

Structured Analysis and Integration of RNA-Seq experiments (de.STAIR)

Fkz 031L0106C

Andrea Bagnacani, Markus Wolfien, Olaf Wolkenhauer

Dept. of Systems Biology & Bioinformatics, University Rostock

Short description of the project

The Galaxy platform is being adopted by a growing body of Life Science communities. However, with its increasing number of tools, it becomes more difficult for users to make informed decisions about the selection of tools and implication of their parameterizations. For this reason, the Galaxy Training Network is organizing an extensive index of best-practice data analysis approaches. This material guides users through the construction of workflows for different types of analyses and experimental setups. However, the absence of a **systematic overview of the alternative tools** of a Galaxy instance does not train users on how to examine alternative best-practice algorithms and parameterizations.

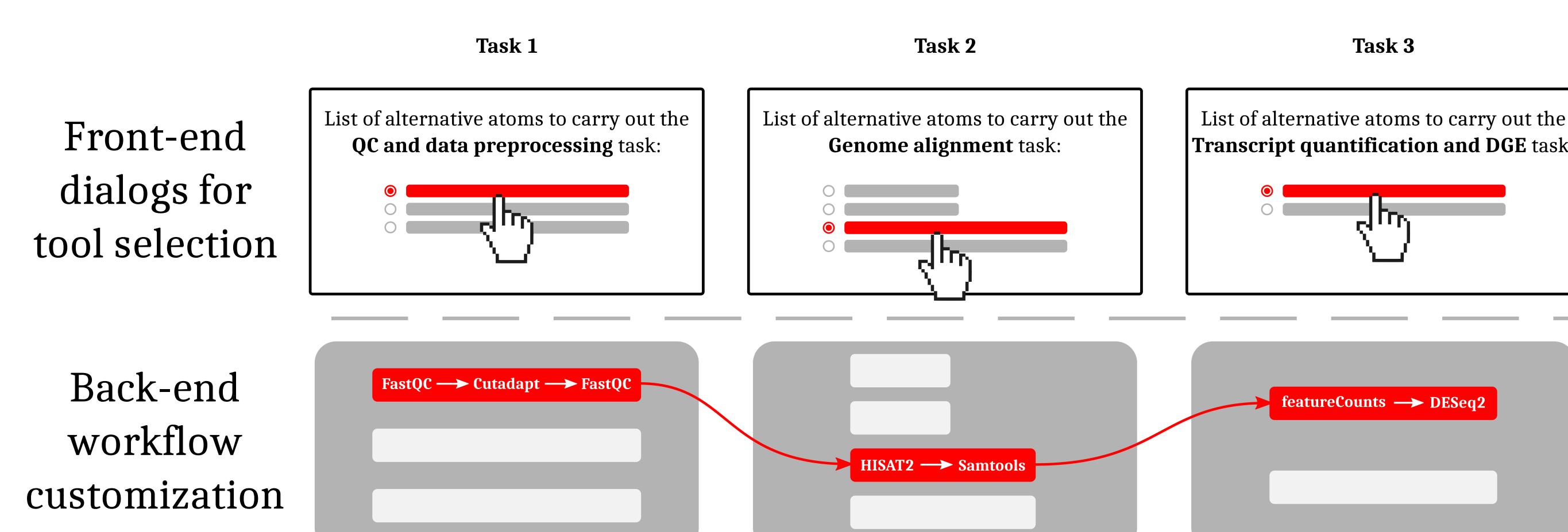
We address this problem by dividing a workflow into its tasks, and provide, for each of them, alternative implementations in the form of interactive Galaxy tours. We call these entities **atoms**. At the same time, we couple each atom with a help section, to **assist users in the selection of the most suitable tool and parameterization**. Finally, we reorganize the training material as a series of user-selectable strategies for the customization of their workflows (See Fkz 031L0106D).

