets; Pathways; Networks... + Custom annotations)

	s1	s2	s3	s4	s5	s6	s7		c1	c2	c3	c4	c5	c6	c7	c8	c9	c10
## g1	9.823383	9.981509	10.911122	9.249729	11.459667	8.369157	9.560181	##	0	1	1	0	0	1	1	1	0	
## g2	9.987826	8.301891	8.509712	9.330750	9.836765	8.369157	9.560181	##	0	1	1	0	0	1	1	1	0	
## g3	9.750415	7.351119	8.204226	9.889991	10.679750	10.266810	10.266810	##	0	1	1	0	0	1	1	1	0	
## g4	12.027450	9.340945	9.826619	9.893645	10.266810	10.266810	10.266810	##	0	1	1	0	0	1	1	1	0	
## g5	7.799825	9.587443	9.410317	10.238429	10.841196	9.395601	10.655145	##	0	1	1	0	0	1	1	1	0	
## g6	10.541859	10.960931	8.932551	12.739604	9.231603	9.395601	10.655145	##	0	1	1	0	0	1	1	1	0	
## g7	9.886592	9.943082	10.846111	9.652113	11.450464	10.266810	10.266810	##	0	1	1	0	0	1	1	1	0	
## g8	9.465699	10.460557	10.687289	9.395601	10.655145	9.395601	10.655145	##	0	1	1	0	0	1	1	1	0	
## g9	8.774980	9.968802	9.973178	9.192533	9.684703	9.395601	10.655145	##	0	1	1	0	0	1	1	1	0	
## g10	8.987706	9.486889	9.925574	9.981692	12.210547	9.395601	10.655145	##	0	1	1	0	0	1	1	1	0	
## g11	9.897302	9.035014	10.687557	9.099124	9.807929	11.450464	10.266810	##	0	1	1	0	0	1	1	1	0	
## g12	9.751261	12.231207	11.440632	8.127683	11.126290	9.395601	10.655145	##	0	1	1	0	0	1	1	1	0	
## g13	7.541331	9.924513	11.811728	10.828893	7.744879	10.266810	10.266810	##	0	1	1	0	0	1	1	1	0	
## g14	10.052642	9.143740	9.786817	9.392912	10.074336	10.266810	10.266810	##	0	1	1	0	0	1	1	1	0	
## g15	11.119820	9.750373	11.246757	8.349641	7.301828	9.395601	10.655145	##	0	1	1	0	0	1	1	1	0	
## g16	9.831945	10.295648	11.664303	9.928730	10.125573	10.266810	10.266810	##	0	1	1	0	0	1	1	1	0	
## g17	9.697817	9.560181	8.354711	0	1	0	0	1	1	0	0	0	0	1	1	0	0	
## g18	10.343394	9.341271	10.030345	1	0	1	0	0	0	1	1	0	0	0	1	0	0	
## g19	10.610743	10.277298	11.329309	1	1	0	0	0	1	0	1	0	1	0	1	1	0	
## g20	10.642697	10.545357	9.680662	0	0	0	1	0	0	1	0	0	1	0	0	1	0	
## g21	10.724463	9.069306	1	1	0	0	1	1	0	0	1	0	0	1	0	0	1	
## g22	11.045247	8.997348	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	
## g23	11.308446	10.980477	0	1	1	0	1	0	1	0	0	1	0	0	1	0	0	
## g24	9.308083	9.727783	1	1	0	1	0	1	0	1	0	1	0	1	0	1	0	
## g25	11.373304	10.752722	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
## g26	10.808457	11.275904	0	1	0	1	0	0	1	0	0	1	0	0	1	0	0	
## g27	10.799104	11.289403	0	1	1	0	1	0	1	0	0	1	0	0	1	0	0	
## g28	9.337232	12.705092	1	1	0	1	0	1	1	0	0	1	0	0	1	0	0	
## g29	9.625649	8.803367	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	
## g30	10.552339	8.632491	1	1	0	0	1	0	0	1	0	0	1	0	0	0	0	
## g31	9.757147	10.776161	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	
## g32	9.613965	10.929590	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	
## g33	8.088190	10.200053	0	1	0	1	1	0	0	0	0	1	0	0	0	1	1	
## g34	11.158111	11.109901	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	
## g35	10.642895	10.191623	0	0	1	1	1	1	0	1	0	1	0	1	0	1	0	
## g36	9.724463	9.151632	1	1	0	0	1	1	0	1	0	1	0	1	0	1	0	
## A1	111.193928	112.458958	11	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## A2	122.288059	123.553432	NA	12	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## A3	115.672151	112.602470	NA	NA	11	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## A4	132.611844	137.078157	NA	NA	NA	7	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## A5	72.494721	70.615121	NA	NA	NA	7	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## A6	73.507199	69.407572	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## A7	83.193258	80.569283	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## A8	78.816520	79.934694	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## A9	89.100083	91.384790	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
## A10	103.263346	103.018521	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

## 2. Analysis of the expanded datasets by the use of contrasted joint Dimension Reduction techniques

- MCIA (*made4*, *omicade*...)
- RGCCA (*miXomics*...)
- Multiple Factor Analysis (*FactoMineR*...)





