CADD v1.7 (minor release)

Why a new model?

Since the CADD v1.6 release in 2021 new methods for the identification of functional regions and the assessment of impact of variation in the human genome have been developed. This includes the development of deep learning methods for scoring coding and non-coding changes, the availability of substantially deeper mammalian sequence alignments and the development of several computational predictors of sequence constraint. In CADD version 1.7, we therefore added new features to improve CADD scores for certain variant effects. This boosts the overall performance of CADD and brings new developments to the community. The full list of new annotations can be found in the section ***New annotations*** in Supplement 1 of this document**.**

What else changed?

We changed the Ensembl Variant Effect Predictor (VEP) version as well as the version of the respective annotation builds from Ensembl release 95 to 110.

Model training parameters

We have trained models in the same way as the previous release. The logistic regression used L2 penalty with C = 1 and training was terminated after thirteen L-BFGS iterations using the *sklearn* library.

Genome-wide availability of CADD scores

We are only supporting the major chromosomes of GRCh38 (1-22, X & Y). Due to the limited or questionable coverage of the annotations used for alternative haplotypes, unplaced contigs and the mitochondrial genome, we stopped supporting variants located on those genomic positions of GRCh37 in version 1.4. Similar considerations apply to chromosome Y, for which we would like to caution whether scores are comparable to those of other chromosomes.

Performance of CADD v1.7 in comparison to previous versions

Generally, the new models are highly similar to the previous release with a Spearman correlation of CADD scores for 100,000 variants drawn randomly from the genome of 0.946 between CADD GRCh38-v1.6 and CADD GRCh38-v1.7. Accordingly, CADD v1.7 models perform very similar to previous models in distinguishing known pathogenic variants (ClinVar) from common variants (gnomAD) throughout the genome (Fig. 1). The improvement in CADD v1.7 becomes apparent when limiting the prediction to variant categories like missense or 3'UTR, for which the latest release has added specific annotations. Here, CADD v1.7 outperforms the previous version as well domain specific models ESM-1v and APARENT2. On specialized datasets like Deep Mutational Scanning of human proteins (DMS) and experimental reporter assays of regulatory variant activity readouts from saturation mutagenesis MPRAs, CADD v1.7 shows higher Pearson correlation as compared to the old CADD v1.6 model (Fig. 2 and Fig. 3).

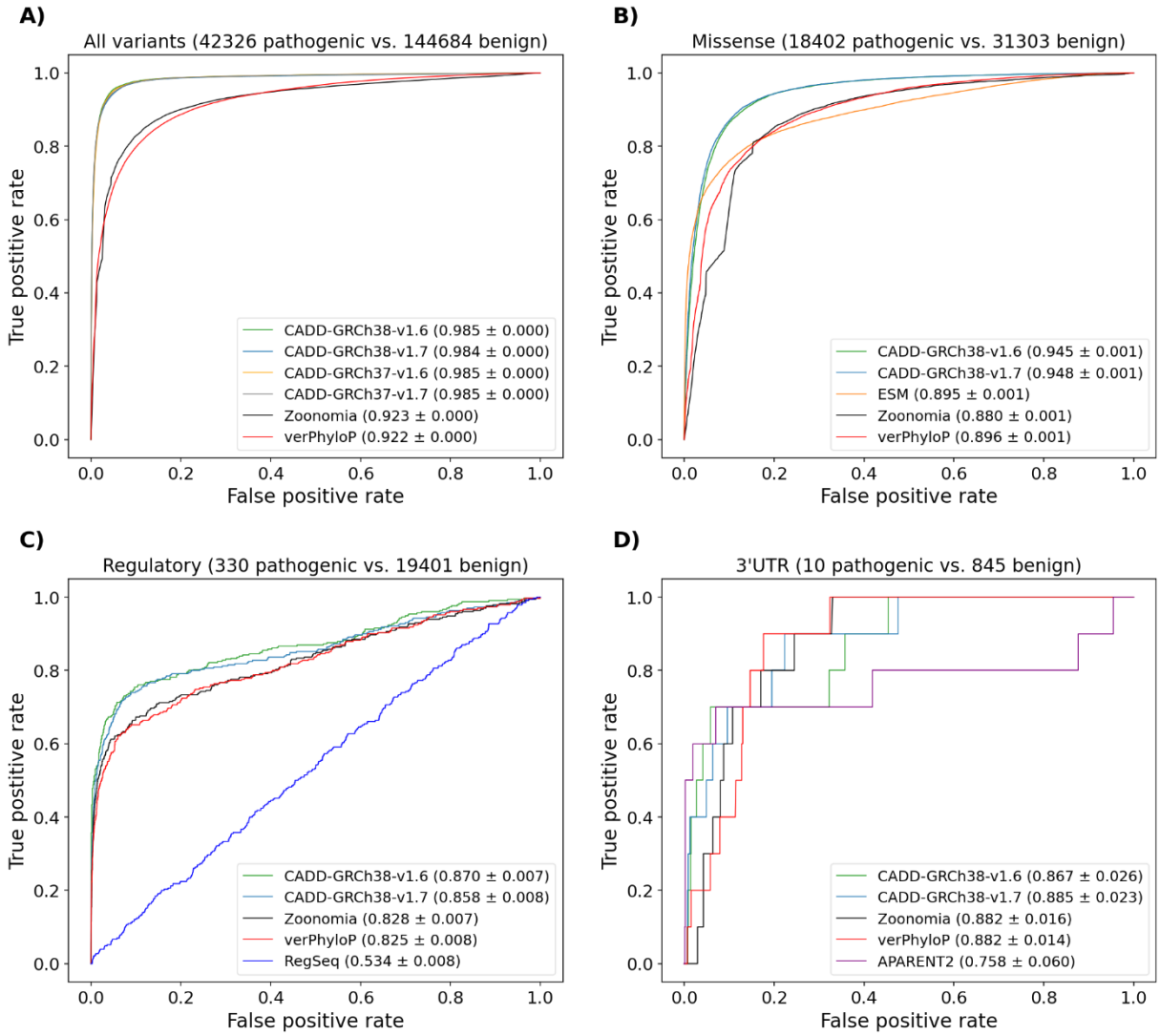


Fig. 1: ClinVar pathogenic vs. common gnomAD variants for different variant effect classes*. CADD v1.7 has comparable performance with CADD v1.6 in terms of AUROC in distinguishing between pathogenic variants from ClinVar and common population variants (gnomAD, MAF > 0.05) for different types of SNV classes. All tested CADD versions are better than conservation scores when tested on all SNVs (A). CADD also outperforms specialized missense score ESM (B), regulatory sequence CNN derived scores (C), and the specialized 3'UTR score APARENT2 (D).*

