Bioinformatic Analyses of Genome-Scale Data
BCHM 612
Syllabus
Summer 2019

Instructor: Dr. Pete E. Pascuzzi
Office: WALC 3053A
Phone: 765-494-3620
email: ppascuzz@purdue.edu
Hours: Wednesday 10:00 - 11:00, after class or by appointment, also via Piazza

Teaching Assistant: Jiaxin Long
Office: WSLR B017
Phone: 765-496-6204
email: long274@purdue.edu
Hours: Friday 1:00 - 2:00, after class or by appointment, also via Piazza

COURSE DESCRIPTION

This course provides a hands-on experience for life science researchers in the bioinformatic analysis of genome-scale data. The various disciplines in the life sciences are generating a wealth of experimental and annotation data. Today's graduate students need experience with modern tools that can help them to access, explore, analyze, interpret and manage the data that they generate in the lab.

Students will use the R programming language and packages from Bioconductor, the R bioinformatics project, as their principal tools for this course. Students will develop workflows in R that bridge established algorithms for bioinformatics such as limma, edgeR or DESeq2, incorporating methods to import, QC, transform and visualize genome-scale datasets derived from next generation sequencing experiments. A critical aspect of bioinformatics that is often inadequate is workflow documentation. This course will use Rmarkdown to integrate computer code, data and results to manage complex bioinformatics projects.

The class has lecture, lab and distance components. Lectures will focus on the theoretical and biological aspects of bioinformatics analysis using recent examples from the literature. In lab, students will work on programming exercises or projects using published datasets. Advanced students will also have the opportunity to work with their own data. Distance instruction will include R tutorials and videos that students can work through at their own pace (subject to completion deadlines). Particular emphasis will be placed on the theoretical and practical limitations of next generation sequencing data.

No prior computer programming experience is required, but it is assumed that students have a firm grasp of the fundamental principles of molecular biology and how they relate to complex processes such as gene expression and chromatin structure.

LEARNING OUTCOMES

- Students will write R scripts that utilize Bioconductor packages for bioinformatic analyses.
- Students will access genome-scale data sets from public repositories and import this data into R for further analysis.

- Students will visualize genome-scale data sets for both quality control and presentation purposes.
- Students will implement strategies to deal with genome-scale datasets including parallel computing.
- Students will be able to critically evaluate the bioinformatic methods and data from publications.
- Students will implement "literate programming" with Rmarkdown to document and share their bioinformatics projects.

TEXTBOOK

No required textbook. However, there are many eBooks on R programming available through the library. Here are a few that are recommended:

R for Data Science, Garret Groolemund and Hadley Wickham
R in a Nutshell, Joseph Adler
Introductory Statistics with R, Peter Daalgard
Bioinformatics with R Cookbook, Paurush Praveen Sinha

LECTURE TIME AND PLACE

Session	Day	Time	Location
Lecture	TR	8:40 - 9:30	NLSN 1225
Lab	TR	9:50 - 11:40	NLSN1225
Distance	110 minutes per week		

ASSESSMENT

This course is offered for a letter grade. Grades will be determined through successful completion of exercises, projects, a data management notebook, and a final. Final grades will be based on the following:

Assignment	Grading
Eight lab exercises	25%
Bioinformatic project	25%
Bioinformatic Notebook	25%
Final exam	25%
Class participation and attendance	0, 2% or 4% bonus points

Exercises and projects will be graded on a scale 0 - 10 with point values clearly indicated. Every effort will be made to give partial credit. Lab exercises and programming projects are not graded on style or efficiency, only on successful completion of the assigned task(s). If your code is well-organized and easy to read, with clear documentation, it is much easier to give you partial credit.

The Bioinformatics Notebook is an open-ended assignment that enables students to synthesize concepts from lecture and lab. The Notebook will be written in Rmarkdown allowing the students to include text narrative, links to web resources, and executeable computer code. The Notebook should include sections ranging from file transfer with Globus, SMB and SCP to quality control metrics for next generation sequencing with clear documentation of how these topics apply to the student's research. The intent is that students will leave the class with a tangible product that will help them with their research.

The final is a traditional exam comprised of short answer questions as well as code chunks that must be completed, edited or corrected.

Perfect attendance is required for 4% bonus points, but one absence is allowed for 2% bonus points. Pre-arranged, excused absences may be allowed at the discretion of the instructor. In either case, active participation in class discussions is expected.

GRADING SCALE

Score	Grade	Score	Grade
98.00 and above	A+	78.00 to 79.99	C+
92.00 to 97.99	A	72.00 to 77.99	C
90.00 to 91.99	A-	70.00 to 71.99	C-
88.00 to 89.99	B+	68.00 to 69.99	D+
82.00 to 87.99	B	62.00 to 67.99	D
80.00 to 81.99	B-	60.00 to 61.99	D-
		59.99 and Below	F

EXTRA CREDIT

Class attendance is the only way to achieve some extra credit. No exceptions will be made.

COURSE MANAGEMENT SYSTEM

We will not use Blackboard for this course.

PIAZZA

There is a Piazza site for this course. Many course materials and resources will be distributed via this site. There is also a Q&A forum where you can post questions. You can post anonymously to you fellow students, but not to the instructors. It is critical that we know who you are so that we can help you effectively.

