COURSE DATA

Data Subject	
Code	36434
Name	Statistics for omics data
Cycle	Grade
ECTS Credits	6.0
Academic year	2023 - 2024

Study (s)

Degree	Center	Acad. year	Period
1406 - Degree in Data Science	School of Engineering	3	Second term

Subject-matter

Degree	Subject-matter	Character
1406 - Degree in Data Science	11 - Health	Obligatory

Coordination

Name	Department
AYALA GALLEGO, GUILLERMO	130 - Statistics and Operational Research
DZUNKOVA, MARIA	194 - Genetics

SUMMARY

An introduction to statistical analysis of omics data is proposed. Among others, we will use data obtained with microarrays and RNA-Seq. We study the previous basic biological concepts. Techniques are discussed omics, pre-processing of information and statistical problems. In particular, we study the problem of multiple comparisons. Then we show its application to the problem of differential expression. We also consider gene set analysis. The software used is R and Bioconductor.

Theoretical classes will be taught in Spanish and the classes practices and laboratory according to the file of the subject available on the web of the degree.

PREVIOUS KNOWLEDGE

Relationship to other subjects of the same degree

There are no specified enrollment restrictions with other subjects of the curriculum.

Other requirements

A first course in Statistics as well as a intermediate level using the R language.

1406 - Degree in Data Science

- (CG01) Knowledge of basic subjects and technologies that enable students to learn new methods and technologies, and to provide them with versatility to adapt to new situations.
- (CG03) Capability to elaborate models, calculations, reports, to plan tasks and other works analogous to the specific field of data science.
- (CG07) Ability to autonomously make decisions and to properly and originally elaborate reasoned arguments, in order to obtain reasonable and contrastable hypotheses.
- (CT01) To be able to access (bibliographical) information tools and appropriately use them in the development of their daily tasks.
- (CT03) Ability to defend your own work with rigor and arguments and to expose it in an adequate and accurate way with the use of the necessary means.
- (CE12) Ability to design and start solutions based on data analysis in the field of medicine and business, taking into account the specific requirements of this type of use cases.
- (CE15) Ability to model and analyse the uncertainty in data-based studies, as well as to know how to interpret and contextualise the results obtained.
- (CB2) Students must be able to apply their knowledge to their work or vocation in a professional manner and have acquired the competences required for the preparation and defence of arguments and for problem solving in their field of study.
- (CB5) Students must have developed the learning skills needed to undertake further study with a high degree of autonomy.

Learn the special characteristics of data in biomedicine. Know the omics data (proteomic, genomic and metabolic). Know the applications of data science in biomedicine. Apply known methods to biomedical problems.
In particular, perform a complete analysis with microarrays and RNA-Seq with R / Bioconductor with special interest in the use of reproducible research tools such as Rmarkdown.
(Assessment of competences CB5, CG7, CT3, CE15)

DESCRIPTION OF CONTENTS

1. Biology part - generation of omics data

Introduction to omic data.
Nucleic acids and proteins.
The genomes of organisms.
Chromosomes, mutations and inheritance.
Nucleic acid extraction and PCR.
Preparation of nucleic acids for sequencing.
Different sequencing platforms.
File types in genetic analysis.
Gene expression and epigenetics.
Mutation and population variation.
General sequence databases.
BLAST program, gene annotation, sequence mapping.
Cancer: gene and chromosomal mutations.
Applications of DNA sequencing.

2. Introduction to omics data

Introduction to omics data. R and Bioconductor. Genome annotation.

3. Data

Microarrays. Obtaining and preprocessing microarray data. RMA method.
GEO and Array Express. RNA-Seq. Obtaining and preprocessing RNA-Seq data.
Alignment and normalization.

4. Differential expression

Marginal (or gene to gene) differential expression.
Microarrays. Non-specific selection. Multiple comparisons. Benjamini-Hochberg method.
The q-value. Limma method for microarray data.
RNA-Seq: edgeR and DESeq2 methods.

5. Gene set analysis

Differential expression considering gene sets.
Gene sets using GO and Kegg. Over-representation analysis. Analysis based on enrichment.

WORKLOAD

ACTIVITY	Hours	% To be attended
Theory classes	32,00	100
Laboratory practices	20,00	100
Classroom practices	8,00	100
Study and independent work	60,00	0
Preparation of evaluation activities	24,00	0
Resolution of online questionnaires	6,00	0
TOTAL	**150,00**	

TEACHING METHODOLOGY

Master classes (45 hours) are proposed as well as practical classes in the Computer room (fifteen). Material developed in a manual is provided. Questionnaires will be proposed weekly online.

(Competency assessment CB5, CT1)

EVALUATION

In the first call, the course will be evaluated according to two aspects:

SE1: Objective test, which will be held at the end of the course and will consist of theoretical and practical questions and problems. This section of the evaluation will account for 70% of the final grade. Of this 70%, 50% (over the total of the subject) corresponds to the Statistics part and the remaining 20% (over the total of the subject) corresponds to the Biology part (evaluation of the competences CB4, CB5, CG1, CE9, CE15). The classes of the biology part are compulsory and the attendance forms 10% of the grade of the biology part.

SE2: Completion of questionnaires during the course evaluating, by means of theory and practice questions, the student's work. This part corresponds completely to the Statistics part and represents 30% of the final grade. This part of the evaluation is not recoverable (evaluation of the competences CB2, CB5, CG1, CT2, CE9, CE15).

The distribution of the evaluations will be done in such a way that the first block of biological contents will have a global evaluation of 20% and the remaining blocks of 80%.

Section SE2 is not recoverable in the second call.

In any case, the evaluation system will be governed by the provisions of the Evaluation and Qualification Regulations of the University of Valencia for Bachelor's and Master's degrees.

(https://webges.uv.es/uvTaeWeb/MuestraInformacionEdictoPublicoFrontAction.do?accion = inicio & idEdictoSeleccionado = 5639).

REFERENCES

Basic

- Bioinformática Estadística. Análisis estadístico de datos ómicos. Guillermo Ayala. https://www.uv.es/ayala/docencia/tami/tami13.pdf
- Pevsner J. (2015) Bioinformatics and Functional Genomics, 3rd Edition Wiley-Blackwell
- Genetics Home Reference (2020). Help Me Understand Genetics (https://ghr.nlm.nih.gov/)