

Syllabus -- Semester Year

The most up-to-date syllabus is located on the guertinlab GitHub page: <https://github.com/guertinlab/meds5420>

Program Information

Class size is limited and priority is given to students supported by the Genomic Medicine T32 and students in either the Systems Biology or Genetics and Developmental Biology area of concentration.

Course and Instructor Information

Course Title: Molecular Genomics Practicum

Course Number: MEDS 5420

Credits: 3

Time: 1:00-3:00 pm, Monday and Wednesday

Room: R1401 Cell and Genome Sciences Building 400 Farmington Ave.

Format: Lecture and immersive in class application. The class will be offered both in-person and remotely at:
redacted

Prerequisites: none

Professor/Instructor/Facilitator: Michael Guertin and Pedro Miura

Pronouns: he/him/his

Email: redacted

Student Hours: Students are encouraged to reach out by email to schedule one-on-one video meetings or in-person traditional office hours.

Course Materials

Required course materials should be obtained before the first day of class.

Students must have a personal computer that is capable of running a Unix or Linux operating system.

Course readings and media are available on GitHub (<https://github.com/guertinlab/meds5420>)

Course Description

Coupling classical molecular biology techniques to high throughput sequencing for nucleic acid detection has revolutionized how scientists study biology. MEDS 5420 will introduce students to the command line and gradually build upon concepts and skills so that students will be capable of building workflows to process and analyze high throughput sequencing data. After basic concepts and command line competency are established, the course will focus on the analysis of ChIP-seq and RNA-seq data. Students will learn to use many common genomics software packages to

perform such tasks as genome alignment, peak calling, motif analysis, and differential expression analysis. If time permits, students will be introduced to the statistical computing language R and perform analyses and visualization using an R interface.

Instructor Thoughts

The study of biology has been completely transformed with the data revolution. Analysis is often rate-limiting scientific progress, not data acquisition. Classically trained molecular biologists, biochemists, and cell biologists are often the best equipped to articulate and ask the most relevant questions regarding a dataset, but they may lack the expertise to carry out the appropriate analyses. The mere thought of analyzing such massive datasets can be intimidating. The main goal of this class is to prepare students so that they are comfortable navigating and analyzing massive datasets and to effectively remove any intimidation barrier.

Course Objectives

By the end of the semester, you should be able to:

1. Comfortably navigate the command line.
2. Use shell scripting to automate parsing, processing, and analysis of genomics data.
3. Design ChIP-seq and RNA-seq analyses, with an emphasis on control experiments/conditions.
4. Align sequencing reads to reference genomes.
5. Retrieve publicly available genomic data sets.
6. Visualize genomics data on a browser.
7. Perform alignment, peak calling, and motif analysis starting of raw ChIP-seq data.
8. Perform alignment, differential expression, and gene set enrichment analysis of raw RNA-seq data.

Course Format

Each class will begin with a short lecture. The instructor will walk through analyses in real-time and the students will spend the majority of class time walking through exercises in parallel with the instructor and learning the skills that are necessary to analyze genomics data.

At the end of each class students may be at different places depending on the speed of their progress during class. Students are expected to finish any of the in-class exercises at home prior to the next class so that everyone will be at the same point at the beginning of the next class. Please come with questions about the previous exercises, but do not expect to use in class time to catch up.

Useful links from the UConn Computational Biology Core

Understanding the UConn Xanadu cluster:

<https://bioinformatics.uconn.edu/resources-and-events/tutorials-2/xanadu/>

Unix basics:

<http://bioinformatics.uconn.edu/unix-basics>

Other CBC tutorials:

<http://bioinformatics.uconn.edu/resources-and-events/tutorials/>

Xanadu server access

Students will have access to a class allocation on the Xanadu server for this course. The instructor will distribute usernames and passwords during the third week of classes. I recommend using this course-associated account even if you have your own already. This will avoid confusion with directory tree structure and problems with access when the server gets busy. **You will need to transfer your data to your own account before the end of the semester.** To request a personal account fill out the form here: <https://bioinformatics.uconn.edu/contact-us/>

