

**BIOL 525: Computational Analyses and Resources in Genomics**  
**University of North Carolina at Chapel Hill**  
**Spring 2012**

Instructor: Terry Furey  
Office: Genetic Medicine Building, 5022  
E-mail: [terry.furey@unc.edu](mailto:terry.furey@unc.edu)  
Office Phone: 966-7033  
Office Hours: TBA

**Description**

A study of computational techniques for the analysis of large-scale genomics data such as from microarrays and high-throughput sequencing technologies. Topics include databases and online genomic resources, programming for standard file processing, and development of analysis pipelines. This course includes a computational lab.

**Weekly Topics**

**Week 1:** Genomic resources on the web and how to use them

**Lab:** Setting up your environment

**Week 2:** Genome Browsers and functionality

**Lab:** Practical demonstration of genome browsers

**Week 3:** Database concepts, structure, and queries

**Lab:** Setting up a MySQL database

**Week 4:** Genome Browser databases, advanced capabilities

**Lab:** The DBD module in Perl for databases

**Week 5:** Introduction to microarray and sequencing technologies

**Lab:** Importing microarray data into your database

**Week 6:** Gene expression microarrays and their analyses

**Lab:** Introduction to R programming

**Week 7:** Genotyping microarrays and analyses

**Lab:** Normalization of microarrays, detecting differential expression

**Week 8:** Copy-number variation; Public repositories of microarray data

**Lab:** Detection of CNV's from genotyping arrays

**Week 9:** RNA-seq and DGE-seq, sequence-based gene expression assays

**Lab:** Programs to detect differential expression in sequence data

**Week 10:** ChIP-seq sequencing assays

**Lab:** Short-read aligners

**Week 11:** Methyl-seq, DNase-seq, FAIRE-seq, part I

**Lab:** Removing contaminants from sequences

**Week 12:** Methyl-seq, DNase-seq, FAIRE-seq, part II

**Lab:** Peak-calling for short-read assays

**Week 13:** Whole-genome sequencing

**Lab:** Setting up a computational processing pipeline

**Week 14:** Introduction to microbiome sequencing

**Lab:** Putting it all together, part I  
**Week 15:** Large-scale consortiums: ENCODE, TCGA, Epigenomics Roadmap  
**Lab:** Putting it all together, part II

### **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.

### **Reading**

The main textbook is *Building Bioinformatics Solutions: with Perl, R and MySQL* by Conrad Bessant, Ian Shadforth, and Darren Oakley, Oxford Biology Press.

Additional reading may be assigned in class and/or posted to the Blackboard site throughout the semester. Students may also find the following books to be useful resources:

Beginning Perl for Bioinformatics by James Tisdall. O'Reilly (2001).

### **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 assignments, both written and computer-based, 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:

- Written and computer-based homework: (50%)
- Midterm exam (20%)

- Final exam (30%) – will concentrate on material from midterm but 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.