ENSEMBL TOOLS: THE VEP (course booklet pages 53-54)

Using VEP (the Variant Effect Predictor) to annotate SNPs and SVs.

Answer 1 – Using VEP to predict the consequence of SNPs on the previous human assembly, GRCh37

Go to the Variant Effect Predictor (VEP) under 'Tools' on the top banner of archive Ensembl page containing GRCh37, i.e. grch37.ensembl.org

Paste the data below in the text box:

15 78889339 78889339 G/A
22 30332586 30332586 T/C
6 31721033 31721033 G/A
5 1260624 1260624 G/A
17 63554591 63554591 A/G
5 1254510 1254510 C/T

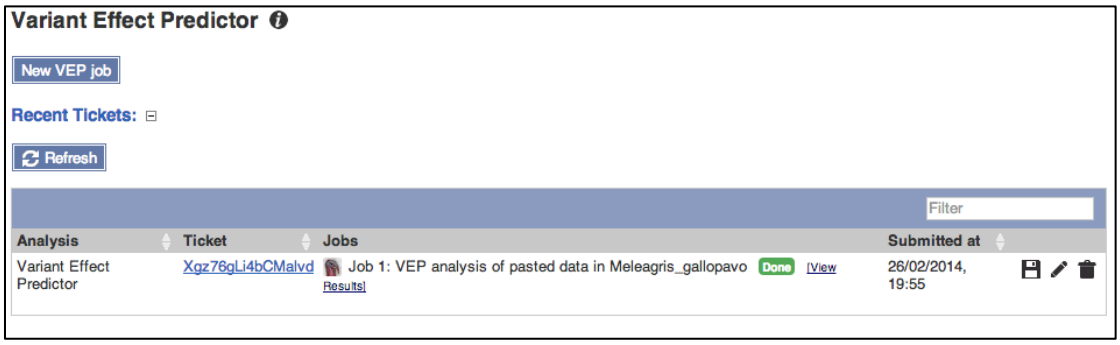
Note that this is the Ensembl default format (chr/start/end/possible alleles). For additional formats accepted by VEP, have a look here:

http://www.ensembl.org/info/docs/tools/vep/vep_formats.html

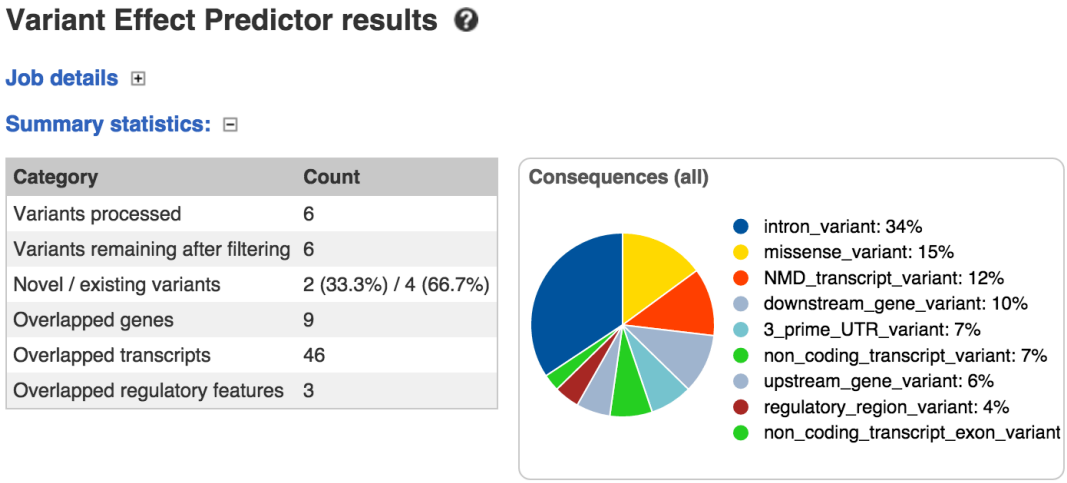
Make sure that 'SIFT predictions' and 'PolyPhen predictions' in 'Extra options' under 'Output options' are selected. Note there are two other tools available through the VEP that can give you predictions on how pathogenic a missense variant might be.

Click on 'Run'.

