dbSNP Columns

RS	dbSNP ID (i.e. rs number)
RSPOS	Chr position reported in dbSNP
RV	RS orientation is reversed
VP	Variation Property. Documentation is at ftp://ftp.ncbi.nlm.nih.gov/snp/specs/dbSNP_BitField_latest.pdf
GENEINFO	Pairs each of gene symbol:gene id. The gene symbol and id are delimited by a colon (:) and each pair is delimited by a vertical bar (|)
dbSNPBuildID	First dbSNP Build for RS
SAO	Variant Allele Origin: 0 - unspecified, 1 - Germline, 2 - Somatic, 3 - Both
SSR	Variant Suspect Reason Codes (may be more than one value added together) 0 - unspecified, 1 - Paralog, 2 - byEST, 4 - oldAlign, 8 - Para_EST, 16 - 1kg_failed, 1024 - other
WGT	Weight, 00 - unmapped, 1 - weight 1, 2 - weight 2, 3 - weight 3 or more
VC	Variation Class
PM	Variant is Precious(Clinical,Pubmed Cited)
TPA	Provisional Third Party Annotation(TPA) (currently rs from PHARMGKB who will give phenotype data)
PMC	Links exist to PubMed Central article
S3D	Has 3D structure - SNP3D table
SLO	Has SubmitterLinkOut - From SNP->SubSNP->Batch.link_out
NSF	Has non-synonymous frameshift A coding region variation where one allele in the set changes all downstream amino acids. FxnClass = 44
NSM	Has non-synonymous missense A coding region variation where one allele in the set changes protein peptide. FxnClass = 42
NSN	Has non-synonymous nonsense A coding region variation where one allele in the set changes to STOP codon (TER). FxnClass = 41
REF	Has reference A coding region variation where one allele in the set is identical to the reference sequence. FxnCode = 8
SYN	Has synonymous A coding region variation where one allele in the set does not change the encoded amino acid. FxnCode = 3
U3	In 3' UTR Location is in an untranslated region (UTR). FxnCode = 53
U5	In 5' UTR Location is in an untranslated region (UTR). FxnCode = 55
ASS	In acceptor splice site FxnCode = 73
DSS	In donor splice-site FxnCode = 75
INT	In Intron FxnCode = 6
R3	In 3' gene region FxnCode = 13
R5	In 5' gene region FxnCode = 15
OTH	Has other variant with exactly the same set of mapped positions on NCBI refernce assembly.
CFL	Has Assembly conflict. This is for weight 1 and 2 variant that maps to different chromosomes on different assemblies.
ASP	Is Assembly specific. This is set if the variant only maps to one assembly
MUT	Is mutation (journal citation, explicit fact): a low frequency variation that is cited in journal and other reputable sources
VLD	Is Validated. This bit is set if the variant has 2+ minor allele count based on frequency or genotype data.
G5A	>5% minor allele frequency in each and all populations

G5	>5% minor allele frequency in 1+ populations
HD	Marker is on high density genotyping kit (50K density or greater). The variant may have phenotype associations present in dbGaP.
GNO	Genotypes available. The variant has individual genotype (in SubInd table).
KGPhase1	1000 Genome phase 1 (incl. June Interim phase 1)
KGPhase3	1000 Genome phase 3
CDA	Variation is interrogated in a clinical diagnostic assay
LSD	Submitted from a locus-specific database
MTP	Microattribution/third-party annotation(TPA:GWAS,PAGE)
OM	Has OMIM/OMIA
NOC	Contig allele not present in variant allele list. The reference sequence allele at the mapped position is not present in the variant allele list, adjusted for orientation.
WTD	Is Withdrawn by submitter If one member ss is withdrawn by submitter, then this bit is set. If all member ss' are withdrawn, then the rs is deleted to SNPHistory
NOV	Rs cluster has non-overlapping allele sets. True when rs set has more than 2 alleles from different submissions and these sets share no alleles in common.
CAF	An ordered, comma delimited list of allele frequencies based on 1000Genomes, starting with the reference allele followed by alternate alleles as ordered in the ALT column. Where a 1000Genomes alternate allele is not in the dbSNPs alternate allele set, the allele is added to the ALT column. The minor allele is the second largest value in the list, and was previuosly reported in VCF as the GMAF. This is the GMAF reported on the RefSNP and EntrezSNP pages and VariationReporter
COMMON	RS is a common SNP. A common SNP is one that has at least one 1000Genomes population with a minor allele of frequency >= 1% and for which 2 or more founders contribute to that minor allele frequency.

