CADD v1.4

Developmental release

What's new?

This version of Combined Annotation Dependent Depletion (CADD) comes with separate models for the most recent human genome build GRCh38 / hg38, in addition to an updated (and similar) model for GRCh37 / hg19. Like in the last developmental releases, CADD v1.4 models are trained using logistic regression using an extended, updated and slightly fixed feature set.

Learner: In contrast to the previous release, the models in CADD v1.4 were trained using the Logistic Regression module from scikit-learn (http://scikit-learn.org). We trained one classifier for each genome build using approximately 15 million human derived variants versus approximately 15 million simulated variants. The logistic regression models used L2 penalty with C = 1 and were terminated after nine (GRCh38) and thirteen (GRCh37) L-BFGS iterations.

GRCh38: We present the first release of a CADD model trained entirely on GRCh38 variants using (mostly) *bona fide* GRCh38 annotations.

In general, we try to keep the annotations between the different genome builds as close as possible. Where possible, we therefore used or generated annotations that are equivalent to those used in the GRCh37 model. If this was not possible, we included close approximations or lifted the annotation from GRCh37. Details about the annotations in the GRCh38 version and how they relate to annotations for GRCh37 can be found in Supplement 1.

For CADD v1.4, we are only supporting the major chromosomes of GRCh38 (1-22, X & Y), but plan to add alternative haplotypes in later releases.

Genome-wide availability of CADD scores: Due to the limited or questionable coverage of the used annotations for alternative haplotypes, unplaced contigs and the mitochondrial genome, we drop support for variants located on those genomic positions. Similar considerations apply to chromosome Y, which is nevertheless included in the genome-wide scores and for which we would like to caution whether scores are comparable to those of the other chromosomes.

Feature set: In contrast to earlier versions and due to some internal restructuring of the CADD code base, CADD v1.4 is able to use recent releases of Ensembl VEP (https://www.ensembl.org/info/docs/tools/vep/index.html; using the v92 database for both builds). In comparison to version 1.3, we made changes for the following annotations:

1) Dropped the mutability index from Michaelson, J.J. et al. Cell 2012 due to over-predictive potential regarding the training data set
2) Dropped DNA shape effects due to calculation problems for INDEL and overfitting potential
3) Updated TargetScanS (Chou C. et al. Nucleic Acids Research 2018) to the latest version 7.1
4) minDistTSS and minDistTSE: fixed imputation of NA values to $\log_{10}(10\text{^}7)$

5) Adjusted representation of domain annotations as provided by current version of VEP (categories now: 'ncoils', 'sigp', 'lcompl', 'ndomain', 'hmmpanther' and 'other')
6) Added measures of genome wide variant density based on the gnomAD (v2.0.1) WGS SNV calls. We evaluate the distance between the two nearest neighboring SNV as well as the number of SNV in a surrounding window of 100, 1000 and 10000 bp for categories frequent (MAF > 5%), medium (MAF < 5%, > 0.1 %) and rare (MAF < 0.1 %). We further include the relative frequencies between the three categories for the three window sizes defined before. These measures are only evaluated for SNVs.
7) Added splice effect prediction from dbscSNV (Jian X. et al. Nucleic Acids Research 2014)

In previous versions, we reported additional annotations (e.g. allele frequencies, mapAbility measures, UCSC segmental duplication score, fitCons score) not included in CADD score calculation in our annotation files. These annotations were used by us to check our model. We removed these annotations to reduce file size. Complete listings of the annotation columns of each model are available as Supplemental 2 and 3.

