FUNDAMENTAL BIOINFORMATICS

Sessions: One 2 hours and 45-minute lecture class per week: Tuesday 3:30-6:15 pm
One 1 hour recitation Thursday 5:00 – 6:00pm

Location: Computer laboratory classroom for recitation required


Instructor: Professor Manpreet S. Katari PhD

Course Aims: The course will introduce students to the field of Bioinformatics, which develops and applies methods to analyze large-scale biological data sets. They will learn about the computational and statistical theories that support fundamental concepts in Bioinformatics, and be able to apply them to high-throughput genomic data using a suite of different applications.

Prerequisites: Principles of Biology I
Principles of Biology II
Molecular and Cell Biology I

Grading:

6 homework assignments: 30%
Midterm Exam 1: 20%
Midterm Exam 2: 20%
Final Exam: 30%

Course Description:

The course has a lecture and a recitation component. During the recitation the students will use popular software packages to analyze publicly available data from genome projects and high-throughput experiments. The course is divided into three sections:
1) Genomics – students will learn about different DNA-sequencing technologies and the challenges behind analyzing the sequences. Students will learn the concepts behind aligning sequences and predicting where genes reside in sequences.
2) Structural Bioinformatics – students will learn the concepts behind predicting RNA and protein structures.
3) Functional Genomics – students will learn about microarray analysis and how it can be used to identify changes in genome structure and gene activity.

Assignments and Projects:

A total of 6 homeworks will be assigned (2 in each section described above) during recitation and will be due the following recitation. Each homework will be worth 5% of the entire grade. Two midterms will be given in class, each worth 20% followed by a cumulative final, which will be worth 30%. Late assignments will be penalized 10% for each day it is late. The midterm and final exams will be composed of short answer questions.
Course Syllabus

Part I: Genomes

Week 1: Definition of Bioinformatics and Biological databases.

   Reading: Chapters 1 and 2.

   Recitation 1: Tour of the NCBI, Ensemble and UCSC browsers.

   Homework 1: Questions related to usage of OMIM (Online Mendelian Inheritance in Man) and Identify disease associated with specific region of the human genome.

Week 2: Sequencing Technologies and Mapping and Assembling Genomes

   Reading: Chapter 17

   Recitation 2: Video demos of different Sequencing technologies.

Week 3: Gene and Promoter Prediction

   Reading: Chapters 8 and 9

   Recitation 3: Predict genes using Genscan and ORF finder

   Homework 2: Given a 200,000 bp sequence, identify all the protein coding genes in the sequence.

Week 4: Pairwise Sequence alignment (BLAST)

   Reading: Chapters 3 and 4

   Recitation 4: Use the different variants of Blast on NCBI’s BLAST to find sequences homologous to one of interest. Statistical relevance of e-value will be discussed.

Week 5: Midterm 1

   Recitation 5: A tour of NYU Center for Genomics and Systems Biology sequencing facility or Guest Speaker

Week 6: Multiple Sequence Alignment and Protein Motif and Domain predictions

   Reading: Chapters 5,6,and 7
Recitation 6: Using ClustalW and MUSCLE to perform multiple sequence alignments. Use Interproscan to identify domains.

Homework 3: Identify protein domains in the genes that were predicted in homework 2.

Week 7: Secondary and Tertiary Protein Structure prediction

Reading: Chapters 12,13,14,15

Recitation 7: Use CN3D to visualize 3d structure of protein.

Week 8: RNA and miRNA structure and prediction

Reading: Chapter 16

Recitation 8: Use tRNASCAN to find RNAs.

Homework 4: Identify and annotate RNA transcripts in your sequence from HW2.

Week 9: Basic Phylogenetics and tree building

Reading: Chapters 10 and 11

Recitation 9: Use PAUP to perform phylogenomic analysis of evolutionary relationships

Week 10: Midterm 2

Recitation 10: Introduction to MeV software for microarray analysis

Week 11: Introduction to Microarrays and determining differentially expressed genes

Reading: Chapter 18

Recitation 11: Load a microarray dataset and normalize the data. Identify genes that are differentially expressed. Statistical methods Students t-test and correction for multiple hypothesis testing will be reviewed.

Homework 5: Load a microarray dataset from public repository. Normalize it and identify genes that are differentially expressed.

Week 12: Clustering and GeneSet Enrichment.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis

**Recitation 12:** Perform clustering and GO-enrichment analysis. Hypergeometric distributions will be used to determine whether a GO-term is enriched.

**Homework 6:** Using the dataset provided in HW5, perform clustering and identify enriched GO-terms in each cluster.

**Week 13: RNA-seq and CHIP-seq**


**Recitation 13:** Perform RNA-seq analysis using MeV. Poisson distributions and their use for determining regions that are differentially represented by sequencing reads will be discussed.

**Week 14: Proteomics and Protein Interactions**

**Reading:** Chapter 19

**Recitation 14:** Use String Database to visualize protein interaction data.