

ASTERICS: A Tool for the ExploRation and Integration of omiCS data.

Élise MAIGNÉ^{1,*}, Céline NOIROT^{1,2,*}, Jérôme MARIETTE^{1,2}, Yaa ADU KESEWAAH^{1,3}, Sébastien DÉJEAN^{3,4}, Camille GUILMINEAU^{1,3}, Julien HENRY^{1,3}, Arielle KREBS^{1,2}, Laurence LIAUBET⁵, Fanny MATHEVET^{1,3}, Hyphen-Stat⁶, Christine GASPIN^{1,3}, and Nathalie Vialaneix¹

¹ Université de Toulouse, INRAE, UR MIAT, 31326, Castanet-Tolosan, France

² Université Fédérale de Toulouse, INRAE, BioinfOmicS, GenoToul Bioinformatics facility, 31326, Castanet-Tolosan, France

³ Plateforme Biostatistique, Genotoul, Toulouse, France

⁴ IMT, UMR5219, Université de Toulouse, CNRS, UPS, 31062, Toulouse, France

⁵ GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326, Castanet-Tolosan, France

⁶ Hyphen-stat, <https://hyphen-stat.com/>, Toulouse, France

(*) These authors contributed equally to the work.

Corresponding author: nathalie.vialaneix@inrae.fr

1 Introduction

The rapid development of omics acquisition techniques has induced the production of a large volume of heterogeneous and multi-level omics datasets measured on the same individuals. Complex information of biological interest is obtained from so-called *integration methods*, which have been increasingly developed in the past few years. Some of these methods are already available in R packages (like **mixOmics** [1] or **mixKernel** [2] to which our team has contributed). However, the use of these packages still requires to learn a programming language and to have access to sufficient statistical knowledge to choose method parameters and interpret outputs.

2 ASTERICS

ASTERICS is a web application that aims at making complex exploratory and integration analysis workflows easily available to biologists. Data edition, exploration and integration menus organize the interface to perform 1/ data edition*, missing value imputation, and normalization*, 2/ data exploration with interactive plots, numerical summaries, PCA, tests, clustering, and self-organizing maps, and 3/ data integration with differential analysis*, MFA, or PLS-based methods. Analyses are adapted* to the most standard omics datasets (RNA-seq or count data from sequencing technologies, microarray, metabolomics, metagenomics or other compositional data).

ASTERICS is also designed to make the analysis flow understandable with a navigable workspace that displays uploaded or obtained datasets and performed analyses in a graph. Finally, it also comes with a documentation for beginners* that helps interpret the results, choose proper options or the next analysis to perform.

ASTERICS is based on Rserve, pyRserve, and flask. R package versions are controlled using **renv**. Frontend is developed in Vue.js and uses the CSS framework Bulma.

A first and limited version of ASTERICS is already available online at <http://asterics.miat.inrae.fr/>. This limited version does not include the features highlighted above with the mark “*” at time of writing of this proposal. ASTERICS will also be released as a docker image. The complete production version of ASTERICS is scheduled for September 2022, with intermediate versions, including an increasing number of features, deployed online meanwhile.

Acknowledgements

This work is funded by Région Occitanie (Grant 20008788 – ASTERICS).

References

- [1] Florian Rohart, Benoît Gautier, Amrit Singh, and Kim-Anh Le Cao. **mixomics**: an R package for omics feature selection and multiple data integration. *PLoS Computational Biology*, 13(11):e1005752, 2017.
- [2] Jérôme Mariette and Nathalie Villa-Vialaneix. Unsupervised multiple kernel learning for heterogeneous data integration. *Bioinformatics*, 34(6):1009–1015, 2018.