

**Identificare il gene a cui appartiene la sequenza (sonda) e la sua posizione sul cromosoma.**

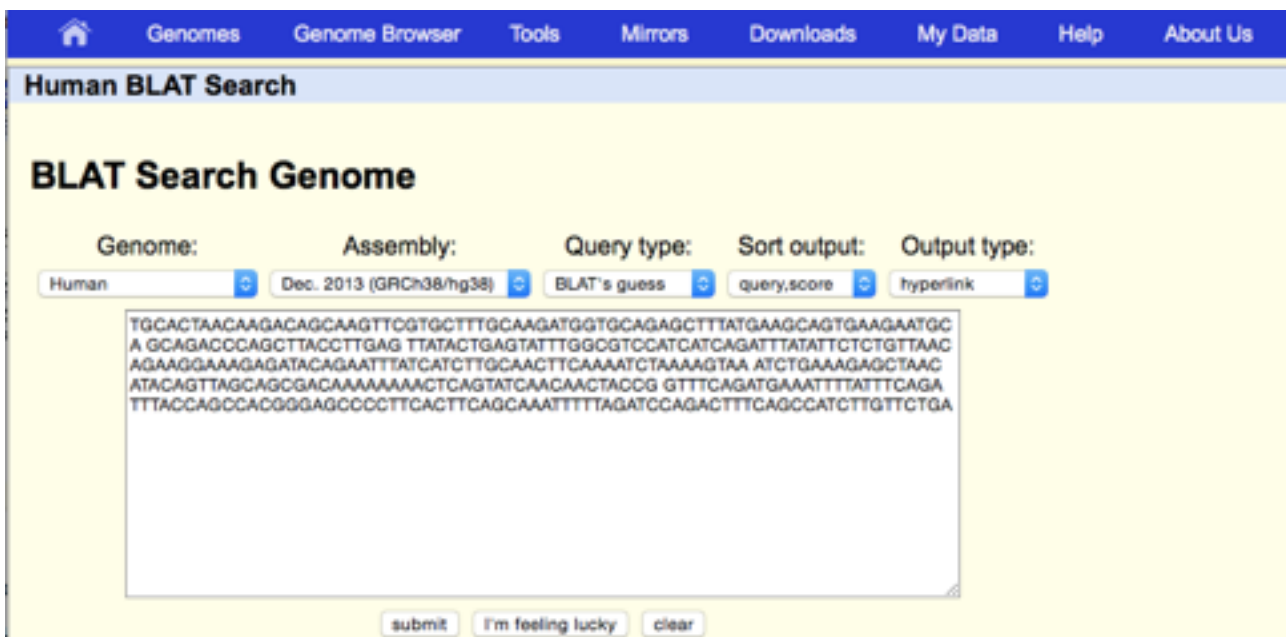
Per raggiungere l'obiettivo della prima parte dell'attività devi usare il software **BLAT** (BLAST-Like Alignment Tool), un algoritmo ottimizzato per confrontare sequenze di cDNA (prive di introni) con intere sequenze genomiche (che contengono introni).

Vai al sito: <http://genome.ucsc.edu/cgi-bin/hgBlat?db=mm2>

**Incolla nel box vuoto la sequenza nucleotidica**

```
TGCACTAACAAAGACAGCAAGTTCGTGCTTTGCAAGATGGTGCAGAGCTTTATGAAGCAGTGAAGAATGCA
GCAGACCCAGCTTACCTTGAG TTATACTGAGTATTTGGCGTCCATCATCAGATTTATATTCTCTGTTAAC
AGAAGGAAAAGAGATACAGAATTTATCATCTTGCAACTTCAAAATCTAAAAGTAA ATCTGAAAGAGCTAAC
ATACAGTTAGCAGCGACAAAAAACTCAGTATCAACAACCTACCG GTTTCAGATGAAATTTTATTCAGA
TTTACCAGCCACGGGAGCCCCCTTCACTTCAGCAAATTTTAGATCCAGACTTTCAGCCATCTTGTCTGA
```

per trovare la sua localizzazione nel genoma umano (controlla che sia selezionata la scelta “human” tra i genomi disponibili e clicca su **“submit”**).

