

Training materials

- Ensembl training materials are protected by a CC BY license
- <http://creativecommons.org/licenses/by/4.0/>
- If you wish to re-use these materials, please credit Ensembl for their creation
- If you use Ensembl for your work, please cite our papers
- <http://www.ensembl.org/info/about/publications.html>



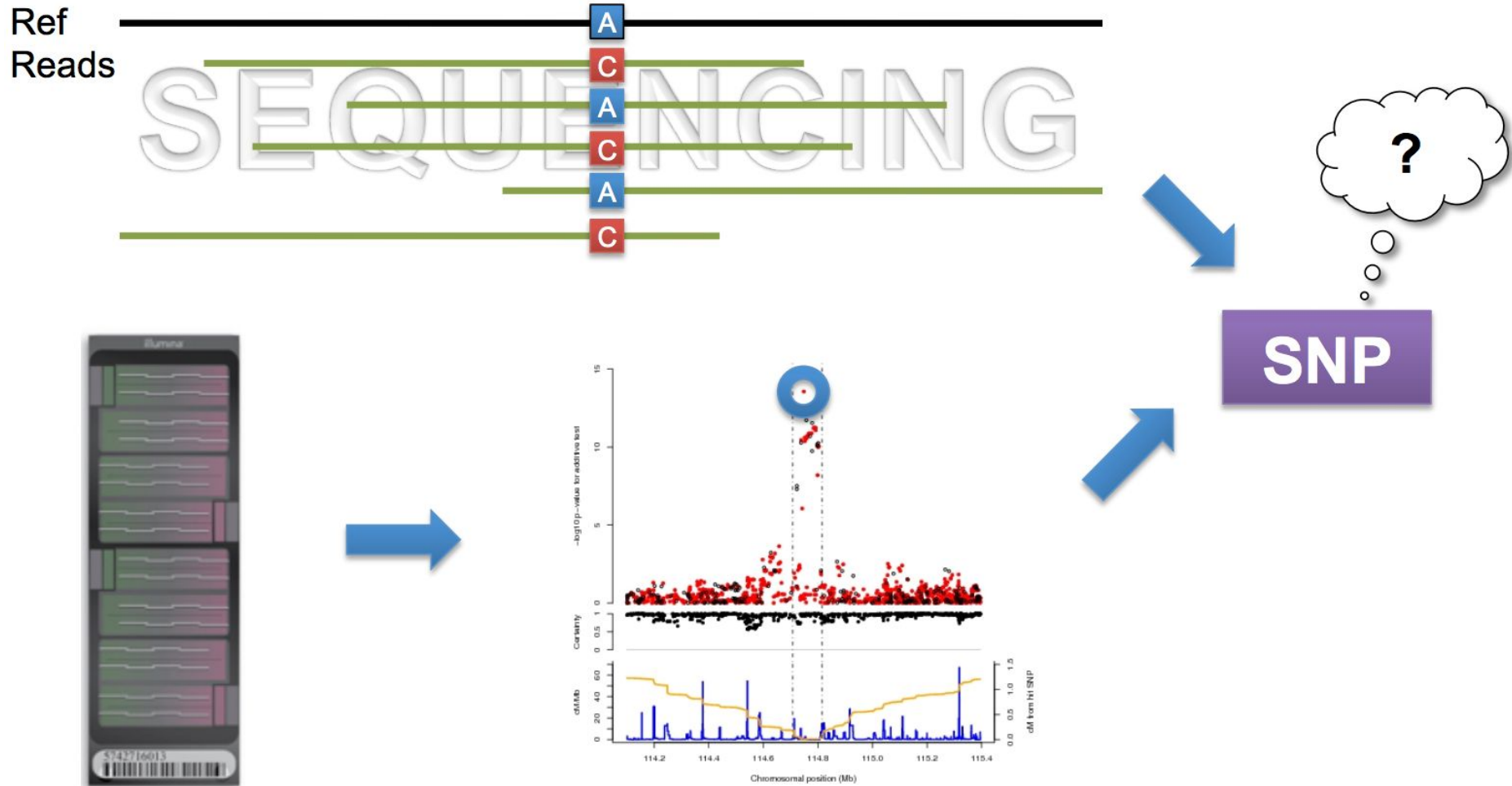
<http://training.ensembl.org/events>

Variant Annotation: Variant Effect Predictor (VEP)

