

BENG 183. Applied Genomic Technologies

Overview of course modules, homework assignments, and exams

1. Course website

<https://systemsbio.ucsd.edu/course/beng183fa24/>

2. Required materials

1. Next-generation sequencing data analysis. CRC Press. Author: Xinkun Wang. **Free access** within UCSD: https://search-library.ucsd.edu/permalink/01UCS_SDI/ld412s/alma991023723967006535

This book will be referred to as the “**NGS book**”.

2. Introduction to Probability, Statistics, and Random Processes. Author: Hossein Pishro-Nik. ISBN-13: 978-0990637202. This textbook is **freely available online**: <https://www.probabilitycourse.com/>
3. BENG 183 Course Handout, https://github.com/Zhong-Lab-UCSD/BENG183_2024Fall_Applied-Genomic-Technologies/tree/main
4. 3D Genome eBook. Xingzhao Wen, Sheng Zhong. **Free access** <https://zhonglab.gitbook.io/3dgenome>

5. [Course slides](#)

Note: Not all chapters of the Books 1-4 are required. Please see the “Required reading materials” for each lecture below.

3. Lecture formats

The lectures will be composed of in-person lectures and remote lectures (pre-recorded videos).

4. An overview of course modules and homework assignments

- Module 1: DNA sequencing, RNA sequencing, and ChIP sequencing
 - Lectures 1 – 4.
 - Homework 1.
- Module 2: Analysis of RNA-seq data
 - Lectures 5 – 7.
 - Homework 2.
- Module 3: Genome interaction, RNA interactions, and a primer to precision medicine
 - Lectures 8 – 10.
 - Homework 3
- Module 4: Random variable and hypothesis testing.
 - Lectures 11 – 13.
 - Homework 4.
- Module 5: A primer to machine learning.
 - Lectures: 14 – 15.
 - Homework 5.
- Module 6: Group presentations.
 - Sessions: 16 – 18.

5. An overview of midterms and final presentation

- Midterm I: covering all materials in Modules 1 and 2.
- Midterm II: covering all materials in Modules 3, 4, and 5.
- Final presentation: please present a mini-discussion session as if you are the TA of this course. You can use this mini-discussion to solve a problem, or introduce a highly useful function from a software or an entire software package, or show a data analysis application that is highly related to the lectures of this course. To discuss a data analysis application, you can and you are welcome to use published papers and results.

6. Course lectures

Module 1: DNA sequencing, RNA sequencing, and precision medicine

Lecture 1. A brief review of molecular biology

- General information about this course
 - Course description and studying goals
 - Course schedule
 - Evaluation
 - Final presentation and final paper
- The following review materials are for you to review or self-study.
- Reading materials:
 - NGS book, Chapter 1.1-1.4.3, Figure 1.1, Chapter 1.5
 - Lecture 1's slides
 - Optional reading materials:
 - 3D Genome eBook, [Chapter 0.1: Molecular Biology](#).
 - YouTube: [Molecular Biology](#)
 - YouTube: [RNA world hypothesis](#)
 - Core concepts:
 - Restriction enzyme
 - Hydrogen bonds
 - Gel electrophoresis
 - PCR
 - DNA sequencing
 - RNA world hypothesis

Lecture 2. DNA sequencing (pre-recorded lecture video)

- Reading materials:
 - NGS book, Chapters 2.1-2.6, 2.8-2.9.1, Figures 2.1, 2.2, 2.3
 - NGS book, Chapters 4.1, 4.3, 4.5, Figures 4.1, 4.2, 4.3, 4.6, Table 4.1
 - 3D Genome eBook, [Chapter 0.2: Sequencing Technologies](#).
 - Lecture 2's slides
 - YouTube: [Sanger Method - Reading DNA Strand Sequences after Gel Electrophoresis](#)
 - Wikipedia: [Sanger sequencing](#)
 - YouTube: [Illumina sequencing by synthesis](#)

