American University of Armenia **June 2016**
Introduction to Bioinformatics

Hands-On Twenty
Transcription, Translation, and GC Content

1) Read, understand, and run transcription_coding.py.
The program performs transcription given a DNA coding strand.
The program uses the for loop and if/else conditional construct, string concatenation with the operator "+" and escape sequence "\t", where t stands for tab.

2) Read, understand, and run translation.py.
The program performs translation of a coding sequence (CDS). A CDS starts with the start codon: ATG, and ends with one of three stop codons: TAA, TAG, or TGA, and its length is a multiple of three.

The program uses the dictionary data structure to store the genetic code, slicing a list, which consists in accessing a subrange of a list, and the built-in python function range().

The function range() creates lists containing arithmetic progressions. It is most often used in `for` loops. Examples of range().

```
>>> range(10)               # range() with 1 argument.  range(stop)
    [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
>>> range(1, 11)            # range() with 2 arguments. range(start, stop)
    [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
>>> range(0, 30, 5)         # range() with 3 arguments.
    [0, 5, 10, 15, 20, 25]                          range(start, stop, step)
>>> range(0, 10, 3)         # range() with 3 arguments
    [0, 3, 6, 9]
```

[https://docs.python.org/2/library/functions.html?highlight=range#]

3) Read and understand draft_gc_content.py. The program does not run. All instances of "XXXXXXXXXX" need to be replaced for the program to run.
The purpose of the program is to compute the percentage of G's and C's in a given DNA sequence: GCAACCGGTTACGCAAACAG.

Replace "XXXXXXXXXX" by appropriate code so as to get the program running properly. Rename the program: gc_content.py.

4) Instead of computing the number of G's & C's of a small DNA sequence, as in Problem 3, draft_gc_content_NCBI.py reads the input sequence from a file retrieved from NCBI.
- Go to http://www.ncbi.nlm.nih.gov/
- Choose "Nucleotide" from the drop-down menu and type "U14680" in the search area
- Click on the blue "Search" button to go to the "BRCA1" CDS record
- Click on "send", choose "coding sequences" and click on "Create File"
- Open the file, remove the first line of information and save the file under "BRCA1.txt"

If you try running draft_gc_content.py, you will get an error message. Copy the path spelled out in the error message into the first argument of open() right before "BRCA1.txt".
Now run the program again.