

dene expression







de.STAIR

Galaxy-based modular workflow generator for guided data analysis

Andrea Bagnacani¹, Konstantin Riege², Steffen C. Lott³, Markus Wolfien¹, Olaf Wolkenhauer¹, Wolfgang R. Hess³, Steve Hoffmann²

FKZ 031L0106

- ¹ Department of Systems Biology & Bioinformatics, University of Rostock, Rostock, Germany
- ² Computational Biology, Leibniz Institute on Aging, Jena, Germany
- ³ Genetics and Experimental Bioinformatics, Faculty of Biology, University of Freiburg, Freiburg, Germany

The project

Shared Galaxy workflows promote the dissemination of best-practice approaches for computational Life Science analyses. Workflows are explained in their sequence of pre-selected tools, and illustrated through manually curated interactive tours, to support the adoption of well established formats, as well as provide a mean for the self-training of a new generation of data analysts. However, tool pre-selection misses to address the use of alternative computational approaches. Our proposed design overcomes this limitation, by defining alternative best-practice approaches for completing established workflow analyses, and lowering the curatorial effort needed to maintain the corresponding set of alternative interactive tours.

Progress

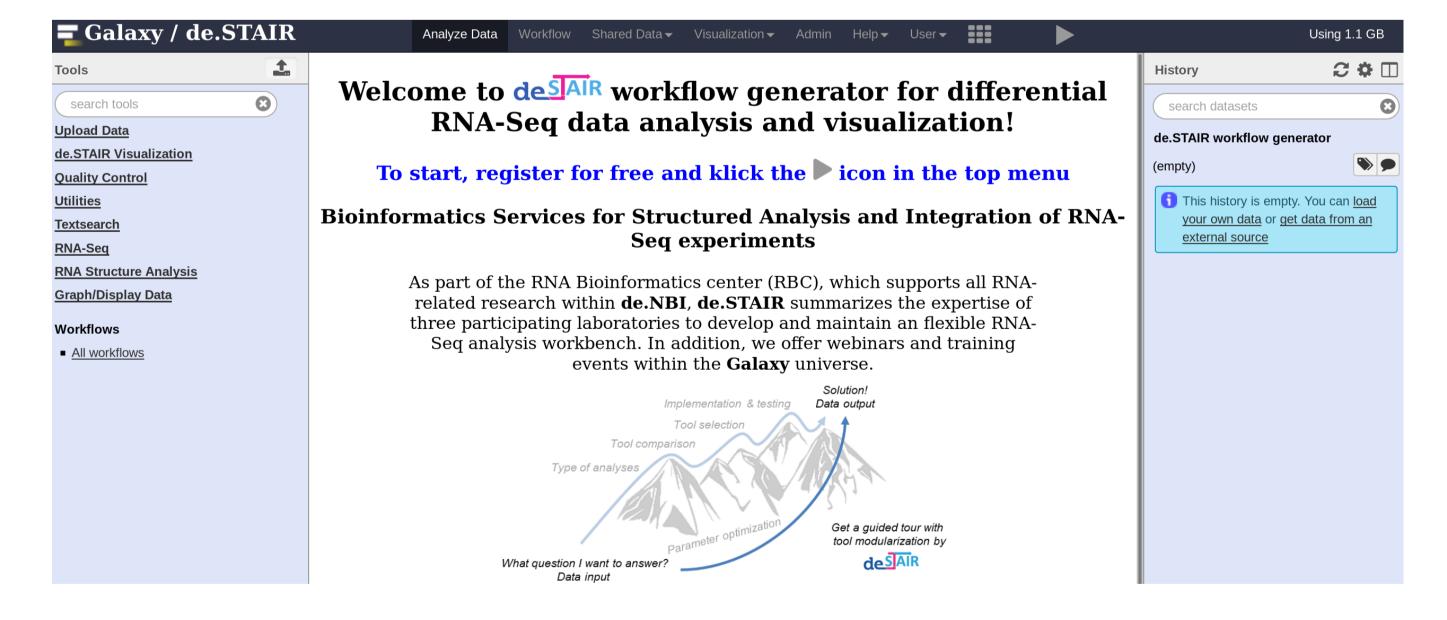
Workflow design

- of differential Organization of established data analyses into modules
- Collection of modules as interchangeable blocks
- Curation of per-module interactive tours
 - Lower curatorial effort:

	Identified modules	Approaches per module	Design	Interactive tours to maintain
	3	4, 4, 2	monolitic	$\prod_{i=1}^{modules} approach_i = 32$
			modular	$\sum_{i=1}^{modules} approach_i = 10$

