BIOC 4723: Introduction to Bioinformatics - exploring information
Fall 2018

Instructor: Charles Chen
Office: 348G Nobel Research Center
Office Tel: 405-744-4025
Email: charles.chen@okstate.edu

Lecture: Monday, Wednesday 1:00 - 2:15 PM
Prerequisites: BIOL1114 and MATH1513
Text: All course material will be available through course web page
No textbook is required.

Course Description
Biology has now stepped into a 'data-rich" era. To comprehend Big Data in Biology, basic computer programming is required for everything from accessing and managing data, to statistical analysis, simulations and numerical modeling. Providing an introduction to programming for biologists, this course is aimed to bring real world research cases to the classroom, discuss biological information presented in research and programming techniques required to solve the problems. Also, all lectures will be followed by computer labs to deliver hands-on computing exercises for students. The course will touch the basics of Shell programming, scripting languages like R and examples of using packages such as Bioconductor. Finally, in order to prepare students for large-volume information, this course will provide an introductory supercomputing experience using OSU's high performance computing clusters. This course is designed to be a stepping-stone to higher-level bioinformatics. No programming background is required.

Course Objectives
In this course you will understand the fundamental aspects of different biological 'data' and basic programming skills that are necessary for conducting biological research. By the end of the course you will be able to use these tools to import data into a high performance computing (HPC) environment, perform analyses, export the results and make them interpretive as graphs, tables and text files, where you can present the biological relevance.

1. Learn to work on the COWBOY, OSU's super computer
2. Write simple data analysis using Unix Shell scripts
3. Learn R and Bioconductor
4. Automate data analyses
5. Apply these tools to address biological questions
This course is not aimed to train professional programmers. However, it is meant to familiarize students with modern day biological information and computational toolsets. Also, learn and understand programming concepts that will help with using other languages.

Tentative Schedule

**W1** Aug 18: **Course Topic:** Course Concept, computer, data and what is informatics
  Aug 20: **Computer Lab:** Visit Cowboy, OSU’s super computer
  Assignment 01 - basics on high performing computing

**W2** Aug 25: **Course Topic:** Understanding Biological Data
  Aug 27: **Computer Lab:** Data resources
  Assignment 02 - basic data query on linux (10%)

**W3** Sep 01: **Course Topic:** What is Unix
  Sep 03: **Computer Lab:** Cowboy new user training
  Assignment 02 – Cowboy account (5%)

**W4** Sep 08: **Course Topic:** from Human Genome Project to 1000 Genomes
  Sep 10: **Computer Lab:** Shell scripting-I
  Assignment 03 - basic Shell on key word search

**W5** Sep 15: **Course Topic:** the chase of BRAC1
  Sep 17: **Computer Lab:** Shell scripting-2
  Assignment 04 - pattern identification, start and end codons

**W6** Sep 22: **Course Topic:** Exploring Patient Information, control-case studies
  Sep 24: **Computer Lab:** Introduction into R, data I/O and basic syntax

**W7** Sep 29: **Course Topic:** applied statistics and graphic functions in R
  Oct 01: **Computer Lab:** Scripting with R, string matching and parsing
  Assignment 05 - sequence comparison

**W8** Oct 06: **Course Topic:** R scripting: variable, expression, condition, loop and function
  Oct 08: **Computer Lab:** scripting in R

**W9** Oct 13: **Course Topic:** Next-generation technologies
  Oct 15: **Computer Lab:** Introduction to Bioconductor
  Assignment 06 - Differential Gene Expression

**W10** Oct 20: **Course Topic:** Bioconductor- Biostrings
  Oct 22: **Computer Lab:** FASTQ file handling

**W11** Oct 27: **Course Topic:** Bioconductor- IRRange and GenomicRange
  Oct 29: **Computer Lab:** fasta, read short sequence analysis
  Assignment 07 – missing data imputation

**W12** Nov 03: **Course Topic:** Associating variants, DNA -> Phenotypes
  Nov 05: **Computer Lab:** Regression, association statistics

**W13** Nov 10: **Course Topic:** Pattern Discovery, biological networks
Nov 12: **Computer Lab:** clustering, heatmap Venn diagram
Assignment 08 – Transcriptome data analysis. Bioconductor

**W14** Nov 17: **Course Topic:** Biological relational database

Nov 19: **Computer Lab:** Gene Ontology Enrichment analysis, network visualization

**W15** Nov 24: **Course Topic:** Artificial Neuron Networks and Machine Learning

Happy Thanksgiving!!!

**W16** Dec 01: Summary lecture
Dec 03: Final assignment and term paper for graduate students.

**Grading**
Grading for this course will be based on student's assignments. There will be 7 equally weighted programming assignments (Assignment 01- Assignment 06, 10% each) and the Assignment 07 and assignment 08 are 20% each. Assignments are due on next Thursday night by 11:59 pm Central Time. Assignments should be submitted via D2L with the file name: Programming Assignment X_NameOfStudent, where X is the number of the assignment. Undergraduate students will be graded based on Assignment 01- 08; a final assignment, a research project and a term paper, is required for graduate students.

Student assignment will receive a thorough code review and a detailed grade as following:

- Produces the correct answer using the requested approach: **100%**
- Generally uses the right approach, but a minor mistake results in an incorrect answer: **90%**
- Attempts to solve the problem and makes some progress using the core concept: **50%**
- Answer demonstrates a lack of understanding of the core concept: **0%**

Late assignments will be docked 25% and will not be accepted after Saturday night at 11:59 pm Central Time except in cases of genuine emergencies that can be documented by the student or in cases where this has been discussed and approved in advance. This policy is based on the idea that in order to learn how to program well students should be programming at least every other day. Time has been allotted in class for working on assignments and you are expected to work on them outside of class. It is intended that you should have finished as much of the assignment as you can based on what we have covered in class by the following class period. Therefore, even if something unexpected happens at the last minute you should already be close to done with the assignment. It also allows me to provide rapid feedback by returning assignments quickly, which is crucial to learning.

Final grades will be assigned on the following scale:

- **A** 90-100
- **B** 80-89
- **C** 70-79
- **D** 60-69
Student’s Responsibilities
Students are expected to read/view assigned material prior to the class for which they are scheduled, attend class, participate in class, complete assignments, take the exam, and ask for help early if they are having trouble.

Instructor’s Responsibility
I expect myself to review the assigned material prior to the class for which they are scheduled, prepare and deliver high quality introductions to the material, prepare exercises and assignments that are relevant to research in biology, prepare exams that are fair and representative of the material covered in class, and provide comments on assignments intended to help students develop their abilities to work with computers and data.

Academic dishonesty or misconduct
Academic dishonesty or misconduct is defined in the Oklahoma State University Policy and Procedures Letter 2-0822. You should become familiar with this document as a matter of self-interest. There is an absolute zero tolerance policy for academic dishonesty or misconduct (without exception). Additional information, policies and support can be found at http://academicintegrity.okstate.edu/.

Disabilities
If any member of this class feels that he/she has a disability and needs special accommodations of any nature, we will work with you and the Office of Student Disability Services, 315 Student Union or http://sds.okstate.edu to provide reasonable accommodations to ensure that you have a fair opportunity to perform in this class. Please advise Dr. Canaan of such disability and the desired accommodations as early as possible to address your needs in a timely manner.

Oklahoma State University Syllabus Attachment: We are very concerned about your success as a student at OSU. The information at the following link is provided to answer questions most often asked by students: http://academicaffairs.okstate.edu/images/Patty/FacultyandStaffResources/Syllabus/spring%202014%20syllabus%20final%20with%20copyright%20statement.pdf. A copy of this course syllabus and the OSU Syllabus Attachment are posted on the OSU D2L website for this course for your convenience.

Class and Classroom Manners
I do not take attendance and therefore I expect that if you are in class you are here to learn. So, please, turn off your cell phones, resist the urge to send emails and text messages, etc. Basically I’m just asking that you be respectful of your fellow students and myself. This class is a collaborative learning experience. If you have already finished with what we are working on then find another student to help.