# Practical assignment

**Module topic: Genomics**

**Contact session title: Genome Browsers**

**Trainer: Astrid Gall**

**Name:**

**Date:**

# Assignment

## Introduction

The Ensembl genome browser allows access to publicly available reference genomic data, such as genes, through tools and a visual interface.

## Tools used in this session

Ensembl Genome browser and BioMart. www.ensembl.org

## Please note

**Hand-in information** If you are formally enrolled in the IBT course, please upload your completed practical assignment to the Vula ‘Practical Assignments’ tab. Take note of the final hand-in date for each practical assignment, which will be indicated on Vula.

## Task 1 – Browsing genomic regions

1. Watch the first video of the session, the introduction to Ensembl and the second video, the demo of browsing genomic regions.
2. In the Ensembl genome browser at www.ensembl.org, navigate to the region 31,937,000 to 32,633,000 bp on human chromosome 13.
   1. On which cytogenetic band is this region located?
   2. How many contigs make up this portion of the assembly (contigs are contiguous stretches of DNA sequence that have been assembled solely based on direct sequencing information)?
3. Configure this page to turn on the LTR (repeat) track in this view.
   1. What tool was used to annotate the LTRs according to the track information?
   2. How many LTRs can you see within the BRCA2 gene? Do any overlap exons?
   3. What does the word “track” mean in the context of genome browsers?
4. Export the genomic sequence of the region you are looking at in FASTA format (only paste in the first three lines to this report).

## 

## Task 2 – Exploring genes and transcripts

1. Watch the third video of the session, the lecture describing the origin of Ensembl genes and transcripts and the fourth video, the demo exploring them in the browser.
   1. What two methods are used to annotate genes in Ensembl?
2. Find the human MYH9 (myosin, heavy chain 9, non-muscle) gene, and go to the Gene tab.
   1. On which chromosome and which strand of the genome is this gene located?
   2. How many transcripts (splice variants) are there and how many are protein coding?
   3. What is the longest transcript, and how long is the protein it encodes?
   4. Which transcript would you take forward for further study?
   5. Click on Phenotypes at the left side of the page. Are there any diseases associated with this gene, according to MIM (Mendelian Inheritance in Man)?
   6. What are some functions of MYH9 according to the Gene Ontology consortium? Have a look at the GO pages for this gene.
3. In the transcript table, click on the transcript ID for MYH9-201, and go to the Transcript tab.
   1. How many exons does it have?
   2. Are any of the exons completely or partially untranslated?
   3. Is there an associated sequence in UniProtKB/Swiss-Prot? Have a look at the General identifiers for this transcript.
   4. Are there microarray (oligo) probes that can be used to monitor expression of ENST00000216181?

## Task 3 – Using BioMart to bulk export gene data

1. Watch the fifth video of the session, with the lecture introducing BioMart and the sixth video, the BioMart demo.
2. Begin a BioMart query in human genes. Filter the data to see only genes on chromosome 9 with transmembrane helices.
   1. How many genes are there?
3. Get the position of the transmembrane domain in the attributes alongside the gene name.
   1. Export the results table as an Excel file. Include the first five lines of the file that show transmembrane helices here.
4. Change your attribute to get the protein sequence instead, with the gene name in the header.
   1. Export the results as a FASTA file. Include the sequence of the first protein here.