

MDSC 403

Computation for Bioinformatics

Instructors:

Quan Long, PhD
quan.long@ucalgary.ca
403-220-5580

Office Hours/Policy on Answering Student Emails

Office Hours: Friday 2pm – 4pm. (Students must arrange a minimum of 3 days in advance to meet.) The meeting could be either in the instructor's office HS 1151 (1173) or a Zoom-based session. Student emails will be answered within 3 business days. Please note that most questions should be posed at the beginning or end of class.

Teaching Assistants:

Qing (Leah) Li
qing.li2@ucalgary.ca

Time and Location:

Thursdays, 3 – 5:50 pm
HSC 1501

Prerequisite/Co-Requisite:

Admission to the BHSc Honours program or permission of Instructor.

Course Description:

This advanced course will focus on current computational and statistical methods in bioinformatics. Topics will shift as computational methods in bioinformatics shift. The underlying philosophy of how to use computational and statistical techniques to solve biological questions will be elaborated using hands-on examples and inquiry-based projects.

Students will also learn how to acquire new techniques in the fast-moving field of bioinformatics by leveraging their 'old-version' skills in conjunction with a general understanding of the field. Currently, the course covers (1) R, as a vehicle for data analysis as well as a real programming language; (2) Statistical techniques frequently used in bioinformatics and data sciences, including hypothesis test, statistical inference, and linear models; and (3) Linux command-line and scripting, including the use of popular bioinformatic tools and high-performance computing (HPC) clusters. The course will build basic skills for both upstream computational processing and downstream statistical characterization. Experienced bioinformaticians will give guest lectures on applying computing techniques to biological projects at the end of the course.

Overarching Theme

Statistics and computation are central to bioinformatics. For many years, statistics and R packages have been seen as essential parts of the toolkit of bioinformaticians. This course will teach basic statistical

models and techniques using R as a demonstrating language. As such, it integrates a theoretical course on statistics and a practical course in R. Additionally, most bioinformatics tools are based on Unix/Linux system. Also, Linux system by itself offered powerful commands that can carry out many *ad hoc* analyses, providing convenient alternatives of writing fully documented programs. Students in this course will develop facility in Linux programming, focusing on widely used commands and approaches. Students will leave this course with the ability to manipulate and analyze data using statistics (R) and computation (Linux), and other bioinformatics tools. There are two formats in this course. The statistical theories of this course will be fully lectured by the instructor. The R/Linux programming will be taught in a blended format with a combination of students using books and online articles at home with in-class lecture content to expand upon a subset of topics.

Global Objectives

By the end of the course, students will have:

- Developed a skillset of statistics
- Understand the use of basic functions in R.
- Can use R to implement statistical models.
- Developed a skillset of computation and execute them in Linux system.
- Gained a basic understanding of some of the key issues surrounding manipulation of very large files, especially with a high-performance computing (HPC) cluster.

Course Learning Outcomes

By the end of this course, students will have:

- Understood basic mathematical theories underlying frequently used statistical techniques in bioinformatics and data science
 1. Hypothesis test;
 2. Statistical inference (parameter estimation);
 3. Linear models;
 4. Power issue in experimental design.
- Experienced programming and solving problems using fundamental features of R
 1. Basic R programming;
 2. Statistical functions;
 3. Graphics functions.
- Learned Linux commands, including
 1. Basic commands necessary for bioinformatics data manipulation;
 2. Advanced commands such as 'awk' and 'sort'.
- Developed the skills of using high-performance computing (HPC) clusters
 1. Submission and manipulation of HPC jobs;
 2. Considerations of optimally using shared resources of an HPC.

Transferable Skill Development:

Many of the skills and abilities that you are developing in your coursework are transferable to the workforce, graduate and professional studies and other facets of life. Employers seek applicants with transferable skills because they can be an asset in the workplace, regardless of industry or sector. Transferable skills are core skills for your success in building your future career.

