Single transcript Output

- **CHROM –** Chrosome id
- **POS -** Position
- **ID –** dbsnp id
- **REF-**Reference allel
- **ALT-** Reference allele as used to get consequences
- **Consequence** - consequence type of this variant
- **IMPACT** - the impact modifier for the consequence type
- **SYMBOL** - the gene symbol
- **Gene** - Ensembl stable ID or RefSeq stable ID of affected gene
- **Feature** - Ensembl stable ID or RefSeq stable ID of feature
- **EXON** - the exon number (out of total number)
- **INTRON** - the intron number (out of total number)
- **HGVSc** - the HGVS coding sequence name
- **HGVSp** - the HGVS protein sequence name
- **Codon** - the alternative codons with the variant base in upper case
- **STRAND** - the DNA strand (1 or -1) on which the transcript/feature lies
- **SIFT** - the SIFT prediction and/or score, with both given as prediction(score)
- **PolyPhen** - the PolyPhen prediction and/or score
- **DOMAINS** - the source and identifer of any overlapping protein domains
- **AF** - Frequency of existing variant in 1000 Genomes
- **AFR_AF** - Frequency of existing variant in 1000 Genomes combined African population
- **AMR_AF** - Frequency of existing variant in 1000 Genomes combined American population
- **EUR_AF** - Frequency of existing variant in 1000 Genomes combined European population
- **EAS_AF** - Frequency of existing variant in 1000 Genomes combined East Asian population
- **SAS_AF** - Frequency of existing variant in 1000 Genomes combined South Asian population
- **AA_AF** - Frequency of existing variant in NHLBI-ESP African American population
- **EA_AF** - Frequency of existing variant in NHLBI-ESP European American population
- **gnomAD_AF** - Frequency of existing variant in gnomAD exomes combined population
- **gnomAD_AFR_AF** - Frequency of existing variant in gnomAD exomes African/American population

- **gnomAD_AMR_AF** - Frequency of existing variant in gnomAD exomes American population
- **gnomAD_ASJ_AF** - Frequency of existing variant in gnomAD exomes Ashkenazi Jewish population
- **gnomAD_EAS_AF** - Frequency of existing variant in gnomAD exomes East Asian population
- **gnomAD_FIN_AF** - Frequency of existing variant in gnomAD exomes Finnish population
- **gnomAD_NFE_AF** - Frequency of existing variant in gnomAD exomes Non-Finnish European population
- **gnomAD_OTH_AF** - Frequency of existing variant in gnomAD exomes combined other combined populations
- **gnomAD_SAS_AF** - Frequency of existing variant in gnomAD exomes South Asian population
- **MAX_AF** - Maximum observed allele frequency in 1000 Genomes, ESP and gnomAD
- **MAX_AF_POPS** - Populations in which maximum allele frequency was observed
- **ada_score** - ensemble prediction score based on ada-boost. Ranges 0 to 1. The larger the score the higher probability the scSNV will affect splicing. The suggested cutoff for a binary prediction (affecting splicing vs. not affecting splicing) is 0.6.
- **rf_score** - ensemble prediction score based on random forests. Ranges 0 to 1. The larger the score the higher probability the scSNV will affect splicing. The suggested cutoff for a binary prediction (affecting splicing vs. not affecting splicing) is 0.6.
- **dpsi_max_tissue** – SPIDEX score
- **dpsi_zscore** – SPIDEX score
- **clinvar_CLNDN** - ClinVar's preferred disease name for the concept specified by disease identifiers in CLNDISDB
- **clinvar_CLNSIG** - Clinical significance for this single variant
- **clinvar_CLNREVSTAT** - ClinVar review status for the Variation ID
- **FORMAT - vcf** FORMAT column, corresponds to **Sample** column.

