



## Introduction

Multi-omics integration data enhances our understanding of biological systems and their underlying mechanisms. This could be achieved by factor based integration methods, including MOFA<sup>+</sup>, FABIA<sup>2</sup>, GFA<sup>3</sup> and MFA<sup>4</sup> that are widely used with great success. Yet the performance of those methods differ and there is a need to assess the level of consistency between those tool in an impartial manner.

## Objective

The aim of this study is to assess four unsupervised factor-based analysis tools; MOFA<sup>+</sup>, FABIA<sup>2</sup>, GFA<sup>3</sup> and MFA<sup>4</sup>, to detect the level of consistency among those tool. For this purpose we have applied both simulated data (with increasing level of complexity) and real biological dataset.

## Methods

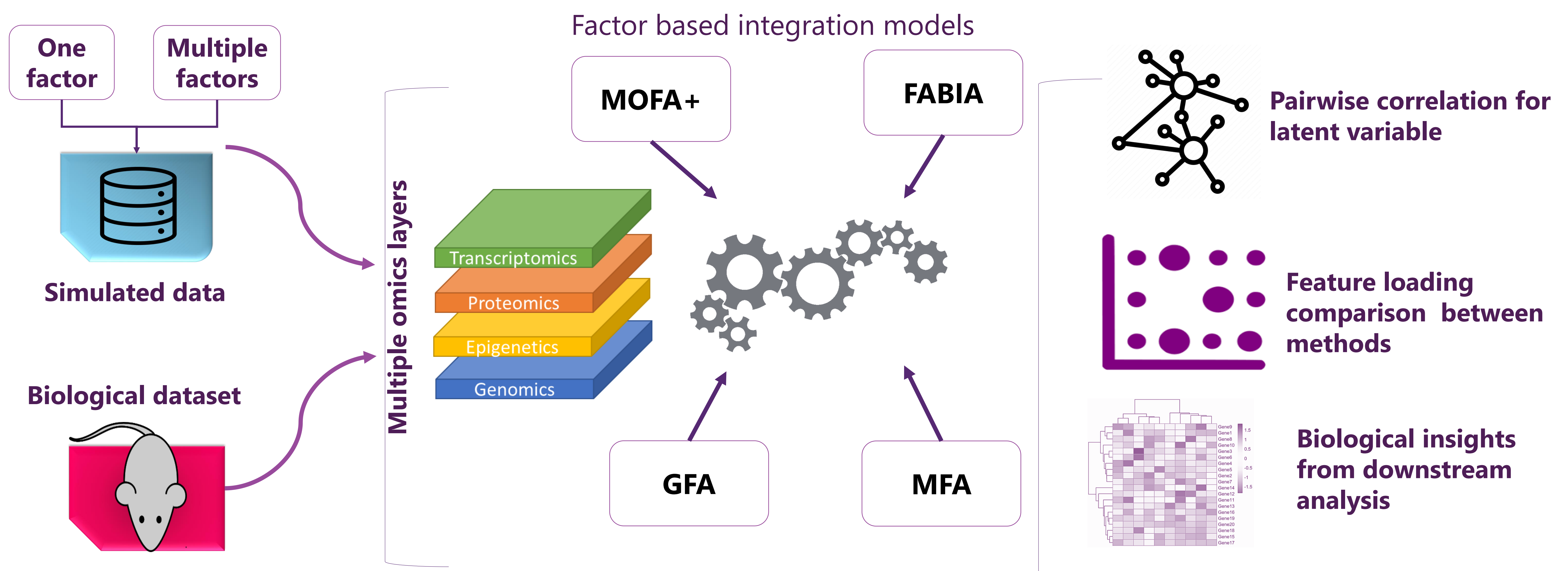


Figure 1 Demonstration of the steps applied to assess the different factor analysis tools

## Results

All methods captured the signal with significant high pairwise correlation between all tools with p-value lower than 0.01 (Figure 2). MOFA factor loadings give different features weight distribution per each omics source (Figure 3), but the rest of methods have the same distribution for feature weights between omics sources.

Figure 3 Correlation plots shows features weights correlation between MOFA and FABIA.

