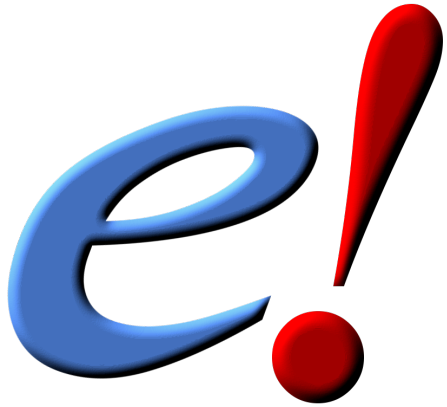
Analysing variants with the

EMBL-EBI

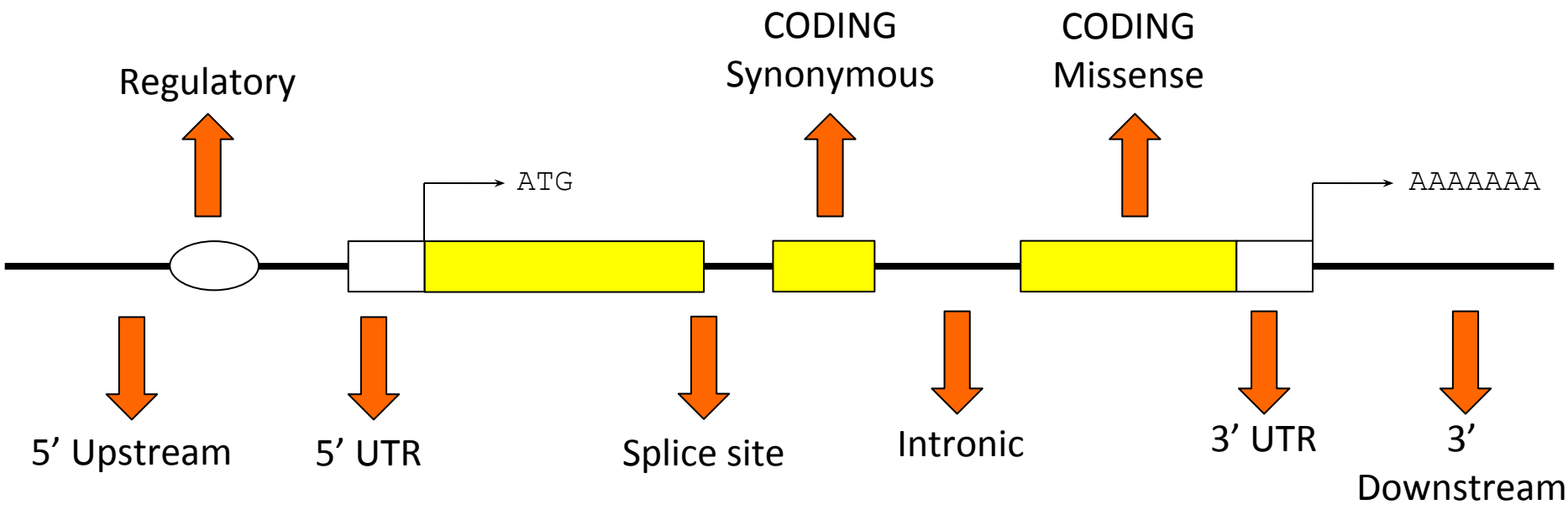
wellcome trust sanger institute



Your own variation data

Variant coordinates	1 881907 881906 -/C + 5 140532 140532 T/C + 12 1017956 1017956 T/A + 2 946507 946507 G/C + 14 19584687 19584687 C/T -
HGVS notation	ENST00000285667.3:c.1047_1048insC 5:g.140532T>C NM_153681.2:c.7C>T ENSP00000439902.1:p.Ala2233Asp NP_000050.2:p.Ile2285Val
VCF	#CHROM POS ID REF ALT 20 14370 rs6054257 G A 20 17330 . T A 20 1110696 rs6040355 A G,T 20 1230237 . T .
Variant IDs	rs41293501 COSM327779 rs146120136 FANCD1:c.475G>A rs373400041

What are the consequences?