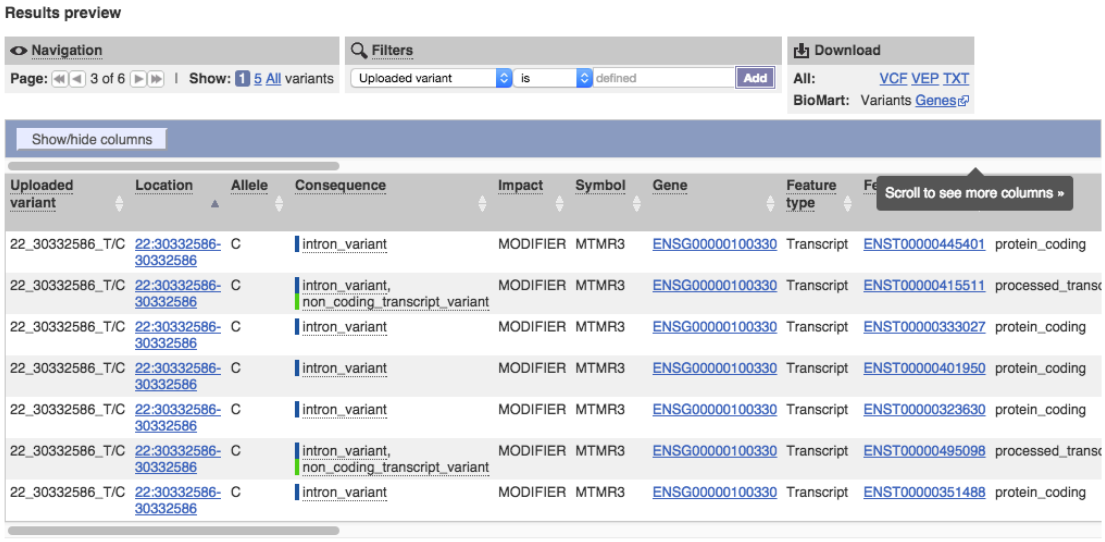
Once the job is done you will see a table similar to this:



Click on 'View' to jump to a page similar to the one below, with a summary statistics and pie charts for quick inspection of the VEP results.



a) Scroll down to view a table listing the different genes/transcript that the six variants map to.



The listed variants map to the several genes and transcripts.

b) The consequence terms can be several such as intron variant, upstream gene variant, missense variant, downstream gene variant, NMD transcript variant, among others.

c) Prediction of pathogenicity from SIFT and PolyPhen are available for missense variants in human. For other species such as mouse, SIFT is the only prediction available.

From the table, you can see that some missense variants will cause a change at the amino acid level that is predicted to be deleterious or probably damaging i.e. rs199422305. The two predictions tools take into account different parameters and algorithms, so discrepancies can indeed be seen (see rs121918664).

For more details on PolyPhen and SIFT have a look at the URLs below:

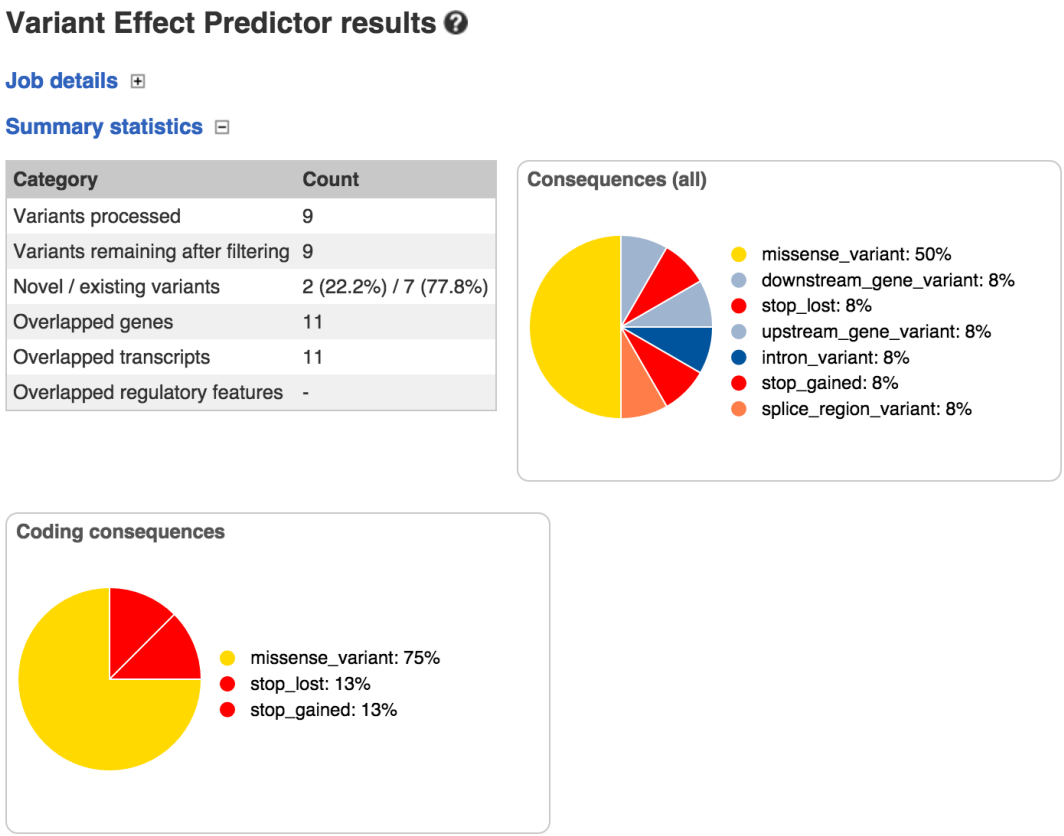
http://sift.jcvi.org/
http://genetics.bwh.harvard.edu/pph2/

Answer 2 – The VEP tool and variants on the bread wheat genome.

