# Computational Genomics

## Course Information and Policies

| | |
|---|---|
| **Semester** | Spring 2019 |
| **Classroom** | TBD |
| **Meeting Times** | Mondays/Thursday 1:40 - 4:40 pm |
| **Course URL** | https://github.com/shahlab/teaching |
| **Course Instructor** | Premal Shah<br>premal.shah@rutgers.edu<br>Ph: (848) 445 9664 |
| **Office hours** | By appointment |
| **Teaching Assistant** | TBD |

## Course Description

The main focus of this course is to learn R programming and apply it towards analysis of genomic datasets. In this course, we will focus on basics of programming, data wrangling, creating user-defined functions, and exploratory graphical data analysis. The primary data sets considered will contain genome sequences, genome annotations, RNA-seq and/or other expression data from multiple model organisms.

## Course Goals

The main objectives of this course are to enable students to learn:

1. Basics of programming in R programming.
2. Manipulation of tabular datasets.
3. Exploratory data analyses using visualization tools.
4. Working with genome sequences and genome annotation files.
5. Basics of RNA-seq.
6. Creating pipelines for trimming/mapping of RNA-seq datasets.
7. Differential gene-expression analysis and identifying gene-ontology and pathway enrichments among gene-sets.
8. Basics of statistical analyses of genomic datasets.

# Course Pre-requisites

1. **Undergraduates**:

   - Gen Bio Lab 01:119:117 or 01:119:102
   - Genetics 01:447:380 or Genetic Analysis I 01:447:384

2. **Graduate students**:
   While there are no pre-requisites for graduate students to join the class, they are STRONGLY encouraged to participate in the two-day DataCarpentry workshop that will be conducted during the winter break.

# Core Curriculum Learning Goals

Information Technology and Research [ITR] goals y and z:

```
y. Employ current technologies to access information, to conduct research, and to communicate
z. Analyze and critically assess information from traditional and emergent technologies.
```

# Course Materials

| Week | Date | Topic | Learning Materials |
|------|------|-------|--------------------|
| 1 | Jan 24 | Introduction and Overview | In-coming survey - Link<br>Presentation<br>R code for presentation<br>Homework |
| 2 | Jan 28 | Variables and data structures I | Presentation<br>R code for presentation<br>Homework |
| 3 | Jan 31 | Variables and data structures II | Presentation<br>R code for presentation<br>Homework |
| 4 | Feb 04 | Quiz 01 and recap | Presentation<br>R code for presentation<br>Homework |
| 5 | Feb 07 | Conditional statements and loops | Presentation<br>R code for presentation |

| Week | Date | Topic | Homework Learning Materials |
|------|------|-------|-----------------------------|
| 6 | Feb 11 | Lists, functions, and strings | Presentation<br>R code for presentation<br>Homework |
| 7 | Feb 14 | Advanced data wrangling – tidyverse I | Presentation<br>R code for presentation<br>Homework |
| 8 | Feb 18 | Quiz 02 and recap | Presentation<br>R code for presentation<br>Homework |
| 9 | Feb 21 | Advanced data wrangling – tidyverse II | Presentation<br>R code for presentation<br>Homework |
| 10 | Feb 25 | Visualization – ggplot2 | Presentation<br>R code for presentation<br>Homework |
| 11 | Feb 28 | Introduction to bioconductor – IRanges | Presentation<br>R code for presentation<br>Homework |
| 12 | Mar 04 | Quiz 03 and recap | Presentation<br>R code for presentation<br>Homework |
| 13 | Mar 07 | Working with genome annotations – GRanges | Presentation<br>R code for presentation<br>Homework |
| 14 | Mar 11 | Working with genome sequences – Biostrings | Presentation<br>R code for presentation<br>Homework |
| 15 | Mar 14 | Advanced sequence analyses | Presentation<br>R code for presentation<br>Homework |
| 16 | Mar 25 | Quiz 04 and recap | Presentation<br>R code for presentation<br>Homework |
| 17 | Mar 28 | Statistics for life scientists I | Presentation<br>R code for presentation<br>Homework |