Benjamin Moore
Ensembl Outreach Officer

Slides available from
training.ensembl.org/events/

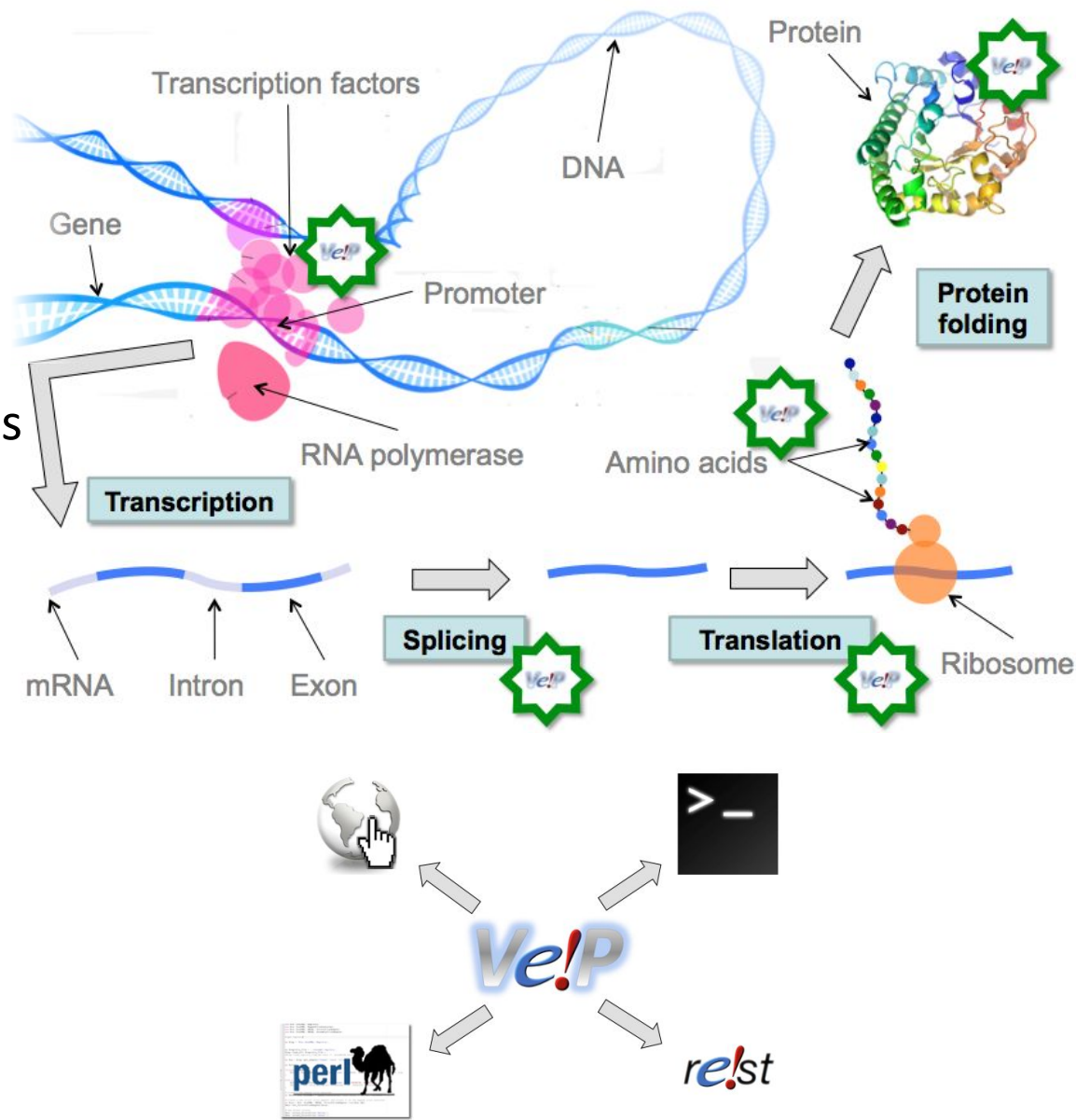
Background



<http://training.ensembl.org/events>

What is VEP?

