

## 04\_Ensembl

**Genom böngészők:** pl. UCSC(University of California Santa Cruz), NCBI (National Center for Biotechnology Information)

A \*-al jelölt részek, csak definíciók, kiegészítések

**Genomböngészés ensembl-el** [www.ensemblgenomes.org](http://www.ensemblgenomes.org)

Az Ensembl egy közös project az EBI (European Bioinformatics Institute) és a Wellcome Trust Sanger Institute részvételével 1999 –ban indult, itt történt az első humán draft genom bemutatása, és azóta is célként szerepel a humán gének annotációja gén modellek alapján. További céluk a gerinchúros és gerinces élőlények (ma ár több mint 100 faj) genomjának annotálása, olyan élőlényeké amelyek fontosaka mezőgazdaságnak, orvostudománynak, veszélyeztetettek, vagy evolúciós jeltősséggel bírnak illetve model állatok (Drosophila, C.elegans, S. cerevisiae)Nagy hangsúlyt fektetnek a humán, egér, és zebradánió fajokra Ezenfelül élesztő és féreg gén szettek is bekerültek ebbe böngészőbe összehasonlító analízis céljából az Ensembl compara csoportnak köszönhetően. A legtöbb annotációt kéthavonta frissítik, ez egyre frisseb Ensembl verzió megjelenést eredményez. További céljuk az automatikus ámde mégis pontos génannotációt. Elérhető genomok, Összefűzni az annotációkat az elérhető biológiai adatokkal. web és program alapú interface. Az Ensembl-ön elérhetőek gének, és egyéb annotációk, mint pl. regulátor régiók, konzervált bázispárok a fajok között és különböző szekvencia variációk. Az ensembl-nek van egy testvér browsere: at [www.ensemblgenomes.org](http://www.ensemblgenomes.org), ahová nem chordata élőlények gerinchúr nélküliek kerülnek. (baktériumok, növények, metazoa-k, egysejtűek).

\* A genom annotáció nem más, mint biológiai információ hozzárendelése egy DNS szekvenciához (általában a génekre értik). A biológiai információ lehet strukturális (pl. kromoszómán való pontos elhelyezkedésének megállapítása), ill. funkcionális(például részvétel a szabályozásban, kölcsönhatásokban). (<http://web.med.u-szeged.hu/mdbio/hun/anyagok/2013-2014/i.felev/smge/6/1.A-Genom-Szerkezete-word.pdf>)

### **A core (alap) adatok:**

- Genom szekvenciák
- Gén / transzkript / protein modellek
- Külső referenciák
- Feltérképezett cDNS-ek, proteinek, microarray probák, BAC klónok, kromoszóma sáv térképek(cytogenetic band) repeat-ek, markerek etc.

Komparatív adatok:

- orthologok and paralogok, fehérje családok, teljes genom illesztések, szinténikus régiók,

Variant adatok:

- szekvencia variációk, strukturális variációk, linkage disequilibrium

Regulációs adatok

- “best guess” (legjobbnek becsült)regulációs elemek sorozata

linkek más adatbázisokhoz(directly or via Distributed Annotated System DAS)

Az Ensembl gén model-ek alapjául az ENA/ GenBank /DDBJ (International Nucleotide Sequence Database Collaboration) INSDC , cDNS, EST,RNAseq az UniProtKB : Swiss-Prot(manuálisan javított),

TrEMBL (fehérje adatbázisok) és NCBI RefSeq adatbázisokban felelhető protein és mRNS szekvenciák, szolgálták melyek transzkriptjeit (és még EST RNA-seq, cDNS, protein etc.,) illesztették aztán klaszterezték úgy, hogy az átfedő kódoló szekvenciák segítségével pontosan meg lehessen adni az Ensembl gének szekvenciáit és helyzetét. Automatikus annotáció: Genome-wide (Teljes genom) meghatározás Ensembl génalapú pipeline (parancssorozatok) felhasználásával történt. *Ab initio* gén predikció zajlott le. Ezenfelül manuálisan annotált géneket vettek át a VEGA/Havana csoporttól, majd az automatikusan és a manuálisan annotált géneket összemergelték, összevették egymással és így jött létre a Genecode geneset. Adatokat importáltak még a FlyBase, WormBase, és yeast (SGD) adatbázisokból is.

\*Az intrinsic (ab initio, de novo) gén predikciós megközelítések a géneket a fehérjekódoló génekre jellemző tulajdonságok felismerése alapján azonosítják a genom szekvenciában. Például az annotációs pipeline megkeresi a promoter régióra, transzkripció start helyre, exon/intron határra, poliadenilációs helyre stb. utaló speciális szekvencia jeleket vagy felismeri a fehérjekódoló gének jellegzetes statisztikus a nem-kódoló részekről különböző tulajdonságait. (<http://math.bme.hu/~gnagy/mmsz/eloadasok/PatthyLaszlo2012.pdf>)

\* [Vertebrate Genome Annotation](#) (VEGA), - [Wellcome Trust Sanger Institute](#), A Havana csoport a consensus-coding sequence (CCDS) együttműködés része, illetve a teljes emberi genom kézi annotációjáért felelős ENCODE project kiterjesztése, ami így elérhető referenciaként genomként illetve komparatív analízisekhez.

### **Gén annotációs folyamat:**

- Targeted build: (cél alapú):

Fajspecifikus proteineket és cDNS-eket illesztnek a genomra, hogy iniciális transzkript modelleket hozzanak létre.

- Similarity build (Azonosság alapú):

Nagyon közelrokon fehérjéit és cDNS-t illesztik fel, hogy annotálhassák a géneket. Az ismeretlen transzkriptumokat nem térképezik a fajspecifikus proteinekhez. Egybevetik a redundáns transzkriptumokat és klaszterekbe rendezve az alternatív transzkriptumokat hozzáteszik a megfelelő génekhez.

A CCDS-ek (core set of protein coding regions) –magas minőségű meghatározás, embernél és egégnél

\*Genome assembly: read-> contig-> scaffold-> chromosome

\* A referencia genom a haploid genom (embernél a 24-féle kromoszóma + a mitokondrium DNS) pszeudoszekvenciáit tartalmazó konszenzus szekvenciák összessége, ami önmagában ilyen formában nem létezik és nem is létezett. A célja az, hogy a többi szekvenciát ehhez viszonyítsuk. A referencia genomszekvenciákat folyamatosan újraépítik, ami azt jelentheti, hogy a szekvencia koordináták elcsúszhatnak, ezért vigyázni kell, hogy ha nem szekvencia, hanem koordináta alapon keresünk például két szekvencia között átfedést, akkor ugyanazt az összerakott genomot használjuk (pl. hg18 vs hg18 vagy mm8 vs mm8). A referencia genomszekvenciákat a nagy genomikai portálokról lehet

(érdeemes) letölteni. Javaslom az ensembl, az NCBI vagy az UCSC webhelyét.  
([http://genomika.blog.hu/2010/04/23/mi\\_az\\_a\\_referencia\\_genom](http://genomika.blog.hu/2010/04/23/mi_az_a_referencia_genom))

Elérhetőek variációs adatok, Összehasonlítókegymással genomok, és a regulatórikus régiók tulajdonságait is felderíthetjük. **Genom szekvencia assembly-k is hozzáférhetőek pl** -Mouse NCBI m37 (mm9)-Zebrafish Zv9 –Human GRCh37 (hg19). folyamatosan frissül, de a korábbi verziók is megtalálhatóak a preensembl-ben az alternatív haplotípusok is megtekinthetőek.

Minden adat ingyen hozzáférhető online [www.ensembl.org](http://www.ensembl.org). A PERL programnyelvet használók pedig közvetlenül hozzátudnak férni az ENSEMBL adatbázisokhoz az Application Programming Interfaces (**Perl APIs**) (felhasználói felület) keresztül. A gén szekvenciák letölthetőek az ENSEMBL browserről közvetlenül vagy a BioMart web felhasználói felület-en keresztül, ami lehetővé teszi, hogy programozói ismeretek nélkül, egyszerű felhasználóként információkat szerezzünk az Ensembl adatbázisairól.

