Variant Effect Predictor

Demo: The Variant Effect Predictor (VEP)

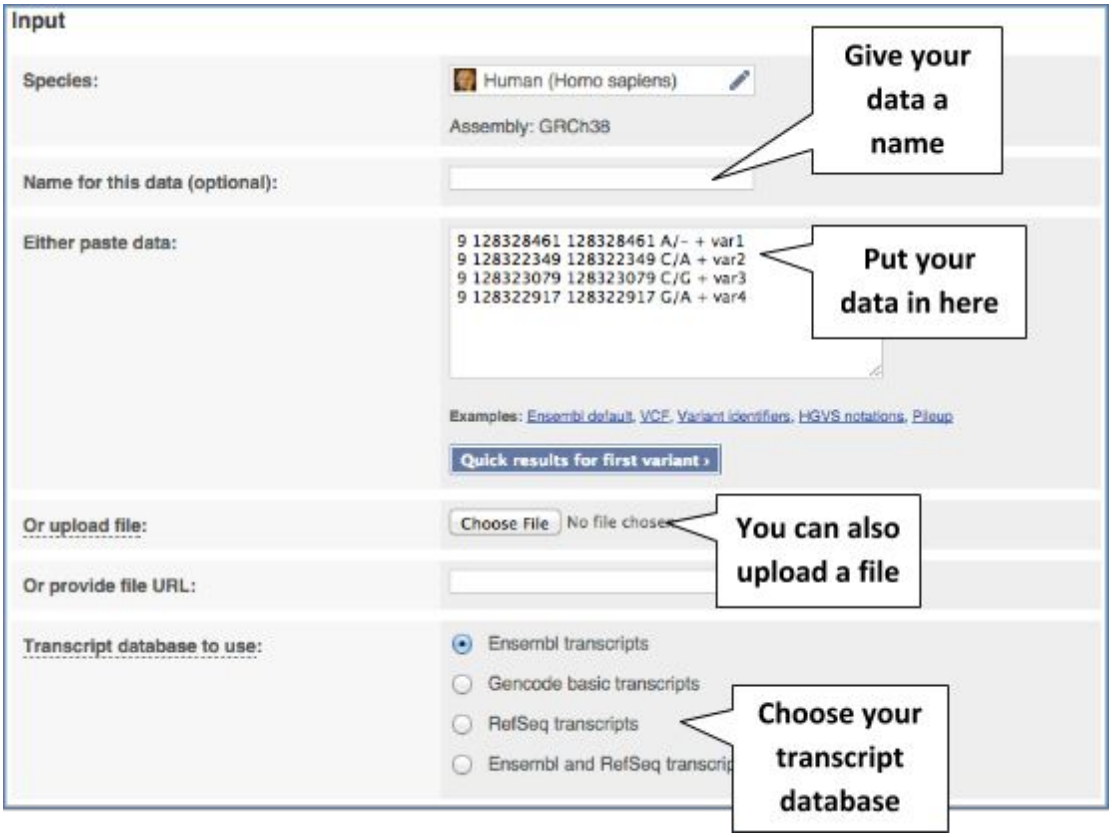
We have identified five variants on human chromosome nine, an A deletion at 128328461, C->A at 128322349, C->G at 128323079, G->A at 128322917 and A->G at 128322495.

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.