Reproducible workflows

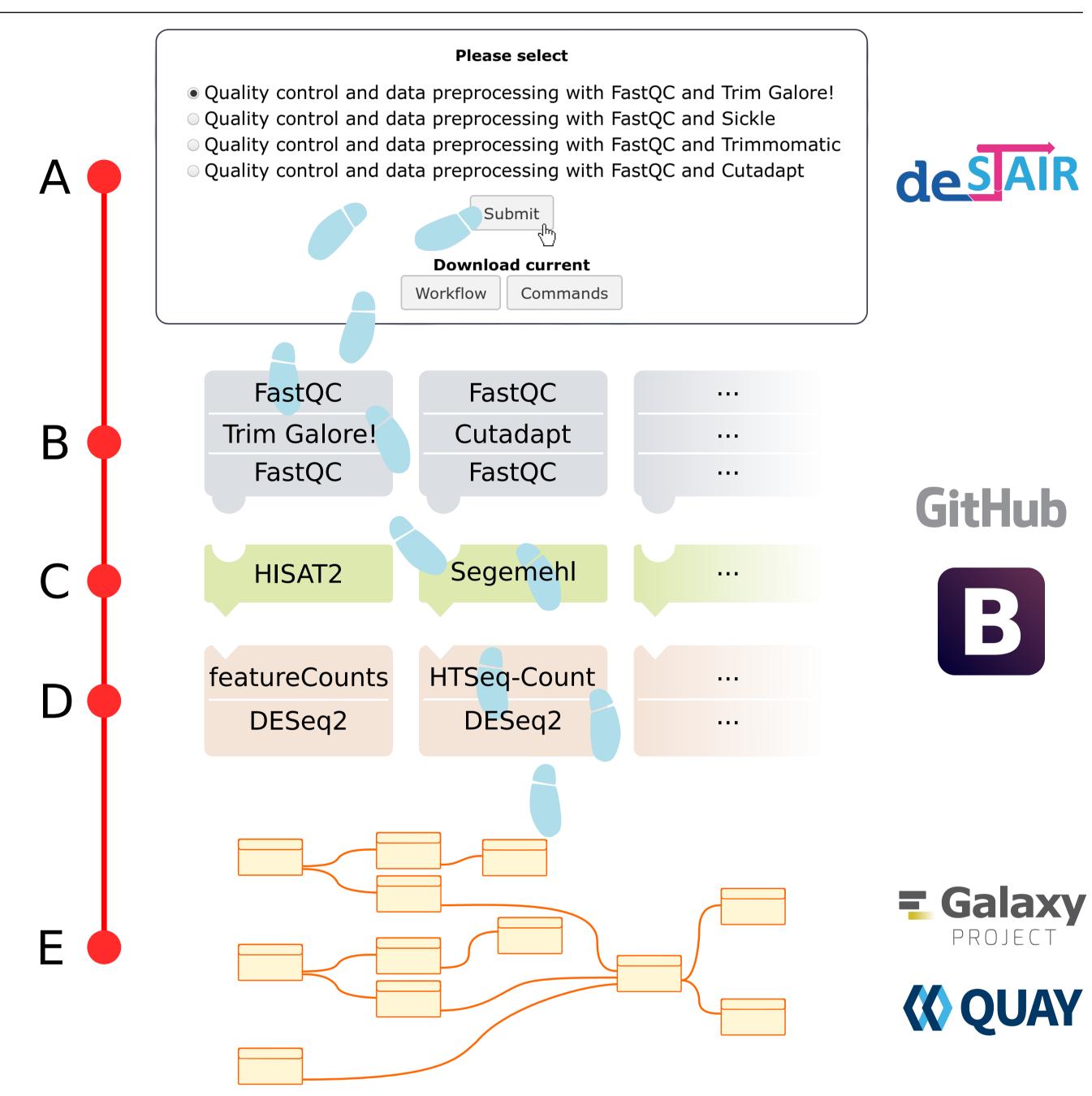
- Development of a system to compose alternative modules
- Development of a Galaxy webhook to trigger each module
- Dockerization
 - Alternative best practice workflows & Self training:



About the project

	Scientist	Andrea Bagnacani
Funded by de.NBI	Scientist	Konstantin Riege
	Scientist	Steffen C. Lott
	Scientist	Markus Wolfien
	Scientist (PI)	Prof. Olaf Wolkenhauer
Other staff involved	Scientist (PI)	Prof. Wolfgang R. Hess
	Scientist (PI)	Prof. Steve Hoffmann

Services



The system asks which approach to use (A) for the first analysis module. Upon completion, the user selects the preferred set of tools (B, C, D) to complete the analysis. The chosen route (blue trail) composes the final workflow (E), which can be exported and shared for downstream analysis.

Training and education

Support to students	Dummerstorf 29.09.2017 - 29.09.2017
A primer for RNA-Seq processing, interpreting and visualization	Freiburg 04.10.2017 - 06.10.2017
Support to students	Rostock 10.10.2017 - 12.10.2017
Introduction to RNA-Seq data analysis with Galaxy	Kiel 07.03.2018 - 07.03.2018
Support to students	Rostock 21.05.2018 - 25.05.2018
A primer for RNA-Seq processing, interpreting and visualization	Jena 27.06.2018 - 29.06.2018
RNA-Seq data analysis with Galaxy for clinical applications (GMDS 2018)	Osnabrück 04.09.2018 - 04.09.2018

Publications

Lott S. C., Wolfien M., Riege K., Bagnacani A., et al. (2017).

Customized workflow development and data modularization concepts for rna-sequencing and metatranscriptome experiments. Journal of biotechnology, 261, 85–96

Afgan E., Baker D., Batut B., Van Den Beek M., et al. (2018).

The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic acids research, 46 (W1), W537-W544.

Batut B., Hiltemann S., Bagnacani A., Baker D., et al. (2018).

Community-driven data analysis training for biology. Cell Systems, 6 (6), 752 - 758.e1.