- Variant Effect Predictor
- Predicts the functional effects of genomic variants
- Interpretation on
 - genes
 - regulatory features
 - known variants
- Developed by Ensembl
 - standalone tool
 - web interface
 - APIs (perl, RESTful)



<http://training.ensembl.org/events>

What is the VEP?

A tool to predict and annotate the functional consequences of variants
(SNPs, insertions, deletions, CNVs or structural variants)

Data input

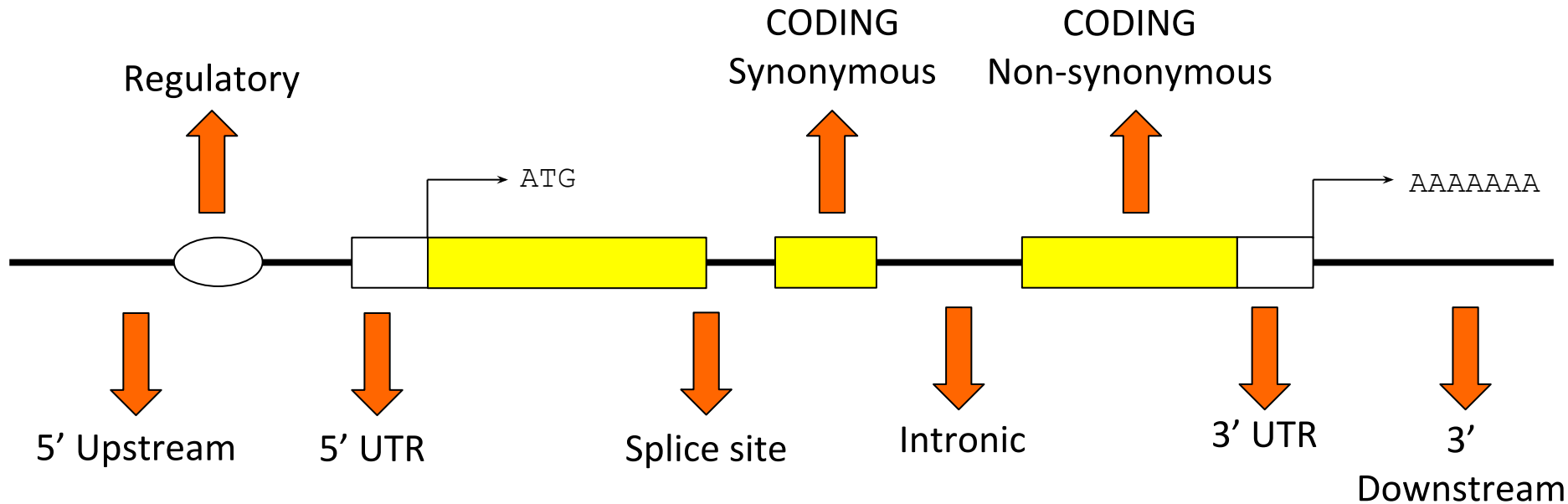
- Variant coordinates
- VCF
- HGVS
- Variant IDs



- Affected gene, transcript and protein sequence
- Pathogenicity 
- Frequency data 
- Regulatory consequences
- Splicing consequences
- Literature citations 

<http://training.ensembl.org/events>

Variation consequences



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 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

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

<http://training.ensembl.org/events>

VEP features

- Over 5,000 species
- Consequence predictions for different transcript sets
- Regulatory region consequence predictions
- Known variants
- Allele frequencies
- Pathogenicity predictions
- Phenotype/disease, clinical significance
- Huge VEP user-base

- 
- Ensembl
 - RefSeq
 - Merged
 - GENCODE basic

- 
- The Ensembl Regulatory Build:
- ENCODE
 - BLUEPRINT
 - NIH Epigenomics Roadmap

Can be limited to regulatory regions observed in specific cell types.

