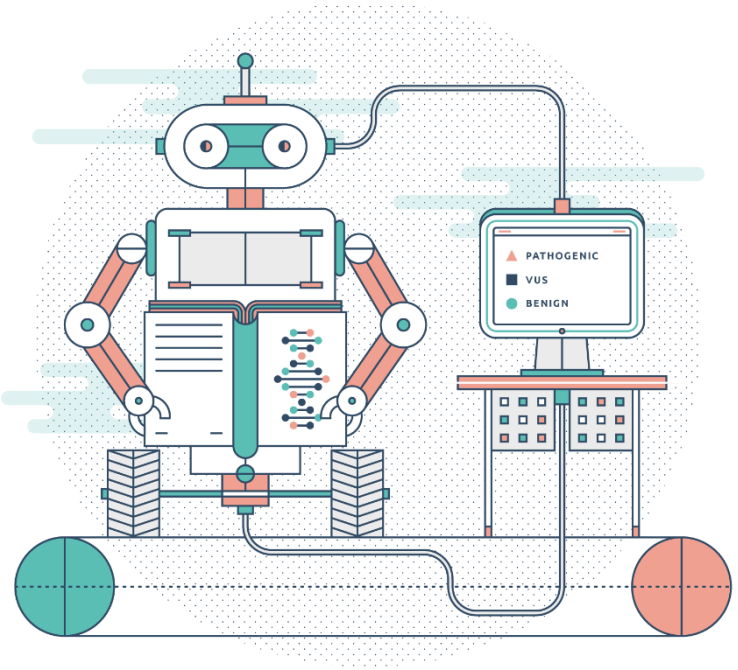
eVai – the Expert Variant Interpreter

v3.4

User Guide



The Expert Variant Interpreter - eVai - is enGenome's cloud solution developed with the goal to support geneticists in the interpretation of germline variants and increase the diagnostic yield of NGS-based tests.

By applying Artificial Intelligence on data gathered from more than 30 judiciously selected omics resources and an internal knowledgebase, eVai annotates genomic variants such as SNVs, INDELs and CNVs, automates the application of ACMG guidelines (Richards et al., 2015; Riggs et al., 2020), accurately classifies and prioritizes every genomic variant and reduces the turnaround time of the whole interpretation process.

Moreover, eVai enables professionals and geneticists to collect and curate findings in their own Laboratory Knowledgebase: a private repository of known variants designed to improve classification accuracy over time.

The underlying versioning system allows to automatically keep track of Knowledgebase modifications and facilitates the revision of uncertain variants whose interpretation may change over time because of software and omics resources updates.

For more information on eVai web interface usage and FAQs visit our Help Center
https://engenome.zendesk.com/hc/en-us

OMICS RESOURCES

Table 1. Versioned omics resources and tools used by eVai.

Name	Version	Description
gnomAD	4.1	Genome Aggregation Database (for SNV/INDEL)
gnomAD	2.1	Genome Aggregation Database (for CNVs to compare mid-size deletions)
gnomAD	3.0	Genome Aggregation Database (for CNVs only)
DGV	2016-05-15	Database of Genomic Variants (for CNVs only)
ExAC	r1	Exome Aggregatium Consortium
ESP	6500SIv2	NHLBI GO Exome Sequencing Project
dbSNP	151	dbSNP used for variant rs id and 1000 Genomes Project frequencies
1000 Genomes Project	Phase3	1000 Genomes Project
ClinVar	2024-06	ClinVar database used for bona-fide pathogenic/benign variants
DECIPHER	2024-06-09	DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources (for CNVs only)
ClinGen	2024-06-25	ClinGen database of curated genes and regions (for CNVs only)
MedGen	2024-06	MedGen database used for gene-conditions relations
Disease Ontology	2024-05-29	Human disease ontology, http://disease-ontology.org/
Human Phenotype Ontology	2024-07-01	Human phenotype ontology, https://hpo.jax.org/app/
Orphanet	2023-12-04	Orphanet database used for condition incidence/prevalence estimation
Repeat Masker	2023-05-09	Interspersed repeats and low complexity DNA sequences, obtained from UCSC tracks
PaPI	v1.1	Pseudo Amino Acid Protein Intolerance Variant Predictor (for coding variants SNVs/INDELs)
GERP++	-	Genomic Evolutional Rate Profile score (embedded in PaPI)
PhyloP	46way(placental)	Evolutionary conservation score (embedded in PaPI)
SiPhy	v0.5	Evolutionary substitution patterns score (embedded in PaPI)

HI Predictions	Version 3	Haploinsufficiency Decipher Index
CADD / CADD-Splice	v1.6	Combined Annotation Dependent Depletion, inlcludes CADD-Splice (for coding/non-coding variants, SNVs and INDELs)
REVEL	May 2021	REVEL (rare exome variant ensemble learner), an ensemble method for predicting the pathogenicity of missense variants on the basis of individual tools.
PolyPhen-2	2.2.2r394(HumVar)	Polymorphism Phenotyping Variant Predictor (for coding variants, SNVs)
SIFT	v2015	Sorting Intolerant From Tolerant Variant Predictor (for coding variants, SNVs)
dbscSNV	1.0	Database of all human SNVs within splicing consensus regions (for coding/non-coding variants, SNVs)
SnpEff	v4.3	Variant annotator
VEP	97.4	Variant Effect Predictor
RefSeq	105 (GRCh37) / 109 (GRCh38)	RefSeq transcript tracks
Ensembl	75 (GRCh37) / 97(GRCh38)	Ensembl transcripts tracks
UniProt	UP000005640_9606 2024_03	UniProt domains (for SNV/INDEL only)
Jaspar	75 (GRCh37) / 97 (GRCh38)	Transcription factor binding profiles database, embedded in SnpEff for GRCh37; obtained from Ensembl for GRCh38.
InterPro	78	InterPro domains (for CNV only)
MitoMap	2024-06	MitoMap database for mitochondrial variants
HmtVar	2022-10	Predictor for mitochondrial variants
MitoTIP	2023-01-15	Predictor for mitochondrial variants
dbSNP-mito	156	dbSNP used for mitochondrial variant rs
gnomAD-mito	3.1	Genome Aggregation Database (for mitochondrial variants only)

INPUT

eVai requires Variant Calling Format (VCF) files as input, optionally gzip compressed (.gz). VCF files can be easily analyzed in parallel batches. For more information visit the Help Center.

https://engenome.zendesk.com/hc/en-us/articles/360006614612-VCF-Upload

IMPORTANT NOTES:

- eVai currently analyzes variants on assemblies GRCh37/hg19/GRCh38/hg38/NC_012920.1 and on standard chromosomes (chr1-chr22, chrX, chrY, chrMT).
- In case of a multisample VCF file, only the first sample is taken into account for analysis.
- Copy number variants are required to have the VCF ALT field set to <DEL>, <DUP> or <CNV>.
- **Please note the following variants are excluded from the analysis**: intergenic SNV/INDEL; complex variants such as inversions (<INV>), translocations (<BND>) and gene fusions; variants on alternative haplotypes; Monomorphic reference events or no-call genotype (e.g. 0/0, ./., ./1).