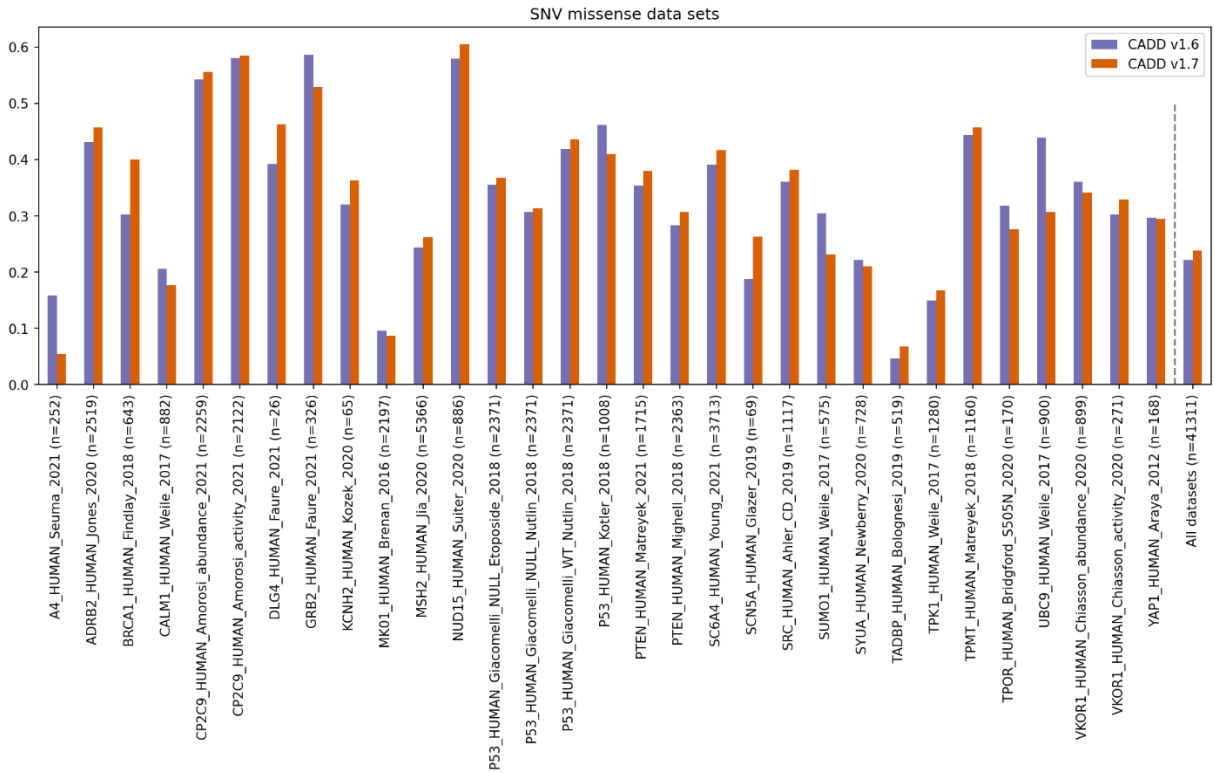


Fig. 2: Spearman correlations of CADD v1.6 and CADD v1.7 for missense variants with experimental effect scores form the ProteinGym. *The average correlation over all datasets increased for CADD v1.7, where among others, ESM protein language model scores for missense variants were included. For more information on the datasets, refer to the CADD v1.7 publication (https://doi.org/10.1093/nar/gkad989).*