### **Mire jó az Ensembl?**

- Megnézheted a gének különböző annotációi a kromoszómán
  - Megtekintheted az adott gén alternatív transzkriptumait (azaz splice variánsait).
  - Megkeresheted a több mint 100 faj (120-25) bármely génjének homológjait, amelyek alapján filogenetikus fákot lehet készíteni
  - Teljes genom alignmenteket és konzervált régiókat hasonlíthatsz össze a fajok között.
  - Láthatod az Ensembl génekhez illeszkedő mikroarray-szekvenciákat
  - Megnézheted meg EST-eket, klónokat, mRNS-eket és fehérjéket bármilyen kromoszomális régióban.
- \*klón genetikailag megegyezik az anyaszervezettel, EST: Expressed sequence tag: A CDNS darabjai.
- Megtálhatod a nukleotid polimorfizmusokat (SNP) a géneken vagy kromoszomális régiókon.
  - Láthatod az SNP-eket a fajok (patkány, egér), a populációk (emberi) vagy fajták (kutyák) között.
  - Megnézheted az Ensembl génekhez illeszkedő mRNS-ek és fehérjék pozícióját és szekvenciáját.
  - Feltöltheted a saját adatait.
  - Használhatod a BLAST-ot vagy a BLAT-ot bármelyik Ensembl genommal szemben.
  - Exportálhatod a szekvenciákat, vagy létrehozhatod egy táblázatokat a génadatokról a BioMart segítségével.
  - Láthatod az a saját génvariánsaid milyen hatással vannak a génekre és a transzkriptumokra a Variant Effect Predictor segítségével.

A felső kék csíkban linkeket mutat számos különböző ensembl alkalmazáshoz. felsorolás mik vannak

A kereső rész search bar lehetőséget ad hogy az ensembl összes fájában keressünk genom koordináták , gén nevek, próba szettek, variant ID-k, fenotípusok, és regulatórikus jellemzők ID –je (azonosítója )alapján.

Fent találhatóak az újdonságok, 2-3 hónapos frissülés miatt, a változások az annotációkban.

Összes genom elérhető:

- [View full list of all Ensembl species](#)

és a genomok régi verziója is a preensembl-n, a keresőben a gén nevek a HGNC azonosító alapján a HUGO gene nomenclature committee által jóváhagyott formában kereshetőek pl BRCA2

Nézzük meg a humán genom honlapját klikk a humán –ra GRCh38 –as assembly-re

Info a genom assembly-ről, a komparatív genomról, a gén annotációról, és a varációkról.

## BRCA2 keresőbe GO

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

Search all species...

New Search Jobs

Current selection: Only searching Human **BRCA2** 19474 results match BRCA2 when restricted to species: Human

Restrict category to:

- Gene: 9
- Transcript: 47
- Phenotype: 6
- Somatic Mutation: 1837
- GeneTree: 1
- ProbeFeature: 75
- Clones & Regions: 1
- Protein Domain: 6
- Protein Family: 2
- Variant: 17490

Per page: 40

Did you mean...  
 BRCA2 (Human Gene)  
 ENSG00000139618 13.32315474-32400266.1  
 BRCA2, DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]  
 LRG\_293 (LRG display in Ensembl gene record; description: Locus Reference Genomic record for BRCA2.) is an external reference matched to Gene ENSG00000139618  
 Variant table • Phenotypes • Location • External Refs • Regulation • Orthologues • Gene tree  
 BRCA2-203 (Human Transcript)  
 ENST00000528762 13.32370971-32379495.1  
 BRCA2, DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]  
 Location • External Refs • cDNA seq • Exons • Variant table • Protein seq • Population • Protein summary  
 BRCA2-204 (Human Transcript)  
 ENST00000530893 13.32315505-32332291.1  
 BRCA2, DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]  
 Location • External Refs • cDNA seq • Exons • Variant table • Population  
 BRCA2-207 (Human Transcript)  
 ENST00000614259 13.32316461-32362659.1  
 BRCA2, DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]

## Kattintsatok az első linkre

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

Search all species...

Human (GRCh38.p12)

Location: 13:32,315,474-32,400,266 Gene: BRCA2 Jobs

Gene-based displays

- Summary
- Splice variants
- Transcript comparison
- Gene alleles
- Sequence
- Secondary Structure
- Comparative Genomics
- Genomic alignments
- Gene tree
- Gene gain/loss tree
- Orthologues
- Paralogues
- Ensembl protein families
- Ontologies
- GO: Biological process
- GO: Molecular function
- GO: Cellular component
- Phenotypes
- Genetic Variation
- Variant table
- Variant image
- Structural variants
- Gene expression
- Pathway
- Regulation

Gene: BRCA2 ENSG00000139618

Description: BRCA2, DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]

Synonyms: BRCC2, FACD, FAD, FAD1, FANCD, FANCD1, XRCC11

Location: Chromosome 13: 32,315,474-32,400,266 forward strand. GRCh38:CM000675.2

About this gene: This gene has 7 transcripts (splice variants), 89 orthologues, is a member of 1 Ensembl protein family and is associated with 100 phenotypes

Transcripts: Hide transcript table

Name	Transcript ID	bp	Protein	Biotype	CCDS	UniProt	RefSeq	Flags
BRCA2-201	ENST00000380152.7	11986	3418aa	Protein coding	CCDS9344	P51587	-	TSL:5 GENCODE basic APPRIS P1
BRCA2-206	ENST00000544455.5	10984	3418aa	Protein coding	CCDS9344	P51587	NM_000059 NP_000050	TSL:1 GENCODE basic APPRIS P1
BRCA2-202	ENST00000470094.1	842	186aa	Nonsense mediated decay	-	H0YE37	-	CDS 5' incomplete TSL:5
BRCA2-203	ENST00000528762.1	495	64aa	Nonsense mediated decay	-	H0YD86	-	CDS 5' incomplete TSL:4
BRCA2-207	ENST00000614259.1	7950	No protein	Processed transcript	-	-	-	TSL:2
BRCA2-204	ENST00000530893.6	2011	No protein	Processed transcript	-	-	-	TSL:1

Gén azonosító Gene:BRCA2 ENSG00000139618 ez minden ensembl release, újabb genom assembly-ben meg fog egyezni!

Gén nevek az Ensembl-ben:

ENS (ENSEMBL)MUS (fajra jellemző, humánnál nincsen!)G/T/P/E /R(gén,transzkript,protein,exon, regulációs sajátosság)11 számjegy

Fenn látható 3 fül amelyekből mi most a gene-ben vagyunk, baloldalon e génre vonatkozó speciális információk találhatóak. pl. gén szekvencia, szekvencia alignmentek, filogenetikei törzsfák, homológiák, gén ontológia, variációs és gén expressziós adatok, itt hét annotált transzkript látható. Show transcript table

Location: 13:32,315,474-32,400,266 Gene: BRCA2 Jobs

**Gene: BRCA2** ENSG00000139618

Description: BRCA2, DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]

Synonyms: BRCC2, FACD, FAD, FAD1, FANCD, FANCD1, XRCC11

Location: Chromosome 13: 32,315,474-32,400,266 forward strand. GRCh38:CM000675.2

About this gene: This gene has 7 transcripts (splice variants), 89 orthologues, is a member of 1 Ensembl protein family and is associated with 100 phenotypes.

