## CADD v1.3 minor/developmental release

## What's new?

This version of CADD fixes a minor issue with overlapping annotation that was identified in the developmental releases v1.1 and v1.2. Like the previous developmental releases, CADD v1.3 was trained using a logistic regression learner as well as an extended and updated feature set. This version also uses a new and simplified training data set. The following document describes the differences to our last major release (v1.0).

*Learner:* For this version we used the Logistic Regression module of GraphLab Create v1.4 (<u>https://dato.com/products/create/</u>). In contrast to previous releases, we trained only one classifier using approximately 15 million human derived variants (newly extracted from EPO 6 primate alignments v75) versus approximately 15 million (newly) simulated variants. We used an L2 penalty of 1.0 and terminated training after 10 iterations.

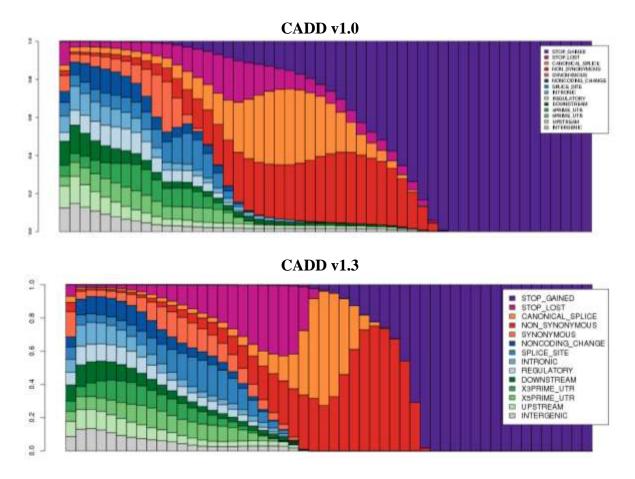
*Feature set:* CADD v1.3 is still based on the GRCh37/hg19 genome build. It uses the same updated version of Ensembl Variant Effect Predictor (VEP, McLaren, W. et al. Bioinformatics 2010) and annotation tracks as CADD v1.1 and v1.2. We are using Ensembl script release v76 (using the v75 database for GRCh37) of VEP. With this VEP version, the functional consequence predictions for insertion/deletion (InDel) events have considerably improved. In comparison to v1.0, we added the following annotation features:

- Genomic segmentation inferred from the combined ENCODE (16 cell types) data and NIH Roadmap Epigenomics (111 cell types) data using ChromHMM (Ernst, J. & Kellis, M. Nature Methods 2012; <u>https://sites.google.com/site/anshulkundaje/projects/epigenome roadmap#TOC-Core-Integrative-chromatin-state-maps-127-Epigenomes-</u>). We added one feature for each state. The value for each feature is the proportion of cell types annotated with this state by ChromHMM.
- Predicted local DNA secondary structure effects as measured by delta HeIT, MGW, ProT, and Roll values for substitutions within a context of 2x11bp using the software described in Zhou, T. et al. Nucleic Acids Research 2013.
- 3) Predicted microRNA binding sites reported in mirSVR (Betel, D. et al. Genome Biol 2010) and TargetScanS (UCSC hg19 track, Lewis, B.P. et al. Cell 2005).
- 4) Genome-wide mutability index from Michaelson, J.J. et al. Cell 2012 coordinate lifted from NCBI36/hg18 to GRCh37/hg19.
- 5) Reduced-level representation (i.e. "ncoils", "tmhmm", "sigp", "lcompl", "ndomain" for other named domains) derived from the domain annotations provided for protein-coding variants by the VEP annotation.

In CADD v1.3, we updated the EPO 6 primate whole genome alignments to Ensembl version 75. The ancestral allele is stored in our annotated files and also informs the isDerived column. Neither column is used by the model for score calculation. We also include fitCons scores (Gulko, B. et al. 2014; doi:10.1101/006825; downloaded on June 30 2014) in our annotation files for exploratory purposes (and do *not* use them for the CADD model). A complete list of the 114 columns present in the annotation files is available in the supplemental table 1. Information on imputation of missing values can be found in supplemental table 2. If not noted otherwise descriptions in Kircher, M & Witten, D.M. et al. Nature Genetics 2014 apply.

## How does performance compare between CADD v1.0 and v1.3?

Even though results reported for many of our previously used validation sets are similar or better (see below), we note that there is a measurable reshuffling of variant ranks between versions. We also see differences in the score ranges that are obtained for certain predicted functional consequences:



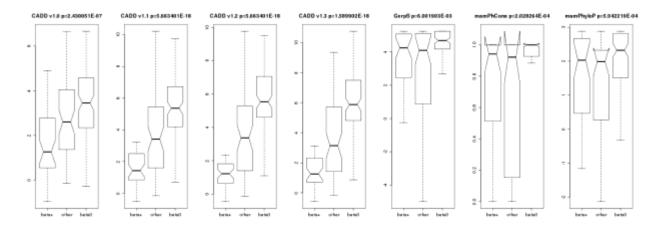
The correlation between derived allele frequency (DAF) in the 1000 Genome data and raw scores is -0.082 SNVs / -0.069 InDels for version 1.0 and -0.076 SNVs / -0.078 InDels for v1.3.

Correlation between the observation frequency of P53 cancer variants in the IARC data base (p53.iarc.fr) and CADD scores changed from 0.387 in v1.0 to 0.458 in v1.3.

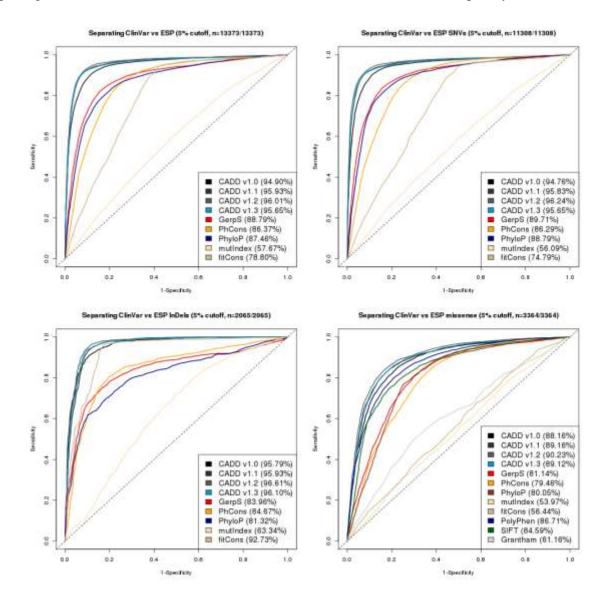
Correlation of CADD scores and log2-fold changes determined from saturation mutagenesis in ALDOB, ECR11 and HBB (Patwardhan, R.P. et al. Nature Biotechnology 2012, Patwardhan, R.P. et al. Nature Biotechnology 2009) regulatory sequences changed as follows:

	ALDOB	ECR11	HBB	ALL
CADD v1.0	0.3588	0.2459	0.2017	0.3123
CADD v1.3	0.4545	0.2055	0.1834	0.3166

The discrimination of HBB variants associated with varying degrees of severity for betathalassemia (Giardine, B. et al. Hum Mutation 2007) looks as follows:



The performance for an updated ClinVar (Landrum, M.J., et al. Nucleic Acids Research 2014) pathogenic vs. ESP (Fu, W. et al Nature 2012) with at least 5% allele frequency looks as follows:



**Supplemental Table 1:** Columns of the extended annotation tables. Parentheses around the column name indicate that the column is not used for model training or prediction of pathogenicity.