ESP Columns

DBSNP	dbSNP version which established the rs_id
EA_AC	European American Allele Count in the order of AltAlleles,RefAllele. For INDELs, A1, A2, or An refers to the N-th alternate allele while R refers to the reference allele.
AA_AC	African American Allele Count in the order of AltAlleles,RefAllele. For INDELs, A1, A2, or An refers to the N-th alternate allele while R refers to the reference allele.
TAC	Total Allele Count in the order of AltAlleles,RefAllele For INDELs, A1, A2, or An refers to the N-th alternate allele while R refers to the reference allele.
MAF	Minor Allele Frequency in percent in the order of EA,AA,All
GTS	Observed Genotypes. For INDELs, A1, A2, or An refers to the N-th alternate allele while R refers to the reference allele.
EA_GTC	European American Genotype Counts in the order of listed GTS
AA_GTC	African American Genotype Counts in the order of listed GTS
GTC	Total Genotype Counts in the order of listed GTS
DP	Average Sample Read Depth
AA	chimpAllele
FG	functionGVS
HGVS_CDNA_VAR	HGVS Coding DNA Variant

HGVS_PROTEIN_VAR	HGVS Protein Variant
CDS_SIZES	Coding DNA Sizes
PH	polyPhen2 result including prediction class and score
CP	scorePhastCons
CG	consScoreGERP
GL	geneList
GS	granthamScore
CA	clinicalAssociation
EXOME_CHIP	Whether a SNP is on the Illumina HumanExome Chip
EA_AGE	Estimated Variant Age in kilo years for the European American Population
AA_AGE	Estimated Variant Age in kilo years for the African American Population
GRCh38_POSITION	GRCh38 chromosomal position liftover from the original GRCh37 chromosomal position. A value of -1 means the GRCh37 position can not be mapped to the GRCh38 build.

ExAC Columns

AC	Allele count in genotypes, for each ALT allele, in the same order as listed
AC_AFR	African/African American Allele Counts
AC_AMR	American Allele Counts
AC_Adj	Adjusted Allele Counts
AC_EAS	East Asian Allele Counts
AC_FIN	Finnish Allele Counts
AC_Hemi	Adjusted Hemizygous Counts
AC_Het	Adjusted Heterozygous Counts
AC_Hom	Adjusted Homozygous Counts
AC_NFE	Non-Finnish European Allele Counts
AC_OTH	Other Allele Counts
AC_SAS	South Asian Allele Counts
AF	Allele Frequency, for each ALT allele, in the same order as listed
AN	Total number of alleles in called genotypes
AN_AFR	African/African American Chromosome Count
AN_AMR	American Chromosome Count
AN_Adj	Adjusted Chromosome Count
AN_EAS	East Asian Chromosome Count
AN_FIN	Finnish Chromosome Count
AN_NFE	Non-Finnish European Chromosome Count
AN_OTH	Other Chromosome Count
AN_SAS	South Asian Chromosome Count
BaseQRankSum	Z-score from Wilcoxon rank sum test of Alt Vs. Ref base qualities
CCC	Number of called chromosomes
ClippingRankSum	Z-score From Wilcoxon rank sum test of Alt vs. Ref number of hard clipped bases
DB	dbSNP Membership
DP	Approximate read depth; some reads may have been filtered