Performance of CADD v1.4 in comparison to v1.3

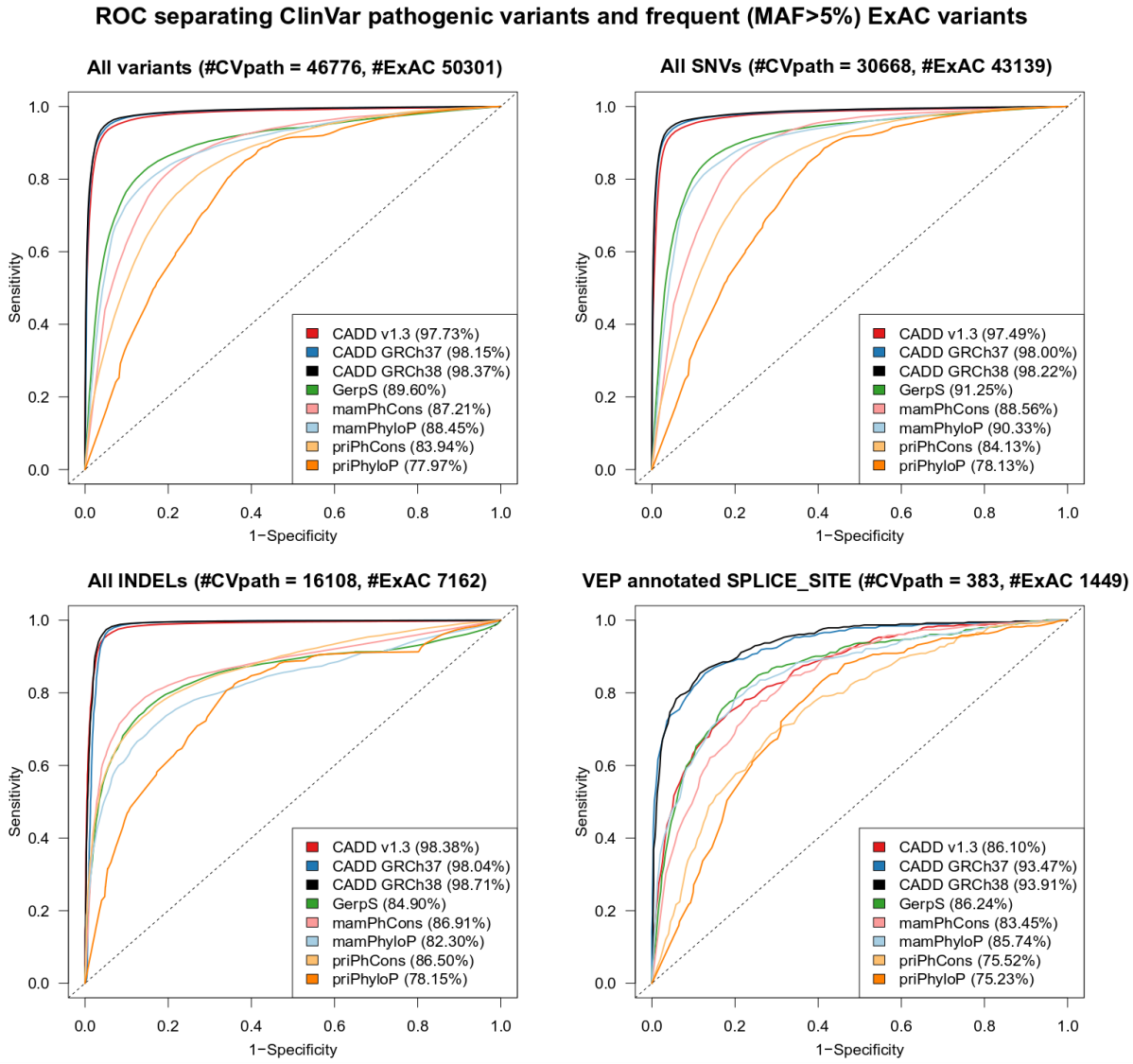
As with previous releases, we see a number of difference in how the model evaluates different variant categories (see figure on page 4). Most notable is that the scores of variants in canonical splice sites differ in this release, probably because we have added more splice site annotations. We further evaluated the CADD v1.4 models against test datasets from the previous release. Here, only variants mapped unambiguously between the genome builds were evaluated to allow comparisons between builds.