	Name	Туре	Description
1	(Chrom)	factor	Chromosome
2	(Pos)	int	Position (1-based)
3	Ref	factor	Reference allele
			Ancestral (e.g. chimp like) base; defined using EPO 6 primate
4	(Anc)	factor	alignments (v75)
5	Alt	factor	Observed allele
6	Туре	factor	Event type (SNV, DEL, INS)
7	Length	int	Number of inserted/deleted bases
8	isTv	bool	Is transversion?
9	(isDerived)	bool	Observed allele is an evolutionary derived allele
10	(AnnoType)	factor	CodingTranscript, Intergenic, MotifFeature,
			NonCodingTranscript, RegulatoryFeature, Transcript 3PRIME_UTR, 5PRIME_UTR, DOWNSTREAM, INTERGENIC, INTRONIC, NON_SYNONYMOUS,
11	Consequence	factor	SYNONYMOUS, REGULATORY, STOP_GAINED, STOP_LOST, SPLICE_SITE, CANONICAL_SPLICE UPSTREAM, NONCODING_CHANGE
12	(ConsScore)	int	Custom deleterious score assigned to Consequence
13	(ConsDetail)	string	Trimmed VEP consequence prior to simplification
14	GC	num	Percent GC in a window of +/- 75bp
15	CpG	num	Percent CpG in a window of +/- 75bp
16	(mapAbility20bp)	num	Mapability of 20bp fragments determined by Duke
17	(mapAbility35bp)	num	Mapability of 35bp fragments determined by Duke
18	(scoreSegDup)	num	UCSC segmental duplication similarity, indicate the percent identity to the highest-similarity segmental duplication event.
19	priPhCons	num	Primate PhastCons conservation score (excl. human)
20	mamPhCons	num	Mammalian PhastCons conservation score (excl. human)
21	verPhCons	num	Vertebrate PhastCons conservation score (excl. human)
22	priPhyloP	num	Primate PhyloP score (excl. human)
23	mamPhyloP	num	Mammalian PhyloP score (excl. human)
24	verPhyloP	num	Vertebrate PhyloP (excl. human)
25	GerpN	num	Neutral evolution score defined by GERP++
26	GerpS	num	Rejected Substitution' score defined by GERP++
27	GerpRS	num	Gerp element score
28	GerpRSpval	num	Gerp element p-Value
29	bStatistic	int	Background selection score
30	mutIndex	num	Mutability index from Michaelson, J.J. et al. Cell 2012
31	dnaHelT	num	Predicted local DNA structure effect on dnaHelT
32	dnaMGW	num	Predicted local DNA structure effect on dnaMGW
33	dnaProT	num	Predicted local DNA structure effect on dnaProT
33 34	dnaRoll	num	Predicted local DNA structure effect on dnaPloT
34 35	mirSVR-Score		mirSVR-Score
35 36	mirSVR-E	num	mirSVR-E
30 37	mirSVR-Aln	num int	mirSVR-Aln
38	targetScan	int	targetScan
38 39	(fitCons)		fitCons score
39 40	cHmmTssA	num	
40 41	cHmmTssAFlnk	num	Proportion of 127 cell types in cHmmTssA state Proportion of 127 cell types in cHmmTssAFlnk state
41	CITIIIII I SSAL IIIK	num	1 roportion of 127 cen types in crimin ssArmik state

