

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

deSTAIR

Fkz 031L0106D

Konstantin Riege and Steve Hoffmann  
Computational Biology, Leibniz Institute on Aging - Fritz Lipmann Institute (FLI), Jena, Germany

## Short description of the project

**Galaxy** is an established platform for biomedical research with a continuously increasing popularity within Life Science communities. The Galaxy project offers share places for workflows and training resources, that implement tools of a large and steadily growing toolshed. However, the training material and workflows **lacking the ability to give a systematic overview** of alternative tools and implications of their parameterizations to address data analysis tasks.

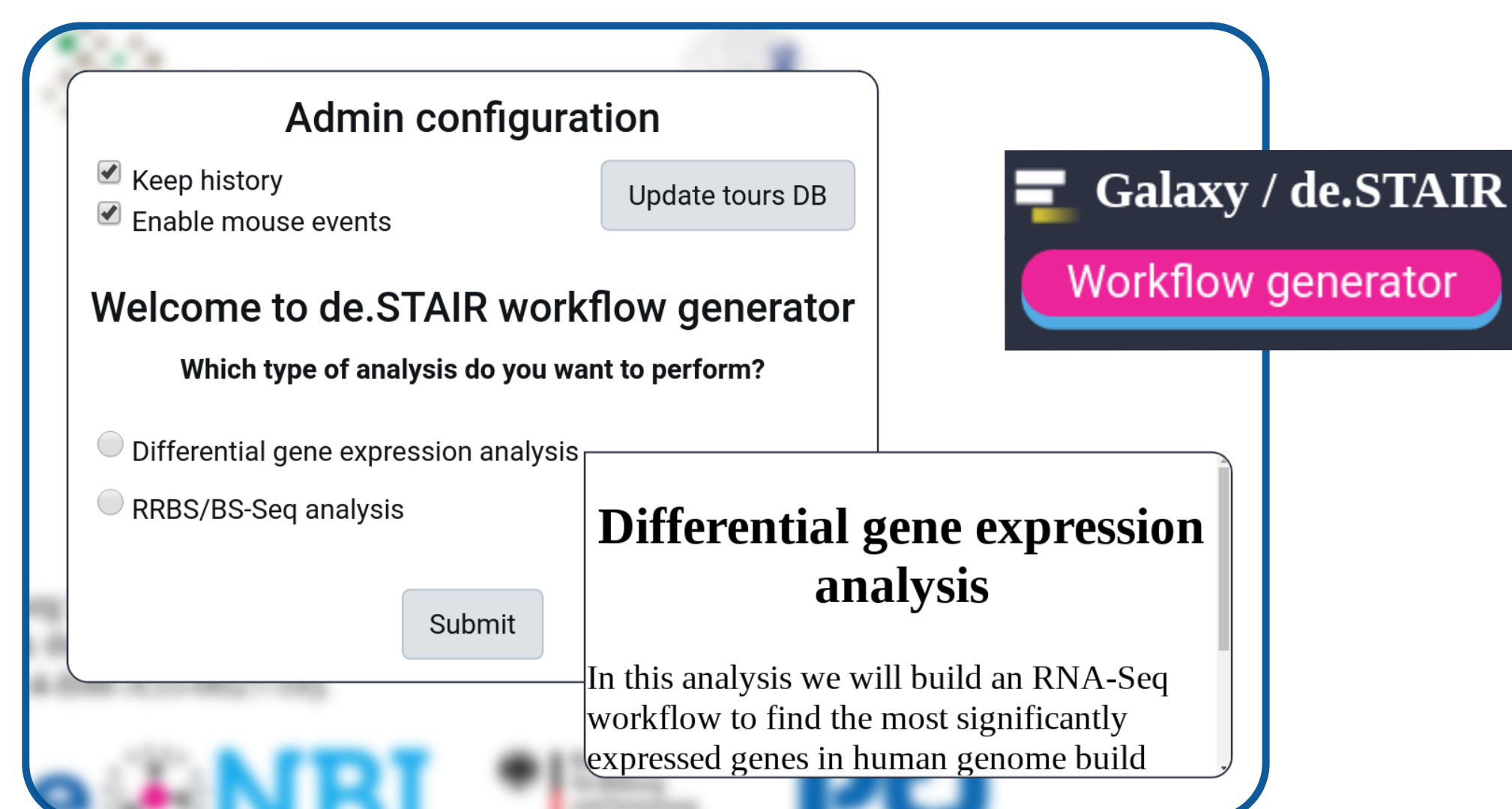
In order to enable users to make informed decisions about the selection and best practice of available software, we developed the **Galaxy workflow generator plugin**. This application leverages on a webhook and utilizes the novel concept of Galaxy **atoms**, *i.e.* new interactive tours that illustrate precise algorithms, which parameterization are tailored for specific experimental setups (see Fkz 031L0106A).