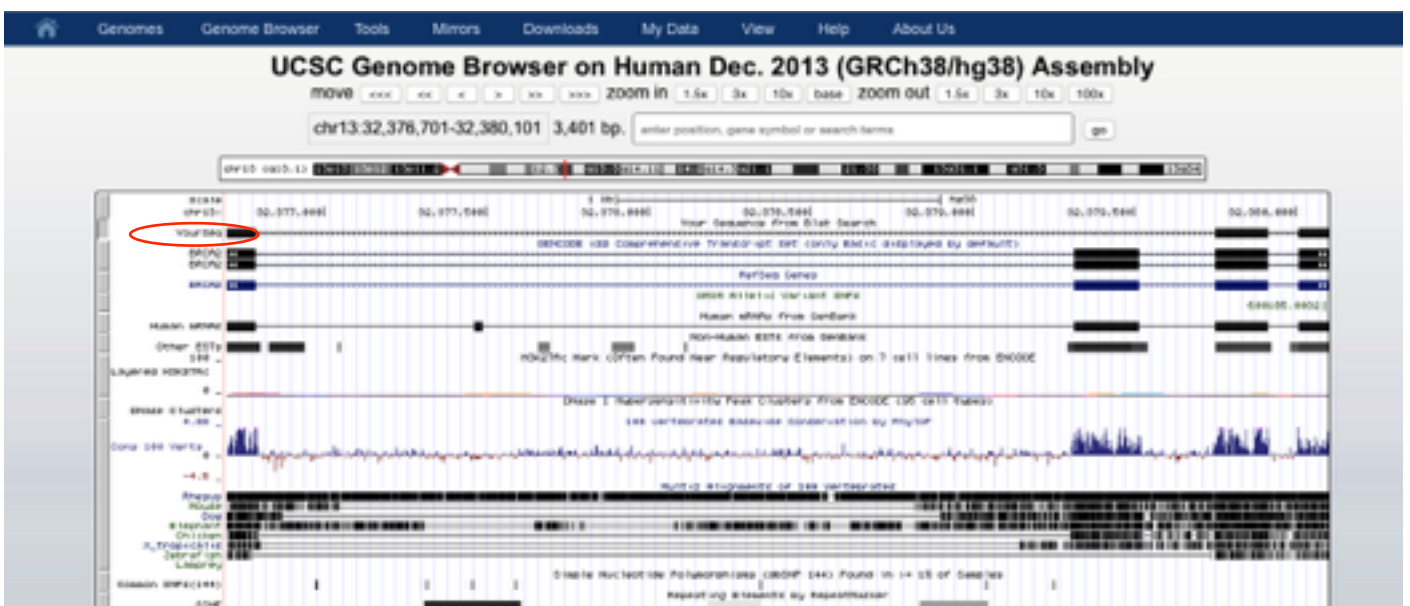


Bisogna attendere alcuni secondi per avere i risultati della ricerca del software BLAT. Scegli il primo risultato che mostri il migliore allineamento definito con un numero detto *score*: più alto è lo *score* migliore è l'allineamento. Scoprirai che la tua sequenza (definita dal software *query*) è composta da 350 nucleotidi e si trova nel cromosoma 13.

