

COURSE ANNOUNCEMENT

STAT 892 Section 2: Statistical Methods for Microarray Data

Class Location and Time: 49 Hardin Hall North, 8:00-9:15, TR.

Instructor: Dr. Dong Wang

Office: 355C Hardin Hall North Phone: 472-4921

Email: dwang3@unlnotes.unl.edu

Scope and Objective:

This course will cover statistical aspects of designing microarray experiments and analyzing microarray data obtained with commonly used technology platforms. Statistical methods for both two-color microarrays and oligonucleotide microarrays will be discussed. Applications will be demonstrated on real data sets with popular software packages. After taking this course, students should gain a sound understanding of important statistical issues related to the microarray technology, be able to design simple microarray experiments for their research projects and analyze the data, as well as to describe the results in a manner suitable for publication.

Who should take this course: graduate students in biological sciences who are interested in the statistical aspects of microarray experiments, want to analyze microarray data sets and effectively collaborate with statisticians; or graduate students in statistics who are interested in the application of statistical methods to microarray experiments.

Prerequisite: a course in statistics such as STAT218 or STAT801.

Topic to be Covered:

Introduction

Basic Biological Concepts

Basics of the Microarray Technology

Image processing

Normalization methods:

Two-color microarrays

Oligonucleotide microarrays

An example of two-sample comparison problems

Multiple testing and false discovery rate

Experimental design issues for microarrays

Mixed linear model analysis

Empirical Bayes methods

Clustering of genes

Related technologies: RT-PCR, protein arrays, tiling arrays, etc. Depending on student interests and available time.

Wrapping up and miscellaneous comments

Softwares: In this course I will use R, Bioconductor, and SAS. R is a language and environment for statistical computing and graphics, it can be obtained freely from <http://www.r-project.org>. Bioconductor is an open source software package for analysis of genomic data. It is mostly based on R and can be obtained at <http://www.bioconductor.org>. SAS is commercial software package available on various sites on campus.

Grading: the grading of this course will be based on the homework assignments, two exams and a project. The final score used for grading will be calculated largely as the following,

Final score=homework(20%)+midterm exam(25%)+final exam(35%)+project(20%).

Homeworks will be handed out approximately every two weeks, some problems require using statistical software to analyze data. Students can discuss the homework problems with the instructor or with other students, but each student has to write his/her own codes and finish the answers independently.

There will be two **exams** for this course. The midterm exam will be held in March with the specific date to be determined. The final exam will be held on May 2. Makeup exam will not be given unless there exist extreme circumstances.

In the **project**, students will analyze a real data set using methods covered in lectures. The instructor will assign students into teams to work on the project. More details will be announced later.