PC user resources

Ubuntu (Linux) is available at Microsoft Store, instructions here:

<https://tutorials.ubuntu.com/tutorial/tutorial-ubuntu-on-windows#0>

Shell terminal is also now available for Windows 10: <https://www.laptopmag.com/articles/use-bash-shell-windows-10>

PuTTY, a SSH tool for connecting to server: <https://www.putty.org/>

<https://mediatemple.net/community/products/dv/204404604/using-ssh-in-putty->

WinSCP, a tool for file transfer between server and user's local: <https://winscp.net/eng/download.php>

Course Requirements and Grading

Summary of Course Grading:

| Course Components | Weight |
|--------------------|--------|
| In class exercises | 20% |
| Homework | 30% |
| Midterm project | 25% |
| Final project | 25% |

Note on the use of chatGPT and similar LLMs:

We highly recommend using the course materials and manual pages of the functions to complete the assignments. This is an optional class, so presumably students are taking the class to learn; using chatGPT to perform the tasks is not productive. ChatGPT and LLMs are really only good at performing basic tasks and learning these fundamentals is important for students to develop domain expertise, which chatGPT does not have. So, ironically, one has to learn all the basics that can be outsourced in order to develop a foundation and develop domain expertise, which cannot be outsourced.

In class exercises

Students will be presented with short tasks throughout the lecture and we will go over the exercises as a class. Questions are encouraged and participation is necessary.

Homework

Two to four homework assignments will be assigned throughout the semester. Verbal collaboration, searching code online, and chatGPT are acceptable if students attribute the source.

Midterm

The midterm project is analysis of ChIP-seq data. Each student is provided with ChIP-seq data from an unknown sequence-specific transcription factor. Students will analyze the data and provide their analysis code and conclusions regarding the identity of the factor.

Final

The final project is analysis of RNA-seq data. Each student is provided with RNA-seq data from unknown treatments. Students will analyze the data and provide their analysis code and conclusions regarding the treatment condition.

For additional information on graduate grading policies see here

<https://gradcatalog.uconn.edu/grad-school-info/academic-regulations/#Grades>

Grading Scale for MED 5420:

| Grade | Letter Grade | GPA |
|---------|--------------|-----|
| 180-200 | A | 4.0 |
| 165-179 | A- | 3.7 |
| 130-164 | B+ | 3.3 |
| 120-129 | B | 3.0 |
| 110-119 | B- | 2.7 |
| 105-109 | C+ | 2.3 |
| 100-104 | C | 2.0 |
| 95-99 | C- | 1.7 |
| 92-94 | D+ | 1.3 |
| 90-91 | D | 1.0 |
| 88-89 | D- | 0.7 |
| <88 | F | 0.0 |

Due Dates and Late Policy

All course due dates are identified here. Deadlines are based on Eastern Time unless otherwise specified. *The instructor reserves the right to change dates accordingly as the semester progresses. All changes will be communicated in an appropriate manner. Reasonable accommodations will be made for late homework and projects if the instructor is informed prior to the due date or extenuating circumstances prevent prior notification.*

Feedback and Grades

I will make every effort to provide feedback and grades within a week of the due dates. To keep track of your performance in the course, refer to My Grades in HuskyCT.

Weekly Time Commitment

You should expect to dedicate 10-15 hours a week to this course--this 10-15 hours includes the 4 hours of class time. This expectation is based on the various course activities, assignments, and assessments and the [University of Connecticut's policy regarding credit hours](#). (More information related to hours per week per credit can be accessed at the [Online Student website](#)).