OUTPUT

eVai internal workflow consists of two sequential steps:

1. Annotation
   Variants are matched against omics resources and information regarding transcripts, genes, protein domains, protein effects, associated diseases and many others are collected.

2. Classification
   eVai processes all information gathered from the annotation step, makes an internal data representation and matches evidences against ACMG/AMP guidelines for SNV/INDEL and ACMG/ClinGen guidelines for CNV, by finally assigning a predicted class and a pathogenicity or ACMG score.

For each analyzed sample, variants can be interpreted through the Variant Browser that reports annotation and classification information and prioritize variants according to the predicted pathogenicity score, thus fastening the interpretation process.
For more information visit the Help Center.
https://engenome.zendesk.com/hc/en-us/articles/360007254071-Variants-Sample-sheet

Moreover, the following two tab-delimited text files that contain the whole set of annotated and classified variants are available for download. These files can be easily imported in spreadsheet software (e.g. Excel).

NOTE: comments and classification adjustments are not reported within these files. To know more about how to export sample variants from the Variant Browser visit the Help Center.
https://engenome.zendesk.com/hc/en-us/articles/360006615112-Export-from-Variants-Browser

- **Annotation file v1.8 (*.tsv) - for SNVs/INDELs only**
  Each line corresponds to a genomic variant and each "value" or field is a useful annotation. See **Table 2a** for a detailed output file description.

  Both RefSeq and Ensembl transcript tracks are used to map variants to genes and compute the transcript-variant effect according to the MISO sequence ontology terms (Eilbeck et al., 2005) (e.g. missense, stop-gained; see EFFECT field in **Table 3**).

  A genomic variant can overlap multiple transcripts. Therefore, the following priority rule is used to report a variant overlapping a transcript on a single line of the TSV output file: for each MISO term associated with a transcript-variant effect (see **Table 3**), a score is assigned. Higher scores correspond to higher functional impacts of the variant on the transcript (e.g. a variant with a stop-gained effect on the transcript A has a higher score than the same variant with a synonymous effect on the transcript B). The transcript with the highest score is the one reported on a line of the TSV file. In case multiple variant-transcripts have the same score, the canonical (Ensembl) or Select (RefSeq) transcript is chosen. RefSeq transcripts have priority over Ensembl ones. Other overlapping transcripts with the corresponding variant effects are reported as well (see OTHER_TRANSCRIPT field in **Table 2a**).

  Genomic variants are finally annotated by population variant databases (e.g. gnomAD and ExAC (Lek et al., 2016)), variant-disease databases (e.g. MedGen and ClinVar (Landrum et al., 2014)), functional prediction tools for coding, splicing and non-coding variants (e.g. PaPI (Limongelli et al., 2015), PolyPhen-2 (Adzhubei et al., 2013), SIFT (Ng and Henikoff, 2003), dbscSNV (Jian and Liu, 2017), CADD (Rentzsch P. et al, 2021)) and other useful resources (see **OMICS RESOURCES** section).

- **CNV Annotation file v1.2 (*.CNV.tsv) - for CNVs only**
  Each line corresponds to a deletion or a duplication event and each "value" or field is a useful annotation. For each CNV the final ACMG/ClinGen class is reported as well, plus the ACMG score and the related conditions according to MedGen. See **Table 2b** for a detailed output file description and look at **ACMG/ClinGen CRITERIA DESCRIPTION** section for more details about ACMG/ClinGen recommendations for CNV interpretation.

  Moreover, look at our Best Practice guideline for a correct understanding of ACMG class and score used by eVai.

- **Mito Annotation file v1.0 (*.MT.tsv) - for SNVs/INDELs only**
  Each line corresponds to a genomic variant classified for a set of conditions, according to ACMG/AMP Mitochondrial recommendations. See **Table 2c** for a detailed output file description and look at **ACMG/ClinGen CRITERIA**

DESCRIPTION section for more details about ACMG/ClinGen recommendations for CNV interpretation.

- **ACMG Classification file v1.1 (*.ACMG.tsv) - for SNVs/INDELs only**

  Each line corresponds to a genomic variant classified for a specific condition, according to ACMG/AMP recommendations. For example, if a gene has been associated in literature (MedGen) to three possible conditions, a variant in that gene is classified and scored for each of the known related conditions, suggesting the most probable one.

  Each line reports variant information regarding gene, HGVS nomenclature, disease, inheritance pattern, observed genotype and a series of flags (ACMG criteria) that are set to “true” or “false” (see **Table 4**). When a certain criteria is set to “true”, it means that the corresponding ACMG criteria has been triggered. Look at **ACMG/AMP CRITERIA DESCRIPTION** section for more details about ACMG/AMP recommendations for SNV/INDEL interpretation.

  Finally, a pathogenicity score is assigned to each variant. This is computed on the basis of triggered ACMG criteria and the corresponding level of evidence. This is particularly useful to prioritize variants of unknown significance (VUS) in case no Pathogenic or Likely Pathogenic variants are found. We suggest to evaluate “Pathogenic” and “Likely pathogenic” variants first, followed by VUS with higher scores.
  Look at our Best Practice guideline for an accurate use of pathogenicity score with eVai.

  NOTE: common variants (MAF>5%) that are defined with the “Benign” class by ACMG guidelines are automatically classified as Benign according to the BA1 standalone criterion. Look at **ACMG CRITERIA DESCRIPTION** section for more details about ACMG criteria.

Table 2a. Annotation file: Tab-Separated Values (TSV) output fields.

Field Name	Description	Value
CHR	Variant chromosome	Text
START	Variant start position on chromosome	Number
END	Variant end position on chromosome	Number
REF	Reference allele	Text [ACGTN-]
ALT	Alternative allele	Text [ACGTN-]

EFFECT	Variant effect on transcript based on MISO (http://www.sequenceontology.org/miso)	Text Look at **Table 3** for a complete list of supported terms (multiple occurrences delimited by “&”)
GENE	Gene symbol according to HUGO Gene Nomenclature Committee (HGNC)	Text
TRANSCRIPT_ID	Transcript name according to the supported RefSeq or Ensembl version	Text
SELECT_CANONICAL	The transcript is “select” according to RefSeq or “canonical” according to Ensembl	Text true|false
TFBS_Id	Transcription factor binding site motif Id (Jaspar database (Khan et al., 2018))	Text
TFBS_name	Transcription factor name (Jaspar database)	Text
EXON_INTRON_NUM	Exon/Intron number	Number
HGVS_C	Coding HGVS nomenclature	Text
HGVS_P	Protein HGVS nomenclature	Text
CDS_DISTANCE	Variant distance (START) from the nearest CDS	Number
CDS_LEN	CDNA length of the current ENS[$version]_TRANSCRIPT	Number
AA_LEN	Protein length for the corresponding ENS[$version]_TRANSCRIPT	Number
OTHER_TRANSCRIPTS	Other transcripts overlapping the variant. For Ensembl: TRANSCRIPT_ID:GENE:EFFECT:HGVS_C:HGVS_P:CDS_DISTANCE For RefSeq: TRANSCRIPT_ID:GENE:select=[true|false];EFFECT:HGVS_C:HGVS_P:CDS_DISTANCE	Text multiple occurrences delimited by “|”
ExAC_AN	Total number of corrected alleles counted in ExAC database for this locus	Number
ExAC_AC	Number of corrected ALT alleles counted in ExAC database	Number
ExAC_AF	ExAC frequency of the ALT allele [0-1]	Number
ExAC_isTarget	The genomic locus is covered by ExAC according to the WES design file (.bed)	Text true|false
DBSNP	The dbSNP id (no rs)	Text
DBSNP_VERSION	The dbSNP version that first included the variant	Text
DBSNP_1TGP_REF_freq	1000 Genomes Project REF allele frequency [0-1] as reported in dbSNP	Number