DS	Were any of the samples downsampled?
END	Stop position of the interval
FS	Phred-scaled p-value using Fisher's exact test to detect strand bias
GQ_MEAN	Mean of all GQ values
GQ_STDDEV	Standard deviation of all GQ values
HWP	P value from test of Hardy Weinberg Equilibrium
HaplotypeScore	Consistency of the site with at most two segregating haplotypes
Hemi_AFR	African/African American Hemizygous Counts
Hemi_AMR	American Hemizygous Counts
Hemi_EAS	East Asian Hemizygous Counts
Hemi_FIN	Finnish Hemizygous Counts
Hemi_NFE	Non-Finnish European Hemizygous Counts
Hemi_OTH	Other Hemizygous Counts
Hemi_SAS	South Asian Hemizygous Counts
Het_AFR	African/African American Heterozygous Counts
Het_AMR	American Heterozygous Counts
Het_EAS	East Asian Heterozygous Counts
Het_FIN	Finnish Heterozygous Counts
Het_NFE	Non-Finnish European Heterozygous Counts
Het_OTH	Other Heterozygous Counts
Het_SAS	South Asian Heterozygous Counts
Hom_AFR	African/African American Homozygous Counts
Hom_AMR	American Homozygous Counts
Hom_EAS	East Asian Homozygous Counts
Hom_FIN	Finnish Homozygous Counts
Hom_NFE	Non-Finnish European Homozygous Counts
Hom_OTH	Other Homozygous Counts
Hom_SAS	South Asian Homozygous Counts
InbreedingCoeff	Inbreeding coefficient as estimated from the genotype likelihoods per-sample when compared against the Hardy-Weinberg expectation
MLEAC	Maximum likelihood expectation (MLE) for the allele counts (not necessarily the same as the AC), for each ALT allele, in the same order as listed
MLEAF	Maximum likelihood expectation (MLE) for the allele frequency (not necessarily the same as the AF), for each ALT allele, in the same order as listed
MQ	RMS Mapping Quality
MQ0	Total Mapping Quality Zero Reads
MQRankSum	Z-score From Wilcoxon rank sum test of Alt vs. Ref read mapping qualities
NCC	Number of no-called samples
NEGATIVE_TRAIN_SITE	This variant was used to build the negative training set of bad variants
POSITIVE_TRAIN_SITE	This variant was used to build the positive training set of good variants
QD	Variant Confidence/Quality by Depth
ReadPosRankSum	Z-score from Wilcoxon rank sum test of Alt vs. Ref read position bias

VQSLOD	Log odds ratio of being a true variant versus being false under the trained gaussian mixture model
culprit	The annotation which was the worst performing in the Gaussian mixture model, likely the reason why the variant was filtered out
DP_HIST	Histogram for DP*
GQ_HIST	Histogram for GQ**
DOUBLETON_DIST	Euclidean distance of carriers of doubletons
AC_MALE	Allele count among males
AC_FEMALE	Allele count among females
AN_MALE	Allele number among males
AN_FEMALE	Allele number among females
AC_CONSANGUINEOUS	Allele count among individuals with F > 0.05
AN_CONSANGUINEOUS	Allele number among individuals with F > 0.05
Hom_CONSANGUINEOUS	Homozygote count among individuals with F > 0.05
CSQ	Consequence annotations from Ensembl VEP***
AC_POPMAX	AC in the population with the max AF
AN_POPMAX	AN in the population with the max AF
POPMAX	Population with max AF
clinvar_measureset_id	Clinvar Measureset ID
clinvar_conflicted	Clinvar Conflicted Status
clinvar_pathogenic	Clinvar Pathogenic Status
clinvar_mut	Clinvar MUT Flag (is disease allele REF?)
K1_RUN	Number of ensuing single nucleotide repeats.
K2_RUN	Number of ensuing di-nucleotide repeats.
K3_RUN	Number of ensuing tri-nucleotide repeats.
ESP_AF_POPMAX	Max allele frequency across populations in ESP
ESP_AF_GLOBAL	Overall allele frequency in ESP
ESP_AC	Allele Count in ESP
KG_AF_POPMAX	Max allele frequency across populations in 1000 Genomes
KG_AF_GLOBAL	Overall allele frequency in 1000 Genomes
KG_AC	Allele Count in 1000 Genomes

* Mids:
2.5|7.5|12.5|17.5|22.5|27.5|32.5|37.5|42.5|47.5|52.5|57.5|62.5|67.5|72.5|77.5|82.5|87.5|92.5|97.5

** Mids:
2.5|7.5|12.5|17.5|22.5|27.5|32.5|37.5|42.5|47.5|52.5|57.5|62.5|67.5|72.5|77.5|82.5|87.5|92.5|97.5

*** Format:
Allele|Consequence|IMPACT|SYMBOL|Gene|Feature_type|Feature|BIOTYPE|EXON|INTRON|HGVSc|HGVSp|cDNA_position|CDS_position|Protein_position|Amino_acids|Codons|Existing_variation|ALLELE_

NUM|DISTANCE|STRAND|VARIANT_CLASS|MINIMISED|SYMBOL_SOURCE|HGNC_ID|CANONICAL|TSL|CCDS|ENSP|SWISSPROT|TREMBL|UNIPARC|SIFT|PolyPhen|DOMAINS|HGVS_OFFSET|GMAF|AFR_MAF|AMR_MAF|ASN_MAF|EAS_MAF|EUR_MAF|SAS_MAF|AA_MAF|EA_MAF|CLIN_SIG|SOMATIC|PHENO|PUBMED|MOTIF_NAME|MOTIF_POS|HIGH_INF_POS|MOTIF_SCORE_CHANGE|LoF_info|LoF_flags|LoF_filter|LoF|context|ancestral

