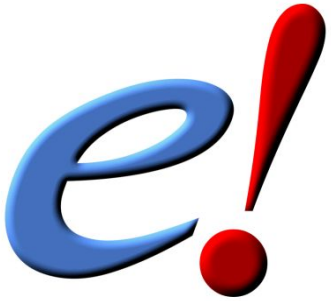
Browsing Genomes with Ensembl



www.ensembl.org
www.ensemblgenomes.org

Exercise Answers v93

http://training.ensembl.org/events/2018/2018-08-27-VEPTC

Johor – 27th-30th August 2018

Exercise answers

Exercise 1 – VEP using variant coordinates

Go to www.ensembl.org and click on the link tools at the top of the page.
Click on Variant Effect Predictor and enter the three variants as below:
7 117530985 117530985 G/A
7 117531038 117531038 T/C
7 117531068 117531068 T/C

Note: Variation data input can be done in a variety of formats. See more details here
http://www.ensembl.org/info/docs/variation/vep/vep_formats.html

Click Run.

When your job is listed as Done, click View Results.

You will get a table with the consequence terms from the Sequence Ontology project (http://www.sequenceontology.org/) (i.e. synonymous, missense, downstream, intronic, 5' UTR, 3' UTR, etc) provided by VEP for the listed SNPs. You can also upload the VEP results as a track and view them on Location pages in Ensembl. SIFT and PolyPhen are available for missense SNPs only. For two of the entered positions, the variations have been predicted to be probably damaging/deleterious (coordinate 117531038) and benign/tolerated (coordinate 117531068). All the three variations have been already described and are known as in rs1800078, rs1800077 and rs35516286 in dbSNP and other sources (databases, literature, etc).

Exercise 2 – viewing structural variants with the VEP

(a) Give your data a name, such as Patient deletion.
Paste 13 32307062 sv1 . <DEL> . . SVTYPE=DEL;END=32908738
into the Paste data field.
Tick Show gene symbols (e.g. HGNC) where available.
Hit Next, then choose HTML.
The HGNC identifiers for the affected genes are *BRCA2* and *ZAR1L*.

(b) Look at the consequence column of the result page. Complete deletions are referred to as transcript_ablation. This consequence is not found.

(c) To view your variant in the browser click on the location link in the results table 13: 32307062-32908738.
The link will open the Region in detail view in a new tab. If you have given your data a name it will appear automatically in red. If not, you may need to Configure this page and add it under Your data.

Exercise 3 – uploading a VCF file for VEP analysis

Go to www.ensembl.org and click on the link tools at the top of the page. Click on Variant Effect Predictor. Enter the name and location of the file into the 'Or provide file URL' field:

http://ftp.ebi.ac.uk/pub/databases/ensembl/training/2018/VEPTC_2018/VEPTC_VCF.vcf

Click Run.

When your job is listed as Done, click View Results.

(a) The summary results at the top of the page tells us that 20 variants were processed in the analysis. Twelve are 'known'. Eight are 'novel'.

(b) Use the results filter found just above the results table. Select 'Phenotype or disease is 1'. Click Add. All variants with a known phenotype association will be displayed, denoted by a '1' in the 'Phenotype or Disease' column of the results table.

Click on each of the individual IDs from the 'existing variant' column from the results table. In the variant tab, click on Phenotype data.

rs10925500 and rs100562 are associated with Alzheimer disease

(c) Use the results filter found just above the results table. Remove the filter applied to answer (b). Select 'Consequence is Regulatory_region_variant'. Click Add.

Four variants are predicted to have the Regulatory_region_variant consequence. Rs943769928 affects a CTCF binding site, while rs567497980, rs1716784 and a novel variant affect promoter regions.

The three promoter regions affected by the variants are: ENSR00000220033, ENSR00000110117 and ENSR00000248250.

Click on the stable ID for the promoter in the results table. This will take you to the Regulation tab. Click on the gene (gold feature) in the figure to open a pop-up window with the gene name and further information.

These promoters overlap the PAXIP1, APOE and WDR44 genes, respectively.

The variants affect three separate TF binding motif features from two of these promoters. These TF binding motif are: MA0139.1, MA0162.2 and PB0010.1.

(d) Use the results filter found just above the results table. Remove the filter applied to answer (c). Select 'Consequence is missense_variant'. Click Add.

HFE and BLMH are affected by two variants from the input list of variants.

Click on the stable ID for both HFE and BLMH. This will take you to the gene tab. Click Phenotype from the menu on the left hand side.

HFE is associated with Alzheimer disease and various hemochromatosis and porphyria disorders.

BLMH is associated with Alzheimer disease and resistance to bleomycine treatment.

(e) Use the results filter found just above the results table. Remove the filter applied to answer (d). Select 'Consequence is stop_lost'. Click Add.

A single variant has a predicted stop_lost consequence, which affects the APBB2 gene.