Create a single python script that performs the following steps.

Assignment

You’re a computational scientist that loves the color red. So much so that you would like to discover all animals that have red hair.

A quick google search tells you a gene responsible for this color (try it yourself). **MC1R**

Sometimes it's not enough to know just the gene name, you need specific ids for certain databases.

**Step 1**. Use the mygene database restful api endpoint to get the ensembl id. Start from the name of the gene (MC1R), and end with the ensembl id. It may take multiple queries to get there. This does not need to be written to a file, you need to use it in your code to do the ensembl database lookup.

<https://www.ncbi.nlm.nih.gov/gene/4157?_ga=2.90939346.911584715.1622506542-741789674.1622506542>

[Ensembl:ENSG00000258839](http://www.ensembl.org/id/ENSG00000258839)

<https://useast.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000258839;r=16:89912119-89920973>

<http://useast.ensembl.org/biomart/martview/e2b3e8eb22848cfa36a6210de85c56b4>

https://rest.ensembl.org/

\*Make sure to only look for human species

Now that you have the ensembl id, use the Ensemble database and RESTful api to get more information about this red hair gene.

**Step 2**. Use the ensembl database endpoints to grab the nucleotide sequence data for MC1R. Write this sequence to a fasta file. (Remember a fasta file always has a header > , followed by a sequence).

Find the longest open reading frame (ORF) and then convert it from DNA to Amino Acid. Feel free to use any function you find online to to get the longest ORF (as long as it doesn't require you to install other pip packages).

Then use **biopython** (from Bio.Seq import Seq) to convert the nucleotide sequence into an amino acid sequence. Write this amino acid sequence to the same fasta file.

**Step 3**. What other species have genes that are homologous to MC1R? Write the full **unique** list of species to a text file call **mc1r\_homology\_list.txt**.

----------------------------------

Your final week8 directory should have a single python script. And two outputs, a fasta file and a homology txt file.

You can use this link to find appropriate endpoints:

[http://rest.ens..embl.org/?content-type=text/html](http://rest.ensembl.org/?content-type=text/html)

Other notes:

Please use the ensembl id for the DNA sequence (not the protein sequences).

Since it is the full DNA sequence it will contain introns, exons, and maybe splice sites.

Find the longest open reading frame (ORF) and then convert it from DNA to Amino Acid. Feel free to use any function you find online to to get the longest ORF (as long as it doesn't require you to install other pip packages)

Assignment 4 due by Tuesday, midnight ET.