CSBL 6095
Analysis and Visualization of Genomic Data
Spring 2018

CLASS DAYS and TIME: Wednesdays, 3:00-5:00

CLASSROOM: GCCRI 2.150

COURSE DIRECTORS:

Yidong Chen, PhD
Professor
Department of Epidemiology of Biostatistics
School of Medicine
Office: GCCRI 4.100.06
Telephone: 210 562-9163
Email: cheny8@uthscsa.edu

Alexander Pertsemlidis, PhD
Associate Professor
Departments of Pediatrics and C&SB
School of Medicine
Office: GCCRI 3.100.22
Telephone: 210 562-9062
Email: pertsemlidis@uthscsa.edu

<div style="border:2px solid black; text-align:center; color:red; font-weight:bold;">
READ THIS DOCUMENT CAREFULLY - YOU ARE RESPONSIBLE FOR ITS CONTENTS.
</div>

COURSE DESCRIPTION AND OBJECTIVES

In recent years, life science researchers have witnessed an explosion of genome-wide profiling technologies that have generated – and continue to generate -- an unprecedented amount of data. These genome-wide data are multi-modal in nature and include measurements spanning DNA to RNA to proteins, tissues in normal or perturbed states, organisms from simple models to humans, providing ample opportunities for biologists to address their investigation objectives. Along with these opportunities come challenges: large amounts of data, data quality issues and noise filtering, data modeling and prediction from high-dimensional biological data, multi-modality and heterogeneous biological data integration and visualization. In this class, we will concentrate on the methods used in various bioinformatics tools, with the purpose of reinforcing *intuition and concepts*, rather than detailed mathematical derivation. The organization of the class reflects the analysis and visualization methods aspects of high-throughput profiling data: (1) common methods, including differential gene expression, clustering, classification and visualization; (2) model-based methods, including network modeling and biological pathway-based analysis; and (3) other advanced topics, such as genome-wide association analysis.

As stated above, this course is designed to address data analysis and visualization issues in genome-wide data generated from high-throughput technologies. Although many methods are shared, methods such as sequence alignment, gene and functional element prediction, RNA/protein folding, and other comparative genomic issues will not be covered.

It is highly recommended that all students use the R statistical programming environment throughout the class and for the homework assignments.

Pre-requisites

Required: CSBL 5095 Experimental Design and Introduction to Statistics

Recommended: INTD 5067 Introduction to Bioinformatics and Computational Biology

Semester credit hours – 2 SCH

By the end of this course, each student should understand:

- current technologies for the generation of large biological datasets
- methods for analysis and visualization of gene expression and differential gene expression data

- methods for analysis and visualization of ChIP-seq and DNA methylation data
- methods for analysis and visualization of genome-wide association data
- methods for analysis and visualization of metagenomics and gene regulation network data

COURSE ORGANIZATION

Three main teaching formats are used to cover the material in this course: **1) Conventional didactic lectures** in which information is delivered to the class; **2) Conference Discussions** which are highly interactive case-based activities, encouraging two-way communication between the instructor and the class, and requiring student active participation in the learning process; and **3) Online review and self-study activities**.

This course is designed to extend statistical bioinformatics concepts at a level appropriate for students who have completed the required genetics/genomics core courses. Students will learn the statistical concepts necessary for high-throughput data analysis and visualization, and gain hands-on experience in conventional statistical/bioinformatics software or web-based applications. By using the well-validate approach of problem-based learning, students will learn through applying the strategies and tools used in bioinformatics to topical problems drawn from ongoing research and applications in a variety of fields. The content and curricula of the course may adjust to evolve with developing technologies and changing topics of interest.

The class will prepare students to handle genome-wide profiling data, and to communicate and collaborate with statisticians and computational biologists in their future experimental research. As a minimum at the end of this course, there should be a solid understanding of the scope of data analysis and visualization. And we fully expect that the basic methods discussed in the class to be applicable in many different fields of interest. Students should gain substantial competency in content, skills, and awareness within the field of bioinformatics, and the confidence to apply rigorous statistical and programming methods for data analysis and visualization. Many of the problems presented in the course will serve as launch points for further inquiry and exploration as students move on into other specific and advanced courses.

