Course Syllabus

High-throughput molecular genomic and epigenomic data analysis

Biostat 646; Bioinf 545; Stats 545

Winter 2013

Class time: TBA
Lecture location: TBA
Lab location: TBA

Office hours: Will be set by the 2nd day of class via doodle poll. Both instructors and the GSI will hold office hours.

Course Instructors:
Dr. Laura Scott, M4134 SPH II, 763-0006, ljst@umich.edu
Dr. Maureen Sartor, 2044B Palmer Commons, 763-8013, sartorma@umich.edu

Teaching Assistant: TBA
The graduate student instructor will hold optional, bi-weekly lab sessions to answer questions regarding labs and homework. Times to be determined.

Prerequisites:
Graduate Standing and Stat 400, Biostat 523, Biostat 553 or equivalent (or permission of instructor). Please contact one of the instructors if you have a question regarding prerequisites.

Recommended Book: TBA

Goals of Class:
The class mainly concentrates on basic analysis of microarrays, RNA-Seq, and ChIP-Seq data including hands-on lab sessions. The class also covers an introduction to the underlying biology and the technologies for measuring RNA levels, transcription factor binding and epigenetic modifications, and quality control of microarray and deep sequencing data.

Grading:
Based on homeworks (some of which are based on lab sessions) and class project
8 Homeworks @ 7.5% x 8 60%
Class Participation 5%
Final Project 35%

Assignments:
Homeworks are due at the start of class (hard copy). If you will miss class please arrange to turn in your homework to the TA before the class (may obtain GSI permission to submit electronically in this case.) We will drop the lowest graded homework for each student. If a homework is skipped, that will be the one dropped. Homeworks submitted 1-2 days late will be deducted by 10%.

**Group Project:**
You will work in assigned groups of students (2-4 individuals/group) on a final project. We will try to assign groups based on mutual interest and complementary skills. Each group will be responsible for identifying an analysis problem and performing an analysis or developing an approach to analyzing a given type of data, creating a thorough scientific write-up, and presenting your work to the class. The project proposals and final projects should be turned in by the assigned due date/time.

**Academic Integrity:**

The following is the School of Public Health statement on academic integrity (Standard of Academic Conduct, University of Michigan School of Public Health, revised 08/08):

“The faculty of the School of Public Health believes that the conduct of a student registered or taking courses in the School should be consistent with that of a professional person. Courtesy, honesty, and respect should be shown by students toward faculty members, guest lecturers, administrative support staff, and fellow students. Similarly, students should expect faculty to treat them fairly, showing respect for their ideas and opinions, and striving to help them achieve maximum benefits from their experience in the School.

Student academic misconduct refers to behavior that may include plagiarism, cheating, fabrication, falsification of records or official documents, intentional misuse of equipment or materials (including library materials), and aiding and abetting the perpetration of such acts. The preparation of reports, papers, and examinations, assigned on an individual basis, must represent each student’s own effort. Reference sources should be indicated clearly…”

The standards of academic conduct and procedures for dealing with alleged violations of these standards are detailed online at the following URL:

http://www.sph.umich.edu/students/handbook/rights.html

Allegations of violation of this honor code and standard of academic integrity will be handled according to SPH procedures as detailed in the web site, http://www.sph.umich.edu/students/handbook/rights.html

**Based on the SPH Competencies this course includes:**

**Core competencies**

1. Describe basic concepts of probability, random variation and commonly used statistical probability distributions.

2. Describe preferred methodological alternatives to commonly used statistical methods when assumptions are not met.
3. Apply descriptive and inferential methodologies according to the type of study design for answering a particular research question.

4. Develop written and oral presentations based on statistical analyses for both public health professionals and educated lay

**Biostat MS competencies**

1. Develop knowledge to communicate and collaborate effectively with scientists in a variety of health-related disciplines to which biostatistics are applied (e.g. public health, medicine, genetics, biology; psychology; economics; management and policy).

2. Become well-versed in the application of core statistical techniques (biostatistical inference, linear regression, generalized linear models, analysis of variance (ANOVA), linear mixed models) and 4-5 selected statistical specialization techniques.

3. Select appropriate techniques and apply them to the processing of data from health studies.


5. Interpret the results of statistical analysis and convert them into a language understandable to the broad statistical community.

6. Develop written and oral presentation skills and other scientific reporting skills, based on statistical analyses for public health, medical and basic scientists and educated lay audiences.

**Course specific competencies**

1. Understand the role of data cleaning and normalization procedures for data generated from microarrays, RNA-Seq, ChIP-Seq and DNA methylation experiments.

2. Basic understanding of the probability distributions (normal, poisson, hypergeometric, negative binomial) that underlie the statistical tests for data generated from microarrays, RNA-Seq, ChIP-Seq and DNA methylation experiments.

3. Understand the impact on statistical testing when probability distributional assumptions are not met.

4. Be able to evaluate type 1 error rate for a specific dataset and statistical test

5. Understand strategies for evaluating the power of a test and comparing the results of different tests for a given dataset.

6. Identify potential biases that may occur in detection of differentially expressed genes or differential binding of DNA binding proteins

7. Be able to analyze data from microarrays, RNA-Seq, and ChIP-Seq datasets, interpret and present the results.
8. Be able to correct for multiple testing, understanding when and why it's necessary

9. Develop ability to critically read the literature and evaluate results