  AD：Allelic depths for the ref and alt alleles in the order listed；
  DP:Read depth
  GQ:Genotype quality
  GT:Genotype
  PGT: Physical phasing haplotype information, describing how the alternate alleles are phased in relation to one another
  PID: Physical phasing ID information, where each unique ID within a given sample (but not across samples) connects records within a phasing group
  PL:The phred-scaled genotype likelihoods rounded to the closest integer
- **Sample -** Sample name
- **pathogenicity –** Clinical significance in Knowledge database
- **disease –** related disease about this variant in Knowledge database

- **pubmed_id -** related literature about this variant in Knowledge database
- **Mim Number –** omim number of omim phenotype
- **Phenotypes –** omim phenotype of this gene

dbNSFP result:

- LRT_score: The original LRT two-sided p-value (LRTori), ranges from 0 to 1.
- LRT_converted_rankscore: LRTori scores were first converted as LRTnew=1-LRTori*0.5 if Omega<1, or LRTnew=LRTori*0.5 if Omega>=1. Then LRTnew scores were ranked among all LRTnew scores in dbNSFP. The rankscore is the ratio of the rank over the total number of the scores in dbNSFP. The scores range from 0.00162 to 0.84324.
- LRT_pred: LRT prediction, D(eleterious), N(eutral) or U(nknown), which is not solely determined by the score.
- LRT_Omega: estimated nonsynonymous-to-synonymous-rate ratio (Omega, reported by LRT)
- MutationTaster_score: MutationTaster p-value (MTori), ranges from 0 to 1. Multiple scores are separated by ";". Information on corresponding transcript(s) can be found by querying http://www.mutationtaster.org/ChrPos.html
- MutationTaster_converted_rankscore: The MTori scores were first converted: if the prediction is "A" or "D" MTnew=MTori; if the prediction is "N" or "P", MTnew=1-MTori. Then MTnew scores were ranked among all MTnew scores in dbNSFP. If there are multiple scores of a SNV, only the largest MTnew was used in ranking. The rankscore is the ratio of therank of the score over the total number of MTnew scores in dbNSFP. The scores rangefrom 0.08979 to 0.81033.
- MutationTaster_pred: MutationTaster prediction, "A" ("disease_causing_automatic"), "D" ("disease_causing"), "N" ("polymorphism") or "P" ("polymorphism_automatic"). The score cutoff between "D" and "N" is 0.5 for MTnew and 0.31713 for the rankscore.
- MutationTaster_model: MutationTaster prediction models.
- MutationTaster_AAE: MutationTaster predicted amino acid change.
- MutationAssessor_UniprotID: Uniprot ID number provided by MutationAssessor.
- MutationAssessor_variant: AA variant as to MutationAssessor_UniprotID.
- MutationAssessor_score: MutationAssessor functional impact combined score (MAori). The score ranges from -5.135 to 6.49 in dbNSFP.
- MutationAssessor_rankscore: MAori scores were ranked among all MAori scores in dbNSFP.The rankscore is the ratio of the rank of the score over the total number of MAori scores in dbNSFP. The scores range from 0 to 1.
- MutationAssessor_pred: MutationAssessor's functional impact of a variant :predicted functional, i.e. high ("H") or medium ("M"), or predicted non-functional, i.e. low ("L") or neutral ("N"). The MAori score cutoffs between "H" and "M", "M" and "L", and "L" and "N", are 3.5, 1.935 and 0.8, respectively. The rankscore cutoffs between "H" and "M", "M" and "L", and "L" and "N", are 0.92922, 0.51944 and 0.19719, respectively.
- FATHMM_score: FATHMM default score (weighted for human inherited-disease mutations with Disease Ontology) (FATHMMori). Scores range from -16.13 to 10.64. The

smaller the score the more likely the SNP has damaging effect. Multiple scores separated by ";", corresponding to Ensembl_proteinid.

