BIO377: Big Data In Biology (49674)

Group Meetings TTh 11-12:15, FNT 1.104

Website `https://wikis.utexas.edu/display/BDIBstream/Big+Data+In+Biology+Stream`

Course Overview

The Big Data in Biology stream explores methods for analyzing large-scale biologic datasets using computational algorithms, statistical tools, and supercomputers. In the spring semester, you were taught some skills required to analyze large biological datasets. In this semester, you will be utilizing those skills to answer real biological questions using large next generation sequencing datasets.

Course Objectives

This semester, you will work closely with the instructors on their research projects with the following objectives:

- Learn how to work in a research setting.
- Conduct independent research in the field of bioinformatics using the skills acquired in the spring.
- Acquire experience reading publications and making scientific presentations.

Course Expectations

Apart from class meetings you are expected to contribute 6 additional hours per week towards meeting research project goals, divided roughly evenly between individual work and group work. Group meeting times will be decided in the first two weeks of class according to instructor, mentor, and student availabilities.

Grading Rubric

Item	Weight
Participation*	**10%**
- In group meetings and journal club	
Documentation*	**20%**
- Well documented analyses and scripts	10%
- Weekly progress updates	10%
Presentations*	**50%**
- Journal club presentation	10%
- Mid term presentation	10%
- Final presentation	30%
Exercises	**10%**
Keck Industry modules	**10%**

*Participation you are expected to be present at class and group meetings and to contribute to project and journal club discussions and group work.

*Documentation you are expected to provide clear documentation of the analyses you perform in order to facilitate replicability of work. This includes both detailed documentation of any code written (what commands run, including exact arguments, as well as where scripts were run and what versions of software were used) as well as weekly progress update using the provided template.

*Presentations you are expected to present one journal article individually and to present project results in groups both at mid-term and at the end of the term.

Late assignments

Late assignments will be accepted within 3 days of the due date and incur a 10% reduction in points for that assignment each late day up to 30% total.

Appeals

The instructors and TA (if any) will make decisions concerning grades, attendance, and other policy matters. Should you disagree with the TA, you are welcome to take the matter to the instructors or an assistant director of the FRI.

Course Schedule (Subject to Change)

Week	Date	Item
1 - Introduction	Aug 25	Syllabus and Course Overview
	Aug 25	Hans Hofmann lecture
2 - Introduction	Aug 30	GSAF tour
	Sep 1	Project Overview
	Sep 1	**Assigned: Select journal articles**
3 - Programming environment	Sep 6	Set up your environment (jupyter at GSAF)
	Sep 6	Unix and python tricks
	Sep 8	**Assigned: Data Analysis Exercise**
4 - How to give good presentations	Sep 13	Tips and discussion
	Sep 15	Journal club and Keck Unit I Discussion
	Sep 15	**Assigned: Keck Unit I Exercise**
5 - Bioinformatics workflow	Sep 20	**In-class: Workflow diagram excercise**
	Sep 22	Research updates
	Sep 22	**Due: Data Analysis Exercise**
6 - Research work	Sep 27	
	Sep 29	Journal club
7 - Research work	Oct 4	
	Oct 6	Research updates
	Oct 6	**Due: Keck Unit I Exercise**
8 - Group Project Overview	Oct 11	**Mid-term Presentation**
	Oct 13	Journal club
9 - Research work	Oct 18	
	Oct 18	Keck Unit II Discussion
	Oct 20	Research updates
10 - Research work	Oct 25	
	Oct 27	Journal club
11 -Research work	Nov 1	
	Nov 3	Research updates
12 - Research work	Nov 8	
	Nov 10	Journal club
13 - Research work	Nov 15	
	Nov 17	Research updates
	Nov 17	**Due: Keck Report**
14 - Thanksgiving holiday	Nov 22	No class meeting
	Nov 24	No class meeting
15 - Final Presentation	Nov 29	**Present your work**
	Dec 1	**Present your work**

Instructor Of Record

Dhivya Arasappan

Tues 12:30–1:30
GDC 7.422
darasasappan@mail.utexas.edu

Dhivya Arasappan is a CCBB research scientist with 6+ years experience analyzing next-generation sequencing data. Her primary focus has been in transcriptomic profiling and de novo genome assembly.

