# ECS 129

## Assignment: Option5

## Genes and Proteins

**Due:** Wednesday, March 2nd, 2022

### Writing a GeneFinder program

***A. Write a program (your choice for the programming language) that will:***
- Read in a DNA sequence, given from 5' to 3', preferably from a file
- Generate the sequence of the complementary strand,
- Find the longest possible "gene" (or open reading frame) for that DNA; Remember, there are 6 possibilities, as we do not know if the input strand is the coding strand, and we do not know the reading frame. If there are more than one "genes", you may decide to output all of them, or choose the longest one.
- Transcribe this "gene" into mRNA,
- Translate this mRNA into an amino acid sequence, using the one letter code for amino acid name. You will assume that there are no introns.

***B. Test this program on the sequence:***

5' TCAATGTAACGCGCTACCCGGAGCTCTGGGCCCAAATTTCATCCACT 3'

The "expected" protein sequence is:
MKFGPRAPGSALH

***You should provide:***
- A report; do remember that I am expecting you to go "beyond" the prompt. Possible extensions:
  - Discuss options to account for the possible presence of introns
  - Can we design the "inverse" tool, i.e. a tool that would generate a coding
  - DNA sequence given a protein sequence
  - Can your tool be used for a full genome?
- The source code of the program itself