# Our proposal

- # 1. Pre-process omics datasets in order to include biological information before the joint analysis

## → Expanded datasets

Applying standard biological significance analysis methods  
(e.g. Gene over-representation; Gene Sets; Pathways; Networks... + Custom annotations)

- ## 2. Analysis of the expanded datasets by the use of contrasted joint Dimension Reduction techniques

→ MCIA (*made4, omicade...*)

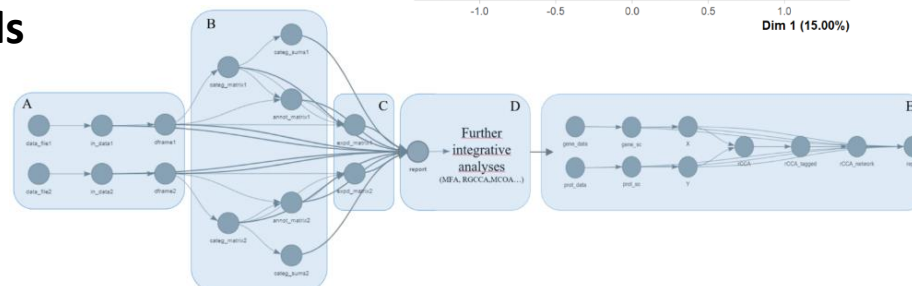
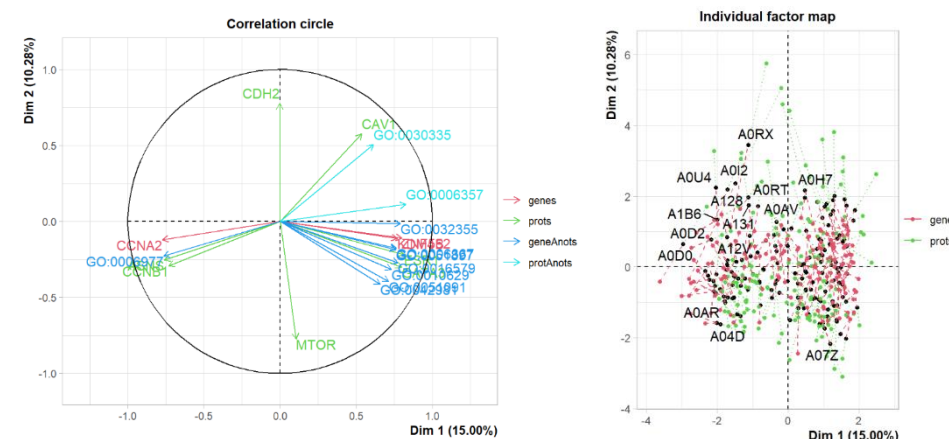
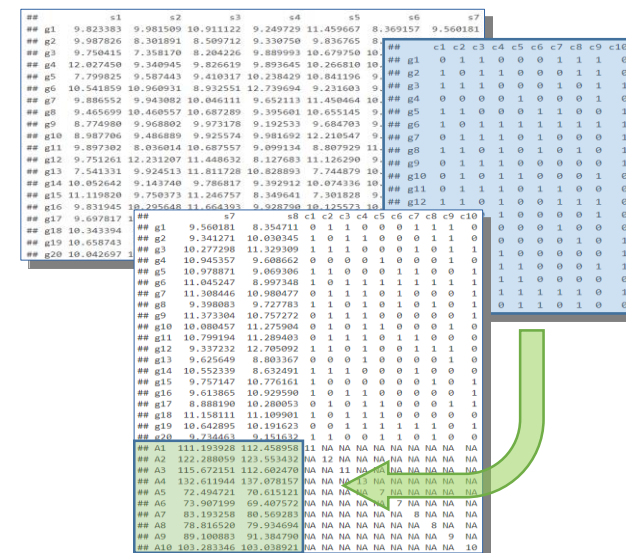
→ RGCCA (*miXomics...*)

### → Multiple Factor Analysis (*FactoMineR*...)

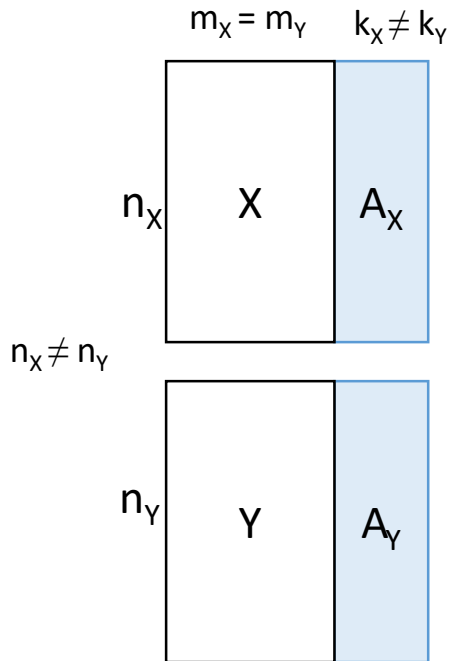
- ### 3. Process (semi-)automation in ease-to-use tools