- Optional reading materials:
 - [Illumina Sequencing by Synthesis](#)
- Core concepts:
 - Sanger sequencing
 - DNA synthesis by primer extension
 - Chain termination of PCR
 - PAGE (Polyacrylamide gel electrophoresis)
 - Automation of Sanger sequencing with fluorescent nucleotides
 - Sequencing by DNA Synthesis

Lecture 3. RNA sequencing

- Reading materials:
 - NGS book, Chapter 3.1-3.3, Figures 3.2, 3.3
 - NGS book, Chapter 5.1-5.2, Figures 5.1, 5.2, 5.3
 - NGS book, Chapter 7.1-7.2, Table 7.1.
 - Course handout, Chapter 3.1
 - Lecture 3's slides
- Core concepts:
 - Genome structure
 - Gene structure
 - Non-coding elements in the genome
 - Gene regulation
 - Genome size paradox
 - The multiple levels of gene regulation
 - Transcription regulation
 - The RNA-seq technology
 - Differential expression
 - Gene expression matrix and heatmap
 - Two-sample comparison

Lecture 4. ChIP sequencing (pre-recorded lecture video)

- Reading materials:
 - NGS book, Chapter 13.1-13.3.2, Figures 13.1, 13.2, 13.3, (Optional) 13.4
 - NGS book, Chapters 5.3.1, 5.3.3-5.3.4, Figures 5.7, 5.8, 5.9, Table 5.2
 - Lecture 4's slides
 - Course handout, Chapter 4.
 - 3D genome eBook, [Chapter 2.2: Experimental techniques for accessing primary order chromatin](#)
 - 3D genome eBook, [Chapter 3.1: Computational analysis, primary order analysis](#)
- Core concepts:
 - Functional features of the genome
 - ChIP-seq technology
 - FASTQ format
 - Quality scores for sequencing data
 - Mapping sequencing reads

Module 2: Analysis of RNA-seq data

- Reading materials for this entire module:
 - 3D Genome eBook, [Chapter 0.3: RNA-seq Data Analysis: Mapping and Quantification](#).
 - 3D Genome eBook, [Chapter 0.4: RNA-seq Data Analysis: Differential Gene Expression](#).

Lecture 5. RNA-seq data analysis, part 1 (pre-recorded lecture video)

- Reading materials:
 - NGS book, Chapter 7.3.1-7.3.3, Figure 7.1
 - NGS book, Appendix I
 - Lecture 5's slides
- Core concepts:
 - Alternative splicing
 - Alignment of sequencing reads to the genome
 - Visualization of the alignment output file (BAM file)
 - Normalization of RNA-seq data
 - RPKM (Reads Per Kilobase of transcript per Million mapped reads)

Lecture 6. RNA-seq data analysis, part 2

- Reading materials:
 - Lecture 6's slides
- Core concepts:
 - RPKM (Reads Per Kilobase of transcript per Million mapped reads)
 - FPKM (Fragments Per Kilobase of transcript per Million mapped reads)
 - TPM (Transcripts Per Million)
 - Relationship between TPM and FPKM

Lecture 7. RNA-seq data analysis, part 3 (pre-recorded lecture video)

- Reading materials:
 - NGS book, Chapters 7.3.5-7.5, Figure 7.4, Table 7.2
 - Lecture 7's slides
- Core concepts:
 - STAR software
 - Reads mapping
 - FeatureCounts software
 - Quantification of gene expression
 - DESeq2 software
 - Detection of differentially expressed genes
 - Metascape software
 - Detection of enriched biological functions

Module 3: Genome interaction and a primer to precision medicine

Lecture 8. Genome organization, Hi-C (pre-recorded lecture video)