- FATHMM_converted_rankscore: FATHMMori scores were first converted toFATHMMnew=1-(FATHMMori+16.13)/26.77, then ranked among all FATHMMnew scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of FATHMMnew scores in dbNSFP. If there are multiple scores, only the most damaging (largest) rankscore is presented. The scores range from 0 to 1.
- FATHMM_pred: If a FATHMMori score is <=-1.5 (or rankscore >=0.81332) the corresponding nsSNV is predicted as "D(AMAGING)"; otherwise it is predicted as "T(OLERATED)".Multiple predictions separated by ";", corresponding to Ensembl_proteinid.
- PROVEAN_score: PROVEAN score (PROVEANori). Scores range from -14 to 14. The smaller the score the more likely the SNP has damaging effect. Multiple scores separated by ";", corresponding to Ensembl_proteinid.
- PROVEAN_converted_rankscore: PROVEANori were first converted to PROVEANnew=1-(PROVEANori+14)/28, then ranked among all PROVEANnew scores in dbNSFP. The rankscore is the ratio of the rank the PROVEANnew score over the total number of PROVEANnew scores in dbNSFP. If there are multiple scores, only the most damaging (largest) rankscore is presented.The scores range from 0 to 1.
- PROVEAN_pred: If PROVEANori <= -2.5 (rankscore>=0.543) the corresponding nsSNV is predicted as "D(amaging)"; otherwise it is predicted as "N(eutral)". Multiple predictions separated by ";", corresponding to Ensembl_proteinid.
- MetaSVM_score: Our support vector machine (SVM) based ensemble prediction score, which incorporated 10 scores (SIFT, PolyPhen-2 HDIV, PolyPhen-2 HVAR, GERP++, MutationTaster, Mutation Assessor, FATHMM, LRT, SiPhy, PhyloP) and the maximum frequency observed in the 1000 genomes populations. Larger value means the SNV is more likely to be damaging. Scores range from -2 to 3 in dbNSFP.
- MetaSVM_rankscore: MetaSVM scores were ranked among all MetaSVM scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of MetaSVM scores in dbNSFP. The scores range from 0 to 1.
- MetaSVM_pred: Prediction of our SVM based ensemble prediction score,"T(olerated)" or "D(amaging)". The score cutoff between "D" and "T" is 0. The rankscore cutoff between "D" and "T" is 0.82268.
- MetaLR_score: Our logistic regression (LR) based ensemble prediction score, which incorporated 10 scores (SIFT, PolyPhen-2 HDIV, PolyPhen-2 HVAR, GERP++, MutationTaster, Mutation Assessor, FATHMM, LRT, SiPhy, PhyloP) and the maximum frequency observed in the 1000 genomes populations. Larger value means the SNV is more likely to be damaging. Scores range from 0 to 1.
- MetaLR_rankscore: MetaLR scores were ranked among all MetaLR scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of MetaLR scores in dbNSFP. The scores range from 0 to 1.
- MetaLR_pred: Prediction of our MetaLR based ensemble prediction score,"T(olerated)" or "D(amaging)". The score cutoff between "D" and "T" is 0.5. The rankscore cutoff between "D" and "T" is 0.81113.

