

Stat 689: Statistical Bioinformatics
Spring 2010

Time and Place: MWF 12:40-1:30pm, Blocker 411; Q&A Sessions T 6:15-7:15pm, Blocker 457

Instructor: Dr. Alan Dabney

Office: 525H Blocker Building

Phone: 862-7581

Email: adabney@stat.tamu.edu

Office Hours: By appointment only

Course Description: High-throughput genomics / proteomics (“-omics”) applications are widely used in modern biological science and have great promise for expanding our understanding of systems biology. Common characteristics of these applications include: (a) very high-dimensional datasets, with thousands of “features” (genes or proteins, for example); (b) small sample sizes (relative to the number of features); and (c) complex technical aspects of the data collection process. These characteristics present great challenges to the data analyst charged with extracting meaningful information from -omics data.

This course will present an overview of the biological concepts, technologies, and statistical challenges of -omics applications. The student will learn how to handle, visualize, analyze, and interpret -omics data, using real-world datasets. We will begin with a brief overview of the biological and technological principles behind modern high-throughput -omic technologies, then move on to the bulk of the course, which will be divided into: (1) exploratory data analysis (or, how to visualize and summarize -omics data), (2) data preprocessing and normalization (or, how to remove unwanted aspects of the data prior to statistical inference), (3) hypothesis testing (for the selection of interesting features from very large candidate lists), and (4) classification and prediction (for using -omics data to predict disease status in new subjects, say).

Course Material: No textbook required. Course reading materials will consist of in-class lectures, online resources, and journal articles.

DoStat: We will use the DoStat course management system. To register, follow the steps below:

1. Go to <http://www.si-sv2846.com/dostat> and click on the “Register here” link.
2. Submit information for your account.
3. After logging in, click on “Add course,” and enter the appropriate information. The course reference ID is DS-58.

Prerequisites: STAT 604, 651, 652, or equivalent. Or prior approval by the instructor.

Computing: All computing will be done with R or other freely-available resources.

Grading: Your grade will be computed as follows:

- **Homework:** 25% - There will be periodic homework assignments. The assignments will typically involve the analysis of real datasets from the literature.
- **Exams:** 50% - There will be two midterm exams, each counting 25% of your grade. The first exam will be on Wednesday, Feb. 24, and the second exam will be on Wednesday, Mar. 31.

- **Research Project:** 25% - You will be assigned a research project requiring a comprehensive analysis of a real -omics dataset. The project will require you to develop your own analysis protocol to explore and interpret your data, beginning with the data in raw format, and ending with a formal statistical writeup.

Incomplete Grade: A temporary grade of I (Incomplete) at the end of a semester indicates that the student has COMPLETED THE COURSE WITH THE EXCEPTION OF A MAJOR QUIZ, FINAL EXAM, OR OTHER WORK. The instructor shall give this grade only when the deficiency is due to an authorized absence or other cause beyond the control of the student.

Statement on Disabilities: The Americans with Disabilities Act (ADA) is a federal anti-discrimination statute that provides comprehensive civil rights protection for persons with disabilities. Among other things, this legislation requires that all students with disabilities be guaranteed a learning environment that provides for reasonable accommodation for their disabilities. If you believe you have a disability requiring an accommodation, please contact the Office of Support Services for Students with Disabilities in Cain Hall. The phone number is 845-1637.

Statement on Plagiarism: The handouts used in this course are copyrighted. As such, you do not have the right to copy them unless I expressly grant permission. As commonly defined, plagiarism consists of passing off as one's own ideas, words, writing, etc., that belong to another. In accordance with this definition, you are committing plagiarism if you copy the work of another person and turn it in as your own, even if you have the permission of that person. If you have any questions, please consult the latest issue of the Texas A&M University Student Rules, under the section "Scholastic Dishonesty."

Academic Integrity Statement: "AN AGGIE DOES NOT LIE, CHEAT, OR STEAL OR TOLERATE THOSE WHO DO." Please refer to the Honor Council Rules and Procedures at <http://www.tamu.edu/aggiehonor>.