<http://training.ensembl.org/events>

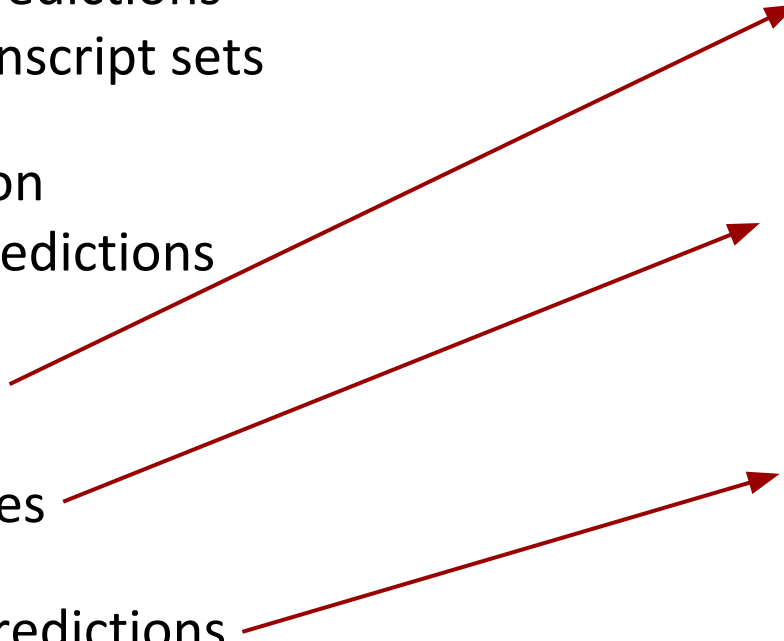
VEP features

- Over 5,000 species
- Consequence predictions for different transcript sets
- Regulatory region consequence predictions
- Known variants
- Allele frequencies
- Pathogenicity predictions
- Phenotype/disease, clinical significance
- Huge VEP user-base

- dbSNP
- Cosmic
- Clinvar
- ESP
- HGMD-Public
- Phencode

- 1000 Genomes
- ESP
- ExAC projects
- GnomAD

- Built in
 - SIFT
 - PolyPhen
- via plugins
 - CADD
 - FATHMM
 - LRT
 - MutationTaster
 - many more!



<http://training.ensembl.org/events>

VEP features

- Over 5,000 species
- Consequence predictions for different transcript sets
- Regulatory region consequence predictions
- Known variants
- Allele frequencies
- Pathogenicity predictions
- Phenotype/disease, clinical significance
- Huge VEP user-base

- OMIM
- Orphanet
- GWAS catalog
- ClinVar

- 1000 Genomes
- EVA
- ExAC
- DECIPHER
- OpenTargets
- LRG
- GnomAD

- Illumina
- CRUK
- Congenica
- Omicia
- Personalis

<http://training.ensembl.org/events>

A Quick Demo - the web interface



<http://training.ensembl.org/events>

Variant Effect Predictor ?

VEP for Human GRCh37

If you are looking for VEP for Human GRCh37, please go to [GRCh37 website](#).

Species:

 Human (Homo sapiens) 

Assembly: GRCh38.p10

Name for this job (optional):

Either paste data:

```
9 128328461 128328461 A/- + var1
9 128322349 128322349 C/A + var2
9 128323079 128323079 C/G + var3
9 128322917 128322917 G/A + var4
```

[Run instant VEP for current line](#)

Examples: [Ensembl default](#), [VCF](#), [Variant Identifiers](#), [HGVS notations](#),
NB: pileup format no longer supported

Or upload file:

no file selected

Or provide file URL:

Transcript database to use:

- ☒ Ensembl transcripts
- ☐ Gencode basic transcripts
- ☐ RefSeq transcripts
- ☐ Ensembl and RefSeq transcripts

Species and
assembly of
interest

Paste, upload or
link to data

<http://training.ensembl.org/events>

Identifiers

Gene symbol:



CCDS:



Protein:



Uniprot:



HGVS:

CSN^(p):Unshifted HGVS^(p):Population
frequency data

Frequency data

Find co-located known variants:

Yes



Frequency data for co-located variants:



1000 Genomes global minor allele frequency



1000 Genomes continental allele frequencies



ESP allele frequencies



gnomAD (exomes) allele frequencies

PubMed IDs for citations of co-located
variants:

