Zayd Hammoudeh  
CS123A – Bioinformatics

**Hands On Assignment #09 – BRCA1 Gene**

**Question #1: For the NCBI sequence with accession number “U14680”, provide the following information:**

1. Organism: Homo sapiens
2. Gene: BRCA1 breast and ovarian cancer susceptibility gene.
3. Locus: 17q21 (This is under the source then “/map”

**Question #2: What type of sequence (e.g. DNA, RNA, protein, etc.) is described from the NCBI record?**

mRNA

**Question #3:**

1. 24 Exons
2. No. The 5’ UTR is from nucleotides 1 to 119. Hence, exon #1 is not a coding exon.

**Question #4:**

Starting from nucleotide 120 with codon start=1, the first twelve bases in the sequence are:

ATG GAU UUA UCU GCG which maps to MDLSA

**Question #5:**

**FASTA** is a format for displaying sequence information.

**Question #6:**

The protein ID is: AAA73985.1. (This is listed under the coding sequence as “/protein\_id”).

**Question #7:**

There are 1863 residues (i.e. amino acids) in the protein. (This is visible in the locus as 1863 aa. Note the coding sequence of a protein can include the stop codon which is not a residue in the protein.)

**Question #8:**

**FASTA** is a format for storing protein sequence information.

**Question #9:**

To compare the nucleotide sequence with the protein sequence, we are using blast2seq with **blastx**. We use blast2seq since we are comparing two sequences together. Blastx compares a nucleotide sequence to a protein sequence.

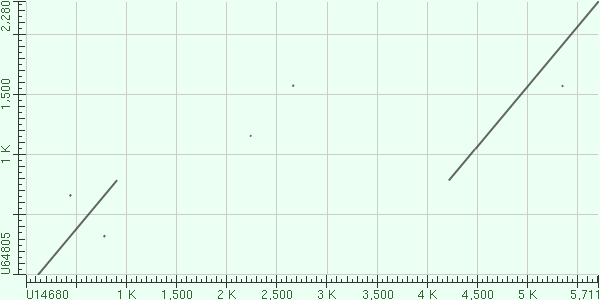
1. The genes have 100% identity (i.e. degree of similarity). What is more, the E-value is 0.
2. Yes this is expected since the protein is translated from the nucleotide sequence.

**Question #10:**

**Wild type:** A strain, gene, or characteristic that prevails among individuals in natural conditions. This is distinct from an atypical **mutant type**.

When using BLAST, you can align not only by providing the sequence information but also by providing the accession numbers.

It appears from this the Dot Matrix view there was a selection in a large section of the gene.



Note the wild type is on the X-axis.

**Reading the ClustalW Alignment:**

* “\*” – Amino acid in all sequences is identical.
* “:” – A strongly conserved sequence.
* “.” – A weakly conserved sequence

**Question #11:**

The Rattus and the Bos are the least closely related (identity of 54.14%).