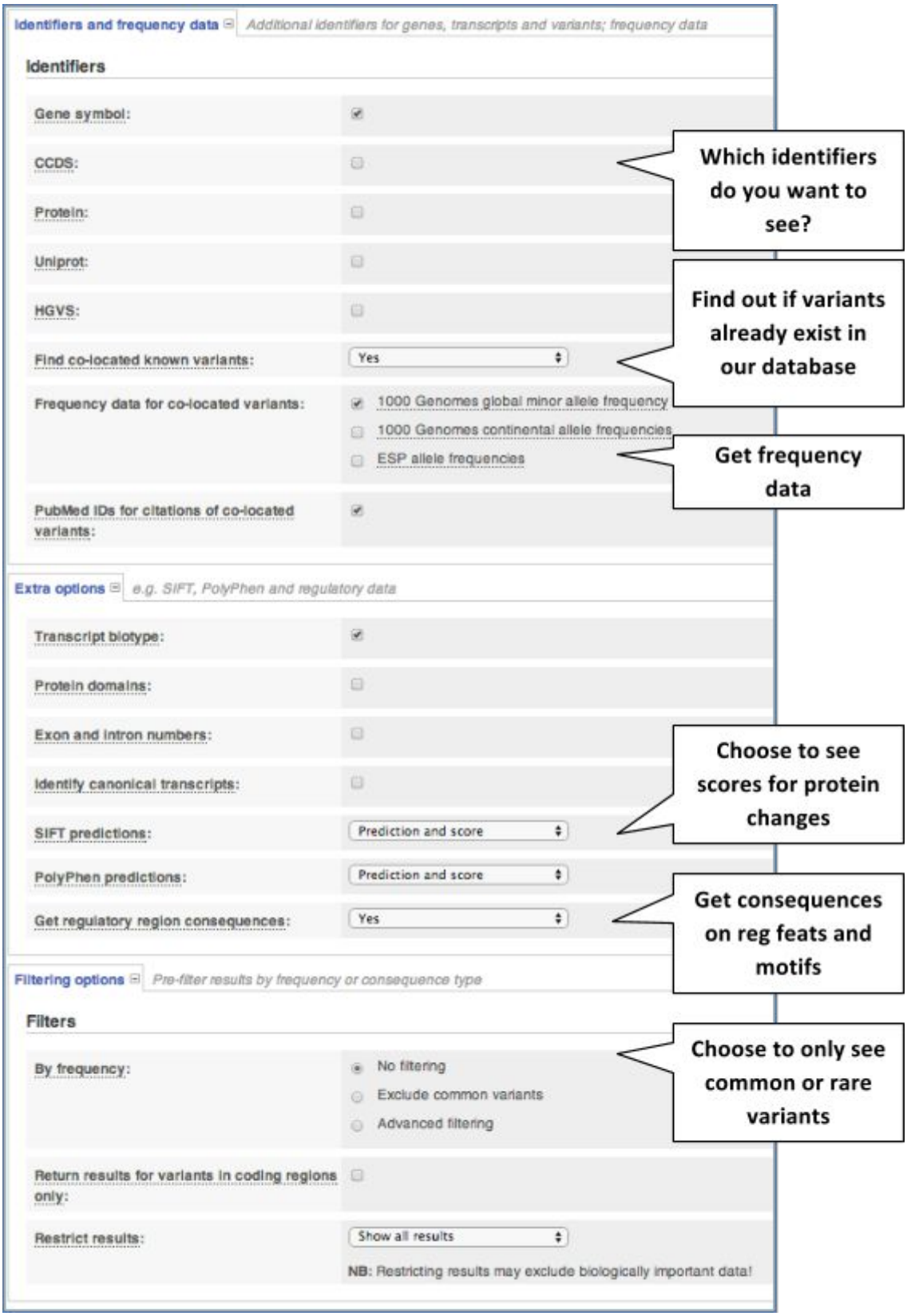


The data is in the format:
Chromosome Start End alleles (ref/alt) strand name

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.