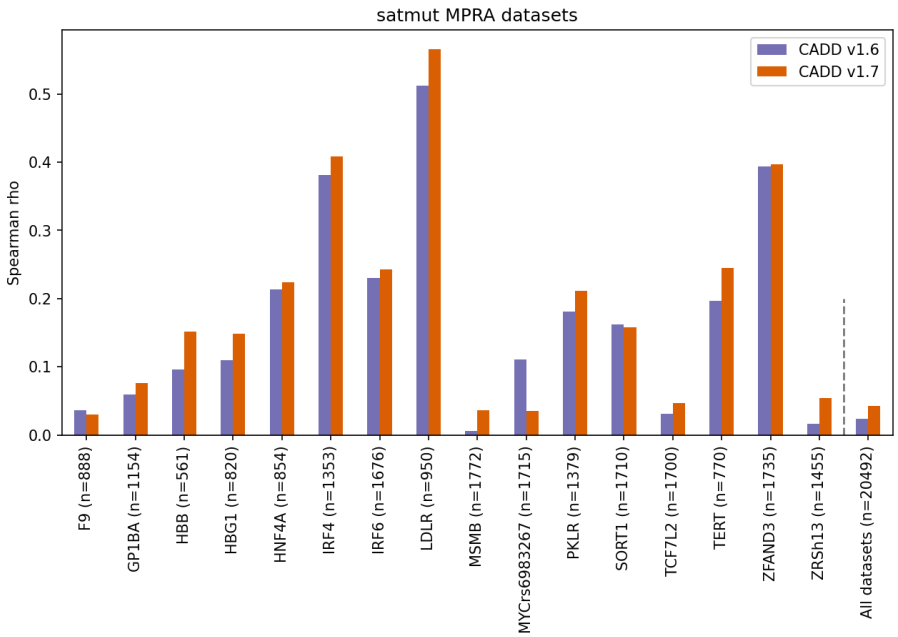
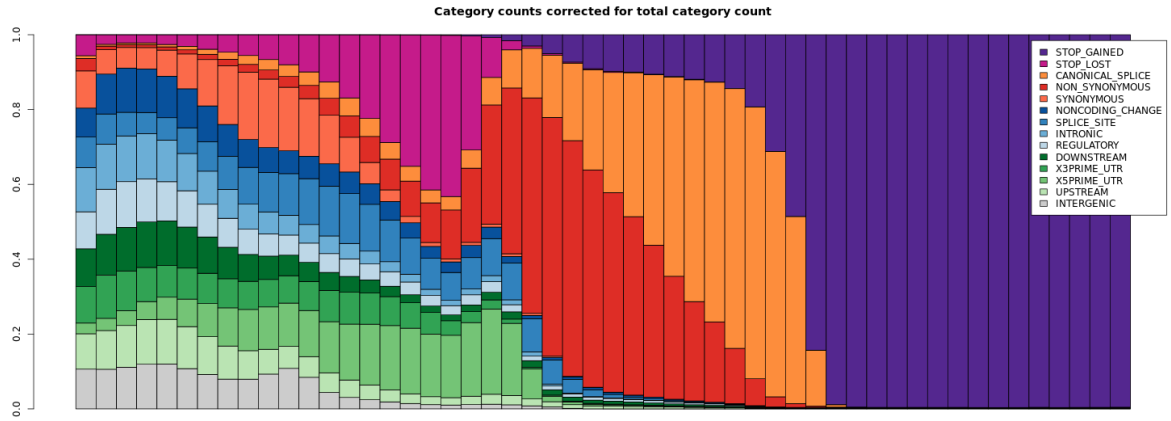
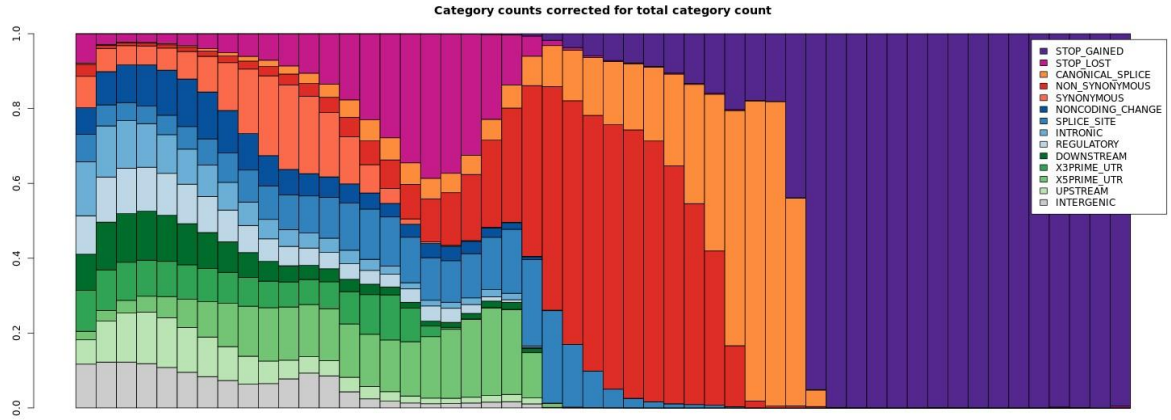


Fig. 3: Spearman correlation of expression effects from saturation mutagenesis MPRA experiments (SNV only). *Correlation of prediction outputs of CADD v1.6 and CADD v1.7 models with measured absolute expression effects. We used the absolute value of the MPRA effect (Kircher M et al. 2019) because CADD is not expected to predict the effect directionality. The average correlation over all datasets increased for CADD v1.7, where among others sequence-based CNN (RegSeq) features were included. For more information on the datasets, refer to the CADD v1.7 publication (https://doi.org/10.1093/nar/gkad989).*