- Reading materials:
 - Lecture 8's slides
 - 3D Genome eBook, [Chapter 1.1: from 2D to 3D nuclear structure](#)
 - 3D Genome eBook, [Chapter 1.3: intra- and inter-chromosomal interactions](#)
 - 3D Genome eBook, [Chapter 2.3: Higher order C-techs](#)
- Core concepts:
 - Enhancer-promoter interaction
 - Folding of the chromosomes
 - Hi-C technology

Lecture 9. RNA-RNA, RNA-DNA interactions (pre-recorded lecture video)

- Reading materials:
 - Lecture 9's slides
 - 3D Genome eBook, Chapter 4.1-4.2
 - [Systematic mapping of RNA-chromatin interactions in vivo](#). Bharat Sridhar et al., Current Biology, 2017, 27(4): 602–609.
 - [Mapping RNA-RNA interactome and RNA structure in vivo by MARIO](#). Tri C. Nguyen et al., Nature Communications, 2016, 7:12023.
- Core concepts:
 - The non-coding parts of the human genome
 - RNA-RNA interaction
 - RNA-DNA interaction

Lecture 10. A primer to precision medicine

- Reading materials:
 - Lecture 10's slides
 - Course handout, Chapter 5.1 – 5.4
 - YouTube: [Personal medicine](#)
- Core concepts:
 - Personalized medicine
 - Gene expression assays
 - Sensitivity and specificity
 - Differential expression
 - Gene expression matrix and heatmap
 - Two-sample comparison

Module 4: Random variable and hypothesis testing.

Lecture 11: Random variable and cumulative distribution function (pre-recorded lecture video)

- Reading materials:
 - [Pishro-Nik book](#) chapter 3.1 – 3.1.4.
 - [Pishro-Nik book](#) chapter 3.2.1.
 - [Pishro-Nik book](#)'s accompanying videos:
 - [Video 3.1 – Introduction to Random Variables: Discrete Random Variables - Part 1](#)

- [Video 3.2 - Discrete Random Variables, PMF, Independent Random Variables](#)
 - [Video 3.5 – CDF for Discrete Random Variables](#)
- Pre-recorded lecture video
- Core concepts:
 - Random variable (RV)
 - Discrete random variable
 - Probability mass function (PMF)
 - Cumulative distribution function (CDF)
- **Not required** contents: book chapters other than 3.1 – 3.1.4 and 3.2.1.

Lecture 12. Hypothesis testing (pre-recorded lecture video)

- Reading materials:
 - [Pishro-Nik book](#) chapter 8.4.1-8.4.2, chapter 8.4.4.
 - Course handout, Chapter 6.1
 - Lecture 12's slides
- Pre-recorded lecture video
- Core concepts:
 - Null and alternative hypotheses
 - Test statistic
 - P-value
- **Not required** contents: Pishro-Nik chapters 8.1 – 8.3.

Lecture 13. Acceptance and rejection regions (pre-recorded lecture video)

- Reading materials:
 - [Pishro-Nik book](#) chapters 8.4.2 and 8.4.4.
 - Course handout, Chapter 6.1
- Core concept:
 - Acceptance region and rejection region
 - Type I error
 - Significance level
 - Type II error
- **Not required** contents: Pishro-Nik chapter 8.4.5.

Module 5: A primer to machine learning.

- Required reading material for this entire module: Course handout, Chapter 7.1

Lecture 14. An introduction to classification

- Reading materials:
 - Lecture 14's slides
- Core concepts:
 - Classification
 - Feature selection
 - Training data and testing data

- True positives and true negatives
- False positives and false negatives
- Precision and recall

Lecture 15. An introduction to clustering

- Reading materials:
 - Lecture 15's slides
- Core concepts:
 - Hierarchical clustering
 - Distance
 - Dendrogram
 - K-means clustering
 - Iterative algorithm

Module 6: Group presentations.

Session 16. Student group presentations, part 1.

Session 17. Student group presentations, part 2.

Session 18. Student group presentations, part 3.

Session 19. Student group presentations, part 4.