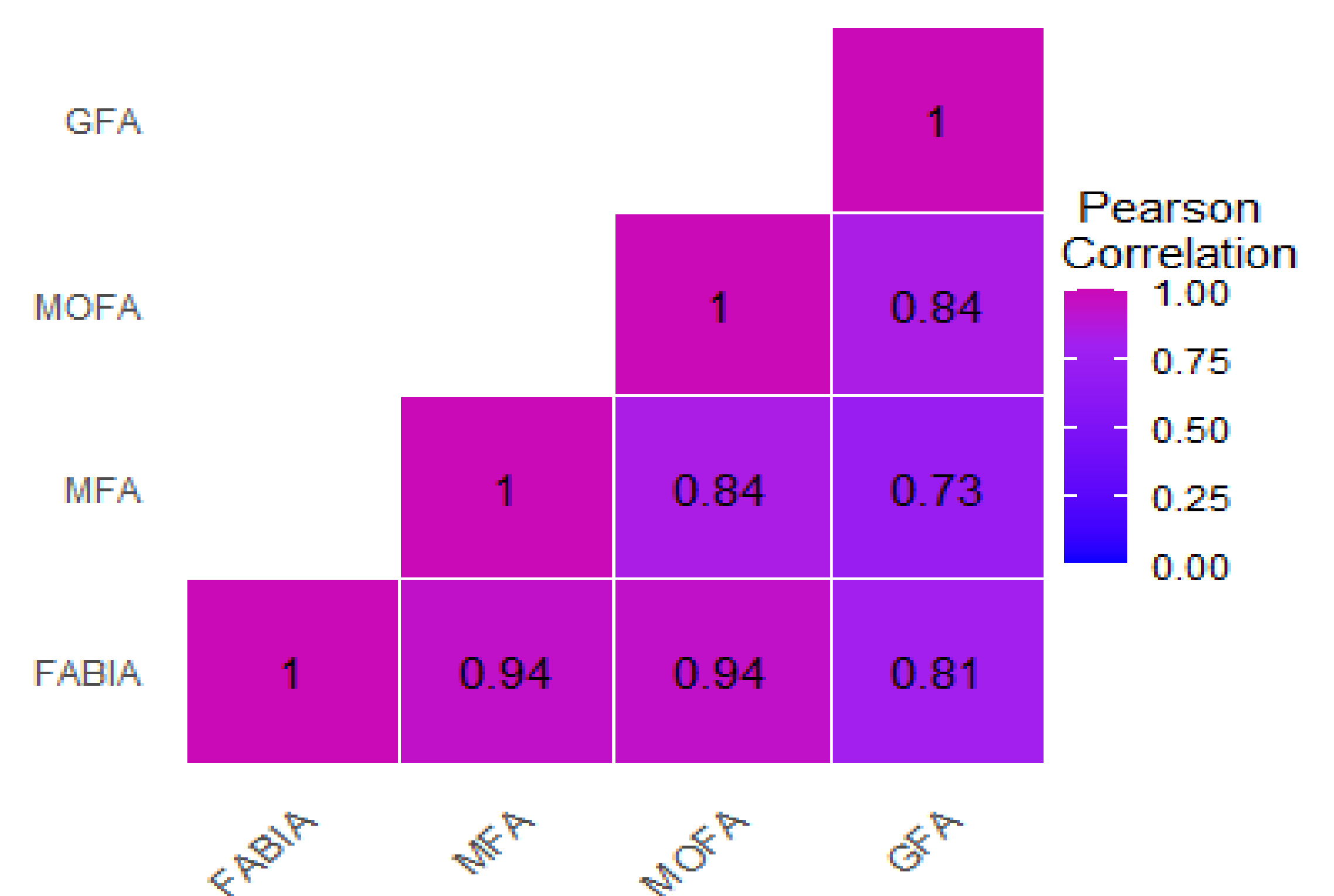
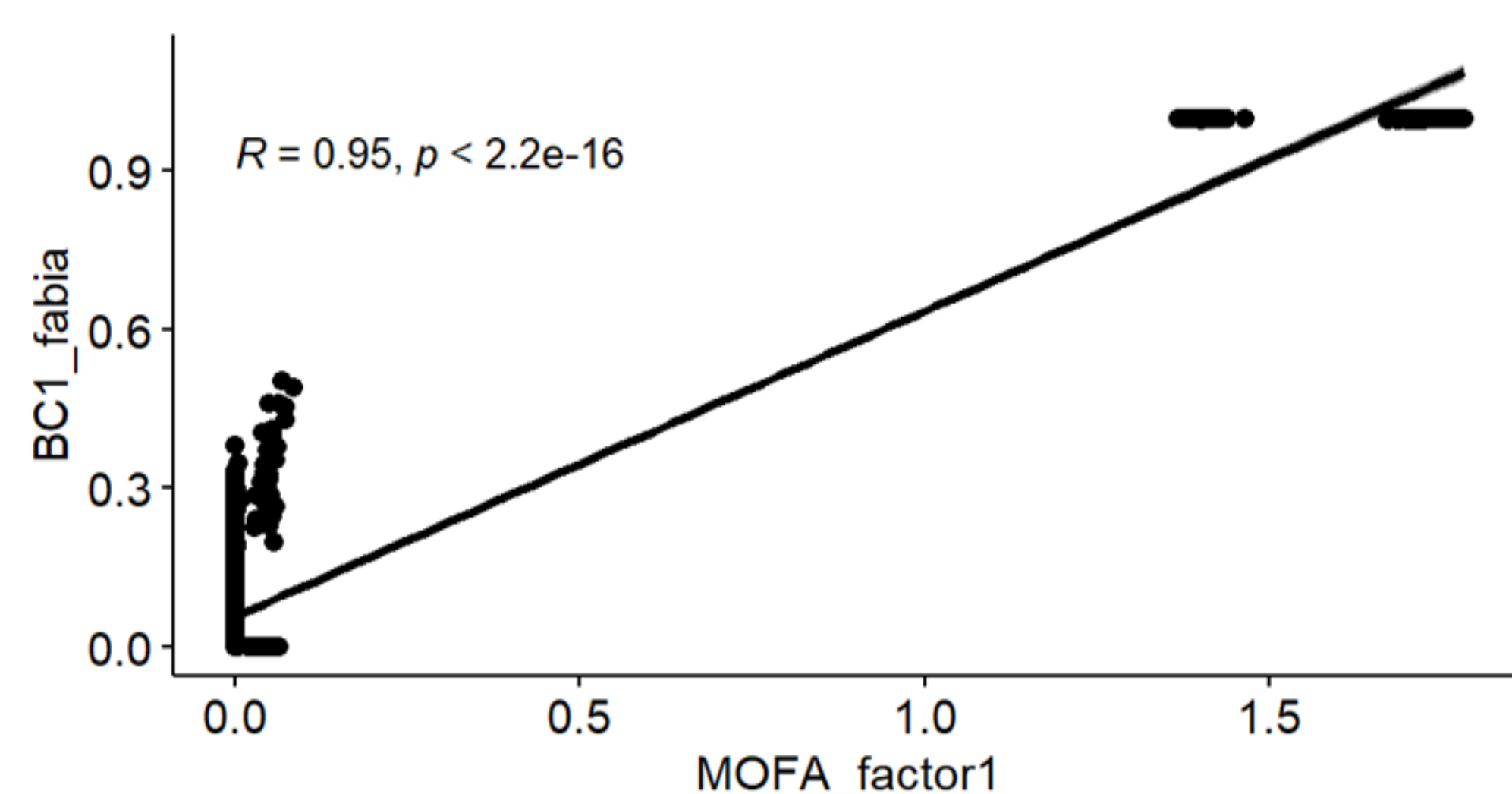


Figure 2 Heatmap represents features weights correlation between all methods on simulated dataset.

## Future work

Factor-based analysis-based methods almost give similar latent variables and features loading, especially when the signal is strong. We are still in preliminary steps, and therefore, will be performing extra analyses to include additional levels of complexity in the simulated data with different level of signals. Moreover, in depth interpretation of the biological data will be done to understand better the identified factors and how they correlate with the biological knowledge.

## References

1. R. Argelaguet *et al.*, (2020) "MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data", *Genome Biology*.
2. S Hochreiter *et al.*, (2010), "FABIA: factor analysis for bicluster acquisition, *Bioinformatics*".
3. E. Leppäaho, *et al.*, (2017), "GFA: Exploratory analysis of multiple data sources with group factor analysis, *journal of machine learning research*".
4. Becue-Bertaut, *et al.*, (2008), "MFA: Multiple factor analysis and clustering of a mixture of quantitative, categorical and frequency data. *Computational Statistics and Data Analysis*".

