

Practical Course Computational Workflows - 2025 -

Mark Polster and Julian Flesch

Introduction to Computational Workflows

Prof. Sven Nahnsen

Course Information

Attendance

Attendance is mandatory!


We start at 09:15 every morning (lunch-break between 11:45 and 12:45)

There might be interfering meetings, that require our room.
We will adapt our breaks to match that, where possible

Friday 03.10. is public holiday: No course day

Structure of the course

Week 1: Analyzing data using nf-core pipelines / Intro to Nextflow



Week 2: Pipeline project



Contact

In person during the course

via Mail:

julian.flesch@uni-tuebingen.de
mark.polster@uni-tuebingen.de

always to both of us, please!

Missing a course date

If you are sick, inform us **before the course starts in the morning**

A doctor's note **must** always be handed in **via mail to Prof. Sven Nahnsen**
sven.nahnsen@uni-tuebingen.de

Unexcused absence will lead to failing the course!

Grading

Attendance is mandatory to get credits

All hand-ins are done via Github as a **Release**.

Any work committed after the deadlines will not count!

Each student is required to hand-in their own **week 1 assignments**

You will hand-in your course project **in teams of two** including:

- 1) A reproducible nextflow/nf-core workflow for RNA-sequencing analysis
- 2) A scientific report of roughly 5 pages

Git Basics: Using your favorite LLM ...

... What is **git** and what are repositories?

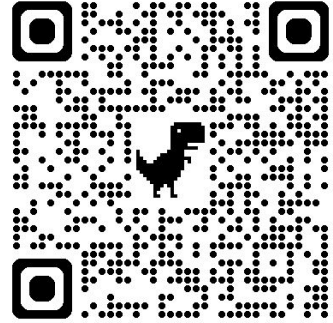
... What is a **fork** on Github?

... What is the difference between
origin and **upstream**?

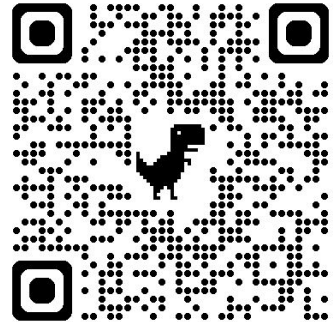


10

min



chat GPT



Gemini

Your typical Day in the Computational Workflows Course

1. You pull today's updates from our **upstream**
2. Work on the assignments
3. Frequently commit and push to your **own fork**
4. Ask questions, if you have problems

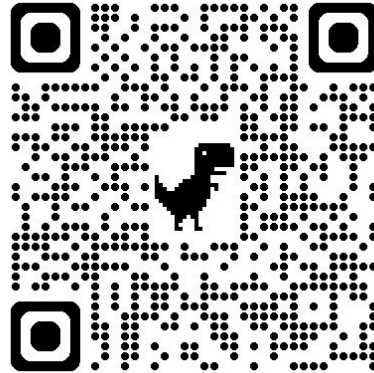


Let's get started!

Demo

Let's get started!

1. Fork the assignment repository on github:



github.com/qbic-pipelines/computational-workflows-2025

Let's get started!

1. Fork the assignment repository
2. Set up ssh-key for Github
3. Clone your **personal fork** to your local machine
4. Set the **assignment repository** as upstream
5. Add us to your fork
6. Start working on today's jupyter notebook
7. Commit and push your work frequently!

```
$ mkdir -p ~/.ssh && cd ~/.ssh
$ ssh-keygen -t rsa
$ nano config

$ cd <your-workspace>
$ git clone git@<your-fork>
$ git remote add upstream <qbic>

$ git add <file1> <file2>
$ git commit -m "adds files"
$ git push origin main
```

Practical Course

Computational Workflows

Day 2

Mark Polster and Julian Flesch

Upfront Information

Please add us to your forks on Github!

