

BISC478: Computational Genome Analysis

Spring 2016

Course overview

Advances in DNA sequencing technology have revolutionized biological science and genomic perspectives are now essential to all areas of biology. Dramatic reductions in the cost of sequencing will soon allow each of us to obtain our own genome sequence. The emerging concept of “personalized medicine” recognizes the value of our genomes in guiding clinical decisions and in some cases lifestyle decisions. To leverage genomics in this way depends not only on our capacity to sequence DNA, but also our ability to extract knowledge from genomic data. This course will introduce students to the concepts of modern genome data analysis. Students will learn the following:

- How current high-throughput sequencing technologies work
- What kinds of information that can be extracted from large genomic data sets
- Theoretical principles and techniques for data analysis
- How to apply those concepts in practical data analysis contexts

Instructors

Prof. Ting Chen	tingchen@usc.edu	RRI408h	213-740-2415	Friday 8-10AM
Prof. Andrew D. Smith	andrewds@usc.edu	RRI408e	213-821-4142	Tuesday 4-6PM

Evaluation

The evaluation will be based on two tests (non-cumulative), along with 5 graded labs. The tests will each be worth 30% of the course grade, and each lab will be worth 8%.

Textbook

There is no required textbook for this course. Lecture notes and other course materials will be given to students by the instructors.

Labs

One lab session will be held each week. A TA will be present to explain material and lab activities, and answer student questions. Students will submit 5 brief written lab reports that will be graded. Students are required to use their own computers to complete the required labwork. If a student cannot complete all labwork during the lab sessions, he/she can do it at home by himself/herself.

Lecture schedule

- (Week 1; January 12) Modern sequencing technologies and Precision Medicine
- (Week 1; January 14) High Performance Computing Hands-on (**bring your laptop to the classroom**)
- (Week 2; January 19) Mapping your genome: steps and a simple algorithm for read mapping
- (Week 2; January 21) Mapping your genome: faster read mapping with k-mer indexing
- (Week 3; January 26) Estimating cost of sequencing: estimating fraction of genome sequenced
- (Week 3; January 28) Estimating cost of sequencing: estimating number of gaps and distribution of read coverage
- (Week 4; Feb 2) Detecting your genome variants: copy number variations
- (Week 4; Feb 4) Detecting your genome variants: SNPs
- (Week 5; Feb 9) Detecting your genome variants: indels and structural variations
- (Week 5; Feb 11) Genome wide association studies
- (Week 6; Feb 16) Gut metagenome disease association studies
- (Week 6; Feb 18) Metagenome assembly: overlap-layout-consensus
- (Week 7; Feb 23) Metagenome assembly: de Bruijn Graph
- (Week 7; Feb 27) Test 1 (Midterm)
- (Week 8; March 1) Sequence alignment: pairwise, global
- (Week 8; March 3) Sequence alignment: pairwise, local
- (Week 9; March 8) Introduction to functional genomics and epigenomics
- (Week 9; March 10) Genome annotations and predicting function by orthology
- Spring Break
- (Week 10; March 22) Multiple sequence alignment
- (Week 10; March 24) Patterns in DNA (1)
- (Week 11; March 29) Patterns in DNA (2)
- (Week 11; March 31) ChIP-seq data analysis
- (Week 12; April 5) Clustering (1)
- (Week 12; April 7) Clustering (2)
- (Week 14; April 12) Gene expression analysis (1)
- (Week 14; April 14) Gene expression analysis (2)
- (Week 15; April 19) Correction for multiple hypothesis testing
- (Week 15; April 21) Phylogenetic trees
- (Week 16; April 26) Genetic variation in populations
- (Week 16; April 28) Test 2 (Final)

Lab schedule

- (Week 1; January 12) Intro to Python
- (Week 2; January 19) Genome annotations and data sources (UCSC Browser)
- (Week 3; January 26) Obtaining and manipulating genome sequence data (1)
- (Week 4; Feb 2) Obtaining and manipulating genome sequence data (2)
- (Week 5; Feb 9) Mapping sequenced reads
- (Week 6; Feb 16) Mapping reads and calling SNPs
- (Week 7; Feb 23) Calling SNPs
- (Week 8; March 1) Sequence alignment (1)
- (Week 9; March 8) Sequence alignment (2)
- (Week 10; March 22) ChIP-seq data analysis (1)
- (Week 11; March 29) ChIP-seq data analysis (2)

- (Week 12; April 5) Sequence motif discovery (1)
- (Week 14; April 12) Sequence motif discovery (2)
- (Week 15; April 19) Gene expression data analysis (1)
- (Week 16; April 26) Gene expression data analysis (2)