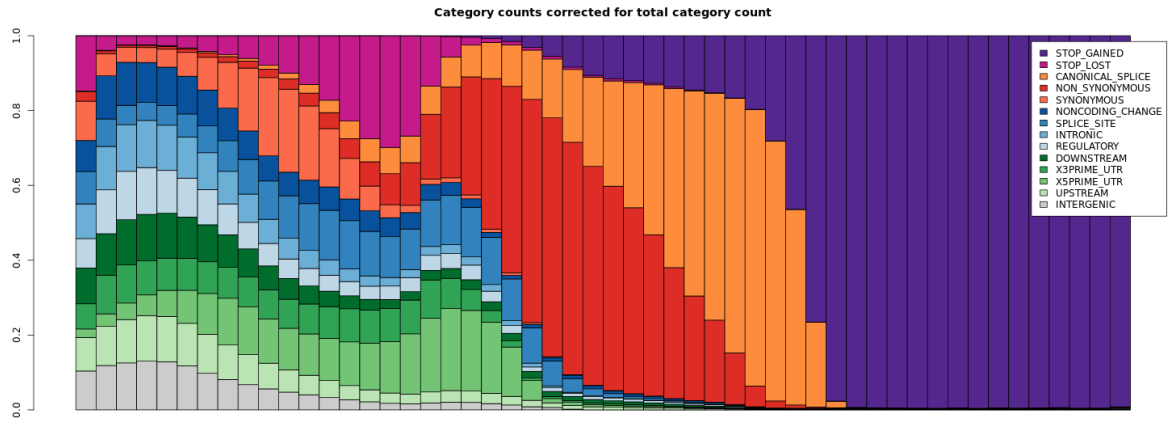
Category distribution CADDv1.7-GRCh38



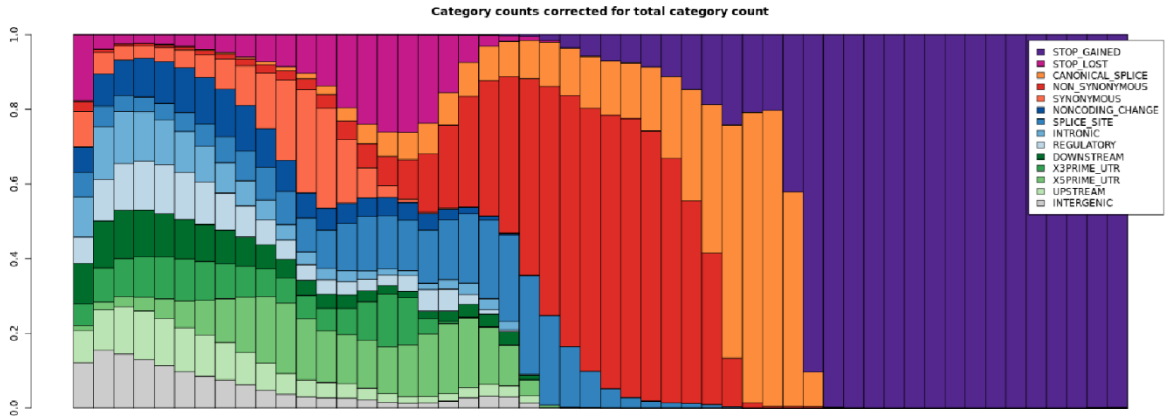
Category distribution CADDv1.6-GRCh38



Category distribution CADDv1.7-GRCh37



Category distribution CADDv1.6-GRCh37



Supplement 1: CADD v1.7 annotation changes compared to CADD v1.6

New annotations

Meta AI Evolutionary Scale Model for variant effects in protein coding sequences (https://github.com/facebookresearch/esm). Details of how annotations were derived from these models are available in the accompanying CADD v1.7 publication (https://doi.org/10.1093/nar/gkad989).

- EsmScoreMissense: ESM-1v derived score for missense variants
- EsmScoreInFrame: ESM-1v derived score for infame InDels
- EsmScoreFrameshift: ESM-1v derived score for InDels frameshift and stop gain

CNN trained on open chromatin sequences of multiple tissues. Details of the model are described in the accompanying publication (https://doi.org/10.1093/nar/gkad989). The following annotations are retrieved from a regulatory sequence model of 7 cell-lines with:

- RegSeq0: Regulatory sequence model HEK293T
- RegSeq1: Regulatory sequence model K562
- RegSeq2: Regulatory sequence model HepG2
- RegSeq3: Regulatory sequence model HeLa-S3
- RegSeq4: Regulatory sequence model MC-7
- RegSeq5: Regulatory sequence model iPS DF 19.11
- RegSeg6: Regulatory sequence model GM23338
- RegSeg7: Regulatory sequence model GC-matched background

Other nucleotide-level scores derive from Aparent2 (https://doi.org/10.1186/s13059-022-02799-4), Zoonomia (https://doi.org/10.1126/science.abn3943) and Roulette mutability scores (https://doi.org/10.1038/s41588-023-01562-0):

- Aparent2: score for 3'UTR based on human polyadenylation
- ZooPriPhyloP: Zoonomia Primate PhyloP conservation score (43 genomes)
- ZooVerPhyloP: Zoonomia Vertebrate PhyloP conservation score (241 vertebrate genomes)
- ZooRoCC: Zoonomia Runs of Contiguous Constraint
- ZooUCE: Zoonomia UltraConserved Elements
- Roulette-MR: Roulette mutation rate estimate
- Roulette-AR: Adjusted Roulette mutation rate estimate
- Roulette-FILTER: filter used by Roulette mutational score

Updates

VEP annotation build: now based on Ensembl Release 110

Supplement 2: Columns in annotation tables of the GRCh38 CADD v1.7 model. Note that parentheses around a column name indicate that this column is not used by the CADD model and only provided for user interpretation.

	Name	**Type**	**Description**
1	(Chrom)	string	Chromosome
2	(Pos)	integer	Position (1-based)
3	Ref	factor	Reference allele (default: N)
4	Alt	factor	Observed allele (default: N)
5	Type	factor	Event type (SNV, DEL, INS)
6	Length	integer	Number of inserted/deleted bases
7	(AnnoType)	factor	CodingTranscript, Intergenic, MotifFeature, NonCodingTranscript, RegulatoryFeature, Transcript
8	Consequence	factor	VEP consequence, priority selected by potential impact (default: UNKNOWN)
9	(ConsScore)	integer	Custom deleterious score assigned to Consequence
10	(ConsDetail)	string	Trimmed VEP consequence prior to simplification
11	GC	float	Percent GC in a window of +/- 75bp (default: 0.42)
12	CpG	float	Percent CpG in a window of +/- 75bp (default: 0.02)
13	motifECount	integer	Total number of overlapping motifs (default: 0)
14	(motifEName)	string	Name of sequence motif the position overlaps
15	motifEHIPos	bool	Is the position considered highly informative for an overlapping motif by VEP (default: 0)
16	motifEScoreChng	float	VEP score change for the overlapping motif site (default: 0)
17	oAA	factor	Reference amino acid (default: unknown)
18	nAA	factor	Amino acid of observed variant (default: unknown)
19	(GeneID)	string	ENSEMBL GeneID
20	(FeatureID)	string	ENSEMBL feature ID (Transcript ID or regulatory feature ID)
21	(GeneName)	string	GeneName provided in ENSEMBL annotation
22	(CCDS)	string	Consensus Coding Sequence ID
23	(Intron)	string	Intron number/Total number of exons
24	(Exon)	string	Exon number/Total number of exons
25	cDNApos	float	Base position from transcription start (default: 0*)
26	relcDNApos	float	Relative position in transcript (default: 0)
27	CDSpos	float	Base position from coding start (default: 0*)
28	relCDSpos	float	Relative position in coding sequence (default: 0)
29	protPos	float	Amino acid position from coding start (default: 0*)
30	relProtPos	float	Relative position in protein codon (default: 0)
31	Domain	factor	Domain annotation inferred from VEP annotation (ncoils, sigp, lcompl, hmmpanther, ndomain = "other named domain") (default: UD)
32	Dst2Splice	float	Distance to splice site in 20bp; positive: exonic, negative: intronic (default: 0)
33	Dst2SplType	factor	Closest splice site is ACCEPTOR or DONOR (default: unknown)

