BIOL 525: Computational Analyses and Resources in Genomics  
Department of Biology, University of North Carolina at Chapel Hill  
Spring 2013

Instructor: Terry Furey  
Office: Genetic Medicine Building, 5022  
E-mail: terry.furey@unc.edu  
Website: http://bio.unc.edu/people/faculty/furey/  
Office Phone: 966-7033  
Office Hours: TBA  
Class Meeting Times: Tues/Thurs 12:30-1:45pm, GSB 1374  
Computer Lab Meeting Time: Tues 2:00-2:250pm, GSB 1378  
Final Exam: Monday, May 6th, 12:00-3:00pm

Description  
This class aims to provide a study of computational techniques for the analysis of large-scale genomics data such as from microarrays and high-throughput sequencing technologies. With advances in technology that enable running experiments that generate massive amounts of genome-wide data, the ability to accurately and efficiently interpret and extract information from these data is critical for many fields of biological research.

Goals  
The goals of this class are to  
   (i) provide a deeper understanding of high-throughput experimentation, the enabling technologies, and the data produced to put these in the larger context of biological knowledge and research;  
   (ii) provide an overview of key aspects and techniques of computational biology research including understanding strengths and limitations of these efforts;  
   (iii) provide practical experience downloading and installing software, writing programs for biological applications, and using web-based tools and resources for the analysis of genomics data.

Prerequisites  
This course is intended for upper-level undergraduate and beginning graduate students in life sciences. Basic knowledge of molecular biology, beginning level programming skills, and familiarity with basic statistical concepts are expected, such as those learned in the following UNC classes or their equivalents:

BIOL 202 – Molecular Biology and Genetics  
COMP 116 – Introduction to Scientific Programming  
STOR 155 – Introduction to Statistics

Students may also request a waiver from the instructor.
Weekly Topics

**Week 1:** Databases, mysql queries, and importing data  
**Lab:** Installing and setting up a MySQL database

**Week 2:** Perl programming, part I  
**Lab:** Installing Perl, navigating operating systems

**Assignment #1, Perl Programming** (Due two weeks)

**Week 3:** Perl programming part II  
**Lab:** UCSC Genome Browser

**Week 4:** Introduction to microarrays  
**Lab:** Installing R programs

**Week 5:** Analyzing microarray data  
**Lab:** R programs for determining differential expression

**Assignment #2, Perl and MySQL** (Due two weeks)

**Week 6:** DNA sequencing technologies  
**Lab:** Introduction to the Galaxy pipeline system

**Week 7:** Sequence-based gene expression experiments (RNA-seq)  
**Lab:** Using Galaxy to import data, assess sequence data quality

**Week 8:** Sequence alignments  
**Lab:** Setting up pipelines in Galaxy

**Week 9:** Sequence-based transcription factor binding experiments (ChIP-seq)  

**Assignment #3, Galaxy Pipelines part 1** (Due two weeks)  
**Lab:** RNA-seq pipelines

**Week 10:** ChIP-seq assays for chromatin modifications  
**Lab:** ChIP-seq pipelines, differential peak calling

**Week 11:** Gene-ontology (GO) analysis  
**Lab:** Using GoSTAT

**Week 12:** Sequence-based open chromatin experiments (DNase-seq, FAIRE-seq)  

**Assignment #4, Galaxy Pipelines part 2** (Due two weeks)  
**Lab:** DNase-seq/FAIRE-seq pipelines

**Week 13:** DNasel Footprinting  
**Lab:** DNase-seq/FAIRE-seq pipelines

**Week 14:** Sequence-based chromatin interaction experiments (3C, ChIA-PET)  
**Lab:** Displaying results in the UCSC Genome Browser

**Week 15:** Detecting DNA variation, allelic imbalance in sequence data  
**Lab:** Review for final exam

**Final Exam**

**Reading and Resources**

Additional reading may be assigned in class and/or posted to the Sakai site throughout the semester to provide background for specific subject areas. Students may also find the following book to be a useful resource:


The Sakai system at UNC ([http://sakai.unc.edu](http://sakai.unc.edu)) will be used extensively to provide instructional material, assignments including student submission and grading of assignments, brief summaries of lectures, and resources related to class topics.
**Computer Labs**
This course includes a weekly one-hour computer lab. The labs will be designed to complement topics in class and provide more practical instructions on specific genomics resources such as at the NCBI and the UCSC Genome Browser. Additional instruction to help complete class assignments will also be provided.

**Grading**
Grades for this course will reflect the ability of the student to master both the theoretical and practical aspects of this course. These will be assessed through graded computer-based assignments, which will involve programming in some instances, and tests. Late homework assignments will be penalized as follows: 25% off first late assignment, 50% off second late assignments, no credit for third or more late assignments. Exceptions will be made by prior approval by instructor. Final grades will be computed as follows:

- Computer-based homework: (75%)
- Final exam (25%) – will include concepts from the whole semester.

While class participation is not formally factored into the final grade, it may be used to decide borderline cases.

**Honor Code**
Computational genomic research is, in general, highly collaborative and open. That being said, students need to learn to independently perform the work assigned in this class. Students are encouraged to help classmates understand general concepts and techniques learned in class, even related to homework assignments, but under no circumstances should complete answers, computer code, or the like for homework be shared. Specific questions about individual homework assignments should be discussed with the professor.

The in-class final exam will be open note, open computer, and open Internet, but must be completed independently by each student.

**Syllabus Changes**
The professor reserves to right to make changes to the syllabus, including project due dates and test dates. These changes will be announced as early as possible.