SO consequence terms

* SO term	SO description	SO accession	Old Ensembl term
transcript_ablation	A feature ablation whereby the deleted region includes a transcript feature	SO:0001893	Transcript ablation
splice_donor_variant	A splice variant that changes the 2 base region at the 5' end of an intron	SO:0001575	Essential splice site
splice_acceptor_variant	A splice variant that changes the 2 base region at the 3' end of an intron	SO:0001574	
stop_gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript	SO:0001587	Stop gained
frameshift_variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three	SO:0001589	Frameshift coding
stop_lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript	SO:0001578	Stop lost
initiator_codon_variant	A codon variant that changes at least one base of the first codon of a transcript	SO:0001582	Non synonymous coding
inframe_insertion	An inframe non synonymous variant that inserts bases into in the coding sequence	SO:0001821	
inframe_deletion	An inframe non synonymous variant that deletes bases from the coding sequence	SO:0001822	
missense_variant	A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved	SO:0001583	
transcript_amplification	A feature amplification of a region containing a transcript	SO:0001889	Transcript amplification
splice_region_variant	A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron	SO:0001630	Splice site
incomplete_terminal_codon_variant	A sequence variant where at least one base of the final codon of an incompletely annotated transcript is changed	SO:0001626	Partial codon
synonymous_variant	A sequence variant where there is no resulting change to the encoded amino acid	SO:0001819	Synonymous coding
stop_retained_variant	A sequence variant where at least one base in the terminator codon is changed, but the terminator remains	SO:0001567	
coding_sequence_variant	A sequence variant that changes the coding sequence	SO:0001580	Coding unknown
mature_miRNA_variant	A transcript variant located with the sequence of the mature miRNA	SO:0001620	Within mature miRNA
5_prime_UTR_variant	A UTR variant of the 5' UTR	SO:0001623	5prime UTR
3_prime_UTR_variant	A UTR variant of the 3' UTR	SO:0001624	3prime UTR
intron_variant	A transcript variant occurring within an intron	SO:0001627	Intronic
NMD_transcript_variant	A variant in a transcript that is the target of NMD	SO:0001621	NMD transcript
non_coding_exon_variant	A sequence variant that changes non-coding exon sequence	SO:0001792	Within non coding gene
nc_transcript_variant	A transcript variant of a non coding RNA	SO:0001619	
upstream_gene_variant	A sequence variant located 5' of a gene	SO:0001631	Upstream
downstream_gene_variant	A sequence variant located 3' of a gene	SO:0001632	Downstream
TFBS_ablation	A feature ablation whereby the deleted region includes a transcription factor binding site	SO:0001895	Tfbs ablation
TFBS_amplification	A feature amplification of a region containing a transcription factor binding site	SO:0001892	Tfbs amplification
TF_binding_site_variant	A sequence variant located within a transcription factor binding site	SO:0001782	Regulatory region
regulatory_region_variant	A sequence variant located within a regulatory region	SO:0001566	
regulatory_region_ablation	A feature ablation whereby the deleted region includes a regulatory region	SO:0001894	Regulatory region ablation
regulatory_region_amplification	A feature amplification of a region containing a regulatory region	SO:0001891	Regulatory region amplification
feature_elongation	A sequence variant that causes the extension of a genomic feature, with regard to the reference sequence	SO:0001907	Feature elongation
feature_truncation	A sequence variant that causes the reduction of a genomic feature, with regard to the reference sequence	SO:0001906	Feature truncation
intergenic_variant	A sequence variant located in the intergenic region, between genes	SO:0001628	Intergenic

