Explanation of fields in variant calling files “SampleID_germline.annot.xlsx” (v.1.0)

Column name	Description	Tool
CHROM	Chromosome or scaffold name (same as in reference genome used for alignment)	GATK
POS	Nucleotide position on the chromosome (start position for insertions/deletions), 1-based	GATK
REF	Reference sequence at the position (sequence of the reference genome used for analysis)	GATK
ALT	Alternative sequence at this position (sequence found in the sequencing reads of the sample); e.g. a SNP: T (Ref: C) or an insertion: AAAC (Ref: A)	GATK
Total_depth	Approximate read depth considering all samples; some reads may have been filtered	GATK
AF	Allele Frequency- for each ALT allele- in the same order as listed	Calculated from AD
AD	Allelic depths for the ref and alt alleles in the order listed	GATK
DP	Approximate read depth; some reads may have been filtered	GATK
GT	Genotype	GATK
Hugo_Symbol	HUGO symbol for the gene (HUGO symbols are always in all caps). "Unknown" is used for regions that do not correspond to a gene	VEP
Entrez_Gene_Id	Entrez gene ID (an integer). "0" is used for regions that do not correspond to a gene region or Ensembl ID	VEP
NCBI_Build	The reference genome used for the alignment (GRCh38)	VEP
Chromosome	The affected chromosome (chr1)	VEP
Start_Position	Lowest numeric position of the reported variant on the genomic reference sequence. Mutation start coordinate	VEP
End_Position	Highest numeric genomic position of the reported variant on the genomic reference sequence. Mutation end coordinate	VEP
Variant_Classification	Translational effect of variant allele	VEP
Variant_Type	Type of mutation. TNP (tri-nucleotide polymorphism) is analogous to DNP (di-nucleotide polymorphism) but for three consecutive nucleotides. ONP (oligo-nucleotide polymorphism) is analogous to TNP but for consecutive runs of four or more (SNP, DNP, TNP, ONP, INS, DEL, or Consolidated)	VEP
Reference_Allele	The plus strand reference allele at this position. Includes the deleted sequence for a deletion or "-" for an insertion	VEP
Alternative_Allele	Primary data genotype for alternative allele 1. A "-" symbol for a deletion represents a variant. A "-" symbol for an insertion represents wild-type allele. Novel inserted sequence for insertion does not include flanking reference bases	VEP
dbSNP_RS	The rs-IDs from the dbSNP database, "novel" if not found in any database used, or null if there is no dbSNP record, but it is found in other databases	VEP
HGVSc	The coding sequence of the variant in HGVS recommended format	vcf2maf
HGVSp	The protein sequence of the variant in HGVS recommended format	vcf2maf
HGVSp_Short	Same as HGVSp, but using 1-letter amino-acid codes	vcf2maf
Transcript_ID	Transcript onto which the consequence of the variant has been mapped	vcf2maf
Exon_Number	the exon number (out of total number)	vcf2maf
all_effects	A semicolon delimited list of all possible variant effects, sorted by priority. This column is relevant to analyses that consider the effect of the variant on all alternate isoforms of the gene, or on non-coding/regulatory transcripts. The effects are sorted first by transcript biotype priority, then by effect severity, and finally by decreasing order of transcript length. Each effect in the list is in the format [SYMBOL,Consequence,HGVSp,Transcript_ID,RefSeq]	vcf2maf