42cHmmTxFlnknumProportion of 127 cell types in cHmmTxFlnk state43cHmmTxnumProportion of 127 cell types in cHmmTx state44cHmmTxWknumProportion of 127 cell types in cHmmTxWk state45cHmmEnhGnumProportion of 127 cell types in cHmmZnfRpts state46cHmmZnfRptsnumProportion of 127 cell types in cHmmZnfRpts state47cHmmZnfRptsnumProportion of 127 cell types in cHmmZnfRpts state48cHmmBivFlnknumProportion of 127 cell types in cHmmBivFlkk state50cHmmReprPCnumProportion of 127 cell types in cHmmReprPC state51cHmmReprPCWknumProportion of 127 cell types in cHmmReprPC state53cHmmReprPCWknumProportion of 127 cell types in cHmmReprPC state54cHmmQuiesnumMaximum ENCODE H3K27 acetylation level55EncH3k27AcnumMaximum ENCODE H3K27 acetylation level56EncOCCintMaximum ENCODE H3K4 trimethylation level58EncOCCnumP-Value (PHRED-scale) of Faire evidence for open chromatin60EncOCCnump-Value (PHRED-scale) of GTCF evidence for open chromatin61EncOCCDNasePValnump-Value (PHRED-scale) of GTCF evidence for open chromatin62EncOCChaiseSignump-Value (PHRED-scale) of GTCF evidence for open chromatin64EncOCCPillPValnump-Value (PHRED-scale) of GTCF evidence for open chromatin65EncOCCraiseSignump-Value (PHRED-scale) of	
44cHmmTxWknumProportion of 127 cell types in cHmmTxWk state45cHmmEnhGnumProportion of 127 cell types in cHmmEnh state46cHmmEnhGnumProportion of 127 cell types in cHmmEnh state47cHmmTskBivnumProportion of 127 cell types in cHmmEnhSistate48cHmmHetnumProportion of 127 cell types in cHmmEnhBiv state49cHmmBivFlnknumProportion of 127 cell types in cHmmBivFlnk state50cHmmReprPCnumProportion of 127 cell types in cHmmReprPC state51cHmmReprPCWknumProportion of 127 cell types in cHmmReprPC state53cHmmReprPCWknumMaximum ENCODE H3K27 acctylation level54cHmmQuiesnumMaximum ENCODE H3K27 acctylation level55EncH3k27AcnumMaximum ENCODE H3K4 methylation level56EncH3k27AcnumMaximum ENCODE Nucelosome position track scor60EncOCCintMaximum of ENCODE Nucelosome position track scor61EncOCCMbPValnump-Value (PHRED-scale) of Faire evidence for open chromatin62EncOCDINasePValnump-Value (PHRED-scale) of GDias evidence for open chromatin64EncOCChiIFValnump-Value (PHRED-scale) of GDias evidence for open chromatin65EncOCPiIFValnump-Value (PHRED-scale) of Faire evidence for open chromatin66EncOCCiFiFValnump-Value (PHRED-scale) of GDias evidence for open chromatin67EncOCCPiIFValnump-Value (PHRED-scale) of Gpe	
45cHmmEnhGnumProportion of 127 cell types in cHmmEnhG state46cHmmZnfRptsnumProportion of 127 cell types in cHmmEnh state47cHmmZnfRptsnumProportion of 127 cell types in cHmmHet state48cHmmHetnumProportion of 127 cell types in cHmmEnhBiv state50cHmmBivFlnknumProportion of 127 cell types in cHmmEnhBiv state51cHmmEnhBivnumProportion of 127 cell types in cHmmReprPC state52cHmmReprPCWknumProportion of 127 cell types in cHmmReprPC state54cHmmQuiesnumMaximum ENCODE H3K4 methylation level55EncExpnumMaximum ENCODE H3K4 trimethylation level56EncH3K4Me1numMaximum ENCODE H3K4 trimethylation level57EncH3K4Me3numMaximum ENCODE H3K4 trimethylation level58EncOCCintMaximum ENCODE H3K4 trimethylation level59EncOCCintMaximum CCTCF evidence for open chr61EncOCCDNasePValnump-Value (PHRED-scale) of Dase evidence for open chr64EncOCCDNaseFValnump-Value (PHRED-scale) of faire evidence for open chr65EncOCCDNaseSignumP-value (PHRED-scale) of faire evidence for open chr66EncOCCDISignumP-value (PHRED-scale) of faire evidence for open chr67EncOCCDNaseSignumPeak signal for Faire evidence of open chromatin68EncOCCDissesnumPeak signal for Taire evidence of open chromatin70EncOCCCI	
46cHmmEnhnumProportion of 127 cell types in cHmmEnh state47cHmmHetnumProportion of 127 cell types in cHmmRpts state48cHmmHetnumProportion of 127 cell types in cHmmTrsBiv state49cHmmBivBiknumProportion of 127 cell types in cHmmBivFihk state50cHmmBivFlnknumProportion of 127 cell types in cHmmRprPC state51cHmmRepPCnumProportion of 127 cell types in cHmmReprPC state52cHmmQuiesnumProportion of 127 cell types in cHmmReprPC state53cHmmQuiesnumProportion of 127 cell types in cHmmReprPC state54cHmmQuiesnumMaximum ENCODE expression value55EncExpnumMaximum ENCODE H3K47 acetylation level56EncH3K4Me1numMaximum of ENCODE Nucelosome position track sco60EncOCCintENCODE combined p-Value (PHRED-scale) of Faire59EncOCCbasePValnump-Value (PHRED-scale) of OTCF evidence for open chromatin61EncOCTairePValnump-Value (PHRED-scale) of OTCF evidence for open chromatin62EncOCTairePValnump-Value (PHRED-scale) of OTCF evidence for open chromatin64EncOCTairePValnump-Value (PHRED-scale) of OTCF evidence for open chromatin65EncOCTilsignump-Value (PHRED-scale) of OTCF evidence for open chromatin66EncOCTairePValnump-value (PHRED-scale) of OTCF evidence for open chromatin70EncOCTaisignump-value (PHRED-scale) o	
47cHmmZnfRptsnumProportion of 127 cell types in cHmmZnfRpts state48cHmmHetnumProportion of 127 cell types in cHmmTssBiv state49cHmmTssBivnumProportion of 127 cell types in cHmmBivFlnk state50cHmmBivFlnknumProportion of 127 cell types in cHmmEnhBiv state51cHmmRptPCWknumProportion of 127 cell types in cHmmReprPC state53cHmmRptPCWknumProportion of 127 cell types in cHmmReprPC state54cHmmQuiesnumProportion of 127 cell types in cHmmReprPC state55EncExpnumMaximum ENCODE expression value56EncH3K4Me1numMaximum ENCODE H3K4 methylation level57EncH3K4Me3numMaximum ENCODE H3K4 methylation level58EncH3K4Me3numMaximum ENCODE H3K4 methylation level59EncOCCombPValnumPvAlue (PHRED-scale) of Daase evidence for open ch61EncOCDNasePValnump-Value (PHRED-scale) of Daase evidence for open ch62EncOCTaireSignump-Value (PHRED-scale) of CTC evidence for open ch64EncOCpollIPValnump-Value (PHRED-scale) of Myc evidence for open ch65EncOCTaireSignumPeak signal for Daase evidence of open chromatin66EncOCpollISignumPeak signal for Myc evidence of open chromatin71EncOCtafreSignumPeak signal for Myc evidence of open chromatin72SegwayfactorReefreence minus alternate allel difference in nu73 </td <td></td>	
48cHmmHetnumProportion of 127 cell types in cHmmHet state49cHmmBixFlnknumProportion of 127 cell types in cHmmBixFlnk state50cHmmBinkinnumProportion of 127 cell types in cHmmEnhBiv state51cHmmEnhBivnumProportion of 127 cell types in cHmmReprPC state53cHmmReprPCnumProportion of 127 cell types in cHmmReprPC state54cHmmQuiesnumProportion of 127 cell types in cHmmReprPCWk state55EncExpnumMaximum ENCODE H3K27 acetylation level56EncH3K4Me1numMaximum ENCODE H3K4 methylation level57EncH3K4Me3numMaximum ENCODE H3K4 methylation level58EncNucleonumMaximum ENCODE H3K4 methylation level59EncNucleonumMaximum of ENCODE Nucelosome position track sco60EncOCCintENCODE combined p-Value (PHRED-scale) of Faire evidence for open chromatin61EncOCCDNasePValnump-Value (PHRED-scale) of Faire evidence for open chromatin62EncOCTPValnump-Value (PHRED-scale) of Faire evidence for open chromatin64EncOCTPValnump-Value (PHRED-scale) of Faire evidence for open chromatin65EncOCTAIPSignump-Value (PHRED-scale) of polII evidence for open chromatin66EncOCTAIPSignumPeak signal for Dase evidence of open chromatin70EncOCTSignumPeak signal for PII evidence of open chromatin71EncOCTSignumPeak signal for PIC evidence of	
49cHmmTssBivnumProportion of 127 cell types in cHmmTssBiv state50cHmmEnhBivnumProportion of 127 cell types in cHmmBivFlnk state51cHmmRepPCnumProportion of 127 cell types in cHmmRepPC state52cHmmRepPCWknumProportion of 127 cell types in cHmmRuperS state53cHmmRepPCWknumProportion of 127 cell types in cHmmRuperS state54cHmmQuiesnumProportion of 127 cell types in cHmmQuies state55EncExpnumMaximum ENCODE H3K47 acetylation level56EncH3K4Me1numMaximum ENCODE H3K4 trimethylation level57EncH3K4Me1numMaximum ENCODE H3K4 trimethylation level58EncH3K4Me3numMaximum ENCODE H3K4 trimethylation level59EncNucleonumMaximum ENCODE H3K4 trimethylation level61EncOCCDNasePValnumPvalue (PHRED-scale) of Dase evidence for open chr62EncOCDNasePValnump-Value (PHRED-scale) of Faire evidence for open chr64EncOCripValnump-Value (PHRED-scale) of Faire evidence for open chr65EncOCripValnump-Value (PHRED-scale) of CTCF evidence for open chr66EncOCryValnumPeak signal for CTCF evidence of open chromatin67EncOCpolIISignumPeak signal for CTCF evidence of open chromatin70EncOCrySignumPeak signal for CTCF evidence of open chromatin71EncOCrySignumPeak signal for CTCF evidence of open chromatin72<	
50cHmmBivFlnknumProportion of 127 cell types in cHmmBivFlnk state51cHmmRenPCnumProportion of 127 cell types in cHmmBehPiv state52cHmmReprPCWknumProportion of 127 cell types in cHmmReprPC state53cHmmReprPCWknumProportion of 127 cell types in cHmmReprPCWk state54cHmmQuiesnumMaximum ENCODE expression value56EncH3K4Me1numMaximum ENCODE H3K4 methylation level57EncH3K4Me1numMaximum ENCODE H3K4 trimethylation level58EncH3K4Me3numMaximum ENCODE H3K4 trimethylation level59EncNucleointENCODE combined p-Value (PHRED-scale) of Faire61EncOCCombPValnump-Value (PHRED-scale) of Dnase evidence for open chr62EncOCDNasePValnump-Value (PHRED-scale) of CTCF evidence for open chr64EncOCpolIIPValnump-Value (PHRED-scale) of CTCF evidence for open chr65EncOCpolIIPValnumPeak signal for Dase evidence of open chromatin68EncOCpaseSignumPeak signal for Faire evidence of open chromatin70EncOCpolIISignumPeak signal for Signal for Myc evidence of open chromatin71EncOCmySignumPeak signal for Signal for Myc evidence of open chromatin72SegwayfactorResturg factor73tOverlapMotifsintNumber of overlapping motifs74motifDistnumPeak signal for Myc evidence of open chromatin75motifEName	
