# Practical assignment

**Module topic: Genomics**

**Contact session title: Human genetic variation**

**Trainer: Astrid Gall**

**Name:**

**Date:**

# Assignment

## Introduction

The Ensembl genome browser allows access to variation data, and annotation of your own variants through the Variant Effect Predictor (VEP) tool.

## Tools used in this session

Ensembl Genome browser and VEP. 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 – Exploring known genetic variation

1. Watch the first video of the session, the introduction to Ensembl variation and the second video, the demo of Ensembl variation.
   1. What is meant by reference and alternative alleles?
2. The SNP rs1738074 in the 5’ UTR of the human TAGAP gene has been identified as a genetic risk factor for a few diseases. Search for this SNP in Ensembl.
   1. In which transcripts is this SNP found?
   2. What is the least frequent genotype for this SNP in the Yoruba (YRI) population from the 1000 Genomes phase 3?
   3. What is the ancestral allele? Is it conserved in other primates?
   4. With which diseases is this SNP associated? Are there any known risk (or associated) alleles?

## 

## Task 2 – Annotating genetic variants with the VEP

1. Watch the third video of the session, the lecture describing common variant formats and the fourth video, the demo using the VEP.
   1. Express the following variant in HGVS, a Leucine to Proline change at position 442 in the protein ENSP00000062863.3 (not a real protein).
   2. Express the following variant in VCF, a Cytosine to Thymine change at position 18365927 on chromosome 10, with the identifier rs7801903 (not a real variant).
2. Run the example file (IBT2020-genomics.human\_genetic\_variation.TaskVEP.vcf) through the VEP online tool.
   1. How many genes have been overlapped by the variants in this file?
   2. How many of the variants in the input file were novel?

## 

## Task 3 – Variant prioritisation

1. Watch the fifth video of the session, with the lecture discussing variant prioritisation and the sixth video, the variant prioritisation demo.
   1. What would you consider when carrying out variant prioritisation?
2. The example file you ran through the VEP in Task 2 (IBT2020-genomics.human\_genetic\_variation.TaskVEP.vcf) was taken from a patient with Cornelia de Lange Syndrome (CdLS). You have a list of known genes associated with CdLS, which is shown below.
   1. Filter the VEP output to identify variants that you would take through to further study. Which ones would you study further and why? Were there others that you considered and rejected?

**CdLS Gene list:**

NIPBL

RAD21

SMC3

SMC1A

HDAC8

SETD5

KMT2A