Quantitative and Computational Biology: November 2nd – December 9th 2026

Course Title: Quantitative and Computational Biology
Course Number: G420
Credits: 1
Course Instructors: Richard Koche kocher@mskcc.org; Maria Sirenko Maria.a.sirenko@gmail.com
Jacob Gutierrez gutierj6@mskcc.org; Sahil Sahni sahnis1@mskcc.org
Course Prerequisites: Introduction to R (G450)
Grading Policy: Pass/Fail

Course Structure: The class will be held 1:30-3pm each Monday and Wednesday from Nov 2nd to Dec 9th, with the exception of Nov 23 and 25 (no class Thanksgiving week).

Summary: Students in this course will learn to apply quantitative exploratory data analysis techniques to different forms of experimental data. **The focus of this course is not coding but rather the hands-on learning of analysis modules which can form the basis of future learning.**

The instructors are available and happy to assist with installing or using R as needed. The course will begin with a review of computing in the R programming language. The course will then cover visualization of The Cancer Genome Atlas (TCGA) data, analysis of bulk and single cell RNA-sequencing data, and conclude with epigenomic and integrative analysis. Students will become comfortable performing exploratory data analysis and will understand how concepts from statistics underlie the tools they use. Overall, the goal of this course is to serve as a practical primer for various bioinformatic analyses and should provide students with the foundation for future self-guided learning and skill acquisition in this discipline. These skills will enable them both to collaborate effectively with computational biologists, as well as begin to carry out their own computational experiments.

Project: The goal of the course project is for you to apply the skills and concepts learned in the course. There are no formal requirements, but we hope that you will challenge yourself to work towards the goal you stated in the pre-course survey. The instructors will be available during the course and office hours to identify suitable project ideas or datasets, and to help with the analysis. Presentations will be short (3-5 minutes, ~5 slides). You will explain the question you tried to answer and what approaches you used in the analysis. **Project presentations are scheduled for Wednesday Dec 9th, 2026.**

Assigned reading/exercises: To facilitate a shared baseline familiarity, we ask that you complete the following self-learning tasks before the course starts. **We will begin teaching the course assuming you have completed these exercises and have a basic familiarity with programming in R.** Please reach out to the instructors if you need any help working through these!

1. **Join our qcbio Slack group** to ask questions / discuss the materials **(Required)**
    - We will send the link once the semester begins.
    - Please consider the slack group as an ongoing office hour - we are here to answer your questions re: the assignments and provide help as needed
2. **Install R and RStudio** (For help with this, check Chapter 2 below) **(Required)**
3. **Computational Genomics with R** https://compgenomr.github.io/book/ **(Required - reading and exercises)**
    - Chapter 1: Intro to Genomics
    - Chapter 2: Intro to R programming
    - Chapter 3: Statistics
    - Chapter 6 & 7: Genomic Ranges and FASTQs (read the content but coding aspect is optional)
    - Chapter 8: RNA-seq (Read the content, coding highly recommended)

4. **Intro to ggplot2 (Plotting in R) (Required)**
    - **https://open.oregonstate.education/computationalbiology/chapter/plotting-data-and-ggplot2/**
5. **Part 1** from this course: **Unix and Perl Primer for Biologists** (*Optional, but highly recommended*) http://korflab.ucdavis.edu/Unix_and_Perl/ . This course is designed for biologists with no bioinformatics experience to learn the basics of programming using the terminal.
    - PDF of course materials: http://korflab.ucdavis.edu/Unix_and_Perl/current.pdf (Part 1 is on Pages 1-48)
    - You may prefer to download all the course materials here: http://korflab.ucdavis.edu/Unix_and_Perl/current.zip which includes the example data files.
    - If you are using a Windows PC - please slack or email the instructors - see #1)
6. If you prefer to watch videos rather than read the book in (3): **YaRrr! Pirates Guide to R** (Chapters 1-6 video series - *optional*)
    - https://www.youtube.com/playlist?list=PL9tt3I41HFS9gmeZFEuNrnu_7V_NFngfJ
    - Accompanying Book: https://bookdown.org/ndphillips/YaRrr/ (chapters numbered differently from video series)
    - Optional but highly recommended:
        i. Read the remaining chapters in **YaRrr! Pirates Guide to R** and/or watch the remaining videos here: https://www.youtube.com/playlist?list=PL9tt3I41HFS8-A8V82yicRIl0uSb9mwEU
        ii. In particular: Chapters 10-15 from the book
        iii. Complete the exercises at the end of each chapter

	Topics
Week 1 (Nov 2, 4)	R basics + Plotting TCGA AML data ● Review Unix and R ● Visualization / plotting TCGA AML ● Survival analysis ● Statistics review Bulk RNA-seq Part I ● Alignment ● Quality Control ● Counting ● Normalization ● PCA ● Batch effects ● Clustering
Week 2 (Nov 9, 11)	Single Cell RNA-seq Part 1 ● Differences from bulk RNA-seq ● Normalization ● PCA ● tSNE/UMAP visualization ● Clustering ● Identifying cluster markers

Week 3 (Nov 16, 18)	Single Cell RNA-seq Part 2 ● Differences from bulk RNA-seq ● Normalization ● PCA ● tSNE/UMAP visualization ● Clustering ● Identifying cluster markers
Week 4 (Nov 30, Dec 2)	Bulk/sc RNA-seq downstream analysis ● Review of RNA-seq workflow ● Differential gene expression ● Differential splicing ● GO terms ● GSEA ● Pathway analysis Epigenomics and integrative analysis Part 1 ● Intro to ChIP-seq, ATAC-seq, and scATAC-seq ● Working with data ● Peak calling ● Normalization ● Using Integrative Genomics Viewer (IGV)
Week 5 (Dec 7)	Epigenomics and integrative analysis Part 2 ● Integrative epigenomic analysis ● Motif signatures ● Intersection with other technologies Additional lecture material TBD
Wednesday Dec 9th	Final Project Presentations

Teaching Fellows

Teaching Fellows, drawn from senior GSK students and the postdoctoral community at MSK, are present in the course sessions. Their role is to act as an additional source of information/assistance, to help keep the discussion sessions moving.

Assignments and Methods for Assessing Student Achievement

This course is graded on a pass/fail basis; class participation, assignments and a final project will form the basis of the grading.

Course Evaluation

Students are expected to complete surveys regarding the lectures and overall course via their student portal. This feedback will be used to evaluate the effectiveness and relevance of the topics and provide direction for the subsequent iterations of the course.

Academic Dishonesty, Plagiarism and Artificial Intelligence

The Policy can be found in the Student and Faculty Handbook linked on the GSK Website.