BIMM 143: Introduction to Bioinformatics (Winter 2026)

Course Instructor: Dr. Barry J. Grant (bjgrant@ucsd.edu)
Course Website: https://bioboot.github.io/bimm143_W26/

***N.B.** This class will be offered in-person for the Winter 2026 quarter*.*
***Version:** 2025-12-09 (09:39:35 PST on Tue, Dec 09)*

Overview:
Bioinformatics - the application of computational and analytical methods to biological problems - 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 provides an introduction to 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. Major topics include: Genomic and biomolecular bioinformatic resources, Advances in sequencing technologies; Genome informatics, Structural informatics, and Transcriptomics. Computational tools, techniques and best practices that foster reproducible bioinformatics research will also be introduced. A comprehensive website containing all reading materials, screencasts and course notes will be maintained throughout the term.

Students completing this course will be able to apply leading existing bioinformatics tools to address 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. A familiarity with basic biomedical concepts is essential (students should have successfully completed BILD1 and BILD4 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.

Schedule:
For Winter 2026 we are offering two sections of BIMM-143. Both sections cover identical material and meet in-person.

Section 1 meets on Tuesday and Thursday at 9:30 - 12:30 pm in **TATA 2501** (Map).
Section 2 meets on Wednesday and Friday at 10:00 - 1:00pm in **York 3050** (Map).

Please check our class website for schedule updates*. Supporting video lectures and screencast hands-on lab sessions will be available on a weekly basis. See the class website for full details.

Class announcements:
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 there. If you have a question or concern you don't feel comfortable posting on Piazza feel free to reach out via email.

***Important notes on staying home if you're feeling ill!**
If you're not feeling well, you should stay home and get tested for COVID-19. When you are able please contact me and I will make every effort to support your continued learning. We are all susceptible to Flu and COVID-19 illness-related disruptions and I have designed this course with resilience in mind.

Office hours:
We will use walk-in (*i.e.* in person) and Zoom on a weekly basis at a time to be determined from student polling. At the time of writing this is set to Thursday at 12:30pm San Diego time. Again, see the class website for full up-to date details.

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.

Piazza "clubhouse" forum:
We will use Piazza to help facilitate course communication. If you have course related questions outside of class time first check if they have already been asked on Piazza and if not post there. I also encourage students to post suggestions and strategies that can help their classmates. Feel free to post something funny that relates to our course materials, your study skill suggestions, memes about our topics, videos, and, yes, the occasional complaint. My hope is that this "clubhouse" forum will allow us to support one another, build camaraderie even outside of traditional in-person class time.

Course scope and syllabus outline:
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.

N.B. Any schedule changes will be posted on the course website and to Piazza. Please make sure to frequently check the class website to keep updated.

Week 1

Introduction to Bioinformatics and Key Online Bioinformatics Resources: NCBI & EBI

Biology is an information science, History of Bioinformatics, Types of data, Application areas: Introduction to upcoming segments, NCBI & EBI resources for the molecular domain of bioinformatics, Focus on GenBank, UniProt, Entrez and Gene Ontology.

Week 2

Sequence Alignment, DNA and Protein Database Searching

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

Week 3

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. Import biomolecular data in various formats (both local and from online sources).

Week 4

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.

Week 5

Writing your own R functions and working with R packages for bioinformatics

Using R scripts and Quarto/Rmarkdown files, Import data in various formats both local and from online sources, The basics of writing your own functions that promote code robustness, reduce duplication and facilitate code re-use, Obtaining R packages from CRAN and Bioconductor, Working with Bio3D for molecular data, Managing genome-scale data with bioconductor..

Week 6

Machine learning for bioinformatics

Unsupervised learning, K-means clustering, Hierarchical clustering, Heatmap representations. Dimensionality reduction, Principal Component Analysis (PCA). Longer hands-on session with unsupervised learning analysis of cancer cells further highlighting practical considerations and best practices for the analysis and visualization of high dimensional datasets.

Week 7

Genome informatics and high throughput sequencing

Searching genes and gene functions, Genome databases, Variation in the Genome, High-throughput sequencing technologies, biological applications, bioinformatics analysis methods. The Galaxy platform along with resources from the EBI & UCSC.

Week 8

Transcriptomics, RNA-Seq analysis, and the interpretation of gene lists

RNA-Seq aligners, Differential expression tests, RNA-Seq statistics, Counts and FPKMs and 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 9

Structural Bioinformatics (AlphaFold)

Comparative structure and sequence analysis in R. The importance of Multiple Sequence Alignments (MSAs). Combining knowledge based and physics based approaches for structure prediction and modeling functional motions, AlphaFold2, LLMs and the new age of structural biology.