The work that you will undertake in this course will help you build the following transferable skills:

- **Verbal Communication:** Learn and share information by presenting, listening, and interacting with others.
- **Creativity and Innovation:** Find different and better ways to do things, being curious, thinking imaginatively.
- **Critical Thinking:** Actively and skillfully conceptualize, apply, analyze, synthesize, and/or evaluate information (data, facts, observable phenomena, and research findings) to make a reasoned judgement or draw a reasonable conclusion.
- **Digital Skills:** Use digital technologies like computers, social media, virtual meeting platforms, and the internet.
- **Information Literacy:** Find, understand, and use information presented through words, symbols, and images
- **Numeracy:** Use mathematical information such as numbers, symbols, words, and graphics to do tasks.
- **Problem solving:** Identify an issue, find and implement a solution, and assess whether the situation has improved.
- **Project Management:** Conceptualize, initiate, plan and execute a plan to achieve a predetermined goal (project) by effectively prioritizing activities and meeting deadlines.
- **Written Communication:** Share ideas and information by using words, images, and symbols.

Learning Resources

Simulated and real world genomic data will be used in the class.

Recommended Textbooks/Readings

There are no required textbooks. Norman Matloff's *The Art of R Programming* is used to be part of the reading materials for R. Casella & Berger's *Statistical Inference* (second edition) contains professional description of many topics covered in the class, although we will learn the materials in a more intuitive and hand-on manner. Several reading materials on statistics, Linux, and HPC will be given in the class.

A Note regarding readings

A list of required readings will be outlined on D2L and links and documents will be made available, where possible. Required readings have been chosen carefully to inform you and enhance the lecture material. **Students are REQUIRED to complete assigned readings BEFORE each lecture.** Instructors will proceed in class on the assumption that students have read completely the assigned readings. Students should be aware that many of the readings they will be assigned may be of an unfamiliar nature and style. Students should allot sufficient time to allow for several reads of the assigned material.

Learning Technology Requirements

Brightspace (by D2L) is located on the University of Calgary server and will be used extensively for communication with students. **It is the student's responsibility to ensure that they receive all posted communications and documents and that they receive emails sent by instructors or fellow students through D2L.** Only your @ucalgary.ca email address may be linked to D2L. Please ensure that you are regularly checking your @ucalgary.ca account.

Evaluation

The University policy on grading and related matters is described in section F of the 2022-2023 Calendar.

In determining the overall grade in the course, the following weights will be used:

- 12% in class participation and performance, including performance on in-class problems
 - This addresses all components of the learning objectives
- 48% after-class assignments
 - This addresses all components of the learning objectives
- 20% Written test
 - This addresses statistical models and its R implementation.
 - Open book and online searching are NOT permitted.
- 20% final project
 - This addresses all components of the learning objectives
 - Open book and online searching are permitted
 - Note that this project component encompasses writing and presentation (see next section)

A student’s final grade for the course is the sum of the separate components. It is not necessary to pass each component separately in order to pass the course.

A Note regarding Writing Assignments:

Writing skills are important to academic study in all disciplines. In keeping with the University of Calgary’s emphasis on the importance of academic writing in student assignments (section E.2 of 2022-23 Calendar), writing is emphasized, and the grading thereof in determining a student’s mark in this course. The Bachelor of Health Sciences values excellence in writing. Competence in writing entails skills in crafting logical, clear, coherent, non-redundant sentences, paragraphs and broader arguments, as well as skills with the mechanics of writing (grammar, spelling, punctuation). Sources used in research papers must be properly documented. The University of Calgary offers instructional services through the Students’ Success Centre’s Writing Support Services (<http://www.ucalgary.ca/writingsupport/>) for students seeking feedback on assignments or seeking to improve their general writing skills. Students are **strongly encouraged** to take advantage of these programs.

