

Elective Module (Wahlpflichtmodul)

Special Topics: 9th October to 13th October 2023
Final Exam on 18th October 2023

Responsible: **Prof. Dr. Axel Blokesch**, blokesch@fb2.fra-uas.de

Introduction to Bioinformatics

by Pablo Riesgo-Ferreiro, Tron gGmbH

M.Sc. in Computer Science, Universidad Politécnica de Madrid, 2008

M.Sc. in Bioinformatics and Biostatistics, Universitat Oberta de Catalunya /
Universidad de Barcelona, 2016

Intended learning outcomes

- Familiarize with common computational genomic pipelines and data formats
- Understand reproducibility and FAIR principles (findability, accessibility, interoperability, and reusability)
- Understand open source, open science and open data initiatives
- Practical experience with Git as a free and open source distributed version control system to track and version analysis code and software
- Understand general principles of workflow managers
- Practical experience with Nextflow as a workflow manager

Timetable:

Day		Topic
Monday	9 th October 2023 9:00 – 10:30 10:45 – 12:15 13:30 – 15:00	Brief introduction to popular computational genomics methods used in current research Alignment Variant calling Differential expression Methylation Single cell approaches Standard genomic data formats: FASTA, FASTQ, BAM, VCF The reproducibility crisis or the death by the thousand cuts
Tuesday	10 th October 2023 9:00 – 10:30 10:45 – 12:15 13:30 – 15:00	Replicate and reproduce FAIR principles (findability, accessibility, interoperability, reusability) Open approaches: open source, open science, open data Data sharing and the Genomic Alliance For Global Health (GA4GH) ➤ Practical use case 1: autopsy of a scientific publication Introduction to software engineering tools to enable

		reproducibility Git as a tool designed to handle everything from small to very large projects with speed and efficiency. Best practices, quality and assisted development See next page.
Day		Topic
Wednesday	11 th October 2023 9:00 - 10:30 10:45 - 12:15 13:30 - 15:00	Automation and continuous integration Software registries (eg: PyPI, CRAN, bioconductor, bioconda) Programming in the biomedical sciences or standing on the shoulders of others ➤ Practical use case 2: develop, install and automate a variant calling pipeline for SARS-CoV-2 in GitHub continuous integration environment Introduction to workflow managers or how to write better pipelines Common features of a workflow manager Python, R, Java or any of them?
Thursday	12 th October 2023 9:00 - 10:30 10:45 - 12:15 13:30 - 15:00	Existing alternatives (eg: Common Workflow Language (CWL), Workflow Description Language (WDL), Galaxy, SnakeMake, Apache Airflow, Nextflow) The Nextflow universe and the NF-core community Discussion of the practical use case on SARS-CoV-2 variants, improve it and run a NF-core pipeline for this issue Processes, subworkflows, channels and the DAG, (Directed Acyclic Graph) Integration via configuration with different computational environments Managing environments with conda and containerization with docker/singularity
Wednesday	19 th October 2023	Final Exam from 8:15 to 9:45 a.m.

Room for Lectures: 7 / 204

Room for the final exam: Still to be announced.

If you have participated actively in the course and worked on the practical use case you are allowed to use your written records in the final exam an complete tasks related to these cases with a rating of 15 % (case 1) and 25 % (case 2). If you did not edit these practical use cases (or only one of them), you have to complete another task in the exam instead.