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
<a href="#">browser details</a>	YourSeq	348	1	350	350	100.0%	13	+	32376701	32380101	3401
<a href="#">browser details</a>	YourSeq	36	157	248	350	97.4%	10	+	67445836	67732799	286964
<a href="#">browser details</a>	YourSeq	27	41	68	350	100.0%	10	-	25998573	25998985	413
<a href="#">browser details</a>	YourSeq	22	183	220	350	73.0%	5	-	124384784	124384820	37
<a href="#">browser details</a>	YourSeq	22	252	273	350	100.0%	3	-	GL383526v1	172430	172451
<a href="#">browser details</a>	YourSeq	21	267	287	350	100.0%	6	+	136777031	136777051	21
<a href="#">browser details</a>	YourSeq	21	325	345	350	100.0%	11	+	59978829	59978849	21
<a href="#">browser details</a>	YourSeq	21	127	153	350	88.9%	1	+	169432716	169432742	27
<a href="#">browser details</a>	YourSeq	20	27	46	350	100.0%	2	+	118262726	118262745	20
<a href="#">browser details</a>	YourSeq	20	176	195	350	100.0%	15	+	85685834	85685853	20
<a href="#">browser details</a>	YourSeq	20	132	151	350	100.0%	12	+	77267786	77267805	20

Se desideri maggiori dettagli sulla struttura esoni-introni della regione genomica in esame, clicca su **“details”** (userai questa funzione anche in seguito); ora però clicca su **“browser”** per trovare la regione cromosomica che contiene la sequenza che ci interessa. Nella schermata che si apre osserva che l'ideogramma del cromosoma ha un tratto rosso nella posizione dove si trova la sequenza che stiamo studiando.

Nella finestra sottostante ritrovi la sequenza (*your seq*) al di sotto della quale c'è la sequenza



genomica che si allinea nel miglior modo possibile; per capire bene le informazioni contenute in questa finestra bisogna rinfrescare le conoscenze sulla struttura del gene. I geni rappresentano solo

l'1% del genoma umano e gli scienziati stanno ancora studiando per capire la composizione e la funzione del restante 99%.

Negli eucarioti i geni sono affiancati da lunghe sequenze non codificanti il cui significato risulta ancora abbastanza oscuro (queste sequenze sono anche dette DNA spazzatura); quando viene localizzato un gene si nota che è suddiviso in piccole parti dette esoni, che sono le parti codificanti proteine o molecole di RNA e che gli esoni sono separati da lunghi tratti non codificanti detti introni. Nella finestra di BLAT gli esoni sono rappresentati da blocchetti neri o rossi, mentre gli introni sono raffigurati come linee sottili.

**Prendi nota della posizione della sonda che stai studiando sul cromosoma 13 (q 13.1) e della lunghezza della regione (chr13:32,376,701-32,380,101)**

Il gene al quale appartiene la tua sonda è quello che mostra il migliore allineamento, indicato sotto a *your seq*: come ci attendevamo, è BRCA2.

**Identificare la mutazione presente nel DNA di Anne e nella sua famiglia**

Confronta la struttura della tua sequenza (*your seq*) con la struttura del gene normale sottostante. **Quanti esoni sono visibili in questa parte di gene? Quanti esoni ci sono nella tua sequenza? Ci sono differenze tra la struttura esoni-introni della tua sequenza e la struttura della sequenza del gene normale? Cosa possiamo dire sulla mutazione genica di Anne?**

In questa pagina ci sono altre informazioni; sulla sinistra ci sono dei blocchetti grigi e blu: l'interfaccia del browser ti permette di personalizzare ed arricchire la tua ricerca visualizzando informazioni diverse, aggiungendone o nascondendone altre. Infatti cliccando sui blocchetti verticali si accede ad altre pagine e si possono cambiare anche i parametri della ricerca.

Osserva attentamente la barra verticale relativa alla conservazione della sequenza (Vertebrate Alignment and Conservation). Nota che le regioni maggiormente conservate corrispondono agli esoni (i rettangoli colorati pieni nella sezione RefSeq genes) e che queste regioni si conservano anche in specie di vertebrati lontane evolutivamente dall'uomo, come in opossum, pollo o rana. Un discreto livello di conservazione si osserva anche in alcune regioni introniche (conservazione delle sequenze non-codificanti) ma ancora non ne è chiaro il ruolo anche se si può supporre che siano zone coinvolte nella regolazione dell'espressione genica. La penultima barra mostra gli SNPs (single nucleotide polymorphisms) e l'ultima, Repeat masker, mostra le sequenze ripetute.

