STAT588/BIOL588

Genomic Data Science

2023 Fall

MW 2:20-3:35p

Instructor Name: Yen-Yi Ho
E-mail: hoyen@stat.sc.edu
Classroom: LeConte 107
Office: LeConte 216A
Office Hours: MW 10-11AM or by appointment (virtual office hour is also available)
Teaching Assistant: Cenxiao Gao (CENXIAO@email.sc.edu)
TA Office Hour: TBA
Course website: https://people.stat.sc.edu/hoyen/STAT588/Stat588.html

Course Description

This course focuses on quantitative knowledge for interdisciplinary applications in genetics as well as “hands-on” experience in analyzing genetic data. Topics include in gene association mapping, analyses of gene expression data generated from high throughput technologies, false discovery rate, gene set enrichment analysis and machine learning algorithms. Advanced topics such as analyses of single-cell, multi-omics, ChIP-seq data will be included if time permits. In this course, students will have programming exercises in using analysis tools to conduct genome-wide analysis, annotation and interpretation of genetic data. The lab sections of this course will focus on using R/Bioconductor software packages as tools for analyses.

Prerequisites

Students will be expected to have taken introductory Statistics course (STAT205 or STAT 201 or STAT515 or equivalent with grade C or better).

Learning Outcomes

After successful completion of this course, you will be able to:

1. demonstrate advanced knowledge of statistical tools applied to human genetics using open source software, including R and Bioconductor
2. demonstrate skills for genomic visualization, reproducible analysis, data architecture, and exploration of large-scale genomic data
3. apply knowledge, analytical skills and interpret findings for a typical genome-scale assays.
4. analyze appropriate literature to develop a literature review based on a selected topic.

Course Materials

Required

1. Gondo (2015): Primer to Analysis of Genomic Data Using R. Springer. ISBN 978-3-319-14475-7
2. Thomas Girke's R & Bioconductor manuals (http://manuals.bioinformatics.ucr.edu/home/R_BioCondManual)
3. Wim P. Krijnen. Applied Statistics for Bioinformatics using R. (https://cran.r-project.org/doc/contrib/Krijnen-IntroBioInfStatistics.pdf)
4. The free "try R" class from Code School is also a good place to start: http://tryr.codeschool.com/
5. Advanced R by Hedley Wickham (http://adv-r.had.co.nz/Data-structures.html)

Recommended alternative sources:

5. Hahne, Huber, Gentleman, and Falcon (2008): Bioconductor Case Studies
6. R tutorials at UCLA: http://www.ats.ucla.edu/stat/r/
7. Introduction to R: http://cran.r-project.org/doc/manuals/R-intro.pdf

All course materials comply with copyright/fair use policies.

Technology Requirements

The lecture presentations, links to articles, assignments, quizzes, and rubrics are located on the Blackboard site for the course. To participate in learning activities and complete assignments, you will need:

- This course involves hand-on data analysis during the class section. It is encouraged that students bring a laptop during class sections.
- Access to a working computer that has a current operating system with updates installed;
- RStudio program and report writing using R Markdown installed; and
- Reliable Internet access and a USC email account;
- A current Internet browser that is compatible with Blackboard (Google Chrome is the recommended browser for Blackboard);
- Reliable data storage for your work, such as a USB drive or Office365 OneDrive cloud storage.

If your computer does not have RStudio program. It is free and can be downloaded from https://www.rstudio.com/products/rstudio/download/
If you have further questions or need help with the software, please contact the Service Desk (https://www.sc.edu/about/offices_and_divisions/university_technology_services/support/servicedesk.php).

Minimal Technical Skills Needed

Minimal technical skills are needed in this course. All work in this course must be completed and submitted using compile pdf file from R Markdown. Therefore, you must have consistent and reliable access to a computer and the Internet. The minimal technical skills you have include the ability to:

- Organize and save electronic files;
- Use USC email and attached files;
- Check email daily;
- Download and upload documents;
- Locate information with a browser; and
- Use Blackboard.

Technical Support

If you have problems with your computer, technology, IT-related questions, support, including Blackboard, please contact the Division of Information Technology (DoIT) Service Desk at (803) 777-1800 or submit an online request through the Self-Service Portal (https://scprod.service-now.com/sp) or visit the Carolina Tech Zone (https://www.sc.edu/about/offices_and_divisions/university_technology_services/support/ctz.php)). The Service Desk is open Monday – Friday from 8:00 AM – 6:00 PM (Eastern Time Zone). If you are located in the Columbia, SC area, the Thomas Cooper Library at USC has computers for you to use in case you encounter computer issues/problems. If you are not located in the Columbia, SC area, most regional campuses and public libraries have computers for public use.

