Genes and Disease Worksheet

Part A: Genetic basis of disease

1. What is the difference between a gene that causes a disease and a gene that predisposes someone for a higher risk of a disease? Give examples.

Part B: Genome sequencing is awesome and cheap(er)

1. Go to genome.gov/sequencingcosts/ - click on the plot with the orange lines titled "Cost per Genome".
2. What was the cost of sequencing the human genome ...
   a. In October 2001?
   b. In October 2005?
   c. In October 2010?
   d. Today in early 2012? (ask)
Part C: Find a disease gene via OMIM

Write down the disease you're exploring:

__________________________________________________

1. Go to omim.org (OMIM = Online Mendelian Inheritance in Man) and search for your disease.

2. Click the top result that starts with a “%” or “#” - this means the result is a phenotype and will give you a list of possible genotypes.

3. Scroll down to Phenotype Gene Relationships and pick a gene to explore.

4. Write down the gene you chose:

Write down the gene you chose:
Part D: Explore gene function via GeneCards

D.1 - GeneCards: Searching and Summaries

1. Go to genecards.org and search for your gene at the top. Click on the top hit for your gene. Make sure the “Symbol” matches the name of your gene.

2. Scroll down to “Summaries for ____ gene” on the left and “Entrez Gene summary for ____”

3. In 1-2 sentences, what is function of the protein coded by this gene? Does it perform some specific chemical reaction?

4. What biological process does it govern? (e.g. DNA Damage for BRCA2)

5. What diseases are associated with this gene? (Including, but not limited to, the disease you’re exploring)

D.2 - GeneCards: Proteins

On the left, open up the “Jump to Section...” drop down menu and choose “Proteins”
1. Under “UniProtKB/Swiss-Prot”, bullet point “Subunit”, what molecules and proteins does your protein interact with? Are there special conditions for any of the interactions?
   a. Size of your protein: _______ amino acids __________ Daltons

2. Is your protein part of a complex? If so, what other subunits are in the complex?

3. Scroll down a bit until you see “Gene Ontology (GO)”. .

4. What is the top GO term for the cellular location of your gene? Does that make sense with its biological function?

D.3 - GeneCards: Genomic Views

In “Jump to Section...”, chose “Genomic Views.”
6. **Where is this gene located in the genome?** The “address” is described by the “cytogenic band,” which looks like this: 9q23.3

7. In a new tab or window, go to [ghr.nlm.nih.gov/handbook/howgeneswork/genelocation](ghr.nlm.nih.gov/handbook/howgeneswork/genelocation)

   a. **What chromosome is your gene on?**
   
   b. **Which arm? (circle one)** long / short
   
   c. **Which position?**

8. Go back to the GeneCards tab. **How many base pairs (bp) are in your gene?**

   Go to: “GeneLoc gene densities for chromosome ____” (under the chromosome diagram)

9. **How many genes are on this chromosome?**

---

**Part E: cDNA transcript analysis via *Ensembl***
Scroll up to the top and click the Ensembl ID, which will look like this: “ENSG00000141510” (this opens in a new tab/window) You should get a page that looks like this:

You’ll explore two transcripts from your gene. You’ll do the first transcript under “Transcript ID” and then another one that you choose--anyone you like!

First transcript ID: __________________________________________

Transcript ID you chose: ______________________________________

How many exons does this transcript have? ______________________

How many exons does this transcript have? ______________________

Click “Exons” on the left (you have to be viewing a transcript to see this option). The green letter sequence is the upstream sequence of the gene, the purple letters are the untranslated region, the black is the exon gene sequence, and the blue is the intron sequence. Note: This is cDNA, a reverse complement copy of the mRNA that this came from.

What are the first three letters of the gene sequence? ______________

What are the first three letters of the gene sequence? ______________

What do these three letters represent?

What are the first two letters of each intron? ______________

What are the first two letters of each intron? ______________

Last two letters of each intron? ______________

Last two letters of each intron? ______________

How long is the longest intron (in bp)? ______________

How long is the longest intron (in bp)? ______________

Shortest intron (in bp)? ______________

Shortest intron (in bp)? ______________

How long is the longest exon (in bp)? ______________

How long is the longest exon (in bp)? ______________

Shortest exon (in bp)? ______________

Shortest exon (in bp)? ______________

Which are longer in general, introns or exons?

Part F: Genomic views via UCSC Genome Browser
F.1 - mRNA transcripts in the UCSC Genome Browser

Go back to the GeneCards site and the “Genomic Views” for your gene. Click “UCSC Golden Path with GeneCards custom track” (this will open in the window/tab that Ensembl opened). You should get a page that looks like this:

There is a LOT of data here. Let’s condense it. Under the red boxes and lines, click “Human ESTs Including Unspliced.” This “rolls up” the track like a window shade. [“ESTs” stand for Expressed Sequence Tag and is an old technology (from 1993) that was used to figure out what mRNAs are present in an organism. Now we have better technology called RNA-Sequencing (publicly unveiled in 2008) which gives us the “Human mRNAs from GenBank” track that we will use.]

Your Genome Browser should now look like this:

The track Human mRNAs from GenBank is a bunch of real mRNAs from real human samples (either biopsies from people or human cell lines). Because RNA-seq isn’t available to everyone, most of the subjects in the database are there because they have a disease or condition that scientists are studying.

The thick boxes are exons and the lines are introns. The longer boxes at the beginning and end are the untranslated regions (remember the first and last rows in the Ensemble table?)

1. Do you see a pattern in what exons and introns are excluded/included in the mRNAs? Explain the pattern

2. Does your gene go from right to left or left to right?

F.2 - Conservation via UCSC Genome Browser

Let’s see how these introns and exons are conserved across species.
Scroll down to the bottom. Under **Comparative Genomics**, there are a bunch of categories. Under **Conservation**, click open the drop down menu and choose **pack**.

Your genome browser should now look like this:

1. **Which organism(s) that came up in conservation is most surprising?**

2. **Which are more conserved, exons or introns? Why?**

3. **Why do you think sequences outside of the gene are conserved?**

4. **If your disease is caused by a sSNP, search for your rs ase in [snpedia.org](http://snpedia.org), then search for that SNP (will look like “rs334”) in the genome browser. How does this SNP affect the protein? Is this site conserved?**