@JulianFlesch @mapo9

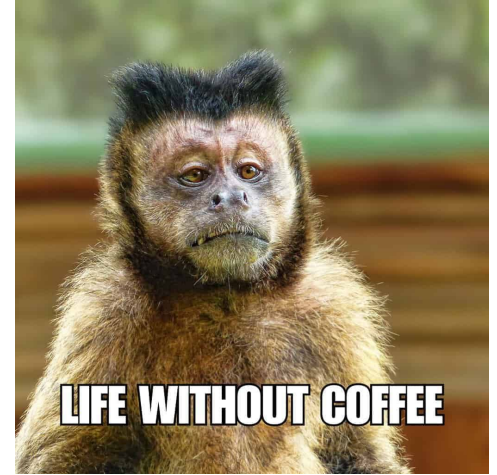
The screenshot shows the GitHub interface for the repository 'computational-workflows-2025-prepared' by user 'JulianFlesch'. The top navigation bar includes links for Code, Pull requests, Actions, Projects, Wiki, Security, Insights, and Settings. A red arrow labeled '1' points to the 'Settings' tab. The left sidebar shows the 'General' tab selected, with a red arrow labeled '2' pointing to the 'Collaborators' link under the 'Access' section. The main content area shows the 'General' settings, including the 'Repository name' field with the value 'computational-workflows-2025-pr' and a 'Rename' button. There is also a checkbox for 'Template repository'.

Upfront Information

Help yourself to coffee, if you want
(list at our desk)

Push your work to Github regularly

Lunch break today from 12:00 to 13:00

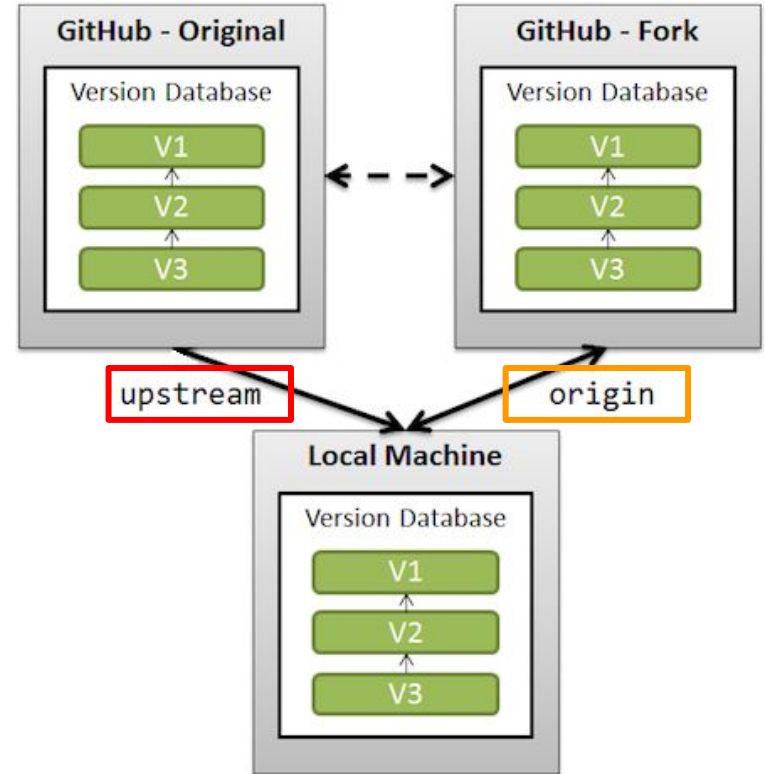


Quick recap: Github Fork and Upstream

We update only the **upstream**

You work on **your fork**, keeping local and remote in sync and pulling in updates

(if we do not change files you worked on, there should be no *merge conflicts*)



Getting your Material for today

Step 1: Save your work to **your fork**

```
$ git add <file1> <file2>

$ git commit -m "adds files"

$ git push
```

Step 2: Get today's notebook **from upstream**

```
# pull from upstream

$ git pull upstream main

# To leave vi:
# ":" -> "q" -> "w" -> "enter"
```

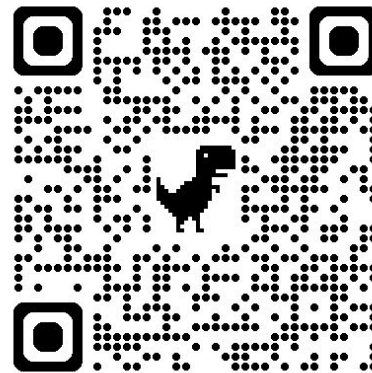
Github Rate Limits?

1. Try setting `GITHUB_TOKEN=<github-token>`
2. Try setting up a **scm** file:

Put this:

```
providers {  
  github {  
    user = 'your-user-name'  
    password = 'your-personal-access-token'  
  }  
}
```

into `$HOME/.nextflow/scm`



nextflow.io/docs

Nextflow Resource Limits?