Alternative atoms are presented as multiple choice options through a **novel interaction layer**. Upon user selection, the plugin interprets extended and **revised interactive tours** and allows **fully automated playback**. Furthermore, the plugin offers one-click export of functional workflows, compatible with *usegalaxy.eu*, commands and tool references. Our workflow generator enables **self-training**, can advice and gives expertise. We compiled a new Galaxy flavor and deployed our system as a **web service** and Docker image.

## de.NBI services

de.STAIR Galaxy flavor  
powered by Docker

<https://destair.leibniz-fli.de>  
[quay.io/destair](https://quay.io/destair)



Galaxy atoms & Workflow generator plugin [github.com/destairdenbi](https://github.com/destairdenbi)  
segemehl [10.1371/journal.pcbi.1000502](https://doi.org/10.1371/journal.pcbi.1000502)  
DIEGO [10.1093/bioinformatics/btx690](https://doi.org/10.1093/bioinformatics/btx690)  
Metilene [10.1101/gr.196394.115](https://doi.org/10.1101/gr.196394.115)

## de.NBI Training and education

A primer for RNA-Seq processing, interpretation and visualization	Freiburg	04.10.2017 - 06.10.2017
A primer for RNA-Seq processing, interpretation and visualization	Jena	27.06.2018 - 29.06.2018
Galaxy for linking Bisulfite sequencing with RNA sequencing	Rostock	06.03.2019 - 08.03.2019
Introduction to RNA-Seq data analysis with Galaxy	Jena	04.04.2019 - 04.04.2019
Galaxy for linking Bisulfite sequencing with RNA sequencing	Freiburg	09.10.2019 - 11.10.2019
Introduction phylogenetic tree inference with Galaxy	Jena	14.01.2020 - 15.01.2020

## Progress report

RNA-Seq analysis	RRBS/BS-Seq analysis
Data upload	
Upload control and treatment datasets, genome reference and features	
Data set renaming	Creation of data set collection
Quality control and data preprocessing	
Quality inspection	Quality inspection, adapter clipping
Read quality trimming and artificial rRNA depletion	
Genome alignment	
Mapping and extraction of uniquely and properly paired alignments	
	Artificial depletion of PCR duplicates
Differential gene expression	Inference of differential methylation
Inference of sequencing strandness	Calculation of CpG methylation rates
Read quantification and test for DGE	Take feature mean methylation level
Report of log2-fold-changes	

**Select the preferred atom**  
Quality control and data preprocessing

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

Submit

**Select the preferred atom**  
Quality control and data preprocessing

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

Submit

**FastQC and Cutadapt**

deSTAIR 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

## General information on the project

Funded by de.NBI

Scientist

Dr. Konstantin Riege

Additional staff involved

Scientist (PI)

Prof. Steve Hoffmann

## Publications

- Bagnacani et al. (2019). Assisted generation of alternative workflows using Galaxy atoms. Submitted
- Riege et al. (2019). Dissecting the DNA binding landscape and gene regulatory network of p63 and p53. Submitted
- Doose et al. (2018). DIEGO: detection of differential alternative splicing using Aitchison's geometry. *Bioinformatics*, 34(6):1066-1068
- Grüning et al. (2017). The RNA workbench: best practices for RNA and high-throughput sequencing bioinformatics in Galaxy. *Nucleic Acids Res*, 45(W1):W560-W566
- Lott et al. (2017). Customized workflow development and data modularization concepts for RNA-sequencing and metatranscriptome experiments. *J. Biotechnol*, 261:85-96