Ora dobbiamo trovare quale è l'esone che manca nella mutazione della famiglia di Anne. Per ottenere questa informazione bisogna trovare la struttura esone-introne completa del gene normale BRCA2. Per fare questo hai bisogno della sequenza codificante completa (CDS) e, con questa, cominciare una nuova ricerca BLAT.

Vai alla pagina <http://www.ncbi.nlm.nih.gov/> per trovare il codice identificativo del gene BRCA2. Seleziona “Nucleotide” dalla tendina che apparirà cliccando su “All databases” e nel campo “search” p digita “BRCA2 homo sapiens”. Clicca poi su Search. Nell’elenco che apparirà, trova BRCA2 *Homo sapiens* breast cancer 2 (BRCA2), mRNA.

Prendi nota del numero di accesso identificativo NM\_000059 (NM significa mRNA naturale). Clicca su NM\_000059 e si aprirà una nuova pagina. difficile da interpretare ma alcune

LOCUS	NM_000059	11386 bp	mRNA	linear	PRI 15-DEC-2015
DEFINITION	Homo sapiens breast cancer 2 (BRCA2), mRNA.				
ACCESSION	NM_000059				
VERSION	NM_000059.3 GI:119395733				
KEYWORDS	RefSeq.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 11386)				
AUTHORS	Zhao W, Vaithiyalingam S, San Filippo J, Maranon DG, Jimenez-Sainz J, Fontenay GV, Ewon Y, Leung SG, Lu L, Jensen RB, Chazin MJ, Wiese C and Sung P.				
TITLE	Promotion of BRCA1-Dependent Homologous Recombination by DSS1 via RPA Targeting and DNA Mimicry				
JOURNAL	Mol. Cell 59 (2), 176-187 (2015)				
PUBMED	26145171				

informazioni sono chiare; prima di tutto ci sono le *Reference*, cliccando sulle quali si possono trovare le citazioni alle pubblicazioni in letteratura scientifica che riguardano la nostra sequenza. Per leggere un abstract dell'articolo che descrive il gene, clicca sul link **PubMed** presente in ogni *Reference*.

Scendendo nella pagina puoi trovare informazioni più dettagliate sulla **mutazione genica**.

**Alla fine della pagina, nella sezione Features, troverai la sequenza di cDNA (CDS).** Questa è la sequenza utile per la ricerca in BLAT. Non si può copiare la sequenza CDS così come si trova



NCBI Resources How To

Nucleotide Nucleotide Advanced

NCBI is phasing out sequence GI numbers in September 2016. Please use accession version! [Read more...](#)

GenBank -

### Homo sapiens breast cancer 2 (BRCA2), mRNA

NCBI Reference Sequence: NM\_000059.3

**FASTA** Graphics

Go to:

LOCUS NM\_000059 11386 bp mRNA linear PRI 15-DEC-2015

DEFINITION Homo sapiens breast cancer 2 (BRCA2), mRNA.

ACCESSION NM\_000059

VERSION NM\_000059.3 GI:119395733

KEYWORDS RefSeq.

SOURCE Homo sapiens (human)

ORGANISM [Homo sapiens](#)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;  
Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 11386)

AUTHORS Zhao W, Vaithiyalingam S, San Filippo J, Maranon DG, Jimenez-Sainz J, Fontenay GV, Evon Y, Leung SC, Lu L, Jensen RB, Charin NF, Wiesse C and Sung P.

TITLE Promotion of BRCA1-Dependent Homologous Recombination by DGS1 via RPA Targeting and SSA Mimicry

JOURNAL Mol. Cell 59 (2), 174-187 (2015)

PUBMED 26145171

nella pagina ma bisogna ottenere un file utile per la ricerca del gene correlato nelle banche dati. Per fare ciò bisogna risalire all'inizio della pagina e selezionare **FASTA**. Copia, incolla e salva su un file di testo con il nome **cdNAFASTA.doc** la sequenza di basi che apparirà sullo schermo. Tornando in banca dati alla pagina con le caratteristiche di NM\_000059, trova il numero identificativo della proteina prodotta dal gene (NP\_000050.2), la sua sequenza di amminoacidi, copiala e salvala in un nuovo file di testo dal nome **BRCA2 protein.doc**. Prendi nota del numero di amminoacidi che costituiscono la proteina normale.

