

# Ensembl Variant Effect Predictor (VEP)

1

We're going to use Ensembl VEP to look at a set of variants:

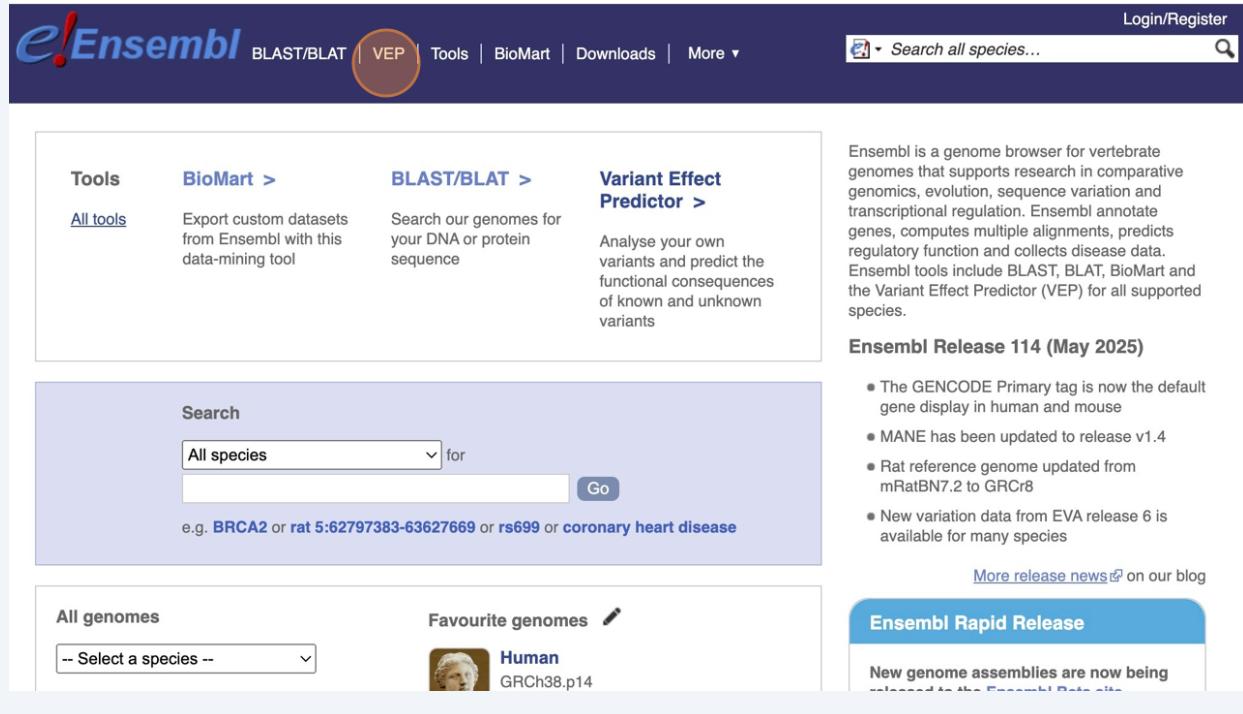
- 14 73219188 73219188 C/T
- 6 161785778 161785778 A/C
- 7 117480097 117480097 G/T
- 7 117587806 117587806 G/A
- 10 62813413 62813413 G/A

Ensembl VEP can help determine if:

- The variants have already been annotated in Ensembl.
- Genes and transcripts affected by these variants.
- Any of the variants affect a protein.

The screenshot shows the Ensembl homepage with a dark blue header. On the left, there's a logo with the word "Ensembl". To the right are links for "BLAST/BLAT", "VEP", "Tools", "BioMart", "Downloads", and "More". A search bar with a magnifying glass icon is on the right. Below the header, there are four main tool sections: "BioMart >" (with a link to "All tools"), "BLAST/BLAT >" (with a link to "Search our genomes for your DNA or protein sequence"), and "Variant Effect Predictor >" (with a link to "Analyse your own variants and predict the functional consequences of known and unknown variants"). To the right of these is a summary text about Ensembl and its tools. Below this is a "Release 114 (May 2025)" section with a bulleted list of changes. Further down are sections for "All genomes" (with a dropdown menu) and "Favourite genomes" (listing "Human GRCh38.p14" with a small profile picture). A "Ensembl Rapid Release" box on the right says "New genome assemblies are now being released to the [Ensembl Beta site](#)".