34	minDistTSS	float	Distance to closest Transcribed Sequence Start (TSS) (default: 5.5)
35	minDistTSE	float	Distance to closest Transcribed Sequence End (TSE) (default: 5.5)
36	SIFTcat	factor	SIFT category of change (default: UD)
37	SIFTval	float	SIFT score (default: 0*)
38	PolyPhenCat	factor	PolyPhen category of change (default: UD)
39	PolyPhenVal	float	PolyPhen score (default: 0*)
40	priPhCons	float	Primate PhastCons conservation score (excl. human) (default: 0.0)
41	mamPhCons	float	Mammalian PhastCons conservation score (excl. human) (default: 0.0)
42	verPhCons	float	Vertebrate PhastCons conservation score (excl. human) (default: 0.0)
43	priPhyloP	float	Primate PhyloP score (excl. human) (default: -0.029)
44	mamPhyloP	float	Mammalian PhyloP score (excl. human) (default: -0.005)
45	verPhyloP	float	Vertebrate PhyloP score (excl. human) (default: 0.042)
46	bStatistic	integer	Background selection score (default: 800)
47	targetScan	integer	targetscan (default: 0*)
48	mirSVR-Score	float	mirSVR-Score (default: 0*)
49	mirSVR-E	float	mirSVR-E (default: 0)
50	mirSVR-Aln	integer	mirSVR-Aln (default: 0)
51	cHmm_E1	float	Number of 48 cell types in chromHMM state E1_poised (default: 1.92*)
52	cHmm_E2	float	Number of 48 cell types in chromHMM state E2_repressed (default: 1.92)
53	cHmm_E3	float	Number of 48 cell types in chromHMM state E3_dead (default: 1.92)
54	cHmm_E4	float	Number of 48 cell types in chromHMM state E4_dead (default: 1.92)
55	cHmm_E5	float	Number of 48 cell types in chromHMM state E5_repressed (default: 1.92)
56	cHmm_E6	float	Number of 48 cell types in chromHMM state E6_repressed (default: 1.92)
57	cHmm_E7	float	Number of 48 cell types in chromHMM state E7_weak (default: 1.92)
58	cHmm_E8	float	Number of 48 cell types in chromHMM state E8_gene (default: 1.92)
59	cHmm_E9	float	Number of 48 cell types in chromHMM state E9_gene (default: 1.92)
60	cHmm_E10	float	Number of 48 cell types in chromHMM state E10_gene (default: 1.92)
61	cHmm_E11	float	Number of 48 cell types in chromHMM state E11_gene (default: 1.92)
62	cHmm_E12	float	Number of 48 cell types in chromHMM state E12_distal (default: 1.92)
63	cHmm_E13	float	Number of 48 cell types in chromHMM state E13_distal (default: 1.92)
64	cHmm_E14	float	Number of 48 cell types in chromHMM state E14_distal (default: 1.92)

65	cHmm_E15	float	Number of 48 cell types in chromHMM state E15_weak (default: 1.92)
66	cHmm_E16	float	Number of 48 cell types in chromHMM state E16_tss (default: 1.92)
67	cHmm_E17	float	Number of 48 cell types in chromHMM state E17_proximal (default: 1.92)
68	cHmm_E18	float	Number of 48 cell types in chromHMM state E18_proximal (default: 1.92)
69	cHmm_E19	float	Number of 48 cell types in chromHMM state E19_tss (default: 1.92)
70	cHmm_E20	float	Number of 48 cell types in chromHMM state E20_poised (default: 1.92)
71	cHmm_E21	float	Number of 48 cell types in chromHMM state E21_dead (default: 1.92)
72	cHmm_E22	float	Number of 48 cell types in chromHMM state E22_repressed (default: 1.92)
73	cHmm_E23	float	Number of 48 cell types in chromHMM state E23_weak (default: 1.92)
74	cHmm_E24	float	Number of 48 cell types in chromHMM state E24_distal (default: 1.92)
75	cHmm_E25	float	Number of 48 cell types in chromHMM state E25_distal (default: 1.92)
76	GerpRS	float	Gerp element score (default: 0)
77	GerpRSpval	float	Gerp element p-Value (default: 0)
78	GerpN	float	Neutral evolution score defined by GERP++ (default: 3.0)
79	GerpS	float	Rejected Substitution score defined by GERP++ (default: -0.2)
80	tOverlapMotifs	float	Number of overlapping predicted TF motifs
81	motifDist	float	Reference minus alternate allele difference in nucleotide frequency within an predicted overlapping motif (default: 0)
82	EncodeH3K4me1-sum	float	Sum of Encode H3K4me1 levels (from 13 cell lines) (default: 0.76)
83	EncodeH3K4me1max	float	Maximum Encode H3K4me1 level (from 13 cell lines) (default: 0.37)
84	EncodeH3K4me2-sum	float	Sum of Encode H3K4me2 levels (from 14 cell lines) (default: 0.73)
85	EncodeH3K4me2max	float	Maximum Encode H3K4me2 level (from 14 cell lines) (default: 0.37)
86	EncodeH3K4me3-sum	float	Sum of Encode H3K4me3 levels (from 14 cell lines) (default: 0.81)
87	EncodeH3K4me3max	float	Maximum Encode H3K4me3 level (from 14 cell lines) (default: 0.38)
88	EncodeH3K9ac-sum	float	Sum of Encode H3K9ac levels (from 13 cell lines) (default: 0.82)
89	EncodeH3K9ac-max	float	Maximum Encode H3K9ac level (from 13 cell lines) (default: 0.41)
90	EncodeH3K9me3-sum	float	Sum of Encode H3K9me3 levels (from 14 cell lines) (default: 0.81)
91	EncodeH3K9me3max	float	Maximum Encode H3K9me3 level (from 14 cell lines) (default: 0.38)