Transcripts: [Hide transcript table](#)

Name	Transcript ID	bp	Protein	Biotype	CCDS	UniProt	RefSeq	Flags
BRCA2-201	ENST00000380152.7	11986	3418aa	Protein coding	CCDS9344	P51587	-	TSL:5 GENCODE basic APPRIS P1
BRCA2-206	ENST00000544455.5	10984	3418aa	Protein coding	CCDS9344	P51587	NM_000059 NP_000050	TSL:1 GENCODE basic APPRIS P1
BRCA2-202	ENST00000470094.1	842	186aa	Nonsense mediated decay	-	H0YE37	-	CDS 5' incomplete TSL:5
BRCA2-203	ENST00000528762.1	495	64aa	Nonsense mediated decay	-	H0YD86	-	CDS 5' incomplete TSL:4
BRCA2-207	ENST00000614259.1	7950	No protein	Processed transcript	-	-	-	TSL:2
BRCA2-204	ENST00000530893.6	2011	No protein	Processed transcript	-	-	-	TSL:1
BRCA2-205	ENST00000533776.1	523	No protein	Retained intron	-	-	-	TSL:3

Summary

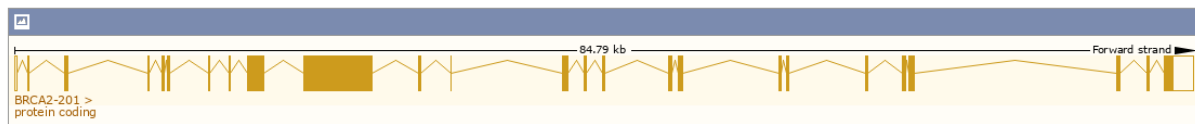
Az első két találat a CCDS consensus-coding sequence szett tagja vagyis ezek olyan protein kódoló transzkriptumok melyek más genom browserekkel (NCBI, UCSC, HAVANA) való együttműködés révén határoztak meg. Established as collaborative effort. A sárga a HAVANA génekkel összevetett verziót mutatja. A kék vonal jelöli a kontigot, ami egy genome szekvencia. A felette lévő trk a vezető forward szálon vannak, az alatta lévőek a maradó szálon lévő reverz transzkriptek. A boxok téglalapok exonokat, az ezeket összekötő vonalak intronokat jelölnek. Ha a boxok színezettek akkor kódoló szekveniát jelölnek ha nem akkor untranszlated régiókat. A transzkriptumok színe a biológiai típusokra utal a sárga az ensembl automatikus annotációs pipeline-al annotált de a havana project kézzel annotált génjeivel összevetett, összefésült transzkriptumot jelöli. A piros csak automatikusan annotált ensembl-ös. A kékek nem kódolóak. Kétféle módon nézhetjük meg a transzkriptet: vagy a Transcript ID alapján vagy

Go to Region in Detail for more tracks and navigation options (e.g. zooming)

HGNC Symbol: BRCA2	
Gene	BRCA2, DNA repair associated <a href="#">ENSG00000139618</a>
Location	<a href="#">Chromosome 13: 32,315,474-32,400,266</a>  
Exon	11 of 27
Transcript	BRCA2-201 <a href="#">ENST00000380152.7</a> <a href="#">Exons</a> <a href="#">cDNA Sequence</a>
Protein	<a href="#">ENSP00000369497</a> <a href="#">Protein Variations</a>
Gene type	Protein coding
Transcript type	Protein coding
Strand	Forward
Base pairs	11,986
Amino acids	3,418
Source	Ensembl/Havana merge

vagy úgy hogy rákattintunk a képre és megjelennek róla az információk. Itt az azonosítót választjuk. kinyílik egy újabb fül és láthatjuk a trk jellemzőit

### Summary



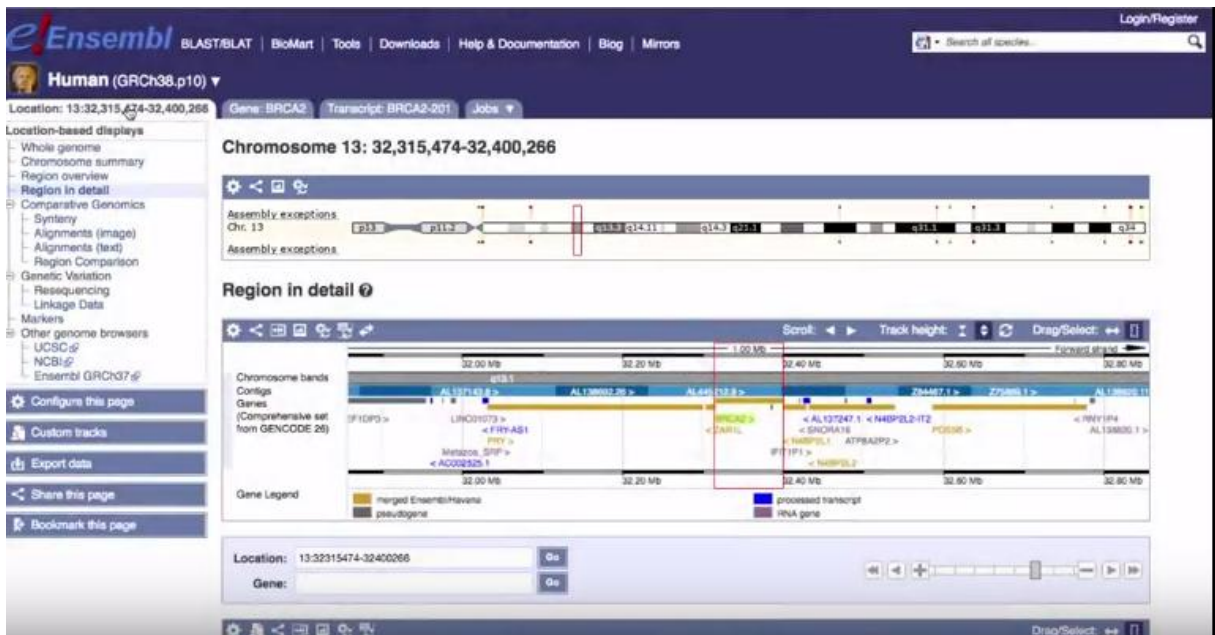
Statistics	Exons: 27, Coding exons: 26, Transcript length: 11,986 bps, Translation length: 3,418 residues
CCDS	This transcript is a member of the Human CCDS set: <a href="#">CCDS9344</a>
Uniprot	This transcript corresponds to the following Uniprot identifiers: <a href="#">P51587</a>
Transcript Support Level (TSL)	<a href="#">TSL:5</a>
Version	ENST00000380152.7
Type	Protein coding
Annotation Method	Transcript where the Ensembl genebuild transcript and the <a href="#">Vega</a> manual annotation have the same sequence, for every base pair. See <a href="#">article</a> .
GENCODE basic gene	This transcript is a member of the <a href="#">Genecode basic</a> gene set.

Ensembl release 92 - April 2018 © [EMBL-EBI](#)

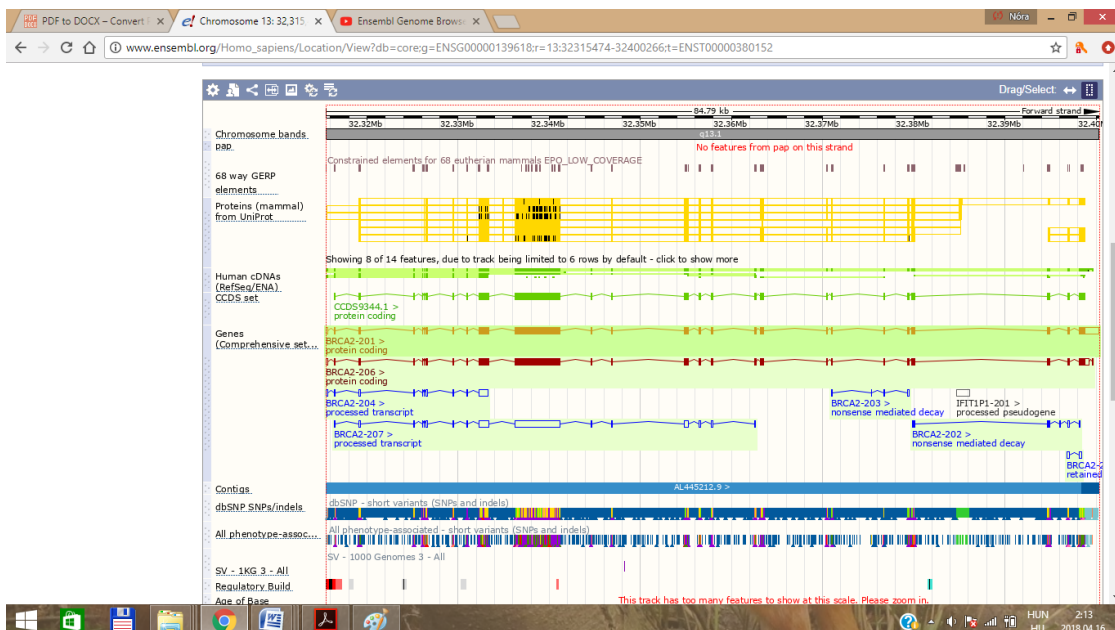
[Permanent link](#) - [View in archive site](#)

ha külső general identifiers re kattintunk akkor láthatjuk e transzkriptnek a találatait más külső adatbázisokban pl UCSC, de megnézhetjük az oligipróbáit, vagy a protein summeryben a proteinjeire jellemző tulajdonságokat,

Most nézzük meg e gén melletti szekvenciákat úgy hogy rámegyünk a Location földre.