51cHmmEnhBivnumProportion of 127 cell types in cHmmEnhBiv state52cHmmReprPCnumProportion of 127 cell types in cHmmReprPC state53cHmmReprPCWknumProportion of 127 cell types in cHmmReprPCWk state54cHmmQuiesnumProportion of 127 cell types in cHmmReprPCWk state55EncExpnumMaximum ENCODE expression value56EncH3K4Me1numMaximum ENCODE H3K4 methylation level57EncH3K4Me3numMaximum ENCODE H3K4 trimethylation level58EncH3K4Me3numMaximum ENCODE H3K4 trimethylation level59EncNucleonumMaximum Of ENCODE Nucelosome position track sco60EncOCCintENCODE combined p-Value (PHRED-scale) of Faire61EncOCpolIIPValnump-Value (PHRED-scale) of Dase evidence for open chr62EncOCpolIIPValnump-Value (PHRED-scale) of CTCF evidence for open chr64EncOCpolIIPValnump-Value (PHRED-scale) of GTCF evidence for open chr65EncOCpolIISignumPeak signal for Dase evidence of open chromatin68EncOCpolISignumPeak signal for polI evidence of open chromatin70EncOCpolISignumPeak signal for PolI evidence of open chromatin71EncOCpolISignumPeak signal for Segmentation algorithm73tOverlapMotifsintNumber of overlapping motifs74motifDistnumPeak signal for CTCF evidence of open chromatin75motifEName	
52cHmmReprPCnumProportion of 127 cell types in cHmmReprPC state53cHmmQuiesnumProportion of 127 cell types in cHmmReprPCWk state54cHmmQuiesnumProportion of 127 cell types in cHmmReprPCWk state55EncExpnumMaximum ENCODE H3K27 acetylation level56EncH3K4Me1numMaximum ENCODE H3K4 methylation level57EncH3K4Me3numMaximum ENCODE H3K4 trimethylation level58EncH3K4Me3numMaximum ENCODE H3K4 trimethylation level59EncNucleonumMaximum of ENCODE Nucelosome position track score60EncOCCintENCODE open chromatin code61EncOCCmbPValnump-Value (PHRED-scale) of Daase evidence for open chr62EncOCchifPValnump-Value (PHRED-scale) of Faire evidence for open chr63EncOCrairePValnump-Value (PHRED-scale) of CTCF evidence for open chr64EncOCpolIIPvalnump-Value (PHRED-scale) of Myc evidence for open chr65EncOCcpolIISignumPeak signal for Faire evidence of open chromatin66EncOCpolIISignumPeak signal for GTCF evidence of open chromatin70EncOCcpolIISignumPeak signal for GTCF evidence of open chromatin71EncOCraireSignumPeak signal for Myc evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73tOverlapMotifsintNumber of overlapping motifs74motifEscoreChng<	
52cHmmReprPCnumProportion of 127 cell types in cHmmReprPC state53cHmmQuiesnumProportion of 127 cell types in cHmmReprPC state54cHmmQuiesnumProportion of 127 cell types in cHmmReprPC state55EncExpnumMaximum ENCODE expression value56EncH3K4Me1numMaximum ENCODE H3K4 methylation level57EncH3K4Me3numMaximum ENCODE H3K4 methylation level58EncH3K4Me3numMaximum ENCODE H3K4 trimethylation level59EncNucleonumMaximum ENCODE Nucelosome position track score60EncOCCintENCODE open chromatin code61EncOCCombPValnump-Value (PHRED-scale) of Daase evidence for open chromatin62EncOCpolIIPValnump-Value (PHRED-scale) of Faire evidence for open chromatin63EncOCpolIIPValnump-Value (PHRED-scale) of CTCF evidence for open chromatin64EncOCpolIISignumPeak signal for Daase evidence of open chromatin65EncOCpolIISignumPeak signal for Faire evidence of open chromatin70EncOCpolIISignumPeak signal for CTCF evidence of open chromatin71EncOCmycSignumPeak signal for CTCF evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73tOverlapMotifsintNumber of overlapping motifs74motifEscoreChngnumVEP score change for the overlapping motif site79TFBSint<	
53cHmmReprPCWk cHmmQuiesnum numProportion of 127 cell types in cHmmReprPCWk state Proportion of 127 cell types in cHmmQuies state54cHmmQuiesnum maximum ENCODE expression value55EncExpnum Maximum ENCODE H3K27 acetylation level56EncH3K27Acnum Maximum ENCODE H3K24 methylation level58EncH3K4Me3num Maximum ENCODE H3K4 trimethylation level59EncNucleonum Maximum ENCODE H3K4 trimethylation level60EncOCCint61EncOCCombPVal EncOCDaisePValnum p-Value (PHRED-scale) of Dase evidence for open chromatin p-Value (PHRED-scale) of Faire evidence for open chromatin p-Value (PHRED-scale) of CTCF evidence for open chromatin p-Value (PHRED-scale) of CTCF evidence for open chromatin p-Value (PHRED-scale) of CTCF evidence for open chromatin p-Value (PHRED-scale) of open chromatin <b< td=""><td></td></b<>	
54cHmmQuiesnumProportion of 127 cell types in cHmmQuies state55EncExpnumMaximum ENCODE expression value56EncH3K4Me1numMaximum ENCODE H3K27 acetylation level57EncH3K4Me1numMaximum ENCODE H3K4 methylation level58EncH3K4Me3numMaximum ENCODE H3K4 trimethylation level59EncNucleonumMaximum ENCODE Nucelosome position track sco60EncOCCintENCODE combined p-Value (PHRED-scale) of Faire61EncOCDNasePValnump-Value (PHRED-scale) of poll evidence for open chromatin62EncOCDNasePValnump-Value (PHRED-scale) of poll evidence for open chromatin63EncOCpolIIPValnump-Value (PHRED-scale) of poll evidence for open chromatin64EncOCpolIISignump-Value (PHRED-scale) of CTCF evidence for open chromatin65EncOCpaireSignumPeak signal for Dnase evidence of open chromatin66EncOCpolIISignumPeak signal for Dnase evidence of open chromatin70EncOCpolIISignumPeak signal for Myc evidence of open chromatin71EncOCmycSignumPeak signal for Myc evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73tOverlapMotifsintNumber of overlapping motifs74motifESumestringName of sequence motif the position overlaps75motifENamestringName of sequence motif the position overlaps74	;
55EncExpnumMaximum ENCODE expression value56EncH3K47AcnumMaximum ENCODE H3K27 acetylation level57EncH3K4Me1numMaximum ENCODE H3K4 trimethylation level58EncH3K4Me3numMaximum ENCODE H3K4 trimethylation level59EncNucleonumMaximum of ENCODE Nucelosome position track sco60EncOCCintENCODE combined p-Value (PHRED-scale) of Faire61EncOCDnasePValnump-Value (PHRED-scale) of Dnase evidence for open chromatin62EncOCDNasePValnump-Value (PHRED-scale) of faire evidence for open chromatin64EncOCpolIIPValnump-Value (PHRED-scale) of Faire evidence for open chro65EncOCtrifValnump-Value (PHRED-scale) of Myc evidence for open chro66EncOCpolIISignumPeak signal for Dnase evidence of open chromatin67EncOCDNaseSignumPeak signal for Faire evidence of open chromatin68EncOCpolIISignumPeak signal for CTCF evidence of open chromatin70EncOCtrifsignumPeak signal for Myc evidence of open chromatin71EncOCmycSignumPeak signal for Myc evidence of open chromatin73tOverlapMotifsintNumber of overlapping predicted TF motifs74motifDistnumReference minus alternate allele difference in nu frequency within an predicted overlapping motif75motifENamestring79TFBSintNumber of overlapping motifs80TFBS	
56EncH3K27AcnumMaximum ENCODE H3K27 acetylation level57EncH3K4Me1numMaximum ENCODE H3K4 methylation level58EncH3K4Me3numMaximum OENCODE Nucelosome position track sco60EncOCCintENCODE open chromatin code61EncOCDNasePValnumP-Value (PHRED-scale) of Dase evidence for open chromatin62EncOCDNasePValnump-Value (PHRED-scale) of Dase evidence for open chromatin63EncOCDIIPValnump-Value (PHRED-scale) of Dase evidence for open chromatin64EncOCpolIPValnump-Value (PHRED-scale) of Dase evidence for open chromatin65EncOCDNaseSignump-Value (PHRED-scale) of Myc evidence for open chromatin66EncOCDNaseSignumPeak signal for Dase evidence of open chromatin67EncOCDNaseSignumPeak signal for poll evidence of open chromatin68EncOCFaireSignumPeak signal for CTCF evidence of open chromatin70EncOCctifSignumPeak signal for Myc evidence of open chromatin71EncOCmycSignumPeak signal for Myc evidence of open chromatin72SegwayfactorReference minus alternate allele difference in nu frequency within an predicted overlapping motif73tOverlapMotifsintNumber of overlapping motifs74motifENamestringName of sequence motif the position overlaps75motifENamestringName of sequence motif the position overlaps79TFBSintN	
57EncH3K4Me1numMaximum ENCODE H3K4 methylation level58EncH3K4Me3numMaximum ENCODE H3K4 trimethylation level59EncNucleonumMaximum of ENCODE Nucelosome position track score60EncOCCintENCODE open chromatin code61EncOCCombPValnumFNCODE combined p-Value (PHRED-scale) of Faire62EncOCDNasePValnump-Value (PHRED-scale) of Dase evidence for open chromatin63EncOCFairePValnump-Value (PHRED-scale) of Faire evidence for open chromatin64EncOCpolIIPValnump-Value (PHRED-scale) of OII evidence for open chromatin65EncOCctfValnump-Value (PHRED-scale) of Myc evidence for open chromatin66EncOCpolIISignumPeak signal for Dase evidence of open chromatin68EncOCpolIISignumPeak signal for polII evidence of open chromatin70EncOCpolIISignumPeak signal for Duse evidence of open chromatin71EncOCmycSignumPeak signal for Myc evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73toverlapMotifsintNumber of overlapping motifs74motifENamestringName of sequence motif the position overlaps79TFBSintint80TFBSPeaksintNumber of different overlapping ChIP transcription factor bind peaks summed over different cell types/tissue81TEBSPeaksMaxintNatimu Paule of overlapping ChIP transcription <td></td>	
58EncH3K4Me3num numMaximum ENCODE H3K4 trimethylation level59EncNucleonumMaximum of ENCODE Nucelosome position track score60EncOCCintENCODE open chromatin code61EncOCCmbPValnumENCODE combined p-Value (PHRED-scale) of Faire62EncOCDNasePValnump-Value (PHRED-scale) of Dasse evidence for open chromatin63EncOCFairePValnump-Value (PHRED-scale) of Faire evidence for open chromatin64EncOCpolIIPValnump-Value (PHRED-scale) of CTCF evidence for open chromatin65EncOCTaireSignump-Value (PHRED-scale) of Myc evidence for open chromatin66EncOCFaireSignumPeak signal for Dasse evidence of open chromatin67EncOCFaireSignumPeak signal for CTCF evidence of open chromatin70EncOCtofSignumPeak signal for CTCF evidence of open chromatin71EncOCmycSignumPeak signal for Myc evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73tOverlapMotifsintNumber of overlapping predicted TF motifs74motifENamestringName of sequence motif the position overlaps75motifENamestringName of sequence motif the position overlaps74motifENamestringName of sequence motif the position overlaps75motifENamestringName of overlapping ChIP transcription factor bind peaks summed over different cell types/tissue79TFBS<	