DBSNP_1TGP_ALT_freq	1000 Genomes Project ALT allele frequency [0-1] as reported in dbSNP	Number
COMMON_1TGP_1_perc	ALT allele frequency is greater than 1% in at least one population with at least two individuals from different families having the same minor allele	Number
ESP[$version]_EA_freq	ALT allele frequency in ESP European population [0-100]	Number
ESP[$version]_AA_freq	ALT allele frequency in ESP African American population [0-100]	Number
ESP[$version]_All_freq	ALT allele frequency in ESP general population [0-100]	Number
gnomAD_AF_ALL	ALT allele frequency in gnomAD considering all populations	Number
gnomAD_Hom_ALL	Number of Homozygotes in gnomAD considering all populations	Number
gnomAD_AF_MAX_POP	Maximum ALT allele frequency in gnomAD considering all populations excluding founder-population (Finnish, Ashkenazi Jewish)	Number
CADD_score	Probability for this variant (both SNVs and INDELs) to be deleterious according to CADD [9]. Both for coding and non-coding genomic variants. Variants can be considered damaging for scores > 20.	Number
dbscSNV_AB_score	Probability for this variant (SNV only) to be deleterious for the nearby splicing site. Score computed by AdaBoost machine learning classifier. Valid for variants at −3 to +8 at the 5' splice site and −12 to +2 at the 3' splice site. Variants can be considered damaging for scores > 0.9.	Number
dbscSNV_RF_score	Probability for this variant (SNV only) to be deleterious for the nearby splicing site. Score computed by Random Forest machine learning classifier. Valid for variants at −3 to +8 at the 5' splice site and −12 to +2 at the 3' splice site. Variants can be considered damaging for scores > 0.9.	Number
PaPI_pred	PaPI (http://papi.unipv.it) prediction for this variant to be damaging/tolerated for the protein structure/function. It is the combined prediction given by PolyPhen-2, SIFT and PseeAC-RF classifiers	Text DAMAGING |TOLERATED
PaPI_score	PaPI (http://papi.unipv.it) score for this variant to be damaging/tolerated for the protein structure/function. It is the combined score given by PolyPhen-2, SIFT and PseeAC-RF classifiers. Variants can be considered damaging for scores > 0.5.	Number
PolyPhen-2_pred	PolyPhen-2 (HumVar) prediction for this variant to be damaging/tolerated for the protein structure/function	Text D|B

PolyPhen-2_score	PolyPhen-2 (HumVar) score for this variant to be damaging/tolerated for the protein structure/function	Number
SIFT_pred	SIFT prediction for this variant to be damaging/tolerated for the protein structure/function	Text D|B
SIFT_score	SIFT score for this variant to be damaging/tolerated for the protein structure/function	Number
PseeAC-RF_pred	Random Forest Pseudo-Amino acidic classifier prediction for this variant to be damaging/tolerated for the protein structure/function	Text D|B
PseeAC-RF_score	Random Forest Pseudo-Amino acidic classifier score for this variant to be damaging/tolerated for the protein structure/function	Number
ClinVar_hotSpot	Variant overlaps a ClinVar “hotspot”, defined as a genomic region at high density of pathogenic variants reported in ClinVar (no conflictual interpretation variants only). It reports: the length in bases of the hotspot region | the number of pathogenic variants clustered in the hotspot | the percentage of altered bases in the hotspot region	Number | Number | Number
ClinVar_RCV	Accession number for ClinVar	Text multiple occurrences delimited by “|”
ClinVar_clinical_significance	Variant classification reported in ClinVar according to different submissions (RCVs)	Text multiple occurrences delimited by “|”
ClinVar_rev_status	Review status in ClinVar	Text multiple occurrences delimited by “|”
ClinVar_traits	List of ClinVar submitted traits	Text multiple occurrences delimited by “|”
ClinVar_PMIDS	List of PUBMED Ids reported in ClinVar for this variant	Text multiple occurrences delimited by “;”
Diseases	Associated diseases to the gene	Text multiple occurrences delimited by “;”
Disease_IDs	ID associated to the disease	Text multiple occurrences delimited by “;”
[$SAMPLE].GENO	The zygosity corresponding to REF/ALT allele. hom=ALT/ALT; het=ALT/REF; multi= multiple alleles (not only REF and ALT) occur at this	Text hom|het|multi

	genomic locus. Extracted from VCF SAMPLE GT if available.	
[$SAMPLE].QUAL	Quality of the variant. Extracted from VCF QUAL field if available.	Number
[$SAMPLE].GENO_QUAL	Genotype quality. Extracted from VCF SAMPLE GQ if available.	Number
[$SAMPLE].FILTER	Extracted from VCF FILTER if available.	Text multiple occurrences delimited by ";"
[$SAMPLE].AF	ALT allele frequency. Extracted from VCF if available or inferred.	Number
[$SAMPLE].AO	Read depth for the ALT allele. Extracted from VCF if available or inferred	Number
[$SAMPLE].RO	Read depth for the REF allele. Extracted from VCF if available or inferred	Number
[$SAMPLE].COV	Read depth for the genomic locus. Extracted from VCF if available or inferred	Number

Table 2b. CNV Annotation file: Tab-Separated Values (TSV) output fields.

Field Name	Description	Value
CHR	Variant chromosome	Text
START	Variant start position on chromosome	Number
END	Variant end position on chromosome	Number
TYPE	Type of the CNV	Text [DEL, DUP]
SIZE	Size of the CNV	Number
[$SAMPLE].CN	Copy Number status, if available in input vcf	Number
[$SAMPLE].CN_CONSISTENCY	Flag which evaluates consistency between Copy Number Status and CNV type. This field has been introduced because some CNV calling tools reports a CN which is not consistent with the CNV type (e.g. deletions with CN=3)	Text true|false|to_check
[$SAMPLE].VCF_GT	Genotype, if available in input vcf	As reported in VCF (e.g. 0/1, 1/1)
[$SAMPLE].COV	Coverage (in case DP field present in vcf file)	Number
[$SAMPLE].READS_RATIO	Reads ratio of the CNV compared to the control reference set (custom “READSRATIO” VCF INFO)	Number
[$SAMPLE].QUAL	QUAL field value of input vcf	Number
[$SAMPLE].FILTER	FILTER field value of input vcf	Text
[$SAMPLE].PRECISE	Flag which evaluate precision of CNV breakpoints. This flag assumes a not-empty value (True or False) if the standard vcf flag (PRECISE or IMPRECISE) is available in input file	Text true|false|.
CONSEQUENCE	List of unique predicted consequences of the CNV based on MISO (http://www.sequenceontology.org/miso)	Text multiple occurrences delimited by “,”
GENES_NUMBER	Number of genes involved in the CNV	Number
GENES	List of gene symbols involved in the CNV. The list reports HGNC, Clone_based_ensembl_gene and Clone_based_vega_gene	Text multiple occurrences delimited by “,”
Ensembl_single-gene_transcript	Informations about the canonical Ensembl transcript when the CNV involves a single gene. If no canonical available, one	Text fields separated by “|”