ClinVar Columns

RS	dbSNP ID (i.e. rs number)
RSPOS	Chr position reported in dbSNP
RV	RS orientation is reversed
VP	Variation Property. Documentation is at ftp://ftp.ncbi.nlm.nih.gov/snp/specs/dbSNP_BitField_latest.pdf
GENEINFO	Pairs each of gene symbol:gene id. The gene symbol and id are delimited by a colon (:) and each pair is delimited by a vertical bar (|)
dbSNPBuildID	First dbSNP Build for RS
SAO	Variant Allele Origin: 0 - unspecified, 1 - Germline, 2 - Somatic, 3 - Both
SSR	Variant Suspect Reason Codes (may be more than one value added together) 0 - unspecified, 1 - Paralog, 2 - byEST, 4 - oldAlign, 8 - Para_EST, 16 - 1kg_failed, 1024 - other
WGT	Weight, 00 - unmapped, 1 - weight 1, 2 - weight 2, 3 - weight 3 or more
VC	Variation Class
PM	Variant is Precious(Clinical,Pubmed Cited)
TPA	Provisional Third Party Annotation(TPA) (currently rs from PHARMGKB who will give phenotype data)
PMC	Links exist to PubMed Central article
S3D	Has 3D structure - SNP3D table
SLO	Has SubmitterLinkOut - From SNP->SubSNP->Batch.link_out
NSF	Has non-synonymous frameshift A coding region variation where one allele in the set changes all downstream amino acids. FxnClass = 44
NSM	Has non-synonymous missense A coding region variation where one allele in the set changes protein peptide. FxnClass = 42
NSN	Has non-synonymous nonsense A coding region variation where one allele in the set changes to STOP codon (TER). FxnClass = 41
REF	Has reference A coding region variation where one allele in the set is identical to the reference sequence. FxnCode = 8
SYN	Has synonymous A coding region variation where one allele in the set does not change the encoded amino acid. FxnCode = 3
U3	In 3' UTR Location is in an untranslated region (UTR). FxnCode = 53
U5	In 5' UTR Location is in an untranslated region (UTR). FxnCode = 55
ASS	In acceptor splice site FxnCode = 73
DSS	In donor splice-site FxnCode = 75
INT	In Intron FxnCode = 6
R3	In 3' gene region FxnCode = 13
R5	In 5' gene region FxnCode = 15
OTH	Has other variant with exactly the same set of mapped positions on NCBI refernce assembly.

CFL	Has Assembly conflict. This is for weight 1 and 2 variant that maps to different chromosomes on different assemblies.
ASP	Is Assembly specific. This is set if the variant only maps to one assembly
MUT	Is mutation (journal citation, explicit fact): a low frequency variation that is cited in journal and other reputable sources
VLD	Is Validated. This bit is set if the variant has 2+ minor allele count based on frequency or genotype data.
G5A	>5% minor allele frequency in each and all populations
G5	>5% minor allele frequency in 1+ populations
HD	Marker is on high density genotyping kit (50K density or greater). The variant may have phenotype associations present in dbGaP.
GNO	Genotypes available. The variant has individual genotype (in SubInd table).
KGPhase1	1000 Genome phase 1 (incl. June Interim phase 1)
KGPhase3	1000 Genome phase 3
CDA	Variation is interrogated in a clinical diagnostic assay
LSD	Submitted from a locus-specific database
MTP	Microattribution/third-party annotation(TPA:GWAS,PAGE)
OM	Has OMIM/OMIA
NOC	Contig allele not present in variant allele list. The reference sequence allele at the mapped position is not present in the variant allele list, adjusted for orientation.
WTD	Is Withdrawn by submitter If one member ss is withdrawn by submitter, then this bit is set. If all member ss' are withdrawn, then the rs is deleted to SNPHistory
NOV	Rs cluster has non-overlapping allele sets. True when rs set has more than 2 alleles from different submissions and these sets share no alleles in common.
CAF	An ordered, comma delimited list of allele frequencies based on 1000Genomes, starting with the reference allele followed by alternate alleles as ordered in the ALT column. Where a 1000Genomes alternate allele is not in the dbSNPs alternate allele set, the allele is added to the ALT column. The minor allele is the second largest value in the list, and was previuosly reported in VCF as the GMAF. This is the GMAF reported on the RefSNP and EntrezSNP pages and VariationReporter
COMMON	RS is a common SNP. A common SNP is one that has at least one 1000Genomes population with a minor allele of frequency >= 1% and for which 2 or more founders contribute to that minor allele frequency.
CLNHGVS	Variant names from HGVS. The order of these variants corresponds to the order of the info in the other clinical INFO tags.
CLNALLE	Variant alleles from REF or ALT columns. 0 is REF, 1 is the first ALT allele, etc. This is used to match alleles with other corresponding clinical (CLN) INFO tags. A value of -1 indicates that no allele was found to match a corresponding HGVS allele name.
CLNSRC	Variant Clinical Chanels
CLNORIGIN	Allele Origin. One or more of the following values may be added: 0 - unknown; 1 - germline; 2 - somatic; 4 - inherited; 8 - paternal; 16 - maternal; 32 - de-novo; 64 - biparental; 128 - uniparental; 256 - not-tested; 512 - tested-inconclusive; 1073741824 - other
CLNSRCID	Variant Clinical Channel IDs
CLNSIG	Variant Clinical Significance, 0 - Uncertain significance, 1 - not provided, 2 - Benign, 3 - Likely benign, 4 - Likely pathogenic, 5 - Pathogenic, 6 - drug response, 7 - histocompatibility, 255 - other
CLNDSDB	Variant disease database name