Week 10

Course wrap up, project completion

Summary of learning goals, Student course evaluation time; Find a gene assignment due. Open study or a student selected topic from those below:

Biological network analysis

Network based approaches for integrating and interpreting large heterogeneous high throughput data sets; Discovering relationships in 'omics' data; Network construction, manipulation, visualization and analysis; Major graph theory and network topology measures and concepts. Hands-on with Cytoscape and igraph packages.

Cancer genomics

Cancer genomics resources and bioinformatics tools for investigating the molecular basis of cancer. Mining the NCI Genomic Data Commons; Immunoinformatics and immunotherapy; Using genomics and bioinformatics to help design a personalized cancer vaccine. Implications for personalized medicine.

Hands-on with git and GitHub

Hands-on introduction to git, currently the most popular version control system. We will learn how to perform common operations with git and RStudio. We will also cover the popular social code-hosting platforms GitHub, BitBucket and GitLab.

Essential statistics for bioinformatics

Review of data summary statistics; Inferential statistics; Significance testing; Two sample T-test; Power analysis; Multiple testing correction; and almost everything you wanted to know about p-values but were afraid to ask! Extensive R examples and applications.

Essential Unix for bioinformatics

Bioinformatics on the command line, Why do we use UNIX for bioinformatics? UNIX philosophy, 21 Key commands, Understanding processes, File system structure, Connecting to remote servers, Redirection, streams and pipes, Workflows for batch processing, Organizing computational projects.

Bioinformatics for systems vaccinology

Systems vaccinology uses "multi-omics" approaches combined with bioinformatics and computational modeling to provide insights into vaccine-induced immunity and vaccine efficacy. Focusing on either Pertussis (whooping cough) or Influenza (flu), students will analyze CDC tracking data and then analyze pre- and post-vaccination multi-omic datasets from the NIH-funded Computational Models of Immunity (CMI-X) project. This is followed by a modeling section that demonstrates how pre-vaccination data can be used to predict post-vaccination immune responses.

<u>The future of bioinformatics</u>
Applications of bioinformatics to translational medicine and the social impacts and ethical implications of how genomic sequence information is used in society. Example topics include: How to best balance privacy concerns with the desire to share and re-use data generated by taxpayer funded research? Should doctors know and preemptively act based on our genetic disease predispositions? How are racial DNA differences impacting racial prejudices?

Course Objectives:
At the end of this course students will:

- Understand the increasing necessity for computation in modern life sciences research.
- Be able to use and evaluate online bioinformatics resources including major biomolecular and genomic databases, search and analysis tools, genome browsers, structure viewers, and select quality control and analysis tools to solve problems in the biological sciences.
- Understand the process by which genomes are currently sequenced and the bioinformatics processing and analysis required for their interpretation.
- Be familiar with 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 existing bioinformatic tools and resources.

Specific Learning Goals
Teaching toward the specific learning goals below is expected to occupy 60%-70% of class time. The remaining course content is at the discretion of the instructor with student body input. This includes student selected topics for peer presentation as well one student selected guest lecture from an industry based genomic scientist.

All students who receive a passing grade should be able to:

1 Appreciate and describe in general terms the role of computation in hypothesis-driven discovery processes within the life sciences.

2 Be able to query, search, compare and contrast the data contained in major bioinformatics databases and describe how these databases intersect (GenBank, GENE, UniProt, PFAM, OMIM, PDB, UCSC, ENSEMBLE).

3 Describe how nucleotide and protein sequence and structure data are represented (FASTA, FASTQ, GenBank, UniProt, PDB).

4 Be able to describe how dynamic programming works for pairwise sequence alignment and appreciate the differences between global and local alignment along with their major application areas.

5 Calculate the alignment score between two nucleotide or protein sequences using a provided scoring matrix and be able to perform BLAST, PSI-BLAST, HMMER and protein structure based database searches and interpret the results in terms of the biological significance of an e-value.

6 Use R to read and parse comma-separted (.csv) formatted files ready for subsequent analysis.

7 Perform elementary statistical analysis on boimolecular and “omics” datasets with R and produce informative graphical displays and data summaries.

8 Appreciate and describe in general terms the rapid advances in sequencing technologies and the new areas of investigation that these advances have made accessible.

9 Understand the process by which genomes are currently sequenced and the bioinformatics processing and analysis required for their interpretation.

10 For a genomic region of interest (e.g. the neighborhood of a particular gene), use a genome browser to view nearby genes, transcription factor binding regions, epigenetic information, etc.

11 Given an RNA-Seq data file, find the set of significantly differentially expressed genes and use online tools to interpret gene lists and annotate potential gene functions.

12 Perforn a GO analysis to identify the pathways relevant to a set of genes (e.g. identified by transcriptomic study or a proteomic experiment).

13 Use the KEGG pathway database to look up interaction pathways.

14 Understand the challenges in integrating and interpreting large heterogenous high throughput data sets into their functional context.

