



We will Study the Gene Sequence of the BRCA1 Gene

The gene named Breast Cancer 1 (BRCA1) is an important tumor suppressor gene that is responsible for repairing DNA. As you might guess from its name, when this gene isn't working right, it can lead to breast cancer. Certain mutations in BRCA1 are responsible for a large proportion of inherited breast and ovarian cancers (which make up only 5-10% of these cancers).

The protein encoded by the BRCA1 gene combines with other tumor suppressors, DNA damage sensors, and signal transducers to form a large multi-subunit protein complex known as the [BRCA1-associated genome surveillance complex \(BASC\)](#). The BRCA1 protein works with [RNA polymerase II](#) and also interacts with [histone deacetylases](#) (which de-acetylates histones and hinders transcription). This makes a lot of sense as cancer is caused by changes or mutations in the DNA. If a gene responsible for correcting DNA mistakes breaks, then mutations won't get fixed and will build up. Eventually a mutation will happen somewhere important and bam, cancer.

The BRCA1 gene isn't just found in humans, it is present in all mammals (and some other animals too!).

Your Pre-Lab Activity: You will use traditional phenotypic information to compare four different organisms to humans, and then construct a cladogram showing the evolutionary relationship between the five organisms.

Lab Activity Using BLAST: In [Part I](#), you will compare a small piece of the BRCA1 gene in these organisms, and determine the DNA similarities and degree of relatedness. In [Part II](#), you will also find out which amino acids in the BRCA1 protein are similar and different among the five species. Finally in [Part III](#), you will study different human BRCA1 mutations and make an educated guess as to whether certain mutations might result in a cancerous cell.

More About BLAST: We will use the sequence comparison tool BLAST (Basic Local Alignment Search Tool) to figure out which DNA sequence belongs to which mammal. BLAST is a powerful tool provided for free to all researchers (and students) by the National Center for Biotechnology Information ([NCBI](#)). This program is used millions of times a month by scientists all over the world to compare unknown DNA sequences with the complete database of public sequences. BLAST also allows you to locate sequences that are similar but not identical to unknown DNA. **A short tutorial on using BLAST will help you in your exploration.** There are two ways to do the tutorial.

1. Go to: apbiostudio.stanford.edu/biocomputation/BLAST-Tutorial.html and then use blast.ncbi.nlm.nih.gov.
2. Use the printed version of the BLAST tutorial and practice the activities that allow you to become familiar with the BLAST website.

Part I of the Lab Activity: After the Tutorial, you are ready to use BLAST to figure out to which animal each of four DNA sequences belongs. Use BLAST to determine the genus and species for each of the unknown sequences and the animal's common name. Here's what you do:

- Go to: apbiostudio.stanford.edu/biocomputation/Unknown-Sequences.html
- Copy the first DNA sequence (just the sequence, not the label).
- Once you have copied the sequence, paste the sequence into the BLAST form as in the tutorial.
- Collect information about this DNA sequence and record in Data Table 1 below.
- Repeat for the other 3 sequences.

Then, you will record your results in a data table, figure out the identity of each of the organisms and draw a cladogram based upon DNA similarity in the BRCA1 gene. You can check your answers by going to the labs [Photo Gallery](#).

Part 2 of the Lab Activity: Once you have figured out which DNA sequence belongs to which animal, you will compare the sequences to determine which parts of the DNA are the most conserved (and so the most important). Then you will translate a small part of the gene to figure out which amino acids are most conserved between these five animals. You will use this information in Part III to make some educated guesses about which mutations in the BRCA1 gene may have led to cancer.

Shown below are the five sequences you've worked with so far. This multiple sequence alignment shows just a small portion of the full BRCA1 gene from the 3' end of the gene.

```

5335 (coordinates of H. sapiens BRCA1 gene)                                5414
|                                                                           |
human TGGTCAATGGAAGAAACCACCAAGGTCCAAAGCGAGCAAGAGAATCCCAGGACAGAAAGATCTTCAGGGGGCTAGAAATC
#1    --GTGAATGGAAGAAACCACCAGGGTCCAAAGAGAGCAAGAGAATCCCGGGACAGGAAGATCTTCAGGGGCCTAGAAGTC
#2    --GTTAACGGAAGAAACCACCAAGGTCCAAAGCGAGCAAGAGAATTCACGACAGAAAGATCTTCGAGGGCCTAGAAGTC
#3    --GTCAATGGAAGAAACCATCAAGGCCCAAACGAGCAAGAGAATCCCGGGACAAGAAGATCTTCAAGGGCCTGGAAATC
#4    ---TCAATGGGAGAAACCACCGAGGCCAGAAAGAGCAAGAGAATCTCAGGGGATGAAGATCTTCAGGGGCCTGGAAATC
      * ** *.***** *..** ***.*..***** * . * .*****..*** **.***.*

