

# Data Science in Bioinformatics

## Day 4 - Snakemake

To help you get familiar with Snakemake and how it works, we'd like you to work through the tutorial on the Snakemake documentation page for this exercise. We know you can find the results for the exercises there as well, so of course you could just copy and paste the code and be done with it. But that wouldn't get you any closer to understanding how Snakemake works. So please try to do the exercises yourself before looking up the code.

The goal of the workflow you will create in this tutorial is to map sequencing samples located in a „data/samples“ folder to the reference genome in „data/genomes.fa“. Then you will call genomic variants over the mapped samples and create a sample plot. At the end, a report will be generated summarising all results.

Snakemake documentation short tutorial:

<https://snakemake.readthedocs.io/en/stable/tutorial/short.html>

We would also like you to add some things to your code, that are not explicitly named in the tutorial:

- 1) There is the possibility to add a config-file „config.yaml“ to your Snakemake workflow. Here you can add not only the names of the samples but many other things like pre-set parameters that could be changed for the workflow, depending on the data used. Please add that to your workflow and fill it with a few parameters needed in your code.
- 2) There is the possibility to set the parameters used in shell commands in a parameter directive, like the input or output directive of the rule. Please set the parameters you use in the shell commands of your workflow in this directive. Then call them in a shell command with {params.Parametername}.
- 3) You can specify a log file for every rule, in which errors can be logged. Please do this for every rule in your workflow under the „log“ directive.
- 4) Think about which files are needed for the report and which files are just intermediate results, that don't need to be saved in the end. Mark the files that should be deleted in the end of the workflow as temporary.