

Problem A1
Simple Sequence Analysis Operations
Due: January 26, 2006

Hand in written answers to the questions or submit via Blackboard a single Word or PDF document containing your answers with any requested printouts embedded in the document (DO NOT submit individual files for each answer or printout).

Questions (Total of 40 points)

1. Do problem 1 from Chapter 2 of the Mount textbook. Hand in answers for parts c (how many hits), d (how many hits), e (how many human heat shock factors), and f (how long is the sequence), and printouts of the protein sequence saved in part j and the mRNA sequence found in part k (NOT the genome sequence). DO NOT print the web pages; save the sequences to a file on your computer and then use a text editor, word processing or other utility to print the file using a *fixed-width font*.
2. Do problem 3 from Chapter 2 of the Mount textbook. Hand in written answers to the questions.
3. Use the Entrez Gene search page (accessible from the Entrez Nucleotide search page using the link “Search for Genes”) to find the entry for the “giantin” gene from mouse (species name: *Mus musculus*)
 - a) What is the gene name that Entrez uses for it?
 - b) What is its Gene ID?
 - c) What is the TrEMBL accession number for its mRNA?

Extra credit (5 points each)

4. Using Entrez Gene, find the “igf1” gene from *Rattus norvegicus*.
 - a) What is the long name of the gene?
 - b) Find the Unigene link for the gene and click on it. What is the percentage of similarity between the rat gene and the corresponding human gene?