Course Assignments and Grading

General Assignment Information

Weekly assignments are due on the day indicated on the course website.
A final project are due on the day indicated on the course website.

Evaluation and Grading Scale

Grading Rubric	
Homework Assignments	70%
Final Project	30%
Total	100%

The grades will be based on performance on homework assignments (70%), a final course project (30%). Final Project is due on Friday December 8, 2023 at 5PM (EST).

Grading Scale

89.5% - 100% = A
84.5% - 89.4% = B+
79.5% - 84.4% = B
74.5% - 79.4% = C+
69.5% - 74.4% = C
64.5% - 69.4% = D+
59.5% - 64.4% = D
0% - 59.4% = F

Course Policies and Procedures

- Please Email your Homework and R code to Cenxiao Gao (CENXIAO@email.sc.edu).
- I encourage you to work together in computing and discussing the problems. However, each student is expected to independently write up the submitted assignment using her or his own computing and giving explanations in her or his own words.
- **Late homework will lose 10% of total points per day**, unless arrangements have been made with the instructor for an extension. Homework will not be accepted after the time at which graded homework are returned.
- Incomplete Contracts: A grade of incomplete "I" shall be assigned at the discretion of the instructor when, due to extraordinary circumstances (e.g., documented illness or hospitalization, death in family, etc.), the student was prevented from completing the work of the course on time. The assignment of an "I" requires that a contract be initiated and completed by the student before the last official day of class, and signed by both the student and instructor. If an incomplete is deemed appropriate by the instructor, the student in consultation with the instructor, will specify the time and manner in which the student will complete course requirements. Extension for completion of the work will not exceed one year (or earlier if designated by the student's college).

Attendance Policy

When you miss class, you miss important information. If you are absent, you are responsible for learning material covered in class. If you are absent when an assignment is due, you must have submitted the assignment prior to the due date to receive credit. If you miss more than 20% of the classes unexcused, your grade will be dropped one letter grade.

Academic Integrity

You are expected to practice the highest possible standards of academic integrity. Any deviation from this expectation will result in a minimum academic penalty of your failing the assignment, and will result in additional disciplinary measures. This includes improper citation of sources, using another student's work, and any other form of academic misrepresentation. The first tenet of the Carolinian Creed is, "I will practice personal and academic integrity."

Below are some websites for you to visit to learn more about University policies:

Carolinian Creed (http://www.sa.sc.edu/creed)
Academic Responsibility (http://www.sc.edu/policies/staf625.pdf)
Office of Student Conduct and Academic Integrity (https://www.sa.sc.edu/academicintegrity/)
Network Guidelines for Responsible Computing (http://www.sc.edu/about/offices_and_divisions/university_technology_services/policies_procedures/networkguideline.php)

Plagiarism

Using the words or ideas of another as if they were one's own is a serious form of academic dishonesty. If another person's complete sentence, syntax, key words, or the specific or unique ideas and information are used, one must give that person credit through proper citation.

Class Conduct

Professionalism will be expected at all times. Because the university classroom is a place designed for the free exchange of ideas, we must show respect for one another in all circumstances. We will show respect for one another by exhibiting patience and courtesy in our exchanges. Appropriate language and restraint from verbal attacks upon those whose perspectives differ from your own is a minimum requirement. Courtesy and kindness is the norm for those who participate in my class.

Instructional Methods

The course will be taught using multiple instructional methods. These methods will include lecture, computer programming exercise sections.

Diversity and Inclusion

The university is committed to a campus environment that is inclusive, safe, and respectful for all persons, and one that fully embraces the Carolinian Creed. To that end, all course activities will be conducted in an atmosphere of friendly participation and interaction among colleagues, recognizing and appreciating the unique experiences, background, and point of view each student brings. You are expected at all times to apply the highest academic standards to this course and to treat others with dignity and respect.

Expectations of the Instructor

I am expected to facilitate learning, answer questions appropriately, be fair and objective in grading, provide timely and useful feedback on assignments, maintain adequate office hours, and treat you as I would like to be treated.

Course Communication

I will be communicating with you regarding grades and assignments. If you need to get in touch with me, the best method is via email.
If you are having trouble with this course or its material, you should contact me via email to set up a appointment to discuss the issues.
Emails will be sent your email address. In addition, announcements may be posted on the course website. If you primarily use another email account, you should make sure that the Blackboard account is linked to that address. It is your responsibility to ensure that your email account works properly in order to receive mail.

Please be sure that the email you check regularly is set in Blackboard:

- Click on the My USC tab along the top of the page in Blackboard
- In the Tools module, click on "Personal Information"
- Click on "Edit Personal Information"
- Scroll down to the listing for Email
- In the box will be listed what Blackboard has as your email address. If you wish to change it, delete the email address in the box and type in the email address you want to use.
- Click on the Submit button at the top or bottom of the page.

