

Course unit name:
BIOINFORMATICS AND COMPUTATIONAL GENOMICS IN CANCER RESEARCH

1.- General information

Code	303011	Plan		ECTS	3
Type	Elective	Course	2023/2024	Periodicity	2 nd Semester
Department	Cancer Research Center (CiC-IBMCC, USAL/CSIC)				
Virtual Platform	Platform:	CICLOUD			
	URL de Acces:	http://cicloud.dep.usal.es/index.php/s/Gp0vghR305Y6glo/authenticate			

Faculty

Professor Coordinator	Dr. Javier De las Rivas Sanz				
Research area	Bioinformatics and Functional Genomics				
Center	Cancer Research Center				
Office	Laboratory 19				
Tutorials	Monday, Tuesdays and wednesdays from 13 to 14 pm				
URL Web	https://www.cicancer.org/grupo?id=42				
E-mail	jrivas@usal.es	Phone	+34 923294819		

2.- The course in the context of the Master's Program

Training Module
Fourth block (out of five) of master program organization.
General aim of the subject
The subject of Bioinformatics, although it is elective, it has a central role in the Master's Student Curriculum, especially in the case that students are going to perform omic data analysis or protein structure-function studies.
Professional specialization
Normally students have little knowledge of the subject, and therefore (the subject is 100% practical) with computers will provide them in their professional profile a good introduction to Bioinformatics tools and methods for using, as well as the management of Bases of Cancer Related Data.

3.- Previous recommendations

No needed.

4.- Aims of the subject

This Subject is focused on the new area of Bioinformatics and Computational Biology aiming to teach students the use of bioinformatic tools, algorithms and computational methods to analyse omic biological data derived from the use of genomic, transcriptomic and proteomic large-scale techniques. The Subject will focus in the study of data obtained mainly in cancer research: both in human clinical studies with patients, and in more basic biomolecular studies focused on certain oncogenes or anti-cancer agents, etc. In addition, special emphasis will be placed on approaches and methods of integrative and relational computational biology in order to generate and explore sets and networks of biological entities (genes, proteins, ncRNAs, mutations, drugs, targets, etc) associated to certain pathological states under study. Finally, another part of the Subject will be focus on protein structural bioinformatics, to deal with methods and algorithms developed to study and analyse protein sequences and protein structures, as well as the inference of function associated to structure.

Learn about the main biological resources and biomolecular databases:

- Genome Databases.
- Sequence Databases (genes and proteins).
- Structural Databases (proteins, nucleic acids, etc).
- Promoter / Regulatory Databases.
- Genomic and Proteomic Databases.
- Metabolism and Pathways Databases.
- Scientific Publications Databases.
- Databases to build and explore Biological Networks.
- Integrated biological resources.

Learn about the main biological resources and databases focused on Cancer:

- Genomic Data Commons (GDC) Data Portal: <https://portal.gdc.cancer.gov/>
- Cancer Gene Census - COSMIC: <https://cancer.sanger.ac.uk/census>
- cBioPortal for Cancer Genomics: <https://www.cbioportal.org/>
- Cancer Cell Map: <http://ccmi.org/>

Understand main bioinformatics tools used for the analysis of biomolecular data:

- Sequence alignment tools.
- Tools for multiple alignment and phylogenetics.
- Tools for motifs and domains finding and prediction.
- Primary, secondary and tertiary protein structure analysis and prediction.
- Protein structure visualization tools.
- Expression analysis tools (transcriptomics).
- Functional annotation and enrichment tools.
- Molecular network analysis tools

5.- Contents

Theoretical-practical lectures:

- Introduction to the Course
- Introduction to R in bioinformatics (<https://www.r-project.org/>): free software environment and programming language for statistical computing and graphics.
- Introduction to BioConductor (<https://www.bioconductor.org/>): Bioconductor provides tools for the analysis and comprehension of high-throughput biomolecular data, mainly genomic data.

Biological Databases and Resources:

- Primary databases of sequences (genes and proteins), access and file characteristics: GenBank, RefSeq, EMBL, UniProt.
- Integrated search systems for biological data: NCBI, Entrez, Gene Cards, etc.
- Homologous sequence search systems: FASTA, BLAST, Psi-BLAST, HMMer.
- Genomes (with special emphasis on human and mouse): genomic navigation in ENSEMBL and other Genome Browsers.
- Genomic and proteomic databases: GEO, Human Protein Atlas (HPA).
- Protein-protein interaction and gene-gene association databases: GeneMania, STRING, APIDinteractomes.
- Ontological and functional annotation databases: Gene Ontology, DAVID, GeneTerm Linker, GSEA.