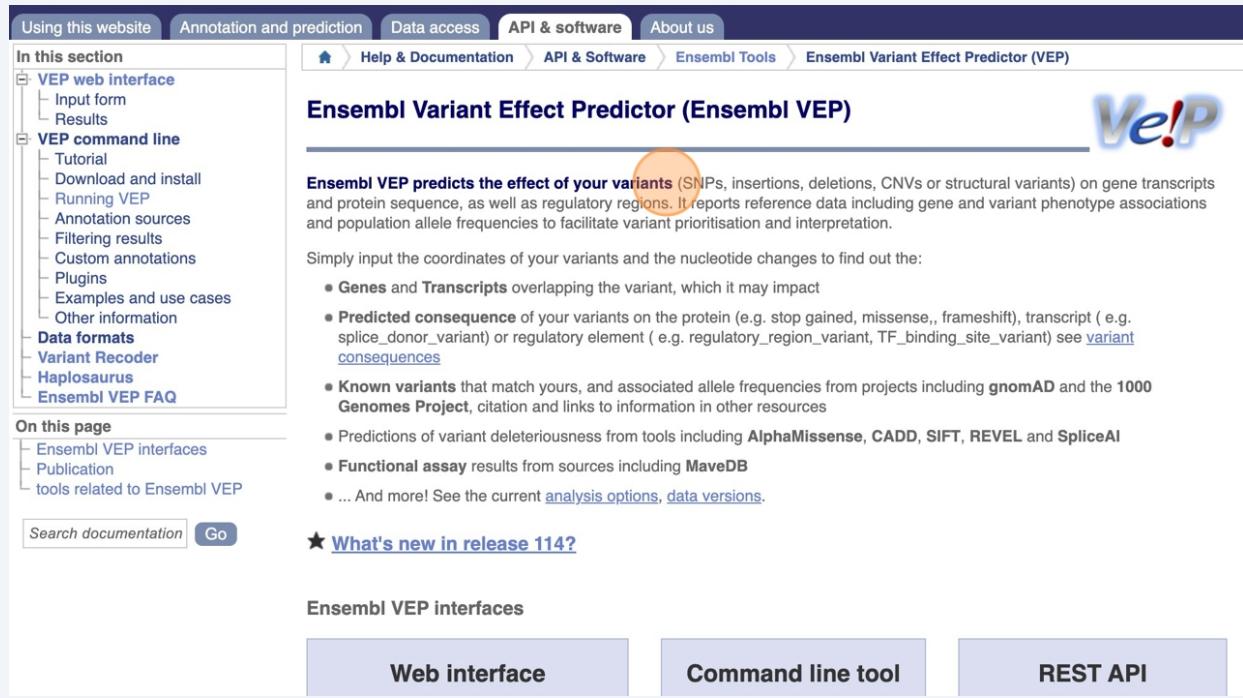
- 2 Go to "www.ensembl.org" and click on the "VEP" available on the blue navigation bar at the top of the page.



The screenshot shows the Ensembl homepage. At the top, there is a dark blue header with the Ensembl logo on the left and a search bar on the right. Below the header, there is a navigation bar with links for BLAST/BLAT, VEP (which is highlighted with a red circle), Tools, BioMart, Downloads, and More. The main content area has several sections: 'Tools' (with links to BioMart and BLAST/BLAT), 'BioMart >' (with a link to 'All tools'), 'BLAST/BLAT >' (with a link to 'All species'), and 'Variant Effect Predictor >' (with a link to 'Analysing variants'). To the right, there is a large text block about Ensembl, followed by a section titled 'Ensembl Release 114 (May 2025)' with a bulleted list of changes. Below this is a 'Search' bar with a dropdown menu set to 'All species' and a 'Go' button. A note below the search bar says 'e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease'. Further down, there are sections for 'All genomes' (with a dropdown menu) and 'Favourite genomes' (listing 'Human' with GRCh38.p14). On the right, there is a 'Ensembl Rapid Release' section with a note about new genome assemblies.

- 3 This page contains information about the Ensembl VEP, including links to download the script version of the tool.

We will use the web interface for this demo. Click on "launch VEP" to open the input form.