1. Define you custom config file:

Configure the local executor:

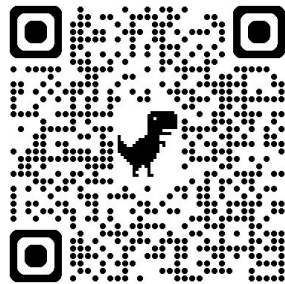
```
executor.cpus = <your-cpus>  
executor.memory = <your-mem>
```

Specify process Limits:

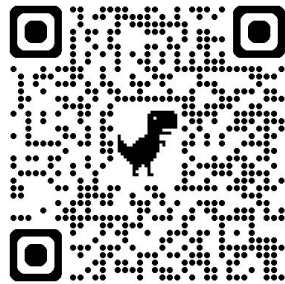
```
process.resourceLimits = [  
    memory: <your-mem>,  
    cpus: <your-cpus>,  
    ... ]
```

2. Use your config:

```
nextflow run ... -c <your-config>
```



Process Resource Limits



Local Executor Settings

Practical Course

Computational Workflows

Day 3

Mark Polster and Julian Flesch

Recall: Paper on the effect of Oxycodon Withdrawl

Sham + oxy



SNI + oxy



Sham + sal



SNI + sal



What we did so far:

Understand experiment and methods

Extract metadata from a lab sheet

Downloaded raw reads with **nf-core/fetchngs**

Build a samplesheet

Process raw sequencing reads with **nf-core/rnaseq**

???

Nextflow Resource Limits?

1. Define you custom config file:

Configure the local executor:

```
executor.cpus = <your-cpus>  
executor.memory = <your-mem>
```

Specify process Limits:

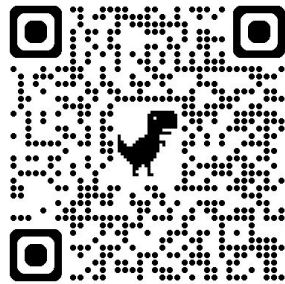
```
process.resourceLimits = [  
    memory: <your-mem>,  
    cpus: <your-cpus>,  
    ... ]
```

2. Use your config:

```
nextflow run ... -c <your-config>
```



Process Resource Limits



Local Executor Settings

Getting your Material for today

Step 1: Save your work to **your fork**

```
$ git add <file1> <file2>

$ git commit -m "adds files"

$ git push
```

Step 2: Get today's notebook **from upstream**

```
# pull from upstream

$ git pull upstream main

# To leave vi:
# ":" -> "q" -> "w" -> "enter"
```

Practical Course

Computational Workflows

Day 4

Mark Polster and Julian Flesch

Upfront Information

Please add us to your forks on Github!

@JulianFlesch @mapo9

The screenshot shows the GitHub repository settings page for the repository 'computational-workflows-2025-prepared' owned by 'JulianFlesch'. The top navigation bar includes links for Code, Pull requests, Actions, Projects, Wiki, Security, Insights, and Settings. A red arrow labeled '1' points to the 'Settings' tab. The left sidebar shows the 'General' tab selected, with a red arrow labeled '2' pointing to the 'Collaborators' link under the 'Access' section. The main content area shows the 'General' settings, including the 'Repository name' field with the value 'computational-workflows-2025-pr' and a 'Rename' button. There is also a checkbox for 'Template repository'.

Upfront Information

No Course tomorrow, because of public Holiday!

Projects start next week. Find a project partner today and tell us about your team (via mail from **one of you**, to **both of us**)

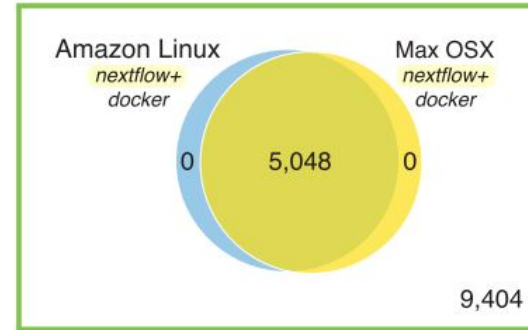
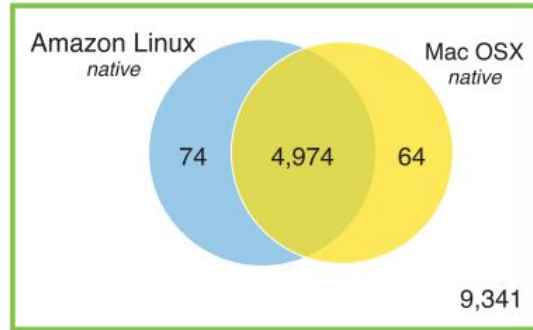
Why even use Docker/Containers?