Felül a kromoszóma látható a piros négyzetek a gén körüli régiókat mutatják, amit kinagyítva látunk a következő képen, ahol 1 Mb nagyságban látszik a gén környezete, a szomszédos gének, a kontigok világosak és sötét kéket, hogy el lehessen különíteni őket. a 3. nagy kép részlet zoom-olható, kifejezetten a génre fókuszál, itt olyan plusz információk is megjelennek mint pl e régióra jellemző fenotípus ezek ael is tüntethetőek a Configure this page opcióval, pl dbSNP variant normal, uniprot keresőbe, nézzünk információt és írjuk be az emlőst







### Információk a variant-ról:

\*Frameshift kereteltolódás: blokk- értelmezése deléció, hol van

Nem okoznak változást a kódolt polipeptid aminosavsorrendjében. Ezek az úgynevezett szinoním (synonymous) vagy más néven csendes (silent) mutációk.

Ha egy pontmutáció eredményeképpen a kódolt aminosav minősége megváltozik, akkor misszensz mutációról beszélünk. aminosav csere

Előfordul, hogy egy aminosavat kódoló kodon a transzlációt termináló, úgynevezett STOP kodonra változik, ezáltal a fehérje transzlációja idő előtt befejeződik. Az ilyen mutációkat nevezzük nonszensz (nonsense , értelmetlen) mutációknak.

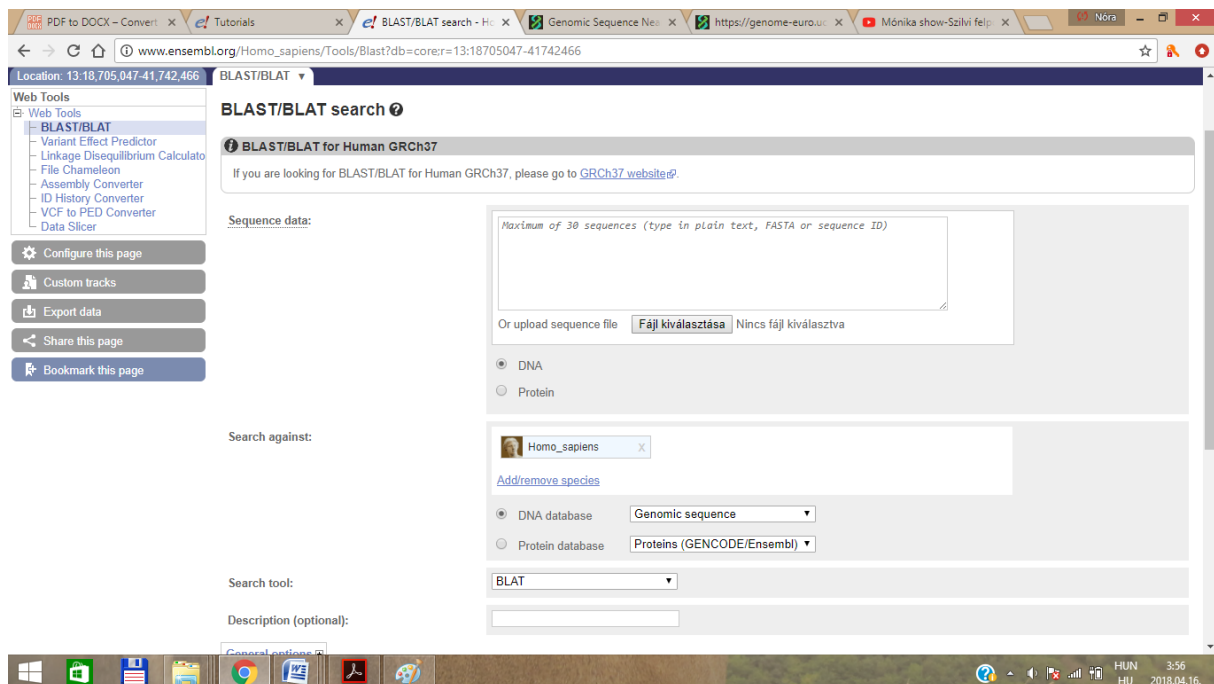
### Regulációs feature információk:

Regulatory Feature		
Stable ID	<a href="#">ENSR00000060894</a>	
Type	Promoter	
Core bp	<a href="#">13:32315000-32316601</a>	
Bounds bp	<a href="#">13:32314402-32317199</a>	
Motif Information		
Name	PWM ID	Score
Gabp	<a href="#">PB0020.1</a>	8.314
Gabp	<a href="#">MA0062.2</a>	9.552
Gabp	<a href="#">MA0062.1</a>	9.503
USF1	<a href="#">MA0093.2</a>	9.831
USF1	<a href="#">MA0093.2</a>	9.235
USF1	<a href="#">MA0281.1</a>	8.184
E2F4	<a href="#">MA0541.1</a>	5.94
E2F4	<a href="#">MA0541.1</a>	8.994
E2F6	<a href="#">MA0471.1</a>	9.773
PU1	<a href="#">PB0058.1</a>	9.023
Pax5	<a href="#">MA0014.2</a>	8.914

A region detail segít abban hogy mi is van az oldalon.



## Blast:/blat:



Például UCSC kivesszük BRCA2-t protein szekvenciát! (de ugyanígy blasztolhatsz saját szekvenciát vagy például az NCBI-ből, ENA-ból sstb.-ből szekvenciákat.)

Állítsuk át proteinre, állítsuk be a chimpanze-t referencia szekvenciaként, mindent kibontunk de alapbeállításon hagyunk tblastn –el futtatunk. megtudjuk nézni a szekvenciát lokációt, de igazán informatív az ha az alignmenteket nézzük meg. A result table-ben a query vagyis a kereső, blasztolandó szekvencia a humán fasta. A reference chimpanze. Töltsük le download result file. HSP distribution on query sequence.

\*A BLAST futás eredményeként olyan találatokat kapunk, amelyek az adatbázisban tárolt szekvenciák közül szignifikáns hasonlóságot mutatnak a célszekvenciával. A program sorba állítja ezeket a szekvenciapárokat, kezdve a legnagyobb hasonlóságot mutatóval. A szignifikanciát egy E-vel jelölt, a véletlen hasonlóság mértékéhez viszonyított várható érték (*expectation*) jelzi, valamint egy „score” érték, ami az azonos, hasonló és „rés” (gap) pozíciókat számolja egy nukleotid vagy aminosav hasonlósági mátrix alapján. Ezek közül leggyakrabban az ún. BLOSUM (BLOcks SUBstitution Matrix) mátrixokat használják. Ha  $E < 0,01$ , akkor a két szekvencia minden bizonnyal homológ (azaz evolúciós rokonságot mutat, közös ősből származik). **BLAST (Basic Local Alignment Search Tool)** : HSP-k (high scoring pair) keres kibővít.

