



ROLLINS
SCHOOL OF
PUBLIC
HEALTH
EMORY

DEPARTMENT: Biostatistics and Bioinformatics

COURSE NUMBER: 540 SECTION NUMBER: 0

SEMESTER:

CREDIT HOURS: 2

COURSE TITLE: Introduction to Bioinformatics

INSTRUCTOR NAME: Zhaohui "Steve" Qin

INSTRUCTOR CONTACT INFORMATION:

EMAIL: zhaohui.qin@emory.edu

COURSE DESCRIPTION (3-4 Sentences)

This course is an introduction to the field of Bioinformatics for students with a quantitative background. The course covers biological sequence analysis, introductions to genomics, transcriptomics, proteomics and metabolomics, as well as some basic data analysis methods associated with the high-throughput data. In addition, the course introduces concepts such as curse of dimensionality, multiple testing and false discovery rate, and basic concepts of networks.

Prerequisites: Bios 500, 501, 506, or permission of instructor

EVALUATION

The grade assignment will be based on:

Participation in class and discussions (10%);

Three homeworks (10% each);

Final presentation (60%). The instructor will assign a published research paper. The presentation should include three parts: (A) a detailed study on the literature cited in the paper, (B) the methods and results of the paper, and (C) more recent developments in the field the paper addresses.

ACADEMIC HONOR CODE

The RSPH requires that all material submitted by a student in fulfilling his or her academic course of study must be the original work of the student.

LEARNING OBJECTIVES OR COMPETENCIES OF THE COURSE

After taking the course, the students should have a general understanding of the data generation and statistical challenges in genomics, transcriptomics, proteomics and metabolomics. Specifically, they should have basic working knowledge on biological sequence alignment, Hidden Markov Model and its estimations, motif discovery, microarray data pre-processing, multiple testing, deep sequencing data structure, pre-processing of mass spectrometry-based proteomic/metabolomic data, and biological networks.

Recommended readings:

Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids
Bioinformatics and Computational Biology Solutions Using R and Bioconductor
Introduction to Computational Biology – Maps, Sequences & Genomes

Schedule:

Lecture 1, Introduction
Lecture 2, Sequencing; Dynamic programming sequence alignment
Lecture 3, BLAST; Hidden Markov Models in sequence alignment (1)
Lecture 4, Hidden Markov Models (2)
Lecture 5, Motif Discovery; Phylogeny
Lecture 6, Gene expression array
Lecture 7, Brief introduction to supervised learning and unsupervised learning
Lecture 8, Feature selection and Multiple Testing
Lecture 9, Array-based analysis of the DNA
Lecture 10, High-throughput sequencing
Lecture 11, MS-based Proteomics & Metabolomics
Lecture 12, Biological networks

LEARNING OBJECTIVES OR COMPETENCIES FOR THE DEPARTMENT OR PROGRAM TO WHICH THE COURSE CONTRIBUTES

The course offers biostatistics MS/PhD students basic knowledge in Bioinformatics. It serves as the first course for students who are interested in pursuing research in this area. It is the basis for more advanced courses in Bioinformatics, including Bioinformatic Machine Learning and other future courses in this area.