## → Pipeline for workflow control / visualization

→ R Package + Interactive application



# Overview of the methods – Biological Annotations

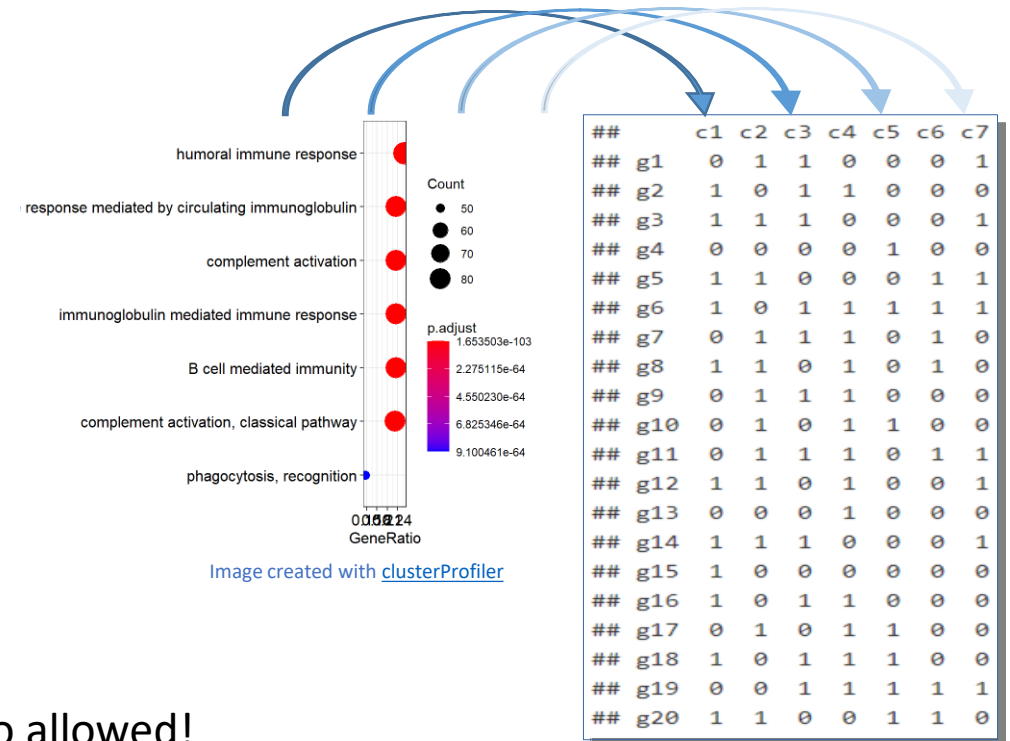


n : features (eg. gene ids, different for X and Y)  
m : samples (individuals, same ones for X and Y)  
k : biological annotations (different for X and Y)