Include flagged variants:

(p) = functionality from [VEP plugin](#)

Literature citations

Miscellaneous

Transcript biotype:	☒
Protein domains:	☐
Exon and intron numbers:	☐
Transcript support level:	☒
APPRIS:	☒
Identify canonical transcripts:	☐
Upstream/Downstream distance (bp):	5000
miRNA structure ^(p) :	☐

Pathogenicity predictions

SIFT:	Prediction and score
PolyPhen:	Prediction and score
dbNSFP ^(p) :	☒ Disabled ☐ Enabled
Condel ^(p) :	☒ Disabled ☐ Enabled
LoFtool ^(p) :	☐

Pathogenicity predictions

Regulatory data

Get regulatory region consequences:	Yes
-------------------------------------	----------------------------------

Splicing predictions

dbSnpSV ^(p) :	☐
MaxEntScan ^(p) :	☐

Splicing predictions

Conservation

BLOSUM62 ^(p) :	☐
---------------------------	--------------------------

^(p) = functionality from [VEP plugin](#)

<http://training.ensembl.org/events>

Filtering options  *Pre-filter results by frequency or consequence type*

Filters

Filter by frequency:

- ☒ No filtering
- ☐ Exclude common variants
- ☐ Advanced filtering

Return results for variants in coding regions only:

☐

Restrict results:

Show all results 

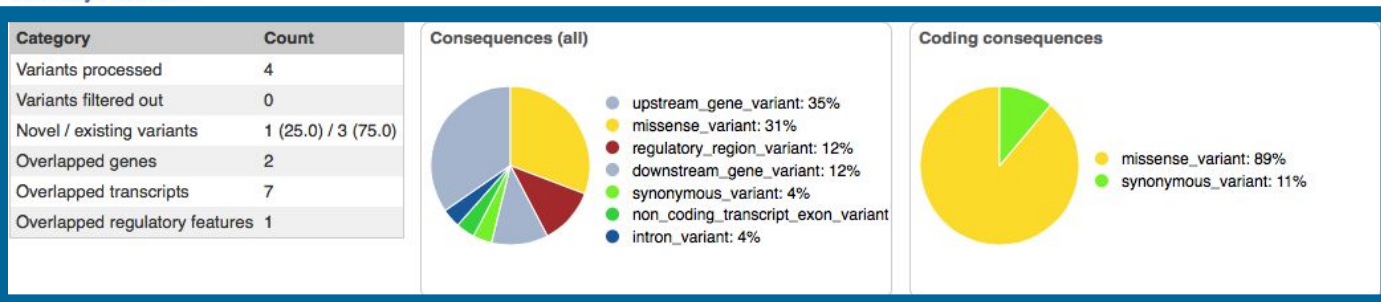
NB: Restricting results may exclude biologically important data!

[Run](#) [Clear](#) [Close form](#)

Variant Effect Predictor results

[Job details](#)

[Summary statistics](#)



Summary statistics

Results preview

Navigation

Page: 1 of 1 | Show: 1 All variants

Filters

Uploaded variant is defined

Download

All: [VCF](#) [VEP](#) [TXT](#)

BioMart: [Variants](#) [Genes](#)

Download data

Results columns

Show/hide columns (3 hidden)

