How to use Variant Effects Report

A. Introduction to Ensembl Variant Effect Predictor
B. Using RefSeq_v1
C. Using TGACv1

A. Introduction

The Ensembl Variant Effect Predictor is a toolset for the analysis, annotation, and prioritization of genomic variants in coding and non-coding regions. There are more than 1 million single nucleotide variants in wheat. There may be only 10 thousand that change the amino acid coding and a smaller subset of these that truncate or produce a loss of function.

The Ensembl Variant Effect Predictor (VEP) can be accessed by web, Perl script, web based API. The report page shows calculations for markers in T3 and provides links to calculation provided by Ensembl Plant

The inputs used for the VEP are
   a. Ensembl or VCF format SNP locations
   b. High confidence (HC) gene predictions
   c. FASTA file of the assembly

The outputs of the VEP are
   a. Feature – transcript (can also include motif and regulatory elements)
   b. Consequence
   c. Impact

A detailed description of the Consequence and Impact values can be found here: http://ensembl.org/info/genome/variation/predicted_data.html

The Ensembl VEP also incorporates SIFT and PolyPhen-2 but these are not used on the T3 website.

**Sorting Intolerant From Tolerant (SIFT)** predicts whether an amino acid substitution is likely to affect protein function based on sequence homology and the physico-chemical similarity between the alternate amino acids.

<table>
<thead>
<tr>
<th>SIFT value</th>
<th>Qualitative prediction</th>
<th>Website display example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 0.05</td>
<td>&quot;Deleterious&quot;</td>
<td>0.01 0.01</td>
</tr>
<tr>
<td>Greater than or equal to 0.05</td>
<td>&quot;Tolerated&quot;</td>
<td>0.8 0.8</td>
</tr>
</tbody>
</table>

**PolyPhen-2** predicts the effect of an amino acid substitution on the structure and function of a protein using sequence homology, Pfam annotations, 3D structures from PDB where available, and a number of other databases and tools (including DSSP, ncoils etc.)
Consequence values for VEP

3_prime_UTR_variant
5_prime_UTR_variant
coding_sequence_variant
coding_sequence_variant,3_prime_UTR_variant
coding_sequence_variant,5_prime_UTR_variant
downstream_gene_variant
frameshift_variant
frameshift_variant,splice_region_variant
frameshift_variant,splice_region_variant,intron_variant
frameshift_variant,start_lost
frameshift_variant,start_lost,splice_region_variant
frameshift_variant,start_lost,start_retained_variant
frameshift_variant,stop_lost
frameshift_variant,stop_lost,splice_region_variant
frameshift_variant,stop_retained_variant
inframe_deletion
inframe_deletion,splice_region_variant
inframe_insertion
inframe_insertion,splice_region_variant
inframe_insertion,stop_retained_variant
intergenic_variant
intron_variant
protein_altering_variant
protein_altering_variant,splice_region_variant
splice_acceptor_variant
splice_acceptor_variant,5_prime_UTR_variant
splice_acceptor_variant,coding_sequence_variant
splice_acceptor_variant,coding_sequence_variant,intron_variant
splice_acceptor_variant,frameshift_variant
splice_acceptor_variant,inframe_insertion
splice_acceptor_variant,intron_variant
splice_donor_variant
splice_donor_variant,5_prime_UTR_variant
splice_donor_variant,coding_sequence_variant
splice_donor_variant,coding_sequence_variant,intron_variant
splice_donor_variant,coding_sequence_variant,5_prime_UTR_variant
splice_donor_variant,frameshift_variant
splice_donor_variant,inframe_insertion
splice_donor_variant,intron_variant
splice_region_variant
splice_region_variant,3_prime_UTR_variant
splice_region_variant,5_prime_UTR_variant
splice_region_variant,intron_variant
start_lost
start_lost,3_prime_UTR_variant
start_lost,inframe_deletion
start_retained_variant
start_retained_variant,5_prime_UTR_variant
stop_gained
stop_gained,frameshift_variant
stop_gained,frameshift_variant,splice_region_variant
stop_gained,inframe_deletion
stop_gained,inframe_insertion
stop_gained,inframe_insertion,splice_region_variant
stop_gained,splice_region_variant
stop_lost,3_prime_UTR_variant
stop_lost,inframe_deletion
stop_retained_variant
upstream_gene_variant
Impact values for VEP
HIGH - disruptive impact in the protein, protein truncation or loss of function
LOW – harmless, unlikely to change protein behaviour
MODERATE - non-disruptive variant that might change protein effectiveness
MODIFIER – usually non-coding variants

General instructions for to use the T3 VEP report
First select a list of markers (limit the selection to under 1000). It will accept a mix of markers from different genotype experiments or a single genotype experiment. The positions of the markers on the genome assembly have been identified either by BLAST or from the coordinates provided when the genotype results were loaded into the database. If the marker position cannot be identified then it will be listed at the bottom of the page as not found.

For markers not found on the map
You can run BLAST against RefSeq and format the output in either Ensembl or VCF format then mail the file to me using the feedback link on the T3 website. Then I can run the Ensembl VEP program on our machine and email you the results.
B. Using RefSeq_v1 assembly – the only markers that have been mapped to RefSeq_v1 are in the RefSeq v1.0 Physical Map and the 2017_WheatCAP genotype experiment.

1. Visit https://triticeaetoolbox.org/wheat/

2. Select markers of interest. Go to Select => Markers
3. View the Variant Effects: Go to Reports => Variant Effects

Selecting the link in the Gene column gives you a report of all markers for that gene.
Selecting the link in the region column gives you a JBrowse view for that marker.
C. Using TGACv1 assembly

1. Select markers of interest. Go to Select => Markers
2. View the Variant Effects: Go to Reports => Variant Effects
3. Select TGACv1 for the Genome Assembly
4. To view the Variant Effect, you can either
   a. Click on the link in the Gene column. This will take you to a table on Ensembl Plant website.
   b. Scroll down the page until you see the second table. Copy these entries and paste them in the data field of the Variant Effect Predictor on the Ensembl Plant website.
Selecting the link in genes column directs you to Ensembl Plant to show you a table of variants for that gene.
On the Ensembl page there is a link to view the location in the Ensembl Browser, which will show you the position and variant types for that gene.
You can run the Ensembl VEP program for your own markers by copying the results from the bottom table of the T3 Variant Effects page and pasting it into the tool on the Ensembl Plant website. http://plants.ensembl.org/Triticum_aestivum/Tools/VEP?db=core