Number of differentially expressed genes
without Nextflow and containerization

Number of differentially expressed genes
with containerization

C

Transcript quantification and differential expression with Kallisto and Sleuth



DI Tommaso, et al. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*.

Recap from yesterday: Different sources of Images

Existing Container Images:

1. Biocontainers: <https://biocontainers.pro/registry>
(hosted on quay: <https://quay.io/organization/biocontainers>)
2. DockerHub: <https://hub.docker.com/>

Building your own Container Image:

1. Seqera Container Builder: <https://seqera.io/containers/>
2. From scratch with ...
 - ... an existing Dockerfile
 - ... **a custom Dockerfile**

Building the Salmon Dockerfile

Missing Libraries

for apple:

`-platform linux/amd64`

Getting your Material for today

Step 1: Save your work to **your fork**

```
$ git add <file1> <file2>

$ git commit -m "adds files"

$ git push
```

Step 2: Get today's notebook **from upstream**

```
# pull from upstream

$ git pull upstream main

# To leave vi:
# ":" -> "q" -> "w" -> "enter"
```


Practical Course

Computational Workflows

Day 05

Mark Polster and Julian Flesch

Upfront Information

Please add us to your forks on Github!

@JulianFlesch @mapo9

The screenshot shows the GitHub interface for the repository 'computational-workflows-2025-prepared' by user 'JulianFlesch'. The top navigation bar includes links for Code, Pull requests, Actions, Projects, Wiki, Security, Insights, and Settings. A red arrow labeled '1' points to the 'Settings' tab. The left sidebar shows the 'General' tab selected, with a red arrow labeled '2' pointing to the 'Collaborators' link under the 'Access' section. The main content area shows the 'General' settings, including the 'Repository name' field with the value 'computational-workflows-2025-pr' and a 'Rename' button. There is also a checkbox for 'Template repository'.

Recap: Nextflow

- Processes
 - Inputs / Outputs
 - Scripts in a specific language
- Channels + **Operators**
- Workflows
- Advanced Input Concept: **Metamap**
= Sample Data and Metadata in one place!

Next-up: nf-core



DEMO

Week 2 Project

Orga:

- Teams of 2
- Fill out the PROJECT.txt in your fork (*Name1, Name2, Project GitHub URL*)
- **Final Deadline is on 24th of October at 6pm**

Week 2 Project

Topic:

„Build your own nf-core pipeline for analyzing RNA-sequencing data“

First steps:

1. Create Project Github Repository with your partner **and add us to it**
2. Create a pipeline from the the nf-core TEMPLATE
3. Think about what kind of samplesheet you need
4. Today's goal: **Read in a samplesheet**

Pipeline Template



DEMO

Getting your Material for today

Step 1: Save your work to **your fork**

```
$ git add <file1> <file2>

$ git commit -m "adds files"

$ git push
```

Step 2: Get today's notebook **from upstream**

```
# pull from upstream

$ git pull upstream main

# To leave vi:
# ":" -> "q" -> "w" -> "enter"
```


Week 2 Project

Topic:

„Build your own nf-core Pipeline for analyzing RNA-sequencing Data“

First steps:

1. Enter your team into the [PROJECT.md](#) file
2. Create Project Github Repository **and add us to it**
3. Create a pipeline from the the nf-core TEMPLATE:
`nf-core pipelines create`
4. Think about what kind of samplesheet you need
5. Today's goal: **Read in a samplesheet**

Practical Course

Computational Workflows

Day 06

Mark Polster and Julian Flesch

Use a SMALL test dataset for development

Few samples in your samplesheet

Few records per sample:

```
$ gunzip big_file.fq.fz
```

```
$ head -n 30 big_file.fq > small_file.fq
```

```
$ tail -n 30 big_file.fq > small_file.fq
```

```
$ gzip small_file.fq
```

Nextflow Resource Limits?