Grading Scheme:

Letter Grade	Description	Percentage
A+	Outstanding performance	96-100
A	Excellent performance	90-95.99
A-	Approaching excellent performance	85-89
B+	Exceeding good performance	80-84
B	Good performance	75-79
B-	Approaching good performance	70-74
C+	Exceeding satisfactory performance	65-69
C	Satisfactory performance	60-64
C-	Approaching satisfactory performance	57-59
D+	Marginal pass	54-56
D	Minimal pass	50-53
F	Did not meet course requirements	0-49

Missed Components of Term Work:

Late assignments may be accepted if within 72 hours; however, students will lose 5% per day late past the deadline for all assignments. In this case, assignments will **NOT** be accepted more than 72 hours

after the posted deadline and students failing to submit any assignment within this time frame will receive a mark of zero.

Traffic jams and late or full buses are common events in Calgary and are NOT acceptable reasons for late arrivals to class, meetings and examinations. Please note that while absences are permitted for religious reasons, students are responsible for providing advance notice and adhering to other guidelines on this matter, as outlined in the University Calendar (<https://www.ucalgary.ca/pubs/calendar/current/e-4.html>).

Course Evaluations and Student Feedback

Student feedback will be sought at the end of the course through the Universal Student Rating of Instruction (USRI) and a qualitative student evaluation. Students are welcome to discuss the process and content of the course at any time with the instructor. Students may also address any concerns they may have with Dr. Ebba Kurz, Associate Dean (Undergraduate Health and Science Education) in the Cumming School of Medicine (kurz@ucalgary.ca).

Attendance

Given the large class performance mark, it is essential that students attend each meeting of this class, unless under conditions indicated above.

Conduct During Lectures

The classroom should be respected as a safe place to share ideas without judgement - a community in which we can all learn from one another. Students are expected to frame their comments and questions to lecturers in respectful and appropriate language, always maintaining sensitivity towards the topic. Students, employees, and academic staff are also expected to demonstrate behaviour in class that promotes and maintains a positive and productive learning environment.

As members of the University community, students, employees and academic staff are expected to demonstrate conduct that is consistent with the University of Calgary Calendar, the Code of Conduct and Non-Academic Misconduct policy and procedures, which can be found at <https://www.ucalgary.ca/legal-services/university-policies-procedures>.

Students are expected to take notes during class and should not rely solely on material supplied by the instructors.

Use of Internet and Electronic Communication Devices in Class

The Bachelor of Health Sciences program aims to create a supportive and respectful learning environment for all students. The use of laptop and mobile devices is acceptable when used in a manner appropriate to the course and classroom activities. However, research studies have found that inappropriate/off-topic use of electronic devices in the classroom negatively affects the learning of others during class time.

Students are responsible for being aware of the University's Internet and email use policy, which can be found at <https://www.ucalgary.ca/policies/files/policies/electronic-communications-policy.pdf>.

UNIVERSITY OF CALGARY POLICIES AND SUPPORTS

Copyright

All students are required to reach the University of Calgary policy on Acceptable Use of Material Protected by Copyright (<https://www.ucalgary.ca/policies/files/policies/acceptable-use-of-material-protected-by-copyright-policy.pdf>) and requirements of the Copyright Act (<https://laws-lois.justice.gc.ca/eng/acts/C-42/index.html>) to ensure they are aware of the consequences of unauthorized sharing of course materials (including instructor notes, electronic versions of textbooks, etc.). Students who use material protected by copyright in violation of this policy may be disciplined under the Non-Academic Misconduct Policy <https://www.ucalgary.ca/pubs/calendar/current/k.html>.

Instructor Intellectual Property

Course materials created by instructors (including course outlines, presentations and posted notes, labs, case studies, assignments and exams) remain the intellectual property of the instructor. These materials may **NOT** be reproduced, redistributed or copied without the explicit consent of the professor. **The posting of course materials to third-party websites such as note-sharing sites without permission is prohibited.** Sharing of extracts of these course materials with other students enrolled in the course **at the same time** may be allowed under fair dealing.

