BIMM 143: **Bioinformatics Lab** Spring 2023

Version: 2023-03-26

Course Basic Information

Website	https://marcos-diazg.github.io/BIMM143_SP23/
Schedule	Wed / Fri 12 – 3 pm YORK 3050 (map)
Office hours	Weekly zoom (link) at a time to be determined from student polling (check class website for up-to-date details)
Instructor	Dr. Marcos Díaz-Gay (mdiazgay@ucsd.edu)
IA	Josef Reimon Rodriguez Urrete (jurrete@ucsd.edu)
Materials	Adapted from Prof. Barry J. Grant

Course Description

Bioinformatics can be defined as the application of computational and analytical methods to complex biological problems. Bioinformatics is a rapidly maturing field that is driving the collection, analysis, and interpretation of the avalanche of data in modern life sciences and medical research.

This upper division 4-unit course is designed for biology majors and introduces the principles and practical approaches of bioinformatics as applied to genes and proteins. An integrated lecture/lab structure with hands-on exercises and small-scale projects emphasizes modern developments in genomics and proteomics. A comprehensive website containing all reading materials, complementary videos, and course notes will be maintained throughout the term.

By completing this course, you will be able to apply leading existing bioinformatics tools to address complex biological questions. Our broader goal is to point towards perspectives that bioinformatics can expose for the integration and analysis of complex biological information.

Audience

Biology majors with upper division standing. Familiarity with basic biomedical concepts is essential (students should have completed BILD 1 and BILD 4 or BIMM 101). No formal programming training or high-level mathematical skills are required.

Requirements

To fully participate in this course, students will be expected to use their own computers to access both class material and bioinformatics software and data resources that are freely available online (mainly based on the R framework). Please check the class website for a computer setup guide.

Course Format

The format of the course for Spring 2023 will be a mix of in-person and online (blended/hybrid learning). Each week, the class will meet in-person on Wednesday and Friday at 12 - 3 pm in **YORK 3050** (map). Please check the course website for schedule updates*. In addition, supporting video lectures and screencast hands-on lab sessions will be available on a weekly basis.

Office hours

In person and Zoom-based (link) office hours will be available on a weekly basis at a time to be determined from student polling. Again, check the class website for full up-to-date details. Office hours are a time when you can come ask clarifying questions about the course material, chat about research, career goals, or about any other topics!

Class Announcements and Communication

All announcements regarding the course will be by email to your UCSD address. We will also be using Piazza to facilitate course communication, particularly around questions and answers. If you have a question outside of class or office hours, first check if it has already been asked on Piazza and, if not, post it there. If you have a question or concern you don't feel comfortable posting on Piazza, feel free to reach out via email.

Textbook

There is no textbook for the course. Lecture notes, homework assignments, grading criteria, video lectures, hands-on session screencasts, and required reading material will be available asynchronously from our public-facing course website.

Course Learning Outcomes

By the end of this course, you will be able to:

- Examine and extract information from online bioinformatics resources, including major biomolecular and genomic databases, search and analysis tools, genome browsers, and structure viewers.
- Select the appropriate bioinformatics tools and resources for quality control and analysis to solve complex problems in the biological sciences and medical research.
- Develop bioinformatics pipelines that collect, analyze and provide insightful visualizations for helping in the interpretation of the results obtained using publicly available sequencing data.
- Recognize the research objectives of the bioinformatics-related sub-disciplines of Genome informatics, Transcriptomics, and Structural informatics.

In short, students will develop a solid foundational knowledge of bioinformatics and be able to evaluate new biomolecular and genomic information using state-of-the-art bioinformatic tools and resources.

Course Schedule

This is a tentative course schedule. The flow of topics might change slightly depending on how quickly/slowly it feels right to progress through them.

Week 1

Introduction to bioinformatics and key online bioinformatics resources

Biology is an information science. History of Bioinformatics. Types of data. Application areas. Introduction to upcoming segments. Hands-on with major Bioinformatics databases and key online NCBI and EBI resources.

Week 2

Sequence alignment, DNA, and protein database searching

Homology. Sequence similarity. Local and global alignment. Database searching with BLAST. Using PSI-BLAST, Profiles and HMMs. Protein structure comparisons.

Bioinformatics data analysis with R

Why do we use R for bioinformatics? R language basics and the RStudio IDE. Major R data structures and functions. Using R interactively from the RStudio console.

Week 3

Data exploration and visualization in R

The exploratory data analysis mindset. Data visualization best practices. Simple base graphics (including scatterplots, histograms, bar graphs, dot chats, boxplots, and heatmaps). Building more complex charts with ggplot.