Instructors

Benni Goetz

Wed 2–3
GDC 7.414D
benni@utexas.edu

Benni Goetz is a bioinformatics consultant at CCBB. He writes custom scripts for analysis of next-generation sequencing data. Recently, transcriptome assembly and annotation, bacterial genome annotation, and next-generation sequence databases have been primary focuses.

Dennis Wylie

Thurs 12:30–1:30
GDC 7.516
denniscwylie@gmail.com

Dennis Wylie focuses on application of statistical and machine learning approaches to the analysis of biological data, especially next-generation sequencing data.

Mentors

Your research team will be assigned a mentor. Your assigned mentor is your first point of contact for questions, help, etc. But you will also be meeting regularly with the instructors to make progress on the research project.

Vinita Gottipati

vinitag.95@gmail.com

Vinita Gottipati is a senior biochemistry major in the Bridging Disciplines Program, and is working to get a CSE certificate. She enjoys statistics as well as topics in healthcare, with specific interests in human genetics and public health.

Santiago Enrique Sanchez

santiago.es@utexas.edu

Santiago Sanchez is a senior Biochemistry and Plan II Honors double major in the Health Science Scholars program. His research background is in physical and computational chemistry as well as cancer biology. Outside of basic science, Santiago is interested in health policy, ethics, and history.

Principal Investigators

Prof. Hans Hofmann

Office Hours By Appointment
PAT 319
hans@utexas.edu

Dr. Hans Hofmann is a Professor of Integrative Biology (`http://cichlid.biosci.utexas.edu/`) and the Director of the Center for Computational Biology and Bioinformatics (`http://ccbb.utexas.edu/`). Recently, Dr. Hofmann has been appointed as the inaugural Director of the new Center for Biomedical Core Facilities at UT Austin. Research in his lab focuses on the neural and molecular basis of social behavior and its evolution.

Prof. Vishy Iyer

Office Hours By Appointment
MBB 3.212A
vishy@utexas.edu

Dr. Vishy Iyer is a Professor in the Department of Molecular Biosciences and Institute for Cellular and Molecular Biology, and co-Director of the Center for Systems and Synthetic Biology. His research interests include the genomics of gene regulation, epigenetics, non-coding RNA, cancer, genetic variation and human disease.

Course Flags

Independent Inquiry

This course carries the Independent Inquiry flag. Independent Inquiry courses are designed to engage you in the process of inquiry over the course of a semester, providing you with the opportunity for independent investigation of a question, problem, or project related to your major. You should therefore expect a substantial portion of your grade to come from the independent investigation and presentation of your own work.

Materials & References

Laptops

A laptop will be very useful for this class. If you do not own a laptop, please come talk to one of the instructors immediately, and we will work accommodate you. UT Libraries have several computer labs on campus that you can use to connect to our dedicated server for this class.

CCBB Short Courses

The Center for Computational Biology and Bioinformatics offers several 3-hour short courses over the course of the semester. Courses cover a variety of topics in biological computing, ranging from an introduction to the Linux command line, to more advanced topics like ChIP-seq and RAD-seq. For more information see: `http://www.ccbb.utexas.edu/shortcourses.html`

References

These aren't required, and we won't be following them, but they may be useful references to get started. But part of this class is exploring, talking to your fellow students, and finding your own sources!