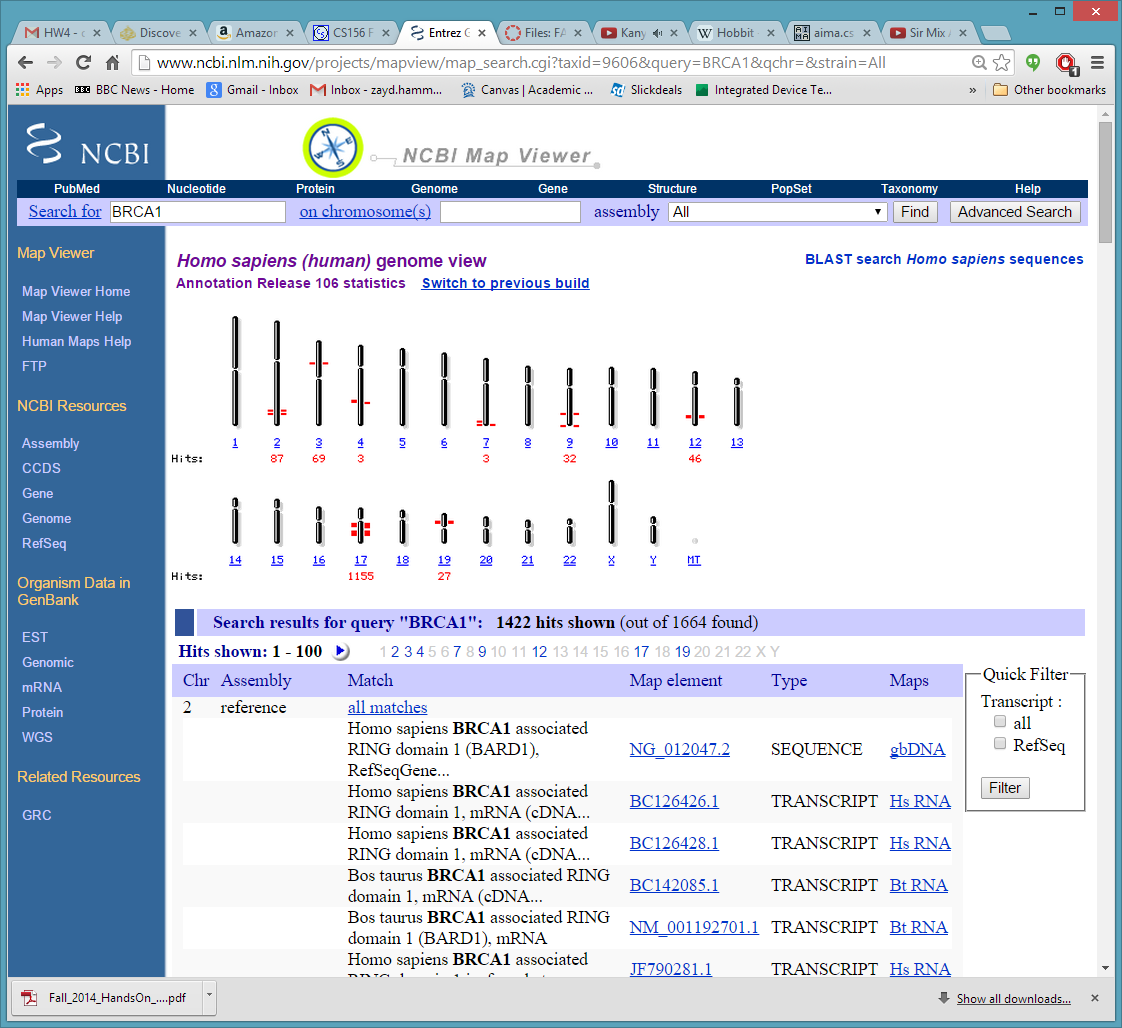
**Question #12:**

The most closely related to human is the BRCA1\_GORGO from the gorilla.

**Hands On Assignment #10 – Using the NCBI Genome Map Viewer**

**Question #13:**

1. There are 1422 hits shown.
2. Chromosome 17 has the most hits with 1155.



1. Chromosome 17.

**Question #14:**

34 hits.

**Question #15:**

|  |  |  |  |
| --- | --- | --- | --- |
| Chromosome | Match | Map Element | Type |
| 2 | hCG\_14827 | BARD1 | Gene |
| 3 | ubiquitin carboxyl-terminal hydrolase BAP1 | BAP1 | Gene |
| 9 | negative elongation factor protein B | NELFB | Gene |
| 12 | renal carcinoma antigen NY-REN-63 | BRAP | Gene |
| 13 | renal carcinoma antigen NY-REN-34 | PHF11 | Gene |
| 19 | new component of the BRCAA1 A complex | BABAM1 | Gene |

**Question #16:**

There are 2010 genes on the chromosome.

**Question #17:**

Locus is 17q21

**Question #18:**

It is BP 42,983K to 43,186K.

**Question #19:**

Other genes include: NBR1, BRCA1P1, RND2, VAT1, RPL27, IFI35, and RUNDC1.

**Question #20:**

No not all genes are on the same strand. Some are on the stand denoted by an (↑ i.e. “left”) while others are on the strand denoted by the (↓ “right”). BRCA1 is on the (↑) “left” strand.

**Question #21:**

There are about 83 million base pairs on the chromosome.

The record was updated November 7, 2014.

**Question #23:**

NM\_007294.3 🡪NP\_009225.1

NM\_007297.3 🡪 NP\_009228.2

NM\_007298.3 🡪 NP\_009229.2

All of these are isoforms that make one susceptible to breast cancer.

**Question #24:**

Nucleotides 43,032,116 to 43,137,660 are shown.

**Question #25:**

The fish bones shown that the 5’ end is on the left side of the viewer.

**Question #26:**

There are five mRNA sequences, one noncoding sequence, and 13 predicted sequences.

NEED MORE INFORMATION

**Question #27:**

Yes, all three sequences from question #23 appear in the genome map (e.g. NM\_007294.3, NM\_007297.3, NM\_007298.3)

NEED MORE INFORMATION

**Question #28:**

They are the four phases in the cell lifetime. G1 is growth. S is DNA synthesis; G2 is growth and preparation for cell division. M is mitosis.

**Question #29:**

NEED MORE INFORMATION

**Question #30:**

BRCA1 is a transcription factor.

**Question #31:**

Under the picture of the mouse, click on the text “Mouse Pathway”.

**Question #32:**

No it is very similar to that of humans.