UNIVERSIDADE DA CORUÑA

Lab 4: Getting information for a particular region of the genome

Fundamentals of Bioinformatics

Pedro Sánchez García

Season 2021-22

Lab 4.- Getting information for a particular region of the genome.

Introduction

The objective of this Lab 4 is to obtain information about a specific region of the genome, which is located on chromosome 2 and comprises positions 156,033,247 y 156,034,080. Then, the SNPs in the region will be determined.

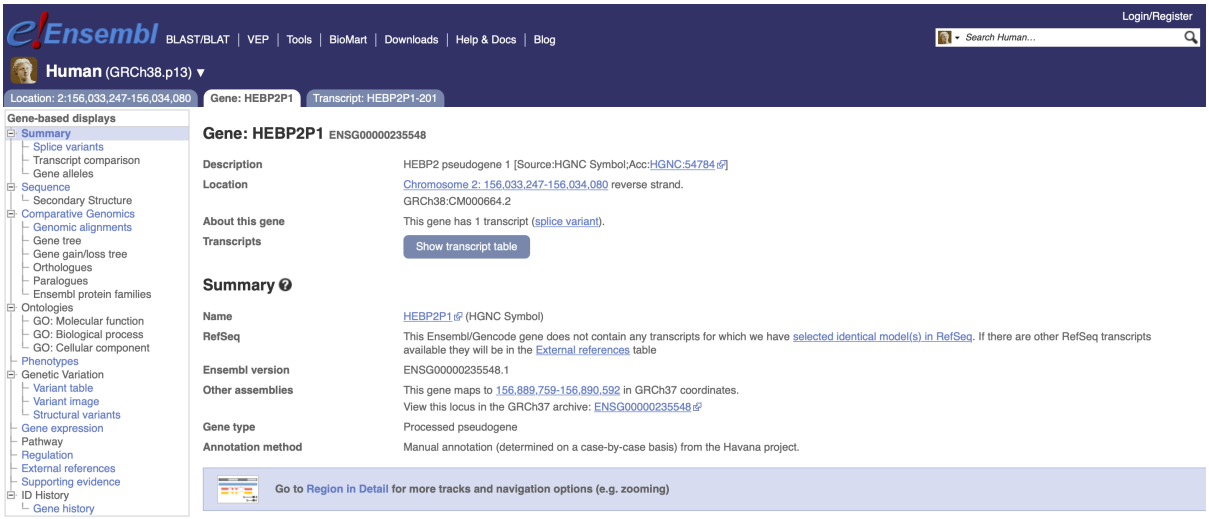
BioMart

1. Identification of the region

First, we access to BioMart and select GRCh38.p13 version of the human genome. Since the region of interest is located on chromosome 2, we filter by chromosome and indicate the corresponding coordinates. We put the attributes *Gene stable ID, Strand, Gene start (bp), Gene end (bp)* and *Gene name*. On the other hand, we verify that the GO term accession and GO term name attributes are checked. The table gives us the following results:

If we look at the coordinates of the second result, they coincide with the problem region, so that the gene ***HEBP2P1 pseudogene 1*** with ENSG00000235548 as *Gene stable ID* is the gen from which we want to obtain information.

2. Gen *HEBP2P1*



As shown in the previous image, we see that the gene has 1 transcript, which has 1 exon associated with 23 allelic variants.

3. Determination of SNPs in the *HEBP2P1* gen

To observe the SNPs that are present in the gene, we select *Variant table* in the left panel, which provides us with a table where we can filter the type of genetic variation in which we are interested. In this case, the *HEBP2P1* gene has 217 SNPs: