



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:

The screenshot shows the Ensembl BioMart interface. The 'Dataset' is 'Human genes (GRCh38.p13)'. The 'Filters' section shows 'Chromosome/scaffold: 2', 'Start: 156033247', and 'End: 156034080'. The 'Attributes' section includes 'Gene stable ID', 'Strand', 'Gene end (bp)', 'Gene start (bp)', and 'Gene name'. The 'Export' section shows 'all results to' and 'Email notification to'. The 'View' section shows '10 rows as HTML'. The resulting table is as follows:

Gene stable ID	Strand	Gene end (bp)	Gene start (bp)	Gene name
ENSG00000226383	-1	156254950	156011530	LINC01878
ENSG00000235548	-1	156034080	156033247	HEBP2P1

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*

The screenshot shows the Ensembl gene page for HEBP2P1. The 'Location' is 'Chromosome 2: 156,033,247-156,034,080 reverse strand'. The 'Gene: HEBP2P1' is selected. The 'Transcript: HEBP2P1-201' is also selected. The 'Summary' section shows the gene name 'HEBP2P1' (HGNC Symbol) and the description 'HEBP2 pseudogene 1 [Source:HGNC Symbol;Acc:HGNC:54784]'. The 'Location' is 'Chromosome 2: 156,033,247-156,034,080 reverse strand'. The 'About this gene' section states 'This gene has 1 transcript (splice variant)'. The 'Transcripts' section shows 'Show transcript table'. The 'Summary' section shows the gene name 'HEBP2P1' (HGNC Symbol) and the description 'This Ensembl/Gencode gene does not contain any transcripts for which we have selected identical model(s) in RefSeq. If there are other RefSeq transcripts available they will be in the External references table'. The 'Ensembl version' is 'ENSG00000235548.1'. The 'Other assemblies' section shows 'This gene maps to 156,889,759-156,890,592 in GRCh37 coordinates. View this locus in the GRCh37 archive: ENSG00000235548'. The 'Gene type' is 'Processed pseudogene'. The 'Annotation method' is 'Manual annotation (determined on a case-by-case basis) from the Havana project'. The 'Go to Region in Detail' button is visible at the bottom.

Human (GRCh38.p13) ▼

Location: 2:156,033,247-156,034,080 Gene: **HEBP2P1** Transcript: **HEBP2P1-201**

Transcript-based displays

- Summary
- Sequence
 - Exons
 - cDNA
 - Protein
- Protein Information
 - Protein summary
 - Domains & features
 - Variants
 - 3D Protein model
- Genetic Variation
 - Variant table
 - Variant image
 - Haplotypes
 - Population comparison
 - Comparison image
- External References
 - General identifiers
 - Oligo probes
- Supporting evidence
- ID History
 - Transcript history
 - Protein history

Configure this page Custom tracks Export data Share this page Bookmark this page

Transcript: HEBP2P1-201 ENST00000450741.1

Description HEBP2 pseudogene 1 [Source:HGNC Symbol;Acc:HGNC:54784]

Location [Chromosome 2: 156,033,247-156,034,080](#) reverse strand.

About this transcript This transcript has 1 [exon](#), is associated with 238 [variant alleles](#) and maps to 10 [oligo probes](#).

Gene This transcript is a product of gene [ENSG00000235548.1](#) [Show transcript table](#)

Summary

Statistics Exons: 1, Coding exons: 0, Transcript length: 834 bps, Transcript Support Level (TSL) [TSL:NA](#)

Version ENST00000450741.1

Type Processed pseudogene

Annotation Method Manual annotation (determined on a case-by-case basis) from the Havana project.

GENCODE basic gene This transcript is a member of the [Gencode basic](#) gene set.

Ensembl release 104 - May 2021 © [EMBL-EBI](#) [Permanent link](#) [View in archive site](#)

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:

Gene: HEBP2P1 ENSG00000235548

Description HEBP2 pseudogene 1 [Source:HGNC Symbol;Acc:HGNC:54784]

Location [Chromosome 2: 156,033,247-156,034,080](#) reverse strand.
GRCh38:CM000664.2

About this gene This gene has 1 transcript ([splice variant](#)).

Transcripts [Show transcript table](#)

Variant table

This table shows known variants for this gene. Use the 'Consequence Type' filter to view a subset of these.

Filter [Global MAF: All](#) [SIFT: All](#) [PolyPhen: All](#) [Consequences: All](#) [Class: SNP](#) [Filter Other Columns](#)

[working](#) [Show/hide columns](#) Search...

Variant ID	Chr: bp	Alleles	Global MAF	Class	Source	Evidence	Clin. Sig.	Conseq. Type	AA	AA co-ord	SIFT	Poly-Phen	CADD	REVEL	MetaLR	Mutation Assessment	Transcript
rs1169442563	2:156033259	T/C	-	SNP	dbSNP		-	non coding transcript exon variant	-	-	-	-	-	-	-	-	ENST00000450741.1
rs191144089	2:156033260	T/C	-	SNP	dbSNP		-	non coding transcript exon variant	-	-	-	-	-	-	-	-	ENST00000450741.1
rs1573867962	2:156033273	T/C	-	SNP	dbSNP		-	non coding transcript exon variant	-	-	-	-	-	-	-	-	ENST00000450741.1
rs1404473378	2:156033274	T/C	-	SNP	dbSNP		-	non coding transcript exon variant	-	-	-	-	-	-	-	-	ENST00000450741.1
rs1316438423	2:156033275	T/C	-	SNP	dbSNP		-	non coding transcript exon variant	-	-	-	-	-	-	-	-	ENST00000450741.1
rs957685215	2:156033279	A/G	-	SNP	dbSNP		-	non coding transcript exon variant	-	-	-	-	-	-	-	-	ENST00000450741.1
rs1272817121	2:156033280	A/G	-	SNP	dbSNP		-	non coding transcript exon variant	-	-	-	-	-	-	-	-	ENST00000450741.1