Academic Accommodations

It is the student's responsibility to request academic accommodations according to the University policies and procedures listed below. The Student Accommodations policy is available at <https://ucalgary.ca/student-services/access/prospective-students/academic-accommodations>. Students needing an accommodation based on disability or medical concerns should contact Student Accessibility Services (SAS) in accordance with the Procedure for Accommodations for Students with Disabilities (<https://www.ucalgary.ca/policies/files/policies/procedure-for-accommodations-for-students-with-disabilities.pdf>). Students who require an accommodation in relation to their coursework based on a protected ground other than disability should communicate this need in writing to their instructor.

SAS will process the request and issue letters of accommodations to instructors. For additional information on support services and accommodations for students with disabilities, visit www.ucalgary.ca/access/.

Academic Misconduct

The University of Calgary is committed to the highest standards of academic integrity and honesty. The University of Calgary has created rules to govern all its members regarding the creation of knowledge and the demonstration of knowledge having been learned.

Academic Misconduct refers to student behaviour that compromises proper assessment of a student's academic activities and includes (but is not limited to): cheating, fabrication, falsification, plagiarism, unauthorized assistance, failure to comply with an instructor's expectations regarding conduct required of students completing academic assessments in their courses, and failure to comply with exam regulations applied by the Registrar. **It also includes using of third-party websites/services to access past/current course material, essay/assignment writing services, or real-time assistance in completing assessments, seeking answers to assessment questions and similar, whether paid, bartered or unpaid.**

For information of the Student Academic Misconduct Policy and Procedures, please visit; <https://ucalgary.ca/policies/files/policies/student-academic-misconduct-policy.pdf>
<https://ucalgary.ca/policies/files/policies/student-academic-misconduct-procedure.pdf>

Additional information is available on the Academic Integrity website at: <https://ucalgary.ca/student-services/student-success/learning/academic-integrity>.

Recording of Lectures

Audio or video recording of lectures (or similar) is prohibited except where explicit permission has been received from the instructor.

Freedom of Information and Protection of Privacy Act

Student information will be collected in accordance with typical (or usual) classroom practice. Students' assignments will be accessible only by the authorized course faculty. Private information related to the individual student is treated with the utmost regard by the faculty at the University of Calgary

Appeals

If there is a concern with the course, academic matter or a grade, first communicate with the instructor. If these concerns cannot be resolved, students can proceed with an academic appeal, as per Section I of the University Calendar. Students must follow the official reappraisal/appeal process and may contact the Student Ombuds' Office (<http://www.ucalgary.ca/ombuds>) for assistance with this and with any other academic concerns, including academic and non-academic misconduct. Students should be aware that concerns about graded term work may only be initiated **within 10 business days** of first being notified of the grade. <https://www.ucalgary.ca/pubs/calendar/current/i-2.html>

MEDIA RECORDING

Please refer to the following statement on media recording of students: https://elearn.ucalgary.ca/wp-content/uploads/2020/05/Media-Recording-in-Learning-Environments-OSP_FINAL.pdf

Sexual and Gender-Based Violence Policy

The University recognizes that all members of the University Community should be able to learn, work, teach and live in an environment where they are free from harassment, discrimination, and violence. The University of Calgary's sexual violence policy guides us in how we respond to incidents of sexual violence, including supports available to those who have experienced or witnessed sexual violence, or those who are alleged to have committed sexual violence. It provides clear response procedures and timelines, defines complex concepts, and addresses incidents that occur off-campus in certain circumstances. Please see the policy available at <https://www.ucalgary.ca/policies/files/policies/sexual-violence-policy.pdf>

