Variation

Demo: Exploring variants in Ensembl

In any of the sequence views shown in the Gene and Transcript tabs, you can view variants on the sequence. You can do this by clicking on Configure this page from any of these views.

Let’s take a look at the Gene sequence view for AGT in human. Search for AGT and go to the Sequence view.

If you can’t see variants marked on this view, click on Configure this page and select Show variants: Yes and show links.

Find out more about a variant by clicking on it.

You can add variants to all other sequence views in the same way.

You can go to the Variation tab by clicking on the variant ID. For now, we’ll explore more ways of finding variants.
To view all the sequence variations in table form, click the **Variant table** link at the left of the gene tab.

You can filter the table to only show the variants you’re interested in. For example, click on **Consequences: All**, then select the variant consequences you’re interested in.

You can also filter by the different pathogenicity scores and MAF, or click on **Filter other columns** for filtering by other columns such as Evidence or Class.

The table contains lots of information about the variants. You can click on the IDs here to go to the Variation tab too.

Let’s look at Structural Variation in the Gene Tab. You’ll find it in the left-hand menu.
You can click on the structural variants (SVs) in the image, or on their IDs in the table to go to the SV tab.

You can also see the phenotypes associated with a gene. Click on Phenotype in the left hand menu.

The Haplotypes view in the transcript tab shows you the actual protein and CDS sequences in 1000 Genomes individuals. Open the transcript table and go to AGT-201 ENST00000366667, then click on Haplotypes in the left hand menu.
Click on one of the haplotypes, we'll go for 207T>M,268M>T, to find out more about it. Here you will see the frequency in the 1000 Genomes subpopulations, the sequence and the 1000 Genomes individuals where this protein is found.

Let's have a look at variants in the Location tab. Click on the Location tab in the top bar.

Configure this page and open Variation from the left-hand menu.
There are various options for turning on variants. You can turn on variants by source, by frequency, presence of a phenotype or by individual genome they were isolated from. Turn on the following sequence variants in Expanded with name:

- 1000 genomes – All
- 1000 genomes – All – common
- All phenotype-associated variants

Also turn on Larger and Smaller Structural variants (all sources) in Expanded.

Click on a variant to find out more information. It may be easier to see the individual variants if you zoom in.

Let’s have a look at a specific variant. If we zoomed in we could see the variant rs699 in this region, however it’s easier to find if we put rs699 into the search box. Click through to open the Variation tab.
The icons show you what information is available for this variant. Click on Genes and regulation, or follow the link at the left.

This variant is found in AGT only. It has been associated with 987 gene expression level changes. It has not been associated with any regulatory features or motifs.

Click on 3D protein model to see the variant on the protein. Like the transcript tab, this is a LiteMol viewer and works in the same way. The variant is highlighted on the structure, but you may have to rotate the protein to see it.
Let’s look at population genetics. Click on Population genetics in the left-hand menu.

These data are mostly from the 1000 genomes, HapMap, GnomAD, TOPMed and UK10K projects in human.

There are big differences in allele frequencies between populations. Let’s have a look at the phenotypes associated with this variant to see if they are known to be specific to certain human populations. Click on Phenotype Data in the left-hand menu.
This variant is associated with various phenotypes including coronary artery disease. Are there other variants in the genome that also cause coronary artery disease? Click on the phenotype coronary artery disease to find out.

Demo: The Variant Effect Predictor (VEP)

We have identified four variants on human chromosome nine, an A deletion at 128328461, C->A at 128322349, C->G at 128323079 and G->A at 128322917.

We will use the Ensembl VEP to determine:
- Have my variants already been annotated in Ensembl?
- What genes are affected by my variants?
- Do any of my variants affect gene regulation?

Go to the front page of Ensembl and click on the Variant Effect Predictor.
This page contains information about the VEP, including links to download the script version of the tool. Click on Launch VEP to open the input form.

The data is in the format:
Chromosome Start End alleles (reference/alternative) strand name

Put the following into the Paste data box:
9 128328461 128328461 A/- + var1
9 128322349 128322349 C/A + var2
9 128323079 128323079 C/G + var3
9 128322917 128322917 G/A + var4

The VEP will automatically detect that the data is in Ensembl default format.

There are further options that you can choose for your output. These are categorised as Identifiers, Variants and frequency data, Additional annotations, Predictions, Filtering options and Advanced options. Let’s open all the menus and take a look.
Select Phenotypes.
Choose to see scores for protein changes.

Choose to see scores for splicing changes.

Choose to see conservation at the variant locus.

Choose to only see common or rare variants.

Limit the number of consequences you see per variant.

Hover over the options to see definitions.

When you’ve selected everything you need, scroll right to the bottom and click Run.
The display will show you the status of your job. It will say Queued, then automatically switch to Done when the job is done, you do not need to refresh the page. You can edit or discard your job at this time. If you have submitted multiple jobs, they will all appear here.

Click View results once your job is done.

In your results you will see a graphical summary of your data, as well as a table of your results.
### Summary of the consequences of all your variants

- Filter your data
- Download your data
- Put your data into BioMart

#### Additional Notes:

- All four affect multiple transcripts of two genes.
- Information about the consequences on genes and positions within them.
- Some variants already exist in our database: click to find out more about them.
- SIFT/ PolyPhen predictions
- Phenotypes linked to the known variants and genes.