| Week | Date | Topic | Learning Materials |
|------|------|-------|--------------------|
| 18 | Apr 01 | Statistics for life scientists II | R code for presentation<br>Homework |
| 19 | Apr 04 | Introduction to RNA-seq | Presentation<br>R code for presentation<br>Homework |
| 20 | Apr 08 | Quiz 05 and recap | Presentation<br>R code for presentation<br>Homework |
| 21 | Apr 11 | Working with Illumina sequencing data | Presentation<br>R code for presentation<br>Homework |
| 22 | Apr 15 | Working with short reads in R | Presentation<br>R code for presentation<br>Homework |
| 23 | Apr 18 | Differential gene expression analyses I | Presentation<br>R code for presentation<br>Homework |
| 24 | Apr 22 | Quiz 06 and recap | Presentation<br>R code for presentation<br>Homework |
| 25 | Apr 25 | Differential gene expression analyses II | Presentation<br>R code for presentation<br>Homework |
| 26 | Apr 29 | Gene Ontology and KEGG Pathway analyses | Presentation<br>R code for presentation<br>Homework |
| 27 | May 02 | Review | Presentation |

# Website and Communication

### Course website

This GitHub repo will serve as the course website.

### Communication

The primary mode of communication in this course (including major announcements), will be the course Slack account https://Sp19CompGenomics.slack.com. All of you should have invitations to join this account in your Rutgers email.

### Emails

Although the bulk of the communication will take place via Slack, at times (rarely), we will send out important course information via email. This email is sent to your Rutgers email address. You are responsible for all information sent out to your University email account, and for checking this account on a regular basis.

## Course Activities

### In-class exercises

Each class will contain 2-4 modules. Each module will involve a ~15 min lecture followed by a programming exercise lasting ~20-30 min. Programming exercises will typically involve 2-5 questions. You will be asked to solve these questions on your own, which will involve applying the functions taught in the lecture part of the module. At the end of the individual exercise portion, I will go over the exercise as a group. Any piece of code written by me will be annotated and commented before being uploaded to the course website.

### Homeworks

In every lab, you will receive a homework assignment, typically a series of programming exercises that you must code based on the material you learn in class that day. You will upload solutions to these homework to Slack as a single R Notebook file before the beginning of the next lab. NO LATE SUBMISSIONS WILL BE ALLOWED. If you miss your due date for the homework, you will receive a 0 grade for that assignment.

Each homework is worth 10 points.

### Discussion channels

Once every 2 classes, a discussion channel (in Slack) will be opened. You will be given a current events article regarding genomic data with at least one open-ended question. The articles may or may not be related to that particular lab's topic. The point of these threads is to generate discussion about the material in the article. Throughout the semester, students are required to participate regularly and provide thoughtful comments. You do not need to limit yourself to only commenting. Thoughtful questions that you may have from reading the article are also valid.

1. Any and all comments must be about science, MUST use facts as evidence to support your claims, and must not contain inappropriate language or content. You are required to participate in these forums at least once every week. However, your participation must be **meaningful**. If you participate

using only comments such as "I agree", "very interesting", etc. you will get a 0 for that weekly forum. Having said that, there are no right questions/comments to be asked/made in these forums. The idea is that you engage in them.

2. Participation will be graded on a 0-to-2 scale every week. Note that, in what follows, the word "comment" may be replaced by "question". Points are given as follows:

   ○ You will receive **0** points if you do nothing or if you comment is not meaningful (as described above).
   ○ You will receive **1** point if you write a meaningful/thoughtful original comment **OR** if you provide a meaningful response to someone else's comment (including TA or Instructor comments).
   ○ You will receive **2** points if you write a meaningful/thoughtful original comment **AND** you provide a meaningful response to someone else's comment (including TA or Instructor comments).

## Quizzes

Quizzes will be conducted once every 4 classes. They will last 1-hr and be similar to in-class exercises. Quizzes will typically be "open-book", i.e., you will be able to refer to your class notes, presentations, and my codes.

Each Quiz is worth 10 points.

## Final Quiz

The final quiz will be cumulative. Half of the questions will be from previous in-class exercises, homeworks, and quizzes. The other half will involve new applications of the material we cover over the course of the semester. There will be a 15-25 questions. Once you start the Quiz, you will have FOUR hours to complete it.

The Final Quiz is worth 20% of the final grade.

