

## Variant Effect Predictor

### **Demo: The Variant Effect Predictor (VEP)**

We have identified five variants on human chromosome nine, an A deletion at 128328461, C->A at 128322349, C->G at 128323079, G->A at 128322917 and A->G at 128322495.

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 [VEP button](#).



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.

**Input**

Species:  Assembly: GRCh38

Name for this data (optional):

Either paste data:

```
9 128328461 128328461 A/- + var1
9 128322349 128322349 C/A + var2
9 128323079 128323079 C/G + var3
9 128322917 128322917 G/A + var4
```

Examples: [Ensembl default](#), [VCF](#), [Variant identifiers](#), [HGVS notations](#), [Pileup](#)

[Quick results for first variant >](#)

Or upload file:  No file chosen

Or provide file URL:

Transcript database to use:

- Ensembl transcripts
- Gencode basic transcripts
- RefSeq transcripts
- Ensembl and RefSeq transcripts

**Callouts:**

- Give your data a name (points to Name for this data)
- Put your data in here (points to Either paste data)
- You can also upload a file (points to Or upload file)
- Choose your transcript database (points to Transcript database to use)

The data is in the format:

Chromosome Start End alleles (ref/alt) 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
9 128322495 128322495 A/G + var5
```

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 and frequency data**, **Filtering options** and **Extra options**. Let's open all the menus and take a look.

**Identifiers and frequency data** Additional identifiers for genes, transcripts and variants; frequency data

**Identifiers**

|  |   |
|--|---|
| Gene symbol:                                     | <input checked="" type="checkbox"/>   |
| CCDS:  | <input type="checkbox"/>  |
| Protein:   | <input type="checkbox"/>  |
| Uniprot:   | <input type="checkbox"/>  |
| HGVS:  | <input type="checkbox"/>  |
| Find co-located known variants:                  | Yes <input type="button" value="v"/>  |
| Frequency data for co-located variants:          | <input checked="" type="checkbox"/> 1000 Genomes global minor allele frequency<br><input type="checkbox"/> 1000 Genomes continental allele frequencies<br><input type="checkbox"/> ESP allele frequencies |
| PubMed IDs for citations of co-located variants: | <input checked="" type="checkbox"/>   |

**Extra options** e.g. SIFT, PolyPhen and regulatory data

|                                     |   |
|-------------------------------------|---|
| Transcript biotype:                 | <input checked="" type="checkbox"/>                   |
| Protein domains:                    | <input type="checkbox"/>                              |
| Exon and Intron numbers:            | <input type="checkbox"/>                              |
| Identify canonical transcripts:     | <input type="checkbox"/>                              |
| SIFT predictions:                   | Prediction and score <input type="button" value="v"/> |
| PolyPhen predictions:               | Prediction and score <input type="button" value="v"/> |
| Get regulatory region consequences: | Yes <input type="button" value="v"/>                  |

**Filtering options** Pre-filter results by frequency or consequence type

**Filters**

|   |  |
|---|--|
| By frequency:                                       | <input checked="" type="radio"/> No filtering<br><input type="radio"/> Exclude common variants<br><input type="radio"/> Advanced filtering |
| Return results for variants in coding regions only: | <input type="checkbox"/>   |
| Restrict results:                                   | Show all results <input type="button" value="v"/>  |

NB: Restricting results may exclude biologically important data!

Which identifiers do you want to see?

Find out if variants already exist in our database

Get frequency data

Choose to see scores for protein changes

Get consequences on reg feats and motifs

Choose to only see common or rare variants

Many of the options have a question mark [?] after the names. Hover over the question mark [?] 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. (Note that some empty columns in the results table have been hidden in the following screenshot to save space.)



## **Exercise**

### **Running the VEP using a VCF file**

There is a VCF file of chr21 variants at <https://www.encodeproject.org/tutorials/encode-users-meeting-2015/> called [VEP\\_input.vcf](#). Run the VEP using the VCF file to find out the consequences of the variants.

- (a) Do any variants affect regulatory features? What kinds of regulatory features?
  
- (b) How many variants affect transcription factor binding motifs? What is the biggest change in the motif score as a result of one of these variants?

## **VEP Exercise Answer**

Open the VEP tool by clicking on the [Variant Effect Predictor](#) button from the Ensembl homepage, then click [Launch VEP](#). Click [New VEP job](#).

Paste the URL into the box [Or provide file URL](#). Scroll to the bottom and click [Run](#).

When your job is listed as [Done](#), click [View Results](#).

(a) To see only `regulatory_region_variant`, use the filters. Select [Consequence](#) from the dropdown, and [is](#), then put in [regulatory\\_region\\_variant](#). Click [Add](#) to apply the filter.

You can see all the variants that hit regulatory features. Some affect promoters, promoter-flanks and CTCF-binding sites, as well as TF binding sites.

(b) Use the filters again, this time select [Consequence](#), [is](#), [TF\\_binding\\_site\\_variant](#). Click [Add](#).

Scroll to the far right to see the Motif Score change.

There are ten variants that are `TF_binding_site_variants`. The biggest score change is -0.203 for rs730996 on the motif Egr1:MA0341.1.