DEPARTMENT: Biostatistics and Bioinformatics

COURSE NUMBER: 516 **SECTION NUMBER:**

CREDIT HOURS: 1 **SEMESTER:** Fall

COURSE TITLE: Introduction to Large-Scale Biomedical Data Analysis

CLASS HOURS AND LOCATION:

INSTRUCTOR NAME: Ying Guo, Suprateek Kundu, Yijuan Hu, Zhaohui Qin, Hao Wu, Tianwei Yu

INSTRUCTOR CONTACT INFORMATION

EMAIL: hao.wu@emory.edu

PHONE: (404)727-8633

SCHOOL ADDRESS OR MAILBOX LOCATION: The Rollins School of Public Health

OFFICE HOURS: By appointment

Teaching Assistant(s):

COURSE DESCRIPTION

This is an overview course for the Bioinformatics, Imaging and Genetics (BIG) concentration in the PhD program of the Department of Biostatistics and Bioinformatics. It aims to introduce students to modern high-dimensional biomedical data, including data in bioinformatics and computational biology, biomedical imaging, and statistical genetics.

This course will be co-taught by several BIG core faculty members, with each faculty member giving one or two lectures. The focus of the course will be on the data characteristics, opportunities and challenges for statisticians, as well as current developments and active areas of the research fields of bioinformatics, biomedical imaging and statistical genetics.

Prerequisites: BIOS 501 or equivalent, or permission from the instructor.

MPH/MSPH FOUNDATIONAL COMPETENCIES:

- Select quantitative data collection methods appropriate for a given public health context.

CONCENTRATION COMPETENCIES:

- Identify biostatistical aspects in contemporary public health issues.
- Communicate the results of statistical analyses to a broad audience.

COURSE LEARNING OBJECTIVES:

Upon completion of the class, the students will be able to:

- Understand the structures and characteristics of some common data types generated for modern biomedical research.
- Understand the challenges and opportunities modern biomedical data poses to the field of statistics.
- Understand some general statistical concepts involved in high-dimensional biomedical data analysis, such as multiple testing, dimension reduction etc.
- Grasp the current trend of the methodological research in the areas of bioinformatics, biomedical imaging and statistical genetics.

EVALUATION

- Participation 10%. The students are expected to attend all classes and actively participate in in-class discussions on various topics of biomedical research.
- Reading assignments 90%. There are three reading assignments. In each reading assignment, students will be given a methodology paper on related topics. The students are required to write a two-page report (1.5x line spacing, 1 inch margins) to briefly summarize the main points of the paper.

COURSE STRUCTURE

MPH/MSPH Foundational Competency assessed	Representative Assignment
Select quantitative data collection methods appropriate for a given public health context.	Write reports for reading assignments
BIOS Concentration Competencies assessed	**Representative Assignment**
Identify biostatistical aspects in contemporary public health issues	Write reports for reading assignments
Communicate the results of statistical analyses to a broad audience.	Write reports for reading assignments

COURSE POLICIES

As the instructor of this course I endeavor to provide an inclusive learning environment. However, if you experience barriers to learning in this course, do not hesitate to discuss them with me and the Office for Equity and Inclusion, 404-727-9877.

RSPH POLICIES

Accessibility and Accommodations

Accessibility Services works with students who have disabilities to provide reasonable accommodations. In order to receive consideration for reasonable accommodations, you must contact the Office of Accessibility Services (OAS). It is the responsibility of the student to register with OAS. Please note that accommodations are not retroactive and that disability accommodations are not provided until an accommodation letter has been processed.

Students who registered with OAS and have a letter outlining their academic accommodations are strongly encouraged to coordinate a meeting time with me to discuss a protocol to implement the accommodations as needed throughout the semester. This meeting should occur as early in the semester as possible.

Contact Accessibility Services for more information at (404) 727-9877 or accessibility@emory.edu. Additional information is available at the OAS website at http://equityandinclusion.emory.edu/access/students/index.html

Honor Code

You are bound by Emory University's Student Honor and Conduct Code. RSPH requires that all material submitted by a student fulfilling his or her academic course of study must be the original work of the student. Violations of academic honor include any action by a student indicating dishonesty or a lack of integrity in academic ethics.

Academic dishonesty refers to cheating, plagiarizing, assisting other students without authorization, lying, tampering, or stealing in performing any academic work, and will not be tolerated under any circumstances.

The RSPH Honor Code states: “Plagiarism is the act of presenting as one’s own work the expression, words, or ideas of another person whether published or unpublished (including the work of another student). A writer’s work should be regarded as his/her own property.” (http://www.sph.emory.edu/cms/current_students/enrollment_services/honor_code.html)

COURSE CALENDAR

COURSE OUTLINE

TENTATIVE SCHEDULE

Lecture 1 **Introduction to next generation sequencing data (1)**
Introduction to next generation sequencing technology. RNA-seq: biological motivations, experimental procedure, challenges and opportunities in statistical data analysis.

Lecture 2 **Introduction to next generation sequencing data (2)**
ChIP-Seq: experimental procedure and statistical data analysis.

Lecture 3 **Introduction to Proteomics and Metabolomics**
LC/MS technology, challenges in data processing. Biological pathways.

Lecture 4 **Introduction to biomedical imaging data analysis**
An overview of modern radiology and imaging techniques and related statistical problems.

Lecture 5 **Statistical Analysis of Neuroimaging Data: A look at current methods and future challenges**
Data characteristics associated with several imaging modalities, data processing techniques, challenges and important directions for future research.

Lecture 6 **A discussion of statistical methods for analyzing large-scale data in Neuroimaging studies**
Adjusting for large number of multiple tests in hypothesis testing in neuroimaging data, and dimension reduction techniques.

Lecture 7 **Basic concepts of molecular genetics and population genetics**
Brief introduction to the biological backgrounds needed for statistical genetics, concepts from population genetics that are most relevant to association analysis.

Lecture 8 **Genetic association studies**

Commonly used tests for association, challenges especially in the context of genome-wide association studies (GWAS), including how to correct for population stratification and multiple testing.