CLNDSDBID	Variant disease database ID
CLNDBN	Variant disease name
CLNREVSTAT	no_assertion - No assertion provided, no_criteria - No assertion criteria provided, single - Criteria provided single submitter, mult - Criteria provided multiple submitters no conflicts, conf - Criteria provided conflicting interpretations, exp - Reviewed by expert panel, guideline - Practice guideline
CLNACC	Variant Accession and Versions

dbNSFP Columns

chr	chromosome number
pos(1-based)	physical position on the chromosome as to hg19 (1-based coordinate)
ref	reference nucleotide allele (as on the + strand)
alt	alternative nucleotide allele (as on the + strand)
aaref	reference amino acid. if the variant is a splicing site SNP (2bp on each end of an intron)
aaalt	alternative amino acid. if the variant is a splicing site SNP (2bp on each end of an intron)
rs_dbSNP141	rs number from dbSNP 141
hg18_pos(1-based)	physical position on the chromosome as to hg18 (1-based coordinate)
hg38_chr	chromosome as to hg38, "." means the same as in the chr column
hg38_pos	physical position on the chromosome as to hg38 (1-based coordinate)
genename	gene name; if the NScan be assigned to multiple genes, gene names are separated by ";"
Uniprot_acc	Uniprot accession number. Multiple entries separated by ";".
Uniprot_id	Uniprot ID number. Multiple entries separated by ";".
Uniprot_aapos	amino acid position as to Uniprot. Multiple entries separated by ";".
Interpro_domain	domain or conserved site on which the variant locates. Domain annotations come from Interpro database. The number in the brackets following a specific domain is the count of times Interpro assigns the variant position to that domain, typically coming from different predicting databases. Multiple entries separated by ";".
cds_strand	coding sequence (CDS) strand (+ or -)
refcodon	reference codon
SLR_test_statistic	SLR test statistic for testing natural selection on codons. A negative value indicates negative selection, and a positive value indicates positive selection. Larger magnitude of the value suggests stronger evidence.

codonpos	position on the codon (1, 2 or 3)
fold-degenerate	degenerate type (0, 2 or 3)
Ancestral_allele	Ancestral allele (based on 1000 genomes reference data). The following comes from its original README file: ACTG - high-confidence call, ancestral state supproted by the other two sequences actg - low-confindence call, ancestral state supported by one sequence only N - failure, the ancestral state is not supported by any other sequence =- - the extant species contains an insertion at this postion . - no coverage in the alignment
Ensembl_geneid	Ensembl gene id
Ensembl_transcriptid	Ensembl transcript ids (separated by ";")
aapos	amino acid position as to the protein -1 if the variant is a splicing site SNP (2bp on each end of an intron)
aapos_SIFT	ENSP id and amino acid positions corresponding to SIFT scores. Multiple entries separated by ";"
aapos_FATHMM	ENSP id and amino acid positions corresponding to FATHMM scores. Multiple entries separated by ";"
SIFT_score	SIFT score (SIFTori). Scores range from 0 to 1. The smaller the score the more likely the SNP has damaging effect. Multiple scores separated by ";".
SIFT_converted_rankscore	SIFTori scores were first converted to SIFTnew=1-SIFTori, then ranked among all SIFTnew scores in dbNSFP. The rankscore is the ratio of the rank the SIFTnew score over the total number of SIFTnew scores in dbNSFP. If there are multiple scores, only the most damaging (largest) rankscore is presented. The rankscores range from 0.02654 to 0.87932.
SIFT_pred	If SIFTori is smaller than 0.05 (rankscore>0.55) the corresponding NS is predicted as "D(amaging)"; otherwise it is predicted as "T(olerated)". Multiple predictions separated by ";"
Polyphen2_HDIV_score	Polyphen2 score based on HumDiv, i.e. hdiv_prob. The score ranges from 0 to 1. Multiple entries separated by ";".
Polyphen2_HDIV_rankscore	Polyphen2 HDIV scores were first ranked among all HDIV scores in dbNSFP. The rankscore is the ratio of the rank the score over the total number the scores in dbNSFP. If there are multiple scores, only the most damaging (largest) of rankscore is presented. The scores range from 0.02656 to 0.89917.

