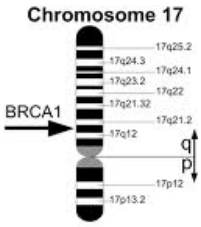
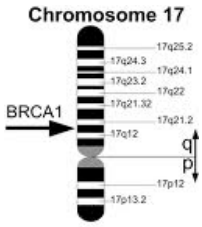


Hands-On Ten
The BRCA1 Gene
(Part One)



Objective: To review transcription, translation, reading frames, mutations, and reading files from GenBank, and to review some of the bioinformatics tools, such as BLAST for pairwise alignment and CLUSTAL OMEGA for multiple sequence alignment.

Assignment: The BRCA1 Gene

A) We would like to retrieve the record with accession number U14680 from NCBI.

- Go to the NCBI web site: www.ncbi.nlm.nih.gov.
- Select "Nucleotide" from the "All Databases" drop-down menu and type "U14680" in the search window.
- Click on the blue "Search"

1. Determine the organism and the name of the gene from which the sequence comes. What is its locus?
   Organism: _____
   Gene: _____
   Locus: _____.

2. What type of sequence (DNA, RNA, protein) is described in this NCBI record? _____.

3. How many exons does the gene have? _____. Are they all coding exons? _____.
   Explain: _____.

4. Why does the sequence under /translation start with following specific amino acids: MDLS?

_____
_____ .

5. Scroll up to the top of the page and click on "FASTA" under "GenBank: U14680.1".

- Copy and save the sequence in FASTA format in a file.

6. Change the display back to the default, GenBank format by clicking on "GenBank" at the top of the page.
   Based on the annotation, what is the accession number of the resulting **protein sequence**? _____.

7. Click on the accession number of the protein. How many amino acid residues are there in the protein? _____.

8. Scroll up to the top of the page and click on "FASTA" under "GenBank: AAA73985.1".
   - Copy and save the sequence in FASTA format in the same file you created in number 5.

9. We will perform a pairwise alignment between translations of the mRNA sequence (from part 5) and the protein sequence (from part 8).

   - Open a second browser window, go to the NCBI home page, and select the BLAST tool by clicking on "BLAST" under "Popular Resources" at the right-hand side of the page.
   - From the main page of BLAST, choose "Align two (or more) sequences using BLAST (bl2seq)" under "Specialized BLAST".
   - From the new page, select the translated BLAST (blastx) from the top of the page
   - Scroll down towards the bottom of the page and click on "Algorithm parameters"
   - Go to "Filter" and deselect "Low complexity regions"
   - Paste the sequence from part 5) in the first window and the amino acid sequence from part 8) in the second window. Alternatively, we can type "U14680" and "AAA73985.1" in the first and second windows, respectively. Note that the order of entering the sequences is important.
   - Choose "Show results in a new window" next to "BLAST".
   - Click on the blue "BLAST".

   The following three questions are about the very first alignment (and most significant one) displayed under "Alignments".

   a) How similar are the two sequences? _____.
   b) Did you expect the degree of similarity you found? _____.

Why?

_____.

10. Pairwise alignment methods can also be used to detect mutations. We are going to align the wild-type BRCA1 mRNA sequence (with accession number U14680) with a mutant version of the mRNA (with accession number U64805).

- Go back the BLAST web page
- Choose "Align two (or more) sequences using BLAST (bl2seq)" under "Specialized BLAST".
- Make sure you have "blastn" at the top of the page.
- To make the alignment, simply enter the accession numbers: U14680 in the first window and U64805 in the second window. You do not have to retrieve and paste the two sequences in the windows.
- Choose "Somewhat similar sequences (blastn)" under "Program Selection"
- Choose "Show results in a new window" next to "BLAST".
- Click on "BLAST".

Examine the top two alignments in the result web page, and answer the following question.

Given the results of your alignment, what type of mutation do you think has occurred to produce the mutant mRNA? (i.e. insertion, deletion, point mutation, etc). Be specific. Retrieve the mutant version of the BRCA1 mRNA and check your answer.
Hint: Click on "Dot Matrix View" and carefully examining the graph produced by BLAST.

B) We are going to use the multiple sequence alignment tool CLUSTALW available at EBI to find regions of similarity between the human BRCA1 wild-type and seven other BRCA1 proteins from other organisms.

- Retrieve the seven sequences: "sequences_BRCA1.txt" from Canvas. Note that the first sequence is the human BRCA1 wild type.
- Go to http://www.ebi.ac.uk/Tools/msa/clustalo
- Paste all seven sequences in the CLUSTAL OMEGA data window.
- Under STEP 2, click on "More options ..." and choose "input" under "ORDER".
- Click on "Submit" to align the seven sequences and wait for the results.
- Scroll down and click on "Show Colors".
- Carefully examine the alignment. Note that:

- "*" at the bottom of a column means that we have an alignment where all the amino acids are identical.
- ":" at the bottom of a column represents strongly conserved alignment.
- "." at the bottom of a column represents a weakly conserved alignment.

- Note that by clicking on "Page Summary" at the top of the page, and by choosing the JAVA alignment viewer and editor "JalView", we get to see a different way of showing the alignment of the eight sequences.

Answer the following questions based on the results of the alignment:

11. Which two sequences are the "least" similar?
12. Which sequence is the most similar to the human BRCA1?