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

<u>Pietro Mandreoli¹</u>, Luca Alessandrì², Marco Antonio Tangaro³, Raffaele Calogero², Federico Zambelli⁴

¹ Dept. of Biosciences, University of Milano - Italy, Email: <u>pietro.mandreoli@unimi.it</u>

- ² Dept. of Molecular Biotechnology and Health Sciences, University of Torino Italy.
- ³ Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, CNR Italy.

⁴Dept. of Biosciences, University of Milano - Italy.

Project Website: <u>https://laniakea-elixir-it.github.io/</u> https://kendomaniac.github.io/rCASC/index.html Source Code: <u>https://github.com/pmandreoli/rCASC wrappers</u> License: <u>GNU GPLv3</u>

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

r CellRangerCount @ x	🗲 SCanonByGtf (in docker) 🛛 🖄 🗙	iopX ⊀	(2 ×
o collection_offaatga.compressed in tar.gz logfile (tixt) = o mattix (cov) = o	input cav file input gif file	topX on input dataset(s	(chimi-file (html)
	SCanorByOlf (in docker) on input dataset(s) html-file (html) SCanorByOlf (in docker) on input dataset(s) log (sd) SCanorByOlf (in docker) on input	topX on input dataset(is topX on input dataset(is	
	dataset(s):tar (tar)		
PermutationClustering (in 😢 🗙	🖌 PermAnalysisGriph 🖉 🗶	ACOMET	(Par)
	PermAnalysisGriph ②x o tar_Clustering	F COVET	Ø×
scker) ar_from_tops Permutation/Clustering (in docker) on	tar_Clustering PermAnalysisGriph on input detassel(s):htmi-file (html) *	F CONET far CONET on input datase (html)	
ocker)	ter_Clustering PermAnalysisGriph on input	COMET on input datase	• • • • • • • • • • • • • • • • • • •

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