The screenshot shows the Ensembl Variant Effect Predictor (VEP) web interface. At the top, there is a dark blue header with a navigation bar containing 'Using this website', 'Annotation and prediction', 'Data access', 'API & software', and 'About us'. Below the header, there is a sidebar on the left with a tree view of links: 'In this section' (VEP web interface, VEP command line, Data formats, Variant Recoder, HaploSaurus, Ensembl VEP FAQ), 'On this page' (Ensembl VEP interfaces, Publication, tools related to Ensembl VEP), and a 'Search documentation' input field with a 'Go' button. The main content area has a title 'Ensembl Variant Effect Predictor (Ensembl VEP)' with a 'VeIP' logo. Below the title, there is a paragraph about Ensembl VEP predicting variants and their impact. It then lists several bullet points about the tool's features, such as inputting variant coordinates, predicted consequences, known variants, and predictions from other tools. At the bottom, there is a section titled 'Ensembl VEP interfaces' with three buttons: 'Web interface' (highlighted with a red circle), 'Command line tool', and 'REST API'.

4

The image shows three cards side-by-side, each with a title, a small icon, a list of features, and a 'Documentation' link.

- Point-and-click interface**
  - Point-and-click interface
  - Suits smaller volumes of data[Documentation](#)
- Command-line interface**
  - More options and flexibility
  - For large volumes of data[Documentation](#)
- re!st**
  - Language-independent API
  - Simple URL-based queries[Documentation](#)

**Launch VEP** button

#### Publication

If you use Ensembl VEP, please cite our most recent publication to help us continue to support development:

[Cite us](#)

5

The input form gives you the option to select the species of interest, name your job, copy and paste or upload your data.

The screenshot shows the Ensembl Variant Effect Predictor (VEP) input form. The top navigation bar includes links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, and More. A search bar is also present. The main content area is titled "Variant Effect Predictor". It features a "New job" button and a "Clear form" button. A dropdown menu for "Species" is set to "Homo\_sapiens" (with an orange circle highlighting the "X" button). Below this, there is a note about assembly: "Assembly: GRCh38.p14" and a link to "Change species". There is also a note: "If you are looking for VEP for Human GRCh37, please go to [GRCh37 website](#)". The form includes fields for "Name for this job (optional)" and "Input data", which is described as "Either paste data:" followed by a large text area.

- 6 Click the "Name for this job (optional):" field.

The screenshot shows the Variant Effect Predictor (VEP) web interface. The top navigation bar has a dropdown menu 'VEP' with a downward arrow. Below it is a sidebar titled 'Web Tools' containing links to 'BLAST/BLAT', 'Variant Effect Predictor' (which is highlighted in purple), 'Linkage Disequilibrium Calculator', 'Variant Recoder', 'File Chameleon', 'Assembly Converter', 'ID History Converter', 'VCF to PED Converter', and 'Data Slicer'. To the right of the sidebar is the main content area. The title 'Variant Effect Predictor' is at the top, followed by a 'New job' button and a 'Clear form' button. A 'Species:' field contains 'Homo\_sapiens' with a small profile icon. Below it, 'Assembly: GRCh38,p14' and a link 'Change species' are shown. A note says 'If you are looking for VEP for Human GRCh37, please go to [GRCh37 website](#)'. A text input field labeled 'Name for this job (optional):' contains 'Demo\_ECCB' and is circled in orange. Below it is a large text area labeled 'Input data:' and 'Either paste data:' with a placeholder 'Examples: Ensembl default, VCF, Variant identifiers, HGVS notations, SPDI'. At the bottom, there are fields for 'Or upload file:' (with a 'Choose file' button showing 'No file chosen') and 'Or provide file URL:' (with an empty input field).

7

The demo input data is in the Ensembl default format. The default format is a simple whitespace-separated format (columns may be separated by space or tab characters), containing five required columns plus an optional identifier column:

- chromosome - just the name or number, with no 'chr' prefix
- start
- end
- allele - pair of alleles separated by a '/', with the reference allele first (or structural variant type)
- strand - defined as + (forward) or - (reverse)
- identifier - this identifier will be used the output. If not provided, Ensembl VEP will construct an identifier from the given coordinates and alleles.

