**Practical Computing for Biologists**

**Introduction**

The CFAR Biostatistics and Computational Biology Core is conducting a four-day workshop for Duke researchers to learn how to use the computer more effectively for scientific work. It is designed for people who need to work with large and complex data sets and suspect that there is a better and faster way to get their work done. The course will use the textbook *Practical Computing for Biologists* by Steven Haddock and Casey Dunn, and CFAR is generously giving each participant a free copy of the book. The main intent of the course is to teach researchers how to use the Unix shell, the Python programming language, databases and image manipulation tools to execute common scientific chores. Participants are expected to use a Mac OS X system since these provide a Unix command line (Linux is fine too). No previous Unix or programming experience is necessary. The course will be limited to 12 participants and will be held at the Surgical Oncology Research Facility (SORF) Beard Conference Room from 29 May 2012 to 1 June 2012.

Please email [cliburn.chan@duke.edu](mailto:cliburn.chan@duke.edu) if you have any enquiries or wish to register for the course. Acceptance will be on a first-come first-serve basis, but CFAR investigators and their trainees will be given priority.

**Course Description**

**29 May 2012 (Tuesday)**

**8-12 AM:** Software installation and working with text editors. We will install the TextWrangler editor, the Enthought Python distribution (Academic license), ImageMagick, ImageJ, MySQL Community Server and MySQL Workbench. Participants are encouraged to install the software ahead of the workshop, but help and troubleshooting will be provided in the morning session if necessary. We will then learn to use the TextWranger editor to understand the basics of regular expressions, and how to reformat text using regular expressions. TextWrangler will also be used to develop programs from Day 2.

**1-5 PM:** Many operations on large file sets, especially for text data, are performed much more efficiently from the command line than from a graphical interface. We will learn how to open a Terminal, and perform text processing, access material from the web, and write simple shell scripts to automate common tasks. We will also learn to transfer and synchronize files with remote computers from the command line, or run programs on remote computers using the command line (ssh) or graphical user interface (VNC).

**30 May 2012 (Wednesday)**

**8-12 AM:** Day 2 introduces you to the Python programming language, a modern dynamic language that is (relatively) easy to learn. The morning session will introduce you to the powerful IPython interpreter, where you will test out code snippets with instant feedback, and learn about the Python documentation and help system.

**1-5 PM:** The afternoon will introduce you to Python scripting, including decisions and loops, reading from and writing to files, organization into modules and libraries, and how to write tests and debug your code.

**31 May 2012 (Thursday)**

**8-12 AM:** Day 3 is for learning to make use of the standard Python scientific computing packages (numpy, scipy, matplotlib). The morning will show how to perform simple numerical calculations, data analysis and generate simple graphics.

**1-5 PM:** In the afternoon, we will continue with learning how to customize graphics, and the use of Python libraries for bioinformatics applications. Optimization techniques to speed up Python programs will be shown.

**01 June 2012 (Friday)**

**8-12 AM:** On the final day, we will examine two very common computing tasks – working with relational databases, and working with images. The morning will be a lesson on the use of MySQL, an open source relational database, including database design, query and updating. Using MySQL from Python will also be shown.

**1-5 PM:** The course will conclude with an introduction to the use of computing techniques to process and extract quantitative data from images, including the use of ImageMagick, ImageJ and Python image manipulation libraries.

**Instructor: Cliburn Chan, Biostatistics and Bioinformatics.**

Cliburn is a computational biologist whose main research interest is in data analysis and modeling of immune responses. He teaches the Introduction to the Practice of Biostatistics I & II courses for the Duke Masters in Biostatistics program, and has been programming in Python for over a decade. Other instructors will be Jacob Frelinger, a PhD student in the Computational Biology and Bioinformatics (CBB) program and Adam Richards, a postdoctoral fellow in the department of Biostatistics and Bioinformatics.