Universidad Zaragoza

Curso Académico: 2023/24

68461 - Big Data en Biología

Información del Plan Docente

Año académico: 2023/24
Asignatura: 68461 - Big Data en Biología
Centro académico: 100 - Facultad de Ciencias
Titulación: 626 - Máster Universitario en Biofísica y Biotecnología Cuantitativa / Master in Biophysics and Quantitative Biotechnology
Créditos: 6.0
Curso: 01
Periodo de impartición: Segundo semestre
Clase de asignatura: Optativa
Materia:

1. Información básica de la asignatura

This course introduces data science applications in bio-medical research, focusing on fields where its applications are more widespread, such as the study of -omics data compiled using Next-Generation Sequencing (NGS) technologies, specially transcriptomics. The theory underlying some of the statistical models more frequently used to analyze NGS data at bulk and single-cell resolutions is discussed, as well as its practical applications to test specific hypotheses in fields such as evolutionary Biology, or immunology, among others.
The course is designed for an interdisciplinary audience including students from formal/quantitative and bio-medical programs alike, and its contents complement the course in “Systems and Sinthetic

Biology”. Strong R programming skills are recommended, and taking the optional course in Bio-statistics and Bio-informatics 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. Resultados de aprendizaje

• 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, from quality control to data modelling 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. Programa de la asignatura

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

4. Actividades académicas

- 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, 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. Sistema de evaluación

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. Coherence and completeness of the analyses proposed, and understanding of the subject, as well as 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 student's progress during the practical sessions, through the correction of the 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, and understanding of the subject, as well as clearness of presentation will be assessed and evaluated.
3.2 Final exam -theory- (25%). Written exam on the theory topics discussed throughout the course.