Stat 555: Statistical Analysis of High Throughput Biology Experiments

Instructor: Qunhua Li, Dept. of Statistics

Office: 514D Wartik Lab

Email: qunhua.li@psu.edu

Office hours: 1:45-3:45pm T

TA: Tao Yang, txy146@psu.edu, 514 Wartik Lab,

Office Hour: TBA

Frank Shen, fvs5027@psu.edu, 514 Wartik Lab,

Office Hour: TBA

The course covers some statistical methods for the design and analysis of bioinformatics experiments.

Prerequisites: Because of the rapid pace of the course, it is recommended that students have background in basic statistics (e.g. Stat 500 or Stat 501) and computing concepts (e.g. loops and functions) prior to starting the course. A basic understanding of cell biology is also recommended. The following courses are helpful:

IBIOS 554 Foundations in Data Driven Life Sciences

BMMB 852 Applied Bioinformatics

STAT 500 Applied Statistics or STAT 501

Tentative Topics

1. Introduction to R and RStudio
2. Introduction to cell biology
3. Introduction to measurement technologies: microarrays, sequencing, and SNPs
4. Basic statistics: descriptive statistics, hypothesis testing, multiple testing, ANOVA and regression, bootstrap
5. GWAS (Regression)
6. Gene Expression Microarrays (Normalization, Regression, ANOVA, multiple testing, empirical Bayes)
7. RNA-seq (Poisson, Negative Binomial, analysis of count data, Generalized Linear Model)
8. ChIP-seq (Analysis of count data)
9. Gene Set enrichment analysis (Categorical data analysis)
10. Clustering and Classification using statistical machine learning
11. Dimension reduction (PCA, etc.)
12. Selected topics

Useful references:

We will use the part of the following two text books and review papers for our class. More reading will be assigned.

Biomedical Data Science **http://genomicsclass.github.io/book/**

An introduction to statistical learning **http://www-bcf.usc.edu/~gareth/ISL/**

Points of significance on Nature methods

http://www.nature.com/collections/qghhqm/pointsofsignificance

Evaluation:

Homework (50%) + Mini Lesson (10%) + Project (40%)

Course Project (40%):

The project is an opportunity for you to work on a problem of interest to you, by utilizing statistical methods and tools for analysis of high-throughput data. You will work in groups of 2-4. You will give a 15min to 20min presentation in class and submit a written report. More information is provided later in the semester.

Mini Lesson (10%):

You will present a statistical tutorial paper from the column of "Point of significance" on Nature Methods in class. The lesson is an opportunity for you to learn some basic statistics in the context of biology and learn to communicate with your peers. Each presentation is about 10min to 15 min in class.

All Penn State and Eberly College of Science policies regarding academic integrity apply to this course. For details, see http://www.science.psu.edu/academic/Integrity/index.html