

**BCH364C/BCH391L Bioinformatics** (course # 54995/55095)  
**Spring 20145 Tues/Thurs 11 – 12:30 PM BUR 212**

Instructor: Prof. Edward Marcotte      marcotte@icmb.utexas.edu  
Office hours: Wed 11 AM – 12 Noon      MBB 3. 148BA

TA: Joe Taft      taft@utexas.edu  
Office hours: Mon/Fri 11 AM – 12 Noon      MBB 3.310      TA Phone: 920-475-7784  
Course web page:      [http://www.marcottelab.org/index.php/BCH391L\\_2015](http://www.marcottelab.org/index.php/BCH391L_2015)

Open to graduate students and upper division undergrads (with permission) 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 systems biology and bioinformatics**, emphasizing quantitative analysis of high-throughput biological data, and covering typical data, data analysis, and computer algorithms. Topics will include introductory probability and statistics, basics of Python programming, protein and nucleic acid sequence analysis, genome sequencing and assembly, proteomics, synthetic biology, analysis of large-scale gene expression data, data clustering, biological pattern recognition, and gene and protein networks.

\*\* Note that this is not a course on practical sequence analysis or using web-based tools (although we'll use a few), but rather on algorithms, exploratory data analyses and their applications in high-throughput biology. \*\*

Most of the lectures will be from research articles and slides. For sequence analysis, there will be an  
**Optional text:** *Biological sequence analysis*, Durbin, Eddy, Krogh, Mitchison, Cambridge Univ. Press  
(available from Amazon, used from \$15.00)

For biologists rusty on their stats, *The Cartoon Guide to Statistics* (Gonick/Smith) is very good (really!).

We will also be learning some Python programming. The following is highly recommended:

Python programming for beginners: <http://www.codecademy.com/tracks/python>

Online homework will be assigned and evaluated using the free bioinformatics web resource Rosalind (<http://rosalind.info/faq/>). Enroll here: <http://rosalind.info/classes/enroll/9c8379d905/>

**No exams will be given. Grades will be based on online homework** (counting 30% of the grade), **3 problem sets** (given every 2-3 weeks and counting 15% each towards the final grade) **and a course project** (25% of final grade), which can be collaborative (1-3 students/project). The course project will consist of a research 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 by midnight, May 4, 2015.**

All projects and homework will be turned in electronically and time-stamped. No makeup work will be given. Instead, all students have 5 days of free “late time” (for the entire semester, NOT per project, and counting weekends/holidays). For projects turned in late, days will be deducted from the 5 day total (or what remains of it) by the number of days late (in 1 day increments, rounding up, e.g. 10 minutes late = 1 day deducted). Once the full 5 days have been used up, assignments will be penalized 10 percent per day late (rounding up), e.g., a 50 point assignment turned in 1.5 days late would be penalized 20%, or 10 points.

Homework, problem sets, and the project total to a possible 100 points. There will be no curving of grades, nor

will grades be rounded up. We'll use the plus/minus grading system: A= 92 and above, A-=90 to 91.99, etc. Here are the grade cutoffs:  $92\% \leq A$ ,  $90\% \leq A- < 92\%$ ,  $88\% \leq B+ < 90\%$ ,  $82\% \leq B < 88\%$ ,  $80\% \leq B- < 82\%$ ,  $78\% \leq C+ < 80\%$ ,  $72\% \leq C < 78\%$ ,  $70\% \leq C- < 72\%$ ,  $68\% \leq D+ < 70\%$ ,  $62\% \leq D < 68\%$ ,  $60\% \leq D- < 62\%$ , F < 60%.

Students are welcome to discuss ideas and problems with each other, but **all programs, Rosalind homework, and written solutions should be performed independently** (except the final collaborative project).

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

## BASICS OF PROGRAMMING

A Python programming primer for non-programmers

**Guest lecture: AppSoma: a collaborative cloud computing environment for UT students**

## BIOLOGICAL SEQUENCE ANALYSIS

Substitution matrices (BLOSSUM, PAM) & sequence alignment

Protein and nucleic acid sequence alignments, dynamic programming

Sequence profiles

BLAST! (the algorithm)

Biological databases

**Guest lecture: Protein 3D structural modeling**

Markov processes and Hidden Markov Models

## GENOMES, PROTEOMES, & "BIG BIOLOGY"

**Guest lecture: Next- (& next-next-) generation DNA and RNA sequencing**

Gene finding algorithms

Genome assembly & how the human genome was sequenced

An introduction to large gene expression data sets

Promoter and motif finding, Gibbs sampling

**Guest lecture: Overview of mass spectrometry shotgun proteomics**

**Guest lecture: Mapping universally shared animal protein complexes**

Clustering algorithms, hierarchical, k-means, self-organizing maps, force-directed maps

Classifiers, k-nearest neighbors, Mahalonobis distance

Principal component analysis and data transformations

## NETWORK & SYNTHETIC BIOLOGY

**Guest lecture: Genome engineering**

Biological networks: metabolic, signaling, graphs, regulatory

Network alignment and comparisons, network organization

Deep homology and the evolution of traits

Designing, simulating, and building gene circuits

Genome design and synthesis

**\*\*\* THE FINAL PROJECT IS DUE by midnight, May 4, 2015 \*\*\***

The final two class days will be devoted to presenting your projects to the rest of the class.

Several lectures (indicated above) will be given by researchers from the UT Center for Systems and Synthetic Biology (CSSB) on topics at the cutting edge of systems biology and bioinformatics.

**Note that there is NO CLASS over spring break (March 17 & March 19).**