More on input formats

here:([https://www.ensembl.org/info/docs/tools/vep/vep\\_formats.html#default](https://www.ensembl.org/info/docs/tools/vep/vep_formats.html#default))

**Species:** Homo\_sapiens

**Name for this job (optional):** Demo\_ECCB

**Input data:**

```
14 73219188 73219188 C/T
6 161785778 161785778 A/C
7 117480097 117480097 G/T
7 117587806 117587806 G/A
10 62813413 62813413 G/A
```

**Transcript database to use:**

Ensembl/GENCODE transcripts

HGNC symbols

HGVS nomenclature

8

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.

The screenshot shows a user interface for a bioinformatics tool. At the top, there is a file input field labeled "Choose file" with the placeholder "NO FILE CHOSEN" and an alternative input field "Or provide file URL:" with a text input box. Below this, under "Transcript database to use:", there is a group of radio buttons. The first option, "Ensembl/GENCODE transcripts", is selected and highlighted with a blue circle. The other options are "Ensembl/GENCODE basic transcripts", "Ensembl/GENCODE primary transcripts", "RefSeq transcripts", and "Ensembl/GENCODE and RefSeq transcripts". Under "Additional configurations:", there is a list of options with descriptive text and small icons. The first option, "Identifiers", is highlighted with an orange circle. The other options are "Variants and frequency data", "Additional annotations", "Predictions", "Filtering options", and "Advanced options". At the bottom right, there is a green "Run >" button.

Choose file NO file chosen  
Or provide file URL:

Transcript database to use:

Ensembl/GENCODE transcripts  
 Ensembl/GENCODE basic transcripts  
 Ensembl/GENCODE primary transcripts  
 RefSeq transcripts  
 Ensembl/GENCODE and RefSeq transcripts

Additional configurations:

**Identifiers** Additional identifiers for genes, transcripts and variants

**Variants and frequency data** Co-located variants and frequency data

**Additional annotations** Additional transcript, protein and regulatory annotations

**Predictions** Variant deleteriousness predictions, e.g. SIFT, PolyPhen

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

**Advanced options** Additional enhancements

Run >

9

You can choose different identifiers for the genes and proteins affected by your variant.

The screenshot shows the 'Identifiers' section of the Ensembl Variant Effect Predictor. It includes checkboxes for Gene symbol, Transcript version, CCDS, Protein (which is highlighted with an orange circle), UniProt, and HGVS. Below this is a 'Run' button. Further down, there are sections for 'Variants and frequency data', 'Additional annotations', 'Predictions', 'Filtering options', and 'Advanced options'.

**Identifiers**

Gene symbol:

Transcript version:

CCDS:

Protein:

UniProt:

HGVS:

**Variants and frequency data** Co-located variants and frequency data

**Additional annotations** Additional transcript, protein and regulatory annotations

**Predictions** Variant deleteriousness predictions, e.g. SIFT, PolyPhen

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

**Advanced options** Additional enhancements

Run >

**Recent jobs**

You have no jobs currently running or recently completed.

Ensembl release 114 - May 2025 © EMBL-EBI

10

By default, VEP will give you the allele frequency data from the 1000 Genomes project. You can also select for gnomAD (exomes) and (genomes) frequencies and data from AllOfUs.

The screenshot shows a configuration panel for VEP. On the left, there are several sections with labels and checkboxes or dropdown menus. On the right, there is a vertical stack of checkboxes for selecting frequency data sources. A red circle highlights the checkbox for 'AllOfUs allele frequencies'.

Find co-located known variants:	<input checked="" type="checkbox"/> Yes
Variant synonyms:	<input type="checkbox"/>
Frequency data for co-located variants:	<input checked="" type="checkbox"/> 1000 Genomes global minor allele frequency <input type="checkbox"/> 1000 Genomes continental allele frequencies <input type="checkbox"/> gnomAD (exomes) allele frequencies <input type="checkbox"/> gnomAD (genomes) allele frequencies <input checked="" type="checkbox"/> AllOfUs allele frequencies
PubMed IDs for citations of co-located variants:	<input checked="" type="checkbox"/>
Include flagged variants:	<input type="checkbox"/>
Paralogue variants:	<input checked="" type="radio"/> Disabled <input type="radio"/> Enabled
Open Targets Genetics:	<input type="checkbox"/>

**11** Click "Additional annotations".

VEP can report if the variant has hit a MANE transcript. MANE is the Matched Annotation from the NCBI and EMBL-EBI. This is a collaborative project that aims to converge on human gene and transcript annotation and to define a genome-wide set of representative transcripts and corresponding proteins for human protein-coding genes.