Copyright/Fair Use Statement

I will cite and/or reference any materials that I use in this course that I do not create. You, as students, are expected to not distribute any of these materials, resources, homework assignments, and final project report.

Disability Services

Student Disability Resource Center (http://www.sa.sc.edu/sds/): The Student Disability Resource Center (SDRC) empowers students to manage challenges and limitations imposed by disabilities.

Students with disabilities are encouraged to contact me to discuss the logistics of any accommodations needed to fulfill course requirements (within the first week of the semester). In order to receive reasonable accommodations from me, you must be registered with the Student Disability Resource Center (1523 Greene Street, LeConte Room 112A, Columbia, SC 29208, 803-777-6142). Any student with a documented disability should contact the SDRC to make arrangements for appropriate accommodations.

Student Success Center

In partnership with USC faculty, the Student Success Center (SSC) offers a number of programs to assist you in better understanding your course material and to aid you on your path to success. SSC programs are facilitated by professional staff, graduate students, and trained undergraduate peer leaders who have previously excelled in their courses. Resources available to you in this course may include:

- Peer Tutoring: You can make a one-on-one appointment with a Peer Tutor (www.sc.edu/success). Drop-in Tutoring and Online Tutoring may also be available for this course. Visit their website for a full schedule of times, locations, and courses.
- Supplemental Instruction (SI): SI Leaders are assigned to specific sections of courses and hold three weekly study sessions. Sessions focus on the most difficult content being covered in class. The SI Session schedule is posted through the SSC website each week and will also be communicated in class by the SI Leader.
- Peer Writing: Improve your college-level writing skills by bringing writing assignments from any of your classes to a Peer Writing Tutor. Similar to Tutoring, you can visit the website to make an appointment, and to view the full schedule of available drop-in hours and locations.
- Success Consultations: In Success Consultations, SSC staff assist you in developing study skills, setting goals, and connecting to a variety of campus resources. Throughout the semester, I may communicate with the SSC via Success Connect, an online referral system, regarding your progress in the course. If contacted by the

SSC, please schedule a Success Consultation. Success Connect referrals are not punitive and any information shared by me is confidential and subject to FERPA regulations.

SSC services are offered to all USC undergraduates at no additional cost. You are invited to call the Student Success Hotline at (803) 777-1000, visit the SSC website (www.sc.edu/success), or stop by the SSC in the Thomas Cooper Library on the Mezzanine Level to check schedules and make appointments.

Writing Center

Writing Center (http://artsandsciences.sc.edu/write/university-writing-center)

This course has many of writing assignments. The University Writing Center is an important resource you should use! It's open to help any USC student needing assistance with a writing project at any stage of development. The main Writing Center is in Byrnes 703.

Library Resources

Library Resources (http://library.sc.edu)

The university library has great resources for finding out how to cite materials in your projects. Remember that if you use anything that is not your own writing or media (quotes from books, articles, interviews, websites, movies – everything) you must cite the source in MLA format.

Blackboard and Technology

Blackboard and Technology (http://www.sc.edu/about/offices_and_divisions/university_technology_services/)

As a student in this course, you have access to support from the Division of Information Technology (DoIT) for Blackboard and computer issues. The service desk can be reached at 803-777-1800.

Counseling Services (https://sc.edu/about/offices_and_divisions/student_health_services/medical-services/counseling-and-psychiatry/index.php): The University offers counseling and crisis services as well as outreach services, self-help, and frquently asked questions.

Course Schedule

Approximate course outline: (Lecture notes will be updated often). Please check the course website regularly for homework/due dates, reading assignments, and announcements.

Week	Topics
Week 1-3	Introduction to Biology and R
Week 4	Review Statistics I/Lab (Data Import/Export, Simple data Manipulation, missing values, writing functions).
Week 5	Review Statistics II/Lab (Statistical Test, SNP data association data).
Week 6	Simple Marker Association Test /Lab (Simple Genomic Data analysis using R)
Week 7	Genome-Wide Association Study
Week 8	Population structure, supervised learning with high dimensional data and categorical predictors
Week 9	Introduction to Bioconductor
Week 10	Introduction to Genomic Technologies
Week 11	Management of Genomic-Scale Data with Bioconductor
Week 12	Genomic Annoation with Bioconductor
Week 13	Inference for Genomics with Bioconductor
Week 14	Case Studies in Functional Genomics: RNA-seq, Single-Cell Data Analysis
Week 15	Machine Learning to Gene Expression Data /Working with high performance computing clusters