## • Gene enrichment (over-representation) | Gene Set Enrichment Analysis

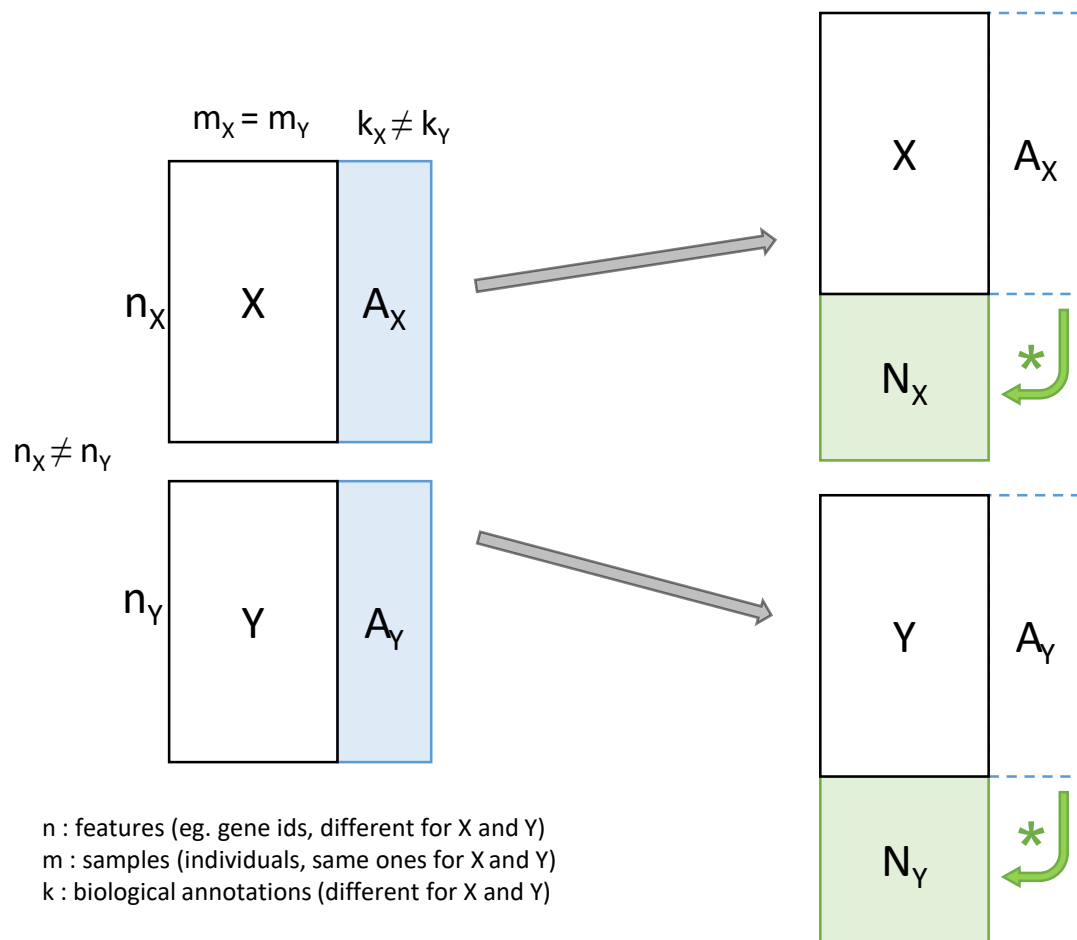
### • Annotations to

- Gene Ontology (GO)
- Gene Sets (MSigDB)
- Networks (DisGeNET...)
- Pathways (KEGG, Reactome)
- other databases (MeSH...)



and custom annotations are also allowed!

## Overview of the methods – Data Expansion (I)

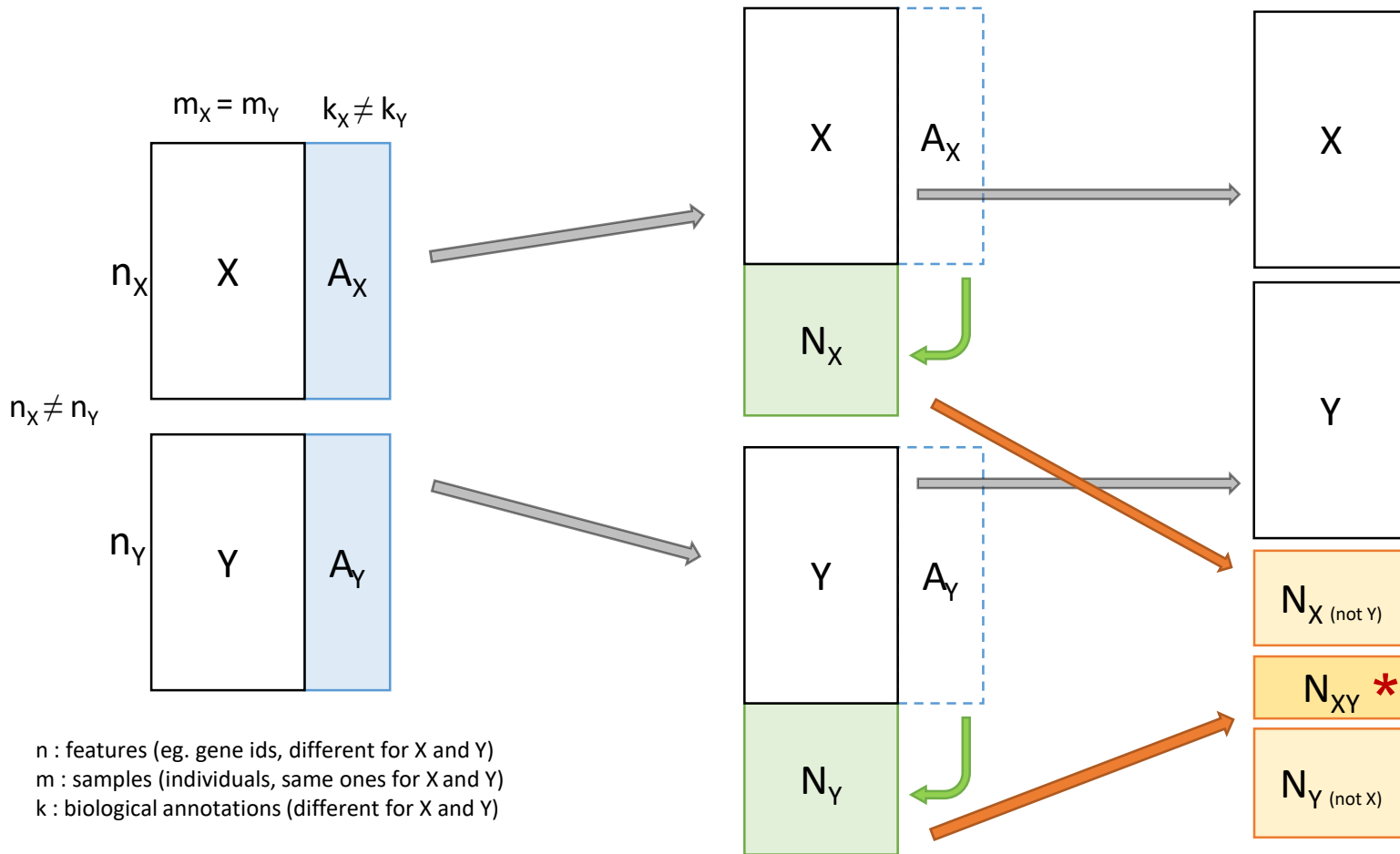
[illegible]

