Fall 2003 Genomics Exam #1
Genomic Medicine and Sequencing Tools

There is no time limit on this test, though I have tried to design one that you should be able to complete within 4 hours, except for typing and web searches. There are three pages for this test, including this cover sheet. You are <u>not allowed discuss the test with anyone</u> until all exams are turned in at 11:30 am on Wednesday October 1. **EXAMS ARE DUE AT CLASS TIME ON WEDNESDAY OCTOBER 1**. You <u>may</u> use a calculator, a ruler, your notes, the book and the internet. I have to say, this is a challenging test, so do NOT put it off too long. You may take it in as many blocks of time as you need to.

The **answers to the questions must be typed on a separate sheet of paper** unless the question specifically says to write the answer in the space provided. If you do not write your answers in the appropriate location, I may not find them. You will need to capture screen images as a part of your answers which you may do without seeking permission since your test answers will not be in the public domain. If you are asked to print out any pages, you do not have to print in color, though it is permitted.

-3 pts if you do not follow this direction.
Please do not write or type your name on any page other than this cover page.
Staple all your pages (INCLUDING THE TEST PAGES) together when finished with the exam.

Name (please print):

Write out the full pledge and sign:

How long did this exam take you to complete (excluding typing)?

1) Start with this partial sequence:

```
atggcaaagg tcacaaactg ggataagttt gtgctgcttt tgtggaagaa
ctggaccctc caatggaacc acaagtggca gatggttata gagctggtgc
tgccagcgat attctccctg ctcctcgttc tagtccgcac cttggtggat
acggagcaaa aaggagtccg gtattataat gagcagaact taacagacct
caatctgctg cagaagaatg gcggcttttc aaaatttgag ttcatcctgt
gctactcgcc cgtcaatccc gtgttgaaga aactggtaga agaggcgtgg
cagagccttg gtaagaacaa aatctgtgag tcggaaaatg ccacccaact
ggagttggat
```

a. What is the species and gene name for the full sequence?
b. Find the human ortholog. (no answer needed for letter b)
c. What is the name of the encoded human protein?
d. Is the human gene/protein associated with any human diseases?
e. How long is the complete mRNA (document your answer with an accession number).
f. How many amino acids are in the largest ORF of the human sequence. Document your answer with data.
g. Where is the human gene located?
h. Are there any STS markers for this gene? Document your answer.
i. Is the human protein post-translationally modified? Explain your answer and support it with data.
j. Describe the human protein's cellular roles in Gene Ontology terminology.
k. Is the human protein an integral membrane protein? Support your answer with evidence showing how many times it spans the membrane or that it does not span the membrane.
l. Document one SNP that produces a disease-causing allele.
m. Are there any non-coding mutations that can cause this disease, or are all documented cases caused by changes in the amino acid sequence? Document your answer.

2. Here is a tricky sequence for you to identify.

```
ATGACAAGGGTCTCAATCGATCGTAATCTTCTTGACCGGCCGTATCAAACAAACCTAAC
GTATATGGTTCATCACCAATCATCACAGTCACCGCATAGTTATCGAACACTGTTGGAAC
ATAGTCGGCTGGAAATTGATTCGTTGTATAGGAGATTAGAAGGCACGTTTTCCCAACAG
CACCATCACCGACAACAACACACTTTAGCGTTTGCATTTTGTGGAAGAGCTAATACGTT
TATTTCTTGTTTTATAAGTTTTGCTTCTCTAATTCGTTTGTTAACCTATCTCCTCAGGA
AAATAGAATAG
```

Your task is to tell me why this is a tricky sequence. Use any online tools you can to document why you should not get a single answer. What I want you to do is tell me the two correct answers. Document how you came to your conclusions.

3) Find the structure file for "Rho transcription terminator"
a. Give me the most recent publication citation for this structure.
b. What analogy do the authors use to explain this protein's mode of action?
c. What is the PDB ID number?
d. How many protein subunits are present? Tell me both the number of subunits and how many different protein sequences are present in your structure.
e. Show me an image of this structure that matches the analogy and shows the number of subunits.

4) Go to the Human Mouse Homology map and focus on Human Chromosome 7.
a. Find a very small region on the human chromosome that has its ortholog on a small piece of a mouse chromosome that is not syntenic with other mouse genes shown on this image. Show me an image of which gene you found.

b. Find a gene in the human map that appears to be mislocated. Name that gene and show a picture of the gene you have chosen. Explain why you think this human gene is mislocated in this map. Verify your choice by further evidence using another source of information.

c. Zoom in on small region of the human chromosome 7 and focus on the human gene ICA1.
What is the location of the mouse orthologs for ICA1's nearest neighbors? What is the problem depicted in this image?

d. Use a database to find the chromosomal location of ICA1, the GC islands in this region and the two flanking neighbors of ICA1. Zoom in as tightly as you can and take a screen shot.

e. Which human gene neighbor is closer to ICA1?

f. Describe the location of the GC islands relative to the genes you tracked down?

g. Does the human mouse homology map show all the genes in this area of the human genome?

h. Given what the authors Eichler and Sankoff (the PDF file I emailed to you a couple weeks ago) said about chromosomal break points, what is surprising about this region of the human genome?

5) a. Find the tissues that express ARHGDIB and DAT1 (DAT1 is also called SLC6A3). Make sure you are working with the two genes located on chromosome 12.

b. Which gene appears to be a housekeeping gene based on expression patterns?

c. What structural feature about housekeeping genes can you detect when comparing these two genes when looking at overall gene structure?