The screenshot shows the Variant Effect Predictor (VEP) interface. At the top, there are several configuration options:

- PubMed IDs for citations of co-located variants:**
- Include flagged variants:**
- Parologue variants:**  Disabled  Enabled
- Open Targets Genetics:**

Below these are four buttons:

- Additional annotations** (highlighted with an orange circle)
- Predictions**
- Filtering options**
- Advanced options**

A green button labeled **Run >** is located at the bottom right of the main panel. Below the main panel, there is a section titled **Recent jobs** with the note: "You have no jobs currently running or recently completed." A small footer at the bottom left reads "Ensembl release 104 - May 2025 © EMBL-EBI".

12

Select "Protein matches" which will give a lists of matching protein domains and families.

Upstream/Downstream distance (bp): 5000

miRNA structure:

NMD:

UTRAnnotator:

RiboseqORFs:

**Protein annotation**

Protein matches:   

mutfunc:  Disabled  Enabled

**Functional effect**

IntAct:  Disabled  Enabled

MaveDB:

13

VEP can provide pathogenicity prediction from SIFT, PolyPhen, CADD and AlphaMissense to name a few.

**Predictions** Variant deleteriousness predictions, e.g. SIFT, PolyPhen

**Pathogenicity predictions**

SIFT: Prediction and score

PolyPhen: Prediction and score

dbNSFP:  Disabled  Enabled

AlphaMissense:   

CADD:  Disabled  Enabled

REVEL:

ClinPred:

EVE:

**Splicing predictions**

14

VEP allows you to pre-filter your results, e.g. by frequency. Note that it is also possible to perform equivalent operations on the results page for VEP, so if you aren't sure, don't use any of these options!

dbSCNV:

MaxEntScan:

SpliceAI:  Disabled  Enabled

**Conservation**

BLOSUM62:

Ancestral allele:

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

**Advanced options** Additional enhancements

**Run >**

**Recent jobs** Recent jobs

You have no jobs currently running or recently completed.

15

You can hover your mouse over the options to see definitions.

We have selected the following configurations:

- Uniprot
- Ensembl Protein
- HGVS (annotation of variants in terms of the transcripts and proteins they affect, commonly-used by the clinical community)
- Protein matches
- Phenotypes

When you have selected everything you need, scroll right to the bottom and click "Run".

The screenshot shows a configuration interface for a bioinformatics tool. At the top, there are three checkboxes: dbscSNV (unchecked), MaxEntScan (unchecked), and SpliceAI (radio button set to 'Disabled'). Below this is a section titled 'Conservation' containing two checkboxes: BLOSUM62 (unchecked) and Ancestral allele (unchecked). At the bottom of the main configuration area are two buttons: 'Filtering options' and 'Advanced options'. A large green button labeled 'Run' is positioned at the very bottom. The entire configuration area is highlighted with a light gray background. A yellow circle is drawn around the 'Run' button, indicating it is the next step. Below the configuration area, there is a section titled 'Recent jobs' with a note stating 'You have no jobs currently running or recently completed.'

## 16 The display will show you the status of your job.

It will say Queued, then automatically switch to Done when the job is complete; 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 completed.

The screenshot shows the Ensembl Variant Effect Predictor (VEP) interface. On the left, there's a sidebar with 'Web Tools' and a list including 'Variant Effect Predictor'. Below the sidebar are buttons for 'Configure this page', 'Custom tracks', 'Export data', 'Share this page', and 'Bookmark this page'. The main area is titled 'Variant Effect Predictor' with a 'New job' button. Under 'Recent jobs', there's a 'Refresh' button. A table lists a single job: 'Variant Effect Predictor' for 'Demo\_ECCB' in 'Homo\_sapiens'. The status is 'Done' (highlighted with a yellow circle). There are buttons for 'View results', 'Edit', and 'Delete'. A message at the bottom says 'This job is finished.' At the bottom of the page, it says 'Ensembl release 114 - May 2025 © EMBL-EBI' and has sections for 'About Us', 'Get help', 'Our sister sites', and 'Follow us'.

