CSC 314, Cancer Biology Assignment (Proposed Methods)

Your biological collaborators (students taking Biology of Cancer, BIO 426) are requesting your help in identifying a set of genes that may be implicated in cancer. They have previously shared with you a summary of this project as well as its objective. In order to accomplish the objective, you will need to use some of the Bioinformatics databases and tools that we have covered in class, which are listed below.

- Databases:
    a. OMIM
    b. GenBank
    c. PubMed
    d. NCBI Protein Database
    e. UCSC Genome Browser
    f. UCSC Table Browser
    g. HomoloGene
    h. Gene

- Biopython:
    a. Working with sequence records in FASTA or GenBank format
    b. Querying and retrieving data from Entrez (NCBI databases),

- Python's regular expression (*re*) module

Assignment

1. In approximately one paragraph, clearly explain the objective of what you need to do, and what information you ultimately will need to send to your collaborators.

2. Clearly explain how you will accomplish the above objective. Specifically, state what databases and tools you will need to use, and for what purpose. Note that you will need to use Python (and BioPython) to accomplish the above, and you should clearly state what Python modules will be needed and for what purpose. For example, if you will be downloading sequences, you should state the database and tools that are needed to do this. You may list these steps in "recipe" form.

The following is a simple example for a different analysis:

1. Objective: get a list of all human HBB RefSeq nucleotide sequences and their lengths.

2. Methods:

    1. The Biopython *Entrez* module will be used to query GenBank for human HBB nucleotide RefSeq sequences. Specifically, the *Entrez.esearch* method will be used to perform the search.
    2. For each sequence (gene id), 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.