59EncNucleonumMaximum of ENCODE Nucelosome position track score60EncOCCintENCODE open chromatin code61EncOCDNasePValnumENCODE combined p-Value (PHRED-scale) of Faire62EncOCDNasePValnump-Value (PHRED-scale) of Dnase evidence for open chromatin63EncOCpolIIPValnump-Value (PHRED-scale) of Faire evidence for open chromatin64EncOCpolIIPValnump-Value (PHRED-scale) of Faire evidence for open chromatin65EncOCtcfPValnump-Value (PHRED-scale) of Myc evidence for open chromatin66EncOCpolIISignump-Value (PHRED-scale) of Myc evidence for open chromatin67EncOCpolIISignumPeak signal for Dnase evidence of open chromatin68EncOCpolIISignumPeak signal for CTCF evidence of open chromatin70EncOCraireSignumPeak signal for Myc evidence of open chromatin71EncOCmycSignumPeak signal for Myc evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73tOverlapMotifsintNumber of overlapping motifs74motifECountintTotal number of overlapping motifs75motifENamestringName of sequence motif the position overlaps76TFBSintNumber of different overlapping motif site79TFBSintNumber of overlapping ChIP transcription80TFBSPeaksintNumber of overlapping ChIP transcription81<	
60EncOCCintENCODE open chromatin code61EncOCCombPValnumPCODE combined p-Value (PHRED-scale) of Faire poIII, CTCF, Myc evidence for open chromatin62EncOCDNasePValnump-Value (PHRED-scale) of Dase evidence for open chromatin63EncOCFairePValnump-Value (PHRED-scale) of Faire evidence for open chromatin64EncOCpoIIIPValnump-Value (PHRED-scale) of Faire evidence for open chromatin65EncOCtrycPValnump-Value (PHRED-scale) of CTCF evidence for open chromatin66EncOCmycPValnump-Value (PHRED-scale) of Myc evidence for open chromatin67EncOCDNaseSignumPeak signal for Dase evidence of open chromatin68EncOCpoIIISignumPeak signal for Dase evidence of open chromatin70EncOCtrycSignumPeak signal for Myc evidence of open chromatin71EncOCmycSignumPeak signal for Myc evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73tOverlapMotifsintNumber of overlapping predicted TF motifs74motifEcountintTotal number of overlapping motifs75motifECountintTotal number of overlapping motifs76motifENamestringName of sequence motif the position overlaps75motifENamestringName of sequence motif the position overlaps76Is the position considered highly informative overlapping motif by VEPVEP score change for the overlapping ChIP transcrip	ore
61EncOCCombPValnumENCODE combined p-Value (PHRED-scale) of Faire polII, CTCF, Myc evidence for open chromatin62EncOCDNasePValnump-Value (PHRED-scale) of Dnase evidence for open chromatin63EncOCpairePValnump-Value (PHRED-scale) of Faire evidence for open chromatin64EncOCpoIIIPValnump-Value (PHRED-scale) of opoli evidence for open chromatin65EncOCpoValnump-Value (PHRED-scale) of CTCF evidence for open chromatin66EncOCmycPValnump-Value (PHRED-scale) of Myc evidence for open chromatin67EncOCpoIIISignumPeak signal for Dnase evidence of open chromatin68EncOCpoIIISignumPeak signal for Doli evidence of open chromatin70EncOCpoIIISignumPeak signal for Myc evidence of open chromatin71EncOCmycSignumPeak signal for Myc evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73tOverlapMotifsintNumber of overlapping motifs74motifECountintTotal number of overlapping motifs75motifECountintTotal number of overlapping motifs76motifESnamestringName of sequence motif the position overlaps79TFBSintNumber of different overlapping ChIP transcription80TFBSPeaksintNumber of overlapping ChIP transcription81TEBSPeaksMaxintMaximum value of overlapping ChIP transcription	510
61EncoccombPvalnumpoIII, CTCF, Myc evidence for open chromatin62EncOCDNasePValnump-Value (PHRED-scale) of Dnase evidence for open chromatin63EncOCFairePValnump-Value (PHRED-scale) of Faire evidence for open chromatin64EncOCpoIIIPValnump-Value (PHRED-scale) of pOIII evidence for open chromatin65EncOCtcfPValnump-Value (PHRED-scale) of CTCF evidence for open chromatin66EncOCpoIISignump-Value (PHRED-scale) of Myc evidence for open chromatin67EncOCpoIISignumPeak signal for Dnase evidence of open chromatin68EncOCpoIISignumPeak signal for CTCF evidence of open chromatin70EncOCpoIISignumPeak signal for Myc evidence of open chromatin71EncOCmycSignumPeak signal for Myc evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73tOverlapMotifsintNumber of overlapping predicted TF motifs74motifECountintTotal number of overlapping motifs75motifECountintTotal number of overlapping motifs76motifEXnumVEP score change for the overlapping motif site78motifEScoreChngnumVEP score change for the overlapping ChIP transcription79TFBSintNumber of overlapping ChIP transcription80TFBSPeaksintMaximum value of overlapping ChIP transcription	e Dnase
62EncOCDNasePValnump-Value (PHRED-scale) of Dnase evidence for open chr63EncOCFairePValnump-Value (PHRED-scale) of Faire evidence for open chr64EncOCpolIIPValnump-Value (PHRED-scale) of polII evidence for open chr65EncOCtcfPValnump-Value (PHRED-scale) of CTCF evidence for open chr66EncOCDNaseSignump-Value (PHRED-scale) of Myc evidence for open chr67EncOCDNaseSignump-Value (PHRED-scale) of Myc evidence for open chr68EncOCFaireSignumPeak signal for Dnase evidence of open chromatin69EncOCpolIISignumPeak signal for CTCF evidence of open chromatin70EncOCtrdSignumPeak signal for Myc evidence of open chromatin71EncOCmycSignumPeak signal for Myc evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73tOverlapMotifsintNumber of overlapping predicted TF motifs74motifECountintTotal number of overlapping motifs75motifECountintTotal number of overlapping motifs76motifEScoreChngnumVEP score change for the overlapping motif site79TFBSintintNumber of overlapping ChIP transcription80TFBSPeaksintintMaximum value of overlapping ChIP transcription	e, Dhase,
63EncOCFairePValnump-Value (PHRED-scale) of Faire evidence for open chromatin64EncOCpolIIPValnump-Value (PHRED-scale) of polII evidence for open chromatin65EncOCctcfPValnump-Value (PHRED-scale) of CTCF evidence for open chromatin66EncOCpolIISignump-Value (PHRED-scale) of Myc evidence for open chromatin67EncOCDNaseSignumPeak signal for Dnase evidence of open chromatin68EncOCpolIISignumPeak signal for Faire evidence of open chromatin69EncOCpolIISignumPeak signal for CTCF evidence of open chromatin70EncOCtcfSignumPeak signal for Myc evidence of open chromatin71EncOCmycSignumPeak signal for Myc evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73tOverlapMotifsintNumber of overlapping predicted TF motifs74motifECountintnum75motifECountint76motifENamestring77motifESoreChngnum78motifEScoreChngnum79TFBSint80TFBSPeaksint81TEBSPaaksMaxint	hromatin
64EncOCpolIIPValnump-Value (PHRED-scale) of polII evidence for open chromatin65EncOCctcfPValnump-Value (PHRED-scale) of CTCF evidence for open chromatin66EncOCmycPValnump-Value (PHRED-scale) of Myc evidence for open chromatin67EncOCDNaseSignumPeak signal for Dnase evidence of open chromatin68EncOCpolIISignumPeak signal for Faire evidence of open chromatin69EncOCpolIISignumPeak signal for OTCF evidence of open chromatin70EncOCpolIISignumPeak signal for Myc evidence of open chromatin71EncOCmycSignumPeak signal for Myc evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73tOverlapMotifsintNumber of overlapping predicted TF motifs74motifECountintTotal number of overlapping motifs75motifECountintTotal number of overlapping motifs76motifENamestringName of sequence motif the position overlaps77motifEScoreChngnumVEP score change for the overlapping motif site79TFBSintNumber of overlapping ChIP transcription80TFBSPeaksintNumber of overlapping ChIP transcription81TEBSPaaksMaxintMaximum value of overlapping ChIP transcription	
65EncOCctcfPValnump-Value (PHRED-scale) of CTCF evidence for open chromatin66EncOCmycPValnump-Value (PHRED-scale) of Myc evidence for open chromatin67EncOCDNaseSignumPeak signal for Dnase evidence of open chromatin68EncOCFaireSignumPeak signal for Faire evidence of open chromatin69EncOCpolIISignumPeak signal for polII evidence of open chromatin70EncOCctcfSignumPeak signal for CTCF evidence of open chromatin71EncOCmycSignumPeak signal for Myc evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73tOverlapMotifsintNumber of overlapping predicted TF motifs74motifDistnumReference minus alternate allele difference in nu frequency within an predicted overlapping motif75motifECountintTotal number of overlapping motifs76motifEHIPosboolIs the position considered highly informative overlapping motif by VEP78motifEScoreChngnumVEP score change for the overlapping motif site79TFBSintnumNumber of overlapping ChIP transcription factor binding sites80TFBSPeaksintNumber of overlapping ChIP transcription factor binding sites81TEBSPeaksMaxintMaximum value of overlapping ChIP transcription	
