This exam is to be completed by the start of class on October 10, at which time you should hand in a printed copy of the exam. Answer each question as completely and clearly as possible. Although you may discuss the questions with classmates or others, all answers must be your own work, and in your own words. If you copy someone else’s answers, both of you will receive a 0 on this exam. If you quote another source, provide its full bibliographic citation.

1. (10 points) The illustration below shows a partial results page from PubMed. Answer questions 1 through e concerning these results:

   a. Are these single studies or review articles? What is the difference?

   b. Are free full-text versions of these articles available through PubMed?

   c. What happens if you click on one of these icons:

   d. When was the most recent article in the list published?

   e. Which article is not published in English?
2. (5 points) Which of the search terms below is narrower (will produce fewer hits) and why?
   a. “human MAP kinase inhibitor”
   b. “MAP kinase inhibitor” AND human

3. (5 points) Explain the difference (in terms of the results you would expect) between the following queries:
   a. HIV AND AIDS
   b. HIV OR AIDS

4. (10 points) EU131382 is a nucleotide sequence. Find the following information about it:
   a. What organism is the entry from?
   b. How many base pairs in the entry? Is it DNA or RNA?
   c. Is the organism eukaryotic or prokaryotic? How do you determine this?
d. How many coding regions in the sequence? What are the coordinates of each?

e. What protein are proteins are associated with this sequence?

5. (10 points) List 3 genes that are associated with human chromosome 12. Explain how you found this information.

6. (15 points) Use the sequence below to perform tasks a through c:

ACTTGAAACAATAAACACTTCTCTTCGACTGAAAGATTTACTGGAGAGACCCAGGATCATTTGGACTAGTAA
GAAACCATAATGAAACTTTAACATCGCAAGTGTTACAAAGAATGGTAGTGATGATAATCTCAAGTATCT
TAATGCTTTGGAAGTTACCTTTGATGGTCAGCAAAACTTTTCAATTGAAGGTTATGATAACAACGGTAGA
GCTTTATATGATATTAACCTTA

a. Does the sequence show evidence of vector contamination? How do you determine this?
b. What is the ratio of AT content to GC content in this sequence?

c. How many ORFs does the sequence contain? What is the length of each?

7. (5 points) Why is it more likely to find a complete gene in a single nucleotide entry for a prokaryotic organism than for a eukaryote?

8. (10 points) Q9S8P4 is a SwissProt accession number. Answer the following questions about the entry it references:
   
a. When was the entry added to the database?

   b. When were annotations last modified?
c. What is the common name of the organism from which the entry was derived?

d. What amino acid (3-letter code) is found in the conflict region of the feature map?

e. What type of tissue was the sequence taken from?

9. (10 points) Suppose you were interested in studying the gene that produces the protein sequence in question 8. Describe how you would find information on this gene. Include the names and addresses of the sites you would use and what search terms you would use.

10. (20 points) Using the sequence from the record you used in question 8, answer the following questions about the protein:

   a. What is the most common amino acid in the sequence, and what percent of the total protein composition does it account for?

   b. What is the half life of this protein?
c. How many transmembrane segments does the protein contain? What is the basis of your prediction?

d. What are the known motifs in your protein? Include only those with a high likelihood of existence, and explain what criteria you used to determine this.