92	EncodeH3K27ac-sum	float	Sum of Encode H3K27ac levels (from 14 cell lines) (default: 0.74)
93	EncodeH3K27ac-max	float	Maximum Encode H3K27ac level (from 14 cell lines) (default: 0.36)
94	EncodeH3K27me3sum	float	Sum of Encode H3K27me3 levels (from 14 cell lines) (default: 0.93)
95	EncodeH3K27me3max	float	Maximum Encode H3K27me3 level (from 14 cell lines) (default: 0.47)
96	EncodeH3K36me3sum	float	Sum of Encode H3K36me3 levels (from 10 cell lines) (default: 0.71)
97	EncodeH3K36me3max	float	Maximum Encode H3K36me3 level (from 10 cell lines) (default: 0.39)
98	EncodeH3K79me2sum	float	Sum of Encode H3K79me2 levels (from 13 cell lines) (default: 0.64)
99	EncodeH3K79me2max	float	Maximum Encode H3K79me2 level (from 13 cell lines) (default: 0.34)
100	EncodeH4K20me1sum	float	Sum of Encode H4K20me1 levels (from 11 cell lines) (default: 0.88)
101	EncodeH4K20me1max	float	Maximum Encode H4K20me1 level (from 11 cell lines) (default: 0.47)
102	EncodeH2AFZ-sum	float	Sum of Encode H2AFZ levels (from 13 cell lines) (default: 0.9)
103	EncodeH2AFZ-max	float	Maximum Encode H2AFZ level (from 13 cell lines) (default: 0.42)
104	EncodeDNase-sum	float	Sum of Encode DNase-seq levels (from 12 cell lines) (default: 0.0)
105	EncodeDNase-max	float	Maximum Encode DNase-seq level (from 12 cell lines) (default: 0.0)
106	EncodetotalRNA-sum	float	Sum of Encode totalRNA-seq levels (from 10 cell lines always minus and plus strand) (default: 0.0)
107	EncodetotalRNA-max	float	Maximum Encode totalRNA-seq level (from 10 cell lines, minus and plus strand separately) (default: 0.0)
108	Grantham	float	Grantham score: oAA,nAA (default: 0*)
109	SpliceAI-acc-gain	float	Masked SpliceAI acceptor gain score (default: 0*)
110	SpliceAI-acc-loss	float	Masked SpliceAI acceptor loss score (default: 0)
111	SpliceAI-don-gain	float	Masked SpliceAI donor gain score (default: 0)
112	SpliceAI-don-loss	float	Masked SpliceAI donor loss score (default: 0)
113	MMSp_acceptorIntron	float	MMSplice acceptor intron (intron 3') score (default: 0)
114	MMSp_acceptor	float	MMSplice acceptor score (default: 0)
115	MMSp_exon	float	MMSplice exon score (default: 0)
116	MMSp_donor	float	MMSplice donor score (default: 0)
117	MMSp_donorIntron	float	MMSplice donor intron (intron 5'))score (default: 0)
118	Dist2Mutation	float	Distance between the closest BRAVO SNV up and downstream (position itself excluded) (default: 0*)
119	Freq100bp	integer	Number of frequent (MAF > 0.05) BRAVO SNV in 100 bp window nearby (default: 0)
120	Rare100bp	integer	Number of rare (MAF < 0.05) BRAVO SNV in 100 bp window nearby (default: 0)
121	Sngl100bp	integer	Number of single occurrence BRAVO SNV in 100 bp window nearby (default: 0)
122	Freq1000bp	integer	Number of frequent (MAF > 0.05) BRAVO SNV in 1000 bp window nearby (default: 0)

123	Rare1000bp	integer	Number of rare (MAF < 0.05) BRAVO SNV in 1000 bp window nearby (default: 0)
124	Sngl1000bp	integer	Number of single occurrence BRAVO SNV in 1000 bp window nearby (default: 0)
125	Freq10000bp	integer	Number of frequent (MAF > 0.05) BRAVO SNV in 10000 bp window nearby (default: 0)
126	Rare10000bp	integer	Number of rare (MAF < 0.05) BRAVO SNV in 10000 bp window nearby (default: 0)
127	Sngl10000bp	integer	Number of single occurrence BRAVO SNV in 10000 bp window nearby (default: 0)
128	EnsembleRegulatoryFeature	factor	Matches in the Ensemble Regulatory Built (similar to annotype) (default: NA)
129	dbscSNV-ada_score	float	Adaboost classifier score from dbscSNV (default: 0*)
130	dbscSNV-rf_score	float	Random forest classifier score from dbscSNV (default: 0*)
131	RemapOverlapTF	integer	Remap number of different transcription factors binding (default: -0.5)
132	RemapOverlapCL	integer	Remap number of different transcription factor - cell line combinations binding (default: -0.5)
133	EsmScoreMissense	float	ESM-1 derived score for missense variants (default: 0)
134	EsmScoreInFrame	float	ESM-1 derived score or infame InDels (default: 0)
135	EsmScoreFrameshift	float	ESM-1 derived score for InDels frameshift and stop gain (default: 0)
136	RegSeq0	float	Regulatory sequence model HEK293T
137	RegSeq1	float	Regulatory sequence model K562
138	RegSeq2	float	Regulatory sequence model HepG2
139	RegSeq3	float	Regulatory sequence model HeLa-S3
140	RegSeq4	float	Regulatory sequence model MC-7
141	RegSeq5	float	Regulatory sequence model iPS DF 19.11
142	RegSeq6	float	Regulatory sequence model GM23338
143	RegSeq7	float	Regulatory sequence model GC-matched background
144	Aparent2	float	Score for 3'UTR based on human polyadenylation (default: -0.0013)
145	ZooPriPhyloP	float	Zoonomia Primate PhyloP conservation score (43 genomes) (default: 0.005)
146	ZooVerPhyloP	float	Zoonomia Vertebrate PhyloP conservation score (241 vertebrate genome) (default: -0.1460)
147	ZooRoCC	int	Zoonomia Runs of Contiguous Constraint (default: 0)
148	ZooUCE	boolean	Zoonomia UltraConserved Elements (default: 0)
149	Roulette-FILTER	str	Roulette filter (default: na)
150	Roulette-MR	float	Roulette mutation rate estimate (default: 0. 0940)
151	Roulette-AR	float	Adjusted Roulette mutation rate estimate (default: 0.1050)
152	RawScore	float	Raw score from the model
153	PHRED	float	CADD PHRED Score

* A Boolean indicator variable was created in order to handle undefined values. Note that often indicators represent more than one annotation. They are created for only (the first) one if the covered genomic regions are identical.

Supplement 3: Columns in annotation tables of the GRCh37 CADD v1.7 model. Note that parentheses around a column name indicate that this column is not used by the CADD model and only provided for user interpretation.

