UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Bioinformatics

2425-1-F0901D043-F0901D087M

Aims

Students will learn the basics of bioinformatics required for their profession.

Contents

This course introduces bioinformatics applied to medical biotechnology. Topics include sequence alignment, phylogenetics, analysis of gene expression data, and omics integration. Students will learn the basics of the R language for data analysis, applying it to case studies on gene expression and cancer bioinformatics. The course also covers machine learning and evolutionary cancer models, including mutational signature analysis. Each module includes practical exercises to strengthen skills.

Detailed program

The Bioinformatics course aims to provide a solid overview of the main techniques and tools used in bioinformatics, with a particular focus on applications in the medical and biotechnological fields. Students will gain practical skills in analyzing biological data, such as genetic sequences, gene expression, genomic variants, and will learn how to use machine learning techniques for applications in oncology and genomics. The course will be structured into three parts, each consisting of four lessons. The topics covered in each part are outlined in detail below.

Part 1: Introduction to Bioinformatics and Sequence Analysis
In the first part, students will be introduced to the fundamentals of bioinformatics, exploring its history, applications, and the main resources used in the field. Bioinformatics is a crucial discipline for analyzing biological data, which includes managing and interpreting vast amounts of data from scientific experiments. The course will discuss how biological databases such as NCBI, Ensembl, and the UCSC Genome Browser are used to collect and organize

data, and students will learn how to use software and analysis tools to address bioinformatics problems.
The second lesson will focus on sequence alignment, a central technique for comparing genetic sequences, used to identify similarities between genes of different species. Students will explore various alignment algorithms, such as those used in BLAST and FASTA, to compare sequences and analyze their evolutionary relationships. Phylogenetic analysis will also be covered, which allows for reconstructing the evolutionary history of species through the analysis of homologous sequences. In this lesson, students will also learn how to build phylogenetic trees by applying maximum parsimony and maximum likelihood methods.
The third lesson of this module will address gene expression data analysis, a critical component for understanding how genes function under different biological conditions. Students will explore analysis techniques such as microarrays and RNA-Seq, emphasizing how expression data can be used to understand gene regulation and identify genes involved in diseases. In addition, genomic variants such as SNPs, indels, and CNVs, and their impact on biological functions, will be discussed. The module will conclude with an introduction to integrative omics analysis, which combines data from different omic techniques, such as genomics, transcriptomics, and proteomics, to gain a more comprehensive view of biological systems.
The last lesson of this module will be dedicated to a practical exercise, where students will have the opportunity to evaluate sequence alignment techniques, phylogenetic analysis, and gene expression data analysis on real datasets, consolidating the theoretical knowledge gained in the previous lessons.

Part 2: Fundamentals of Programming and Gene Expression Data Analysis
In the second part, students will begin to familiarize themselves with the R programming language, one of the most powerful tools for statistical analysis of biological data. The first two lessons will be dedicated to teaching the fundamentals of R, starting from the basics of syntax and data management to advanced data manipulation techniques. Students will learn how to write and execute scripts in R, manage variables, use functions and libraries, and visualize data through graphs and tables. They will also be introduced to statistical data analysis, with particular focus on common techniques such as regression analysis and statistical tests.
The third lesson of the module will focus on a practical case study related to gene expression data analysis. Students will work with real data from RNA-Seq or microarray experiments and will learn how to perform data preprocessing, including normalization and filtering steps, to prepare the data for analysis. They will then focus on identifying differentially expressed genes and performing multivariate data analysis, using techniques such as PCA (Principal Component Analysis) and clustering to uncover hidden patterns in the data.
The module will conclude with a practical exercise where the techniques learned in the lessons will be discussed and applied.

Part 3: Machine Learning, Cancer, and Mutations
In the third and final part, students will be introduced to machine learning, a discipline that is playing a central role in modern bioinformatics. They will learn the basics of machine learning, distinguishing between supervised and unsupervised techniques, and how to apply algorithms such as decision trees, support vector machines (SVM), and k-NN to analyze and make predictions on biological data. The lesson will also focus on the use of machine learning to address complex bioinformatics problems, such as classifying genomic data and identifying patterns in omic data.
The next lesson will explore cancer evolution models, analyzing how tumor cells accumulate mutations over time and how these mutations influence cancer behavior and progression. Students will learn about cancer evolutionary models and how genomic analyses can be used to reconstruct the evolutionary trajectories of tumor cells, using computational tools to study the distribution of somatic mutations.
The penultimate lesson will focus on mutational signatures, a fundamental concept for understanding the causes and progression of cancer. Mutational signatures are characteristic patterns of mutations that occur in specific types of cancer, and analyzing these signatures is crucial for personalizing cancer treatments. Students will learn to identify and analyze these signatures using advanced bioinformatics and machine learning techniques.
The final meeting of the course will be a practical exercise that will allow students to apply what they have learned regarding genomic mutation analysis and mutational signature analysis.

Prerequisites

Basics biological and mathematical knowledge.

Teaching form

Frontal lessons.

Textbook and teaching resource

Slides and teaching materials provided by the instructor.

Semester

First semester.

Assessment method

Exam mode: Written test.
Question type: Multiple-choice quiz with 5 options, one correct answer, for a total of 16 questions.
Grading results: The score is expressed on a scale of thirty, with a minimum score of 18/30 and a maximum score of 30/30 with honors.
Topics covered in the questions: All topics covered in the lessons may be asked.
Exam duration: The student has 45 minutes to complete the exam.

Office hours

By appointment requested via e-mail.

Sustainable Development Goals

QUALITY EDUCATION