



**HAL**  
open science

## Leveraging multi-omics data integration in the study of Chornobyl tree frogs

Elen Goujon, Olivier Armant, Mélodie Noyau, Sandrine Frelon, Luc Camoin,  
Jean-Marc Bonzom, Arthur Tenenhaus, Imène Garali

► **To cite this version:**

Elen Goujon, Olivier Armant, Mélodie Noyau, Sandrine Frelon, Luc Camoin, et al.. Leveraging multi-omics data integration in the study of Chornobyl tree frogs. ICRER2024 - 6th International Conference on Radioecology & Environmental Radioactivity, Nov 2024, Marseille, France. . irsn-04835908

**HAL Id: irsn-04835908**

**<https://asnr.hal.science/irs-n-04835908v1>**

Submitted on 13 Dec 2024

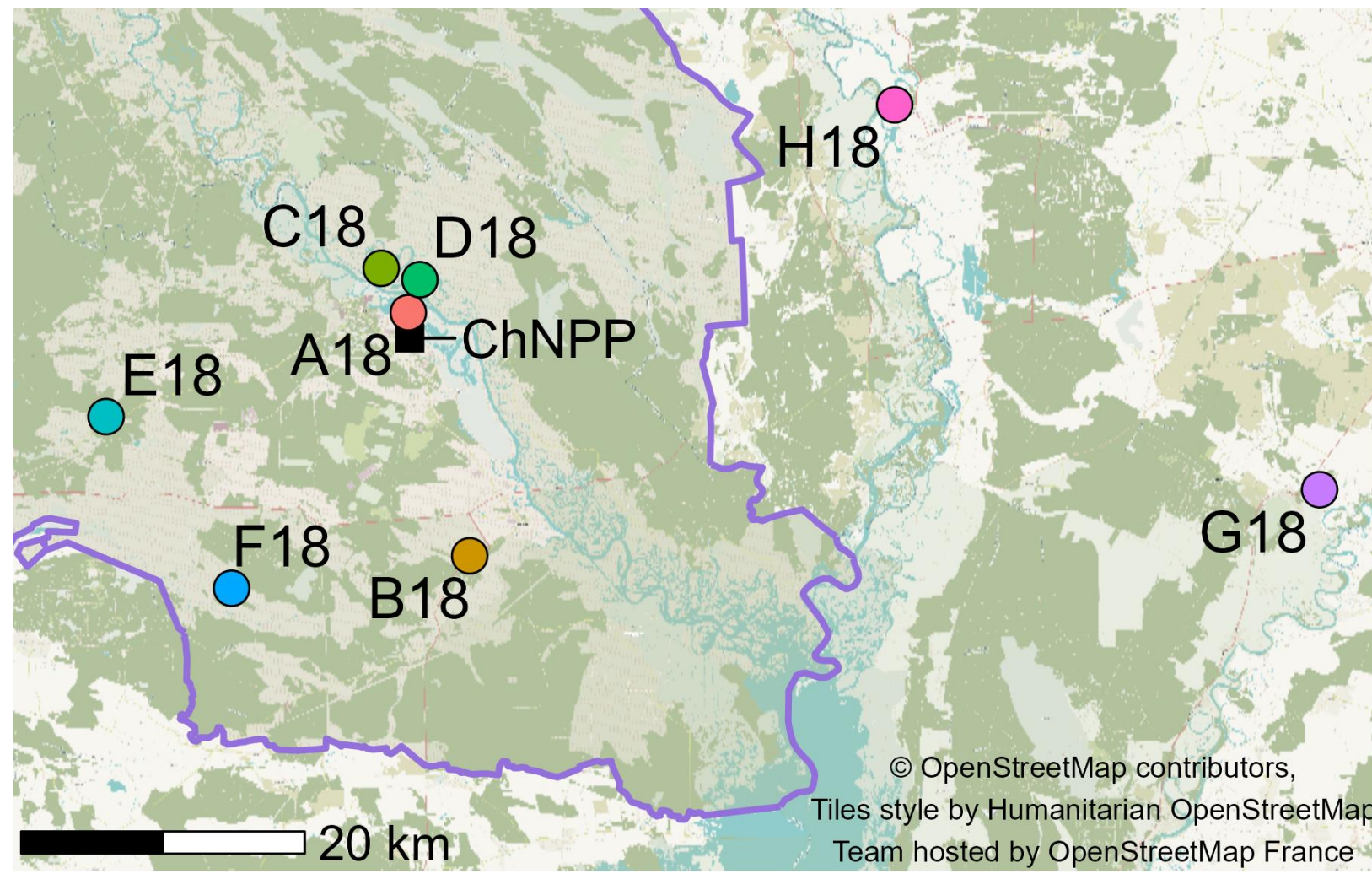
**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

## Context

**1986:** Nuclear power plant accident in Chernobyl, Ukraine: release of ionizing radiation and environmental contamination



**Fig1.** Sampling locations of Eastern tree frog *Hyla orientalis* populations in the CEZ in 2018 (ChNPP = Chernobyl nuclear power plant, — = CEZ) [1]

**2018:** Collection of tree frog population samples

- Male tree frogs sampled from 8 sites, including 6 within the Chernobyl Exclusion Zone (CEZ)
- Exposed to various levels of contamination (ambient dose rate range: 0.044 to 16.2  $\mu\text{Gy/h}$ )
- Omics and non-omics data extraction (n=94):

- Dosimetry: total dose rate with internal and external contributions (in  $\mu\text{Gy/h}$ )
- Age, phenotype (mass, dimensions)
- Genomics (n=87)
- Transcriptomics (mRNA sequencing) (n=87)
- Proteomics (n=76)

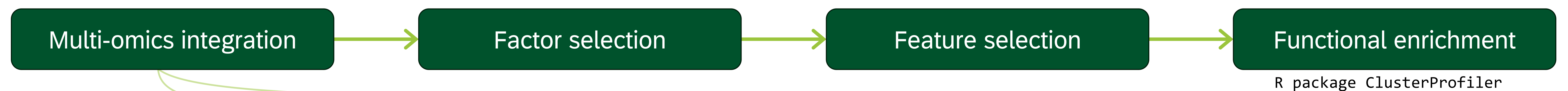
## Objectives

In previous studies, Chernobyl tree frogs have been examined from different perspectives [1-3]. These research efforts have demonstrated various adverse effects in the frogs living in the most contaminated areas and have produced several datasets.

In the presented work, we aim to **identify molecular signatures of exposure** to advance the understanding of low-dose radiation effects on wildlife. We propose an **integrative approach based on the multi-omics datasets** to provide a holistic view of the effects of chronic exposure to low-dose radiation.

Here, we study the **transcriptomics and proteomics datasets** using two integration methods based on joint dimensionality reduction (**R/SGCCA** and **MOFA**). We then extract factors and select variables to perform functional enrichment analysis.

## Methods

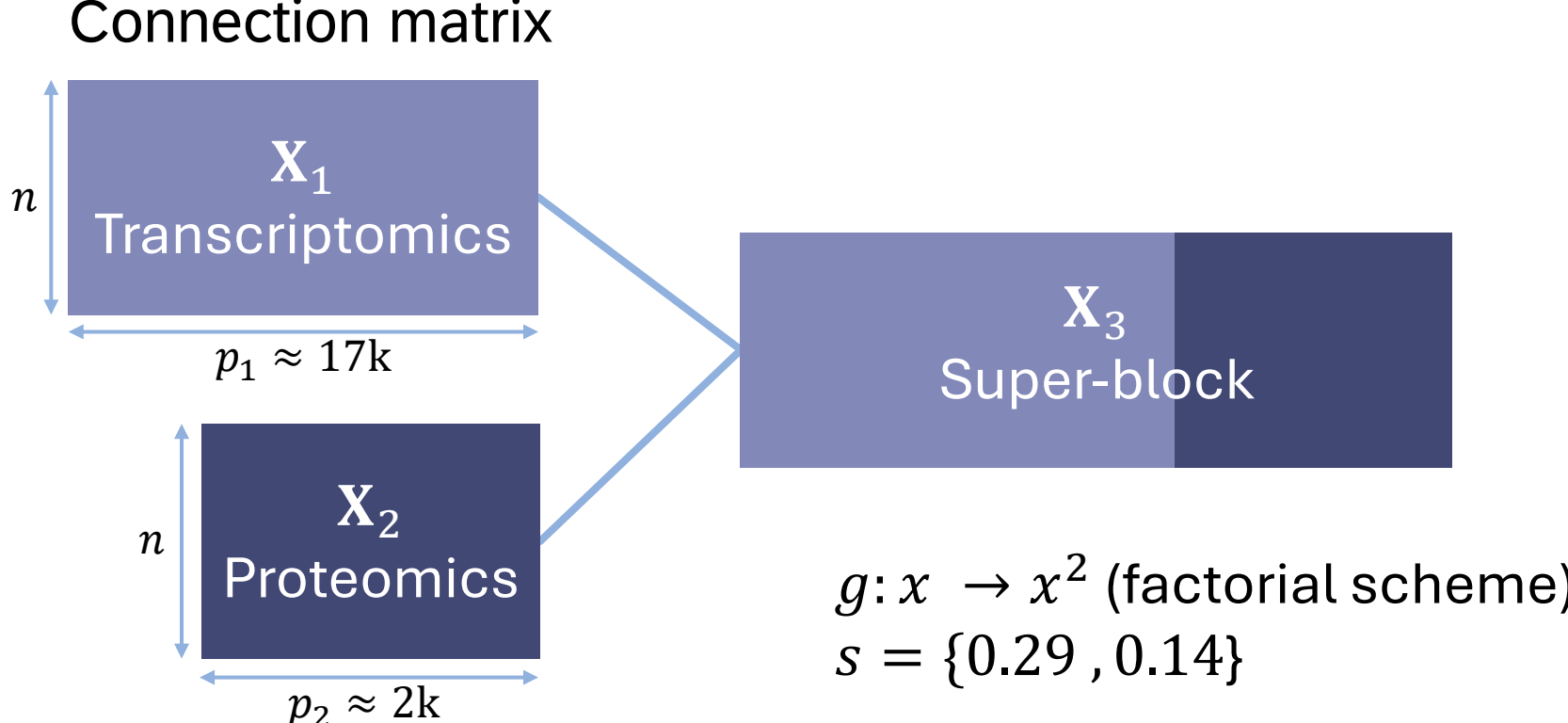


### SGCCA

Regularized Generalized Canonical Correlation Analysis with variable selection [4]  
*Multivariate relationship modeling in a multi-block structure*  
R package RGCCA

$$\begin{cases} \operatorname{argmax}_{\mathbf{w}_1, \dots, \mathbf{w}_J} \sum_{k,l=1, k \neq l}^J c_{kl} g(\operatorname{cov}(\mathbf{X}_k \mathbf{w}_k, \mathbf{X}_l \mathbf{w}_l)) \\ \text{subject to } \|\mathbf{w}_j\|_2 = 1 \text{ and } \|\mathbf{w}_j\|_1 \leq s_j, j = 1, \dots, J \end{cases}$$

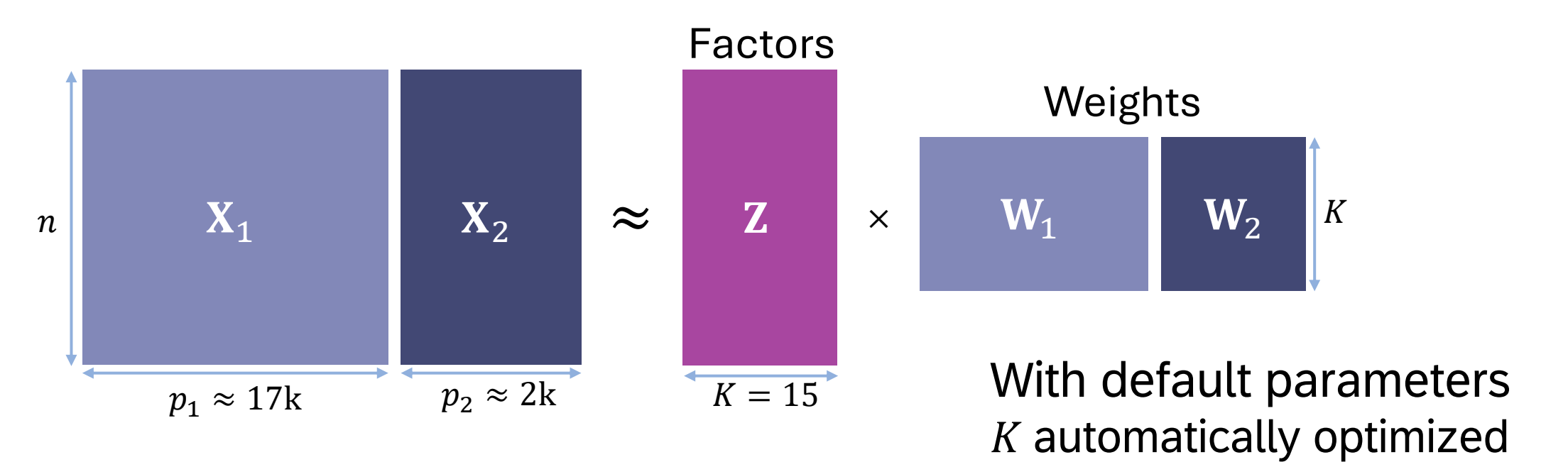
Tested parameters:  
Connection matrix



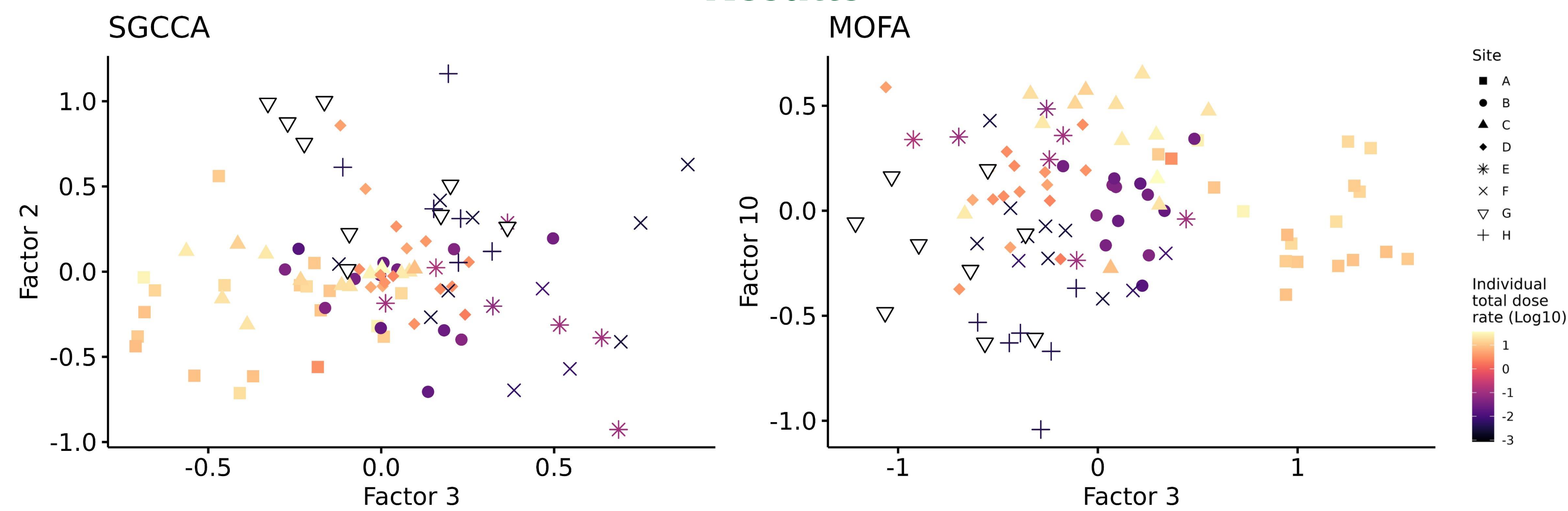
### MOFA

Multi-Omics Factor Analysis [5]  
*Identification of common factors based on Bayesian modelling*  
R package MOFA2

$$\mathbf{X}_j = \mathbf{Z} \mathbf{W}_j^T + \varepsilon_j$$



## Results



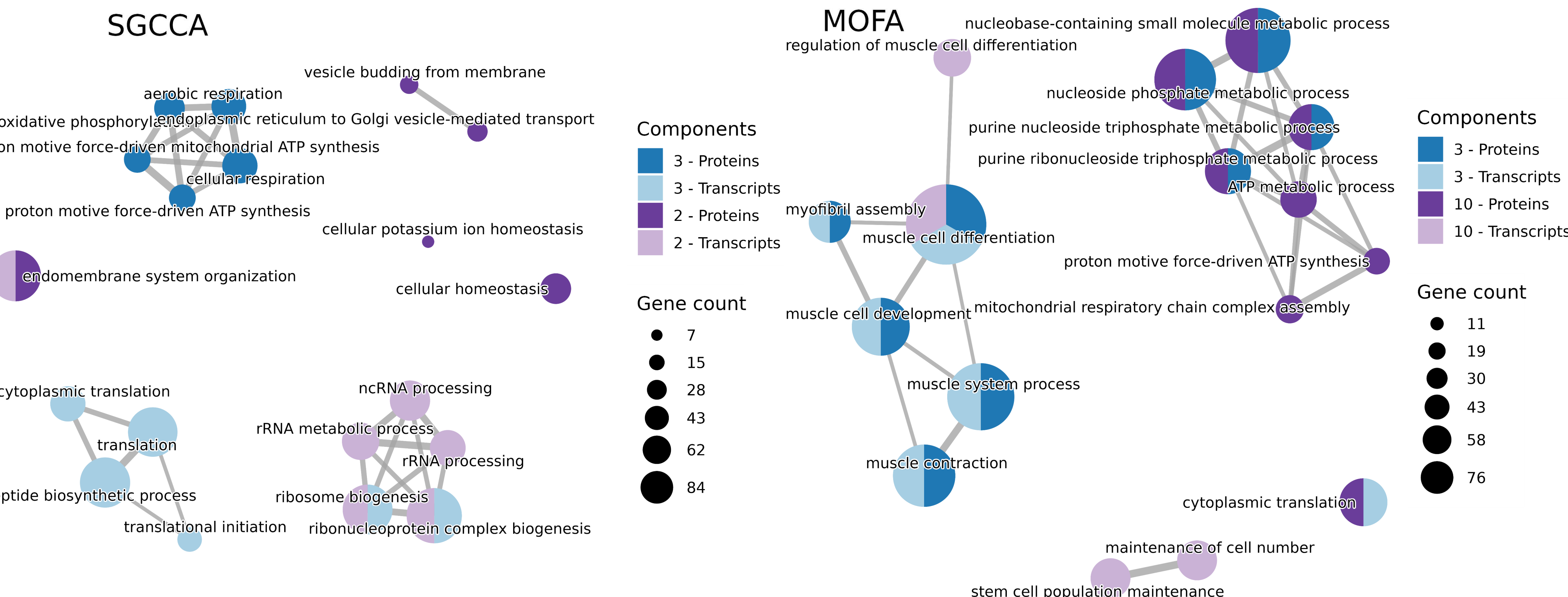
**Fig2.** Individuals represented in the plane of selected factors. Factors were selected according to the highest and second highest correlation with individual dose rate, using the absolute value of Spearman correlation. **(left)** SGCCA superblock common scores of ranks 3 and 2. **(right)** MOFA factors of ranks 3 and 10.