Polyphen2_HDIV_pred	Polyphen2 prediction based on HumDiv, "D" ("porobably damaging", HDIV score in [0.957,1] or rankscore in [0.52996,0.89917]), "P" ("possibly damaging", HDIV score in [0.453,0.956] or rankscore in [0.34412,0.52842]) and "B" ("benign", HDIV score in [0,0.452] or rankscore in [0.02656,0.34399]). Score cutoff for binary classification is 0.5 for HDIV score or 0.35411 for rankscore, i.e. the prediction is neutral if the HDIV score is smaller than 0.5 (rankscore is smaller than 0.35411), and "deleterious" if the HDIV score is larger than 0.5 (rankscore is larger than 0.35411). Multiple entries are separated by ";".
Polyphen2_HVAR_score	Polyphen2 score based on HumVar, i.e. hvar_prob. The score ranges from 0 to 1. Multiple entries separated by ";".
Polyphen2_HVAR_rankscore	Polyphen2 HVAR scores were first ranked among all HVAR scores in dbNSFP. The rankscore is the ratio of the rank the score over the total number of the scores in dbNSFP. If there are multiple scores, only the most damaging (largest) rankscore is presented. The scores range from 0.01281 to 0.9711.
Polyphen2_HVAR_pred	Polyphen2 prediction based on HumVar, "D" ("porobably damaging", HVAR score in [0.909,1] or rankscore in [0.62955,0.9711]), "P" ("possibly damaging", HVAR in [0.447,0.908] or rankscore in [0.44359,0.62885]) and "B" ("benign", HVAR score in [0,0.446] or rankscore in [0.01281,0.44315]). Score cutoff for binary classification is 0.5 for HVAR score or 0.45998 for rankscore, i.e. the prediction is "neutral" if the HVAR score is smaller than 0.5 (rankscore is smaller than 0.45998), and "deleterious" if the HVAR score is larger than 0.5 (rankscore is larger than 0.45998). Multiple entries are separated by ";".
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 0.00166 to 0.85682.
LRT_pred	LRT prediction, D(eleterious), N(eutral) or U(nknown), which is not solely determined by the score.
MutationTaster_score	MutationTaster p-value (MTori), ranges from 0 to 1.

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. The rankscore is the ratio of the rank of the score over the total number of MTnew scores in dbNSFP. The scores range from 0.09067 to 0.80722.
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 MTori and 0.31655 for the rankscore.
MutationAssessor_score	MutationAssessor functional impact combined score (MAori). The score ranges from -5.135 to 6.49 in dbNSFP. Please refer to Reva et al. (2011) Nucl. Acids Res. 39(17): e118 for details.
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 0.92924, 0.51945 and 0.19692, respectively.
FATHMM_score	FATHMM default score (weighted for human inherited-disease mutations with Disease Ontology) (FATHMMori). Scores range from -18.09 to 11.0. Multiple scores separated by ";" Please refer to Shihab et al. (2013) Human Mutation 34(1): 57-65 for details.
FATHMM_rankscore	FATHMMori scores were ranked among all FATHMMori scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of FATHMMori 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 <=0.81415) the corresponding NS is predicted as "D(AMAGING)"; otherwise it is predicted as "T(OLERATED)". Multiple predictions separated by ";"

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 0.83357.
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 0.82268.
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.
VEST3_score	VEST 3.0 score. Score ranges from 0 to 1. The larger the score the more likely the mutation may cause functional change. In case there are multiple scores for the same variant, the largest score (most damaging) is presented. Please refer to Carter et al., (2013) BMC Genomics. 14(3) 1-16 for details. Please note this score is free for non-commercial use. For more details please refer to http: //wiki.chasmsoftware.org/index.php/SoftwareLicense. Commercial users should contact the Johns Hopkins Technology Transfer office.

