ISMB (MOLB 7900)

Jay Hesselberth

2019-04-21

Contents

Course Overview

Informatics and Statistics for Molecular Biologists (MOLB 7900) is offered at the University of Colorado School of Medicine and teaches students how to design and analyze common molecular biology experiments.

The **bookdown** package can be installed from CRAN or Github:

```r
install.packages("bookdown")
# or the development version
# devtools::install_github("rstudio/bookdown")
```

Syllabus

0.1 Course Overview

Informatics and Statistics for Molecular Biologists (MOLB 7900) teaches students to design and analyze experiments commonly used in molecular biology. The course is organized around the Central Dogma (DNA > RNA > Protein) wherein each block presents 2-3 experimental approaches. Each week, a new experiment is introduced with a discussion of appropriate design and statistical considerations. The remaining weeks' classes are devoted to digging into the analysis of a sample data set.

The course begins with a 4 week "boot camp" designed to get students familiar with and bring them up to speed on using software for shell programming and data analysis with R and Python. We also establish accounts on Github for problem set submission.

0.1.1 Block 0: Bootcamp

```
A: Shell programming
B: R Studio
C: Python
D: Integration with Git
```

0.1.2 Block 1: DNA

```
A: Chromatin accessibility
B: Motif finding
C: Variant Calling
```

0.1.3 Block 2: RNA

A: Quantitative PCR B: Bulk mRNA-seq C: RNA-protein interaction (CLIP)

0.1.4 BLOCK 3: PROTEIN

```
A: Mass spectrometry (counts)
B:
C: Fluorescent protein localization (Chad)
```

0.2   Course Objectives

- Cater to students of many backgrounds (more computational or more biological)
- Be able to formulate questions that are testable with computational techniques
- Understand the limitations of sequencing-based techniques
- Be fluent in statistical considerations for different approaches and design well-controlled experiments that can be analyzed using statistical tests
- Be fluent in command-line programming, scripting (Python) and data analysis / viz (R / R Studio)
- Understand the value of internet-based analysis tools (NCBI BLAST etc)
- Use reproducible software development approaches (Github) and dynamic documents (Rmarkdown)
- Independently conceive of and implement a soup to nuts reanalysis of an existing data set, which are presented to the class.

0.3   Class Schedule

Monday, Wednesday, and Friday, 1-2:30pm from Sept 10 to Dec 10.

Chapter 1

Bootcamp

Chapter 2

Block 1: DNA

Chapter 3

Block 2: RNA

Chapter 4

Block 3: Protein

Chapter 5

Exercises