$$N_G = \phi(G, A_G), \quad G \in X, Y, \dots,$$

$$N_X = \frac{1}{r_X}(X \times A'_X),$$

$$N_Y = \frac{1}{r_Y}(Y \times A'_Y),$$

# Overview of the methods – Data Expansion (II)



\* Distinct weights can be applied for factors X and Y of the common N<sub>XY</sub> expanded matrix (e.g. in the same way as datasets are combined in MFA).

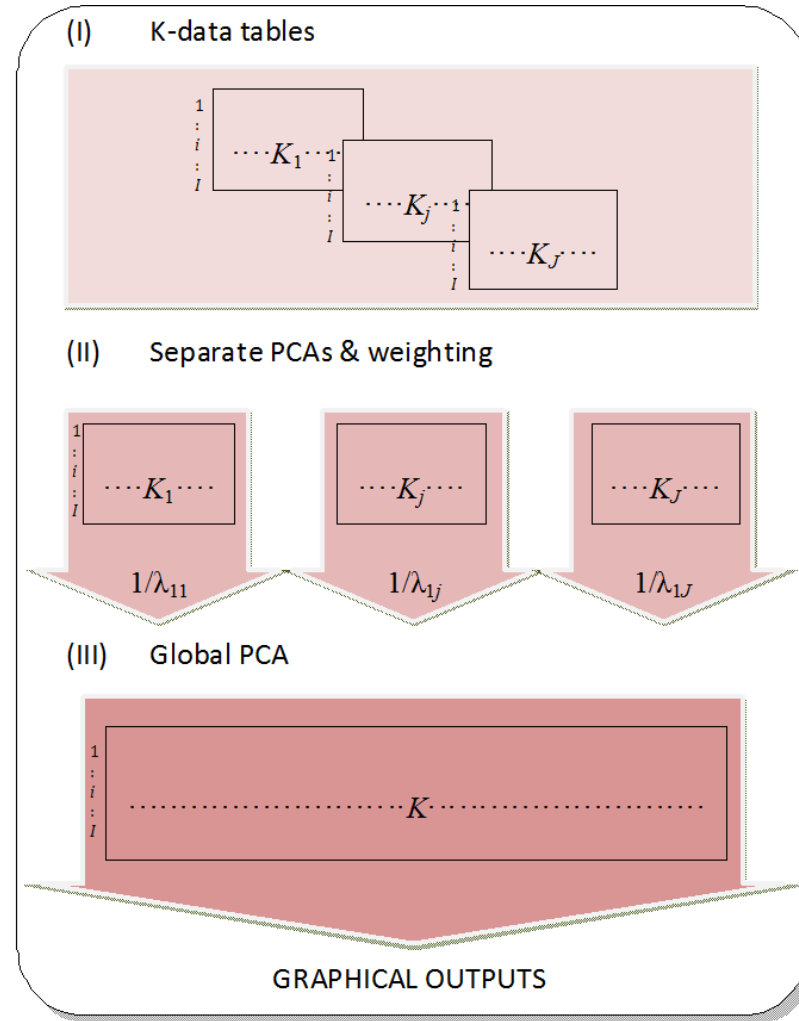
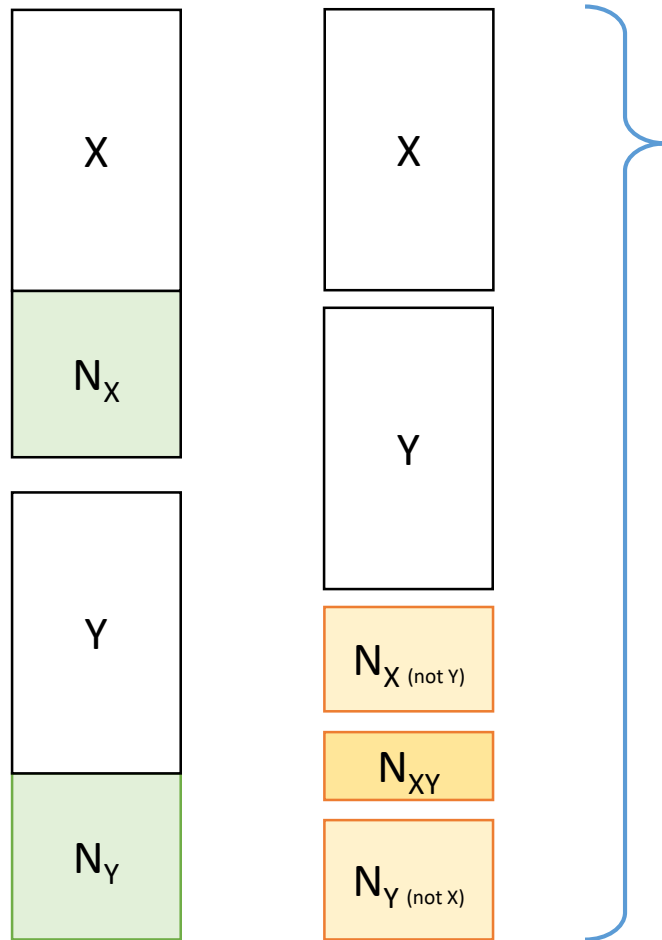
$$N_G = \phi(G, A_G), \quad G \in X, Y, \dots,$$