```

Here is more information that will help you interpret the DNA sequences:

- **Coordinates:** There are two numbers that mark the ends of the DNA, 5335 and 5414 (5335 means it is the 5,355th nucleotide in BRCA1). These coordinates are important, and you will use them later in Part 3.
- **Sequences aren't perfectly aligned:** T is the 1st nucleotide in the human sequence at position 5335. Note that the other sequences are not all the same length (dashes are a space holder representing a gap or missing nucleotide).
- **Interpreting the * and _ symbols in the last row:** The last row of asterisks and dots shows which bases are conserved in the five sequences. For example, at human coordinate 5338, all five sequences have a T and a * appears in the last row. At position 5412, you find an A in three of the sequences and a G in the other two. These are nearly the same, and an _ appears in that position. If there is no conservation, then there is no mark like at position 5339 and 5442.

The next step: Figure out which amino acids are conserved in the protein. This is a bit trickier because we are in the middle of this gene so we can't just identify the usual DNA start codon (ATG) and begin there.

Here's a helpful trick: recall that multiple codons can code for the same amino acid and that the differences tend to be in the third nucleotide. For example, GCU, GCC, GCA, and GCG all code for the same amino acid, alanine. Knowing this, take a look at the sequences above. Do you see a pattern that might help you figure out which is the correct reading frame? In other words, is there a pattern of two conserved and one not conserved repeated over and over?

You will try to find the first amino acid and record your answer on your worksheet. The first codon shared by all five sequences is AAT, which corresponds to the amino acid asparagine (asn).

Now you are ready to translate the proteins. To make things easier, you will work in groups of 5 students and each of you will translate one of the DNA sequences. Then, check over each other's work and record all the group's answers on your worksheet. Finally, look at your translated sequences. Are there amino acids that are the same in all five sequences? This sort of thing happens for a reason. Amino acids that stay conserved across a wide range of animals tend to be important for the protein to work properly. Mark all the conserved amino acids as instructed on your worksheet.

Lab Activity Part 3: The DNA sequences you've seen so far are what scientists call the reference sequences. They do not show any of the differences that might happen between you and your friend or two different porcupines. They are really a list of the most common bases at each position. We all have millions of DNA differences sprinkled across our genomes. While most of these have little effect, a sizeable minority do. It is these that make you look and act a bit different than your friend.

These kinds of differences from the reference sequence can also help scientists figure out if a change has an effect. For example, if all redheads have a certain difference in their MC1R gene, then this might explain their redhead (it does by the way). The argument for this mutation being the cause of red hair is made even stronger if the mutation changes a conserved amino acid. Looking for changes in conserved amino acids doesn't just help explain red hair. As you'll see below, it can also help scientists figure out which DNA changes (or mutations) led to a cell becoming cancerous. Or at least put that cell on the path to becoming a tumor...

For the first part of this lab, you will determine which mutations might have a negative effect on the function of BRCA1. The following four mutations could be found in actual breast cancer tumors:

- a. 5346 A>C
- b. 5376 G>T
- c. 5389 G>T
- d. 5382 ins C

To figure out what this means, you need to know that DNA variants can be defined with the following shorthand where the number represented the nucleotide coordinate. A change in the nucleotide is written with the reference letter, an arrow (just a '>') and the changed letter. For example, 300G>C says that nucleotide 300 is changed from a 'G' to a 'C'. Two other types of single nucleotide changes can be expressed in this shorthand. An insertion ('ins') or deletion ('del') of one nucleotide. An example of an insertion is 300insA, adding an 'A' after nucleotide number 300. An example deletion is 300delG, deleting nucleotide 300 which happens to be a G.

For each of four variations indicate the nucleotide change, the change to the protein sequence if any, and the potential effect on the function of the protein. Refer to the Standard Genetic Code table identify the changes that result from the variation.