

Assignment #2
Due Thur, Jan 24

1. Write a Python function that takes as input a sequence string and returns a list with 4 entries that are the number of A, C, G, and T in the sequence.
2. Write a Python function that takes two inputs: a sequence string and a string of two letters (e.g., "CG" or "CT"). This function returns the number of times the two letters occur consecutively in the sequence.
3. Write a function that takes as input a sequence string and returns a list with 16 entries that are the outputs of function #2 for all 16 possible two letter strings.

If you are not familiar with it, explore the NCBI website. Go to the following two pages and download the FASTA files for the human gene *PTPN11* and it's *Drosophila* orthologue *csw*.

https://www.ncbi.nlm.nih.gov/nuccore/NM_002834

https://www.ncbi.nlm.nih.gov/nuccore/NM_057783.3

Turn in the code for the three functions, and for each file the output of functions #1 and #3.

Since humans have much higher rates of methylation than *Drosophila*, we would expect to see far fewer CpGs in humans. Is this what we see?