(<http://elte.prompt.hu/sites/default/files/tananyagok/BevBiokemGyak/ch11s03.html>)

<i>Program</i>	<i>Kereső szekvencia</i>	<i>Célszekvencia</i>
BLASTN	Nukleotid	Nukleotid
BLASTP	Protein	Protein
BLASTX	Nukleotid, 6 leolvasási keretben	Protein
TBLASTN	Protein	Nukleotid, 6 leolvasási keretben
TBLASTX	Nukleotid, 6 leolvasási keretben	Nukleotid, 6 leolvasási keretben

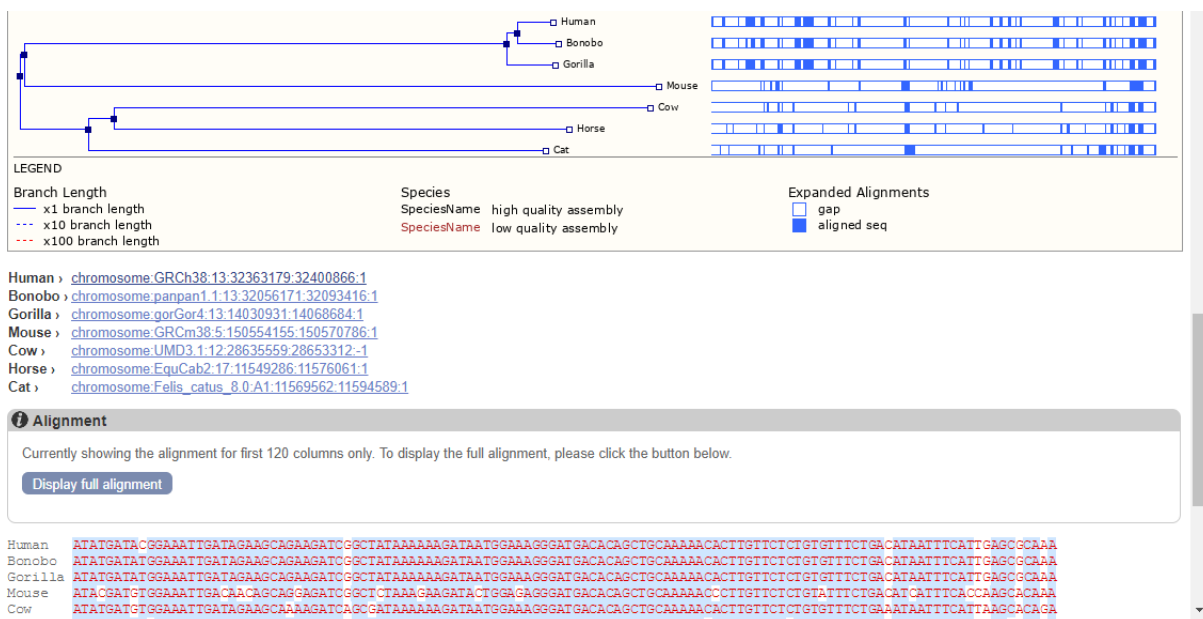
Adatok exportálása:

BRCA2 transcript, export data CSV formátumban, és FASTA formátumban, bekattintani amit lehet

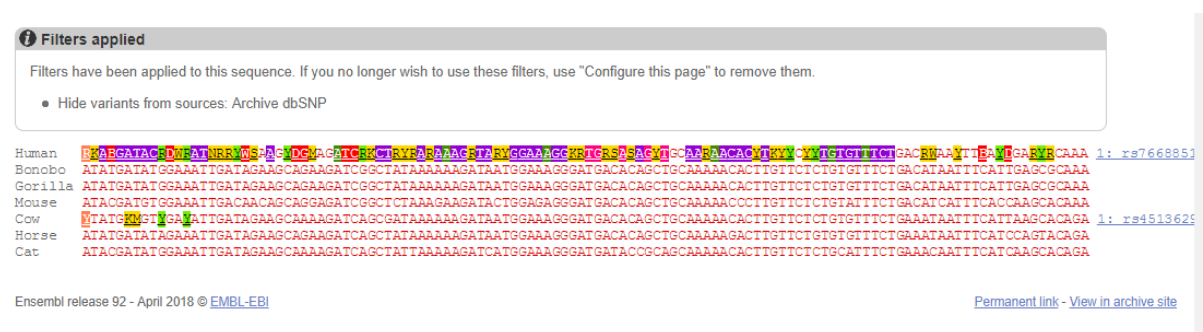
Az BRCA2 gén összehasonlítása emberben és más fajokban.

A gén

-Genomic alignment- select on alignment kiválasztjuk az állatokat! block-ot 1 configure, az állatok passzoló régiói ktiírva chr start end, orientáció, alignments text ,szekvenca :piros exont jelent, vonal nincs semmi, konzervált régiók aláhúzása: configurate dispage display opció, Show conservation regions késsel kiemelve



Lássuk a variant-okat Configure this page show variációk yes and show links



**Variation: rs451362933**

Class	SNP
Location	12:28653312
Source	dbSNP
Alleles	T/C
Consequences	splice region variant synonymous variant

[Explore this variant](#)  
[Gene/Transcript Locations](#)

Nézzük meg a cow Y-t –SNP –tovább az rs-es nevére

Variant: rs451362933 Jobs

**rs451362933** SNP

Most severe consequence | splice region variant | [See all predicted consequences](#)

Alleles | T/C

Location | [Chromosome 12:28653312](#) (forward strand) | VCF: 12 28653312 rs451362933 T C

HGVS names | This variant has 3 HGVS names - [Show](#)

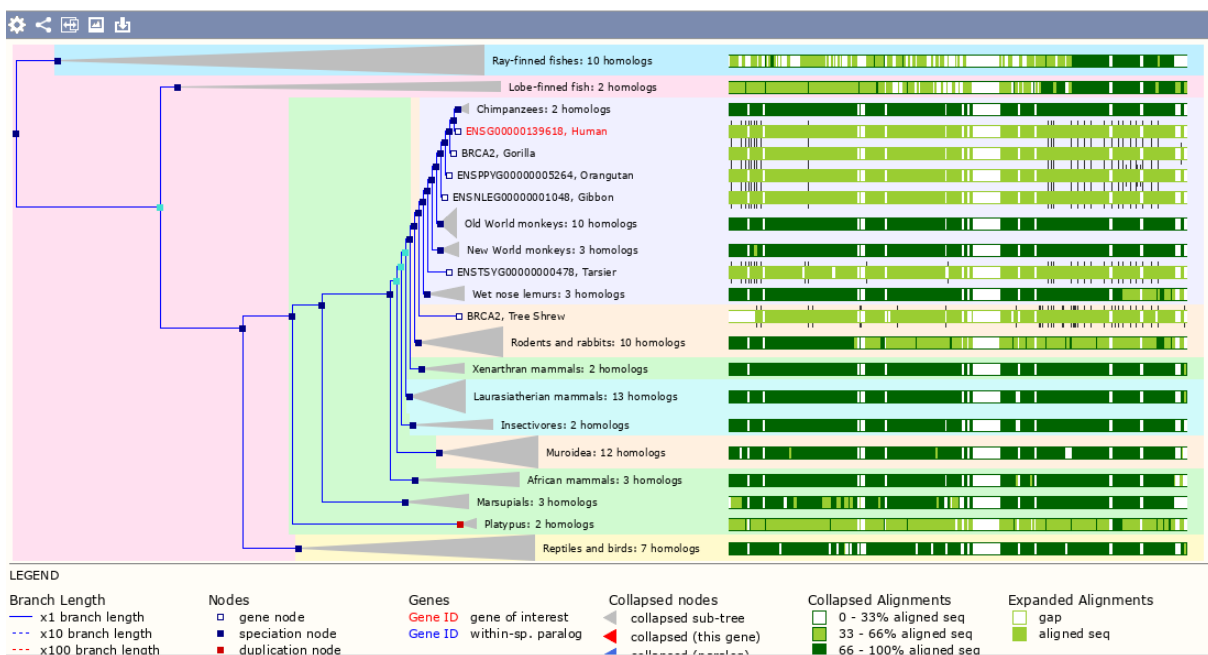
Original source | Variants (including SNPs and indels) imported from dbSNP (release 150)

About this variant | This variant overlaps [1 transcript](#)