Adesso che hai la sequenza codificante completa (cDNA) del gene BRCA2 puoi cercarne la sequenza genomica usando il software BLAT. Come ricorderai, questo software ti permette di

confrontare sequenze di cDNA con sequenze genomiche complete e di identificarne la struttura in esoni ed introni.

Ritorna alla pagina <http://genome.ucsc.edu/cgi-bin/hgBlat?db=mm2> ed incolla la sequenza di cDNA nella finestra di BLAT.

Human BLAT Results											
BLAT Search Results											
Go back to <a href="#">chr13:32376701-32380101</a> on the Genome Browser.											
ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
<a href="#">browser details</a>	YourSeq	11358	1	11386	11386	100.0%	13	+	32315480	32399672	84193
<a href="#">browser details</a>	YourSeq	95	10744	10861	11386	90.6%	22	-	26413153	26413274	122
<a href="#">browser details</a>	YourSeq	95	10734	10866	11386	84.1%	11	+	36179712	36179843	132

Clicca “*submit*” e quando comparirà la nuova pagina clicca su “*details*” alla sinistra del primo record in elenco (score 11.358, 11.386 nucleotides, 100% identity).

Nella pagina che si apre troverai la sequenza di cDNA, scorrendo nella pagina troverai la sequenza di DNA genomico nella quale sono evidenziati in blu scuro e con le lettere maiuscole gli esoni, in nero e con le lettere minuscole gli introni e in azzurro i siti di splicing. Clicca sui links, nella

**Alignment of YourSeq**

- [YourSeq](#)
- [Human chr13](#)
- [block1](#)
- [block2](#)
- [block3](#)
- [block4](#)
- [block5](#)
- [block6](#)
- [block7](#)
- [block8](#)
- [block9](#)
- [block10](#)
- [block11](#)
- [block12](#)
- [block13](#)
- [block14](#)
- [block15](#)
- [block16](#)
- [block17](#)
- [block18](#)
- [block19](#)
- [block20](#)
- [block21](#)
- [block22](#)
- [block23](#)
- [block24](#)
- [block25](#)
- [block26](#)
- [block27](#)
- [together](#)

**Alignment of YourSeq and chr13:32315480-32399672**

Click on links in the frame to the left to navigate through the alignment. Matching bases mark the boundaries of gaps in either sequence (often splice sites).

