**BIOINF 184/BENG 184/BIMM 184/CHEM 184/CSE 184:**
Computational Molecular Biology
UCSD Winter Quarter 2006

**Instructor:** Professor Wei Wang, wei-wang@ucsd.edu, 4254 Urey Hall
Office Hour: 11:50 am - 12:50 pm MWF

**Teaching Assistant:** Robert Shoemaker, rfshoema@ucsd.edu, 4249 Urey Hall
Office Hour: TBA

**Description:**
This course will introduce the challenging questions in the modern molecular biology, genomic and proteomic technologies as well as computational methods that have been developed to address these questions in the post-genome era. This course will survey cutting-edge research in the contemporary bioinformatics and systems biology at the introductory level and aims at exposing students to the up-to-date research topics. It emphasizes the combination of biological insights and statistical techniques to solve real research problems.

The topics include inference of transcriptional regulatory network, which covers gene microarray and ChIP-chip analysis, identification of regulatory motifs, combinatorial regulation of transcription factors and network topology analysis, and protein interaction analysis, which covers protein folding, protein structure prediction, protein motif/domain identification, protein structure classification, and prediction of protein-protein interactions. The computational algorithms that will be introduced in the course include, for example, hierarchical clustering, k-means, support vector machine, expectation maximization and naïve Bayes net. The course consists of core lectures, homework assignments, programming projects with midterm and final exams.

**Prerequisites:**
BIOINF 181 and 182 or basic understanding of molecular biology and familiar with one of the following programming languages, Perl, Java, Python and C.

**Grading:**
The course will be graded by performance on four homework assignments (approximately 40%), two programming projects (approximately 40%), midterm and final exams (approximately 20%).

**Late Policy:**
20% of the score will be taken off for each additional day after the due date.

**Partner Policy:**
Students may discuss ideas with others. However, programs are to be completed independently and should be original work. Code may not be shared. Names of students with whom programming ideas were discussed should be included with assignment.
Readings:
Optional readings will be suggested throughout the course (check the course website for updates).
No required textbook.

Recommended books:

Schedule

**Introduction and overview (1-2 lecture)**
Central dogma of biology and challenging problems in the post-genome era
High-throughput technologies: microarray, ChIP-chip, yeast two hybrid, etc.
Introduction to computational molecular biology and systems biology

**Identification of regulatory elements (4-5 lectures)**
Combination of gene expression and motif search
Single genome statistic analysis
Comparative genomics analysis
Methods based on clustering of DNA motifs

**Gene microarray analysis (3 – 4 lectures)**
Clustering algorithms: hierarchical clustering and k-means
Classification algorithms: nearest neighbor and SVM
Combination of sequence and gene expression

**Inference of combinatorial regulation and network topology (5-6 lectures)**
Review of combinatorial regulation in biology
Network realization upon conditions
Network topology statistics
Network motif identification

**Protein folding and structure prediction (~3 lectures)**
Protein folding
Basic concept of protein structure
Computational modeling: from lattice model to all-atom model
Structure prediction
Introduction to force field, molecular dynamics and Monte Carlo
Homology modeling
Protein structure alignment and motif identification (~3 lectures)
  Protein structure database and classification: PDB, SCOP, CATH
  Protein structure alignment
  Protein motif identification: pfam, BLOCK, PRINT etc.

Protein specificity (3-4 lectures)
  Calculation of binding free energy
  Potential of mean force
  High-throughput technologies
  Combination of noisy high-throughput data

Protein interaction networks (5-6 lectures)
  Protein network topology analysis
  Network motif identification
  Future directions and several challenging problems