Spearman correlation between the observation frequency of TP53 cancer variants in the IARC database (p53.iarc.fr) and CADD scores changed from 0.458 in v1.3 to 0.440 for GRCh37 and 0.474 for GRCh38 in v1.4

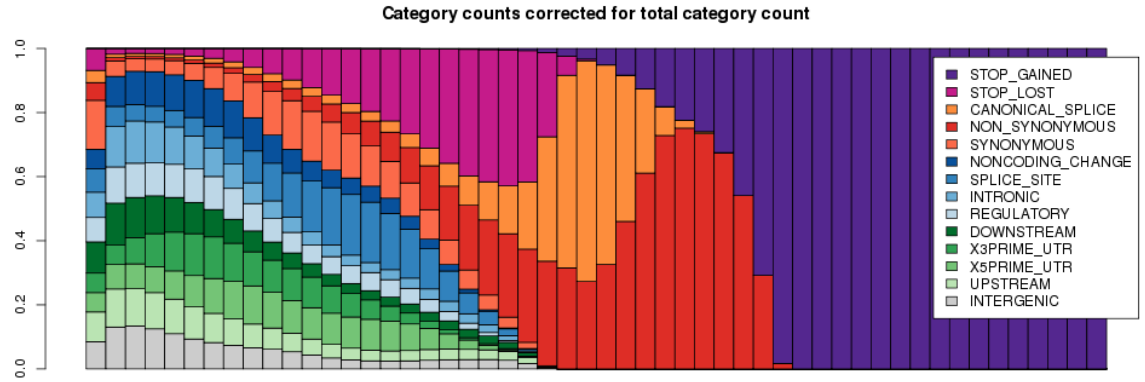
Spearman correlation of CADD scores with absolute $\log_2$-fold changes determined from saturation mutagenesis in ALDOB, ECR11 and HBB regulatory sequences (Patwardhan, R.P. et al. Nature Biotechnology 2012, Patwardhan R.P. et al. Nature Biotechnology 2009) changed as follows:

	ALDOB	ECR11	HBB
CADD v1.3	0. 3385	0. 0886	0.1556
CADD v1.4 – GRCh37	0.4995	0.2028	0.1550
CADD v1.4 – GRCh38	0.4338	0.0199	0.2385

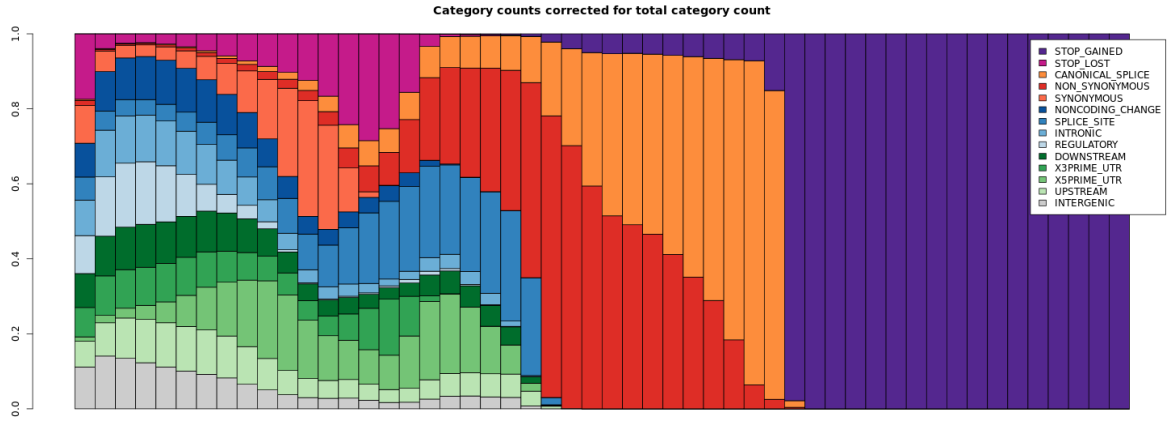
The performance in distinguishing known pathogenic ClinVar (Landrum, M.J. et al. Nucleic acids Research 2014) variants from frequent variants (allele frequency > 5%) in ExAC (Lek, M., Karczewski K. et al. Nature 2016) has changed as shown in the plots below. The prediction of splice sites improved due to the addition of dbscSNV as annotation.



