

## Towards automating workflow analyses in Galaxy

*Andrea Bagnacani<sup>1</sup>, Markus Wolfien<sup>1</sup>, Martin Scharm<sup>1</sup>, Olaf Wolkenhauer<sup>1</sup>*

<sup>1</sup>Systems Biology and Bioinformatics, University of Rostock, Ulmenstr. 69, 18051 Rostock

### Introduction

The Galaxy community is promoting RNA-Seq protocols and best practices through the reuse of existing tools, and the consolidation of a Training Network to provide guidance to researchers through example datasets, tutorials, and interactive tours. However, the more tools and techniques are showcased, the more complex the options for tool chaining and parametrization become.

### Objectives

- Assist researchers in carrying out their analyses
- Integrate the Galaxy framework with an interactive recommendation system leveraging on community consolidated best practices as well as EDAM annotated tools
- Promote the adoption of well established pipelines
- Allow room for experimental tools
- Consolidate protocols and reproducibility

### Materials and Methods

Tool chaining and parametrization through Galaxy's interactive tours



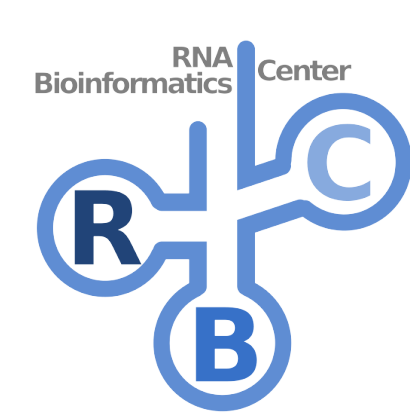
[galaxyproject.org](http://galaxyproject.org)

Tool operations and input / output formats through Elixir's bio.tools

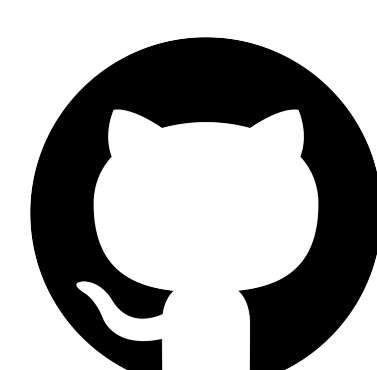


[edamontology.org](http://edamontology.org)

