UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Bioinformatics

2526-1-F0602Q114

Aims

The course is focused on genomics technologies, on how they are used to generate omics data, and on the fundamental concepts and pitfalls in the analysis of these data. The course focuses on technologies relying on high-throughput sequencing approaches. Particular emphasis will be devoted to the study of genome-wide gene expression programs, DNA-binding factors and epigenetic modifications.

Knowledge and understanding.
Students will familiarize with the key genomic technologies and data types. They will learn good practise and key workflows in the field of genomics. Interactive lessons will focus on Galaxy, a suite of tools freely available online for the analysis of omics data. Eventually, the students will acquire key concepts in the field of bioinformatics and genomics, and they will familiarize with selected resources to which they shall reference in case of need during their future research activities.

Applying knowledge and understanding.
Students will be able to apply the acquired knowledge in other courses and in laboratory experiences and to use the comprehension skill in subsequent studies and/or research activities.

Making judgments.
Students will be able to process what they have learned and be able to recognize situations and problems in which the acquired knowledge can be used.

Communication skills.
At the end of the course, students will be able to express themselves appropriately in the description of issues related to genomics and bioinformatics by using language properties, as well as self-confidence.

Learning skills
At the end of the course, students will be able to survey the scientific literature related to the topics covered and will be able to analyse, apply and integrate the acquired knowledge with what learned in related courses.

Contents

1. Gene expression and its determinants
2. From cells to a sequencing library
3. High-throughput sequencing technologies
4. Processing sequencing data, common steps
5. ChIP-seq analysis workflow
6. RNA-seq analysis workflow
7. Genome Browsers
8. Clustering and heatmaps
9. Introduction to Galaxy
10. ChIP-seq data analysis with Galaxy
11. RNA-seq data analysis with Galaxy
12. Dynamics of transcriptional regulation
13. Single cells transcriptional profiling
14. Chromatin conformation

Detailed program

1. Gene expression and its determinants: (i) gene expression programs, (ii) Binding of regulatory factors to chromatin, (iii) Epigenetic marks, (iv) RNA polymerases, (v) RNA modifications
2. From cells to a sequencing library: ChIP-seq, RIP-seq and RNA-seq
3. High-throughput sequencing technologies: (i) Short reads sequencing (Ion Torrent, Illumina), (ii) Long reads sequencing (Pacific Biosciences, Oxford Nanopore Technologies)
4. Processing sequencing data, common steps: (i) FASTQ files, (ii) quality checks, (iii) reads alignment, (iv) SAM/BAM files, (v) inspecting alignments with a genome browser
5. ChIP-seq analysis workflow: (i) peaks calling, (ii) false positives and false negatives, (iii) motif enrichment, (iv) GeneOntology and gene sets enrichment, (v) differential peaks calling
6. RNA-seq analysis workflow: (i) absolute expression and normalization, (ii) differential expression, (iii) splicing detection, (iv) RNA modifications
7. Genome Browsers: IGV/IGB, UCSC Genome Browser, data upload, download pics, playing with ENCODE tracks, UCSC Table
8. Clustering and heatmaps
9. Introduction to Galaxy: From peaks to genes
10. ChIP-seq data analysis with Galaxy
11. RNA-seq data analysis with Galaxy
12. Dynamics of transcriptional regulation: (i) Nascent RNA, (ii) Post-transcriptional regulation, (iii) RNA polymerase life cycle
13. Single cells transcriptional profiling, scRNA-seq
14. Chromatin conformation

Prerequisites

Background: basic notions of Molecular Biology.
Prerequisites: none

Teaching form

- 12 2-hour delivery-focused frontal teaching lectures
- 9 2-hour frontal teaching lectures conducted in lecture mode in the initial part, which is aimed at engaging students interactively in the subsequent part. The lectures conducted in interactive mode involve the use of the Galaxy data analysis system. This part includes optional individual or group project work by the students on analyses carried out independently with Galaxy. The project work concludes with a short presentation by the students and constitutes a partial examination test.

Textbook and teaching resource

Selected scientific papers, slides, and recording of the lessons will be available at the e-learning platform of the course.
Recommended textbooks:

- M.H. Citterich et al. “Fondamenti di Bioinformatica” Zanichelli

Semester

Second semester

Assessment method

Colloquium on the topics covered in the lecture. The discussion may, at the student's choice, include the discussion of a scientific article relevant to the topics covered during the course, assessing the ability and critical analysis to interpret the results.
The assessment integrates the result of the partial examination test, for those students who have chosen to take it, an optional project work aimed at checking disciplinary problem-solving skills.

Office hours

By appointment through e-mail request to the lecturer.

Sustainable Development Goals

GOOD HEALTH AND WELL-BEING