Gene	Stable Ensembl ID of affected gene	VEP
Feature	Stable Ensembl ID of feature	VEP
Feature_type	Type of feature. Currently one of Transcript, RegulatoryFeature, MotifFeature	VEP
Consequence	Consequence type of this variation (http://useast.ensembl.org/info/genome/variation/predicted_data.html#consequence). This may contain multiple terms that are comma-separated. For example, a synonymous mutation might be close enough to an intron to alter splicing. VEP will report both "synonymous_variant" and "splice_region_variant" in this column, for such variants.	VEP
cDNA_position	Relative position of base pair in cDNA sequence	VEP
Codons	The alternative codons with the variant base in upper case	VEP
Existing_variation	Known identifier of existing variation. If the variant was seen in some other somatic/germline DB, its ID will be listed here (dbSNP, clinVar, COSMIC...). You can search any identifier in Ensembl (http://www.ensembl.org/index.html)	VEP
ALLELE_NUM	Allele number from input; 0 is reference, 1 is first alternate etc	VEP
DISTANCE	Shortest distance from variant to transcript	VEP
STRAND_VEP	The DNA strand (1 or -1) on which the transcript/feature lies	VEP
SYMBOL	The gene symbol	VEP
SYMBOL_SOURCE	The source of the gene symbol	VEP
HGNC_ID	Gene identifier from the HUGO Gene Nomenclature Committee	VEP
BIOTYPE	Biotype of transcript	VEP
CANONICAL	A flag (YES) indicating that the VEP-based canonical transcript, the longest translation, was used for this gene. If not, the value is null	VEP
CCDS	The CCDS identifier for this transcript, where applicable	VEP
ENSP	The Ensembl protein identifier of the affected transcript	VEP
SWISSPROT	UniProtKB/Swiss	VEP
TREMBL	UniProtKB/TrEMBL identifier of protein product	VEP
UNIPARC	UniParc identifier of protein product	VEP
RefSeq	RefSeq identifier for this transcript	VEP
SIFT	The SIFT prediction and/or score, with both given as prediction (score): **tolerated**: Not likely to have a phenotypic effect; **tolerated_low_confidence**: More likely to have a phenotypic effect than 'tolerated'; **deleterious**: Likely to have a phenotypic effect; **deleterious_low_confidence**: Less likely to have a phenotypic effect than 'deleterious'	VEP
PolyPhen	The PolyPhen prediction. **Probably damaging (PR)**: It is with high confidence supposed to affect protein function or structure; **possibly damaging (PO)**: It is supposed to affect protein function or structure; **benign (BE)**: Most likely lacking any phenotypic effect; **unknown (UN)**: When in some rare cases, the lack of data does not allow PolyPhen to make a prediction	VEP
EXON	the exon number (out of total number)	VEP
INTRON	The intron number (out of total number)	VEP
DOMAINS	The source and identifier of any overlapping protein domains	VEP
1KG_AF	Non-reference allele and frequency of existing variant in 1000 Genomes	VEP
1KG_AFR_AF	Non-reference allele and frequency of existing variant in 1000 Genomes combined African population	VEP
1KG_AMR_AF	Non-reference allele and frequency of existing variant in 1000 Genomes combined American population	VEP
1KG_ASN_AF	Non-reference allele and frequency of existing variant in 1000 Genomes combined Asian population	VEP

1KG_EAS_AF	Non-reference allele and frequency of existing variant in 1000 Genomes combined East Asian population	VEP
1KG_EUR_AF	Non-reference allele and frequency of existing variant in 1000 Genomes combined European population	VEP
1KG_SAS_AF	Non-reference allele and frequency of existing variant in 1000 Genomes combined South Asian population	VEP
1KG_AA_AF	Non-reference allele and frequency of existing variant in NHLBI-ESP African American population	VEP
1KG_EA_AF	Non-reference allele and frequency of existing variant in NHLBI-ESP European American population	VEP
CLIN_SIG	Clinical significance of variant from dbSNP, which includes ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/docs/clinsig/). Find important variants by looking for tags like pathogenic, likely_pathogenic, or drug_response.	VEP
SOMATIC	Somatic status of each ID reported under Existing_variation	VEP
PUBMED	Pubmed ID(s) of publications that cite existing variant	VEP
IMPACT	The impact modifier for the consequence type calculated by VEP: **HIGH** (H): The variant is assumed to have high (disruptive) impact in the protein, probably causing protein truncation, loss of function, or triggering nonsense mediated decay; **MODERATE** (M): A non-disruptive variant that might change protein effectiveness; **LOW** (L): Assumed to be mostly harmless or unlikely to change protein behavior; **MODIFIER** (MO): Usually non-coding variants or variants affecting non-coding genes, where predictions are difficult or there is no evidence of impact	VEP
PICK	indicates if this block of consequence data was picked by VEP's pick feature	VEP
VARIANT_CLASS	Sequence Ontology variant class	VEP
TSL	Transcript support level, which is based on independent RNA analyses. More info here: https://www.ensembl.org/info/genome/genebuild/transcript_quality_tags.html#tsl	VEP
HGVS_OFFSET	Indicates by how many bases the HGVS notations for this variant have been shifted	VEP
PHENO	Indicates if existing variant is associated with a phenotype, disease or trait	VEP
GENE_PHENO	Indicates if gene that the variant maps to is associated with a phenotype, disease or trait	VEP
FILTER	Copied from input MAF/VCF, with ExAC-based common_variant tag added, as explained below	VEP
flanking_bps	The reference allele per VCF specs, and its 2 flanking base pairs	VEP
gnomAD_AF	Frequency of existing variant in gnomAD exomes combined population	VEP
gnomAD_AFR_AF	Frequency of existing variant in gnomAD exomes African/American population	VEP
gnomAD_AMR_AF	Frequency of existing variant in gnomAD exomes American population	VEP
gnomAD_ASJ_AF	Frequency of existing variant in gnomAD exomes Ashkenazi Jewish population	VEP
gnomAD_EAS_AF	Frequency of existing variant in gnomAD exomes East Asian population	VEP
gnomAD_FIN_AF	Frequency of existing variant in gnomAD exomes Finnish population	VEP
gnomAD_NFE_AF	Frequency of existing variant in gnomAD exomes Non-Finnish European population	VEP
gnomAD_OTH_AF	Frequency of existing variant in gnomAD exomes combined other combined populations	VEP
gnomAD_SAS_AF	Frequency of existing variant in gnomAD exomes South Asian population	VEP