- Reliability_index: Number of observed component scores (except the maximum frequency in the 1000 genomes populations) for MetaSVM and MetaLR. Ranges from 1 to 10. As MetaSVM and MetaLR scores are calculated based on imputed data, the less missing component scores, the higher the reliability of the scores and predictions.
- M-CAP_score: M-CAP score (details in DOI: 10.1038/ng.3703). Scores range from 0 to 1. The larger the score the more likely the SNP has damaging effect.
- M-CAP_rankscore: M-CAP scores were ranked among all M-CAP scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of M-CAP scores in dbNSFP.
- M-CAP_pred: Prediction of M-CAP score based on the authors' recommendation, "T(olerated)" or "D(amaging)". The score cutoff between "D" and "T" is 0.025.
- MutPred_score: General MutPred score. Scores range from 0 to 1. The larger the score the more likely the SNP has damaging effect.
- MutPred_rankscore: MutPred scores were ranked among all MutPred scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of MutPred scores in dbNSFP.
- MutPred_protID: UniProt accession or Ensembl transcript ID used for MutPred_score calculation.
- MutPred_AAchange: Amino acid change used for MutPred_score calculation.
- MutPred_Top5features: Top 5 features (molecular mechanisms of disease) as predicted by MutPred with p values. MutPred_score > 0.5 and p < 0.05 are referred to as actionable hypotheses.MutPred_score > 0.75 and p < 0.05 are referred to as confident hypotheses.MutPred_score > 0.75 and p < 0.01 are referred to as very confident hypotheses.
- fathmm-MKL_coding_score: fathmm-MKL p-values. Scores range from 0 to 1. SNVs with scores >0.5 are predicted to be deleterious, and those <0.5 are predicted to be neutral or benign. Scores close to 0 or 1 are with the highest-confidence. Coding scores are trained using 10 groups of features. More details of the score can be found in doi: 10.1093/bioinformatics/btv009.
- fathmm-MKL_coding_rankscore: fathmm-MKL coding scores were ranked among all fathmm-MKL coding scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of fathmm-MKL coding scores in dbNSFP.
- fathmm-MKL_coding_pred: If a fathmm-MKL_coding_score is >0.5 (or rankscore >0.28317) the corresponding nsSNV is predicted as "D(AMAGING)"; otherwise it is predicted as "N(EUTRAL)".
- fathmm-MKL_coding_group: the groups of features (labeled A-J) used to obtained the score. More details can be found in doi: 10.1093/bioinformatics/btv009.
- Eigen_coding_or_noncoding: Whether Eigen-raw and Eigen-phred scores are based on coding model or noncoding model.
- Eigen-raw: Eigen score for coding SNVs. A functional prediction score based on conservation, allele frequencies, and deleteriousness prediction using an unsupervised learning method (doi: 10.1038/ng.3477).
- Eigen-phred: Eigen score in phred scale.

- Eigen-PC-raw: Eigen PC score for genome-wide SNVs. A functional prediction score based on conservation, allele frequencies, deleteriousness prediction (for missense SNVs) and epigenomic signals (for synonymous and non-coding SNVs) using an unsupervised learning method (doi: 10.1038/ng.3477).
- Eigen-PC-phred: Eigen PC score in phred scale.
- Eigen-PC-raw_rankscore: Eigen-PC-raw scores were ranked among all Eigen-PC-raw scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of Eigen-PC-raw scores in dbNSFP.
- GenoCanyon_score: A functional prediction score based on conservation and biochemical annotations using an unsupervised statistical learning. (doi:10.1038/srep10576)
- GenoCanyon_score_rankscore: GenoCanyon_score scores were ranked among all integrated fitCons scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of GenoCanyon_score scores in dbNSFP.
- integrated_fitCons_score: fitCons score predicts the fraction of genomic positions belonging to a specific function class (defined by epigenomic "fingerprint") that are under selective pressure. Scores range from 0 to 1, with a larger score indicating a higher proportion of nucleic sites of the functional class the genomic position belong to are under selective pressure, therefore more likely to be functional important. Integrated (i6) scores are integrated across three cell types (GM12878, H1-hESC and HUVEC). More details can be found in doi:10.1038/ng.3196.
- integrated_fitCons_rankscore: integrated fitCons scores were ranked among all integrated fitCons scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of integrated fitCons scores in dbNSFP.
- integrated_confidence_value: 0 - highly significant scores (approx. p<.003); 1 - significant scores (approx. p<.05); 2 - informative scores (approx. p<.25); 3 - other scores (approx. p>=.25).
- GM12878_fitCons_score: fitCons score predicts the fraction of genomic positions belonging to a specific function class (defined by epigenomic "fingerprint") that are under selective pressure. Scores range from 0 to 1, with a larger score indicating a higher proportion of nucleic sites of the functional class the genomic position belong to are under selective pressure, therefore more likely to be functional important. GM12878 fitCons scores are based on cell type GM12878. More details can be found in doi:10.1038/ng.3196.
- GM12878_fitCons_rankscore: GM12878 fitCons scores were ranked among all GM12878 fitCons scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of GM12878 fitCons scores in dbNSFP.
- GM12878_confidence_value: 0 - highly significant scores (approx. p<.003); 1 - significant scores (approx. p<.05); 2 - informative scores (approx. p<.25); 3 - other scores (approx. p>=.25).
- H1-hESC_fitCons_score: fitCons score predicts the fraction of genomic positions belonging to a specific function class (defined by epigenomic "fingerprint") that are under selective pressure. Scores range from 0 to 1, with a larger score indicating a higher proportion of nucleic sites of the functional class the genomic position belong to are

under selective pressure, therefore more likely to be functional important. GM12878 fitCons scores are based on cell type H1-hESC. More details can be found in doi:10.1038/ng.3196.