VEST3_rankscore	VEST3 scores were ranked among all VEST3 scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of VEST3 scores in dbNSFP. The scores range from 0 to 1. Please note VEST score is free for non-commercial use. For more details please refer to http://wiki.chasmsoftware.org/index.php/SoftwareLicense. Commercial users should contact the Johns Hopkins Technology Transfer office.
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 ";". Details can be found in DOI: 10.1371/journal.pone.0046688
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.
PROVEAN_pred	If PROVEANori <= -2.5 (rankscore>=0.59) the corresponding NS is predicted as "D(amaging)"; otherwise it is predicted as "N(eutral)". Multiple predictions separated by ";"
CADD_raw	CADD raw score for funtional prediction of a SNP. Please refer to Kircher et al. (2014) Nature Genetics 46(3): 310-5 for details. The larger the score the more likely the SNP has damaging effect. Please note the following copyright statement for CADD: CADD scores (http//cadd.gs.washington.edu/) are Copyright 2013 University of Washington and Hudson-Alpha Institute for Biotechnology (all rights reserved) but are freely available for all academic, non-commercial applications. For commercial licensing information contact Jennifer McCullar (mccullaj@uw.edu).
CADD_raw_rankscore	CADD raw scores were ranked among all CADD raw scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of CADD raw scores in dbNSFP. Please note the following copyright statement for CADD: "CADD scores (http//cadd.gs.washington.edu/) are Copyright 2013 University of Washington and Hudson-Alpha Institute for Biotechnology (all rights reserved) but are freely available for all academic, non-commercial applications. For commercial licensing information contact Jennifer McCullar (mccullaj@uw.edu)."
CADD_phred	CADD phred-like score. This is phred-like rank score based on whole genome CADD raw scores. Please refer to Kircher et al. (2014) Nature Genetics 46(3):310-5 for details. The larger the score the more likely the SNP has damaging effect. Please note the following copyright statement for CADD: "CADD

	scores (http//cadd.gs.washington.edu/) are Copyright 2013 University of Washington and Hudson-Alpha Institute for Biotechnology (all rights reserved) but are freely available for all academic, non-commercial applications. For commercial licensing information contact Jennifer McCullar (mccullaj@uw.edu)."
GERP++_NR	GERP++ neutral rate
GERP++_RS	GERP++ RS score, the larger the score, the more conserved the site.
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.
phyloP46way_primate	phyloP (phylogenetic p-values) conservation score based on the multiple alignments of 10 primate genomes (including human). The larger the score, the more conserved the site.
phyloP46way_primate_rankscore	phyloP46way_primate scores were ranked among all phyloP46way_primate scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of phyloP46way_primate scores in dbNSFP.
phyloP46way_placental	phyloP (phylogenetic p-values) conservation score based on the multiple alignments of 33 placental mammal genomes (including human). The larger the score, the more conserved the site.
phyloP46way_placental_rankscore	phyloP46way_placental scores were ranked among all phyloP46way_placental scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of phyloP46way_placental 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.
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.
phastCons46way_primate	phastCons conservation score based on the multiple alignments of 10 primate genomes (including human). The larger the score, the more conserved the site.
phastCons46way_primate_rankscore	phastCons46way_primate scores were ranked among all phastCons46way_primate scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of phastCons46way_primate scores in dbNSFP.
phastCons46way_placental	phastCons conservation score based on the multiple alignments of 33 placental mammal genomes (including human). The larger the score, the more conserved the site.

