Universidad
de Navarra

*Bioinformatics and Next Generation Sequencing (MIB)*

*Guía docente 2025-26*

# PRESENTACIÓN

**Breve descripción**:

The main objective of the subject is for the student to acquire enough knowledge so as to be able to analyze Next Generation Sequencing data.

- **Titulación**: Máster en Ingeniería Biomédica
- **Módulo/Materia**: Especialidad en Análisis de Datos / Análisis de Datos Biomédicos y Moleculares (Especialidad Tecnologías, Sistemas y Dispositivos Biomédicos: Módulo Optativo/Materia Optatividad)
- **ECTS**: 5 ECTS
- **Curso, semestre**: 1º, Segundo
- **Carácter**: Obligatorio
- **Profesorado**:
    - Rubio Díaz-Cordovés, Ángel - Email: arubio@tecnun.es / Catedrático
    - Ortiz Estévez, María - Email: mortiz@ceit.es / Invitado (Colab.Docente)
    - Otaegui Bichot, David - Email: dotaeguibic@external.unav.es / Invitado (Colab.Docente)
- **Idioma**: Inglés

# RESULTADOS DE APRENDIZAJE (Competencias)

- CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación
- CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio
- CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades
- CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.
- CG02 - Realizar investigación, desarrollo e innovación en productos, procesos y /o métodos en ingeniería biomédica.
- CG05 - Proyectar, calcular y diseñar productos, procesos, instalaciones y sistemas de control en el área de la ingeniería biomédica.
- CE09 - Generar nuevo conocimiento para aplicaciones en ingeniería biomédica integrando datos derivados de distintos análisis tanto clínicos como de biología molecular

# PROGRAMA

This subject is composed of three different parts. After the ending of each part, the students are required to present an assigment where they apply the content of each of these parts. In December, there is an exam with short questions.

### Introduction to Genomics.

- Overall review of the topic
- Study of RNA, DNA and proteins.
- Structure of a gene.
- Hallmarks of cancer

### Topic 1. RNA analysis.

- What is Gene expression?
- Introduction to techniques to measure gene expression.
- Analysis of gene expression: linear models and the R Limma packages.
  - Getting useful information from gene expression experimentsThe Gene Ontology
  - KEGG, Panther and other pathway databases
  - Enrichment analysis
- Downloading expression experiments: GEO and ArrayExpress.

Assignment: analysis of an expression experiment downloaded from GEO. This analysis must include a design matrix with more than two conditions and GO enrichment analysis.

### Topic 2. Prognosis in cancer. Machine learning techniques.

- Survival analysis.
  - Censored data
  - Survival and hazard functions
  - Cox Proportional Hazard regression
- Description of the third assignment: prediction of the outcome in breast cancer.
- Data analysis techniques:
  - Dimensionality reduction: SVD and PCA
- Feature selection:
    - Filter, wrapper and embedded methods
  - Lasso regression
  - Supervised classification: logistic regression, SVM, randomForests

Assigment: Contest on the prediction of the outcome of the given cancer using genomic and clinical data.

### Topic 3. DNA analysis.

- DNA variants across individuals
- Types of mutations
  - Intergenic, intronic, nonsense, missense,...
  - Passenger and driver mutations.
- Sequence analysis. Blast and other matching techniques.
- Studying genotype and alterations of the genome.
  - SNP arrays
  - DNA sequencing
- Visual inspection of the aberrations: IGV software.

**Assigment: Study on copy number alterations and mutation on severity of cancer.**

## ACTIVIDADES FORMATIVAS

There is schedule for the clases available in the Contents folder.

The classes will consist of

- Solving the quizes of the previous clases (15 m approx),
- Theoretical foundations (sequencing technologies, genomics, statistics and bioinformatics) (1-2h)
- Hands on and application of these foundations (1-2h).
- Each student will have access to a Virtual Machine and to the Supercomputer of the DIPC. In this Virtual Machine most of the software is already preinstalled. In the first classes there will be an introduction to Linux and Virtual Machines.

## EVALUACIÓN

### CONVOCATORIA ORDINARIA

During the subject I will send some quizzes to solve: 70%

There is also a final project (30%).

- The final project will be done in group. Each member of the group must state clearly which is the part develped by him/her.

### CONVOCATORIA EXTRAORDINARIA

An exam on the theoretical content of the subject (60%)

- Reevaluation of the final project (40%). In this case, each student will be assigned a specific project that should be solved individually.

## HORARIOS DE ATENCIÓN

Angel Rubio (arubio@unav.es)

The office hours are Wednesday from 15:30 to 18:30.

- Any other time is also OK if arranged previously with the lecturers.

## BIBLIOGRAFÍA Y RECURSOS

This is an evolving topic and most bibliography sources become obsolete in few years.

- We will use papers and Internet resources instead.