- H1-hESC_fitCons_rankscore: H1-hESC fitCons scores were ranked among all H1-hESC fitCons scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of H1-hESC fitCons scores in dbNSFP.
- H1-hESC_confidence_value: 0 - highly significant scores (approx. p<.003); 1 - significant scores (approx. p<.05); 2 - informative scores (approx. p<.25); 3 - other scores (approx. p>=.25).
- HUVEC_fitCons_score: fitCons score predicts the fraction of genomic positions belonging to a specific function class (defined by epigenomic "fingerprint") that are under selective pressure. Scores range from 0 to 1, with a larger score indicating a higher proportion of nucleic sites of the functional class the genomic position belong to are under selective pressure, therefore more likely to be functional important. GM12878 fitCons scores are based on cell type HUVEC. More details can be found in doi:10.1038/ng.3196.
- HUVEC_fitCons_rankscore: HUVEC fitCons scores were ranked among all HUVEC fitCons scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of HUVEC fitCons scores in dbNSFP.
- HUVEC_confidence_value: 0 - highly significant scores (approx. p<.003); 1 - significant scores (approx. p<.05); 2 - informative scores (approx. p<.25); 3 - other scores (approx. p>=.25).
- GERP++_NR: GERP++ neutral rate
- GERP++_RS: GERP++ RS score, the larger the score, the more conserved the site. Scores range from -12.3 to 6.17.
- GERP++_RS_rankscore: GERP++ RS scores were ranked among all GERP++ RS scores in dbNSFP.The rankscore is the ratio of the rank of the score over the total number of GERP++ RS scores in dbNSFP.
- phyloP100way_vertebrate: phyloP (phylogenetic p-values) conservation score based on the multiple alignments of 100 vertebrate genomes (including human). The larger the score, the more conserved the site. Scores range from -20.0 to 10.003 in dbNSFP.
- phyloP100way_vertebrate_rankscore: phyloP100way_vertebrate scores were ranked among all phyloP100way_vertebrate scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of phyloP100way_vertebrate scores in dbNSFP.
- phyloP20way_mammalian: phyloP (phylogenetic p-values) conservation score based on the multiple alignments of 20 mammalian genomes (including human). The larger the score, the more conserved the site. Scores range from -13.282 to 1.199 in dbNSFP.
- phyloP20way_mammalian_rankscore: phyloP20way_mammalian scores were ranked among all phyloP20way_mammalian scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of phyloP20way_mammalian scores in dbNSFP.
- phastCons100way_vertebrate: phastCons conservation score based on the multiple alignments of 100 vertebrate genomes (including human). The larger the score, the more conserved the site. Scores range from 0 to 1.

- phastCons100way_vertebrate_rankscore: phastCons100way_vertebrate scores were ranked among all phastCons100way_vertebrate scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of phastCons100way_vertebrate scores in dbNSFP.
- phastCons20way_mammalian: phastCons conservation score based on the multiple alignments of 20 mammalian genomes (including human). The larger the score, the more conserved the site. Scores range from 0 to 1.
- phastCons20way_mammalian_rankscore: phastCons20way_mammalian scores were ranked among all phastCons20way_mammalian scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of phastCons20way_mammalian scores in dbNSFP.
- SiPhy_29way_pi: The estimated stationary distribution of A, C, G and T at the site, using SiPhy algorithm based on 29 mammals genomes.
- SiPhy_29way_logOdds: SiPhy score based on 29 mammals genomes. The larger the score, the more conserved the site. Scores range from 0 to 37.9718 in dbNSFP.
- SiPhy_29way_logOdds_rankscore: SiPhy_29way_logOdds scores were ranked among all SiPhy_29way_logOdds scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of SiPhy_29way_logOdds scores in dbNSFP.