ROLLINS SCHOOL OF PUBLIC HEALTH
EMORY

DEPARTMENT: Biostatistics and Bioinformatics

COURSE NUMBER: 555 **SECTION NUMBER:** **SEMESTER:** Fall

CREDIT HOURS: 2

COURSE TITLE: High-throughput data analysis using R and BioConductor

INSTRUCTOR NAME: Hao Wu

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

BRIEF COURSE DESCRIPTION

This course covers the basics of microarray and second-generation sequencing data analysis using R/BioConductor and other open source software. Topics include gene expression microarray, RNA-seq, ChIP-seq and general DNA sequence analyses. We will introduce technologies, data characteristics, statistical challenges, existing methods and potential research topics. Students will also learn to use proper Bioconductor packages and other open source software to analyze different types of data and deliver biologically interpretable results.

Prerequisite:
BIOS 501 or equivalents. Basic programming experience in R.

EVALUATION

Homework: 60%
Final project: 30%
Class participation: 10%

LEARNING OBJECTIVES OR COMPETENCIES OF THE COURSE

1. To learn the key capabilities of BioConductor packages.
2. To understand the biological motivations and technological procedures of high-throughput experiments including different types of microarrays and second-generation sequencing.
3. To understand statistical challenges and existing methods for analyzing the data generated from high-throughput experiments.
4. To develop skills for analyzing high-throughput data using R/BioConductor and some other open source software.

LEARNING OBJECTIVES OR COMPETENCIES FOR THE DEPARTMENT OR PROGRAM TO WHICH THE COURSE CONTRIBUTES

Upon completion of the PhD degree in Biostatistics the graduate will be able to:

- Use a variety of statistical computer packages.
- Conduct appropriate statistical analyses.

TENTATIVE SCHEDULE

- **Lecture 1**: Introduction. Brief introduction of molecular biology, high-throughput experiments, R and Bioconductor.
- **Lecture 2**: Lab 1: exploratory analysis of human Refseq genes. (**Homework 1**)
- **Lecture 3**: Gene expression microarray I. Experimental procedures and data pre-processing methods for Gene expression microarrays.
- **Lecture 4**: Gene expression microarray II and other types of microarrays. Differential expression analysis from GE arrays. Introduction to tiling arrays.
- **Lecture 5**: Lab 2: analyzing gene expression array data from MAQC. (**Homework 2**)
- **Lecture 6**: Handling genome data using Bioconductor I. Introduce Biostrings and BSgenome Bioconductor packages.
- **Lecture 7**: Handling genome data using Bioconductor II. Introduce GenomicRanges and GenomicFeatures Bioconductor packages.
- **Lecture 8**: Lab 3: Analyzing human genome. (**Homework 3**)
- **Lecture 9**: Introduction to second generation sequencing. Introduce second generation sequencing technologies, statistical challenges, and software tools for alignment, variant calling and visualization.
- **Lecture 10**: RNA-seq. Experimental procedure and data analysis for RNA-seq data. Normalization and differential expression detection. DEseq and edgeR Bioconductor packages.
- **Lecture 11**: ChIP-seq. Experimental procedure of ChIP-seq. Peak calling methods. Joint analysis of multiple ChIP-seq. Joint analysis of ChIP-seq and RNA-seq.
- **Lecture 12**: Lab 4: Handling second generation sequencing data, RNA- and ChIP-seq analyses. (**Homework 4**)
- **Lecture 13**: Other types of sequencing data. Briefly introduce other types of sequencing experiments: Bisulfite sequencing, Hi-C and CLiP-seq.