	random transcript will be reported. The field reports gene symbol, transcript_id, gene_id, exons and introns involved, CNV consequence on transcript according to MISO, canonical flag and source (HGNC or Clone_based).	
RefSeq_single-gene_transcript	Informations about the main RefSeq transcript when the CNV involves a single gene. If no transcript variant 1 available, one random transcript will be reported. The field reports gene symbol, transcript_id, gene_id, exons and introns involved, CNV consequence on transcript according to MISO, transcript description and source (HGNC or Clone_based).	Text fields separated by "|"
HI_gene_patho	List of established haploinsufficiency pathogenic genes involved, as curated by ClinGen	Text multiple occurrences delimited by ","
HI_gene_ben	List of established haploinsufficiency benign genes involved, as curated by ClinGen	Text multiple occurrences delimited by ","
TS_gene_patho	List of established triplosensitivity pathogenic genes involved, as curated by ClinGen	Text multiple occurrences delimited by ","
TS_gene_ben	List of established triplosensitivity benign genes involved, as curated by ClinGen	Text multiple occurrences delimited by ","
ClinGen_region	ClinGen regions IDs overlapping the CNV	Text multiple occurrences delimited by ","
ClinGen_overlap(%)	Percentage]0-100] of overlap of the examined CNV compared to the ClinGen region	Number multiple occurrences delimited by ","
ClinGen_classification	ClinGen region classification. If CNV is a deletion, classification fits haploinsufficiency score; otherwise classification fits triplosensitivity score	Text multiple occurrences delimited by ","
ClinGen_phenotype	Disease ID associated to ClinGen region	Text multiple occurrences delimited by ","
ClinVar_patho	Similar ClinVar pathogenic or likely pathogenic variants ID overlapping the CNV. Only variants of same type of the CNV are reported.	Text multiple occurrences delimited by ","
ClinVar_ben	Similar ClinVar benign or likely benign variants ID overlapping the. Only variants of same type of the CNV are reported.	Text multiple occurrences delimited by ","

ClinVar_uncertain	ClinVar variants ID overlapping the CNV with uncertain significance associated. Only variants of same type of the CNV are reported.	Text multiple occurrences delimited by “,”
ClinVar_conflicting	ClinVar variants ID overlapping the CNV with conflicting clinical significance interpretation. Only variants of same type of the CNV are reported.	Text multiple occurrences delimited by “,”
ClinVar_other	ClinVar variants ID overlapping the CNV with other interpretation associated (e.g. drug response). Only variants of same type of the CNV are reported.	Text multiple occurrences delimited by “,”
Decipher_patho	Similar Decipher pathogenic or likely pathogenic variants ID overlapping the CNV. Only variants of same type of the CNV are reported.	Text multiple occurrences delimited by “,”
Decipher_ben	Similar Decipher benign or likely benign variants ID overlapping the CNV. Only variants of same type of the CNV are reported.	Text multiple occurrences delimited by “,”
Decipher_uncertain	Decipher variants ID overlapping the CNV with uncertain significance associated. Only variants of same type of the CNV are reported.	Text multiple occurrences delimited by “,”
Decipher_unknown	Decipher variants ID overlapping the CNV with unknown significance associated. Only variants of same type of the CNV are reported.	Text multiple occurrences delimited by “,”
gnomAD_ID	ID of most frequent and similar gnomAD variant overlapping the CNV. Other gnomAD variants might overlap the CNV under evaluation but are not reported.	Text
gnomAD_AF	Allele Frequency [0-1] of the gnomAD variant reported	Number
DGV_ID	ID of most frequent and similar DGV (Database of Genomic Variants) variant overlapping the CNV. Other DGV variants could overlap the CNV under evaluation but are not reported	Text
DGV_AF	Allele Frequency [0-1] of the DGV variant reported	Number
other_single-gene_transcripts	Informations about other transcripts when the CNV involves a single gene. The field reports gene symbol, transcript_id, gene_id, exons and introns involved, CNV consequence on transcript according to MISO, canonical flag for ensembl transcripts and description for refseq ones, and source (HGNC or Clone_based). Transcripts records are semicolon separated. Transcripts which are reported in Ensembl_single-gene_transcript or RefSeq_single-gene_transcript, will not be repeated in this field.	Text fields separated by “|” multiple occurrences delimited by “;”

ACMG_score	ACMG/ClinGen score derived from the sum of points from the single ACMG criteria	Number
ACMG_classification	ACMG/ClinGen class as a consequence of the ACMG score. Pathogenic (score >= 0.99); Likely pathogenic (0.9 <= score <= 0.98); Uncertain significance (-0.89 <= score <= 0.89); Likely benign (-0.98 <= score <= -0.9); Benign (score <= -0.99)	Text [Pathogenic | Likely pathogenic | Likely benign | Benign | Uncertain significance]
MedGen_Conditions	MedGen CUI(s)	multiple occurrences delimited by “,”
Inheritances	Inheritance patterns for each reported condition	multiple occurrences delimited by “,”

Table 2c. Mitochondrial Annotation file: Tab-Separated Values (TSV) output fields.

Field Name	Description	Value
CHR	Variant chromosome	Text
START	Variant start position on chromosome	Number
END	Variant end position on chromosome	Number
REF	Reference allele	Text [ACGTN-]
ALT	Alternative allele	Text [ACGTN-]
VCF_ORIG	Variant representation in VCF format in the original VCF file (CHR:POS:REF:ALT)	Text
EFFECT	Variant effect on transcript based on MISO (http://www.sequenceontology.org/miso)	Text Look at **Table 3** for a complete list of supported terms (multiple occurrences delimited by “&”)
GENE	Gene symbol according to HUGO Gene Nomenclature Committee (HGNC)	Text
TRANSCRIPT_ID	Transcript name according to the supported Ensembl version	Text
HGVS_M	HGVS mitochondrial notation	Text

HGVS_P	Protein HGVS nomenclature	Text
HGVS_C	Coding HGVS nomenclature	Text
OTHER_TRANSCRIPTS	Other transcripts overlapping the variant. TRANSCRIPT_ID:GENE:EFFECT:HGVS_C:HGVS_P:CDS_DISTANCE	Text multiple occurrences delimited by "|"
DBSNP	The dbSNP id (no rs)	Text
DBSNP_VERSION	Allele Frequency [0-1] of the gnomAD variant reported	Number
Mitomap_Poly_AF	MitoMap polymorphism	Text
gnomAD_*	ALT allele frequency in gnomAD considering all populations	Number
Mitotip_score	The Mitochondrial tRNA Informatics Predictor (MitoTIP). Variants are considered damaging for scores >= 12.66.	Text
Mitotip_prediction	MitoTIP prediction	Text
HmtVar_score	The Mitochondrial tRNA Informatics Predictor (MitoTIP). Variants are considered damaging for scores >= 12.66.	Text
HmtVar_prediction	HmtVar prediction	Text
Mitomap_disease_clinical_significance	MitoMap disease	Text
ClinVar_RCV	Accession number for ClinVar	Text multiple occurrences delimited by "|"
ClinVar_clinical_significance	ClinVar_clinical_significance	Variant classification reported in ClinVar according to different submissions (RCVs)
[$Sample].*	Sample's related info about genotype, quality etc. See **Table 2a**.	
MEDGEN_CUI	MedGen CUI(s)	multiple occurrences delimited by ", "
FINAL_CLASSIFICATION	ACMG/AMP variant classification	Text

