Universidad Zaragoza

Academic Year/course: 2024/25

68461 - Big Data in Biology

Syllabus Information

Academic year: 2024/25
Subject: 68461 - Big Data in Biology
Faculty / School: 100 - Facultad de Ciencias
Degree: 626 - Máster Universitario en Biofísica y Biotecnología Cuantitativa / Master in Biophysics and Quantitative Biotechnology
ECTS: 6.0
Year: 01
Semester: Second semester
Subject type: Optional
Module:

1. General information

This course introduces data science applications in biomedical research, focusing on fields where its applications are most widespread, such as the study of -omics data compiled using Next-Generation Sequencing (NGS) technologies, especially transcriptomics. The theory underlying some of the statistical models frequently used to analyze NGS data at bulk and single-cell resolutions is discussed, along with their practical applications for testing hypotheses in fields such as evolutionary biology and immunology, among others.

The course is designed for an interdisciplinary audience, including students from formal/quantitative and biomedical programs alike, and its content complements the course in “Systems and Synthetic Biology.” Strong R programming skills are recommended, and the optional course on biostatistics and bioinformatics is advised. The course aligns with several Sustainable Development Goals of the United Nations 2030 Agenda, including Goals 2, 3, 9, 14, and 15.

2. Learning results

- Understanding the main applications of data science in contemporary Biomedical research and the impact of massive data collection and analysis in Biology and Medicine.
- Acquiring the theoretical, mathematical, and statistical foundations underlying the main pipelines for NGS data analysis in Biomedicine.
- Learning how to use the most popular computational implementations of these pipelines at each stage of the analyses: from quality control to data modeling and biological interpretation of the results.
- Acknowledging the importance of the conceptual implications that certain technical aspects of data analysis may exert in Biological and Medical sciences.
- Identify and avoid problematic practices that compromise reproducibility in biomedical data analyses.

3. Syllabus

Topic 1: General overview of the main fields and applications of Big data science in Bio-medicine.
Topic 2: Cleaning up the data: Quality Control.
Topic 3: Modeling the data (1): Statistical frameworks to model -omics data.
Topic 4: Modeling the data (2): complex designs, confounders, and batch effects.
Topic 5: Interpreting the results: Enrichment analyses.
Topic 6: Single-cell -omics technologies: data generation, and analytic strategies in complex tissues.
Topic 7: Ethics and reproducibility in biomedical data analysis.

4. Academic activities

- Theoretical lectures (12h) using slides, R-markdown documents, or analogous materials, deal with the explanation of theory and methods, organized according to the syllabus of the course.
- Practical lectures (18h), where examples of computational implementations of the analysis described in the theory sessions will be presented to the students, using a combination of slides, R markdown documents, and code scripts.
- Computer lab sessions (24h), where students will be asked to solve specific problems, and implement analytical pipelines applying what was presented in theoretical and practical lectures.
- The presentation of short seminars (6h) on a practical case previously proposed.
- Individual (not in-person) work (90h).

5. Assessment system

The course will be evaluated using a global system, comprising the following activities:

1: Seminar: (10%). Within the first month of the course, the teacher will introduce a series of scientific studies involving the collection and analysis of big-data sets in disparate fields. The students will prepare a short presentation on one of these works before the class. Depth and clearness of presentation will be assessed and evaluated. The datasets in the studies presented will be later used extensively during the practice sessions as case examples.

2. Practice reports: (40%). Continuous evaluation of the students' progress during the practical sessions, through the correction of practice reports.

3.1 Final exam -practice- (25%). Towards the end of the course, the teacher will provide materials about a practical case, about which the students will prepare a written report including the description of the analytic strategy used, and the interpretation of the results. Coherence and completeness of the analyses proposed, understanding of the subject and accuracy in addressing the proposed question will be assessed and evaluated.

3.2 Final exam -theory- (25%). Written exam on the theory topics discussed throughout the course.

6. Sustainable Development Goals

2 - Zero Hunger
3 - Good Health & Well-Being
9 - Industry, Innovation and Infrastructure