15 Have an appreciation for the social impacts and ethical implications of how genomic sequence information is used in our society

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 video screen-casts. All lab and assignment submissions will be via GradeScope.

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 online 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 65% of the course grade.

A total of 35% 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 (see additional details online). Further details will also be given in class.

There will be no final exam for the Winter 2026 quarter.

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 for my students 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 in-person office hors or electronically) or submit anonymous feedback if you have any suggestions to improve the quality of the course materials or delivery mechanisms.

AI Policy:

Generative AI technologies such as ChatGPT, Claude, GitHub Copilot, NoteBookLM and TritonGPT (see next section) are rapidly becoming part of professional practice in bioinformatics. In this course, I will encourage you to engage with these tools, and in some assignments you will be required to use them. Learning to use AI effectively is an emerging skill, and I will provide resources and guidance as required. I am also available to discuss effective and ethical AI use strategies during office hours.

Expectations for Use

Encouraged Use: You may use AI to explore ideas, test knowledge, generate code, improve writing, or visualize concepts. AI can help you experiment, avoid simple errors, and deepen your engagement with the course material.

Critical Responsibility: AI outputs may contain errors, bias, or repetition, and often need refinement. You are responsible for validating information, testing code thoroughly, and ensuring your final submission meets the standards of the course.

Documentation and Transparency is key: For your assignments, including the find-a-gene project, include a short paragraph at the end of your assignment explaining how you used AI and provide the key prompts. For code, include the prompt(s) and significant parts of the AI's response in your comments. Transparency is key: we are learning how to use these tools effectively together.

Things to Keep in Mind

Garbage in, garbage out: Minimal prompts will produce low-quality results. Refining prompts takes thought and practice.

AI is not a fact-checker: Assume that factual information, numbers, or citations may be wrong unless you verify them independently.

AI is not always appropriate: Use it when it enhances your work, but recognize when other tools, resources, and *your own reasoning* are more suitable.

Why We Do This

Our goal as a learning community is not only to master the course content, but also to understand how emerging AI tools can augment and support our research, study, and professional performance. By using AI openly, responsibly, and critically, we are preparing for the environments you will encounter in your academic and professional futures.

Custom TritonGPT AI assistant for BIMM143

This quarter you will have access to a custom AI instructional assistant powered by TritonGPT, our campus-built and secure generative AI. This is powered by OpenAI's leading model and is seeded with custom BIMM143 specific information.

How to Access It:

1. Go to https://tritongpt.ucsd.edu.
2. Click Explore Assistants in the left-hand menu.
3. Scroll and look for "**BIMM143 Grant - Student Assistant**".
4. Click **Pin** to add it to your sidebar for quick access going forward.
5. Ask away, e.g. "what is lab 2 all about?", "what is the answer to q19 in lab 1?", "what is a porphyrin group?", "what are the due dates for assignments?", etc.

What It Can Do:

This custom BIMM143 AI assistant is designed to support your learning in a Socratic, teaching-focused way. It's designed not to just answer your questions, but to help guide your thinking, reinforce course concepts, and clarify material. Its responses are based on our course materials

- which is super cool! Have a look at the examples in this TritonGPT_example.pdf. It can also help you review, quiz yourself, and think through challenging topics and general in-class/ homework problems.

Feedback:

This pilot is focused on exploring responsible and helpful uses of AI in teaching and learning, and your input is essential. Please try out different models (like GPT-5 or LLaMA 3 Scout) to see how their responses compare. Use the thumbs up or thumbs down icons to rate responses and add optional comments when you can. This helps us understand what's working and what isn't.

If you run into any tech issues, please reach out to me and the custom email ai-pilots@ucsd.edu. This will create a ServiceNow ticket and be routed to the appropriate ITS team.

Some important information about privacy & safety to bear in mind:

Nothing you submit is used to train the AI models. I do have access to a general (anonymous) dashboard showing what students are asking, so I can see where people might be struggling and adjust my teaching accordingly. These prompts are never linked to any particular student. If you have any questions or concerns, please don't hesitate to reach out.

Ethics Code:

You are encouraged to collaborate with your fellow students. However, all material submitted to the instructor must be your own work. AI tools can be used to augment and support your work. However any use of AI must be openly acknowledged (see previous section).

> *"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".*

Please do not risk your future by cheating. You can learn more about academic integrity at academicintegrity.ucsd.edu (Source: UCSD Academic Integrity Office, 2024).

Food Support for Students:

If you are skipping and stretching meals, or having difficulties affording or accessing food, you may be eligible for CalFresh, California's Supplemental Nutrition Assistance Program, that can provide up to $292 a month in free money on a debit card to buy food. Students can apply at benefitscal.com/r/ucsandiegocalfresh. The Hub Basic Needs Center empowers all students by connecting them to resources for food, stable housing and financial literacy. Visit their site at basicneeds.ucsd.edu