



HAL
open science

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

Élise Maigné, Céline Noirot, Julien Henry, Yaa Adu Kesewaah, Sébastien D Éjean, Camille Guilmineau, Arielle Krebs, Fanny Mathevet, Entreprise Hyphen-Stat, Christine Gaspin, et al.

► **To cite this version:**

Élise Maigné, Céline Noirot, Julien Henry, Yaa Adu Kesewaah, Sébastien D Éjean, et al.. ASTERICS: A Simple Tool for the ExploRation and Integration of omiCS data. Journée Régionale de Bioinformatique et Biostatistique, Génopole Toulouse, Nov 2023, Toulouse, France. 10.57745/TCKSTD . hal-04746181

HAL Id: hal-04746181

<https://hal.inrae.fr/hal-04746181v1>

Submitted on 21 Oct 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.

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

Élise MAIGNÉ^{1,*†}, Céline NOIROT^{1,2,*}, Julien HENRY^{1,3}, Yaa ADU KESEWAAH^{1,3}, Sébastien DÉJEAN^{3,4},
Camille GUILMINEAU^{1,3}, Arielle KREBS^{1,2}, Fanny MATHEVET^{1,3}, Hyphen-Stat⁵, Christine GASPIN^{1,2},
Laurence LIAUBET⁶, and Nathalie Vialaneix^{1,3}

¹ Université de Toulouse, INRAE, UR MIAT, 31326, Castanet-Tolosan, France
`{prenom.nom}@inrae.fr`

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
`sebastien.dejean@math.univ-toulouse.fr`

⁵ Hyphen-stat, <https://hyphen-stat.com/>, Toulouse, France

⁶ GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326, Castanet-Tolosan, France
`laurence.liaubet@inrae.fr`

`david.colongo@hyphen-stat.com`

* These authors contributed equally to the work. – † Presenting author.

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

2.1 Features

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,
3. data integration with differential analysis, MFA, or PLS-based methods.

Analyses, especially normalization and differential 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 DAG (Directed Acyclic Graph; see Figure 1). In addition, ASTERICS offers options to export results (as CSV or RDATA files) and to generate HTML reports on the performed analyses.

Finally, ASTERICS also comes with

- a documentation for beginners that helps interpret the results, choose proper options or the next analysis to perform https://asterics.pages.mia.inra.fr/user_documentation/;

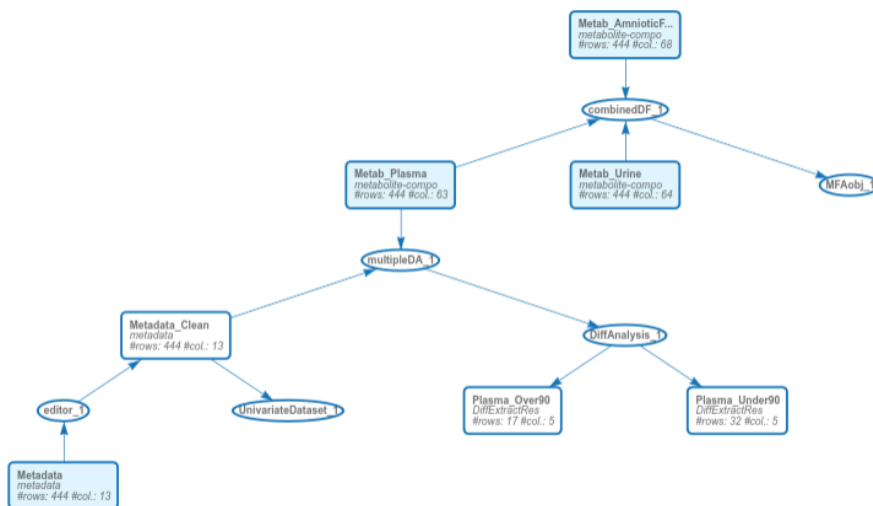


Figure 1. Example of a workspace as displayed in ASTERICS.

- two use cases, including one that is directly available as a ready-to-use workspace https://asterics.miat.inrae.fr/workspace/piglet_usecase and described in our companion article [3]. This use case can also be reproduced using data available at <https://doi.org/10.57745/TCKSTD> (outputs of ASTERICS analysis, including HTML reports generated by ASTERICS, are also available in that data repository).

2.2 Technical information

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.

ASTERICS is available online at <https://asterics.miat.inrae.fr/> and its installation can be performed locally using the three docker images available at https://forgemia.inra.fr/asterics/asterics/container_registry. Installation instructions and source code are provided on our github repository <https://forgemia.inra.fr/asterics/asterics>.

Acknowledgements

This work is funded by Région Occitanie (Grant 20008788 – ASTERICS). The authors are very grateful to Jérôme Mariette for his substantial contribution to the first version of the tool. They are also grateful to Valentine Rossi for her contribution to one of the use cases.

References

- [1] F. Rohart, B. Gautier, A. Singh, and K.-A. Le Cao, “mixomics: an R package for ’omics feature selection and multiple data integration,” *PLoS Computational Biology*, vol. 13, no. 11, p. e1005752, 2017.
- [2] J. Mariette and N. Villa-Vialaneix, “Unsupervised multiple kernel learning for heterogeneous data integration,” *Bioinformatics*, vol. 34, no. 6, pp. 1009–1015, 2018.
- [3] E. Maigné, C. Noirot, J. Henry, Y. Adu Kesewaah, L. Badin, S. Déjean, C. Guilmineau, A. Krebs, F. Mathévet, A. Segalini, L. Thomassin, D. Colongo, C. Gaspin, L. Liaubet, and N. Vialaneix, “ASTERICS: A Simple Tool for the ExploRation and Integration of omiCS data,” *BMC Bioinformatics*, 2023. Forthcoming.