1. Define a custom config file:

Configure the local executor:

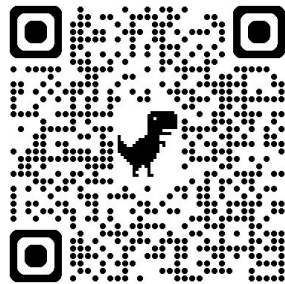
```
executor.cpus = <your-cpus>  
executor.memory = <your-mem>
```

Specify process Limits:

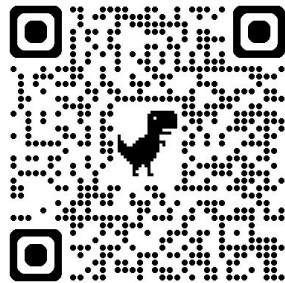
```
process.resourceLimits = [  
    memory: <your-mem>,  
    cpus: <your-cpus>,  
    ... ]
```

2. Use your config:

```
nextflow run ... -c <your-config>
```



Process Resource Limits



Local Executor Settings

Nextflow config

DEMO

The flexibility of the meta map

Practical Course

Computational Workflows

Day 07

Mark Polster and Julian Flesch

Nextflow Resource Limits?

1. Define a limits:

Configure the local executor:

```
executor.cpus = <your-cpus>  
executor.memory = <your-mem>
```

Specify process Limits:

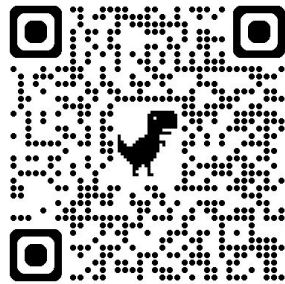
```
process.resourceLimits = [  
    memory: <your-mem>,  
    cpus: <your-cpus>,  
    ... ]
```

2. Place this in nextflow.config:

```
profiles {  
    <your-custom-profile {  
        . . .
```



Process Resource Limits



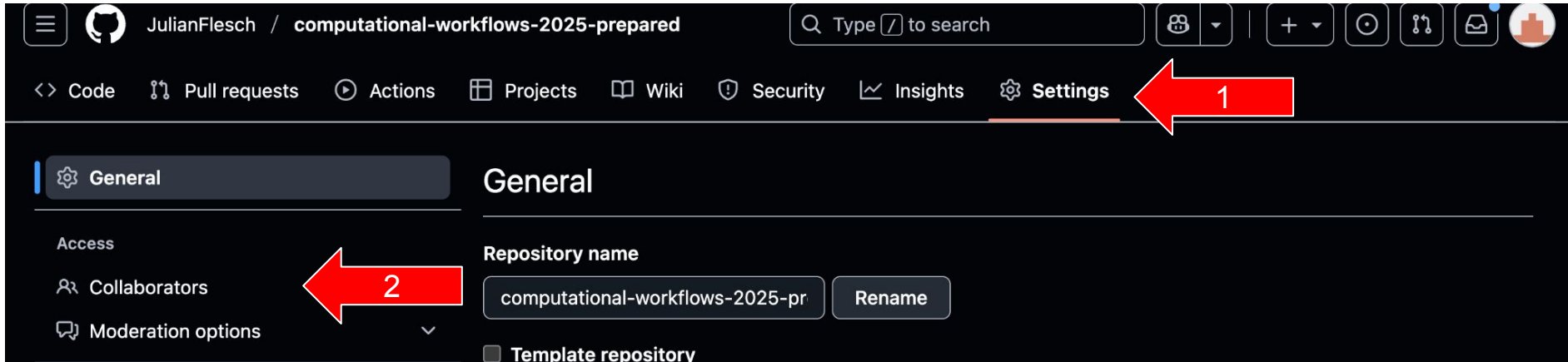
Local Executor Settings

 Happy Coding 

Upfront Information

Please add us to your **forks** and **team projects** on Github!

@JulianFlesch @mapo9



The screenshot shows the GitHub interface for a repository named 'computational-workflows-2025-prepared' owned by 'JulianFlesch'. The top navigation bar includes links for Code, Pull requests, Actions, Projects, Wiki, Security, Insights, and Settings. A red arrow labeled '1' points to the 'Settings' tab. The 'Settings' page is open, showing the 'General' tab. In the left sidebar, under the 'Access' section, a red arrow labeled '2' points to the 'Collaborators' link. The main content area shows the 'Repository name' as 'computational-workflows-2025-pr' with a 'Rename' button, and a checkbox for 'Template repository'.

