Porting the rCASC workflow for scRNA-Seq data analysis to Galaxy and the Laniakea Galaxy on-demand system

Pietro Mandreoli[1], Luca Alessandrì[2], Marco Antonio Tangaro[3], Raffaele Calogero[2], Federico Zambelli[4]

[1] Dept. of Biosciences, University of Milano - Italy, Email: pietro.mandreoli@unimi.it

[2] Dept. of Molecular Biotechnology and Health Sciences, University of Torino - Italy.

[3] Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, CNR - Italy.
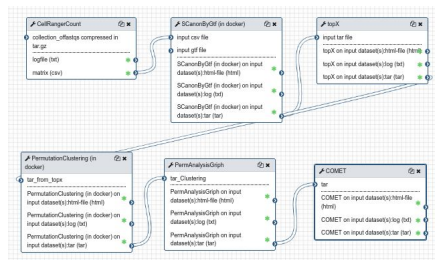
[4] Dept. of Biosciences, University of Milano - Italy.

rCASC[1] is a workflow for scRNA-Seq data analysis providing an integrated analysis environment that exploits Docker containerization to achieve both functional and computational reproducibility

of the data analysis process. rCASC modular architecture consists of 39 Docker images, each one tailored to perform a specific function, e.g., quality control, clustering, and feature selection. While this Docker-based implementation ensures a reliable framework for long-term reproducibility, rCASC is currently available only as a stand-alone software with a custom GUI, or as a command-line tool. To improve its availability and accessibility, a porting of rCASC to Galaxy is in progress to provide end-users with the possibility to automatically download, deploy, and run rCASC within the familiar Galaxy environment over the cloud. This operation is non-trivial due to the internal architecture of rCASC, composed of highly interconnected, Galaxy unfriendly, containerized modules. The porting requires, for example, several tweaks to rCASC's input/output functions, as well as specific configurations for Galaxy and the Docker engine. We have therefore identified a set of rules for rCASC/Galaxy integration that could be easily expanded and applied to the more general task of porting or designing containerized tools for cloud-based Galaxy instances. Following these integration rules, the first rCASC workflow (Fig:1) has been successfully ported to Galaxy. We are now working on the migration of the others. The final aim is to make rCASC available also as a Galaxy flavour into the Laniakea [3] Galaxy on-demand system, providing a medium to deploy a Galaxy instance pre-loaded with rCASC tools without effort. Ultimately, this will result in a more user-friendly instrument for reproducibility-oriented scRNA-Seq data analysis, that can also be seamlessly supported by the cloud resources offered by any Laniakea-based service.

References:

[1] https://doi.org/10.1093/gigascience/giz105

[2] https://doi.org/10.1093/nar/gky379

[3] https://doi.org/10.1093/gigascience/giaa033