## Discussion & Conclusion

- With both integration strategies, feature selection and functional enrichment brought out **biological pathways that are consistent with previous findings** (in particular, energy metabolism and muscle processes) [1,2].
- **Components capture information that is either shared** between datasets or factors (mostly the case with MOFA's output), **or complementary** between modalities (as seen with SGCCA components).
- **Biological pathways related to energy metabolism are not enriched in the transcriptomics components**, although they were in the simple omics analysis [1]. Further analysis could make the analysis more reliable by selecting more variables and using bootstrap resampling to check their stability.

## Perspectives

- Design a new methodological approach to perform **joint dimensionality reduction while simultaneously accounting for confounding variation**, instead of the usual 2-step correction approach [1].
- **Add another molecular level to the integrative analysis:** genomic information was also obtained and studied in [2] in the form of SNPs of the transcriptome, and we plan on enriching the multi-omics integration study with this dataset.
- Compare results obtained from integrative method vs single-omic studies using a common protocol to **test the holistic approach's added value** and check each model's robustness.



**Fig3.** Networks of Gene Ontology biological pathway terms enriched in the variables that contribute the most to the selected factors. **(left)** Enrichment analysis performed on variables with non-zero weights in the SGCCA factors 3 and 2 for the Transcriptomics (~500 selected variables) and Proteomics blocks (~200 selected variables). **(right)** Enrichment analysis performed on variables with highest weights in absolute value in MOFA common factors 3 and 10 (top 500 for transcripts and top 200 for proteins).

### References

- [1] Goujon, E., Armant, O., Car, C., Bonzom, J.M., Tenenhaus, A., Garali, I. (2024). Batch Effect Correction in a Confounded Scenario: a Case Study on Gene Expression of Chernobyl Tree Frogs. *Computational Methods in Systems Biology*. LNCS, 14971. [https://doi.org/10.1007/978-3-031-71671-3\\_8](https://doi.org/10.1007/978-3-031-71671-3_8)
- [2] Car, C., Gilles, A., Goujon, E. et al. (2023). Population transcriptogenomics highlights impaired metabolism and small population sizes in tree frogs living in the Chernobyl Exclusion Zone. *BMC Bioinformatics*, 24, 164. <https://doi.org/10.1186/s12915-023-01659-2>
- [3] Burraco, P., Car, C., Bonzom, J.M. et al. (2021) Assessment of exposure to ionizing radiation in Chernobyl tree frogs (*Hyla orientalis*). *Sci Rep* 11, 20509. <https://doi.org/10.1038/s41598-021-00125-9>
- [4] Tenenhaus, A., Philippe, C., Guillemot, V. et al. (2014). Variable selection for generalized canonical correlation analysis. *Biostatistics*, 15, 3. <https://doi.org/10.1093/biostatistics/kxu001>
- [5] Argelaguet, R., Arnol, D., Bredikhin, D. et al. (2020). MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology*, 21, 111. <https://doi.org/10.1186/s13059-020-02015-1>