Annotation

The directory contains two folders: one named "vep" and the other named "vcfanno".

Vcf file : test1.vcf

VEP

The Test1.vcf file was processed using VEP on a Windows system because my server couldn't support VEP annotation, leading to a failure midway through the process.

Run on the basis of this : ./vep --af --af_gnomade --af_gnomadg --appris --biotype --buffer_size 500 --check_existing --distance 5000 --mane --plugin GO,[path_to]/ --polyphen b --pubmed --regulatory --show_ref_allele --sift b --species homo_sapiens --symbol --transcript_version --tsl --uploaded_allele --cache --input_file [input_data] --output_file [output_file]

Output file : New_vep.vcf and VEP_result_ENSEMBL.txt

Analysis :

i. Python code file name : 1_filter_vep.py
   Input file : VEP_result_ENSEMBL.txt
   Output file : 1.tsv

This code removes the columns that are not used for reporting purposes.

ii. Python code file name : 2_QC_VEP.py
   Input file : 1.tsv
   Output file : 2.tsv

This code filters the Consequence column to provide missense variants and also filters the CLIN_SIG column to provide variants according to our ACMG guidelines, such as pathogenic, uncertain significance, etc.

iii. Python code file name : 3_Final_variant.py
   Input file : 2.tsv
   Output file : 3.tsv

This code generates the final set of variants for reporting by filtering out rows with missing values and allele frequencies greater than 0.05.

vcfAnno

Vcf file Given by : test2.vcf

Run by linux

./vcfanno -lua /home/akkey/Akshay/wellytics/test/clinvar/vcfanno/example/clinvar.lua conf.toml test2.vcf > 1.vcf

Output file after run vcfanno : 1.vcf

The folder contains the Clinvar.lua and conf.toml files.

iv. Python code file name : vcfanno_filter_clinvar.py
    Input file : 1.vcf
    Output file : akj_vcfanno.tsv

This code specifically provides ClinVar details, although many blank entries are present. Therefore, when opening the file in Excel, it's advisable to utilize the filter to examine the ClinVar results. Additionally, information from gnomAD and dbSNP is also included, but it hasn't been filtered out as we've already obtained it from VEP.