		[Pathogenic | Likely pathogenic | Likely benign | Benign | Uncertain significance]
SCORE_OF_PATHOGENICITY	Weighted score of pathogenicity based on triggered ACMG criteria and their level of evidence	Number
FLAG	Description of an event: "no_support" for variants of Uncertain significance because not enough evidences (acriteria) to assign a class; "no_medgen" value is reported for those variants overlapping genes that are not associated yet to conditions according to the literature (MedGen); "conflict" is reported for variants having conflicting criteria (both a pathogenic and benign classification is supported).	Text multiple occurrences delimited by "|"
CRITERIA	Activated ACMG criteria	Text multiple occurrences delimited by "|"
OTHER_CONDITIONS OTHER_MEDGEN_CUI	Other MedGen CUI(s) for annotation purposes	Text multiple occurrences delimited by "|"
OTHER_FINAL_CLASSIFICATIONS	Other possible "ACMG/AMP variant classification" for each condition	Text multiple occurrences delimited by "|"
OTHER_SCORES_OF_PATHOGENICITY	Other scores for each condition/classification	Text multiple occurrences delimited by "|"

Table 3. Annotation file: EFFECT field explained. For more details look at MISO documentation (http://www.sequenceontology.org/miso). Following terms can be combined together in the annotated file and separated by the "&" character. The MISO terms are here reported according to a descending order in terms of the transcript-impact-score.

Term	Description

stop_gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript
stop_lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript
frameshift_variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three
start_lost	A codon variant that changes at least one base of the canonical start codon
exon_loss / exon_loss_variant	A sequence variant whereby an exon is lost from the transcript
splice_acceptor_variant	A splice variant that changes the 2 base region at the 3' end of an intron
splice_donor_variant	A splice variant that changes the 2 base pair region at the 5' end of an intron
disruptive_inframe_insertion	An inframe increase in cds length that inserts one or more codons into the coding sequence within an existing codon
disruptive_inframe_deletion	An inframe decrease in cds length that deletes bases from the coding sequence starting within an existing codon
inframe_insertion	An inframe non synonymous variant that inserts bases into in the coding sequence
inframe_deletion	An inframe non synonymous variant that deletes bases from the coding sequence
missense_variant	A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved
initiator_codon_variant	A codon variant that changes at least one base of the first codon of a transcript
splice_region_variant	A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron
start_retained_variant	A sequence variant where at least one base in the start codon is changed, but the start remains
non_canonical_start_codon	A start codon that is not the usual AUG sequence
stop_retained_variant	A sequence variant where at least one base in the terminator codon is changed, but the terminator remains
synonymous_variant	A sequence variant where there is no resulting change to the encoded amino acid
intron_variant	A transcript variant occurring within an intron
5_prime_UTR_premature_start_codon_gain_variant	A 5' UTR variant where a premature start codon is gained
3_prime_UTR_truncation	A sequence variant that causes the reduction of a the 3' UTR with regard to the reference sequence
5_prime_UTR_truncation	A sequence variant that causes the reduction of a the 5' UTR with regard to the reference sequence

3_prime_UTR_variant	A UTR variant of the 3' UTR
5_prime_UTR_variant	A UTR variant of the 5' UTR
non_coding_exon_variant / non_coding_transcript_exon_variant	A sequence variant that changes non-coding exon sequence in a non-coding transcript.
intragenic_variant	A variant that occurs within a gene but falls outside of all transcript features. This occurs when alternate transcripts of a gene do not share overlapping sequence

Table 4. ACMG Classification file: Tab-Separated Values (TSV) output.

Field Name	Description	Value
CHR	Variant chromosome	Text
START	Variant start position on chromosome	Number
STOP	Variant end position on chromosome	Number
REF	Reference allele	Text [ACGTN-]
ALT	Alternative allele	Text [ACGTN-]
PRIOR_TRANSCRIPT	Transcript name according to the RefSeq/Ensembl	Text
HGVS_P	Protein HGVS nomenclature	Text
HGVS_C	Coding HGVS nomenclature	Text
SELECT_CANONICAL	The transcript is "canonical" according to Ensembl supported version or "select" according to RefSeq	Text true|false
OTHER_TRANSCRIPT_HGVS	Other transcripts overlapping with the current TRANSCRIPT TRANSCRIPT:HGVS_P:HGVC_C	Text multiple occurrences delimited by "|"
GENOTYPE	The zygosity corresponding to REF/ALT allele. hom=ALT/ALT; het=ALT/REF; multi= multiple alleles for the same genomic locus	Text hom|het|multi
GENE	Gene symbol according to HUGO Gene Nomenclature	Text
PHENOTYPE	Disease description accordingly to MedGen	Text
MEDGEN_CUI	MedGen Concept ID	Text

INHERITANCE	Inheritance pattern according to MedGen	Text multiple occurrences delimited by "|"
FINAL_CLASSIFICATION	ACMG/AMP variant classification	Text [Pathogenic | Likely pathogenic | Likely benign | Benign | Uncertain significance]
SCORE_OF_PATHOGENICITY	Weighted score of pathogenicity based on triggered ACMG criteria and their level of evidence	Number
FLAG	Description of an event: "no_support" for variants of Uncertain significance because not enough evidences (criteria) to assign a class; "no_medgen" value is reported for those variants overlapping genes that are not associated yet to conditions according to the literature (MedGen); "conflict" is reported for variants having conflicting criteria (both a pathogenic and benign classification is supported).	Text multiple occurrences delimited by "|"
NOTE	General notes (e.g. "classification_from_frequency" means the variant has been classified by evaluating the Benign standalone BA1 criterion only)	Text multiple occurrences delimited by "|"
VCF_ORIG	Variant representation in VCF format in the original VCF file (CHR:POS:REF:ALT)	Text
PVS1, PS1, PS3, PM1, PM2, PM4, PM5, PP2, PP3, PP4, PP5, BA1, BS1, BS3, BP1, BP3, BP4, BP6, BP7, BP8	ACMG Criteria (look at **ACMG CRITERIA DESCRIPTION** section). In case a criteria has a modified level of evidence, the modified level of evidence is reported in brackets, e.g. TRUE(STRONG).	Text true|false

ACMG/AMP Criteria description (SNVs/INDELs only)

18 out of 28 ACMG/AMP criteria are automatically evaluated by the eVai while one supporting benign criteria has been added (see BP8).

