QBIO 490: Directed Research – Multi-omic Data Analysis

General Information
Time: ██████████████████

Faculty Instructors: ██
██████████

Course Director: ███████████████████████
Course Instructors: ██
Teaching Assistants: ██
████████████████████████████

Office Hours: ████████████████████████
█████████████████████████
█████████████████████████
█████████████████████████
█████████████████████████
█████████████████████████

Course Sign Ups
To sign up for this course, please fill out ██████████████████████████████ **Additionally,**
please reach out to ██ **to**
request D-Clearance.

Course Description
This 2-unit course is designed to introduce students to computational biology research through
hands-on bioinformatics projects. Students will learn coding and research skills that will allow
them to more easily transition into a computational research lab. Skills are taught through an
analysis of publicly available data of cancer patients through The Cancer Genome Atlas (TCGA)
and the Clinical Proteomic Tumor Analysis Consortium (CPTAC). *Students will explore the*
effects of various clinical factors on the molecular profiles of cancer. This course was founded
by a QBIO grad–now Keck medical student–and is still student taught, with faculty support from
QCB.

Our overall goal is to provide a hands-on introduction to bioinformatics research and give
students the tools they need to transition from classes to research. Past students have gone on
to many different labs!

Course Learning Objectives

1. Scientific communication
   ○ Interpret and present the main findings of current literature.
   ○ Effectively present original scientific results.
   ○ Practice using scientific writing techniques to write a scientific paper.
2. Foundational skills in QBIO research

- Understand the central dogma of molecular biology and associated data types.
- Use Linux commands to navigate the file system.
- Gain a foundational understanding of Github for managing software.
- Access data from the Genomic Data Commons (GDC) in R and Python.
- Create a job submission on the QCB cluster.

3. Skills in R
   - Use clinical files to learn basic R commands in RStudio.
   - Understand the structure of MAF and HTSeq files.
   - Use `maftools` to create basic plots from MAF files.
   - Perform a differential expression analysis in R using `DESeq2`.
4. Skills in Python
   - Access proteomic data in Python to learn basic Python commands and how to use Jupyter Notebooks.
   - Implement basic machine learning approaches to supervised and unsupervised learning using commonly used libraries.
   - Learn and utilize Python data visualization tools to effectively communicate scientific concepts with an emphasis on multi-omic analysis.

Note about QBIO 490 Credit
This course is two units of QBIO 490 and can only be taken once. **If preferred, students can still take this course without receiving credit**.

Materials
- Personal laptop, Mac or PC is fine!
- No textbook

Prerequisites
None! This course serves as an introduction to coding and data analysis. If you are already comfortable in R and Python, we recommend you transition directly into a research lab rather than taking this course.

Grading Policy
40% Attendance and Participation
30% Homework
15% Final Project
5% R Review Project
5% Python Review Project
5% Literature Presentation

Student Expectations - Attendance and Participation
Twice a week, 50 minute sessions are *mandatory*. If you will be absent, you must email the course instructor(s) to inform them of your absence prior to the session. For participation, we

encourage students to participate as much as possible. This course is unique as it is peer taught! We hope this establishes an open environment that is conducive to learning.

Student Expectations - Homework
This class requires around 6 hours of outside work per week. Some of this will be individual work and some will be group assignments. Most assignments will be coding, though there are additional written and presentation components to strengthen students' skills in scientific communication. These assignments provide the opportunity to put into practice the skills taught in the in-person sessions. Most of the learning is done through these hands-on activities!

Student Expectations - Concept Reviews and Final Project
The R and Python Review Projects and Final Project are data analysis projects that explore publicly available data. The Final Project will be coded in R and Python. The R and Python Review projects are individual projects, while the Final Project is a group project with a written and presentation component. The Final Project will be presented to fellow students as well as a QCB faculty audience!

Academic Integrity
The University of Southern California is foremost a learning community committed to fostering successful scholars and researchers dedicated to the pursuit of knowledge and the transmission of ideas. Academic misconduct is in contrast to the university's mission to educate students through a broad array of first-rank academic, professional, and extracurricular programs and includes any act of dishonesty in the submission of academic work (either in draft or final form).

This course will follow the expectations for academic integrity as stated in the USC Student Handbook. All students are expected to submit assignments that are original work and prepared specifically for the course/section in this academic term. You may not submit work written by others or "recycle" work prepared for other courses without obtaining written permission from the instructor(s). Students suspected of engaging in academic misconduct will be reported to the Office of Academic Integrity.

Other violations of academic misconduct include, but are not limited to, cheating, plagiarism, fabrication (e.g., falsifying data), knowingly assisting others in acts of academic dishonesty, and any act that gains or is intended to gain an unfair academic advantage.

The impact of academic dishonesty is far-reaching and is considered a serious offense against the university and could result in outcomes such as failure on the assignment, failure in the course, suspension, or even expulsion from the university.

For more information about academic integrity see the student handbook or the Office of Academic Integrity's website, and university policies on Research and Scholarship Misconduct.

Policy on the Use of AI Generators

In this course, the use of artificial intelligence (AI)-powered programs is permitted on coding assignments but not written ones. You should be aware that AI generation tools may present incorrect information, biased responses, and incomplete analyses, and while your code may run, it will likely not be optimized nor will it be up to the standards of this course. To adhere to our university values, you must cite any AI-generated material included or referenced in your work and provide the prompts used to generate the content. Please use APA style citations. See the USC LibGuides resource for further citation assistance. Using an AI tool to generate content without proper attribution will be treated as plagiarism and reported to the Office of Academic Integrity.