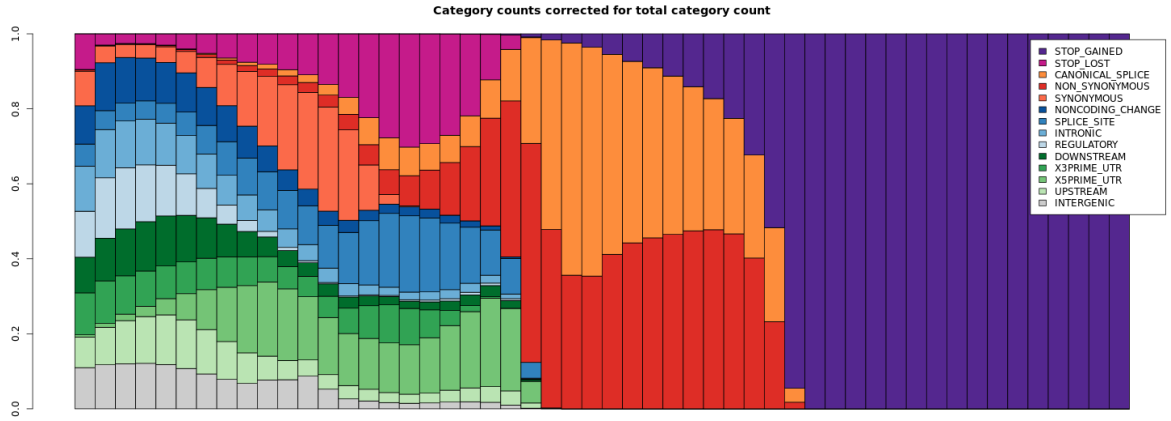
CADD v1.3 (GRCh37)



Catgory distribution CADDv1.4-GRCh37



Category distribution CADDv1.4-GRCh38



Supplement 1: Annotations in the CADD model for GRCh38

Annotations based on VEP output: same processing as for GRCh37

Grantham: unspecific of genome build

phastCons, phyloP, GERP++: novel score generated based on multiz100way alignment of hg38 from UCSC genome browser (excluding the human genome sequence in calculation)

bstatistic: liftover from hg18

mirSVR: liftover from hg19

targetScan: liftover from hg19

chromHMM: uses now chromHMM 25 state model, ENCODE/Roadmap cell types only

Encode expression, nucleosome position, histone modification, open chromatin: Encode reference epigenome data from up to 14 cell types for totalRNAseq, DNase-seq and Chip-Seq of 10 different histone modifications, for each: sum of all cell types and maximum, signal log-transformed

Segway: replaced by Ensembl regulatory build (which is also a Segway model)

tOverlapMotifs: liftover from hg19

TFBS: replaced by data from ReMap2, we are counting for every genomic position the number of different TF and the number of TF-cell-type hits