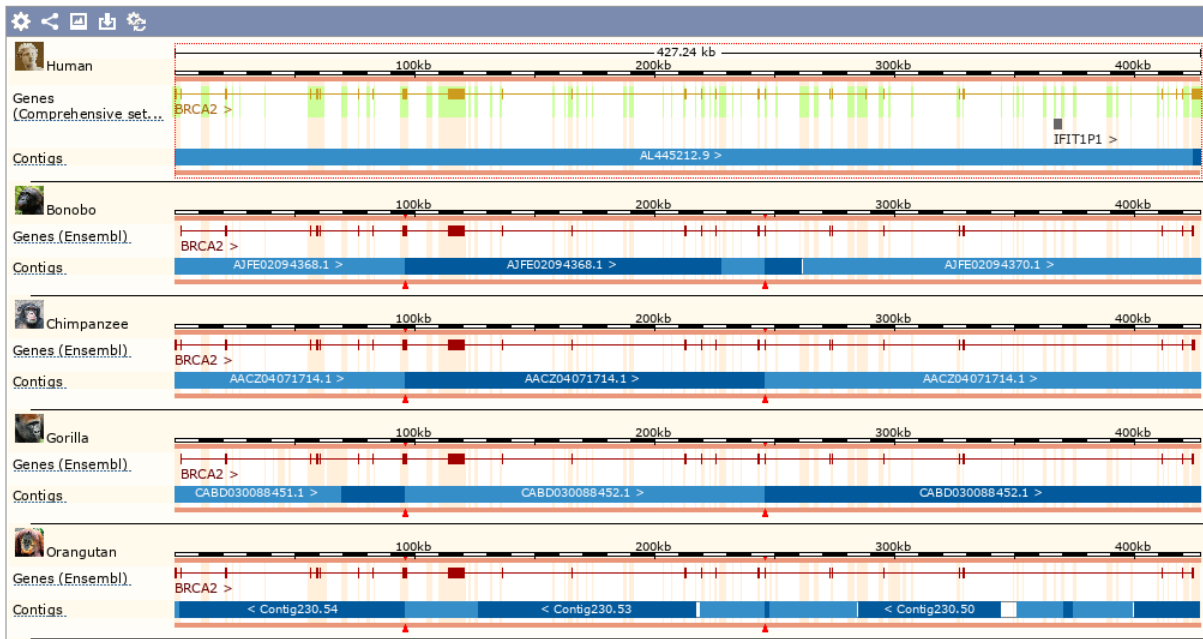
**Explore this variant**

- Genomic context
- Genes and regulation
- Flanking sequence
- Population genetics
- Phenotype data
- Sample genotypes
- Linkage disequilibrium
- Phylogenetic context
- Citations

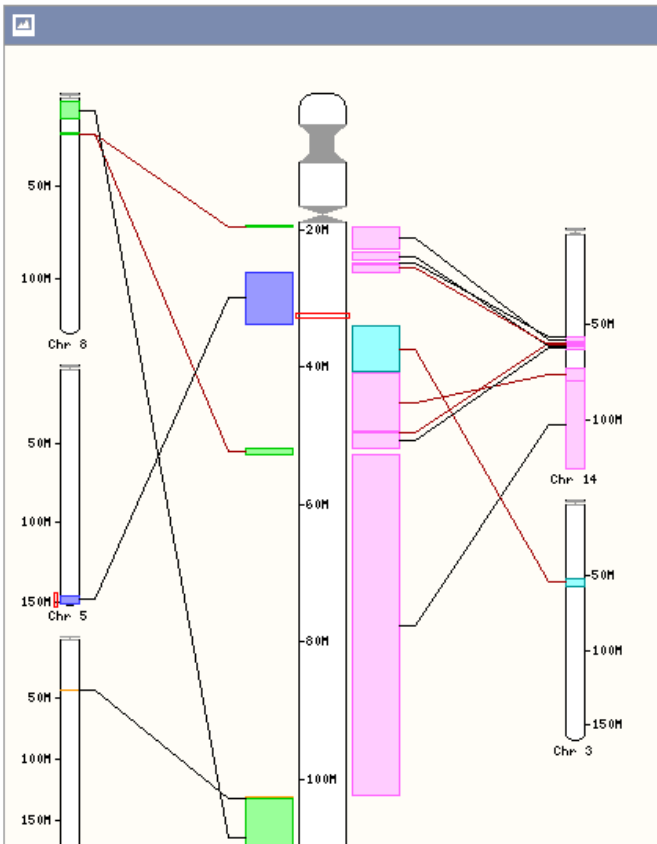
Információk róla pl. phylogenetic context



90 faj BRCA2 génje alapján gene tree a zöld szín intenzitása jelzi az illesztés erősségét, a piros a mi génünk, a tölcser összegyűjtött mód, fehér gap-eket jelöl, a node-kra kattintva megkapjuk azok jellemzőit. A piros node duplikációt jelöl, világos kék kétértelmű kék speciális. az evolúciós távolságok is láthatóak View alignment image: kékek a kontigok pirosak a breaing pointok,



Synteny: A szinténia boxok színes különböző kromoszómáról származó blokk különböző színű, fajokot lehet változtani



Nézzük az ortholog-okat: válasszuk a primates-eket-> selected orthologs, nézzük meg a bonobo alignmentjét

Species	Type	Orthologue	dN/dS	Target %id	Query %id	GOC Score	WGA Coverage	High Confidence
Angola colobus ( <i>Colobus angolensis palliatus</i> )	1-to-1	ENSCANG00000028513 <a href="#">View Gene Tree</a> <a href="#">Compare Regions</a> (KN981104.1:1,541,886-1,632,495-1) <a href="#">View Sequence Alignments</a>	0.53567	91.38 %	91.22 %	100	58.69	Yes
Black snub-nosed monkey ( <i>Rhinopithecus bieti</i> )	1-to-1	ENSRBIG00000001621 <a href="#">View Gene Tree</a> <a href="#">Compare Regions</a> (MCGX01034939.1:7,696-8,856-1) <a href="#">View Sequence Alignments</a>	0.42170	87.56 %	9.89 %	50	55.36	No
Bonobo ( <i>Pan paniscus</i> )	1-to-1	ENSPPAG00000002166 <a href="#">View Gene Tree</a> <a href="#">Compare Regions</a> (13.32.010,066-32,091,332.1) <a href="#">View Sequence Alignments</a>	0.31967	99.12 %	99.09 %	100	100.00	Yes
Bushbaby ( <i>Otolemur garnettii</i> )	1-to-1	BRCA2 ( <a href="#">ENSOGAG00000010588</a> ) <a href="#">View Gene Tree</a> <a href="#">Compare Regions</a> (GL873639.1:315,573-415,772-1) <a href="#">View Sequence Alignments</a>	0.49387	72.61 %	72.35 %	100	56.13	Yes
Capuchin	1-to-1	ENSCCAG000000033858	0.55600	88.17 %	87.62 %	100	100.00	Yes

Species	Gene ID	Peptide ID	Peptide length	% identity (cDNA)	% coverage	Genomic location
Human ( <i>Homo sapiens</i> )	ENSG00000139618	ENSP00000369497	3418 aa	99 %	99 %	13:32315474-32400266
Bonobo ( <i>Pan paniscus</i> )	ENSPPAG00000002166	ENSPPAP00000000836	3417 aa	99 %	100 %	13:32010066-32091332

CLUSTAL W (1.81) multiple sequence alignment

```

ENSP00000369497/1-10254   ATGCGTATTGGATCCAAAGAGAGGCCAACAATTTTITGAAATTTTAAAGACACGCTGCAAC
ENSPPAP00000000836/1-10251 ATGCGTATTGGATCCAAAGAGAGGCCAACAATTTTITGAAATTTTAAAGACACGCTGCAAC
*****

ENSP00000369497/1-10254   AAAGCAGATTTAGGACCAATAAGTCCTTAMTGGTTTGAAGAAGTCTTTCAGAGCTCCA
ENSPPAP00000000836/1-10251 AAAGCAGATTTAGGACCAATAAGTCCTTAMTGGTTTGAAGAAGTCTTTCAGAGCTCCA
*****

ENSP00000369497/1-10254   CCCTATAATTTCTGAACTGCAGAGAAATCTGAACTAAAACAAACAAATTAAGAACCAAC
ENSPPAP00000000836/1-10251 CCCTATAATTTCTGAACTGCAGAGAAATCTGAACTAAAACAAACAAATTAAGAACCAAC
*****

ENSP00000369497/1-10254   CTATTTAAACTCCACAAGGAAACCAATCTATAATCAGCTGGCTCAACTCCAATAATA
ENSPPAP00000000836/1-10251 CTATTTAAACTCCGCAAGGAAACCAATCTATAATCAGCTGGCTCAACTCCAATAATA
*****

ENSP00000369497/1-10254   TTCAAAGAGCAAGGGCTGACTCTGCGCTGTACCAATCTCCTGTAAAGAATTAGATAAA
ENSPPAP00000000836/1-10251 TTCAAAGAGCAAGGGCTGACTCTGCGCTGTACCAATCTCCTGTAAAGAATTAGATAAA
*****

