

**HOMEWORK -- Please don't try to write on this sheet!** Put your answers on another sheet of paper and turn that in. I encourage you to explore the NCBI site in groups, but I expect each of you to find the answers to these questions by yourself, working alone, after you feel comfortable with the site. **This is due Thursday, Sept. 25.**

**Resources: my own handout** ([www.life.umd.edu/labs/mount/Handout-Internet.pdf](http://www.life.umd.edu/labs/mount/Handout-Internet.pdf))

**NCBI tutorial:** [www.ncbi.nlm.nih.gov/Education/](http://www.ncbi.nlm.nih.gov/Education/)

1. (2 points) Find the protein associated with accession number NP\_001776.  
What is the name of this protein and in what species is this protein found?

Questions 2-6 are related (1 point each)

2. What **disease** is associated with the *PRPF8* gene?

3. What **gene** is associated with Tay-Sachs disease? Provide the human gene name (in the form *ABCI*).

4. What name is given to the protein product of this gene?

5. Give a **refseq** accession number for the **protein** product of that gene.

6. What is the degree of amino acid **identity** between that human protein (your answer to number 5) and the corresponding mouse protein? (give the number of identical amino acids / length of match for the entire alignment).

7. (1 point) What is the map position for the **gene** encoding the protein in question 1 (NP\_001776) in humans (e.g. 10q11 would be a map position)

8. (2 points in two parts) Use **tblastn -- protein query and nucleotide database** -- to search the *Homo sapiens* genomic DNA sequence (with the protein sequence CAA57975).

(1 point) What result do you get (describe your top hits and report their E values)?

(1 point) What gene is the best match (use *ABCI* format)?

9. (2 points) Now use **blastn -- nucleotide query and nucleotide database** -- to search the same database with the nucleotide accession X82649 (the nucleotide sequence from which the protein accession used in 8 was generated).

(1 point) What result do you get (describe your top hits and report their E values)?

(1 point) Does this query find the same region? Why or why not? Which is more accurate?

10. (2 points) The human U1A protein (NP\_004587) binds U1 snRNA while the homologous human U1B" protein (NP\_003083) binds U2 snRNA. **What is the evidence for this statement?** \*

11. (6 points in four parts) In some species there is a single homolog for U1A and U2B". An example is *Drosophila melanogaster*.

(1 point) What is the **refseq accession number** for the *Drosophila* protein that is homologous to both?

(1 point) Does this protein bind U1 snRNA, U2 snRNA or both?

(2 points) **What is the evidence for this statement?** \*

(2 points) Make an alignment of these three proteins (e.g. with clustalw). Attach your result.

\* **Evidence.** Please write between 100 and 400 words (no more) **including a citation to the primary literature and a description of the experiment that makes the point.** Failure to cite primary literature and describe the nature of the evidence will result in no credit for these questions.