$$N_X = \frac{1}{r_X} (X \times A'_X),$$

$$N_Y = \frac{1}{r_Y} (Y \times A'_Y),$$

$$N_{XY} = \frac{w_X^*}{r_{XY}} (X \times A'_{XY}) + \frac{w_Y^*}{r_{XY}} (Y \times A'_{XY}).$$

# Overview of the methods – Multi-Block Analysis



## e.g. Multiple Factor Analysis

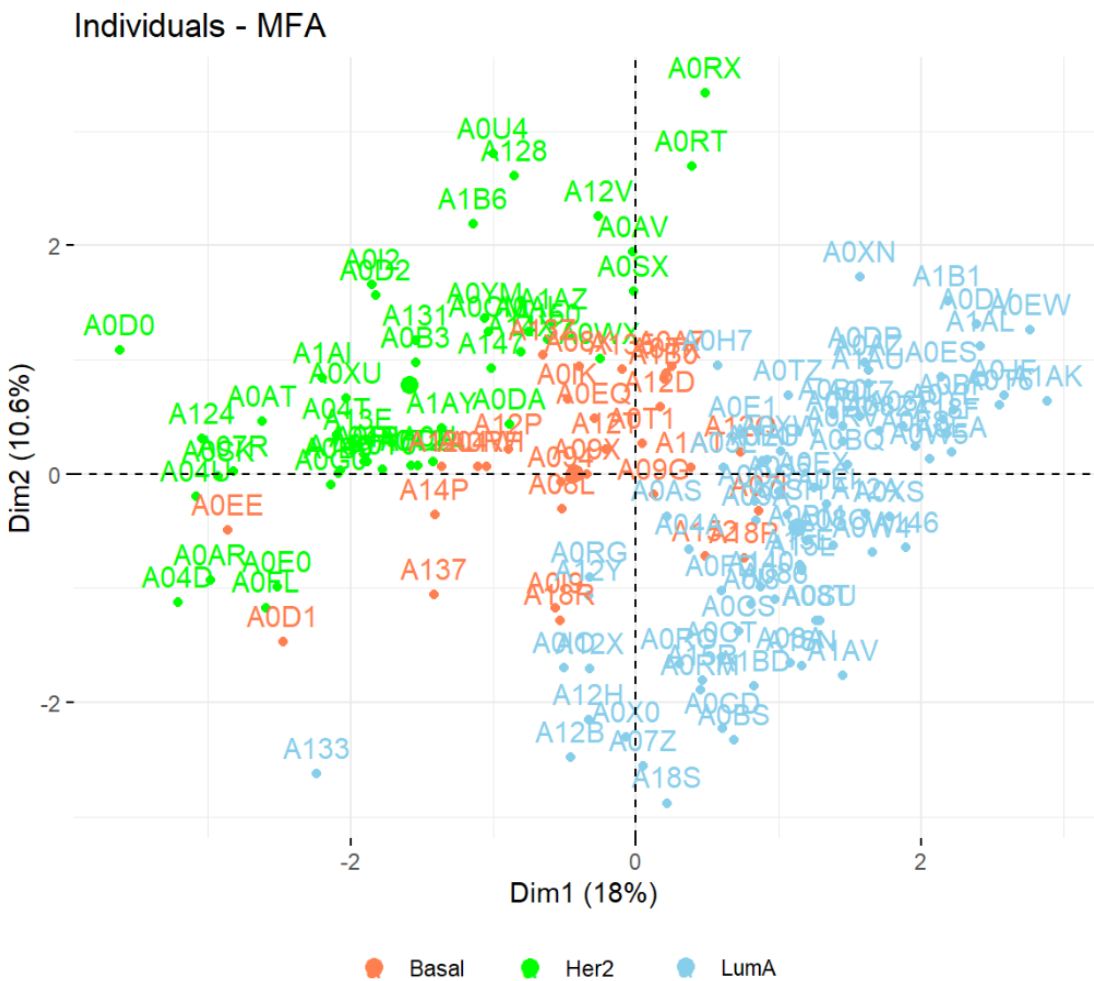