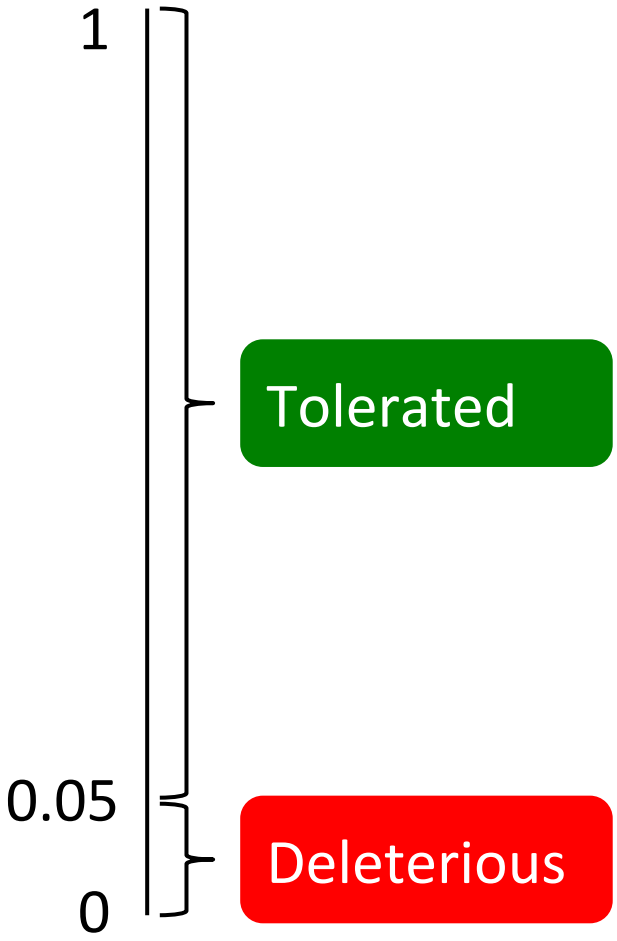
http://www.ensembl.org/info/docs/variation/predicted_data.html

SIFT and PolyPhen

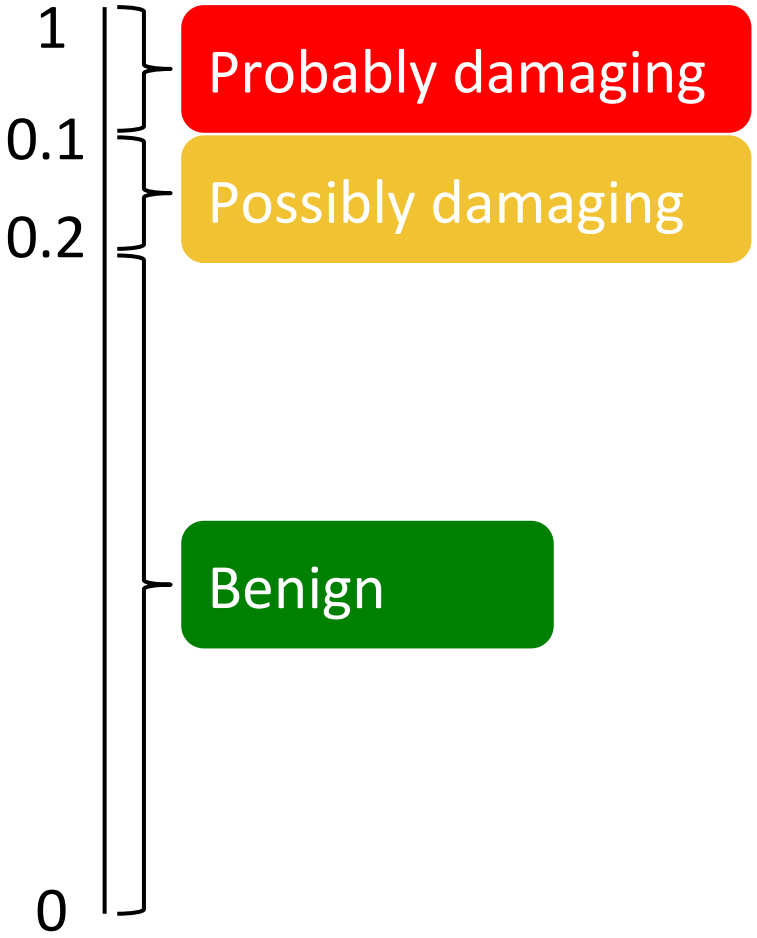
SIFT and PolyPhen score changes in amino acid sequence based on:

- How well conserved the protein is
- The chemical change in the amino acid

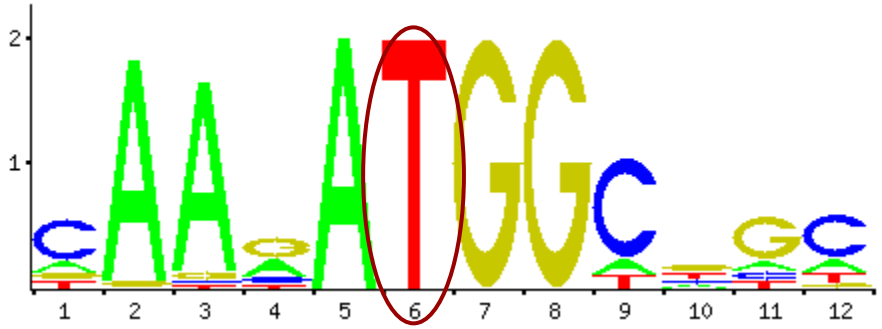
SIFT



PolyPhen



Does it affect TF binding?



GAACATGGCGGC

Score = 10.414

T/C

GAACACGGCGGC

Score = 10.329

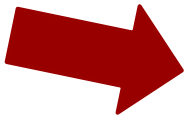
Is it known?

http://www.ensembl.org/info/docs/variation/sources_documentation.html

Use the VEP

Variant Effect Predictor

VeP



Web interface

- Point-and-click interface
- Suits smaller volumes of data

Documentation
Launch the web interface

Standalone perl script

- More options, more flexibility
- For large volumes of data

Documentation
Download latest version

http://www.ensembl.org/info/docs/tools/vep/index.html

Set up a cache

- Speed up your VEP script with an offline cache.
- Use prebuilt caches for Ensembl species.
- Or make your own from GTF and FASTA files - even for genomes not in Ensembl.

http://www.ensembl.org/info/docs/tools/vep/script/vep_cache.html

Hands on

- We're going to look at a set of four variants to find out what genes they hit and what effect they have on them.

  9 128328461 128328461 A/- + var1

  9 128322349 128322349 C/A + var2

  9 128323079 128323079 C/G + var3

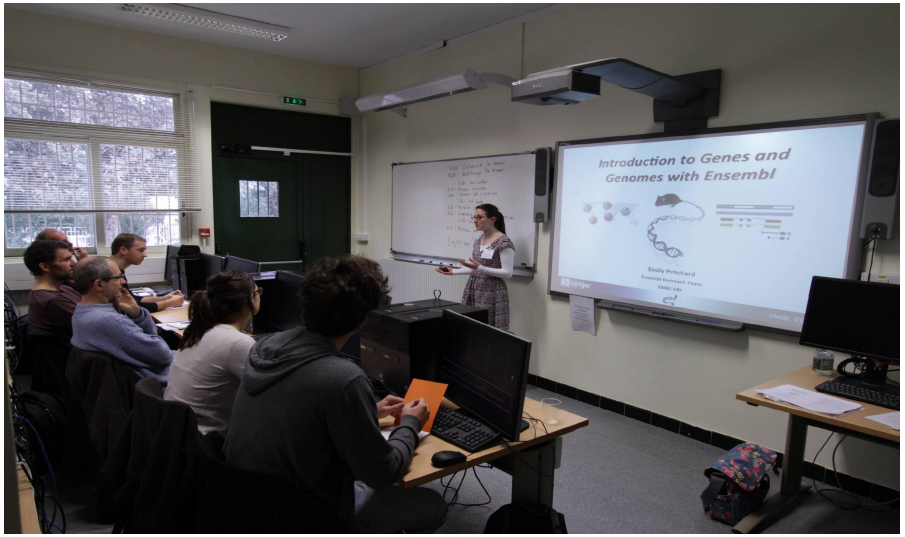
  9 128322917 128322917 G/A + var4

  9 128322495 128322495 A/G + var5

Host an Ensembl course

We can teach an Ensembl course at your institute for free (except trainers' expenses).

