# UAB
## Universitat Autònoma de Barcelona

## Big Data Analysis in Bioinformatics

Code: 104886
ECTS Credits: 6

**2024/2025**

| Degree | Type | Year |
|---|---|---|
| 2503852 Applied Statistics | OT | 4 |

## Contact

Name: Angel Gonzalez Wong

Email: angel.gonzalez@uab.cat

## Teachers

Gianluigi Caltabiano

Angel Gonzalez Wong

Juan Ramon Gonzalez Ruiz

## 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. CM14 (Competence) Propose the statistical model needed to analyse data sets belonging to real studies.
2. KM17 (Knowledge) Recognise the statistical models for the analysis of data with different structures and complexities that frequently appear in different fields of application.
3. KM18 (Knowledge) Recognise the language of applications of economics and finances, biomedical science and engineering, provided by research and innovation in the field of statistics.
4. KM18 (Knowledge) Recognise the language of applications of economics and finances, biomedical science and engineering, provided by research and innovation in the field of statistics.

5. SM16 (Skill) Select appropriate sources of information for the statistical work.
6. SM19 (Skill) Analyse complex data, whether this is due to their characteristics or their size.

## Content

PART 1. Big Data in Drug Discovery

1. Introduction to Big Data in Life Sciences.
2. Databases 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.

## 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.

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 |
| --- | --- | --- | --- | --- |
| Practicum Reports Preparation | 60 | 1 | 0.04 | CM14, KM17, KM18, SM19 |
| Presentation Research Project | 20 | 2 | 0.08 | CM14, KM17, KM18, SM16, SM19 |
| Theoretical-Practical Exam | 20 | 2 | 0.08 | CM14, KM17, KM18, SM19 |

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

- Practical Exercises (30%)

- Presentation of a Bioinformatics 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 |