Python `http://www.greenteapress.com/thinkpython/`

Bioinformatics `https://wikis.utexas.edu/display/bioiteam/Home`

Statistics `http://www.greenteapress.com/thinkstats/`

Canvas

We will use the Canvas system for sending and receiving emails and posting important announcements. Please check the Canvas website regularly for course information. You will submit assignments using Canvas. Canvas will serve as the official repository of the BIO 377 scores that will comprise your total BIO 377 final grade. It is your responsibility to ensure that your scores for all of the various BIO 377 assignments are correctly displayed in Canvas. Canvas is accessible at: `http://canvas.utexas.edu/`

University Policies and Notices

University Policy On Scholastic Dishonesty

Students who violate University rules on scholastic dishonesty are subject to disciplinary penalties, including the possibility of failure in the course and/or dismissal from the University. At minimum, the incident will be reported to the Deans office and SJS, where a permanent record of the violating incident will be placed in the students file, and a score of zero will be given for the assignment or exam in question. Since such dishonesty harms the individual, all students, and the integrity of the University, policies on scholastic dishonesty will be strictly enforced. You must read and be familiar with the expected standards of conduct for students (`catalog.utexas.edu/general-information/appendices/appendix-c/student-discipline-and-conduct/#subchapter11400.prohibitedconduct`), as well as University policies on scholastic dishonesty (`deanofstudents.utexas.edu/sjs/acint_student.php`).

Academic Accommodations For Students With Disabilities and Other Needs

The University provides, upon request, appropriate academic accommodations for qualified students with disabilities. Any student with a documented disability who requires academic accommodations should contact the Division of Diversity and Community Engagement, Services for Students with Disabilities at 512-471-6259 (voice) or 512- 410-6644 (Videophone) (or see `http://www.utexas.edu/diversity/ddce/ssd`) as soon as possible to request an official letter outlining authorized accommodations. Accommodations for religious reasons will be made as per University policy, if you notify me at least 14 days prior to your pending absence.

Policy On Religious Observance

The University of Texas at Austin policy is that the student must notify each instructor at least fourteen days prior to the classes scheduled on dates he or she will be absent to observe a religious holy day. For religious holy days during the beginning of the semester, the notice should be given on the first day of the semester. The student may not be penalized for these excused absences but the instructor may appropriately respond if the student fails to notify *prior* to the observance or complete the missed assignment or examination within a reasonable time after the excused absence.

Emergency/Safety Procedures

Familiarize yourself with procedures for emergency evacuation, available at the Office of Campus Safety and Security, 512-471-5767 (`www.utexas.edu/safety`). You can sign up to receive emergency text alerts at `www.utexas.edu/cellphonealert`. Occupants of buildings on the UT Austin campus are required to evacuate buildings when a fire alarm is activated. Alarm activation or announcement requires exiting and assembling outside. Familiarize yourself with all exit doors of each classroom and building you may occupy. Remember that the nearest exit door may not be the one you used when entering the building. Students requiring assistance in evacuation must inform their instructor in writing during the first week of class. In the event of an evacuation, follow the instruction of faculty or class instructors. Do not re-enter a building unless given instructions by the following: Austin Fire Department, UT Austin Police Department, or Fire Prevention Services office. Information on emergency evacuation routes and emergency procedures can be found at `www.utexas.edu/emergency`. The UT Austin Behavior Concerns Advice Line (BCAL) is: 512-232-5050.

Behavior Concerns Advice Line (BCAL)

If you are worried about someone who is acting differently, you may use the Behavior Concerns Advice Line to discuss by phone your concerns about another individual's behavior. This service is provided through a partnership among the Office of the Dean of Students, the Counseling and Mental Health Center (CMHC), the Employee Assistance Program (EAP), and The University of Texas Police Department (UTPD). Call 512-232-5050 or visit `http://www.utexas.edu/safety/bcal`.

Q drop Policy

The State of Texas has enacted a law that limits the number of course drops for academic reasons to six (6). As stated in Senate Bill 1231: "Beginning with the fall 2007 academic term, an institution of higher education may not permit an undergraduate student a total of more than six dropped courses, including any course a transfer student has dropped at another institution of higher education, unless the student shows good cause for dropping more than that number."