	Name	**Type**	**Description**
1	(Chrom)	string	Chromosome
2	(Pos)	integer	Position (1-based)
3	Ref	factor	Reference allele (default: N)
4	Alt	factor	Observed allele (default: N)
5	Type	factor	Event type (SNV, DEL, INS)
6	Length	integer	Number of inserted/deleted bases
7	(Annotype)	factor	CodingTranscript, Intergenic, MotifFeature, NonCodingTranscript, RegulatoryFeature, Transcript
8	Consequence	factor	VEP consequence, priority selected by potential impact (default: UNKNOWN)
9	(ConsScore)	integer	Custom deleterious score assigned to Consequence
10	(ConsDetail)	string	Trimmed VEP consequence prior to simplification
11	GC	float	Percent GC in a window of +/- 75bp (default: 0.42)
12	CpG	float	Percent CpG in a window of +/- 75bp (default: 0.02)
13	motifECount	integer	Total number of overlapping motifs (default: 0)
14	(motifEName)	string	Name of sequence motif the position overlaps
15	motifEHIPos	bool	Is the position considered highly informative for an overlapping motif by VEP (default: 0)
16	motifEScoreChng	float	VEP score change for the overlapping motif site (default: 0)
17	oAA	factor	Reference amino acid (default: unknown)
18	nAA	factor	Amino acid of observed variant (default: unknown)
19	(GeneID)	string	ENSEMBL GeneID
20	(FeatureID)	string	ENSEMBL feature ID (Transcript ID or regulatory feature ID)
21	(GeneName)	string	GeneName provided in ENSEMBL annotation
22	(CCDS)	string	Consensus Coding Sequence ID
23	(Intron)	string	Intron number/Total number of exons
24	(Exon)	string	Exon number/Total number of exons
25	cDNApos	float	Base position from transcription start (default: 0*)
26	relcDNApos	float	Relative position in transcript (default: 0)
27	CDSpos	float	Base position from coding start (default: 0*)
28	relCDSpos	float	Relative position in coding sequence (default: 0)
29	protPos	float	Amino acid position from coding start (default: 0*)
30	relProtPos	float	Relative position in protein codon (default: 0)
31	Domain	factor	"Domain annotation inferred from VEP annotation (ncoils, sigp, lcompl, hmmpanther, ndomain = ""other named domain"") (default: UD) "
32	Dst2Splice	float	Distance to splice site in 20bp; positive: exonic, negative: intronic (default: 0)
33	Dst2SplType	factor	Closest splice site is ACCEPTOR or DONOR (default: unknown)
34	minDistTSS	float	Distance to closest Transcribed Sequence Start (TSS) (default: 5.5)

35	minDistTSE	float	Distance to closest Transcribed Sequence End (TSE) (default: 5.5)
36	SIFTcat	factor	SIFT category of change (default: UD)
37	SIFTval	float	SIFT score (default: 0*)
38	PolyPhenCat	factor	PolyPhen category of change (default: UD)
39	PolyPhenVal	float	PolyPhen score (default: 0*)
40	priPhCons	float	Primate PhastCons conservation score (excl. human) (default: 0.115)
41	mamPhCons	float	Mammalian PhastCons conservation score (excl. human) (default: 0.079)
42	verPhCons	float	Vertebrate PhastCons conservation score (excl. human) (default: 0.094)
43	priPhyloP	float	Primate PhyloP score (excl. human) (default: -0.033)
44	mamPhyloP	float	Mammalian PhyloP score (excl. human) (default: -0.038)
45	verPhyloP	float	Vertebrate PhyloP score (excl. human) (default: 0.017)
46	bStatistic	integer	Background selection score (default: 800)
47	targetScan	integer	targetscan (default: 0*)
48	mirSVR-Score	float	mirSVR-Score (default: 0*)
49	mirSVR-E	float	mirSVR-E (default: 0)
50	mirSVR-Aln	integer	mirSVR-Aln (default: 0)
51	cHmmTssA	float	Proportion of 127 cell types in cHmmTssA state (default: 0.0667*)
52	cHmmTssAFlnk	float	Proportion of 127 cell types in cHmmTssAFlnk state (default: 0.0667)
53	cHmmTxFlnk	float	Proportion of 127 cell types in cHmmTxFlnk state (default: 0.0667)
54	cHmmTx	float	Proportion of 127 cell types in cHmmTx state (default: 0.0667)
55	cHmmTxWk	float	Proportion of 127 cell types in cHmmTxWk state (default: 0.0667)
56	cHmmEnhG	float	Proportion of 127 cell types in cHmmEnhG state (default: 0.0667)
57	cHmmEnh	float	Proportion of 127 cell types in cHmmEnh state (default: 0.0667)
58	cHmmZnfRpts	float	Proportion of 127 cell types in cHmmZnfRpts state (default: 0.0667)
59	cHmmHet	float	Proportion of 127 cell types in cHmmHet state (default: 0.0667)
60	cHmmTssBiv	float	Proportion of 127 cell types in cHmmTssBiv state (default: 0.0667)
61	cHmmBivFlnk	float	Proportion of 127 cell types in cHmmBivFlnk state (default: 0.0667)
62	cHmmEnhBiv	float	Proportion of 127 cell types in cHmmEnhBiv state (default: 0.0667)
63	cHmmReprPC	float	Proportion of 127 cell types in cHmmReprPC state (default: 0.0667)
64	cHmmReprPCWk	float	Proportion of 127 cell types in cHmmReprPCWk state (default: 0.0667)
65	cHmmQuies	float	Proportion of 127 cell types in cHmmQuies state (default: 0.0667)