```

Nézzük meg a protein familiest, az összes izoforma Markov clusterezése adja meg ezeket a protein családokat. 1 találat van! Nézzük meg a tagjait:

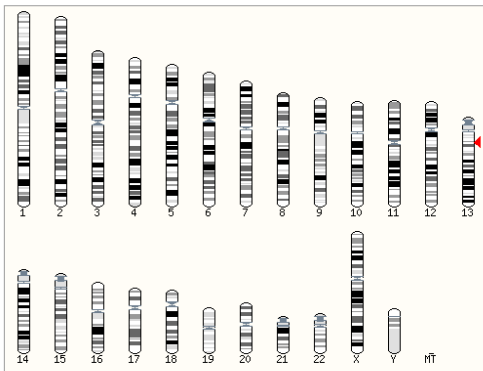
### Ensembl protein families

Family ID	Consensus annotation	Other BRCA2 proteins in this family	Multiple alignments
<a href="#">PTHR11289_SF0</a> (2 genes)	BREAST CANCER TYPE 2 SUSCEPTIBILITY HOMOLOG FANCONI ANEMIA GROUP D1 HOMOLOG	<ul style="list-style-type: none"> <li><a href="#">ENSP00000369497</a> (BRCA2-201)</li> <li><a href="#">ENSP00000434898</a> (BRCA2-202)</li> <li><a href="#">ENSP00000439902</a> (BRCA2-206)</li> </ul>	126 Ensembl members of this family <a href="#">Wasabi viewer</a>  1170 members of this family <a href="#">Wasabi viewer</a>

Menjünk a lokáció fölé nézzük meg a kromoszóma sávtérképét itt a gén piros boxsal jelölt.

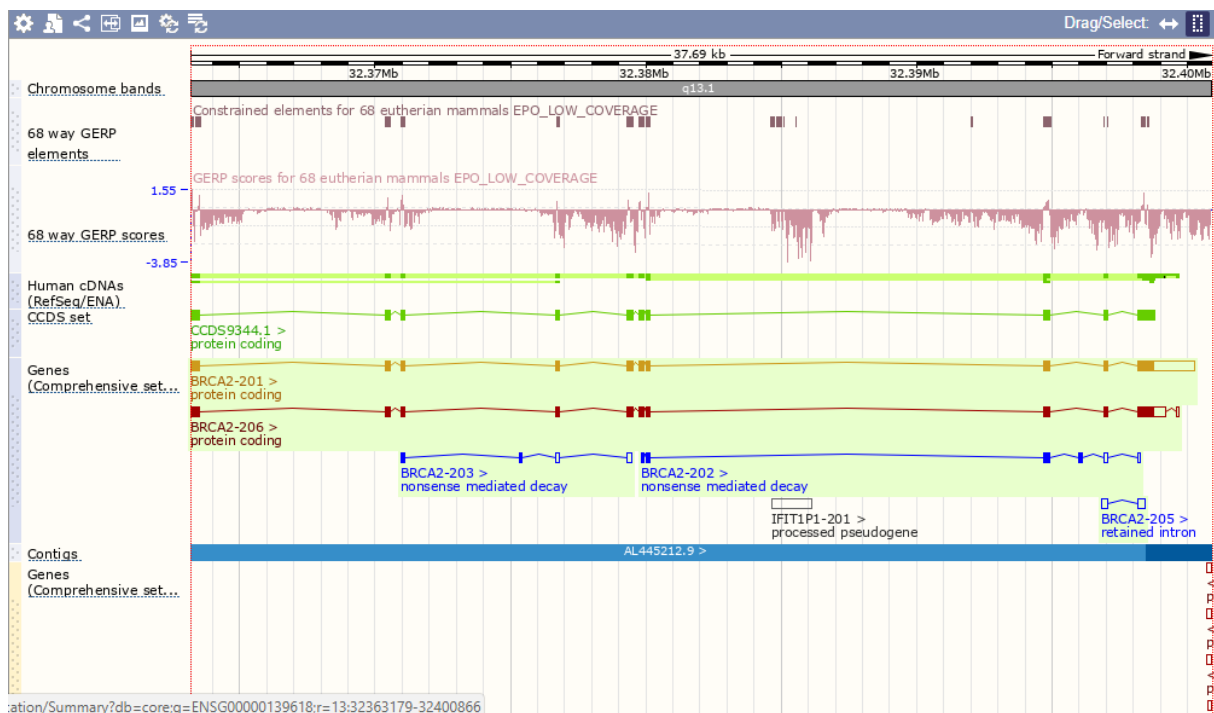
### HUMAN genes in this family

Ensembl genes containing proteins in family [PTHR11289\\_SF0](#)



Gene ID and Location	Gene Name	Description (if known)	Protein ID(s)
<a href="#">ENSG00000139618</a> Chromosome 13: 32.32m	<a href="#">BRCA2</a>	BRCA2, DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]	
<a href="#">LRG_293</a> LRG_293: 5.00k	<a href="#">BRCA2</a>	BRCA2, DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]	

Config this page: Comparative genomics menü, conservation régiók beállítása



\*Az izozimek vagy izoenzimek kifejezés olyan enzimekre utal, melyek aminosav-szekvenciájukban eltérnek, de ugyanazt a kémiai reakciót katalizálják különböző alléljai által kódolt enzimekre, és az izozimek jelentik a különböző génekből származó, de ugyanazt a reakciót katalizáló enzimeket, a két szót általában felcserélhető módon használják. (wikipédia)



\* Homológia a közös evolúciós eredetet jelenti (szekvenciák esetében). Nincs mértéke (nem fejezhető ki %-ban!!!), két szekvencia vagy homológ, vagy nem. A homológia két típusa: ortológia és paralógia

\*Ortológia:Két gén ortológ, ha két különbözőfajban találhatóak, és egy közös ős-génből származnak, mely a két faj közös ősében volt jelen. Ugyanazt a funkciót szolgálják, a két fajban. Példa: laktát dehidrogenáz (ill. génje) az emberben és az egérben.

\*Paralógia:Két gén paralóg, ha ugyanabban az organizmusban találhatóak, és egy közös ős-génből génduplikáció és azt követő-divergens evolúció útján alakultak ki. Többnyire különböző, de egymással összefüggésben lévőfunkciójuk van. Példa: a hisztidin-bioszintézis enzimei (ill. ezek génjei) emberben (nagyon hasonló szerkezetűek, de más-más reakciót katalizálnak). (wikipédia)

### Biomart:

- Ez egy alkalmazás a böngészőn belül
- Nem kell programozni az adatok kiexportálásához
- Néhány klikkel kereső szekvenciát hozhatsz létre
- Az eredményt megnézheted online vagy letöltheted fasta, xls, csv, tsv html formátumokban

