

NCBI Worksheet: **See the handout on our website.**

Name _____ GSI & Sect # _____ Station # _____

1. Name three databases other than Entrez Gene that you can use to find information related to your gene. What kind of information is available on these databases?

2. a) What are the score and E-value for your top hit?

b) What is the percent identity between your query and top hit?

c) How many amino acids align?

3. a) What is the official symbol and full name of your gene?

b) What specific biochemical processes does the protein encoded by your gene perform?

c) Name one fact from the published literature relevant to your gene.

4. a) List the disease(s) associated with your gene.

b) Describe how mutations in a single gene could cause multiple diseases.

c) Briefly describe a disease associated with your gene. Include clinical features like age of onset when available. Describe new or medical terms in your own words.

5. a) How can one genomic sequence encode multiple protein products?

b) List the genes located near yours.

c) Record the gene map locus of your gene.

6.
 - a) What is the GenBank record number of the mRNA for your gene?
 - b) How many base pairs are in the mRNA sequence?

7.
 - a) Which reading frame gives the correct protein?
 - b) How many amino acids are in the virtual sequence?
 - c) What is the molecular weight of the protein?
 - d) Name one function of an untranslated region.

8.
 - a) After deleting a block of nucleotides from the mRNA sequence of your gene, describe the mutated protein product in terms of number of amino acid residues and molecular weight compared to the wild type protein.
 - b) When you make a single base pair substitution, why are silent mutations so common?
 - c) List the amino acid change for a missense mutation using proper notation (for example, A223V signifies that the 223rd amino acid residue has been changed from alanine to valine.)
 - d) How can an insertion or deletion in the mRNA not affect the protein sequence?

9.
 - a) Discuss the phenotype of mutations in your gene in another organism.

b) How different is the mouse ortholog of your gene?

c) Why does d increase across the species listed, while d_N/d_S hits a plateau?

d) Which would you expect to have a higher d_N/d_S , orthologs or paralogs? Why?