Why, when and how of writing your own R functions

The basics of writing your own functions that promote code robustness, reduce duplication, and facilitate code re-use. Extending functionality and utility with R packages from CRAN and Bioconductor. Working with Bio3D for molecular data.

Week 4

Machine learning for bioinformatics

Unsupervised learning. K-means clustering. Hierarchical clustering. Heatmap representations. Dimensionality reduction. Principal Component Analysis (PCA). Hands-on project session with unsupervised learning analysis of cancer cells.

Week 5

Structural Bioinformatics

Protein structure function relationships. Protein structure and visualization resources. Modeling energy as a function of structure. Homology modeling. AlphaFold. Predicting functional dynamics. Inferring protein function from structure.

Week 6

Genome informatics and high-throughput sequencing

Searching genes and gene functions. Genome databases. Variation in the Genome. High-throughput sequencing technologies, biological applications, and bioinformatics analysis methods. The Galaxy platform along with resources from the EBI & UCSC. Hands-on analysis of cancer driver mutations and mutational signatures with R. NCI Genomic Data Commons, cBioPortal, and COSMIC databases. Cancer genomics visualizations.

Week 7

Transcriptomics and the analysis of RNA-Seq data

RNA-Seq aligners. Differential expression tests. RNA-Seq statistics. Counts and FPKMs. Avoiding P-value misuse. Hands-on analysis of RNA-Seq data with R. Gene function annotation. Functional databases (KEGG, InterPro, GO ontologies) and functional enrichment analysis.

Week 8

Hands-on with git and GitHub

Why you should use a version control system. How to perform common operations with Git. Creating and working with your own GitHub repos and navigating and using those of others.

Unix for bioinformatics

Bioinformatics on the command line. Understanding processes. File system structure. Connecting to remote servers. Redirection, streams and pipes. Workflows for batch processing. Organizing computational projects.

Week 9

Mini projects

Topical mini projects based on publicly available data. Query data from biological databases. Web-scraping. Advanced dplyr and ggplot. Practical considerations and best practices for exploratory analyses. Development of bioinformatics pipelines to analyze biological data using state-of-the-art tools and generating complex visualizations.

Week 10

Course wrap-up and portfolio building

Course summary and review. Making a public-facing GitHub pages portfolio of your bioinformatics work. Student course evaluation time. Find a gene project assignment due.

Homework assignments and project

Weekly homework will consist of online knowledge assessment quizzes and application assignments (a.k.a. "hands-on lab sections") together with pre-class reading and short video lectures.

Specific grading criteria (assessment rubrics) for each homework will be given at the time of assignment. Weekly grades will be posted online. Each student is responsible for checking to ensure that a grade has been entered for their submissions. Documents submitted by email do not always arrive at their intended destination, and late submissions will not be accepted after one week past the original due date. Collectively, **homework** performance will account for **85%** of the course grade.

A total of **15%** of the course grade will be assigned based on the ***find-a-gene project assignment***. The purpose of this project assignment is for you to grasp the principles of database searching, sequence analysis, functional annotation, and exploratory data analysis with R that we cover in the course. Further details will be given in class.

There will be no final exam for the Spring 2023 quarter. I understand this is a challenging time and that you may have difficulties with accessing the course material and adapting to in-person learning. I also understand the stress and uncertainty of our current daily reality. My goals are to reach, teach and engage you in the course material. I also aim to fairly test your knowledge of this material, and grade you accordingly while keeping these challenges in mind.

The following grading scheme will be used (cutoffs may be lowered at the instructor's discretion). The course is not graded on a curve (i.e., 20% of students getting A, B, C, and such). Thus, the ability to do well in this course is not dependent on others doing poorly.

A+	100-97%	A	97-93%	A-	93-90%
B+	90-87%	B	87-83%	B-	83-80%
C+	80-77%	C	77-73%	C-	73-70%
D	70-60%	F	< 60%		

Accommodations

Students requesting accommodations and services due to a disability for this course need to provide a current Authorization for Accommodation (AFA) letter issued by the Office for Students with Disabilities (OSD), prior to eligibility for requests. Receipt of AFAs in advance is necessary for appropriate planning for the provision of reasonable accommodations. Please note that instructors are unable to provide accommodations unless they are first authorized by OSD. For more information, contact the OSD at (858) 534-4382 (voice), osd@ucsd.edu, or visit osd.ucsd.edu.

