

# BIostatistics 784 – Introduction to Computational Biology

## Fall 2017

Instructor: Dr. Michael Love (Mike)  
Contact: [milove@email.unc.edu](mailto:milove@email.unc.edu) (please add BIOS784 to subject line)  
4115E McGavran-Greenberg  
Class Time: MW 9:05-10:20  
Office Hours: TBD @ McG 4115E

### Course website:

Sakai for paper and slide PDFs, course calendar, etc.

Public material will be presented at <http://biodatascience.github.io/compbio>

### Course Description:

An introduction to computational biology, with emphasis on **practical examples of statistical analysis** of real biological datasets, and on **statistical methods and frameworks** commonly used in modern biological and biomedical research. We will introduce **data science tools** for code management, reproducible analyses, data cleaning and plotting, and packages specific to genomic data analysis. Examples of methodological topics covered include high dimensional data normalization, multiple testing, resampling strategies, network analysis, expectation-maximization (EM), linear and hierarchical models, hidden Markov models (HMM), and smoothing.

### Course Goals:

At the conclusion of the course, a student will be able to:

- Make exploratory plots of modern biological datasets
- Perform basic statistical analyses of biological data
- Comprehend computational Methods sections of biological/genomic publications
- State the goals and scope of current biological/genomic research
- Postulate new methods building on the frameworks presented in BIOS 784

### Prerequisites:

- BIOS 661 and 663 or permission of instructor.
- Working knowledge of *basic commands in R* (working with vectors, matrices and data frames, indexing and subsetting) is required at the beginning of the course, as we will be using R throughout to explore datasets and statistical methods. We won't have time to re-teach the very

basics of R. A self-evaluation quiz will be distributed before the class so students can ascertain if they are ready for BIOS 784.

### **Course Requirements:**

Reading of designated articles. Homeworks will be assigned typically every week. A take-home mid-term, similar to the homeworks but longer. A written final project and an in-class presentation of the final project (done in groups of 3). R programming will be used for the homework, mid-term and final project. More details below.

### **Not covered in BIOS 784:**

While sometimes part of a *Computational Biology* syllabus, the following topics will **not** be covered in this class due to lack of time: proteomics (measurement and analysis of protein abundances), protein or RNA folding, molecular dynamics, population genetics, evolution, imaging of molecules, cells or tissues, MRI, or computational neuroscience.

### **Optional references:**

There is no required book/reference for the course. The following are optional references:

- Online R Classes and Resources

<http://genomicsclass.github.io/book/pages/resources.html>

- Rafael Irizarry and Michael Love, “Data Analysis for the Life Sciences”

Free PDF <https://leanpub.com/dataanalysisforthelifesciences>

HTML <http://genomicsclass.github.io/book/>

- Kasper Hansen “Bioconductor for Genomic Data Science”

HTML <https://kasperdanielhansen.github.io/genbioconductor/>

### **Grading Policy:**

Weighting of Assignments:

Homeworks: 45%

Midterm: 15%

Final project: 40%

Note: Attendance is mandatory so not included in the weighting above. Unexcused absences will result in points deducted from final grade.

Graduate courses in the School of Public Health use the following grading system:

H: Clear excellence  
P: Entirely satisfactory  
L: Low passing  
F: Fail

The SPH grading system is designed so that the mode of the grading distribution is P.

### **Descriptions of Graded Work:**

**Homework** - Homework problems are designed to ensure that material from the lectures and course notes has been understood and mastered. Homeworks will typically be statistical analyses of real data that has already been presented in class, or implementation of methods that have already been shown in class. It will extend on analyses and methods shown in class. Homework will be given approximately every week, and students are allowed to talk about ideas and approaches to problems in groups, though students should write up the code for the assignments independently. Homework is due at 11 PM on the designated days listed on course calendar available in Sakai. *The lowest homework score will be dropped and the remainder of scores averaged.*

**Midterm** - The take-home midterm will be similar to a longer homework, and more time will be given to work on it. It will likely be due in mid October (more detail on course website and Sakai).

**Final project** - In groups of 3, students will start working on a final project in November, which will be a re-analysis of genomic data from a publication, or re-implementation of a method in a computational biology paper. The analysis or implementation will be written up as an Rmarkdown document (this format covered in the class). We will pick groups and topics early in the course. The last weeks in November will be devoted to guest lectures or in-class project time (no more homeworks), and the last 3 class periods reserved for student presentations on their work.

### **Student Honor Code:**

Students are encouraged to work together on homework or on the midterm, but verbatim copying of someone else's final work (e.g. copy + pasting all the code and turning in the identical homework or midterm) creates an honor code violation. Students suspected of academic misconduct will be referred to the Honor Court. Honor Court sanctions can include receiving a zero for the assignment, failing the course and/or suspension from the University.

For more information on the UNC Honor Code and the Honor Court see [honor.unc.edu](http://honor.unc.edu).

### **Attendance Policy & Late Work**

Regular, punctual attendance are required. Late work submission is not allowed, excepting excused absences, reserved for serious issues that prevent turning in work on time. Remember, the lowest homework score will be dropped.

**Accommodations:**

The instructor will provide accommodations for students registered with UNC Accessibility Resources and Services with written documentation. The instructor will also provide accommodations (make-up tests, adjustment of due dates) for religious observances with documentation.

**Class Rules and Etiquette:**

Put cell phones on vibrate. Keep cell phone out of sight and do not access except in case of emergency. If you receive an emergency call, leave class quietly. No texting.

Use laptop for taking notes/classwork only. Note-taking, using R, searching for information about course content is fine.

Email, Twitter, Facebook, shopping should not be open on your computer during class.

**Reuse of Materials:**

The materials for this course are to be used only for students currently enrolled in the course. Materials should not be further disseminated (exception for material already made publicly available by the instructor). Please don't transmit or post materials from this course that are not made public by the instructor – they are for your personal use. Please don't share materials (such as tests, homework, projects) with students who may take the course in the future.