Email **emily@ebi.ac.uk**



Browser course

½-2 day course on the Ensembl browser, aimed at wet-lab scientists.

One trainer.

API course

1-4 day course on the Ensembl Perl API, aimed at bioinformaticians.

1-4 trainers.

Help and documentation

Course online http://www.ebi.ac.uk/training/online/subjects/11

Tutorials www.ensembl.org/info/website/tutorials

Flash animations

www.youtube.com/user/EnsemblHelpdesk

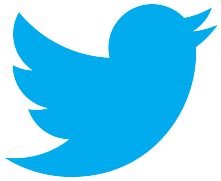
http://u.youku.com/Ensemblhelpdesk

Email us helpdesk@ensembl.org

Ensembl public mailing lists dev@ensembl.org, announce@ensembl.org

Follow us

www.facebook.com/Ensembl.org



@Ensembl

www.ensembl.info

Publications

http://www.ensembl.org/info/about/publications.html

Cunningham, F. *et al*

Ensembl 2015

Nucleic Acids Research

http://nar.oxfordjournals.org/content/early/2014/10/28/nar.gku1010.full?keytype=ref&ijkey=VOWPjAJSu1p1S5M

Xosé M. Fernández-Suárez and Michael K. Schuster
Using the Ensembl Genome Server to Browse Genomic Sequence Data.
Current Protocols in Bioinformatics 1.15.1-1.15.48 (2010)
www.ncbi.nlm.nih.gov/pubmed/20521244

Giulietta M Spudich and Xosé M Fernández-Suárez
Touring Ensembl: A practical guide to genome browsing
BMC Genomics **11**:295 (2010)
www.biomedcentral.com/1471-2164/11/295

Ensembl 2015

Acknowledgements

Ensembl 2015

Fiona Cunningham[1], M. Ridwan Amode[1], Daniel Barrell[1,2], Kathryn Beal[1], Konstantinos Billis[1], Simon Brent[2], Denise Carvalho-Silva[1], Peter Clapham[2], Guy Coates[2], Stephen Fitzgerald[1], Laurent Gil[1], Carlos García Girón[1], Leo Gordon[1], Thibaut Hourlier[1], Sarah E. Hunt[1], Sophie H. Janacek[1], Nathan Johnson[1], Thomas Juettemann[1], Andreas K. Kähäri[2], Stephen Keenan[1], Fergal J. Martin[1], Thomas Maurel[1], William McLaren[1], Daniel N. Murphy[1,2], Rishi Nag[1], Bert Overduin[1], Anne Parker[1], Mateus Patricio[1], Emily Perry[1], Miguel Pignatelli[1], Harpreet Singh Riat[1], Daniel Sheppard[1], Kieron Taylor[1], Anja Thormann[1], Alessandro Vullo[1], Steven P. Wilder[1], Amonida Zadissa[1], Bronwen L. Aken[1], Ewan Birney[1], Jennifer Harrow[2], Rhoda Kinsella[1], Matthieu Muffato[1], Magali Ruffier[1], Stephen M.J. Searle[2], Giulietta Spudich[1], Stephen J. Trevanion[1], Andy Yates[1], Daniel R. Zerbino[1] and Paul Flicek[1,2,*]

Funding

wellcome trust

EMBL

National Human Genome Research Institute

BBSRC bioscience for the future

European Commission
Framework Programme 7

BLUEPRINT epigenome

EURATRANS

SEVENTH FRAMEWORK PROGRAMME

HELIX NEBULA THESCIENCECLOUD