BENG 183. Applied Genomic Technologies

Syllabus

Instructor: Sheng Zhong, Ph.D. see CANVAS for instructor and TAs' emails.
Lectures: TuTh 3:30-4:50pm. For security, Zoom ID for lectures is provided on CANVAS.

TAs and emails:

- Xingzhao Wen (Irene)
- Jingtian Zhou
- Rodolfo Salido Benitez, see CANVAS for TAs' emails.

TA office hours:

- Xingzhao Wen (Irene): Tu 5-6pm, For security, Zoom ID is provided on CANVAS.
- Jingtian Zhou: Fri 4-5pm, For security, Zoom ID is provided on CANVAS.
- Rodolfo Salido Benitez: Thu 2-3pm, For security, Zoom ID is provided on CANVAS.

Prerequisites

BIMM 100 or CHEM 114C, or consent of department.

Course Description

This course introduces genomics technologies and their biomedical applications. Students will learn and genome sequencing, RNA sequencing, and the computational methods for analyses of these data. Probability theory and applications together with the basics of machine learning methods will be introduced. Students will be able to formulate hypotheses and perform statistical tests. Students will be able to design gene expression analysis experiments, transform next-generation sequencing data into summary data matrix, identify differentially expressed genes, perform K-nearest neighbor classification, perform hierarchical clustering and K-means clustering. Contemporary research developments on personalized medicine and liquid biopsy analyses will be discussed.

Schedule

This course has six modules. This course will have a mixture of interactive Zoom lectures (Modules 1, 2, 3, 5), pre-recorded lecture videos (Module 4), and live student presentations (Module 6).

All the interactive Zoom lectures will be recorded and released on CANVAS as soon as the recordings become available (usually in the same day as the lecture date).

You are welcome to study the pre-recorded lecture videos (Module 4) at any time, but no later than our recommended studying dates (see below).

- Module 1: DNA sequencing, RNA sequencing, and ChIP sequencing
    - Lecture 1 – 4: 10/01, 10/06, 10/08, 10/13: These are interactive Zoom lectures.
- Module 2: Analysis of RNA-seq data
    - Lectures 5 – 7: 10/15, 10/20, 10/22. These are interactive Zoom lectures.
- Module 3: Genome interaction and a primer to precision medicine

    - Lectures 8 – 10: These are interactive Zoom lectures.
    - Recommended studying dates: 10/27, 11/03, 11/05.
- Module 4: Random variable and hypothesis testing
    - Lectures 11 – 13: Pre-recorded lecture videos.
    - Recommended studying dates: 11/10, 11/12, 11/17.
- Module 5: A primer to machine learning and applications to cancer classification.
    - Lectures 14 – 16: 11/19, 11/24, 12/01. These are interactive Zoom lectures.
- Module 6: Group presentations.
    - Sessions 17 – 18: 12/08, 12/10. Both sessions are interactive Zoom sessions.

Textbook and Other Materials

Required Reading

1. Applied Genomic Technologies (course handout), https://github.com/Zhong-Lab-UCSD/Applied-Genomic-Technologies
2. 3D Genome (eBook), Xingzhao Wen and Sheng Zhong, Open Access Publications from the University of California (eScholarship), https://zhonglab.gitbook.io/3dgenome
3. Introduction to Probability, Statistics, and Random Processes, Hossein Pishro-Nik, Kappa Research, LLC. August 24, 2014. ISBN-10: 0990637204. ISBN-13: 978-0990637202. Please note that the textbook is also freely available online: https://www.probabilitycourse.com/

Additional Materials

In addition to the textbook listed above, there is a plethora of available information on the Internet. This information includes peer-reviewed manuscripts, Wikipedia articles, YouTube videos of lectures from UCSD and/or other universities, and much more. A course handout will be provided. Students are strongly encouraged to review additional online materials prior to each lecture.

Course Outcomes

After successfully completing this class, a student will be able to:

1. Understand the technologies of Sanger sequencing, sequencing by synthesis, RNA sequencing, and extracellular-RNA sequencing.
2. Design genomic technologies to identify gene expression biomarkers.
3. Apply probability theory to solve real-life problems.
4. Formulate statistical hypotheses and perform hypothesis testing.
5. Identify differentially expressed genes from gene expression data.
6. Cluster data from a data matrix.
7. Perform classification of data from a data matrix.
8. Analyze RNA sequencing data with bioinformatic software.