Diversity and inclusion

I would like to create a learning environment that supports a diversity of thoughts, perspectives and experiences, and honors your identities (including race, gender, class, sexuality, religion, ability, etc.) To help accomplish this:

- If you have a name and/or set of pronouns that differ from those that appear in your official UCSD records, please let me know!
- If you feel like your performance in the class is being impacted by your experiences outside of class, please don't hesitate to contact me. I want to be a resource for you. Remember that you can also submit anonymous feedback (which will lead to me making a general announcement to the class, if necessary, to address your concerns).
- I (like many people) am still in the process of learning about diverse perspectives and identities. If something was said in class or online (by anyone) that made you feel uncomfortable, please talk to me about it. Again, anonymous feedback is always an option.
- As a participant in course discussions, you should also strive to honor the diversity of your classmates.
- Please contact me (in our office hours or electronically) or submit anonymous feedback if you have any suggestions to improve the quality of the course materials or delivery mechanisms.

Ethics code

You are encouraged to collaborate with your fellow students. However, all material submitted must be your own work.

"Academic Integrity is expected of everyone at UC San Diego. This means that you must be honest, fair, responsible, respectful, and trustworthy in all of your actions. Lying, cheating or any other forms of dishonesty will not be tolerated because they undermine learning and the University's ability to certify students' knowledge and abilities. Thus, any attempt to get, or help another get, a grade by cheating, lying or dishonesty will be reported to the Academic Integrity Office and will result in sanctions.

Sanctions can include an F in this class and suspension or dismissal from the University. So, think carefully before you act. Before you act, ask yourself the following questions: a) is my action honest, fair, respectful, responsible & trustworthy and, b) is my action authorized by the instructor? If you are unsure, don't ask a friend—ask your instructor, instructional assistant, or the Academic Integrity Office".

You can learn more about academic integrity at academicintegrity.ucsd.edu (Source: UCSD Academic Integrity Office, 2023).

Resources for support and learning

There are a variety of resources available to students at UC San Diego. If you need help in any capacity, please reach out to any of the references below.

Basic Needs at UCSD
Any student who has difficulty accessing sufficient food to eat every day, or who lacks a safe and stable place to live is encouraged to contact: foodpantry@ucsd.edu | basicneeds@ucsd.edu | (858) 246-2632

Counseling and Psychological Services
Confidential counseling and consultations for psychiatric service and mental health programming.
(858) 534-3755

Campus Community Centers
As part of the Office of Equity, Diversity, and Inclusion, the campus community centers provide programs and resources for students and contribute toward the evolution of a socially just campus.
diversity@ucsd.edu | (858) 822-3542

Office for Students with Disabilities
Supports students with disabilities and accessibility across campus.

Get Involved
Student organizations, clubs, service opportunities, and many other ways to connect with others on campus.

Undocumented Student Services
Programs and services are designed to help students overcome obstacles that arise from their immigration status and support them through personal and academic excellence.

* Important notes on COVID-19

At the time of writing, we are still in a major COVID-19 pandemic crisis. While we are making every effort to support your learning in a safe in-person environment, we expect and require your help. Keeping our classroom community healthy takes all of us following campus safety requirements. You are expected to follow the university's public health requirements and pursue personal protection practices to protect yourself and the others around you from infection. These include:

- **Participating in the university's daily screening process**. Everyone must complete a Daily Symptom Screener.
- **Participating in the university's testing program**. All students can benefit from the COVID-19 Testing program in accordance with their vaccination status.
- **Wearing a well-fitting mask** that covers your nose and mouth if you have been exposed to COVID-19 or are having symptoms but test negative, according to the

current campus recommendations. I encourage anyone who feels more comfortable continuing to wear a mask to do so, as masking remains one of the easiest and most effective ways to reduce the spread of the virus.

- **Stay home if you're feeling ill.** If you're not feeling well, complete the symptom screener and, if needed, get tested for COVID-19. *Do not come to campus unless given the all-clear*. When you are able, please contact me, and I will make every effort to support your continued learning. We are all susceptible to COVID-19 illness-related disruptions, and I have designed this course with pandemic resilience in mind.

I will do everything possible to help ensure that instruction is as safe as possible. This means that if conditions change and in-person instruction is no longer deemed as safe as it could be by the university and public health officials, we will pivot immediately to remote teaching. My hope is that this course can remain in-person and, regardless of delivery mode, serve as a venue where we build a vibrant, supportive learning community. The following UCSD website has many useful resources on returning to in-person learning.