

Degree	Type	Year
2503740 Computational Mathematics and Data Analytics	OT	4

Contact

Name: Angel Gonzalez Wong

Email: angel.gonzalez@uab.cat

Teachers

Gianluigi Caltabiano

Angel Gonzalez Wong

Teaching groups languages

You can view this information at the [end](#) of this document.

Prerequisites

Basic knowledge of the English language is required as part of the teaching material is in that language. Recommended to have taken the Bioinformatics course.

Objectives and Contextualisation

The course aims to provide a view on the possibilities of big data analysis in bioinformatics.

The course consists of two parts: 1) computational methodologies applied to drug discovery and 2) analysis of omics data. The course is part of the Mention in Statistics for Health Sciences.

Learning Outcomes

1. CM34 (Competence) Propose suitable statistical models for epidemiological studies.
2. CM34 (Competence) Propose suitable statistical models for epidemiological studies.
3. CM35 (Competence) Write technical reports that clearly express the results and conclusions of a bioscience study using vocabulary specific to the field of application.
4. KM29 (Knowledge) Recognise the most used statistical inference methods in bioinformatics.
5. KM30 (Knowledge) Identify the use of statistical knowledge in bioinformatics and in health science.
6. KM30 (Knowledge) Identify the use of statistical knowledge in bioinformatics and in health science.
7. KM31 (Knowledge) Identify the most used statistical inference methods in epidemiology studies.

8. SM36 (Skill) Analyse data corresponding to epidemiological studies or clinical trials.
9. SM36 (Skill) Analyse data corresponding to epidemiological studies or clinical trials.
10. SM36 (Skill) Analyse data corresponding to epidemiological studies or clinical trials.
11. SM37 (Skill) Apply statistical methods to the analysis of gene expression data.
12. SM38 (Skill) Use the most common databases in the field of health science.

Content

PART 1. Big Data in Drug Discovery

1. Introduction to Big Data in Life Sciences.
2. Database and representation of biological components and chemical compounds.
3. Analysis, clustering, and visualization of chemical and pharmacological substances.
4. Virtual Screening in Drug Discovery.

PART 2. Big Data in Omics Analysis

1. Introduction to Bioconductor and bioinformatics tools for the analysis of omic data.
2. Genetic Association Studies and GWAS (Genome-wide association studies).
3. Multivariate Methods for Omics Data Analysis Integration and Big Data.

*Unless the requirements enforced by the health authorities demand a prioritization or reduction of these contents.

Activities and Methodology

Title	Hours	ECTS	Learning Outcomes
Type: Directed			
Practical sessions	21	0.84	
Presentation of Research Project	3	0.12	
Theory classes	21	0.84	
Type: Supervised			
Tutoring	10	0.4	
Type: Autonomous			
Preparation of Research Project	20	0.8	
Study	70	2.8	

The course is organized in sessions of 3 hours. Each session consists of a theoretical part (theory classroom) that will introduce the new concepts followed by a practical part (computer room) where the students will work on the implementation of concepts explained in the theoretical part. In each session the teacher will indicate the students some tasks to do autonomously, such as reading articles or sending reports. The material used by the teachers will be available on the Virtual Campus of the course.

***The proposed teaching methodology may experience some modifications depending on the restrictions to face-to-face activities enforced by health authorities.**

Annotation: Within the schedule set by the centre or degree programme, 15 minutes of one class will be reserved for students to evaluate their lecturers and their courses or modules through questionnaires.

Assessment

Continous Assessment Activities

Title	Weighting	Hours	ECTS	Learning Outcomes
Presentation Research Project	20	0.5	0.02	CM34, CM35, KM29, KM30, KM31, SM36, SM37, SM38
Presentation of practicum reports	40	0.5	0.02	CM34, CM35, KM29, KM30, KM31, SM36, SM37, SM38
Theoretical-practical exams	40	4	0.16	CM34, CM35, KM29, KM30, KM31, SM36, SM37, SM38

PART 1. Big Data in Drug Design (50%):

- practical exercises (30%)
- bioinformatics oral presentation of a project (20%)

BLOCK 2. Big Data in Data Analysis (50%):

- practical exercises (30%)
- theoretical-practical test (20%)

The minimum global qualification required to pass the subject will be 5 points. The minimum mark of each of the evaluated activities must be equal to or greater than 4 points. Students who have any of the parts suspended will be able to do the recovery exam where they can be re-examined from the suspended part.

Bibliography

- Lesk A.M. *Introduction to Bioinformatics*. Oxford University Press 2005.
- Attwood, T.K., Parry-Smith, D.J., *Introducción a la Bioinformática*. Pearson Education, 2002.
- Foulkes A.S. *Applied Statistical Genetics with R. For Population-based Association Studies*. Springer Dordrecht Heidelberg London New York. ISBN 978-0-387-89553-6
- Gonzalez JR, Cáceres A. *Omic association studies with R and Bioconductor*. Chapman and Hall/CRC, ISBN 9781138340565, 2019.
- <https://www.bioconductor.org/>

Software

R: <https://www.r-project.org/>

Rstudio: <https://www.rstudio.com/>

Language list

Name	Group	Language	Semester	Turn
(PLAB) Practical laboratories	1	Catalan	first semester	afternoon
(TE) Theory	1	Catalan	first semester	afternoon