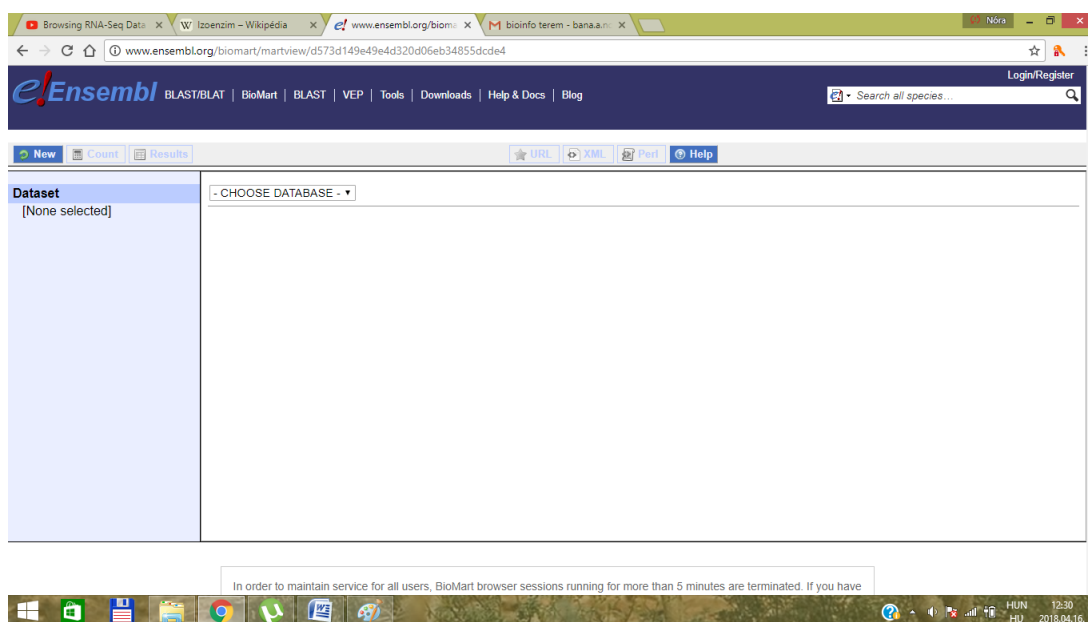
choose database ensembl-choose dataset human genes-Filters gene Input external refernces HGNC symbolt írjuk be a BRCA2, ATM

count: gyors összegzés

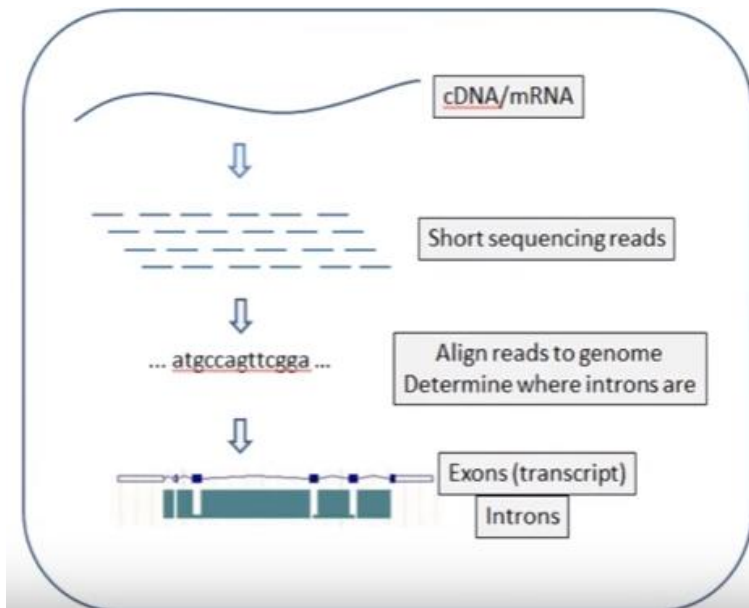
Attributes: External HGNC symbol , Refseq peptid predicted ID,

unqque results only , hogy a multiple transzkriptek redundanciáját csökkentsük

export : letölthető zippelt formátumban, vagy el lehet küldeni saját emai-címre is.



## RNA-seq adatok vizualizációja



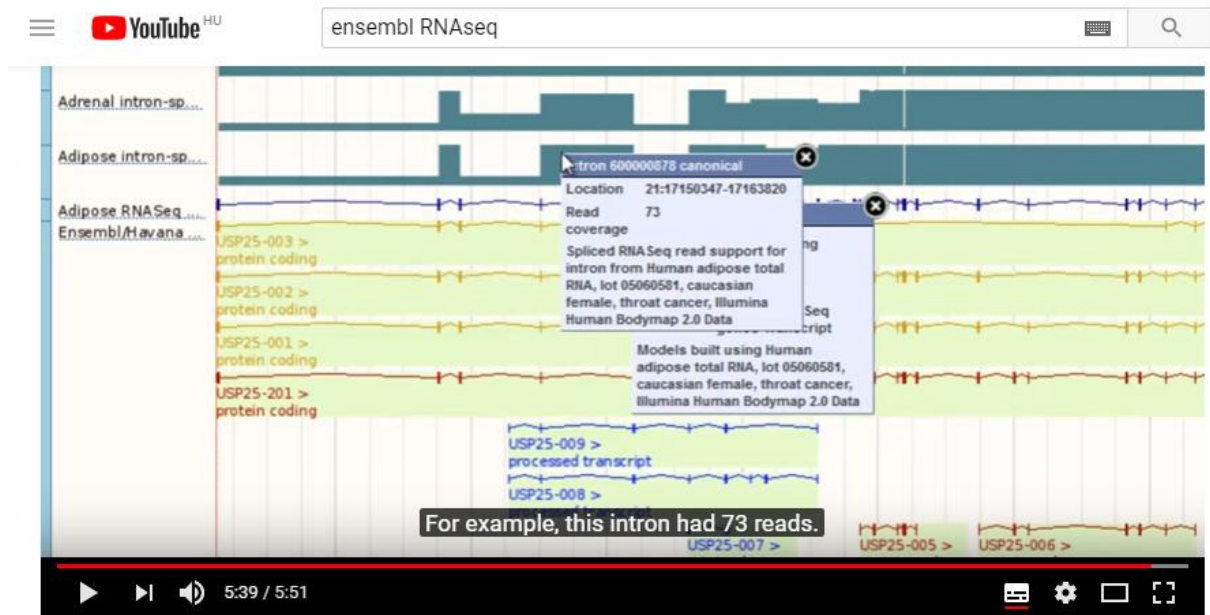
A human USP25 a zsírszövetben expresszálódik –e? Configure this page RNA-seq models az összeset klikkeld be

The screenshot shows a web interface for configuring RNA-seq models. On the left, there is a sidebar with a tree view of tracks. The 'Genes and transcripts' track is expanded, showing 'RNA-Seq models' with 32/32 items. On the right, there is a list of RNA-seq models for various tissues, each with a star icon and a blue 'i' icon. The list includes:

- Adipose intron-spanning reads
- Adrenal intron-spanning reads
- Blood intron-spanning reads
- Brain intron-spanning reads
- Breast intron-spanning reads
- Colon intron-spanning reads
- Heart intron-spanning reads
- Kidney intron-spanning reads
- Liver intron-spanning reads
- Lung intron-spanning reads
- Lymph intron-spanning reads
- Ovary intron-spanning reads
- Prostate intron-spanning reads
- Skeletal muscle intron-spanning reads
- Testes intron-spanning reads

Az RNA-seq-ben is látható intron exon rész, Az adipose gén model lehelyezem a sárgákhoz ezek cDNS-ekből származnak. A consensus coding sequence-t beázom CCDS, így látszik hogy a az adipose RNA-seq-em nagyon hasonlít (ugyanaz) mint a USP25-001 transzkript, vagyis ez a splice variant az zsírszövetben expresszálódik-> hipotézis igaz,

## Intron modellek:



Ontology és phenotypes cancer gene

## Több segítség:

Ensembl [documentációk](#), [videó segédlet \(tutorials\)](#) a YouTube-on, [FAQs](#), [gyakorlati példák](#), [publikációk](#), [online kurzusok](#). email a helpdesk-nek: [helpdesk@ensembl.org](mailto:helpdesk@ensembl.org), kövesd az Ensembl [blogot](#), [íratkozz fel a levelező listára](#)

Flicek, P. *et al* **Ensembl 2013** Nucleic Acids Res. Advanced Access (DatabaseIssue)  
<http://www.ncbi.nlm.nih.gov/pubmed/23203987>

**Ensembl Methods Series** <http://www.biomedcentral.com/series/ENSEMBL2010>

Xosé M. Fernández---Suárez and Michael K. Schuster **Using the Ensembl Genome Server to Browse Genomic Sequence Data**. UNIT 1.15 in Current Protocols in Bioinformatics, Jun 2010.

Giulietta M Spudich and Xosé M Fernández---Suárez **Touring Ensembl: A practical guide to genome browsing** BMC Genomics 2010, 11:295 (11 May 2010)