**CSC 314, Bioinformatics Lab #2:** **Name:**_____
**DNA and Complementary Base Pairing**

DNA is a double stranded molecule composed of complementary base pairs. If the sequence of one strand is known, the *complementary sequence* (***complement***) can be determined based on the following rules: adenine (A) binds with thymine (T) and vice-versa; and cytosine (C) binds with guanine (G) and vice-versa.

If a sequence is read from its 5' to 3' end, its ***reverse*** is the same sequence read from its 3' to 5' end (and vice-versa).

The ***reverse-complement*** of a DNA sequence is the reverse of its complement.

1. Find the specified sequences based on the sequence below. Note that your answers must include labels for the 5' and 3' ends.

   5'-TAGATGCAT-3'

a) Find the reverse sequence

b) Find the complement of the original sequence

c) Find the reverse-complement of the original sequence

d) If the original sequence (5'-TAGATGCAT-3') was an RNA sequence instead of a DNA sequence, what would its sequence be?

2. Write a Python program that prompts the user to enter a DNA sequence, from its 5' to 3' end, and then outputs the following:
   a. The sequence entered by the user (formatted for invalid characters – see below).
   b. The length of the sequence
   c. The complementary sequence (don't forget to label the 5' and 3' ends)
   d. The reverse complement (don't forget to label the 5' and 3' ends)

   Note: Your program should add the 5' and 3' labels, but these labels should not be entered by the user, nor stored by the program. For example, a user entering the sequence in question (1) would enter `TAGATGCAT` and the program would output `5'-TAGATGCAT-3'`. Your program should also work for sequences containing either lower- or uppercase letters. This perhaps is most easily accomplished by converting the user's sequence to lowercase or uppercase.

   For parts (b) - (d), if a character in the sequence is not valid (i.e., is not a T, C, G, or A), then the invalid character should be replaced with a '-'.

3. Respond to the questions posted on Piazza, related to the Implications of Cheap Genomic Sequencing.