Go the plants.ensembl.org and click on Tools on the top banner of the page. Click on VEP and select the species *T. aestivum.* Type in the following input data:

2D 89551917 89551917 G/A
2D 148408765 148408765 G/T
3D 113574123 113574123 C/A
3D 93827883 93827883 G/A
3B 727928129 727928129 C/T
3B 736734474 736734474 C/T
6A 196872409 196872409 T/G
6A 196153918 196153918 A/G
6A 196774882 196774882 G/C

Scroll down in that page and click ‘Run’. When the job is complete (it may take a few seconds), click on ‘View results’.



a) Two variants are completely novel and therefore not known in public databases. The known variants can be found under the column ‘Existing variant’ in the Results table.

b) Several genes and transcripts have been annotated in the region where these variants map to such as Traes_2DL_776B9B786 and Traes_3DL_661D28EB1.

c) The consequences for these variants are missense, splice region, stop loss, downstream gene, among others. The missense variants do cause a change at the amino acid level such as variants vcZ2NV8SV and vcZ2NXCA5.

Answer 3 – VEP in Bacterial genomes

Go to bacteria.ensembl.org and click on tools to get to the page where the VEP (Variant Effect Predictor) among other tools is listed.

Type in '638R' into the species search box, select the full name of the species from the drop down menu and enter the following data into the input box:

FQ312004 4945507 4945510 TGCA/-

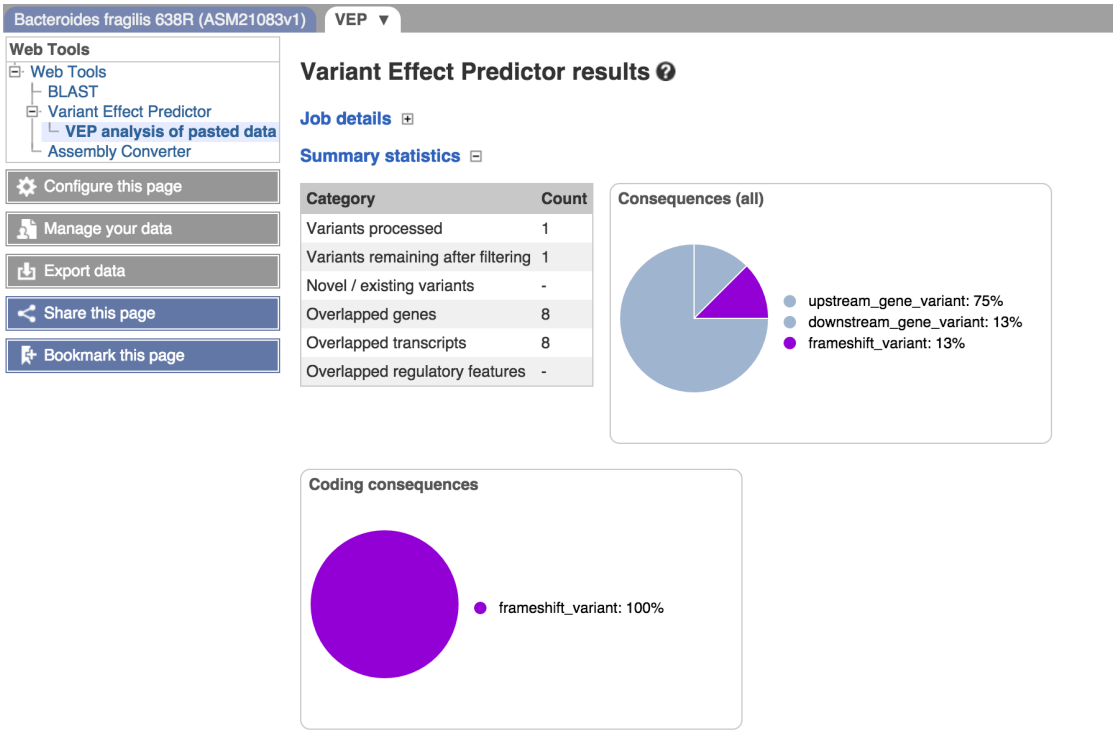
The alternative allele is a deletion (-) compared to the reference allele (TGCA).

For the different input formats VEP can accept have a look at the help below:

http://www.ensembl.org/info/docs/tools/vep/vep_formats.html#input

Click 'Run'.

You will be able to see the different consequences of this deletion on this bacterial genome.



Note that you've entered one variant but it overlaps eight genes such as BF638R_4208. For each of them, the consequence may be different. The only coding consequence is a 'frameshift_variant'.

Hover over the consequence name in the table to find what it means: