# Spring 2014 GNET 742

# Introduction to UNIX and Perl Programming for biomedical data analysis

An introductory course designed to present the fundamentals of UNIX operating system followed by introduction to Perl programming with an emphasis on analysis of bio-medical data. Class will utilize a combination of lectures complemented with practical computer instruction.

### *Calendar*

| Date (2014) | | Instructor | Topic |
|---|---|---|---|
| | | | |
| 13-Jan | M | Hemant Kelkar | Intro to UNIX |
| 14-Jan | Tu | Hemant Kelkar | UNIX operating system contd. |
| 15-Jan | W | Joel Parker | See below |
| 20-Jan | M | MLK Day | No Class |
| 21-Jan | Tu | Joel Parker | See below |
| 22-Jan | W | Joel Parker | See below |
| 27-Jan | M | Joel Parker | See below |
| 28-Jan | Tu | Joel Parker | See below |
| 29-Jan | W | Joel Parker | See below |
| 3-Feb | M | Joel Parker | See below |
| 4-Feb | Tu | Joel Parker | See below |
| 5-Feb | W | Joel Parker | See below |
| 10-Feb | M | Joel Parker | See below |
| 11-Feb | Tu | Joel Parker | See below |
| 12-Feb | W | Joel Parker | See below |

## Location and time

Time: 11 am – 12 pm, MTueW

Lectures and computer lab: Biogen Idec Classroom 307, Health Sciences Library.

Materials:  All learning materials will be posted on "Sakai" at Sakai https://sakai.unc.edu.

## Instructors

**Hemant Kelkar, Ph.D.**
Center for Bioinformatics and Department of Genetics
4159 Genome Science Building,
919-962-4458, hkelkar@unc.edu
**Joel Parker, Ph.D.**
5115 Bioinformatics Building
Lineberger Cancer Center and Department of Genetics
919-966-9614, parkerjs@email.unc.edu

**Teaching Assistant for GNET742**

Austin Hepperla
hepperla@unc.edu
Weekly hour for help: TBD

## Class description

This class is designed to teach the fundamentals of UNIX operating system and Perl programming using practical "hands-on" computer instruction.  This module will concentrate on applications of Perl programming to biomedical data/analysis. Target audience is biomedical scientists who are interested in getting familiar with computer clusters for manipulating, parsing, analyzing biological text format data.

## Class Syllabus

(HK) Lecture 1 – Introduction to UNIX operating system

Accessing UNIX account using the secure shell (SSH) program. Basic UNIX commands followed by navigation of the UNIX file system. Transferring files from local desktop to the UNIX server using secure file transfer protocol.

(HK) Lecture 2 – UNIX computing resources at UNC-CH.

Introduction to the X-Windows interface and graphical output from the UNIX server. Using cluster computing resources at UNC-CH. Introduction to the "kure" compute cluster and the load sharing facility (LSF) used for cluster job management.

(JP) Introduction to Perl including history, advantages, and disadvantages of the language

(JP) Definition of a variable and basic uses. Understanding of variable types including scalars, strings, arrays, and hashes

(JP) Reading and writing files and processing of command line arguments

(JP) Understanding basic logic and program control with conditional statements and loops

(JP) Complex data structures focused on processing large genomic data

(JP) Subroutines, regular expressions and modules

(JP) Review and knowledge application

### General Class Policies

Since we are going to be learning about a new operating system and a programming language learning will be in incremental steps. If you miss a class then you may not be able to understand material being covered next day. That said, we understand that sometimes there are unexpected emergencies. If you miss a class please email the instructors with an explanation as soon as practical. Unexplained absences may lead to a penalty reduction in final class grade. Programming logic comes to some intuitively where as some others will find it hard. If you think you are in the latter category then please get in touch with an instructor early on. If you wait till the last week of class to do this then we may not be able to provide additional help.

### Class grade (subject to change)

100 % Assignments. We will have at least 2 (or more) assignments. Expect the last assignment to build upon all you will learn in class and to take 4-5 h of effort to complete.

There will be NO FINAL EXAM for this class.

### Auditing and class size

Class is limited to twenty (20) students.  Auditing will not be allowed due to the hands-on computer instruction.

### Recommended readings

#### General readings

Tisdall, J. (2001) Beginning perl for bioinformatics. O'Reilly and associates.

### Homework assignments

Homework assignments will be due by 6 pm on Friday of the following the week they were originally assigned in. You can generally expect to have the grades for the assignments available a week after the due date.

We encourage talking with the instructors or teaching assistant (via email may work best) if you have questions about assignments. Searching the web for a specific question/step is acceptable but you should not copy and paste perl code available on the internet as is.

### Class Examinations

There will be **NO FINAL EXAM** for this class.

### Linux server access for class

Compute cluster access will be provisioned for this class with assistance from ITS-Research Computing and will make use of the "kure" cluster. If you are registered student for this module then an account will be automatically added for you on the "kure" cluster. Logins will be based on your "onyen". Access to this server will remain available through the duration of the module and will allow you to work outside the class hours for assignments/projects.