7. Bioinformatika gyakorlat

### [Ensembl genome browser 94](https://www.ensembl.org/)

Nyiss meg egy új Word dokumentumot, melybe a kérdésekre adott válaszokat írod és illeszted! (Vagy a válaszaidat ebbe a file-ba írod. Ha Shift+Enterrel töröd a sorokat, nem változik meg a sorszámozás.)

Hozz létre egy 07\_gyak mappát és abba mentsd le a doksit.

1. Nyisd meg https://www.ensembl.org/index.html

### Írd le a legújabb szarvasmarha genom néhány jellemzőjét! All genomes -> Select species cow

Hány bázis párból áll az assembly? More information and statistics->Statistics /Summary

Mennyi kódoló génje van? More information and statistics->Statistics/Gene counts

### Mennyi kromoszómája van? View karyotype

1. Keresd meg a legújabb human assembly 16 –os kromoszómájának 55169750-55388934 régiójában milyen merged Ensembl/Havana jelzésű géneket (színkód) találsz itt? Írd le ezek nevét (HGNC) és számát.

Search-> human for chr16:55,169,750-55,388,934 GO

1. Keresd meg emberben a saját proteined génjét. (ERK1-nél MAPK3-al keress)

Hol van? (génnév Human Gene) chr:start-end: orientáció (Melyik szálon található?)

Mennyi transzkriptuma van? About this gene vagy Kattints a génnévre -> Show transcript table

1. Az 1. kódoló transzkriptumnak, mennyi exonja van ebből mennyi kódoló? Milyen hosszú a transzkriptum (bázispár) és a belőle létrejövő protein (aminosav)?

Kattints a táblázatban a Transcript ID oszlopban az első transzkriptumra.

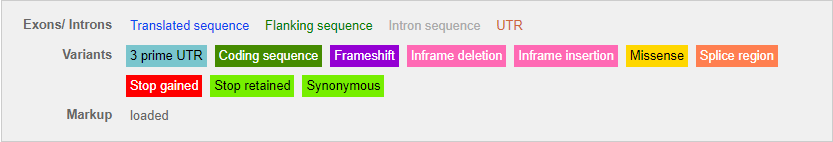
A képen a boxok vagy téglalapok exonokat, az ezeket összekőtő vonalak intronokat jelölnek. Ha a boxok színesek akkor kódoló szekvenciát jelölnek, ha nem akkor nem transzlálódó régiókat. A transzkriptumok színe a biológiai típusukra vonatkozik: a sárga az ensembl automatikus, annotációs pipeline-al annotált és ezenfelül a havana projectben azonosított génjeivel összefésült transzkriptum (jó minőségű). A piros csak az ensembl automatikusan annotált transzkriptumát jelöli. A kékek nem kódolóak.

1. Mi az ensembl azonosítója az 1. transzkriptumnak, ami fehérjét kódol, és mi a protein és a gén azonosítója? Van -e az 1. transzkriptum első exonjában olyan SNP ami stop kodont eredményezhet, mi a neve, milyen báziscsere és aminosav csere történik történik?

ENS + (ENSEMBL) + MUS (fajra jellemző 3 betű itt egérre utal, humánnál nincsen!) + G/T/P/E/R (gén/transzkriptum/protein/exon,/regulációs sajátság) + 11 számjegy

Kattints az első transzkriptumra, aztán baloldalon Summary fül-> Exons,

A lenti táblázat alapján az 1. ENSE(11 szám) szekvenciájában stop (Stop gained) variáns keresése, kattints rá, és a megjelenő ablakból szedd ki az információkat.



1. Keresd meg az első frameshift mutációt, ha van! mi az SNP-jének a neve, mi volt az ősi nukleotid? Keresd meg az első lila nukleotidot, kattints rá, a megjelenő ablakban kattints az Explore this variant opcióra.
2. Menj vissza a génedre (fül:Gene:a géned neve). Menj baloldalon lévő Sequence opcióra. Válaszd ki a Marked-up sequence alatt a BLAST this sequence gombot. A Search against lehetőségnél az Ad remove your protein opciót válaszd ki. A Species selector-ba gépeld be Chimpanzee. Nyomd meg az Apply gombot. DNS szekvenciával keress DNS szekvenciát! A search tool-ba válaszd a BLASTN programot. Futtasd a programot a Run gombbal. Melyik csimpánz kromoszómán kaptál találatot?

8) Információszerzés a kromoszóma régióról az Ensembl BioMart segítségével

1. A BioMart Filterét állítsd arra a kromoszómára és régióra (kar, szám), ahol a fehérjéd van (Filters, Region, Chromosome/scaffold és Karyotype band (start, stop ugyanaz)).

Hány gén található ezen a szakaszon (Count gomb)?

1. A következő dolgokat kérdezd le a génekről (Attributes):

(GENE) Gén neve (Gene Name), Ensembl génazonosítója (Gene stable ID)

a hozzá kapcsolódó fenotípus, betegség (EXTERNAL: MIM Morbid Description)

Mentsd le (Results) az egyedi (pipa Unique results only) eredmények táblázatát TSV formátumban és küld magadnak email-ben (Compressed web file notify by email).

Van olyan fenotípus vagy betegség, ami több génre, fehérjére jellemző?

+2 feladat

9) Nézd meg az géned alapján melyik primates emlős áll legközelebb az emberhez? Készíts törzsfát! (Gene-based displays-[Comparative Genomics](https://www.ensembl.org/Homo_sapiens/Gene/Compara?db=core;g=ENSG00000175387;r=18:47808957-47931146)- [Gene tree](https://www.ensembl.org/Homo_sapiens/Gene/SpeciesTree?db=core;g=ENSG00000175387;r=18:47808957-47931146) és [Gene gain/loss tree](https://www.ensembl.org/Homo_sapiens/Gene/SpeciesTree?db=core;g=ENSG00000175387;r=18:47808957-47931146) )

10) Milyen RNA-seq szekvenciák esnek az a géned régiójába, melyik szövetben expresszálódik az 5 közül (agy, zsír, here, vér, vese) nagymértékben? (Configure this page -RNA-seq models, lásd tutorials)