Best practices and user-tracked data of RBC's Galaxy instance



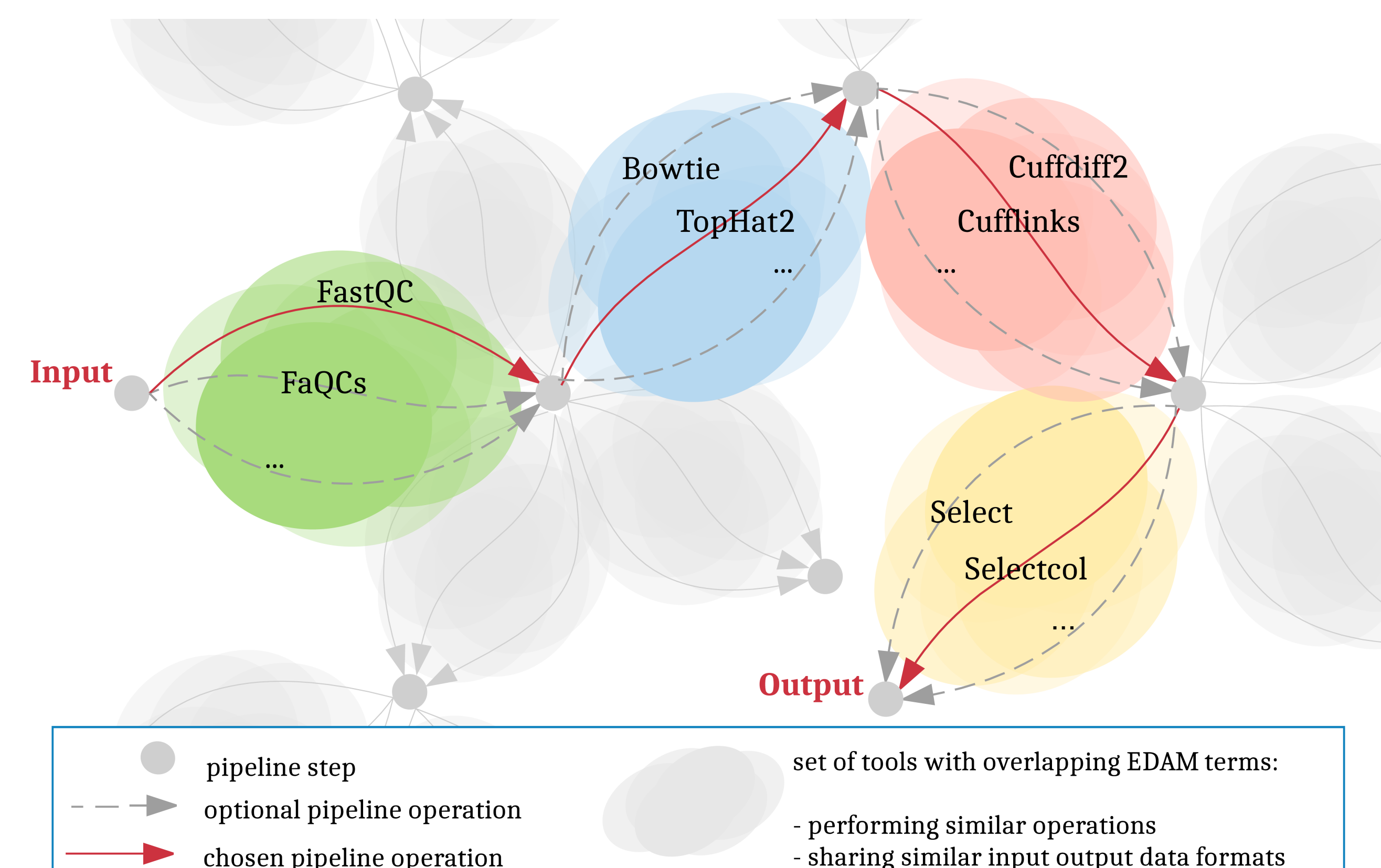
[denbi.de/rbc](http://denbi.de/rbc)



destairdenbi  
[destair.bioinf.uni-leipzig.de](http://destair.bioinf.uni-leipzig.de)



### Results and discussion



- Galaxy tools are grouped by function
- Each tool function bridges two different states of data
- Tools are chained on their input / output data formats
- Galaxy tours recommends pertinent tools step by step
- Users decide which tool to select and parametrize

### Conclusion

A recommendation system enhances the visibility of each Galaxy tool, relieving the user from browsing tool categories, or sticking to the usual analysis tools

Bioinformatics

Name (HTML)	Database search	Database search results (HTML)
Transcription factors and regulatory sites > Functional, regulatory, and non-coding RNA > RNA		
RNA family identifier (Textual format)	Sequence assembly visualisation	Concentration (PNG)
Gene name (Textual format)	Data retrieval	Sequence composition plot (PNG, Dot-bracket format, SVG)
Pathway or network name (Textual format)	Filtering	Database search results (CSV, JSON, HTML)
Organism name (Textual format)	Aggregation	Pathway or network (PNG)
	Pathway or network visualisation	

Tool pertinence is inferred from manually curated EDAM annotations, therefore a tool's pertinence is as accurate as its bio.tools annotation

### References

Lott SC et al. Customized workflow development and data modularization concepts for RNA-Sequencing and metatranscriptome experiments. Journal of Biotechnology, 2017. [10.1016/j.jbiotec.2017.06.1203](https://doi.org/10.1016/j.jbiotec.2017.06.1203)



Systems Biology and Bioinformatics  
University of Rostock  
Ulmenstr. 69, 18051 Rostock  
[www.sbi.uni-rostock.de](http://www.sbi.uni-rostock.de)



#forschungscamp2017 #galaxy  
#elixir #RNA #EDAM #AI #denbi