The course is based on classroom lectures. A 2-hour class (with 1 break) will be held once per week. As determined by each instructor, an in-class computer lab may be required. The R statistical programming environment and specific Bioconductor packages and a personal computer/notebook are highly recommended.

Materials –

Computer Access – The R statistical programming environment and specific Bioconductor packages and a personal computer/notebook are highly recommended. Computer hardware/software requirement: Laptop computer may be used for software demonstration/distribution. R/Bioconductor installation is recommended.

Reading Assignments – Required reading assignments are posted in the schedule of class meetings (shown below) and are never considered optional. Unless specifically noted by the instructor, anything in the required readings, whether emphasized in class or not, is considered testable on exams. Mandatory readings are primarily found in the required text book (see below). However, occasionally a reading assignment will be given that is posted online or sent to you via email attachment.

Lectures – Many of the presentations are given in the common lecture format and are accompanied by slides. You are responsible for all information included in the lecture materials. However, you should not assume that all testable lecture material is found only in the posted materials. That is, lectures may be expanded and enhanced during in-class presentations. **So take good notes because any information discussed in class is considered testable.**

ATTENDANCE

Attendance at lectures, discussions and lab sessions is required.

In order to achieve the expected level of competency, students must be fully engaged. Therefore, attendance for every class session is expected.

It is recognized that a student may occasionally arrive late to class due to unexpected traffic problems or inclement weather. However, chronic lateness is considered an unprofessional behavior that disrupts the learning environment for everyone else in the classroom.

TEXTBOOKS

Recommended:

Lee JK. *Statistical Bioinformatics: A Guide for Life and Biomedical Science Researchers*. Hoboken, N.J.: Wiley-Blackwell, 2010.

Hahne F, Huber W, Gentleman R, Falcon S. *Bioconductor Case Studies*. Springer Publishing Company, 2008.

Mathur SK. *Statistical Bioinformatics with R*. Academic Press, 2010.

GRADING POLICIES AND EXAMINATION PROCEDURES

Attendance - Students are required to attend the class and in-class labs, and will be graded on the basis of participation.

Homework - Assigned after each class, to be turned in before the subsequent class. Most assignments will be to perform computational tasks or programming for additional points. No makeup homework will be allowed.

Final Exam/Project/Presentation - Testable material comes from 3 main sources: Lectures, readings and in-class labs.

Examination Protocol – No electronic devices, extra paper, books, backpacks, etc. are permitted in the testing area. Hats must be removed. You will **not be allowed to ask questions of the proctor** once the examination has started (except to point out potential typographical errors in the exam).

Late Arrival to Exams - Exams will be timed. **If you arrive late** to an exam, and are given permission to take the exam, you will not be given additional time to complete your test. If you arrive after another student has finished the exam and has departed the exam room, you will not be allowed to take the exam. If you miss an exam, you may be eligible for taking a make-up exam (see below).

Grading Procedures – Exam results will be provided to students as quickly as possible. No "challenges" are allowed. However, a time will be scheduled outside of class so that students may review concepts of concern to them.

Make-up Examinations – A student who must miss a scheduled exam for a serious reason must request an excused absence from the Course Director. Acceptable "serious reasons" usually involve serious illness or injury to the student (doctor's excuse may be required) or the student's family member. Examples of unacceptable reasons include: Not prepared or incomplete studying, over-sleeping, hangover, heavy traffic or any travel delays, other appointments or scheduled professional or personal commitments.

If it is determined that missing an exam is justified, a make-up examination will be scheduled. The make-up exam will be given as soon as possible at a time designated by the Course Director. Any student who misses an exam and does not receive an excused absence **will receive a grade of zero for that exam**.

<u>Grading System</u>
Final letter grades are primarily based on class participation, homework assignments, and the final exam/presentation. Grading is based on the following scale:

A = 90-100%   B = 80-89%   C = 70-79%   F = < 69%

REQUESTS FOR ACCOMODATIONS FOR DISABILITIES

In accordance with policy 4.2.3, **Request for Accommodation Under the ADA and the ADA Amendments Act of 2008 (ADAAA)**, any student requesting accommodation must submit the appropriate request for accommodation under the American with Disabilities Act (ADA, form 100). to his/her appropriate Associate Dean of their School and a copy to the ADA Coordinator. Additional information may be obtained at http://uthscsa.edu/eeo/request.asp.