66EncOCmycPValnum nump-Value (PHRED-scale) of Myc evidence for open chromatin67EncOCDNaseSignumPeak signal for Dnase evidence of open chromatin68EncOCFaireSignumPeak signal for Faire evidence of open chromatin69EncOCpolIISignumPeak signal for CTCF evidence of open chromatin70EncOCtcfSignumPeak signal for Myc evidence of open chromatin71EncOCmycSignumPeak signal for Myc evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73tOverlapMotifsintNumber of overlapping predicted TF motifs74motifDistnumReference minus alternate allele difference in nu frequency within an predicted overlapping motif75motifECountintTotal number of overlapping motifs76motifEStamestringName of sequence motif the position overlaps77motifEScoreChngnumVEP score change for the overlapping motif site79TFBSintNumber of overlapping ChIP transcription80TFBSPeaksintMumber of overlapping ChIP transcription81TEBSPeakcMaxintMaximum value of overlapping ChIP transcription	
67EncOCDNaseSignumPeak signal for Dnase evidence of open chromatin68EncOCFaireSignumPeak signal for Faire evidence of open chromatin69EncOCpolIISignumPeak signal for polII evidence of open chromatin70EncOCtcfSignumPeak signal for CTCF evidence of open chromatin71EncOCmycSignumPeak signal for Myc evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73tOverlapMotifsintNumber of overlapping predicted TF motifs74motifDistnumReference minus alternate allele difference in m frequency within an predicted overlapping motif75motifECountintTotal number of overlapping motifs76motifENamestringName of sequence motif the position overlaps77motifEScoreChngnumVEP score change for the overlapping motif site79TFBSintNumber of overlapping ChIP transcription binding sites80TFBSPeaksintMumber of overlapping ChIP transcription factor binding sites	
68EncOCFaireSignumPeak signal for Faire evidence of open chromatin69EncOCpolIISignumPeak signal for polII evidence of open chromatin70EncOCctcfSignumPeak signal for CTCF evidence of open chromatin71EncOCmycSignumPeak signal for Myc evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73tOverlapMotifsintNumber of overlapping predicted TF motifs74motifDistnumReference minus alternate allele difference in m frequency within an predicted overlapping motif75motifECountintTotal number of overlapping motifs76motifENamestringName of sequence motif the position overlaps77motifEStnumVEP score change for the overlapping motif site78TFBSintNumber of overlapping ChIP transcription binding sites80TFBSPeaksintNumber of overlapping ChIP transcription factor bind peaks summed over different cell types/tissue81TEBSPeakcMaxintMaximum value of overlapping ChIP transcription	omatin
69EncOCpolIISignumPeak signal for polII evidence of open chromatin70EncOCctcfSignumPeak signal for CTCF evidence of open chromatin71EncOCmycSignumPeak signal for Myc evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73tOverlapMotifsintNumber of overlapping predicted TF motifs74motifDistnumReference minus alternate allele difference in nu frequency within an predicted overlapping motif75motifECountintTotal number of overlapping motifs76motifENamestringName of sequence motif the position overlaps77motifEHIPosboolIs the position considered highly informative overlapping motif by VEP78motifEScoreChngnumVEP score change for the overlapping motif site79TFBSintNumber of overlapping ChIP transcription factor bind peaks summed over different cell types/tissue80TFBSPeaksint81TEBSPeaksMaxint	
70EncOCctcfSignumPeak signal for CTCF evidence of open chromatin71EncOCmycSignumPeak signal for Myc evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73tOverlapMotifsintNumber of overlapping predicted TF motifs74motifDistnumReference minus alternate allele difference in nu frequency within an predicted overlapping motif75motifECountintTotal number of overlapping motifs76motifENamestringName of sequence motif the position overlaps77motifEHIPosboolIs the position considered highly informative overlapping motif by VEP78motifEScoreChngnumVEP score change for the overlapping motif site Number of different overlapping ChIP transcription binding sites80TFBSPeaksintNumber of overlapping ChIP transcription factor bind peaks summed over different cell types/tissue Maximum value of overlapping ChIP transcription	
71EncOCmycSignum factorPeak signal for Myc evidence of open chromatin72SegwayfactorResult of genomic segmentation algorithm73tOverlapMotifsintNumber of overlapping predicted TF motifs74motifDistnumReference minus alternate allele difference in nu frequency within an predicted overlapping motif75motifECountintTotal number of overlapping motifs76motifENamestringName of sequence motif the position overlaps77motifEHIPosboolIs the position considered highly informative overlapping motif by VEP78motifEScoreChngnumVEP score change for the overlapping motif site79TFBSintNumber of overlapping ChIP transcription binding sites80TFBSPeaksintmuther of overlapping ChIP transcription peaks summed over different cell types/tissue Maximum value of overlapping ChIP transcription	
72Segway tOverlapMotifsfactor intResult of genomic segmentation algorithm Number of overlapping predicted TF motifs73tOverlapMotifsintNumber of overlapping predicted TF motifs Reference minus alternate allele difference in nu frequency within an predicted overlapping motif74motifDistnumInt75motifECountintTotal number of overlapping motifs76motifENamestringName of sequence motif the position overlaps77motifEHIPosboolIs the position considered highly informative overlapping motif by VEP78motifEScoreChngnumVEP score change for the overlapping motif site Number of different overlapping ChIP transcription binding sites80TFBSPeaksintNumber of overlapping ChIP transcription factor bind peaks summed over different cell types/tissue Maximum value of overlapping ChIP transcription	
73tOverlapMotifsintNumber of overlapping predicted TF motifs74motifDistnumReference minus alternate allele difference in nu frequency within an predicted overlapping motif75motifECountintTotal number of overlapping motifs76motifENamestringName of sequence motif the position overlaps77motifEHIPosboolIs the position considered highly informative overlapping motif by VEP78motifEScoreChngnumVEP score change for the overlapping motif site79TFBSintNumber of different overlapping ChIP transcription binding sites80TFBSPeaksintNumber of overlapping ChIP transcription peaks summed over different cell types/tissue Maximum value of overlapping ChIP transcription	
74motifDistnumReference minus alternate allele difference in nu frequency within an predicted overlapping motif75motifECountintTotal number of overlapping motifs76motifENamestringName of sequence motif the position overlaps77motifEHIPosboolIs the position considered highly informative overlapping motif by VEP78motifEScoreChngnumVEP score change for the overlapping motif site79TFBSintNumber of different overlapping ChIP transcription binding sites80TFBSPeaksintNumber of overlapping ChIP transcription factor bind peaks summed over different cell types/tissue Maximum value of overlapping ChIP transcription	
75motifECountintfrequency within an predicted overlapping motif75motifENameintTotal number of overlapping motifs76motifENamestringName of sequence motif the position overlaps77motifEHIPosboolIs the position considered highly informative overlapping motif by VEP78motifEScoreChngnumVEP score change for the overlapping motif site79TFBSintNumber of different overlapping ChIP transcription binding sites80TFBSPeaksintNumber of overlapping ChIP transcription factor bind peaks summed over different cell types/tissue Maximum value of overlapping ChIP transcription	ucleotide
76motifENamestringName of sequence motif the position overlaps77motifEHIPosboolIs the position considered highly informative overlapping motif by VEP78motifEScoreChngnumVEP score change for the overlapping motif site79TFBSintNumber of different overlapping ChIP transcription binding sites80TFBSPeaksintNumber of overlapping ChIP transcription factor bind peaks summed over different cell types/tissue81TEBSPeaksMaxint	
77motifEHIPosboolIs the position considered highly informative overlapping motif by VEP78motifEScoreChngnumVEP score change for the overlapping motif site79TFBSintNumber of different overlapping ChIP transcription binding sites80TFBSPeaksintNumber of overlapping ChIP transcription peaks summed over different cell types/tissue Maximum value of overlapping ChIP transcription	
77MotifEHIPosbooloverlapping motif by VEP78motifEScoreChngnumVEP score change for the overlapping motif site79TFBSintNumber of different overlapping ChIP transcription binding sites80TFBSPeaksintNumber of overlapping ChIP transcription factor bind peaks summed over different cell types/tissue81TEBSPeaksMaxint	
78motifEScoreChngnumVEP score change for the overlapping motif site79TFBSintNumber of different overlapping ChIP transcription80TFBSPeaksintNumber of overlapping ChIP transcription factor bind peaks summed over different cell types/tissue81TEBSPeaksMaxint	for an
79IFBSIntbinding sites80TFBSPeaksintbinding sites81TEBSPeaksMaxintMaximum value of overlapping ChIP transcription	6
80TFBSPeaksintNumber of overlapping ChIP transcription factor bind peaks summed over different cell types/tissue Maximum value of overlapping ChIP transcription	on factor
	-
binding site peaks across cell types/tissue	on factor
82 (isKnownVariant) bool Position is observed as being variable in 1000G or ESP	P?
83 (ESP_AF) num Average ESP frequency for alternative alleles at site	
84 (ESP_AFR) num Average ESP African ancestry frequency	
85 (ESP_EUR) num Average ESP European ancestry frequency	
86 (TG_AF) num Average 1000 Genomes frequency for alternative allele	es at site
87 (TG_ASN) num Average 1000 Genomes Asian population frequency	
88 (TG_AMR) num Average 1000 Genomes South American population fre	requency

