# Genetics 211 Final Take Home Exam 2014

**a) Abstract of final project due by Monday 9:00am February 24th.**
**b) Final project write-up due by Sunday 9:00am March 16th.**

The purpose of the Genomics final exam is for you to think about how the extraction and manipulation of genomic-scale data can supplement or inform your research. We are therefore asking you to conduct a small but cohesive set of bioinformatics analyses (including, but certainly not restricted to, writing some Python). The goal is to answer a particular question that is relevant to your research, is part of a foray into potential research projects, or is simply interesting to you. You are encouraged to consider extending and/or automating tasks that you currently perform frequently (for example, clicking through some particular database, designing primers, retrieving sequences or literature of interest, etc.) Your question should not simply be a problem to which Python can be applied, but a problem to which the application of Python makes sense and makes your work easier.

If you cannot think of anything that would be relevant to your research, you are welcome to use a previous problem set or one of the papers we discussed as a starting point. However in this case you will need to extend the goal of the scripts in a significant way.

**Please write an abstract (~1 page) describing your planned analysis**. In addition to thinking about an interesting biological question, below are other questions to consider as you're writing the abstract:

1. What are the starting data, from where are they drawn, and in what format are they?

2. What are the steps in your data analysis pipeline? What kinds of identifiers will you use to keep track of objects? How many scripts will be required, and consider how much compute time might be necessary? Will external software be needed? (We will try to accommodate requests when possible.)

3. How would your analysis scale if you were to expand the project by 10X or more to even larger datasets?

4. What will the resultant data look like, what format will they be in, and how will these data help you answer the question you propose?

**Please e-mail your abstract to gene211-win1314-staff@lists.stanford.edu by 9:00am February 24th**. You will be contacted by February 27th only if we think your project needs rethinking. If you don't hear from us by then (through e-mail), your abstract has been approved and you should go ahead as proposed.

You must use Python for some of your analyses. The minimal requirements for your script(s) are that you must use **all** of the following features to further familiarize yourself with their use:

1. You should run your code through a checker like pych.atomidata.com/code to make sure your python format is consistent; ignore any obviously silly warnings as there will be some.
2. Dictionary, sets and arrays
3. Subroutines
4. Regular Expressions

In addition, your script(s) must also use **at least two** of the following features:

1. Python os.system() or the pipe form of the os.popen() to invoke a program that takes input and generates output that is then processed by the calling script (for help see http://www.cyberciti.biz/faq/python-execute-unix-linux-command-examples/)
2. BioPython or other Object-Oriented modules
3. Python's urllib

We have set up a page on the class web site that contains some potentially useful hyperlinks to resources. The URL is

http://www.stanford.edu/class/gene211/handouts/final/

The submission of your final should consist of two parts:

1. An electronic document of at most 2500 words (including figure legends and references, but not scripts and output). Your document should follow the structure described below and be **emailed** by Sunday 9:00am on **March 17th**. It should be double-spaced and use a 12-point font.

    - Abstract (150 words)
    - Introduction (1-2 pages)
        - What is the question? Why is it interesting? What is known from the literature?
    - Methods (1 page)
        - Describe your data, analysis steps, programs and/or databases used, and Python scripts written
    - Results (1-2 pages)
        - 1-2 figures, summary tables and description of results
    - Discussion (1 page)
        - What has been learned and what are potential future questions?
    - References (1 page)

2. Submit all Python script(s) you wrote for this final project using the usual command:

    */usr/class/gene211/bin/submit.pl*

When asked which problem set you will be submitting, enter ‘5’. Your script(s) must be well commented, and they must contain your name and an explanation of what it accepts as input and what it outputs. We do not need files generated by the programs unless they are critical for understanding the flow of the analyses. Please contact the TAs if you feel it is necessary to submit a large data file.