Practical Course

Computational Workflows

Day 09

Mark Polster and Julian Flesch

 Happy Coding 

Minimal Viable Pipeline

1. some QC
2. Gene count tables per sample (raw counts or tpm or ...)
3. Combined gene count table over all samples

Test Config

```
!nextflow run nf-core/differentialabundance -r 1.5.0 -profile test docker --outdir test_diffab
```

- today we want to build our own test config to run our pipeline

[rnaseq](#) / [conf](#) / **test.config** 



pinin4fjords Remove commented rsem index 

Code

Blame

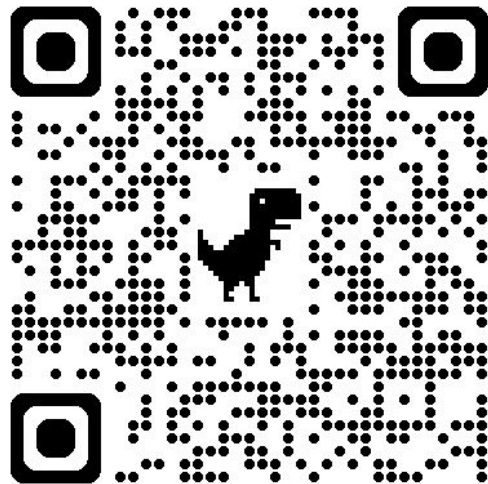
50 lines (41 loc) · 2.47 KB

Metro Maps and Documentation

- part of every good pipeline is thorough documentation. We want to use the README
 - we will grade the docs!
-
- it's a good idea to create your own metro map (e.g. on draw.io)
 - useful for your paper and the docs

Metro Maps and Documentation

- part of every good pipeline is thorough documentation. We want to use the README
- we will grade the docs!
- it's a good idea to create your own metro map (e.g. on draw.io)
- useful for your paper and the docs



Practical Course

Computational Workflows

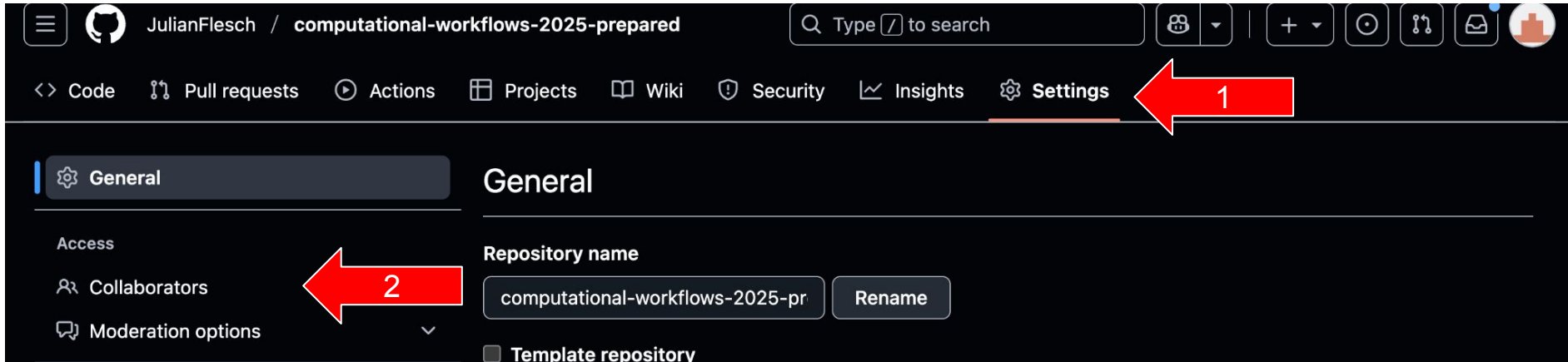
Day 10

Mark Polster and Julian Flesch

Upfront Information

Please add us to your **forks** and **team projects** on Github!

@JulianFlesch @mapo9



The screenshot shows the GitHub repository settings page for the repository 'computational-workflows-2025-prepared' owned by 'JulianFlesch'. The top navigation bar includes links for Code, Pull requests, Actions, Projects, Wiki, Security, Insights, and Settings. A red arrow labeled '1' points to the 'Settings' tab. The left sidebar shows the 'General' tab selected, with sub-sections for Access, Collaborators, and Moderation options. A red arrow labeled '2' points to the 'Collaborators' section. The main content area shows the 'General' settings, including the 'Repository name' field with the value 'computational-workflows-2025-pr' and a 'Rename' button. There is also a checkbox for 'Template repository'.

Upfront Information

Push your **assignment forks** and **team projects** to Github!

Coffee

Please pay up your coffee bill.

1 coffee == 50 cents



 Happy Coding 