Course Outline

Module 1: Command Line

Module 2: ChIP-seq analysis

Module 3: RNA-seq analysis

Calendar and Class Meeting Schedule Meeting times: MW 1:00-3:00pm

| | |
|------------|--|
| January 21 | Overview of Molecular Genomics and High Throughput Sequencing Technology |
| 26 | Introduction to the Command Line: directories, head, wc, etc. |
| 28 | Introduction to the Command Line: pipes, compression, and grep |

| | |
|------------|--|
| February 2 | Introduction to the Command Line: find, cut, variables, scripting, and permissions (Homework 1 assigned) |
| 4 | Introduction to the Command Line: awk |
| 9 | Introduction to the Command Line: Logical operators, loops, and Xanadu (Miura) |
| 11 | Introduction to the Command Line: Batch scripts, interactive sessions, and software installs |
| 16 | Illumina data format, QC, and preprocessing (HW1 due) |
| 18 | Illumina preprocessing: fastx tools |
| 23 | Aligning Illumina data (Homework 2 assigned) |
| 25 | Transcription Factors and ChIP-seq lecture |
| March 2 | Post-mapping processing with samtools and bedtools; UCSC genome browser |
| 4 | ChIP-seq Analysis: ChIP-seq peak calling and bedtools |
| 9 | ChIP-seq Analysis: Analyze ChIP-seq peaks awk; motif analysis (HW 2 due) (Midterm assigned) |
| 11 | ChIP-seq analysis (pre-recorded; possible review) Guertin is away |
| 23 | Introduction to R and R studio (Miura Starts) |
| 25 | R bioinformatics packages (tentative: HW3 assigned) |
| 30 | RNA-Seq 1 (Intro to RNA-Seq: experimental design, library prep) (Midterm due-Guertin) |
| April 1 | RNA-Seq 2 (alignment, pseudo-alignment methods) (tentative: HW4 assigned) |
| 6 | RNA-Seq 3 (track visualization- IGV/UCSC) |
| 8 | Differential Gene Expression, GO Enrichment Analysis, R Markdown |
| 13 | Alternative Splicing Analysis (tentative: Final Assigned) |
| 15 | Long Read RNA-Seq analysis (Alternative Polyadenylation analysis) |
| 20 | Single Cell RNA-Seq 1 (Introduction of single cell concepts and methodology) |
| 22 | Single Cell RNA-Seq 2 (Analysis– Seurat) |
| 27 | Other Resources and tools (Conda, NextFlow), Review |
| 29 | TBD |

Resources for Students Experiencing Distress

The University of Connecticut is committed to supporting students in their mental health, their psychological and social well-being, and their connection to their academic experience and overall wellness. The university believes that academic, personal, and professional development can flourish only when each member of our community is assured equitable access to mental health services. The university aims to make access to mental health attainable while fostering a community reflecting equity and diversity and understands that good mental health may lead to personal and professional

growth, greater self-awareness, increased social engagement, enhanced academic success, and campus and community involvement. Students experiencing personal difficulties during the course may utilize the following services. Please familiarize yourself with these offices and the services they provide.

Student Behavioral Health Services:

<https://health.uconn.edu/student-affairs/health-and-wellness/student-behavioral-health-service/>

Office of Institutional Equity: <http://equity.uconn.edu/>

UConn Ombuds Office: <http://ombuds.uconn.edu/>

Accommodations for Illness or Extended Absences

Please stay home if you are feeling ill and please go home if you are in class and start to feel ill. If illness prevents you from attending class, it is your responsibility to notify me as soon as possible. You do not need to disclose the nature of your illness, however, you will need to work with me to determine how you will complete coursework during your absence.

COVID-19 Specific Information: People with COVID-19 have had a wide range of symptoms reported – ranging from mild symptoms to severe illness. These symptoms may appear 2-14 days after exposure to the virus and can include:

- Fever,
- Cough,
- Shortness of breath or difficulty breathing
- Chills
- Repeated shaking with chills
- Muscle pain
- Headache
- Sore throat
- New loss of taste or smell

If you have symptoms or have had contact with someone who tests positive for COVID-19, please contact **the UConn Health Call Center at 860-679-3199**. They will determine whether you need to be tested and/or quarantine.