Analysis	Jobs	Submitted at (GMT)
Variant Effect Predictor	VEP analysis of Demo_ECCB in Homo_sapiens <b>Done</b> <a href="#">View results</a>	24/06/2025, 15:57

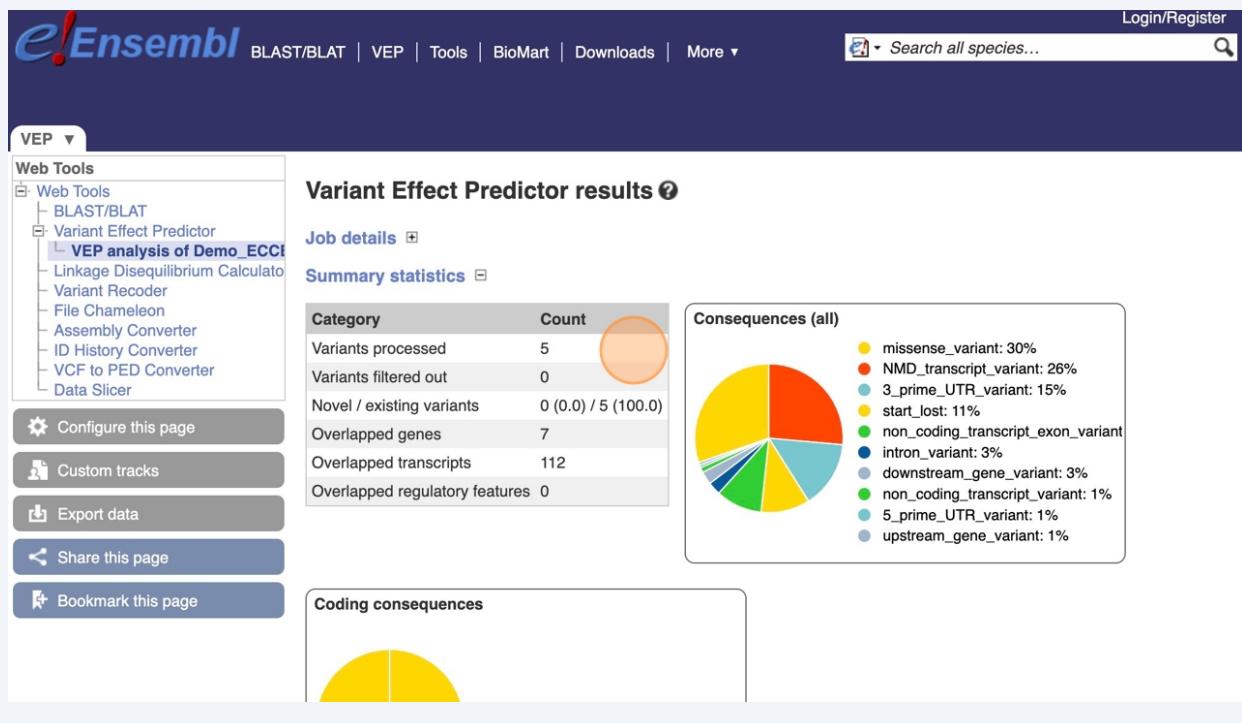
This job is finished.

About Us  
Get help  
Our sister sites  
Follow us

Contact us  
Citing Ensembl  
Privacy policy  
Using this website  
Adding custom tracks  
Downloading data  
Video tutorials  
Ensembl Bacteria  
Ensembl Fungi  
Ensembl Plants  
Ensembl Protists  
Blog  
Twitter  
Facebook

17

In your results, you will see a graphical summary of your data, as well as a detailed table of your results.



18

The first column is Uploaded variant. If your input data contains IDs, like ours does, the ID is listed here. If your input data is only loci, this column will contain the locus and alleles of the variant. You will notice that the variants are not necessarily in the order they were in your input.

You will also see that there are multiple lines in the table for each variant, with each line representing one transcript or other feature the variant affects.

You can mouse-over any column name to get a definition of what is shown.

The next few columns give the information about the feature the variant affects, including the consequence. Where the feature is a transcript, you will see the gene symbol and stable ID and the transcript stable ID and biotype. Where the feature is a regulatory feature, you will get the stable ID and type.



