GRUP: David Marquez, Jaume Jurado, Eloi Vilella, Jan Izquierdo, Adrià Navarro

Consider the gene *ARL6*:

You will first study its location in the human genome and its transcriptional structure, according to the Ensembl genome browser. Go to the browser at https://www.ensembl.org/ and look for the gene symbol *ARL6* using the search box. Answer the following questions:

- **Symbol and name of the gene**
  ADP ribosylation factor like GTPase 6 / ARL6

- **Chromosomal location**. You must include the chromosome, start coordinate, end coordinate, and strand.

  Chromosome 3: 97,764,521-97,801,229 forward strand. GRCh38:CM000665.2

Navigate to the *Location* tab:

- **Genomic context**. Closer upstream (closest to its 5' end) and downstream (closest to its 3' end) protein-coding genes, presence of non-protein coding genes between the nearest upstream and downstream protein-coding genes.

  The protein-coding gene closer to the 5’ end is the EPHA6-201 but in between there is a IncRNA. The protein-coding gene closer to the 3’ end is the CRYBG3 but there is a IncRNA in between.

Return to the *Gene* tab. Regarding the transcriptional structure of the gene (consider only *GENCODE basic* transcripts):

- **Number of different transcripts and their lengths** (how many of them encode a protein?) there are 8 transcripts and the length of each is, 3988, 1630,1590, 937, 675, 1705, 851, 553.

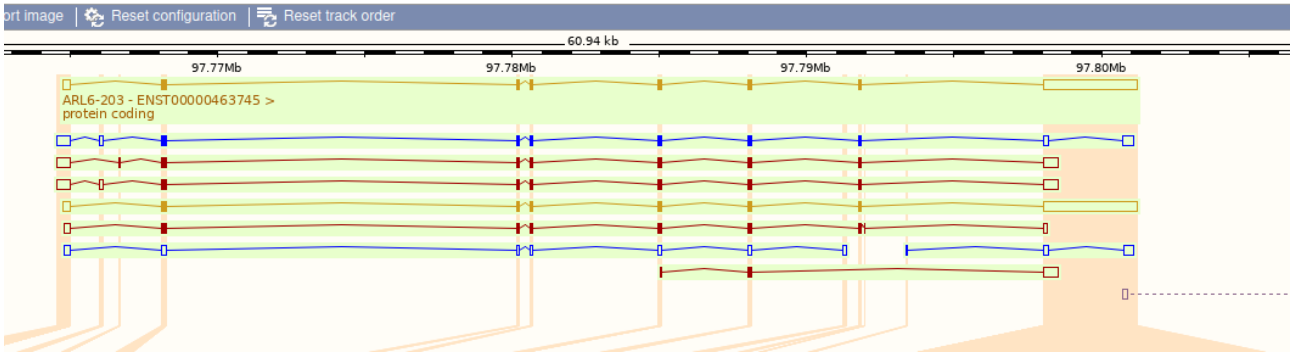
- **Number of different proteins**
  there are 3 different type of protein.

- **Number of constitutive exons** (present in all the transcripts) **and alternative exons** (only present in some transcripts)

  there are 3 constitutive exons and there are 9 alternative exons.

- **Alternative splicing mechanisms**

Include a caption of the gene region as displayed in Ensembl and indicate consitutive and alternative exons, as well as alternative splicing mechanisms.



To determine which are constitutive we look for the exons that is present in all the transcripts, for the alternative exons we did the same but we look for the ones that only present in some transcripts, and the alternative splicing mechanisms present in this gene are exon skipping and alternative 5’ or 3’ splice sites.

Highlight some other relevant information of the gene you may find **in other databases** discussed in this topic.

According to UniProt there are 2 isoforms, being considered one of them the canonical sequence. However, there are 3 computationally mapped isoforms for the entry of the protein.

According to NCBI there are 2 current annotation releases and another one that was the old one