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 Tuesday 2-4PM
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 13) Modern sequencing technologies and applications
- (Week 1; January 15) Mapping sequenced reads to a genome: comparison and k-mer indexing
- (Week 2; January 20) Probabilistic models of whole genome sequencing: estimating # gaps
- (Week 2; January 22) Probability distributions for sequenced reads: Binomial and Poisson
- (Week 3; January 27) Estimating sequencing errors: base-calling errors and insertion/deletion errors
- (Week 3; January 29) SNP-calling (haploid/homozygous): Binomial, Likelihood ratio and Bayesian
- (Week 4; Feb 3) SNP-calling (diploid/multi-ploid): Binomial, Likelihood ratio and Bayesian
- (Week 4; Feb 5) Genome assembly: overlap-layout-consensus
- (Week 5; Feb 10) Genome assembly: de Bruijn graph
- (Week 5; Feb 12) Genome assembly: scaffolding
- (Week 6; Feb 17) Genome re-sequencing: detecting small insertions and deletions
- (Week 6; Feb 19) Genome re-sequencing: detecting copy number variations
- (Week 7; Feb 24) Genome re-sequencing: detecting structural variations
- (Week 7; Feb 26) Test 1 (Midterm)
- (Week 8; March 3) Sequence alignment: pairwise, global
- (Week 8; March 5) Sequence alignment: pairwise, local
- (Week 9; March 10) Introduction to functional genomics and epigenomics
- (Week 9; March 12) Genome annotations and predicting function by orthology
- (Week 10; March 24) Multiple sequence alignment
- (Week 10; March 26) Patterns in DNA (1)
- (Week 11; March 31) Patterns in DNA (2)
- (Week 11; April 2) ChIP-seq data analysis
- (Week 12; April 7) Clustering (1)
- (Week 12; April 9) Clustering (2)
- (Week 14; April 14) Gene expression analysis (1)
- (Week 14; April 16) Gene expression analysis (2)
- (Week 15; April 21) Correction for multiple hypothesis testing
- (Week 15; April 23) Phylogenetic trees
- (Week 16; April 28) Genetic variation in populations
- (Week 16; April 30) Test 2 (Final)

Lab schedule
- (Week 1; January 13) Intro to Python
- (Week 2; January 20) Genome annotations and data sources (UCSC Browser)
- (Week 3; January 27) Obtaining and manipulating genome sequence data (1)
- (Week 4; Feb 3) Obtaining and manipulating genome sequence data (2)
- (Week 5; Feb 10) Mapping sequenced reads
- (Week 6; Feb 17) Mapping reads and calling SNPs
- (Week 7; Feb 24) Calling SNPs
- (Week 8; March 3) Sequence alignment (1)
- (Week 9; March 10) Sequence alignment (2)
- (Week 10; March 24) ChIP-seq data analysis (1)
- (Week 11; March 31) ChIP-seq data analysis (2)
- (Week 12; April 7) Sequence motif discovery (1)
- (Week 14; April 14) Sequence motif discovery (2)
- (Week 15; April 21) Gene expression data analysis (1)
- (Week 16; April 28) Gene expression data analysis (2)