Criteria annotated as “MANUAL” are not triggered automatically, due to additional information needed on patient clinical records or co-segregation. However, it is possible to “adjust” the predicted classification by adding or modifying ACMG criteria through the Variant Browser (see Help Center, https://engenome.zendesk.com/hc/en-us/articles/360006615032-Interpretation-Editing).
Criteria annotated as “CUSTOM” has been introduced to extend and improve the guidelines.

Pathogenic Criteria

VERY STRONG LEVEL OF EVIDENCE FOR PATHOGENICITY

PVS1

Null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease. Moreover, a previously established pathogenic variant has been observed in the same exon.

NOTE: this criterion is declassed to Strong level of evidence if no other pathogenic variant has been observed in the same exon.

STRONG LEVEL OF EVIDENCE FOR PATHOGENICITY

PS1

Same amino acid change as a previously established pathogenic variant regardless of nucleotide change

PS2 (MANUAL)

De novo (both maternity and paternity confirmed) in a patient with the disease and no family history

PS3

Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product

PS4 (MANUAL)

The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls

MODERATE LEVEL OF EVIDENCE FOR PATHOGENICITY

PM1

Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation

PM2

Absent from controls (or at extremely low frequency if recessive) in population databases (e.g. gnomAD, ExAC)

PM3 (MANUAL)

For recessive disorders, detected in trans with a pathogenic variant

PM4

Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants

PM5

Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before

PM6 (MANUAL)

Assumed de novo, but without confirmation of paternity and maternity

SUPPORTING LEVEL EVIDENCE FOR PATHOGENICITY

PP1 (MANUAL)

Co-segregation with disease in multiple affected family members in a gene definitively known to cause the disease

PP2

Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease

PP3

Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.)

PP4 (MANUAL)

Patient's phenotype or family history is highly specific for a disease with a single genetic etiology

PP5

Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation.

Benign Criteria

STAND ALONE LEVEL EVIDENCE FOR BENIGNITY

BA1

Allele frequency is >5% in population databases (e.g. gnomAD)

STRONG LEVEL EVIDENCE FOR BENIGNITY

BS1

Allele frequency is greater than expected for disorder

BS2

Observed in a healthy adult individual for a recessive (homozygous), dominant heterozygous), or X-linked (hemizygous) disorder, with full penetrance expected at an early age

BS3

Well-established in vitro or in vivo functional studies show no damaging effect on protein function or splicing

BS4 (MANUAL)

Lack of segregation in affected members of a family

SUPPORTING LEVEL EVIDENCE FOR BENIGNITY

BP1

Missense variant in a gene for which primarily truncating variants are known to cause disease

BP2 (MANUAL)

Observed in trans with a pathogenic variant for a fully penetrant dominant gene/disorder or observed in cis with a pathogenic variant in any inheritance pattern

BP3

In-frame deletions/insertions in a repetitive region without a known function

BP4

Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc.)

BP5 (MANUAL)

Variant found in a case with an alternate molecular basis for disease

BP6

Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation

BP7

A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved

BP8 (CUSTOM)

Same amino acid change, but not the same nucleotide change, has been reported as benign variant by a reputable source.

ACMG/ClinGen Criteria description (CNV only)

21 out of 36 ACMG/ClinGen criteria for Copy Number Loss (DEL) and 24 out of 40 ACMG/ClinGen criteria for Copy Number Gain (DUP) events are automatically evaluated by the eVai.

Criteria annotated as “NA” are not triggered automatically, due to additional information needed on patient clinical records or co-segregation.

1A (DEL, DUP)

Contains protein-coding or other known functionally important elements

1B (DEL, DUP)

Does NOT contain protein-coding or any known functionally important elements

2A (DEL)

Complete overlap of an established Haploinsufficient (HI) gene/genomic region (or non-HI gene for recessive condition)

2A (DUP)

Complete overlap of an established Triplosensitive (TS) gene/genomic region

2B (DEL)

Partial overlap of an established Haploinsufficient (HI) genomic region

2B (DUP)

Partial overlap of an established Triplosensitive (TS) region

2C (DEL)

Partial overlap with the 5' end of an established Haploinsufficient (HI) gene (3' end of the gene not involved). For recessive condition also non-HI genes are considered

2C (DUP)

Identical in gene content to the established benign copy number gain

2D (DEL)

Partial overlap with the 3' end of an established Haploinsufficient (HI) gene (5' end of the gene not involved). For recessive condition also non-HI genes are considered

2D (DUP)

Smaller than established benign copy number gain, breakpoint(s) does not interrupt protein-coding genes

2E (DEL)

Both breakpoints are within the same Haploinsufficient (HI) gene. For recessive condition also non-HI genes are considered

2E (DUP)

Smaller than established benign copy number gain, breakpoint(s) potentially interrupts protein-coding gene

2F (DEL)

Completely contained within an established benign CNV region

2F (DUP)

Larger than known benign copy number gain, does not include additional protein-coding genes

2G (DEL)

Overlaps an established benign CNV, but includes additional genomic material

2G (DUP)

Overlaps a benign copy number gain but includes additional genomic material

2H (DEL)

Two or more HI predictors suggest that at least one gene in the interval is haploinsufficient (HI)

2H (DUP)

Haploinsufficient (HI) gene fully contained within observed copy number gain. For recessive condition also non-HI genes are considered

2I (DUP)

Both breakpoints are within the same Haploinsufficient (HI) gene (possibly resulting in loss of function (LOF). For recessive condition also non-HI genes are considered

2J (DUP) (NA)

One breakpoint is within an established Haploinsufficient (HI) gene, the patient's phenotype is either inconsistent with what is expected for LOF of that gene OR unknown. For recessive condition also non-HI genes are considered

2K (DUP)

One breakpoint is within an established Haploinsufficient (HI) gene, patient's phenotype is highly specific and consistent with what is expected for LOF of that gene

2L (DUP)

One or both breakpoints are within gene(s) of no established clinical significance

3A (DEL)

Up to 24 protein-coding genes wholly or partially included in the copy number event

3A (DUP)

Up to 34 protein-coding genes wholly or partially included in the copy number event

3B (DEL)

Between 25 and 34 protein-coding genes wholly or partially included in the copy number event

3B (DUP)

Between 34 and 49 protein-coding genes wholly or partially included in the copy number event

3C (DEL)

More than 34 protein-coding genes wholly or partially included in the copy number event

3C (DUP)

More than 49 protein-coding genes wholly or partially included in the copy number event

4A (DEL)

Reported proband (from literature, public databases, or internal lab data) has either: a complete deletion of the gene (or a LOF variant within), encompassed by the observed copy number loss or an overlapping copy number loss similar in genomic content to the observed copy number loss and the reported phenotype is highly specific and relatively unique to the gene or genomic region

4A (DUP)

Reported proband has either: complete duplication of one or more genes within the observed copy number gain OR an overlapping copy number gain similar in genomic content to the observed copy number gain and the reported phenotype is highly specific and relatively unique to the gene or genomic region

4B (DEL)

Reported proband (from literature, public databases, or internal lab data) has either: a complete deletion of the gene (or a LOF variant within), encompassed by the observed copy number loss or an overlapping copy number loss similar in genomic content to the observed copy number loss and the reported phenotype is consistent with the gene/genomic region, is highly specific, but is not necessarily unique to the gene/genomic region

4B (DUP)