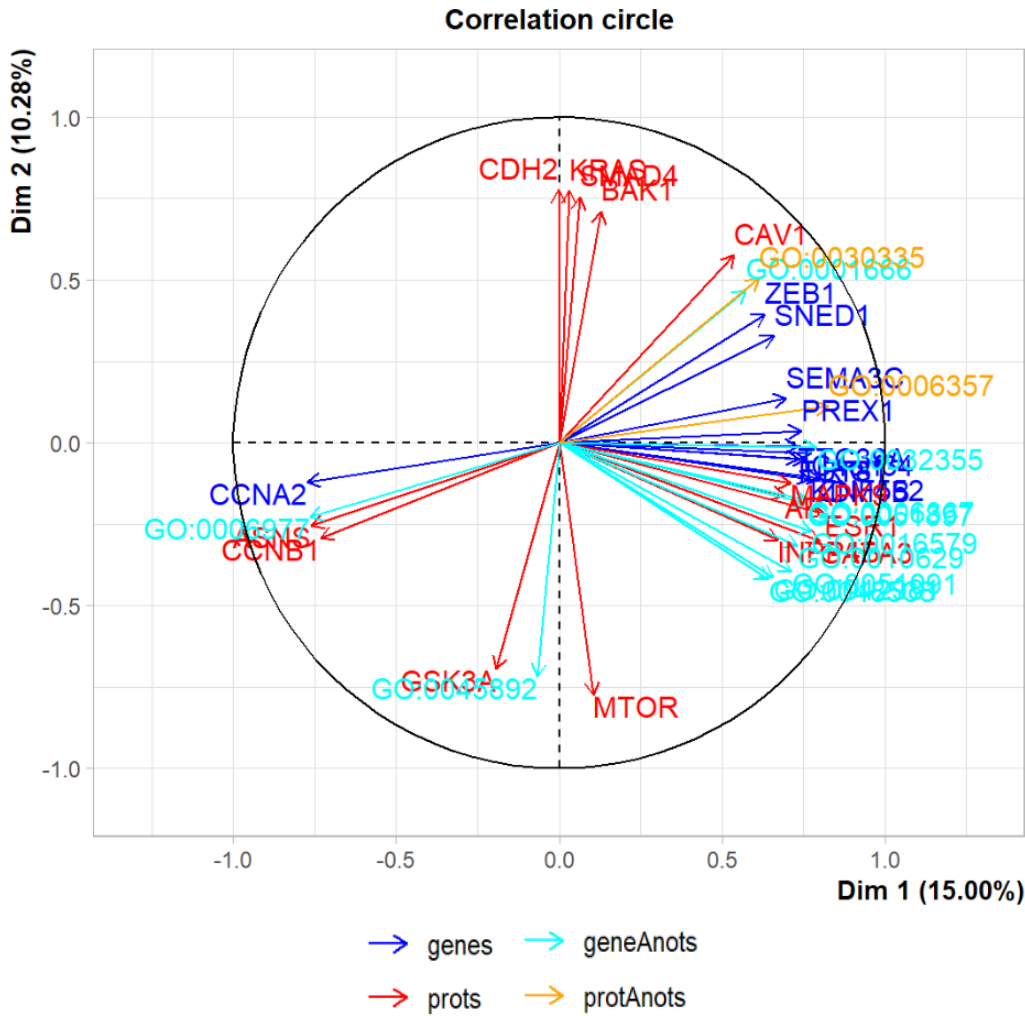
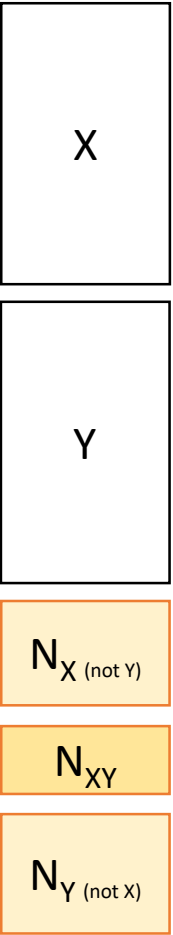
The core of MFA is a PCA applied to the whole set of variables,

Each group of variables is weighted, rendering possible the analysis of different points of view by taking them equally into account.

