Research methodologies in bioinformatics - 225.4408 (223.3020)

Lectures – Tuesdays, 12:15-14:00.

Exercises – Tuesdays, 14:15-16:00.

Lecturers:
Dr. Eyal Privman   Phone: 054-5224888   Email: eprivman@univ.haifa.ac.il
Prof. Mickey Kosloff   Phone: 04-8288996   Email: kosloff@sci.haifa.ac.il
Dr. Martin Mikl   Phone: 04-8240751   Email: MiklLabHaifa@gmail.com

Teaching assistants:
Anna Bakhman   Email: annabakhman@gmail.com
Yoann Pellen   Email: pellen.yoann@gmail.com

Course credit: 4

Prior requirements:
(Biochemistry A+B, molecular biology, genetics)

Requirements:
70% final exam (or project) and 30% written tutorials. The exam can be substituted with a final project, **if** approved in writing by lecturers no later than 2 weeks before the end of the semester.

Attendance:
80% mandatory (both in lessons and excercises). If the lesson/exercise is in Zoom, students are required to have their cameras on *throughout the lesson* – failing to do so will be considered an absence.

Course program:
Students will learn basic concepts, approaches and tools used for bioinformatic analysis of sequence data (DNA and protein) and structural data (proteins). We will discuss the design and analysis of prokaryotic and eukaryotic sequences, structures and genomes, as well as the environmental and evolutionary forces shaping them. Methodological topics covered include: genomic architecture, gene identification characterization and annotation, sequence alignment, phylogeny reconstruction, protein 3D structure evolution visualization and comparison, structure prediction, quantitative methods for protein structure analysis, genome/transcriptome sequencing and assembly methods, and gene/protein networks. The course will include lectures and hands-on computer tutorials.

Final grades:
70% final exam (or project) and 30% written tutorials (the exam can be substituted with a final project, but only if approved by lecturers in writing no later than 2 weeks before the end of the semester)

Note:
Students that have official letters stating they need academic adjustments are welcome to contact us by email as early in the semester as possible.

Research methodologies in bioinformatics - Course outline

1) **Introduction – genes, proteins, and bioinformatics** (25/10/22 – **Eyal**)
   a. Genetic code(s)
   b. The central paradigm: DNA-RNA-protein
   c. The building blocks of genes and proteins.
   d. Non-coding RNAs, non-coding regions, codon ratios and RNA stability
   e. What is bioinformatics and a historical perspective

Elections (1/11/22)

2) **From gene to protein** (8/11/22 – **Eyal**)
   a. Does a gene determine a protein? What determines function?
   b. Identification, characterization and annotation of genes and proteins
   c. Gene and protein databases.
   d. Protein motifs.
   e. **Tutorial: sequence databases** (NCBI, Uniprot etc.)

3) **Pairwise sequence alignment** (15/11/22 - **Eyal**)
   a. Pairwise sequence alignment.
   b. The various flavors of BLAST.
   c. Sequence motif searches.
   d. **Tutorial: ncbi BLAST, Prosite.**

4) **Multiple sequence alignment** (22/11/22 - **Eyal**)
   a. Sequence alignment – the basis of evolutionary analysis
   b. Progressive multiple sequence alignment
   c. Phylogeny-aware alignment (PRANK/PAGAN)
   d. **Tutorial**: **Multiple sequence alignment and visualization (MAFFT; Jalview, PAGAN)**

5) **Alignment and phylogeny** (29/11/22 - **Eyal**)
   a. Nothing in sequence analysis makes sense except in the light of evolutionary trees
   b. Distance-based phylogenetic methods
   c. Bootstrap confidence scores
   d. Likelihood-based methods
   e. Rogue taxa
   f. **Tutorial: Multiple alignment continued (PAGAN) and alignment filtering (Guidance)**

6) **Introduction to Protein structure** (6/12/22 - **Mickey**)
   a. 3D structures of proteins – introduction and biophysical basis, from primary to quaternary structures.
   b. Protein folding.
   c. 3D structure determination and visualization.
   d. **Tutorial: Protein sequence and structure, PDB (PDBe), 3D structure visualization online.**

7) **Proteins – from structure space to subdomains** (13/12/22 - **Mickey**)
   a. Structural domains.
   b. Structure/sequence/function "spaces".
   c. Structural similarity, comparisons and alignment.
   d. Structural effects of mutations.
   e. Sequence-level domain and motif identification, non-globular motifs (IUPs, fibrous proteins, transmembrane domains, coiled coils, Interpro).
   f. **Tutorial: 3D structure and mutation visualization, structure comparison.**

8) **Protein structure prediction** (20/12/2022 - **Mickey**)
   a. Secondary structure prediction.
   b. 3D Structure prediction.
   c. Homology modeling
   d. Deep learning based predictions (Alphafold2 etc.)
   e. **Tutorial: structure prediction**

9) **What makes a genome and how to sequence it** (27/12/2022 - **Martin**)
   a. Genome types: linear, circular, plasmids, viruses
   b. Origins, operons, rRNA
   c. "Junk DNA" and repetitive elements
   d. Sequencing methods – their advantages and disadvantages: Sanger, next-generation sequencing (Illumina), Nanopore, PacBio
   e. Assembling a genome
   f. **Tutorial: Genome browsers/Mapping sequences**

10) **Eukaryotic and comparative genomics** (3/1/2023 - **Martin**)
   a. What makes a gene?
   b. Introns and exons
   c. Gene structure annotation
   d. Genome evolution
   e. Homologs, Paralogs and Orthologs
   f. Core genome and pan genome
   g. Comparative genomics on individuals
   h. **Tutorial:** TBD

11) **Transcriptomics** (10/1/2023 - Martin)
   a. Mapping transcriptomes to genomes
   b. De-novo transcriptome assembly

    c. P value vs Q value
    d. Differential expression analysis
    e. Functional categories (GO, COG, KEGG)
    f. Enrichment analysis
    g. **Tutorial:** Transcriptome analysis - functional enrichment

12) Functional genomics (17/1/2023) - Martin)
    a. Sequences regulating transcription, splicing, translation,…
    b. Regulation in cis and in trans
    c. How to find and study gene regulatory regions
    d. Predicting gene expression
    e. Non-coding genetic variants – a key to understanding diseases
    f. **Tutorial:** Identify regulatory sequence elements and predict their activity