	Name	Туре	Description	
89	(TG_AFR)	num	Average 1000 Genomes African population frequency	
90	(TG_EUR)	num	Average 1000 Genomes European population frequency	
91	minDistTSS	int	Distance to closest Transcribed Sequence Start (TSS)	
92	minDistTSE	int	Distance to closest Transcribed Sequence End (TSE)	
93	(GeneID)	string	ENSEMBL GeneID	
94	(FeatureID)	string	ENSEMBL feature ID (Transcript ID or regulatory feature ID)	
95	(CCDS)	string	Consensus Coding Sequence ID	
96	(GeneName)	string	GeneName provided in ENSEMBL annotation	
97	cDNApos	int	Base position from transcription start	
98	relcDNApos	num	Relative position in transcript	
99	CDSpos	int	Base position from coding start	
100	relCDSpos	num	Relative position in coding sequence	
101	protPos	int	Amino acid position from coding start	
102	relProtPos	num	Relative position in protein codon	
103	Domain	atring	Domain annotation inferred from VEP annotation (ncoils,	
105	Domani	string	tmhmm, sigp, lcompl, ndomain = "other named domain")	
104	Det 28 milion	int	Distance to splice site in 20bp; positive: exonic, negative:	
104	Dst2Splice	IIIt	intronic	
105	Dst2SplType	factor	Closest splice site is ACCEPTOR or DONOR	
106	(Exon)	string	Exon number/Total number of exons	
107	(Intron)	string	Intron number/Total number of exons	
108	oAA	factor	Reference amino acid	
109	nAA	factor	Amino acid of observed variant	
110	Grantham	int	Grantham score: oAA,nAA	
111	PolyPhenCat	factor	PolyPhen category of change	
112	PolyPhenVal	num	PolyPhen score	
113	SIFTcat	factor	SIFT category of change	
114	SIFTval	num	SIFT score	