Classroom/Virtual Classroom Guidelines

Recording Lectures

Classes for this semester's course will be conducted in person and over Zoom. As the host of any virtual classes, I will record these sessions using Zoom's recording feature. The recording feature for others in attendance will be disabled so that no one else will be able to record a session. In order to protect student privacy and intellectual property rights, students are prohibited from recording any session, or any portion of a session, by other means. At my discretion and in accordance with University policies and guidelines, I may share one or more of the recorded sessions with the class to provide students with an additional opportunity to review course content. The sharing of any recorded content without my written permission is prohibited. If you would like to ensure your likeness is not captured during an online class, please turn your camera off. For recordings conducted in person, please alert me to any concerns so that I may take steps to help ensure you are not recorded.

Please remember that the unauthorized recording or sharing of course content may be considered a violation of the law, University policy, and/or The Student Code.

The web-based video delivery of each class in this course is for sole use of the students enrolled in this course. Any other use of these class videos or any pictures or derivatives of the class videos without the written consent of the course's professor is prohibited.

The videos created by students as part of this course are for sole use of the students enrolled in this course. Any other use of these videos or any pictures or derivatives of the videos without the written consent of the video creator is prohibited.

Student Responsibilities and Resources

As a member of the University of Connecticut student community, you are held to certain standards and academic policies. In addition, there are numerous resources available to help you succeed in your academic work. Review these important [standards, policies and resources](#), which include:

- The Student Code
 - Academic Integrity
 - Resources on Avoiding Cheating and Plagiarism
- Copyrighted Materials
- Credit Hours and Workload
- Netiquette and Communication
- Adding or Dropping a Course
- Academic Calendar
- Policy Against Discrimination, Harassment and Inappropriate Romantic Relationships
- Sexual Assault Reporting Policy

Students with Disabilities

The University of Connecticut is committed to protecting the rights of individuals with disabilities and assuring that the learning environment is accessible. Students who require accommodations should contact the Center for Students with Disabilities, Wilbur Cross Building Room 204, (860) 486-2020 or <http://csd.uconn.edu/>.

Blackboard measures and evaluates accessibility using two sets of standards: the WCAG 2.0 standards issued by the World Wide Web Consortium (W3C) and Section 508 of the Rehabilitation Act issued in the United States federal government." (Retrieved March 24, 2013 from [Blackboard's website](#))

Software/Technical Requirements (with Accessibility and Privacy Information)

In addition to the computing environment outlined above, the software/technical requirements for this course include:

- Equipment Recommendations (<https://remotework.uconn.edu/equipment-recommendations/>)
- HuskyCT/Blackboard ([HuskyCT/ Blackboard Accessibility Statement](#), [HuskyCT/ Blackboard Privacy Policy](#))
- [Adobe Acrobat Reader](#) ([Adobe Reader Accessibility Statement](#), [Adobe Reader Privacy Policy](#))
- Google Apps ([Google Apps Accessibility](#), [Google for Education Privacy Policy](#))
- Dedicated access to high-speed internet with a minimum speed of 1.5 Mbps (4 Mbps or higher is recommended).
- WebCam

Privacy Statement: For information on managing your privacy at the University of Connecticut, visit the [University's Privacy page](#). NOTE: This course has NOT been designed for use with mobile devices.

Help

Contact AITS at UConn Health <https://health.uconn.edu/aits/technical-support/> for technical and academic assistance.

This course uses the learning management platform, [HuskyCT](#). If you have difficulty accessing HuskyCT, you have access to the in person/live person support options available during regular business hours through the [Help Center](#). You also have [24x7 Course Support](#) including access to live chat, phone, and support documents.

Evaluation of Course Experience

Students will be given an opportunity to provide feedback on their course experience and instruction through a survey that will be available at the end of the course. Please complete the course survey. Courses are modified according to student feedback to provide a positive and effective learning experience.