

Assignment 8: Due 2pm 10/30

At NCBI website, find the gene page for human gene BRCA1 (BRCA1, DNA repair associated). Answer the following questions. You can use other sources like UCSC genome browser, and Ensembl. Most of the information is available at the gene page at NCBI.

1. Brief description of its function in your own words.
2. What type of tissues express this gene?
3. On what chromosome is this gene located?
4. The exact location of this gene on the genome according to GRCh38.p7 assembly.
5. Names of the 3 genes neighboring this gene on the chromosome.
6. How many papers have been published on this gene?
7. What genetic diseases is associated with mutations of this gene?
8. What proteins interact with the BRCA1 protein? Name 2 with references.
9. What molecular pathways is the BRCA1 protein part of? Name 2.
10. How many alternatively spliced transcripts (variants) does this gene have?
11. How many protein isoforms does this gene encode?
12. For the transcript called NM_007294.3, how many exons and introns?
13. Retrieve the following sequences for the transcript NM_007294.3. First report their length and then paste the FASTA formatted sequence.
 - a. Spanning genomic DNA sequence
 - b. mRNA sequence
 - c. Corresponding protein sequence
 - d. First exon (Hint: easier to find in Ensembl.org, see below)
 - e. 5' Untranslated region (UTR)
 - f. 3' untranslated region (UTR)
 - g. 1000bp promoter sequence upstream of transcription start site(TSS)

Hint: Finding genomic sequences in Ensembl.

1. From the NCBI gene page, find the corresponding transcript ID and click on the direct link to Ensembl.

mRNA and Protein(s)

1. [NM_007294.3](#) → [NP_009225.1](#) **breast cancer type 1 susceptibility protein isoform 1**

[See identical proteins and their annotated locations for NP_009225.1](#)

Status: REVIEWED

Description	Transcript Variant: This variant (1, also known as BRCA1a) represents the more frequently occurring transcript. It encodes the full-length BRCA1 protein (isoform 1), which is also known as p220.
Source sequence(s)	AL701927 , BC072418 , BU617173 , BU679389 , U14680
Consensus CDS	CCDS11453.1
UniProtKB/Swiss-Prot	P38398
Related	ENSP00000350283.3 , OTTHUMP0000021214 , ENST00000357654.8 , OTTHUMT00000348798

Click here

2. To get the sequences of each exons, click on the 23 exons:

e!Ensembl BLAST/BLAT | BioMart | VEP | Tools | Downloads | Help & Docs | Blog

Human (GRCh38.p12) ▼

Location: 17:43,044,295-43,125,483 Gene: BRCA1 Transcript: BRCA1-203

Transcript-based displays

- Summary
- Sequence
 - Exons
 - cDNA
 - Protein
- Protein Information
 - Protein summary
 - Domains & features
 - Variants
- Genetic Variation
 - Variant table
 - Variant image
 - Haplotypes
 - Population comparison
 - Comparison image
- External References
 - General identifiers

Transcript: BRCA1-203 ENST00000357654.8

Description BRCA1, DNA repair associated [Source:HGNC Synt
Synonyms BRCC1, FANCS, PPP1R53, RNF53
Location [Chromosome 17:43,044,295-43,125,483](#) reverse str:
About this transcript This transcript has **23 exons**, is annotated with [50 do](#)
Gene This transcript is a product of gene [ENSG000000120](#)

Summary ⓘ

3. We now can see the different types of sequence shown in different colors. Also, variants are highlighted. (see next page)

Exons ⓘ

[Download sequence](#)

Color coded by type

Exons/ Introns Translated sequence Flanking sequence Intron sequence UTR

Variants Coding sequence Frameshift Inframe deletion Inframe insertion Missense Splice acceptor

Splice donor Splice region Start lost Start retained Stop gained Synonymous

Markup loaded

Variants are highlighted

Exon IDs

Show All entries Show/hide columns

No.	Exon / Intron	Start	End	Start Phase	End Phase	Length	Sequence
	5' upstream sequence					cagagcccccagagacgcttggtctct
1	ENSE00001946433	43,125,483	43,125,271	-	-	213	GTACCTTGTTTCGTATCTGAGAGGCTGCTGCTTA CAACGGAAAGCGCGGGAATTACAGATATTAAAA CTGAGACTTCCTGACCGGGGACAGGCTGTGGGGTT TCAGAGGCCCTCACCTCTGCTCTGCGTAAAG
	Intron 1-2	43,125,270	43,124,116			1,155	gtagtagagtcccggaaggga.....t
2	ENSE00003559512	43,124,115	43,124,017	-	2	99	TTCATTGGGACAGAAAGAAATGGAATATCTGCTCT CATTAATGCTATCAGAAATCTTAGAGTGTCCCA

5' UTR

2nd Exon

variants