Bioinformatics and Genomics:

- Genomic and transcriptomic, proteomic large-scale platforms: types of high density microarrays; deep sequencing methods (DNA-seq, RNA-seq, etc); proteomic methods.
- Gene expression analysis: basic concepts and parameters. Analysis of the data provided by different platforms.
- Practical search for genes of significant expression in databases: identification of isoforms, identification of orthologs and paralogs, identification of homologues.
- Computational analysis of lists of genes or proteins to search for significant biological-functional annotation: Functional Enrichment Analyses.

Bioinformatics and Proteomics:

- Analysis of protein data: sequences, motifs, domains, three-dimensional structures (UniProt, Expaty, PROSITE, Pfam, InterPro, PDB, PDBsum).
- Multiple alignment methods of protein sequences: family profiles, recognition and significance. Construction of multiple alignments with CLUSTALX.
- Molecular and structural analysis of protein families with tree development and evolutionary implications (using UGENE).
- Analysis and prediction of protein structure by bioinformatic methods: secondary structure and tertiary structure (threading).
- Methods of visualization and analysis of 3D three-dimensional structures (mainly using PyMOL, but also knowing about Jmol, SwissPDB viewer, VMD).
- Protein-Ligand & Protein-Protein Interactions visualization and analysis: PDBsum, RCSB PDB viewer, NLDB, Protein-Ligand Interaction Profiler. Docking analysis: AutoDock; PatchDock & FireDock; .
- Construction of biomolecular protein interaction networks (using Cytoscape).

Student Seminars:

Each student working in a team with one or two peers (2 maximum) will have to prepare a 30-minute presentation-seminar in which they will explain a useful resource-database or bioinformatic application tool (mainly selected from the special issues of the journal Nucleic Acids Research Database Issues or Web Server Issues: <http://nar.oxfordjournals.org/>), including a specific example of its use that shows the utilities and functions of the bioinformatics tool.

6.- Skills to be acquired

Specific skills

- Identify the contents and the way of access to the main biological resources and biomolecular databases.
- Know how to use the main bioinformatics tools for the analysis of biomolecular data of both singular genes or proteins and of groups or families of genes and proteins of interest in a study.
- Learn how to interpret and particularize the global (omic) data, present in the main computer servers, for the analysis of genes or proteins.

7.- Teaching methodology

- The majority of the course is practical because classes are held in a classroom equipped with computers (a computer for each student), with internet access and with a whole series of bioinformatics tools installed. Each student must attend all evaluable theoretical-practical sessions of the course (30 hours) having previously read and understood the recommended bibliography.
- The first session will focus on presenting the structure and parts of the course, explaining the different sessions and the organization, as well as the tasks that each student will have to develop both individually and in teams.
- Students will be organized in groups or work teams (of 2 members) to prepare a seminar with a theoretical-practical presentation of a selected work topic (chosen by the students within a series proposed by the professor or proposed by the students).
- The student must attend all the seminars (at least 2 sessions of 3 hours, 6 hours), in which other students present their work, usually based on one or several published bioinformatic articles of current research in the area.
- Each student will have access to personal tutorials for questions regarding the development of the Subject, the preparation of their seminar and other topics of interest for their curricular development.

8.- Estimated learning time

	Hours tutored by the teacher		Individual work (hours)	TOTAL HOURS
	Attendance required (hours)	Distance learning (hours)		
Lectures	15		15	30
Practices	- In classroom			
	- In laboratory			
	- In computer classroom	15		15
	- Countryside			
- Visualization classroom				
Seminars	6		9	15
Work presentations and debates				
Tutorials	3			3
Online activities	9			9
Work preparation				
Other activities				
Exams - evaluation	3			3
TOTAL	51		24	75

9.- Materials

Books
Other bibliographical, electronic references or any other type of resource

10.- Assessment

Assessments on the performance of the student
<p>-Evaluation of the presentation of the Seminar prepared by each student (40% of the final grade).</p> <p>-Final Evaluation, depending on the number of students: (i) written exam with the responses corresponding to the resolution of some bioinformatics practical problems; (ii) presentation of a written report (max. 20 pages) including a full bioinformatic analysis of a selected omic dataset or a full structural-functional analysis of a query protein or family of proteins (60% of the final grade).</p>
Recommendations