Homework **DUE May 6th 2016 12:00 Midnight by EMAIL to mjbuck@buffalo.edu**

Your good friend, Evelyn, a fellow graduate student preformed 3 ChIP-seq experiment for human p63. After performing these three experiments she received the three 100 GB files from the Next-generation sequencing core. The only analysis she was able to complete was running the data through the MACS program to find peaks. She now has three files with the peaks from her experiment and she needs your help. She has a committee meeting on May 15th and she need to determine which genes are bound by p63 and which peaks are found for all three experiments. You can do this assignment with python, perl, or use bed tools.

* Download the three peak datafiles from UBlearns and the refGenes file.
  1. Determine all peaks which are bound in all three experiments. The peaks must overlap each other at least 20%.
  2. For the peaks found in all three experiments what average center across all three peaks? USE average peak center for all following questions.
  3. Annotate each overlapping peak to gene starts within 10 kb of the peak
  4. What is the nearest gene to each peak
  5. For the peaks which are in common for all experiments what terms do they enrich with GREAT?

For this assignment you do the whole assignment in a single python program using the pybedtools module. Please turn in your python program you used and the answers to the above problems by email on May 6th. For problems 1-4 there should be one bedfile turn in.

Before you can access pybedtools you need to do the following on Rush.

module load python

module load gcc

module load bedtools

module pybedtools