Resources for Support of Student Learning, Success, Safety and Wellness

Student Success Centre <http://www.ucalgary.ca/ssc/>
Student Wellness Centre <http://www.ucalgary.ca/wellnesscentre/>
Student Advocacy and Wellness Hub (CSM students)
 <https://cumming.ucalgary.ca/student-advocacy-wellness-hub/home>
Distress Centre <http://www.distresscentre.com/>
Library Resources <http://library.ucalgary.ca>

Wellness and Mental Health Resources

The University of Calgary recognizes the pivotal role that student mental health plays in physical health, social connectedness and academic success, and aspires to create a caring and supportive campus community where individuals can freely talk about mental health and receive supports when needed. We encourage you to explore the excellent mental health resources available throughout the university

community, such as counselling, self-help resources, peer support or skills-building available through the SU Wellness Centre (<https://www.ucalgary.ca/wellnesscentre/services/mental-health-services>) and the Campus Mental Health Strategy (<http://www.ucalgary.ca/mentalhealth/>).

Student Ombuds' Office

The Student Ombuds' Office supports and provides a safe, neutral space for students. For more information, please visit www.ucalgary.ca/ombuds/ or email ombuds@ucalgary.ca

BHSc Student Faculty Liaison Committee (SFLC)

The BHSc SFLC, with elected representatives from all majors, serves to raise issues of interest to BHSc students to the program administration, including items pertaining to curriculum, scheduling and events. A list of current representatives can be found on the BHSc website.

Student Union (SU) Information

The SU Vice-President Academic can be reached at (403) 220-3911 or suvpaca@ucalgary.ca; the SU representatives for the Cumming School of Medicine can be reached at medrep1@su.ucalgary.ca or medrep2@su.ucalgary.ca.

Student Success Centre

The Student Success Centre provides services and programs to ensure students can make the most of their time at the University of Calgary. Our advisors, learning support staff, and writing support staff assist students in enhancing their skills and achieving their academic goals. They provide tailored learning support and advising programs, as well as one-on-one services, free of charge to all undergraduate and graduate students. For more information visit: <https://www.ucalgary.ca/student-services/student-success>

Emergency Evacuation/Assembly Points

As part of the University of Calgary Emergency Evacuation plan, students, faculty, and staff should locate the closest Assembly Point in case of Fire Alarm. Safety signage is posted throughout the campus showing the locations and the possible route to these locations. All students, faculty, and staff are expected to promptly make their way to the nearest Assembly Point if the Fire Alarm is activated. No one is to return into campus facilities until an all clear is given to the warden in charge of the Assembly Area. For more information, see <https://www.ucalgary.ca/emergencyplan/building-evacuation/assembly-points>

Safewalk

Campus security will escort individuals, day or night, anywhere on campus (including McMahon Stadium, Health Sciences Centre, Student Family Housing, the Alberta Children's Hospital and the University LRT station). Call 403-220-5333 or visit <http://www.ucalgary.ca/security/safewalk>. Use any campus phone, emergency phone or the yellow phone located at most parking lot pay booths. Please ensure your personal safety by taking advantage of this service.

Class Schedule

Date MM/DD/YY	Module / Topics	Instructor/Guest Lecturer	Assignments & Due Dates
09/08/22	1: R: Data structures and basic functions.	Quan Long	
09/15/22	2: R: Operations and data frame	Quan Long	
09/22/22	3: R: Programming	Quan Long	
09/29/22	4: Random variables	Quan Long	
10/06/22	5: Hypothesis test	Quan Long	
10/13/22	6: Statistical inference	Quan Long	
10/20/22	7: Linear models	Quan Long	Final project handout
10/27/22	8: Power estimation	Quan Long	
11/03/22	9: R Statistics & Graphics	Quan Long	
11/17/22	10: Linux: Basic commands & Written exam	Quan Long	
11/24/22	11: Guest lecture: Advanced topics in bioinformatics research	Paul Gordon	
12/01/22	12: Linux: advanced commands and HPC	Quan Long	
			Project due 12/03/2022 by 5 pm