**cDNA YourSeq**

```

CTGCCCGAG CTCTGAAAC TAGCCCGCAG AGCCCGAGCC CCTGTGCCAG 50
TCCTGCGCCT CTGCTGCCCC TCGGGTGTCT TTTGCGGCGG TGGGTCCCGC 100
CCGGGAGAGG CGTGAGGGGA CAGATTTGTC ACCGGCCCGG TTTTGTCCAG 150
CTTACTCCGG CCAAAAAAGA ACTGCACCTC TGGAGCGGAC TTATTTACCA 200
AGCATTGGAG GAATATCCTA GGTAAAAATG CCTATTGGAT CCAAGAGAG 250
GCCAACAATTT TTTGAAATTT TTAGACACCG CTCACACAAA GCAATTTAG 300
GACCAATAAG TCTTAATGCG TTTGAAAGAC TTTCTTCAGA AGCTCCACCC 350
TATAATTCCTG AACCTGCAGA AGAATCTGAA CATAAAAACA ACAATTACGA 400
ACCAAACTTA TTTAAAACTC CACAAAGGAA ACCATCTTAT AATCAGCTGG 450
CTTCAACTCC AATAATATTC AAAGAGCAAG GGCTGACTCT GCCCGTGTAC 500
CAATCTCCCTG TAAAGAAATF AGATAAATTC AAATTAGACT TAACAAGGAA 550
TTTTCCCAAT AGTAGACATA AAAGTCTTGG CACAGTGAAA ACTAAAAATGG 600
ATCAAGCAGA TGATTTTCCG TGTCCACTTC TAAATTCCTG TCTTAGTGAA 650
AGTCCCTGTTG TTTTACAATG TACACATGTA ACACACAAAA GAGATAAGTC 700
AGTGGTATGT GGGAGTTTGT TTCATACACC AAAGTTTGTG AAGGTCCTC 750
AGACACAAA ACATATTTCT GAAAGTCTAG GAGCTGAGGT GGATCCCTGAT 800
ATGTCTTGGT CAAATTCCTF AGCTACACCA CCCACCTTFA GTTCTACTGT 850
GCTCATAATC AGAAATGAAG AAGCATCTGA AACTGTATTT CCTCATGATA 900
CTACTCTTAA TGTGAAAAGC TATTTTTCCT ATCATGATGA AAGTCTGAAG 950
AAAAATGATA GATTTATCCG TTCTGTGACA GACAGTGAAA ACACAAATCA 1000
AAGAGAAAGT GCAAGTCATG GATTTGGAAA AACATCAAGG AATTCATTTA 1050
AAGTAAATAG CTCGAAAGAC CACATTGGAA AGTCAATGCC AAATGTCTTA 1100
GAAGATGAAG TATATGAAC AGTTGTAGAT ACCTCTGAAG AAGATAGTTT 1150
TTCATTATGT TTTTCTAAAT GTAGAACAAA AAATCTACAA AAGTAAAGAA 1200
CTAGCAAGAC TAGAAAAAAA ATTTTCCATG AAGCAAAACC TGATGAATGT 1250
GAAAAATCTA AAAACCAAGT GAAAGAAAAA TACTCAATTTG TATCTGAAGT 1300
GGAACCAAAAT GATACTGATC CATTAGATTC AAATGTAACA AATCAGAAGC 1350
CCTTTGAGAG TGGAGTGCAC AAAATCTCCA AGGAAGTTGT ACCGCTTTTG 1400
GCCTGTGAAT GGTCTCAACT AACCCPTTCA GGTCTAAATG GAGCCAGAT 1450

```

colonna di sinistra, per navigare nella sequenza. Cliccando sui vari blocchi, ti appare ad inizio pagina l'esone corrispondente nella sequenza genomica; noterai che ci sono 27 esoni nel gene BRCA2. Il tuo compito ora è di identificare quale è l'esone che manca nel gene mutato di BRCA2. Nella sequenza genomica ricerca i residui nucleotidici che corrispondono alla posizione della tua sonda sul cromosoma 13 (ricorda: **chr13:32,376,701-32,380,101**), semplicemente cliccando uno di seguito all'altro i vari blocchi azzurri di sinistra e leggendo i numeri dei nucleotidi dell'esone che apparirà in alto nella pagina.

**Sei in grado di rispondere alla domanda: qual è l'esone che manca nel gene mutato BRCA2? Arriverai alla conclusione che Anne è portatrice di una mutazione che consiste nella mancanza dell'esone 22 nel gene BRCA2.** La perdita dell'esone è dovuta a una mutazione di un sito di splicing (sostituzione di G con A) nell'introne 21, che porta alla perdita completa dell'esone 22 (Wagner TM et al 1999, PMID: 9971877).

**Copia la sequenza dell'esone mancante e salvala in un file di testo dal nome "missing exon.doc"**

esone 22 mancante:

**GGTTATTTTCAGTGAAGAGCAGTTAAGAGCCTTGAATAATCACAGGCAAATGTTGAATGATA  
GAAACAAGCTCAGATCCAGTTGGAAATTAGGAAGGCCATGGAATCTGCTGAACAAAAGGAA  
CAAGGTTTATCAAGGGATGTCACAACCGTGTGGAAGTTGCGTATTGTAAGCTATTCAAAA  
AAGAAAAGATTCAG**