ENG 183. Applied Genomic Technologies

Recorded lectures

All the lecture videos are accessible at:
https://drive.google.com/drive/folders/17uAGEWxUHuxOTaoZ0VV44OTBcWeJD0or?usp=sharing

Module 1: DNA sequencing and RNA sequencing

Lecture M1-2. DNA sequencing

- Required reading materials:
  - 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 videos
  - Part 1: A primer on Sanger sequencing.
  - Part 2: Reading the gel of Sanger sequencing
  - 2.1. YouTube video: Sanger Method - Reading DNA Strand Sequences after Gel Electrophoresis
  - 2.2. Q&A: how to read the sequencing gel.
  - Part 3: Automation of the Sanger sequencing.
  - Part 4: Illumina sequencing.
  - 4.1. YouTube: Illumina sequencing by synthesis
  - Part 5: Q&A on Illumina sequencing.
  - 5.1: How long does the Illumina sequencing take?
  - 5.2: How does the sequencer limit only one nucleotide to each cycle?
  - 5.3: How to assemble short sequences from the Illumina sequencer into a long sequence?
  - 5.4: How Illumina determines the quality score at each base?

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 M2-1. RNA-seq data analysis, part 1

- Reading materials:
  - 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 videos:
    - Part 1. An overview of the course progress.
    - Part 2. Transcription and Alternative splicing.
    - Part 3. Alignment of sequencing reads to the genome.
    - Part 4.1. Visualization of the alignment output file (BAM file).
    - Part 4.2. Q&A and practicing reading the alignment result figure.
    - Part 5.1: Calculating RNA expression level from RNA-seq data.
    - Part 5.2. Q&A: normalization of RNA-seq data by penalizing gene length.
    - Part 6: RPKM (Reads Per Kilobase of transcript per Million mapped reads).
    - Part 7: Q&A: normalization and RNA isoforms.

Lecture M2-3. RNA-seq data analysis, part 3.

- Reading materials:
    - Lecture 7's slides
- Core concepts:
    - STAR software
    - Reads mapping
    - FeatureCounts software
    - Quantification of gene expression
    - DESeq2 software
    - Detection of differentially expressed genes
    - Detection of enriched biological functions
- Lecture videos:
    - Part 1. STAR software.
    - Part 2. FASTQ file and quality score.
    - Part 3. Reads mapping and BAM file.
    - Part 4: Detection of differential expression.
    - Part 5: Data analysis Q&A.
    - Part 6: Output of DEseq2 software.
    - Part 7: Functional enrichment analysis.

Module 3: Genome interaction and a primer to precision medicine

Lecture M3-1. ChIP sequencing

- Reading materials:
    - 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
- Lecture videos:
    - Part 1: A brief review of the course modules.

- Part 2: DNA sequencing as a tool to interrogate the functional features of the genome.
- Part 3: Transcription factor and functional features of the genome.
- Part 4: ChIP sequencing.
- Part 4.1: JoVE video: ChIP sequencing (this video is in the course slides folder).
- Part 4.2: Toward narrating the ChIP-seq technology video.
- Part 5: FASTQ files, sequencing quality, sequencing mapping.

Lecture M3-2. Genome organization, Hi-C

- 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 videos:
  - Part 1. Genome folding.
  - Part 2. Enhancer-promoter interaction and units of genome organization.
  - Part 3. Methods for studying genome organization.
  - Part 4. The Hi-C technology.
  - Part 5. Q&A: Does Hi-C require a special restriction enzyme?
  - Part 6. Q&A: Is self-ligation a problem in Hi-C?

Lecture M3-3. RNA-RNA, RNA-DNA interactions

- 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 videos:
  - Part 1. Non-coding genome and RNA-RNA interactions.
  - Part 2. The MARIO technology.
  - Part 3. MARIO data analysis.
  - Part 4. Sequence conservation of RNA-RNA interaction sites.
  - Part 5. Q&A: How the linker is used in the bioinformatic analysis?
  - Part 6. Q&A: What is the protein purification necessary for MARIO?
  - Part 7. Q&A: Does MARIO have to selectively label RNA-binding proteins?

Module 4: A primer to machine learning.

(Optional). Lecture M4-3. Cancer classification with extracellular RNA biomarker

Module 5: Random variable and hypothesis testing.

Lecture M5-1: Random variable and cumulative distribution function

The purpose of this lecture is to review several key concepts in probability. If you are certain that you have mastered the following concepts, you are welcome to skip this lecture.

- 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.

Section 1
Chapter 3.1.1. Random variables.
1. Definition.
2. Example 3.1.
3. Example 3.2.

Section 2
Chapter 3.1.3. Probability mass function
1. Definition
2. Example 3.3
3. Plotting a PMF: Figure 3.1

Section 3
4. Example 3.4

Section 4
Chapter 3.2.1. Cumulative distribution function
1. Definition 3.10

Lecture M5-2. Hypothesis testing

- Reading materials:
  - Pishro-Nik book chapter 8.4.1-8.4.2, chapter 8.4.4.
  - Couse 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.

Section 1
An overview of the major steps of hypothesis testing.

Section 2
Forming two competing hypotheses, called **the null (H0) and the alternative hypothesis (H1)**.

Section 3
Generating or getting data.

Section 4
Summarizing the data into a **Test Statistic**.

Section 5
Calculating the **p-value**.

Section 6
Making a decision based on p-value.

Lecture M5-3. Acceptance and rejection regions

- Reading materials:
  - Pishro-Nik book chapters 8.4.2 and 8.4.4.
  - Couse 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.

(Optional) Section 1
Review of the major steps for hypothesis testing.

Section 2
An alternative way of making a decision:

- Acceptance region

Section 3

- Type I error

- Type II error
- Significance level