

**HOMEWORK** -- Please put your answers on another sheet of paper and turn that in class. 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. 29.**

**Resources:**

**NCBI tutorial:** <http://www.ncbi.nlm.nih.gov/Education/index.html>

**Young text:** <http://www.whfreeman.com/young/> (**registration and login required**)

1. (1 point) Use Entrez to find the protein associated with accession number NP\_002006. What is the name of this protein and in what species is this protein found?
2. (1 point) What conserved domains (not COGs or KOGs) are found in this protein (NP\_002006)?
3. (a 1 point) A paper was published in 2003 with N. Libina, J.R. Berman and C. Kenyon as authors. What gene is named in the title of this article?  
(b 1 point) What species was used for the research described in these articles (provide the latin name)
4. (Questions 4-8 are related) (1 point) What **disease** is associated with OMIM entry 602081?
5. (1 point) What **gene** is associated with the disease described under OMIM entry 602081? Provide the human gene name (in the form ABC1).
6. What protein is the product of this gene? (6a. 1 point) Give a **refseq** accession number for the **protein**.  
(6b. 1 point) Give the name for the **protein**.  
(6c. 1 point) What is the Homologene ID? (HomoloGene:??)
7. (1 point) What is the most closely related mouse (*mus musculus*) **protein** (your answer is the **refseq** accession number for the protein)? You can get this from BLAST, Blink, or HomoloGene); try them all!
8. (1 point) What is the degree of amino acid **identity** between the mouse and human proteins? (give the number of identical amino acids / length of match).
9. What **protein** in the yeast *Saccharomyces cerevisiae* is most related to the protein encoded by Mendel's round pea locus (see your text)?
  - a. (1 point) What is the protein name (in the form Abc1p)
  - b. (1 point) What is the refseq accession number for the protein?
  - c. (1 point) What is the degree of amino acid **identity** between the pea and yeast proteins? Give the number of identical amino acids / length of match.
  - d. (1 point) What is the E-value for this blastp hit and how is this value influenced by your choice of filtering options?
10. (1 point) What is the map position for the **gene** encoding the protein in question 1 (NP\_002006) in humans (e.g. 7q11 would be a map position)
  - 11a. (1 point) Use **tblastn -- protein query and nucleotide database** -- to search the *Homo sapiens* DNA sequence with the protein sequence XP\_689969. What result do you get (describe your top hits and report their E values)?
  - 11b. (1 point) Now use **blastn -- nucleotide query and nucleotide database** -- to search the *Homo sapiens* DNA sequence with the nucleotide accession XP\_689969 (the nucleotide sequence from which the protein accession used in 11a was generated). What result do you get?
  - 11c. (2 points) Compare your results from 11a and 11b. Why do you get different results? How are you sure? Which type of search is better (tblastn vs. blastn) when searching for this kind of homolog? Why?