phastCons46way_placental_rankscore	phastCons46way_placental scores were ranked among all phastCons46way_placental scores in dbNSFP. The rankscore is the ratio of the rank of the score over the total number of phastCons46way_placental 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.
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.
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.
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.
LRT_Omega	estimated nonsynonymous-to-synonymous-rate ratio (Omega, reported by LRT)
UniSNP_ids	rs numbers from UniSNP, which is a cleaned version of dbSNP build 129, in format: rs number1;rs number2;…
1000Gp1_AC	Alternative allele counts in the whole 1000 genomes phase 1 (1000Gp1) data.
1000Gp1_AF	Alternative allele frequency in the whole 1000Gp1 data.
1000Gp1_AFR_AC	Alternative allele counts in the 1000Gp1 African descendent samples.
1000Gp1_AFR_AF	Alternative allele frequency in the 1000Gp1 African descendent samples.
1000Gp1_EUR_AC	Alternative allele counts in the 1000Gp1 European descendent samples.
1000Gp1_EUR_AF	Alternative allele frequency in the 1000Gp1 European descendent samples.
1000Gp1_AMR_AC	Alternative allele counts in the 1000Gp1 American descendent samples.
1000Gp1_AMR_AF	Alternative allele frequency in the 1000Gp1 American descendent samples.
1000Gp1_ASN_AC	Alternative allele counts in the 1000Gp1 Asian descendent samples.
1000Gp1_ASN_AF	Alternative allele frequency in the 1000Gp1 Asian descendent samples.

ESP6500_AA_AF	Alternative allele frequency in the Afrian American samples of the NHLBI GO Exome Sequencing Project (ESP6500 data set).
ESP6500_EA_AF	Alternative allele frequency in the European American samples of the NHLBI GO Exome Sequencing Project (ESP6500 data set).
ARIC5606_AA_AC	Alternative allele counts in 2403 exomes of African Americans from the Atherosclerosis Risk in Communities Study (ARIC) cohort study.
ARIC5606_AA_AF	Alternative allele frequency of 2403 exomes of African Americans from the Atherosclerosis Risk in Communities Study (ARIC) cohort study.
ARIC5606_EA_AC	Alternative allele counts in 3203 exomes of European Americans from the Atherosclerosis Risk in Communities Study (ARIC) cohort study.
ARIC5606_EA_AF	Alternative allele frequency of 3203 exomes of European Americans from the Atherosclerosis Risk in Communities Study (ARIC) cohort study.
ExAC_AC	Allele count in total ExAC samples (~60,706 unrelated individuals)
ExAC_AF	Allele frequency in total ExAC samples
ExAC_Adj_AC	Adjusted Alt allele counts (DP >= 10 & GQ >= 20) in total ExAC samples
ExAC_Adj_AF	Adjusted Alt allele frequency (DP >= 10 & GQ >= 20) in total ExAC samples
ExAC_AFR_AC	Adjusted Alt allele counts (DP >= 10 & GQ >= 20) in African & African American ExAC samples
ExAC_AFR_AF	Adjusted Alt allele frequency (DP >= 10 & GQ >= 20) in African & African American ExAC amples
ExAC_AMR_AC	Adjusted Alt allele counts (DP >= 10 & GQ >= 20) in American ExAC samples
ExAC_AMR_AF	Adjusted Alt allele frequency (DP >= 10 & GQ >= 20) in American ExAC samples
ExAC_EAS_AC	Adjusted Alt allele counts (DP >= 10 & GQ >= 20) in East Asian ExAC samples
ExAC_EAS_AF	Adjusted Alt allele frequency (DP >= 10 & GQ >= 20) in East Asian ExAC samples
ExAC_FIN_AC	Adjusted Alt allele counts (DP >= 10 & GQ >= 20) in Finnish ExAC samples
ExAC_FIN_AF	Adjusted Alt allele frequency (DP >= 10 & GQ >= 20) in Finnish ExAC samples

ExAC_NFE_AC	Adjusted Alt allele counts (DP >= 10 & GQ >= 20) in Non-Finnish European ExAC samples
ExAC_NFE_AF	Adjusted Alt allele frequency (DP >= 10 & GQ >= 20) in Non-Finnish European ExAC samples
ExAC_SAS_AC	Adjusted Alt allele counts (DP >= 10 & GQ >= 20) in South Asian ExAC samples
ExAC_SAS_AF	Adjusted Alt allele frequency (DP >= 10 & GQ >= 20) in South Asian ExAC samples
clinvar_rs	rs number from the clinvar data set
clinvar_clnsig	clinical significance as to the clinvar data set 2 - Benign, 3 - Likely benign, 4 - Likely pathogenic, 5 - Pathogenic, 6 - drug response, 7 - histocompatibility. A negative score means the the score is for the ref allele
clinvar_trait	the trait/disease the clinvar_clnsig referring to
COSMIC_ID	ID of the SNV at the COSMIC (Catalogue Of Somatic Mutations In Cancer) database
COSMIC_CNT	number of samples having this SNV in the COSMIC database