**Supplementary Table 2:** Imputation of missing values for model training and prediction. An asterisk (\*) indicates that a Boolean indicator variable was created in order to handle undefined values for that feature.

Name	Value	Name	Value
isTv	0.5	EncH3K4Me3	0
GC	0.418	EncNucleo	0
CpG	0.024	EncOCC	5
priPhCons	0.115	EncOCCombPVal	0
mamPhCons	0.079	EncOCDNasePVal	0
verPhCons	0.094	EncOCFairePVal	0
priPhyloP	-0.033	EncOCpolIIPVal	0
mamPhyloP	-0.038	EncOCctcfPVal	0
verPhyloP	0.017	EncOCmycPVal	0
GerpN	1.909	EncOCDNaseSig	0
GerpS	-0.200	EncOCFaireSig	0
GerpRS	0	EncOCpolIISig	0
GerpRSpval	1	EncOCctcfSig	0
bStatistic	800.261	EncOCmycSig	0
mutIndex	0	Segway	undefined
dnaHelT	0	tOverlapMotifs	0
dnaMGW	0	motifDist	0
dnaProT	0	motifECount	0
dnaRoll	0	motifEHIPos	FALSE
mirSVRs*	0	motifEScoreChng	0
mirSVRe	0	TFBS	0
mirSVRa	0	TFBSPeaks	0
targetScan*	0	TFBSPeaksMax	0
cHmmTssA	0.0667	minDistTSS	LOG(1000000)
cHmmTssAFlnk	0.0667	minDistTSE	LOG(1000000)
cHmmTxFlnk	0.0667	cDNApos*	0
cHmmTx	0.0667	relcDNApos*	0
cHmmTxWk	0.0667	CDSpos*	0
cHmmEnhG	0.0667	relCDSpos*	0
cHmmEnh	0.0667	protPos*	0
cHmmZnfRpts	0.0667	relProtPos*	0
cHmmHet	0.0667	Domain*	undefined
cHmmTssBiv	0.0667	Dst2Splice*	0
cHmmBivFlnk	0.0667	Dst2SplType*	undefined
cHmmEnhBiv	0.0667	oAA	undefined
cHmmReprPC	0.0667	nAA	undefined
cHmmReprPCWk	0.0667	Grantham*	0
cHmmQuies	0.0667	PolyPhenCat	undefined
EncExp	0	PolyPhenVal*	0
EncH3K27Ac	0	SIFTcat	undefined
EncH3K4Me1	0	SIFTval*	0