Progress report

Interactive comparison of the available alternative atoms

Task	1			2		3	
Operation	Quality control	Data preprocessing	Quality re-check (optional)	Genome alignment	Output sorting (optional)	Transcript quantification	Differential gene expression
Atoms	FastQC	Cutadapt ^x Cutadapt ^y PRINSEQ Trimmomatic Trim Galore! ...		BWA HISAT2 ^z HISAT2 ^w segemehl STAR ...	Samtools	featureCounts HTSeq-counts	DESeq2
Sample atoms to complete a task in the analysis		FastQC → Cutadapt ^x FastQC → Cutadapt ^y FastQC → Cutadapt → FastQC FastQC → Trimmomatic ...		HISAT2 ^z HISAT2 ^w HISAT2 → Samtools STAR ...		featureCounts → DESeq2 HTSeq-counts → DESeq2	

Workflow customization



General information on the project

Funded by de.NBI staff

Andrea Bagnacani (Scientist)

Additional staff involved

Prof. Olaf Wolkenhauer (PI)

Markus Wolfien (Scientist)

Dr. Martin Scharm (IT administrator)

de.NBI services

Select the preferred atom

Quality control and data preprocessing

- FastQC → Cutadapt → FastQC
- FastQC → PRINSEQ → FastQC
- FastQC → Trimmomatic → FastQC
- FastQC → Trim Galore! → FastQC

Select the preferred atom

Quality control and data preprocessing

- FastQC → Cutadapt → FastQC
- FastQC → PRINSEQ → FastQC
- FastQC → Trimmomatic → FastQC
- FastQC → Trim Galore! → FastQC

Galaxy / de.STAIR Workflow generator

FastQC and Cutadapt

de.STAIR Tool rating

Ease of use:	● ● ● ○ ○
User base:	● ● ● ○ ○
Sensitivity:	● ● ● ○ ○
Speed:	● ● ● ○ ○
Configurability:	● ● ● ○ ○

Description

FastQC is one of the preferred tools for quality control. This tool analyses FASTQC data, and produces graphs to allow fast inspection of sequencing data sets.

Cutadapt is one of the first tools to perform trimming and filtering of NGS data. This tool has been widely adopted, and has also been used as a baseline for other data preprocessing tools, such as Sickle and Trim Galore!.

References

FastQC
Cutadapt

Galaxy Atoms & Workflow generator plugin

github.com/destairdenbi

10.1186/s12859-015-0873-9

[10.1186/s12859-015-0873-9](https://doi.org/10.1186/s12859-015-0873-9)

TRAPLINE

TriplexRNA

de.NBI Training and education

Introduction to RNA-Seq data analysis with Galaxy	EASyM Berlin 26.10.2016 - 26.10.2016
Introduction to RNA-Seq data analysis with Galaxy	CASyM Ljubljana 29.03.2017 - 29.03.2017
RNA-Seq data analysis with Galaxy for clinical applications	GMDS Oldenburg 18.09.2017 - 18.09.2017
A primer for RNA-Seq processing, interpretation and visualization	Freiburg 04.10.2017 - 06.10.2017
Introduction to RNA-Seq data analysis with Galaxy	Kiel 07.03.2018 - 07.03.2018
A primer for RNA-Seq processing, interpretation and visualization	Jena 27.06.2018 - 29.06.2018
RNA-Seq data analysis with Galaxy for clinical applications	GMDS Osnabrück 04.09.2018 - 04.09.2018
Galaxy for linking Bisulfite sequencing with rna sequencing	Rostock 06.03.2019 - 08.03.2019
RNA-Seq data analysis with Galaxy for clinical applications	GMDS Dortmund 08.09.2019 - 08.09.2019
Galaxy for linking Bisulfite sequencing with rna sequencing	Freiburg 09.10.2019 - 11.10.2019

Publications

- Bagnacani et al. (2019). Tools for understanding miRNA-mRNA interactions for reproducible data analysis. Springer, 199-214
- Wolfien et al. (2019). Workflow Development for the Functional Characterization of ncRNAs. Springer, 111-132
- Fallmann et al. (2019). The RNA workbench 2.0: next generation RNA data analysis. Nucleic Acids Research, 47, W511-W515
- Batut et al. (2018). Community-driven data analysis training for biology. Cell Systems, 6(6), 752-758
- Lott et al. (2017). Customized workflow development and data modularization concepts for rna-sequencing and metatranscriptome experiments. Journal of Biotechnology, 261, 85-96