ACADEMIC MISCONDUCT

Academic misconduct of any kind will not be tolerated in any course offered by the Department of Biochemistry. For specifics, please refer to Purdue's Regulations Governing Student Conduct

Any incidence of academic misconduct will be reported to the Office of the Dean of Students. Academic misconduct may result in disciplinary sanctions including expulsion, suspension, probated suspension, disciplinary probation, and/or educational sanctions. In addition, such misconduct will result in punitive grading such as:

- receiving a lower or failing grade on the assignment, or
- assessing a lower or failing grade for the course

Punitive grading decisions will be made after consultation with the Office of the Dean of Students. Please note reported incidences of academic misconduct go on record for reference by other instructors. Further, a record of academic misconduct is likely to influence how current/future situations are handled.

To provide you with an unambiguous definition of academic misconduct, the following text has been excerpted from "Academic Integrity: A Guide for Students", written by Stephen Akers, Ph.D., Executive Associate Dean of Students (1995, Revised 1999, 2003), and published by the Office of the Dean of Students in cooperation with Purdue Student Government, Schleman Hall of Student Services, Room 207, 475 Stadium Mall Drive West Lafayette, IN 47907-2050.

Purdue prohibits "dishonesty in connection with any University activity. Cheating, plagiarism, or knowingly furnishing false information to the University are examples of dishonesty." [Part 5, Section III-B-2-a, Student Regulations] Furthermore, the University Senate has stipulated that "the commitment of acts of cheating, lying, and deceit in any of their diverse forms (such as the use of substitutes for taking examinations, the use of illegal cribs, plagiarism, and copying during examinations) is dishonest and must not be tolerated. Moreover, knowingly to aid and abet, directly or indirectly, other parties in committing dishonest acts is in itself dishonest." [University Senate Document 72-18, December 15, 1972]

More specifically, the following are a few examples of academic dishonesty which have been discovered at Purdue University.

- substituting on an exam for another student
- substituting in a course for another student
- paying someone else to write a paper and submitting it as one's own work
- giving or receiving answers by use of signals during an exam
- copying with or without the other person's knowledge during an exam
- doing class assignments for someone else
- plagiarizing published material, class assignments, or lab reports
- turning in a paper that has been purchased from a commercial research firm or obtained from the internet
- padding items of a bibliography
- obtaining an unauthorized copy of a test in advance of its scheduled administration
- using unauthorized notes during an exam
- collaborating with other students on assignments when it is not allowed
- obtaining a test from the exam site, completing and submitting it later
- altering answers on a scored test and submitting it for a regrade
- accessing and altering grade records
- stealing class assignments from other students and submitting them as one's own
- fabricating data
- destroying or stealing the work of other students

Plagiarism is a special kind of academic dishonesty in which one person steals another person's ideas or words and falsely presents them as the plagiarist's own product. This is most likely to occur in the following ways:

- using the exact language of someone else without the use of quotation marks and without giving proper credit to the author
- presenting the sequence of ideas or arranging the material of someone else even though such is expressed in one's own words, without giving appropriate acknowledgment
- submitting a document written by someone else but representing it as one's own

CLASS ATTENDANCE

In accordance with University policy, you are expected to attend every scheduled class. If you have a valid reason for missing class such as a University-sponsored activity, religious observances, illness, or family emergency, the instructor or TA will assist you in obtaining information and materials you may have missed. Students who skip class without a valid excuse should not expect the instructor or TA to supply class notes or provide special help. For more information see the Purdue Regulations Governing Classes and the Class Absence page from the Office of the Dean of Students.

EMERGENCY PREPAREDNESS

In the event of a major campus emergency, course requirements, deadlines and grading percentages are subject to changes that may be necessitated by a revised semester calendar or other circumstances. To get information about changes in this course consult the class PURR site or e-mail or phone the instructor.

ON-LINE COURSE EVALUATIONS

During the last two weeks of the semester, you will be provided an opportunity to evaluate this course and your instructor(s). To this end, Purdue has transitioned to online course evaluations. On Monday of the fifteenth week of classes, you will receive an official email from evaluation administrators with a link to the online evaluation site. You will have two weeks to complete this evaluation. Your participation in this evaluation is an integral part of this course. Your feedback is vital to improving education at Purdue University. I strongly urge you to participate in the evaluation system.

NON-DISCRIMINATION POLICY STATEMENT

Purdue University's non-discrimination policy will be upheld in this classroom. Purdue University is committed to maintaining a community which recognizes and values the inherent worth and dignity of every person; fosters tolerance, sensitivity, understanding, and mutual respect among its members; and encourages each individual to strive to reach his or her own potential. In pursuit of its goal of academic excellence, the University seeks to develop and nurture diversity. The University believes that diversity among its many members strengthens the institution, stimulates creativity, promotes the exchange of ideas, and enriches campus life.

Purdue University views, evaluates, and treats all persons in any University related activity or circumstance in which they may be involved, solely as individuals on the basis of their own personal abilities, qualifications, and other relevant characteristics.

For more information, refer to the Purdue Nondiscrimination Policy Statement.

CLASS SCHEDULE (subject to change)

Week	Topic
May 14 & 16	Overview of bioinformatics, logistics and introduction to R
May 21 & 23	Biological data types and manipulating small data with R
May 28 & 30	Genome assemblies and annotation
June 4 & 6	Manipulating sequences and genomic intervals with Bioconductor
June 11 & 13	Gene Regulation and RNA-seq
June 18 & 20	Functional annotation and pathway analysis
June 25 & 27	Bioinformatics projects
July 2	Final exam