Uploaded variant	Location	Allele	Consequence	Impact	Symbol	Gene
10_62813413_G/A	<a href="#">10:62813413-</a>	A <a href="#">62813413</a>	missense_variant	MODERATE	EGR2	<a href="#">ENSG0000012287</a>
10_62813413_G/A	<a href="#">10:62813413-</a>	A <a href="#">62813413</a>	missense_variant	MODERATE	EGR2	<a href="#">ENSG0000012287</a>
14_73219188_C/T	<a href="#">14:73219188-</a>	T <a href="#">73219188</a>	missense_variant	MODERATE	PSEN1	<a href="#">ENSG0000008081</a>
14_73219188_C/T	<a href="#">14:73219188-</a>	T <a href="#">73219188</a>	missense_variant	MODERATE	PSEN1	<a href="#">ENSG0000008081</a>
14_73219188_C/T	<a href="#">14:73219188-</a>	T <a href="#">73219188</a>	missense_variant	MODERATE	PSEN1	<a href="#">ENSG0000008081</a>
14_73219188_C/T	<a href="#">14:73219188-</a>	T <a href="#">73219188</a>	missense_variant	MODERATE	PSEN1	<a href="#">ENSG0000008081</a>
14_73219188_C/T	<a href="#">14:73219188-</a>	T <a href="#">73219188</a>	missense_variant	MODERATE	PSEN1	<a href="#">ENSG0000008081</a>

19

The table shows details on the effects on transcripts, including the position of the variant in terms of the exon number, cDNA, CDS and protein, the amino acid and codon change, transcript flags, which can be used to choose a single transcript for variant reporting, and pathogenicity scores.

The pathogenicity scores are shown as numbers with coloured highlights to indicate the prediction, and you can mouse-over the scores to get the prediction in words.

		Download	New job								
		All: <a href="#">VCF</a> <a href="#">VEP</a> <a href="#">TXT</a>	BioMart: <a href="#">Variants</a> <a href="#">Genes</a>								
		Show/hide columns (31 hidden)									
UNIPROT ISOFORM	Transcript source	SIFT	PolyPhen	Protein matches	HGVSc offset	AF	Clinical signi				
<a href="#">307</a> <a href="#">P11161-1</a>	-	<span style="background-color: red; border: 1px solid black; padding: 2px;">0</span>	<span style="background-color: red; border: 1px solid black; padding: 2px;">0.996</span>	Alphafold model 8 Protein matches ● AFDB-ENSP_mappings: AF-P11161-F1	-	-	-	pathogenic, _uncertain_si pathogenic			
<a href="#">307</a> <a href="#">P11161-1</a>	-	<span style="background-color: red; border: 1px solid black; padding: 2px;">0</span>	<span style="background-color: red; border: 1px solid black; padding: 2px;">0.996</span>	Alphafold model 8 Protein matches	-	-	pathogenic, _uncertain_si pathogenic				
<a href="#">15F</a> <a href="#">P49768-1</a>	-	<span style="background-color: red; border: 1px solid black; padding: 2px;">0</span>	-	Protein Structure View Alphafold model 34 Protein matches	-	-	not_provided				
<a href="#">15F</a> <a href="#">P49768-1</a>	-	<span style="background-color: red; border: 1px solid black; padding: 2px;">0</span>	-	Protein Structure View	-	-	not_provided				

20

The Protein matches list the protein domains found, and where available, provides a link to the 3D protein model, which will launch a LiteMol viewer, highlighting the variant position.

Expand Protein matches to see the variant on PDBe structures or AlphaFold model to see the variant on a predicted structure.

The screenshot shows the VEP tool interface with the following details:

- Navigation (per variant):** Show: 1 5 All variants
- Filters:** Uploaded variant is defined
- Download:** VCF VEP TXT
- BioMart:** Variants Genes
- Table Headers:** IFT, PolyPhen, Protein matches, HGVS offset, AF, Clinical significance, Phenotype or disease
- Row 1 (highlighted):**
  - IFT: 0
  - PolyPhen: 0.996
  - Protein matches: Alphafold model (highlighted)
  - HGVS offset: -
  - AF: -
  - Clinical significance: pathogenic, uncertain\_significance, pathogenic
  - Phenotype or disease: 1, 1
- Row 2 (highlighted):**
  - IFT: 0
  - PolyPhen: 0.996
  - Protein matches: Alphafold model (highlighted)
  - HGVS offset: -
  - AF: -
  - Clinical significance: pathogenic, uncertain\_significance, pathogenic
  - Phenotype or disease: 1, 1
- Row 3 (highlighted):**
  - IFT: 0
  - PolyPhen: -
  - Protein matches: Protein Structure View (highlighted)
  - HGVS offset: -
  - AF: -
  - Clinical significance: not\_provided
  - Phenotype or disease: 1, 1

