

BE 561 - Fall 2009: Syllabus  
(order and topics may change or evolve)

**Lab Sessions Topics (taught by TA)**

- 2 Introduction to Probability (1)
- 2-1 Introduction to Probability (2)
- 2-2 Problem Solving for Probability
- 2-3 Hypothesis Testing
- 2-4 Databases and Tools at NCBI, UCSC, TRANSFAC, PDB
- 3 Perl programming

First lecture: review biological problems and sequence analysis

- 5-1 Introduction to Bioinformatics and Comp. Biology
  - \*The key biology problems (\* covered in course or homeworks)
  - \*Genome Assembly Light (\*as application of multiple alignment & greedy algs).
  - \* Gene Identification (Glimmer \*)
    - Comparative Gene Identification (\*Mummer)
  - \*Identification of functional elements in genomes (CIS elements, microRNAs, repeats)
  - \*Assigning Genes with Function
    - Prob. Prediction using sequence and Naive Bayes
  - \*Protein-Protein Interaction (Networks)
    - Prediction using sequence and Naive Bayes
  - \*Protein - DNA Interaction (TRACTOR ?)
  - \*Pathways and Networks
    - Mapping multiple modalities into networks e.g Zheng et al
  - \*Probabilistic Functional Linkage Networks
    - e.g. Letovsky / Kasif
  - \*Reverse Engineering of Pathways
    - (simple algs if time permits).

**Regular Lectures**

- 5-2 Genome Features:
  - Read and review in lab two genome papers
  - Lander et al
  - Bacterial Genome
- 6 Probabilistic Models / Predictive systems/
  - Markov Models
- 7 Amino Acid Scoring Matrices: BLOSUM and PAM
- 8 Pairwise Sequence Alignment
- 9 Global Dynamic Programming with Affine Gaps, BLAST

- 10 BLAST Statistics (may be dropped )
- 11 Multiple alignment, PSI  
BLAST  
Positional Specific Sequence Profiles
- 11-1 HMMs: PFAM
- 11-2 Blocks
- 12 Phylogenetic Tree Construction:  
UPGMA, Neighbor Joining and Maximum Likelihood
- 13 Statistical Evaluation of Algorithm Performance  
(ROC curves, AUC etc) (may be  
lab topic)

**Midterm Review (Lab Session)**

**Midterm Exam (move to appropriate time slot)**

- 14 Comparative Genomics and Phylogenetic Footprinting
- 15 Supervised & Unsupervised Learning Methodologies
- 15-1 Naive Bayes application to Gene Finding and Protein Analysis
- 15-2 Bayes Networks and applications to Splicing
- 16-1 Markov Models (5th and IMM) and Gene Finding  
Genemark (Borodovsky)  
Glimmer (Delcher, Kasif, Salzberg)
- 16-2 Hidden Markov Models (HMM) and Gene Finding  
Krogh / Haussler  
Genmark (ref not covered)
- 17 Posterior Decoding and Learning of HMM
- 17-1 Problem Solving for Markov Models and HMM  
(comparative gene finding using pair-hmms)
- 18 Protein - DNA prediction  
In-Silico Prediction of Cis Elements  
Fritz/Weng, TRACTOR/Schaffer et al
- 18-2 Multiple Local Alignment: Local Search,  
Simulated Annealing and Gibbs Sampling (applications to multiple  
alignment) and protein-DNA motif discovery
- 18-3 Conserved Sequence Motif and Sequence Analysis  
Mummer -- large scale alignments

Conservation of binding sites in promoters -- small scale alignments

**Thanksgiving Break: No Class**

- 19 Algorithms Performance and Evaluation (speed)
- 20 Review : Pathways and Complexes: Protein - Protein Interaction and Protein - DNA Interactions Scale Free Networks Brandon Xia/Gerstein Explanation
- 21 Protein-Protein Interactions Prediction (sequence based and Naive Bayes)
- 22 Mapping Information into Networks (Zheng, Kasif, Roberts, etc)
- 23 Probabilistic Functional Linkage Graphs (Letovsky / Kasif)
- 24 Prob. Evidence Integration for Prediction (application to microRNAs)

**Dec ? 09 Exam Review**

**Dec ? 09 FINAL**

**Grading: 30% MidTerm, 30% Final, 30% Homeworks, 10% Level of Activity**