CH364C/CH391L Bioinformatics (course # 52716/52950)

Spring Semester, 2013

Instructor: Edward Marcotte		marcotte@icmb.utexas.edu		
Office: MBB 3.148BA		Phone: 471-5435		
Course lectures:	**T/Th**	**2:00 – 3:30 PM**	**WEL 3.260**	
Office hours:	Wednesdays	2:00 – 3:00 PM	MBB 3. 148BA	
TA: John Woods	john.woods@marcottelab.org			
TA Office hours:	Tuesday/Friday 10:00 – 11:00 AM		MBB 3.128	TA Phone: 232-3919
Course web page:	**http://www.marcottelab.org/index.php/CH391L_2013**			

Open to graduate students and upper division undergraduates in natural sciences and engineering.
Prerequisites: Basic familiarity with molecular biology, statistics & computing, but realistically, it is expected that students will have extremely varied backgrounds.

An introduction to computational biology and bioinformatics. The course covers typical data, data analysis, and algorithms encountered in computational biology. Topics will include introductory probability and statistics, basics of programming, protein and nucleic acid sequence analysis, genome sequencing and assembly, synthetic biology, analysis of large-scale gene expression data, data clustering, biological pattern recognition, and biological networks.

** Note that this is not a course on practical sequence analysis or using web-based tools. Although we will use a number of these to help illustrate points, the focus of the course will be on learning the underlying algorithms and exploratory data analyses and their applications. **

Most of the lectures will be from research articles and handouts, with some material from the...

Recommended text *Biological sequence analysis*
(for sequence analysis): R. Durbin, S. Eddy, A. Krogh, G. Mitchison
Cambridge University Press
Avail. from Amazon.com ($38.83)

For non-molecular biologists, I highly recommend (really!) *The Cartoon Guide to Genetics* (Gonick/Wheelis)
For biologists rusty on their stats, *The Cartoon Guide to Statistics* (Gonick/Smith) is also very good.

Some online references:

Online bioinformatics course: http://lectures.molgen.mpg.de/
Bioinformatics resources on the web: http://zlab.bu.edu/zlab/links.shtml
Python programming for beginners: http://www.codecademy.com/tracks/python
Online probability texts: http://omega.albany.edu:8008/JaynesBook.html
http://www-users.york.ac.uk/~mb55/pubs/pbstnote.htm
http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/pdf.html

No exams will be given. Grades will be based on 4 problem sets (given every 2 weeks and counting 15% each towards the final grade) **and a course project** (40% of final grade), which can be individual or collaborative. If collaborative, cross-discipline collaborations are encouraged. The course project will consist of a research paper or project on a bioinformatics topic chosen by the student (with approval by the instructor) containing an element of independent computational biology research (e.g. calculation, programming, database analysis, etc.). This will be turned in as a link to a web page.

The final project is due on April 30, 2013.

Outline for CH364C/CH391L Bioinformatics

1/15 Introduction + next-generation genome sequencing sample collection
1/17 A biology primer for non-molecular biologists (DNA, RNA & proteins)
1/22 A Python programming primer for non-programmers
1/24 Start biological sequence analysis
1/29 NO CLASS
1/31 Biological sequence analysis
1/29, 4/9 NO CLASS (also no class on Spring Break, 3/12, 3/14)

We'll cover the following topics, roughly in this order:

BIOLOGICAL SEQUENCE ANALYSIS

Substitution matrices (BLOSSUM, PAM) & sequence alignment
Protein and nucleic acid sequence alignments, dynamic programming
Sequence profiles
BLAST! (the algorithm)
Biological databases
Markov processes and Hidden Markov Models
Gene finding algorithms

GENOMES, DNA SEQUENCING, & "BIG BIOLOGY"

Gene finding algorithms & GASP
An introduction to genome sequences & shotgun sequencing
Genome assembly & how the human genome was sequenced
Next- (& next-next-) generation DNA sequencing & the revolution in genomics
An introduction to large gene expression data sets
Clustering algorithms, hierarchical, k-means, self-organizing maps, force-directed maps
Classifiers, k-nearest neighbors, Mahalonobis distance
Promoter and motif finding, Gibbs sampling
Principal component analysis and data transformations

BIOLOGICAL NETWORKS & SYNTHETIC BIOLOGY

Biological networks: metabolic, signaling, graphs, regulatory
Properties of biological networks
Network alignment and comparisons, network organization
Analogies between biological networks and electrical circuits
Designing, simulating, and building gene circuits
Synthetic genome engineering

Presentation of selected projects

***** FINAL PROJECT DUE on April 30, 2013 *****