Computational Workflows Course 2025 - Paper

1 Course Project

1.1 Develop a Workflow

For your final project, you will work in teams of two to develop an nf-core workflow for the analysis of RNA sequencing (RNA-seq) data. Each team should begin by creating a dedicated GitHub repository for the project and granting access to the instructors. You are also required to complete the PROJECT.md file in your respective assignment forks on GitHub to provide a concise description of your project objectives and setup.

The development of the workflow should begin with the creation of a template using the nf-core tool installed earlier in the course. Building upon this foundation, you will implement the necessary processes or incorporate existing nf-core modules to perform a comprehensive analysis of the provided RNA-seq dataset. All components should be integrated into a single, well-structured workflow file that adheres to nf-core best practices.

Finally, you are expected to define and produce meaningful results that demonstrate the functionality and utility of your workflow on the data from week 1. Consider what types of outputs, metrics or maybe even visualizations best capture the key findings of your analysis.

Grading Criteria:

1. Functionality

2. Clear and readable code

3. Reproducibility

4. Modularity

5. Usage of proper Nextflow Operators

6. Documentation on Github (Inputs / Ouputs, Parameters, brief description of pipeline steps)

1.2 Write a Report

To obtain credit points for this course, you are required to write a short scientific paper focusing on the topics covered during the two course weeks and on the nf-core workflow you developed. Your paper should demonstrate an understanding of key concepts, including reproducibility, nf-core, and RNA sequencing, and should clearly relate these to your own workflow implementation.

The main body of the paper should be approximately five pages in length, excluding figures, tables, and references. You are expected to use a consistent scientific citation style throughout the document and to reference all sources appropriately. Proper formatting, clear structure, and academic rigor will be considered in the evaluation.

To guide your writing process, you may structure your paper as follows:

1. Introduction
   (a) Pipelines
   (b) FAIR principles and reproducibility
   (c) nf-core
   (d) RNA-seq

2. Methods
   (a) How is an nf-core pipeline structured, explain the most important files
   (b) where / how did you compute, which packages etc did you use (be FAIR when describing)
   (c) ...

3. Results
   (a) describe all aspects of your pipeline
   (b) Bonus: describe aspects like performance with respect to parallelization (potentially compare for a couple of files to running plain bash... (figures!))

4. Discussion
   (a) How does nf-core achieve reproducibility and FAIR pipelines with focus on what your pipeline would still need to be able to be FAIR, reprod.
   (b) what is missing to your pipeline to be a strong RNA-seq workflow. what would be the next steps - Outlook

2 Grading

Again, most important for getting credit, you have to add us to both your assignment fork as well as the project repository on Github. You also need to tell us about your team and the project Github url in the `PROJECT.md` file in your assignment fork.

We will be grading the assignments you worked, the pipelines you developed, as well as the final report you wrote.

2.1 How to hand-in your results

Simply send **both of us** (julian.flesch@uni-tuebingen.de and mark.polster@uni-tuebingen.de) an e-mail titled "**Computational workflow results**" containing your name and the link to your fork of our assignment repository on Github.

You will have to hand in everything latest at the *24th of October, 6pm*