MFA allows to look for common factors by providing a representation of each matrix of variables.

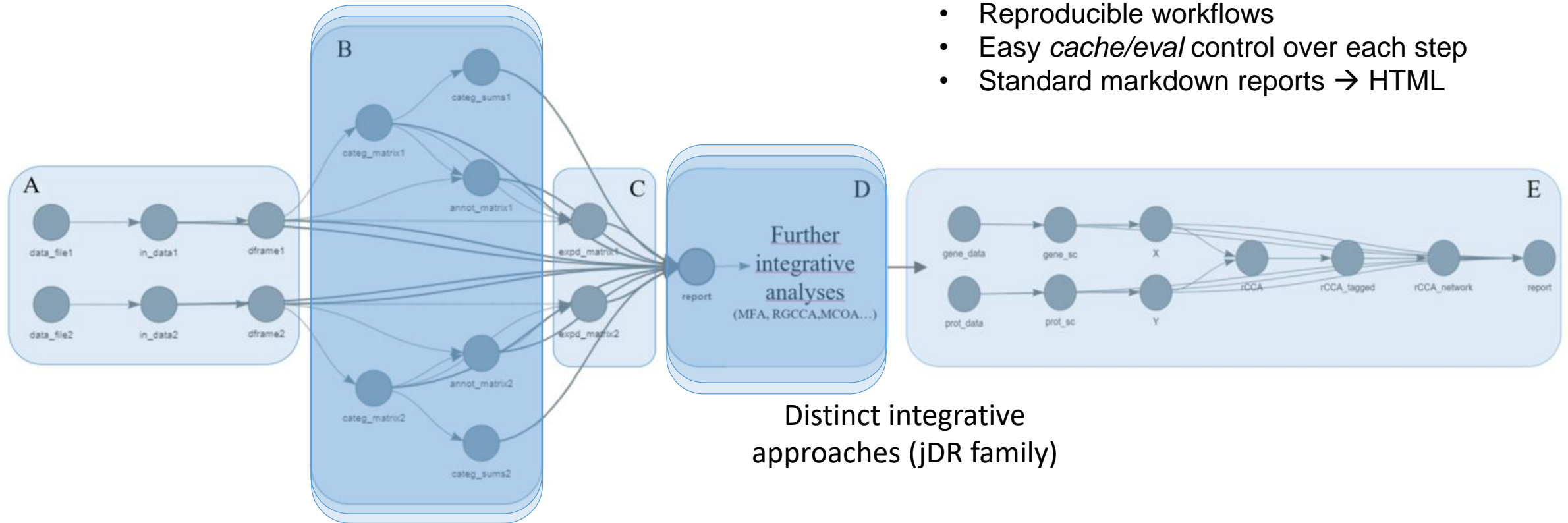


# Example results – 150 TCGA breast cancer samples – MFA



# Example results – *Targets* Pipeline Implementation

Reproducible workflow with a make-like pipeline toolkit *targets* package (<https://books.ropensci.org/targets/>)



- Reproducible workflows
- Easy *cache/eval* control over each step
- Standard markdown reports → HTML

Flexible annotation  
& expansion options

Distinct integrative  
approaches (jDR family)



# Conclusions

- ❑ The need for a **better biological interpretation of multi-omics integrative methods** let us to consider the inclusion of biological information during (not after) the analysis process
- ❑ We propose a method focused on the **expansion of the starting omics datasets, by adding new annotation-derived features to that matrices**, before applying the integrative analysis
- ❑ This approach allows the inclusion of relevant information from the **main biological annotation tools**, as well as any custom annotation, **combined with the use our preferred Dimension Reduction techniques**
- ❑ We have implemented a **pipeline for reproducible and easy-to-use execution**, that facilitates the control of each step, the visualization of results and their reporting to PDF/HTML formats

# Acknowledgements

Staff of the [Statistics and Bioinformatics Unit \(UEB, VHIR\)](#)

Special thanks to **Ricardo Gonzalo Sanz**



[Neurovascular Diseases Laboratory, VHIR](#)

**Teresa García-Berrocso, Alba Simats, Joan Montaner**



[Grup de Recerca en Bioestadística i Bioinformàtica, GRBIO](#)

Universitat de Barcelona & Universitat Politècnica de Catalunya



Spanish Ministerio de Ciencia e Innovación, PID2019-104830RB-I00



Departament d'Economia i Coneixement de la Generalitat de Catalunya,  
2017SGR622



# XVIII Congreso de Biometría CEBMADRID

25, 26 y 27 mayo 2022, Madrid

## *Integrative Analysis of Multi-Omics Data with Addition of Biological Knowledge*

Ferran Briansó & Alex Sánchez-Pla

# *THANKS FOR YOUR ATTENTION*



UNIVERSITAT DE  
BARCELONA

