Quantitative Bioinformatics   **Homework 1**   **Due on Tuesday 01/28/2014 at 6pm**

> **An Important Note on Homework Assignments**: graduate students are to submit their own answers/report; undergraduate students are to work in assigned teams of two-three students (hopefully the same team throughout the semester). Thus, undergraduate students will submit one solution report per team and the homework grade will apply to all students in that group.

Directions: Clearly, concisely and accurately answer the following exercises; type up your solutions or write *very* neatly. As noted, email submissions are not accepted – turn in a hard copy of your solutions.

1. [All Students] Read in the text: Chapter 1 (pp.1-36), Chapter 2 (pp.37-65), and using R, read & work through Appendix A: A Brief Introduction to R (pp.479-498). (Nothing to turn in for this exercise, just understand the material and get familiar with R.)

2. [All Students] Sequences:
    (a) What is the complementary sequence to the following string of nucleotides? Be sure to label the 5' and 3' ends of the sequence that you write: **5'-GGATCGTAGCCTA-3'**
    (b) What is the complementary **RNA** sequence to the following string of nucleotides? Again, be sure to label the 5' and 3' ends of the sequence that you write: **5'-ATAGCATGCAGACCATGACTTCGTAGTGCG-3'**.
    (c) Assume that the RNA sequence obtained in part (b) is prokaryotic. Give the corresponding sequence of amino acids (in the expression order).

3. [All Students] Do exercise 1 on pp.62-3.

4. [All Students] Do exercise 4 on p.63. In parts (b) and (c) write down your R program (code) and results.

5. [Graduate Students only] Using the DNA sequence below (or one with the mapping A=1, C=2, G=3, T=4), write an R program to (a) read in the data, (b) tally the respective number of A, C, G and T, and (c) tally the number of transitions from {A, C, G and T} to {A, C, G and T} in all 2-tuples arranging the results in a 4×4 table (similar to the one given in Equation (2.27) on p.54). Provide your R program and results.

```
ACAAGATGCCATTGTCCCCCGGCCTCCTGCTGCTGCTGCTCTCCGGGGCCACGGCCACCGCTGCCCTGCC
CCTGGAGGGTGGCCCCACCGGCCGAGACAGCGAGCATATGCAGGAAGCGGCAGGAATAAGGAAAAGCAGC
CTCCTGACTTTCCTCGCTTGGTGGTTTGAGTGGACCTCCCAGGCCAGTGCCGGGCCCCTCATAGGAGAGG
AAGCTCGGGAGGTGGCCAGGCGGCAGGAAGGCGCACCCCCCCAGCAATCCGCGCGCCGGGACAGAATGCC
CTGCAGGAACTTCTTCTGGAAGACCTTCTCCTCCTGCAAATAAAACCTCACCCATGAATGCTCACGCAAG
TTTAATTACAGACCTGAA
```