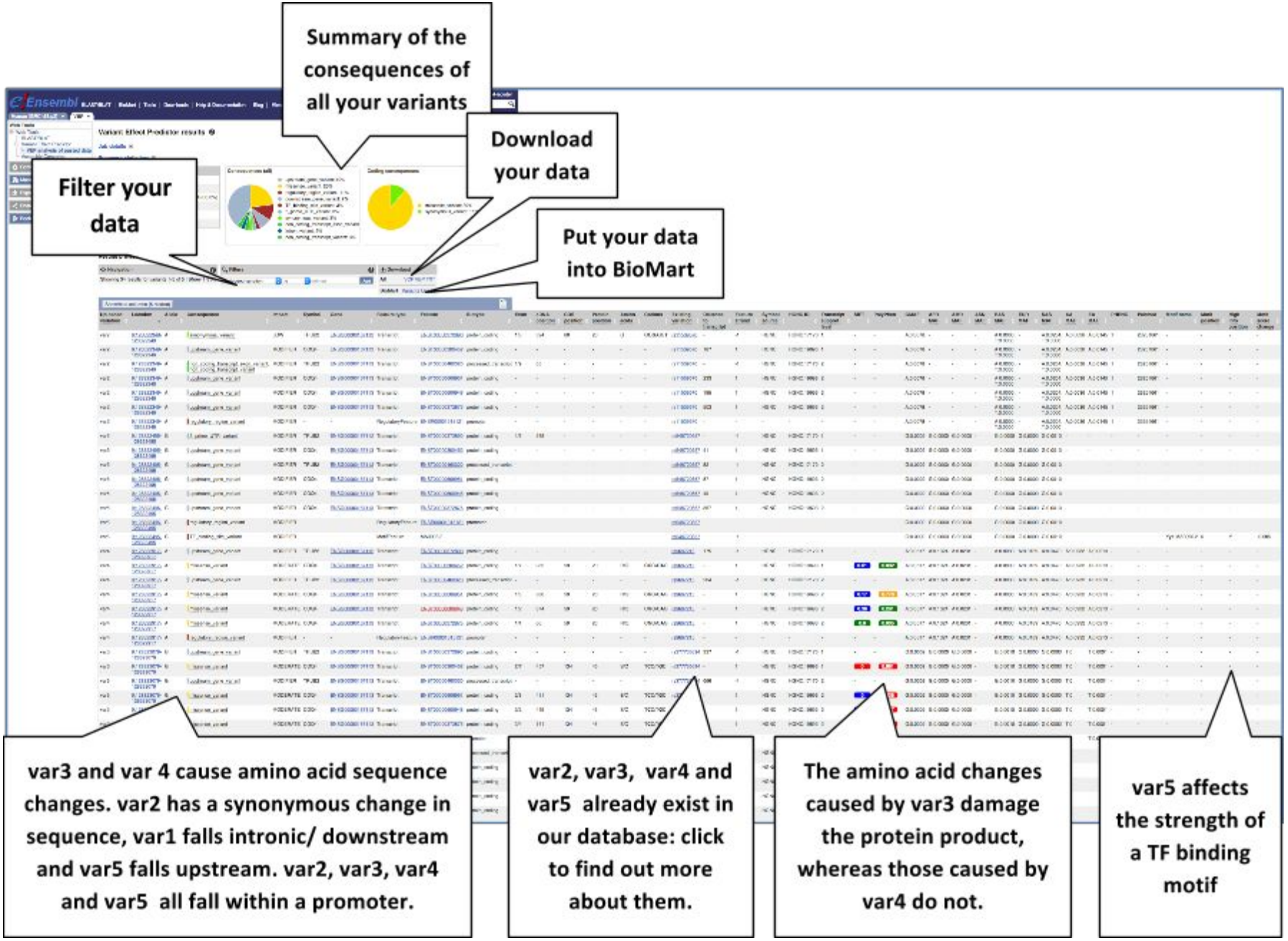
Many of the options have a question mark **[?]** after the names.
Hover over the question mark **[?]** to see definitions.

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.)


Summary of the consequences of all your variants
Download your data
Filter your data
Put your data into BioMart
var3 and var 4 cause amino acid sequence changes. var2 has a synonymous change in sequence, var1 falls intronic/ downstream and var5 falls upstream. var2, var3, var4 and var5 all fall within a promoter.
var2, var3, var4 and var5 already exist in our database: click to find out more about them.
The amino acid changes caused by var3 damage the protein product, whereas those caused by var4 do not.
var5 affects the strength of a TF binding motif

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.