

Course Syllabus: Understanding Bioinformatic Pipelines - BESE 394A

Offering Department	Biological & Environmental Sc. and Eng.
Course Number	BESE 394A
Course Title	Understanding Bioinformatic Pipelines
Academic Semester	Spring 2023/2024
Semester Start Date	01/21/2024
Semester End Date	05/15/2024
Class Schedule (Days & Time)	Understanding Bioinformatic Pipelines Lecture BESE 394A Tue 08:30 - 15:30

Instructor(s)				
Name	Email	Phone	Office Location	Office Hours
David Gomez-Cabrero	DAVID.GOMEZCABRERO@KAUST.EDU.SA			

Teaching Assistant(s)	
Name	Email
Vincenzo Lagani	vincenzo.lagani@kaust.edu.sa
Robert Lehmann	robert.lehman@kaust.edu.sa

Course Information	
Course Description	<p>Data analysis has become a fundamental part of biological and biomedical research. The rapid development of bioinformatics during the last three decades is a clear evidence for that. In order to master bioinformatics the first steps is to learn programming and statistics. After that, bioinformatics becomes as heterogeneous as the experimental protocols generating the data. Consequently, for each data type there is the need to develop, implement and use specific analytical frameworks. Fortunately, there are common traits that accomunate the vast majority of such analytical frameworks, and these common traits can be used for devising appropriate data-analysis pipelines whenever a novel data-type appears.</p> <p>The idea of this course is NOT teaching a set of pipelines for preselected data-types, but to teach how to think when preparing an analysis pipeline. That includes, among others, learning how to ask the right questions to the data depending on the goal of the study, setting quality control steps, understanding the statistics behind the data, properly visualizing the data, and how to optimize the programming implementations.</p>

	<p>The course is conceived as an 8 week course, where every week there will be different pipelines, data-sets and aspects to learn. Through the course, the student will develop the necessary practical skills to develop new pipelines, and review existing ones. A particular aim is to provide the participants with long-term programming skills and the guidelines for improving their knowledge of it.</p>
Learning Outcomes	<p>Understanding and applying bioinformatic pipelines:</p> <ul style="list-style-type: none"> • Understanding and conducting quality controls. • Understanding and conducting pre-processing. • Understanding and conducting normalization. • Understanding and conducting differential analysis. • From data analysis to biomedical understanding: what tools and concepts are available? <p>Understanding the different data types:</p> <ul style="list-style-type: none"> • Understanding the different types of data. • Understanding the appropriate statistical tools for each data type. <p>Implementing bioinformatic frameworks:</p> <ul style="list-style-type: none"> • Using shell, R or Python when necessary. • Automatization and parallelization (e.g. Snakemake). • Generating quality reports. • Data and code can be shared: by following proper sharing protocols. • Sharing in Github.
Textbook/Materials	<ol style="list-style-type: none"> 1. https://snakemake.readthedocs.io/en/stable/tutorial/basics.html. 2. http://www.r-project.org 3. Introductory Statistics with R. Dalgaard P. Springer. 4. A Handbook of Statistical Analyses Using R. Everitt BS, Hothorn T. Chapman and Hall/CRC. 5. https://www.python.org/about/gettingstarted/ 6. Python Programming: An Introduction to Computer Science, 3rd Ed. Zelle J. Franklin, Beedle & Associates. 2016. 7. Learning Python, 5th Ed. Lutz M. O'Reilly Media. 8. https://www.statlearning.com
Method of Assessments	<p>40.00% - Homework /Assignments 20.00% - Midterm exam 20.00% - Final exam 20.00% - Active participation</p>
Nature of the Assignments	<p>There will be four types of assignments:</p> <ul style="list-style-type: none"> • Short exercises: aimed for students to address their understanding of the concepts develop during the class. To be completed during out-of-class hours. • Additional short/medium length exercises: aimed for highly motivated students that aim for deeper insights into the topic. • Exam per data type: a data-set will be provided to the students, along with a set of queries that can be answered by analyzing the data. For example, the students may be asked to analyze single-cell RNA-seq related to hematopoiesis and tasked with identifying the possible factors that drive cell differentiation. Two documents must be provided: <ol style="list-style-type: none"> 1) A 2-pages data-analysis report: summarizing the key findings identified and the steps to reach such findings. 2) The code (annotated) used to reach such conclusions. • Final exam: <ol style="list-style-type: none"> 1) R, Python and HPC must be used in the analysis. 2) Pipeline implementation and usage must be documented. 3) A 4-pages report. 4) Replicate the analysis of a paper or generate a complete new analysis. <p>The evaluation of the exam will be based on the appropriate use of the tools lectured during the course.</p>
Course Policies	<ul style="list-style-type: none"> - Unauthorized absence may be considered a reason to fail the course. - The course final exam must be finalized, and the final report must be submitted approximately one week after the end of the course. In exceptional cases, additional time may be permitted.

Additional Information**Tentative Course Schedule**

(Time, topic/emphasis & resources)

Week	Lectures	Topic
1	Tue 01/23/2024	Pipeline analysis: what is it and general guidelines for mastering pipelines. Investigating our first pipeline: RNA-seq. RNA-seq hands-on.
2	Tue 01/30/2024	Setting a pipeline: code, Github, tracing back analysis, Reimplement the RNA-seq pipeline over the new concepts. Review the statistics behind count data I.
3	Tue 02/06/2024	ATAC-seq. Understanding and implementing the pipeline Review the statistics behind count data II.
4	Tue 02/13/2024	Single-cell RNA-seq. Understanding and implementing the pipeline.
5	No schedule	No class
6	Tue 02/27/2024	Single-cell multi-ome. Understanding and implementing the pipeline.
7	Tue 03/05/2024	No class
8	Tue 03/12/2024	Setting an integrative pipeline.
9	Tue 03/19/2024	DNA Methylation. Understanding and implementing the pipeline.
10	Tue 03/26/2024	No class
11	Tue 04/02/2024	No class
12	No schedule	No class
13	Tue 04/16/2024	No class
14	Tue 04/23/2024	No class
15	Tue 04/30/2024	No class
16	Tue 05/07/2024	Per groups a single-pipeline: HiC, CITE-seq, CyToF, etc.

Note

The instructor reserves the right to make changes to this syllabus as necessary.