66	GerpRS	float	Gerp element score (default: 0)
67	GerpRSpval	float	Gerp element p-Value (default: 0)
68	GerpN	float	Neutral evolution score defined by GERP++ (default: 1.91)
69	GerpS	float	Rejected Substitution score defined by GERP++ (default: -0.2)
70	TFBS	float	Number of different overlapping ChIP transcription factor binding sites (default: 0)
71	TFBSPeaks	float	Number of overlapping ChIP transcription factor binding site peaks summed over different cell types/tissue (default: 0)
72	TFBSPeaksMax	float	Maximum value of overlapping ChIP transcription factor binding site peaks across cell types/tissue (default: 0)
73	tOverlapMotifs	float	Number of overlapping predicted TF motifs (default: 0)
74	motifDist	float	Reference minus alternate allele difference in nucleotide frequency within an predicted overlapping motif (default: 0)
75	Segway	factor	Result of genomic segmentation algorithm (default: unknown)
76	EncH3K27Ac	float	Maximum ENCODE H3K27 acetylation level (default: 0)
77	EncH3K4Me1	float	Maximum ENCODE H3K4 methylation level (default: 0)
78	EncH3K4Me3	float	Maximum ENCODE H3K4 trimethylation level (default: 0)
79	EncExp	float	Maximum ENCODE expression value (default: 0)
80	EncNucleo	float	Maximum of ENCODE Nucleosome position track score (default: 0)
81	EncOCC	integer	ENCODE open chromatin code (default: 5)
82	EncOCCombPVal	float	ENCODE combined p-Value (PHRED-scale) of Faire, Dnase, polII, CTCF, Myc evidence for open chromatin (default: 0)
83	EncOCDNasePVal	float	p-Value (PHRED-scale) of Dnase evidence for open chromatin (default: 0)
84	EncOCFairePVal	float	p-Value (PHRED-scale) of Faire evidence for open chromatin (default: 0)
85	EncOCpolIIPVal	float	p-Value (PHRED-scale) of polII evidence for open chromatin (default: 0)
86	EncOCctcfPVal	float	p-Value (PHRED-scale) of CTCF evidence for open chromatin (default: 0)
87	EncOCmycPVal	float	p-Value (PHRED-scale) of Myc evidence for open chromatin (default: 0)
88	EncOCDNaseSig	float	Peak signal for Dnase evidence of open chromatin (default: 0)
89	EncOCFaireSig	float	Peak signal for Faire evidence of open chromatin (default: 0)
90	EncOCpolIISig	float	Peak signal for polII evidence of open chromatin (default: 0)
91	EncOCctcfSig	float	Peak signal for CTCF evidence of open chromatin (default: 0)

92	EncOCmycSig	float	Peak signal for Myc evidence of open chromatin (default: 0)
93	Grantham	float	Grantham score: oAA,nAA (default: 0*)
94	SpliceAI-acc-gain	float	Masked SpliceAI acceptor gain score (default: 0*)
95	SpliceAI-acc-loss	float	Masked SpliceAI acceptor loss score (default: 0)
96	SpliceAI-don-gain	float	Masked SpliceAI donor gain score (default: 0)
97	SpliceAI-don-loss	float	Masked SpliceAI donor loss score (default: 0)
98	MMSp_acceptorIntron	float	MMSplice acceptor intron (intron 3') score (default: 0)
99	MMSp_acceptor	float	MMSplice acceptor score (default: 0)
100	MMSp_exon	float	MMSplice exon score (default: 0)
101	MMSp_donor	float	MMSplice donor score (default: 0)
102	MMSp_donorIntron	float	MMSplice donor intron (intron 5'))score (default: 0)
103	Dist2Mutation	float	Distance between the closest gnomAD SNV up and downstream (position itself excluded) (default: 0*)
104	Freq100bp	integer	Number of frequent (MAF > 0.05) gnomAD SNV in 100 bp window nearby
105	Rare100bp	integer	Number of rare (MAF < 0.05) gnomAD SNV in 100 bp window nearby
106	Sngl100bp	integer	Number of single occurrence gnomAD SNV in 100 bp window nearby
107	Freq1000bp	integer	Number of frequent (MAF > 0.05) gnomAD SNV in 1000 bp window nearby
108	Rare1000bp	integer	Number of rare (MAF < 0.05) gnomAD SNV in 1000 bp window nearby
109	Sngl1000bp	integer	Number of single occurrence gnomAD SNV in 1000 bp window nearby
110	Freq10000bp	integer	Number of frequent (MAF > 0.05) gnomAD SNV in 10000 bp window nearby
111	Rare10000bp	integer	Number of rare (MAF < 0.05) gnomAD SNV in 10000 bp window nearby
112	Sngl10000bp	integer	Number of single occurrence gnomAD SNV in 10000 bp window nearby
113	dbscSNV-ada_score	float	Adaboost classifier score from dbscSNV (default: 0*)
114	dbscSNV-rf_score	float	Random forest classifier score from dbscSNV (default: 0*)
115	EsmScoreMissense	float	ESM-1 derived score for missense variants (default: 0)
116	EsmScoreInFrame	float	ESM-1 derived score or infame InDels (default: 0)
117	EsmScoreFrameshift	float	ESM-1 derived score for InDels frameshift and stop gain (default: 0)
118	RegSeq0	float	Regulatory sequence model HEK293T
119	RegSeq1	float	Regulatory sequence model K562
120	RegSeq2	float	Regulatory sequence model HepG2
121	RegSeq3	float	Regulatory sequence model HeLa-S3
122	RegSeq4	float	Regulatory sequence model MC-7
123	RegSeq5	float	Regulatory sequence model iPS DF 19.11
124	RegSeq6	float	Regulatory sequence model GM23338
125	RegSeq7	float	Regulatory sequence model GC-matched background
126	Aparent2	float	Score for 3'UTR based on human polyadenylation (default: -0.0013)

127	ZooPriPhyloP	float	Zoonomia Primate PhyloP conservation score (43 genomes) (default: 0.005)
128	ZooVerPhyloP	float	Zoonomia Vertebrate PhyloP conservation score (241 vertebrate genome) (default: -0.1460)
129	ZooRoCC	int	Zoonomia Runs of Contiguous Constraint (default: 0)
130	ZooUCE	boolean	Zoonomia UltraConserved Elements (default: 0)
131	Roulette-FILTER	str	Roulette filter (default: na)
132	Roulette-MR	float	Roulette mutation rate estimate (default: 0. 0940)
133	Roulette-AR	float	Adjusted Roulette mutation rate estimate (default: 0.1050)
134	RawScore	float	Raw score from the model
135	PHRED	float	CADD PHRED Score