



STATS M254 / BIOINFO M271 / BIOMATH M271

Statistical Methods in Computational Biology

Mon & Wed, 2:00-3:15 PM, 2748 PAB

Instructor: Jingyi Jessica Li

Email: jli@stat.ucla.edu

WWW: <https://ccle.ucla.edu/course/view/15S-STATSM254-1>

<http://www.stat.ucla.edu/~jingyi.li/stats-m254-spring-2015.html>

Office: 8951 Math Sciences Bldg.

Office Hours: Mon 1:00-2:00 PM, and by appointment

Telephone: (310) 206-8375

*This syllabus is subject to change. Changes, if any, will be announced in class.

Course Description:

Introduction to statistical and computational methods in computational biology and bioinformatics. Emphasize on the understanding of basic statistical concepts and the ability to use statistical inference to solve biological problems. The course covers topics including gene expression data, regulatory sequence analysis, ChIP-chip/seq data, RNA-seq data, and their applications in gene regulation analysis. Statistical methods include multivariate methods, statistical sequence analysis, machine learning, Markov chain Monte Carlo, etc.

Prerequisite(s): STATS 100A (Probability) or STATS 200A (Probability) or BIOINFO 260A (Bioinformatics).

Credit Hours: 4

Text(s):

- Lecture notes (to be posted on CCLE)
- Reference papers (to be posted on CCLE)

Course Topics:

1. Introduction and Data: molecular biology of gene regulation, typical data.
2. Gene expression analysis: multiple tests, false discovery rate, clustering methods, liquid association, RNA-Seq data analysis.
3. Statistical Sequence Analysis: Bayesian inference, hidden Markov model, missing data, Monte Carlo, motif discovery, ChIP-Seq data analysis.

Course Grades:

1. **Scribe lecture notes.** I will assign 12 lectures for scribing. Every student will be assigned to scribe one lecture's note and check another lecture's scribed note by others. Please fill in your names in the assignment sheet. A LaTeX template will be posted on CCLE. The scribed and checked notes will be due **1 week** after the lecture.
2. **Reading assignments.** A reading paper will be posted to CCLE after every lecture as an interactive discussion topic. For every paper, please post one question or comment under the topic and read others' questions and comments.

3. **Attendance.** Please sign in for every lecture.
4. **Project.** Students form a group of no more than two students. The project datasets will be assigned in the 4th week. Each group will formulate a question based on the dataset. The question can be either a biological question or a statistical question, or both. Then each group will answer the question by using but not limited to the statistical methods we cover in this course. You may come to my office hours to discuss about your project. Each group will write a paper (total length ≤ 8 pages including figures and tables, font size ≥ 12 points, single space) and do a 20-min oral presentation in class during the last two weeks. Papers will be due by **Friday of Week 10**.

The course grade will be based on the calculation:

Scribing lecture notes	15%
Reading assignments	10%
Attendance	10%
Oral presentation	25%
Paper	40%

Communication:

Please keep a current e-mail address with my.UCLA.edu in order to receive class announcements and reminders.

Important Dates:

First day of classes: March 30, 2014.

Last day of classes: June 3, 2014.