## Final projects

Towards the end of the course, all students will be split into teams of 2-3 individuals. Each team will complete a finals research project. Each team can select one of the two broad research topics lsited below:

1. **Comparative genomic analyses**:
   This project will involve downloading genome sequences and annotations of 2 or more organisms and performing various sequence level analyses.

2. **RNA-seq analyses**:
   This project will involve downloading raw sequencing datasets from GEO database and identification of differentially expressed genes and their functional characterization.

Detailed instructions for the finals projects will be given 2 weeks prior to the last class. The final project accounts for 25% of the final grade.

## DataCamp for Graduate Students

All graduate students taking this course will be assigned one DataCamp module every two weeks. The modules will be parallel to the materials covered in the class. The students will be required to submit their certificates of completion from DataCamp. Failure to complete all DataCamp modules will lead to an automatic loss of one grade.

While not required, undergraduates are also encouraged to work through DataCamp modules. No submission of completion certificates is required.

## Grading Information

| Activity | Percentage |
| --- | --- |
| Programming homeworks | 25% |
| Discussion channels | 15% |
| Quizzes | 15% |
| Final Exam | 20% |
| Final project | 25% |

### Grading Scale

| Grade | Percentage |
| --- | --- |
| A | ≥ 90% |
| B+ | ≥ 85% |
| B | ≥ 80% |
| C+ | ≥ 75% |
| C | ≥ 70% |
| D+ | ≥ 65% |
| D | ≥ 60% |
| F | < 60% |

| Grade | Percentage |

**Note:** Grades will not be curved. Your grade is based on your own effort and progress, not based on competition with your classmates.

**Note for graduate students:** A grade of **D+** or below will automatically be marked as an **F**.

## Academic Integrity

We encourage you to interact with other students while doing assignments in class, and in some cases, may be required to work with one another. However, any assignments or homeworks you turn in for grading must represent your own work – it may not be copied from another person's work. Never use a piece of code unless you know the exactly how it works. If we suspect plagiarism, we will ask you to come in and explain your answers or writing.

The official Rutgers policy on cheating can be found here.

There are at least 5 categories of violations: cheating, fabrication, plagiarism, denying others access to information or material, and facilitating Violations of Academic Integrity. Students who violate the Rutgers Integrity policies will be reported to the Office of Student Conduct. Sanctions will be determined by the Office of Student Conduct according to the procedures described in the University Policy on Academic Integrity.

## Student-Wellness Services

**Just in Case Web App**
The web app can be found here. It provides access to helpful mental health information and resources for yourself or a friend in a mental health crisis on your smartphone or tablet and easily contact CAPS or RUPD.

**Counseling, ADAP & Psychiatric Services (CAPS)**
(848) 932-7884
17 Senior Street, New Brunswick, NJ 08901
rhscaps.rutgers.edu/

CAPS is a University mental health support service that includes counseling, alcohol and other drug assistance, and psychiatric services staffed by a team of professional within Rutgers Health services to support students' efforts to succeed at Rutgers University. CAPS offer a variety of services that include: individual therapy, group therapy and workshops, crisis intervention, referral to specialists in the community and consultation and collaboration with campus partners.

**Violence Prevention & Victim Assistance (VPVA)**
(848) 932-1181
3 Bartlett Street, New Brunswick, NJ 08901

[vpva.rutgers.edu](vpva.rutgers.edu)

The Office for Violence Prevention and Victim Assistance provides confidential crisis intervention, counseling and advocacy for victims of sexual and relationship violence and stalking to students, staff and faculty. To reach staff during office hours when the university is open or to reach an advocate after hours, call 848-932-1181.

**Disability Services**
(848) 445-6800
Lucy Stone Hall, Suite A145, Livingston Campus, 54 Joyce Kilmer Avenue, Piscataway, NJ 08854
[https://ods.rutgers.edu/](https://ods.rutgers.edu/)

The Office of Disability Services works with students with a documented disability to determine the eligibility of reasonable accommodations, facilitates and coordinates those accommodations when applicable, and lastly engages with the Rutgers community at large to provide and connect students to appropriate resources.

**Scarlet Listeners**
(732) 247-5555
[http://www.scarletlisteners.com/](http://www.scarletlisteners.com/)

Free and confidential peer counseling and referral hotline, providing a comforting and supportive safe space.