Reported proband has either: complete duplication of one or more genes within the observed copy number gain OR an overlapping copy number gain similar in genomic content to the observed copy number gain and the reported phenotype is consistent with the gene/genomic region, is highly specific, but is not necessarily unique to the gene/genomic region

4C (DEL)

Reported proband (from literature, public databases, or internal lab data) has either: a complete deletion of the gene (or a LOF variant within), encompassed by the observed copy number loss or an overlapping copy number loss similar in genomic content to the observed copy number loss and the reported phenotype is consistent with the gene/genomic region, but not highly specific and/or with high genetic heterogeneity

4C (DUP)

Reported proband has either: complete duplication of one or more genes within the observed copy number gain OR an overlapping copy number gain similar in genomic content to the observed copy number gain and the reported phenotype is consistent with the gene/genomic region, but not highly specific and/or with high genetic heterogeneity

4D (DEL) (NA)

Reported proband (from literature, public databases, or internal lab data) has either: a complete deletion of the gene (or a LOF variant within), encompassed by the observed copy number loss or an overlapping copy number loss similar in genomic content to the observed copy number loss and the reported phenotype is NOT consistent with what is expected for the gene/genomic region or not consistent in general

4D (DUP) (NA)

Reported proband has either: complete duplication of one or more genes within the observed copy number gain OR an overlapping copy number gain similar in genomic content to the observed copy number gain and the reported phenotype is NOT consistent with the gene/genomic region or not consistent in general

4E (DEL,DUP)

Reported proband has a highly specific phenotype consistent with the gene/genomic region, but the inheritance of the variant is unknown

4F (DEL,DUP)

Segregation among similarly affected family members in literature (3-4 segregations)

4G (DEL,DUP)

Segregation among similarly affected family members in literature (5-6 segregations)

4H (DEL,DUP)

Segregation among similarly affected family members in literature (more than 7 segregations)

4I (DEL,DUP) (NA)

Variant is NOT found in another individual in the proband's family AFFECTED with a consistent, specific, well-defined phenotype (no known phenocopies)

4J (DEL,DUP) (NA)

Variant IS found in another individual in the proband's family UNAFFECTED with the specific, well-defined phenotype observed in the proband

4K (DEL,DUP) (NA)

Variant IS found in another individual in the proband's family UNAFFECTED with the non-specific phenotype observed in the proband

4L (DEL,DUP) (NA)

Statistically significant increase amongst observations in cases (with a consistent, specific, well-defined phenotype) compared to controls

4M (DEL,DUP) (NA)

Statistically significant increase amongst observations in cases (without a consistent, non-specific phenotype OR unknown phenotype) compared to controls

4N (DEL,DUP) (NA)

No statistically significant difference between observations in cases and controls

4O (DEL,DUP)

Overlap with common population variation

5A (DEL,DUP) (NA)

CNV is de novo evaluating the patient's family history (same score as 4A-4D).

5B (DEL) (NA)

Patient with specific, well-defined phenotype and no family history. CNV is inherited from an apparently unaffected parent

5B (DUP) (NA)

Patient with a specific, well-defined phenotype and no family history. Copy number gain is inherited from an apparently unaffected parent

5C (DEL) (NA)

Patient with non-specific phenotype and no family history. CNV is inherited from an apparently unaffected parent

5C (DUP) (NA)

Patient with non-specific phenotype and no family history. Copy number gain is inherited from an apparently unaffected parent

5D (DEL,DUP) (NA)

CNV segregates with a consistent phenotype observed in the patient's family (same score as 4F-4H)

5E (DEL,DUP) (NA)

No segregation among affected family members (same score as 4I-4K)

5F (DEL,DUP) (NA)

Inheritance information is unavailable or uninformative

5G (DEL,DUP) (NA)

Inheritance information is unavailable or uninformative. The patient phenotype is non-specific, but is consistent with what has been described in similar cases

5H (DEL,DUP) (NA)

Inheritance information is unavailable or uninformative. The patient phenotype is highly specific and consistent with what has been described in similar cases

REFERENCES

Adzhubei I, Jordan DM, Sunyaev SR. 2013. Predicting functional effect of human missense mutations using PolyPhen-2. Current Protocols in Human Genetics Chapter 7:Unit7.20.

Eilbeck K, Lewis SE, Mungall CJ, Yandell M, Stein L, Durbin R, Ashburner M. 2005. The Sequence Ontology: a tool for the unification of genome annotations. Genome Biology 6:R44.

Jian X, Liu X. 2017. In Silico Prediction of Deleteriousness for Nonsynonymous and Splice-Altering Single Nucleotide Variants in the Human Genome. Methods Mol Biol 1498:191–197.

Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, Lee R van der, Bessy A, Chèneby J, Kulkarni SR, Tan G, Baranasic D, Arenillas DJ, et al. 2018. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Research 46:D260–D266.

Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. 2014. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res 42:D980–D985.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285–291.

Limongelli I, Marini S, Bellazzi R. 2015. PaPI: pseudo amino acid composition to score human protein-coding variants. BMC Bioinformatics 16:123.

Ng PC, Henikoff S. 2003. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Research 31:3812–3814.

Rentzsch P, Schubach M, Shendure J, Kircher M. 2021. CADD-Splice—improving genome-wide variant effect prediction using deep learning-derived splice scores. Genome Medicine 13, 31.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, et al. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17:405–424.

Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, Raca G, Ritter DI, South ST, Thorland EC, Pineda-Alvarez D, Aradhya S, et al. 2020. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genetics in Medicine: Official Journal of the American College of Medical Genetics 22:245–257.

Elizabeth M McCormick et al. 2020.Specifications of the ACMG/AMP standards and guidelines for mitochondrial DNA variant interpretation. Human Mutation Dec;41(12):2028-2057.

Lott, M.T., Leipzig, J.N., Derbeneva, O., Xie, H.M., Chalkia, D., Sarmady, M., Procaccio, V., and Wallace, D.C. 2013. mtDNA variation and analysis using MITOMAP and MITOMASTER. Current Protocols in Bioinformatics 1(123):1.23.1-26.

Preste R et al., HmtVar: a new resource for human mitochondrial variations and pathogenicity data. Nucleic Acids Res. 2019 Jan 8;47(D1):D1202-D1210.

Sonney S, Leipzig J, Lott MT, Zhang S, Procaccio V, et al. (2017) Predicting the pathogenicity of novel variants in mitochondrial tRNA with MitoTIP. PLOS Computational Biology 13(12).