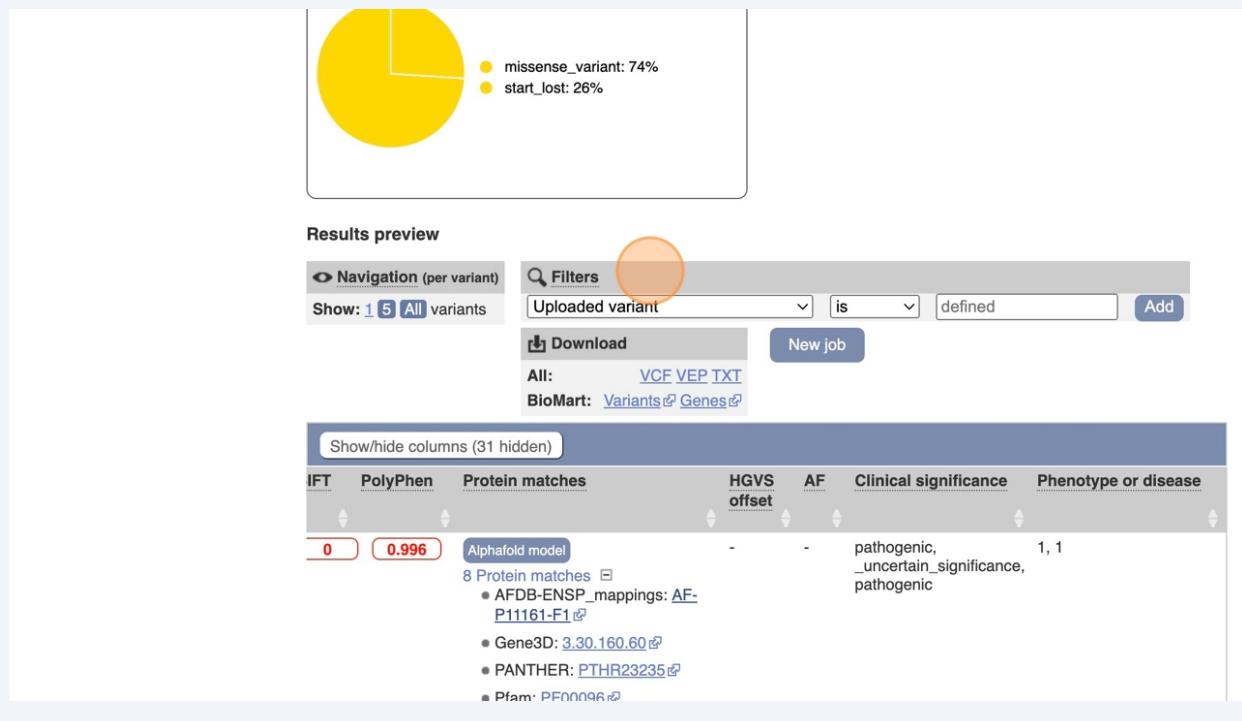
A callout bubble highlights the "Protein matches" section for the first row, showing a list of domain mappings:

- AFDB-ENSP\_mappings: AF-P11161-F1
- Gene3D: 3.30.160.60
- PANTHER: PTHR23235
- Pfam: PF00096
- PROSITE\_patterns: PS00028
- PROSITE\_profiles: PS50157
- SMART: SM00355
- Superfamily: SSF57667

21

Above the table is the Filter option, which allows you to filter by any column in the table.

You can select a column from the drop-down, then a logic option from the next drop-down, and then type in your filter in the following box. For example, a filter of "consequence is missense" will give only missense variant consequences, which might change protein effectiveness. You'll notice that as you type missense, the VEP will make suggestions for an autocomplete.



22

You can export your VEP results in various formats, including VCF. When you export as VCF, you'll get all the VEP annotations listed under CSQ in the INFO column.

After filtering your data, you'll see that you have the option to export only the filtered data. You can also drop all the genes you've found into the Gene BioMart, or all the known variants into the Variation BioMart to export further information about them.