mutationDensity, nearestMutation: same processing based on BRAVO/TOPMed freeze 5

dbscSNV: liftover from hg19

Supplement 2: Columns in annotation tables of the GRCh37 CADD v1.4 model

	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	Dist2Mutation	float	Distance between the closest gnomAD SNV up and downstream (position itself excluded) (default: 0*)
95	Freq100bp	integer	Number of frequent (MAF > 0.05) gnomAD SNV in 100 bp window nearby (default: 0)
96	Rare100bp	integer	Number of rare (MAF < 0.05) gnomAD SNV in 100 bp window nearby (default: 0)
97	Sngl100bp	integer	Number of single occurrence gnomAD SNV in 100 bp window nearby (default: 0)
98	Freq1000bp	integer	Number of frequent (MAF > 0.05) gnomAD SNV in 1000 bp window nearby (default: 0)
99	Rare1000bp	integer	Number of rare (MAF < 0.05) gnomAD SNV in 1000 bp window nearby (default: 0)
100	Sngl1000bp	integer	Number of single occurrence gnomAD SNV in 1000 bp window nearby (default: 0)
101	Freq10000bp	integer	Number of frequent (MAF > 0.05) gnomAD SNV in 10000 bp window nearby (default: 0)
102	Rare10000bp	integer	Number of rare (MAF < 0.05) gnomAD SNV in 10000 bp window nearby (default: 0)
103	Sngl10000bp	integer	Number of single occurrence gnomAD SNV in 10000 bp window nearby (default: 0)
104	dbscSNV-ada_score	float	Adaboost classifier score from dbscSNV (default: 0*)
105	dbscSNV-rf_score	float	Random forest classifier score from dbscSNV (default: 0*)
106	RawScore	float	Raw score from the model
107	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 GRCh38 CADD v1.4 model

	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	EncodeH3K4me1-max	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	EncodeH3K4me2-max	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	EncodeH3K4me3-max	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	EncodeH3K9me3-max	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	EncodeH3K27me3-sum	float	Sum of Encode H3K27me3 levels (from 14 cell lines) (default: 0.93)
95	EncodeH3K27me3-max	float	Maximum Encode H3K27me3 level (from 14 cell lines) (default: 0.47)
96	EncodeH3K36me3-sum	float	Sum of Encode H3K36me3 levels (from 10 cell lines) (default: 0.71)
97	EncodeH3K36me3-max	float	Maximum Encode H3K36me3 level (from 10 cell lines) (default: 0.39)
98	EncodeH3K79me2-sum	float	Sum of Encode H3K79me2 levels (from 13 cell lines) (default: 0.64)
99	EncodeH3K79me2-max	float	Maximum Encode H3K79me2 level (from 13 cell lines) (default: 0.34)
100	EncodeH4K20me1-sum	float	Sum of Encode H4K20me1 levels (from 11 cell lines) (default: 0.88)
101	EncodeH4K20me1-max	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	Dist2Mutation	float	Distance between the closest BRAVO SNV up and downstream (position itself excluded) (default: 0*)
110	Freq100bp	integer	Number of frequent (MAF > 0.05) BRAVO SNV in 100 bp window nearby (default: 0)
111	Rare100bp	integer	Number of rare (MAF < 0.05) BRAVO SNV in 100 bp window nearby (default: 0)
112	Sngl100bp	integer	Number of single occurrence BRAVO SNV in 100 bp window nearby (default: 0)
113	Freq1000bp	integer	Number of frequent (MAF > 0.05) BRAVO SNV in 1000 bp window nearby (default: 0)
114	Rare1000bp	integer	Number of rare (MAF < 0.05) BRAVO SNV in 1000 bp window nearby (default: 0)
115	Sngl1000bp	integer	Number of single occurrence BRAVO SNV in 1000 bp window nearby (default: 0)
116	Freq10000bp	integer	Number of frequent (MAF > 0.05) BRAVO SNV in 10000 bp window nearby (default: 0)
117	Rare10000bp	integer	Number of rare (MAF < 0.05) BRAVO SNV in 10000 bp window nearby (default: 0)
118	Sngl10000bp	integer	Number of single occurrence BRAVO SNV in 10000 bp window nearby (default: 0)
119	EnsembleRegulatory-Feature	factor	Matches in the Ensemble Regulatory Built (similar to annotype) (default: NA)
120	dbscSNV-ada_score	float	Adaboost classifier score from dbscSNV (default: 0*)
121	dbscSNV-rf_score	float	Random forest classifier score from dbscSNV (default: 0*)
122	RemapOverlapTF	integer	Remap number of different transcription factors binding (default: -0.5)
123	RemapOverlapCL	integer	Remap number of different transcription factor - cell line combinations binding (default: -0.5)
124	RawScore	float	Raw score from the model
125	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.