ACADEMIC INTEGRITY AND PROFESSIONALISM

Any student who commits an act of academic dishonesty is subject to discipline as prescribed by the UT System Rules and Regulations of the Board of Regents. Academic dishonesty includes, but is not limited to, cheating, plagiarism, collusion, the submission for credit of any work or materials that are attributable in whole or in part to another person, taking an exam for another person, signing attendance sheets for another student, and any act designed to give unfair advantage to a student or the attempt to commit such an act. Additional information may be obtained at http://catalog.uthscsa.edu/generalinformation/generalacademicpolicies/academicdishonestypolicy.

TITLE IX AT UTHSCSA

Title IX Defined:
Title of the Education Amendments of 1972 is a federal law that prohibits sex discrimination in education. It reads "no person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving Federal financial assistance."

University of Texas Health Science Center San Antonio's Commitment:
University of Texas Health Science Center San Antonio (UTHSCSA) is committed to maintaining a learning environment that is free from discriminatory conduct based on gender. As required by Title IX, UTHSCSA does not discriminate on the basis of sex in its education programs and activities, and it encourages any student, faculty, or staff member who thinks that he or she has been subjected to sex discrimination, sexual harassment (including sexual violence) or sexual misconduct to immediately report the incident to the Title IX Director.

In an emergency, victims of sexual abuse should call 911. For non-emergencies, they may contact UPD at 210-567-2800. Additional information may be obtained at http://students.uthscsa.edu/titleix.

EMAIL POLICY

Every student is issued a University e-mail address and account at the time of enrollment. As a matter of University Policy, communications between students and faculty that occur using the student's University e-mail address is considered official business. Therefore, students are expected to check their university email inboxes on a regular basis so that any announcements, instructions, or information regarding this course will be received in a timely way. Missed communications due to inadequate monitoring of incoming emails on the University's email server will never be a valid excuse for unsatisfactory academic progress.

USE OF RECORDING DEVICES

Recording of lectures and other learning activities in this course by any means (e.g., video, audio, etc.) is only permitted if approved by the instructor or required for compliance with Americans with Disabilities Act (ADA).

ELECTRONIC DEVICES

Computers, tablets use is allowed only for participating in classroom activities (e.g., viewing slides presented in lecture, reading assignments or articles, data analysis, etc.). Activities unrelated to course objectives such as using phone or texting are not allowed.

CLASS SCHEDULE
CSBL 6095, Every Wednesday, January – May 2018
Analysis and Visualization of Genomic Data

WEEK	DATE	TOPIC	Assignment	Instructor and Modality
Week 1	1/10	Course introduction		Chen, Pertsemlidis
Week 2	1/17	Introduction to statistics with R		Gelfond
Week 3	1/24	Overview of high-throughput genomic technologies		Pertsemlidis
Week 4	1/31	Analysis and visualization of gene expression data		Chen
Week 5	2/7	Statistical analysis of differential expression		Chen
Week 6	2/14	Discussion		Chen, Pertsemlidis
Week 7	2/21	Statistical methods for ChIP-seq and gene regulation		Jin
Week 8	2/28	Visualization of genomic alteration		Chen
Week 9	3/7	Visualization of big data		Chen, Pertsemlidis
Week 10	3/21	Discussion		Chen, Pertsemlidis
Week 11	3/28	Analysis and visualization of genetic variation		Pertsemlidis
Week 12	4/4	Statistical analysis of genetic variation		Pertsemlidis
Week 13	4/11	Statistical methods for gene sets		Chen
Week 14	4/18	Clustering and classification		Chen
Week 15	4/25	Metagenomics		Chen
Week 16	5/2	Discussion		Chen, Pertsemlidis
Week 17	5/9	Exam/Presentations		Chen, Pertsemlidis