CHANGE LOG

v3.4

- REVEL score for missense variants introduced
- PP3/BP4 criteria calibrated according to ClinGen recommendation (V. Pejaver et al, 2022)
- New inheritance tags for family analysis introduced: Germline mosaicism (Gm), X linked Germline mosaicism (Gx), Private (P), Ambiguous (A), X linked Ambiguous (Ax), Y linked Ambiguous (Ay)
- Minor bug fix for VCF compatibility (PhiX contig, CADD run optimization)

v3.3

- Following resources have been updated: ClinVar (2024-06), MedGen (2024-06), Disease Ontology (2023-05-24), HPO (2024-07-01), UniProt (UP000005640_9606 2024_03), Decipher (2024-06-29), ClinGen (2024-06-25), MitoMap (2024-06)
- Bug fix for an underestimation of deleted number of bases for intragenic deletion (ie. 2E-PVS1 triggered more often with a Strong level of evidence instead of Moderate).
- Improved Orphanet matches for conditions' inheritance

v3.1

- Following resources have been updated: ClinVar (2023-10), MedGen (2023-10), Disease Ontology (2023-09-28), HPO (2023-09-01), UniProt (UP000005640_9606 2023_10), Decipher (2023-11-05), ClinGen (2023-11-02), MitoMap (2023-06-21)
- Improved PP5 criterion in case of (likely) pathogenic variants in ClinVar according to the rating star (level of evidence increased to strong or very strong)
- PP5 and BP6 now are mutually exclusive
- Somatic variants from ClinVar are now taken into account
- Introduced functional studies from DMS studies for the following genes: BRCA1, PTEN and SCN5A
- PVS1 score for CNV increased to 0.9
- Big CNV events (more than 50 conditions) are interpreted by ClinGen region only if at least one condition is associated with it
- CNVs events with "./." genotype are filtered out only in case of multisample VCF
- Bug fix to manage SVLEN float exp (e.g. SVLEN could be reported in VCF as 1.3e5)
- CNV DEL and DUP concurrency bug in multisample VCF
- Bug fixed preventing for failing when coverage is reported as "." in multisample VCF
- Bug fix for mitochondrial family analysis when one of the sample does not hold MT variants

v3.0

- Introduced mitochondrial interpretation according to ACMG/AMP guidelines for
- Added the following resources for mitochondrial interpretation: MitoMap (2023-01-05), HmtVar (2022-10), MitoTIP (2023-01-15), gnomAD v3.1, dbSNP v156.

v2.8

- Following resources have been updated: ClinVar (2023-04), MedGen (2023-04), Disease Ontology (2022-04-13), HPO (2023-01-27), UniProt (UP000005640_9606 2023_05), Decipher (2023-02-26), ClinGen (2023-04-23), RepeatMasker (2023-05-09)
- Fixed bug for missing PaPI, Polyphen2 and and SIFT scores in case only precomputed predictions are used
- Improved application of PS1, PM5, PP5 on different conditions in case of conflicting Clinvar submissions
- Improved AF inference from VCF in case of AF is present in INFO filed

v2.7

- Following resources have been updated: ClinVar (2022-01), MedGen (2022-11), Disease Ontology (2022-11-01), HPO (2022-10-05), UniProt (UP000005640_9606 2022_11), Decipher (2022-10-30), ClinGen (2022-11-02)
- Introduced CADD and dismissed DANN
- Fixed bug in reporting for particular cases ClinVar matches for the same amino acid change
- Extended gene-conditions associations to pathogenic ClinVar hits with at least 2 stars
- Introduced a PP3/BP4 application constraint: in case of conflict predictors (considering PaPI, CADD, dbscSNV only), none of these two criteria is applied.

v2.6

- Following resources have been updated: ClinVar (2022-06), MedGen (2022-06), Disease Ontology (2022-06-07), HPO (2022-06-12), UniProt (UP000005640_9606 2022_02), Decipher (2022-06-05), ClinGen (2022-06-21)
- Introduced BA1 exception list according to Ghosh et al (PMID: 30311383)
- Introduced the automatic application of BS2

v2.5

- Following resources have been updated: gnomAD (for SNVs and INDELs, v2.1.1)
- Optimized data processing (annotation step for gnomAD and PaPI functional predictions)

v2.4

- Following resources have been updated: ClinVar (2022-01), MedGen (2022-01), Disease Ontology (2021-12-15), HPO (2021-10-10), Orphanet (2022-01), RepeatMasker (GRCh38, 2022-01-28)
- SNV/INDEL with absent genotype sample format in VCF (GT) are considered heterozygous by default

v2.3

- Following resources have been updated: ClinVar (2021-06), MedGen (2021-06), Disease Ontology (2021-06-08), HPO (2021-06-13), Orphanet (2021-01), RepeatMasker (2021-06-29)
- Following resources have been added: RefSeq (105 for GRCh37, 109 for GRCh38); UniProt (UP000005640_9606)
- Improved application of “PM1” criterion for SNV/INDEL (Uniprot domains with known pathogenic variants)
- Implemented “4B” criterion for CNV duplications and deletions.

v2.2

- Implemented CNV criterion 2L for duplication.
- Following resources have been updated: ClinVar (2020-12), MedGen (2020-11), Disease Ontology (2020-12-02), HPO (2020-10-12), DECIPHER (2020-11-15), ClinGen (2020-11-21).
- Improved accuracy in gene-condition association.

v2.1

- Introduced CNV interpretation according to ACMG/ClinGen guidelines and visualization in Variant Browser.

v2.0

- Introduced CNV annotation for <DEL> and <DUP> events if present in the VCF file.

v0.8

- Following resources have been updated: ClinVar (2020-06-02), MedGen (2020-06-23), Disease Ontology (2020-06-18), HPO (2020-06-08), Orphanet (2020-01).
- PVS1 rule improved to take into account only coding transcripts and pathogenic null variants in the same or following exons of the gene.
- Improved PP5 and PS1 in case of indels (ClinVar match).

v0.7

- Following resources have been updated: ClinVar (2020-01), MedGen (2020-01-22), Disease Ontology (2020-01-15), HPO (2019-11-08).
- Family analysis now include de novo variants as possible het compound candidates

v0.6

- Added hg38/GRCh38 support

v0.5

- Following resources have been updated: ExAC (version r1), ClinVar (2019-03), MedGen (2019-03-13), Orphanet (2019-01), Disease Ontology (2019-03), RepeatMasker (2019-03-14).
- Annotation file updated to version 1.5: modified field name “ClinVar_ID” in “ClinVar_RCV”; Fileds “ClinVar_RCV” and “ClinVar_rev_status” can now hold multiple values separated by “|”.
- Improved the way in dealing with ClinVar RCV with multiple submissions and “conflicting in interpretation of pathogenicity” clinical significance.

v0.4.2

- Added support for family analysis
- Improved PM2 application

v0.4.1

- Improved application for PVS1 and BP1 ACMG/AMP criteria
- Optimized elaboration of coverage info from VCF files

v0.4

- Added gnomAD r2.0.2 (Total/Populations Allele Frequencies and Number of Homozygotes over Whole Exome and Whole Genome samples)
- Enhanced support of different VCF types and versions, gVCF included
- Following resources have been updated: ClinVar (2018-07), MedGen (2018-08-02), Orphanet (2018-06), Disease Ontology (2018-08), RepeatMasker (2018-07-11)
- Improved application of BS1 criterion for recessive conditions
- Added field “CDS_DISTANCE” in the Annotation file
- Included the VCF original and normalized variant representation

v0.3.1

- Variants calls on mitochondrial and alternative contigs (e.g. chrMT, MT, GL*, chrUn*) are ignored

v0.3

- Variants can be queried and visualized on the Variant Browser by the web interface

v0.2.1

- Homozygous reference VCF calls (e.g. 0/0 or ./.) genotypes are skipped

v0.2

- Pathogenicity score added
- Following resources have been updated: MedGen (2017_03), ClinVar (2017_04)