CSC 314
Cancer Biology Assignment (Proposed Methods)

The objective of this analysis will be provided by either myself or your collaborators.

In order to meet this objective, you will need to use the Bioinformatics databases and tools described in this class, which are listed below.

- GenBank
- NCBI Protein Database
- UCSC Genome Browser
- UCSC Table Browser
- BioMart
- Biopython tools
  a. *Seq* objects
  b. *SeqIO*
  c. *Entrez.esearch* and *Entrez.esummary*
- Python's regular expression (*re*) module

Assignment

1. Clearly explain the objective of what you need to do, and what information you will need to provide to your collaborators.

2. Clearly explain how you will accomplish the above task. Specifically, state what libraries, databases, and tools you will need to use, and for what purpose. For example, if you will be downloading sequences, which database or tool will you use to get the sequences from? You may list these steps in "recipe" form.

The following is a simple example for a different analysis:

1. Objective: the objective of the analysis is to get a list of all human HBB nucleotide sequences and their lengths.

2. Methods:
   a. The Biopython *Entrez* module will be used to query GenBank for human HBB nucleotide sequences. Specifically, the *Entrez.esearch* method will be used.
   b. For each sequence, we will then use the *Entrez.efetch* method to retrieve each sequence in FASTA format, and find the length of each sequence. The ID and length of each sequence will be written to a file.