Uploaded variant	Location	Allele	Consequence	Impact	Symbol	Gene	Feature type	Feature	Biotype	Exon	cDNA position	CDS position	Protein position	Amino acids
var4	9:128322917-128322917	A	missense_variant	MODERATE	COQ4	ENSG00000167113	Transcript	ENST00000300452	protein_coding	1/7	382	59	20	R/Q
var4	9:128322917-128322917	A	missense_variant	MODERATE	COQ4	ENSG00000167113	Transcript	ENST00000372875	protein_coding	1/4	66	59	20	R/Q
var4	9:128322917-128322917	A	upstream_gene_variant	MODIFIER	TRUB2	ENSG00000167112	Transcript	ENST00000372890	protein_coding	-	-	-	-	-
var4	9:128322917-128322917	A	upstream_gene_variant	MODIFIER	TRUB2	ENSG00000167112	Transcript	ENST00000460320	processed_t	-	-	-	-	-
var4	9:128322917-128322917	A	missense_variant	MODERATE	COQ4	ENSG00000167113	Transcript	ENST00000608951	protein_coding	1/3	336	59	20	R/Q
var4	9:128322917-128322917	A	missense_variant	MODERATE	COQ4	ENSG00000167113	Transcript	ENST00000609948	protein_coding	1/2	374	59	20	R/Q
var4	9:128322917-128322917	A	regulatory_region_variant	MODIFIER	-	-	RegulatoryFeature	ENSR00000241858	promoter	-	-	-	-	-
var3	9:128323079-128323079	G	missense_variant	MODERATE	COQ4	ENSG00000167113	Transcript	ENST00000300452	protein_coding	2/7	457	134	45	S/C
var3	9:128323079-128323079	G	missense_variant	MODERATE	COQ4	ENSG00000167113	Transcript	ENST00000372875	protein_coding	2/4	141	134	45	S/C
var3	9:128323079-128323079	G	upstream_gene_variant	MODIFIER	TRUB2	ENSG00000167112	Transcript	ENST00000372890	protein_coding	-	-	-	-	-
var3	9:128323079-128323079	G	upstream_gene_variant	MODIFIER	TRUB2	ENSG00000167112	Transcript	ENST00000460320	processed_transcript	-	-	-	-	-

More columns!

<http://training.ensembl.org/events>

The VEP script

* exact path may vary

```
$ cd ~/ensembl-tools/scripts/variant_effect_predictor/  
$ perl variant_effect_predictor.pl --help  
  
$ perl variant_effect_predictor.pl --cache -i example_GRCh38.vcf  
$ less variant_effect_output.txt
```

- Documentation:
 - <http://www.ensembl.org/info/docs/tools/vep/script/index.html>
 - http://www.ensembl.org/info/docs/tools/vep/script/vep_options.html

<http://training.ensembl.org/events>

#Uploaded_variation	Location	Allele	Gene	Feature	Feature_type
rs116645811	21:25587758	A	ENSG00000154719	ENST00000307301	Transcript
rs116645811	21:25587758	A	ENSG00000154719	ENST00000352957	Transcript
rs116645811	21:25587758	A	ENSG00000260583	ENST00000567517	Transcript
rs1135638	21:25592836	A	ENSG00000154719	ENST00000307301	Transcript
rs1135638	21:25592836	A	ENSG00000154719	ENST00000352957	Transcript
rs1135638	21:25592836	A	ENSG00000154719	ENST00000419219	Transcript
rs10576	21:25592860	C	ENSG00000154719	ENST00000307301	Transcript
rs10576	21:25592860	C	ENSG00000154719	ENST00000352957	Transcript
rs10576	1:25592860	C	ENSG00000154719	ENST00000419219	Transcript

chr:start[-end]

Consequence	cDNA_position	CDS_position	Protein_position	Amino_acids	Codons
missense_variant	1043	1001	334	T/M	aCg/aTg
intron_variant	-	-	-	-	-
upstream_gene_variant	-	-	-	-	-
synonymous_variant	939	897	299	G	ggC/ggT
synonymous_variant	939	897	299	G	ggC/ggT
synonymous_variant	876	867	289	G	ggC/ggT
synonymous_variant	815	873	291	P	ccA/ccG
synonymous_variant	873	873	291	P	ccA/ccG
synonymous_variant	852	843	281	P	ccA/ccG

Sequence ontology

Variant alleles uppercase

Existing_variation	Extra
rs116645811	IMPACT=MODERATE;STRAND=-1
rs116645811	IMPACT=MODIFIER;STRAND=-1
rs116645811	IMPACT=MODIFIER;DISTANCE=4432;STRAND=-1
rs1135638,COSM5618718,COSM5618719	IMPACT=LOW;STRAND=-1;SOMATIC=0,1,1;PHENO=0,1,1
rs1135638,COSM5618718,COSM5618719	IMPACT=LOW;STRAND=-1;SOMATIC=0,1,1;PHENO=0,1,1
rs1135638,COSM5618718,COSM5618719	IMPACT=LOW;STRAND=-1;FLAGS=cds_end_NF;SOMATIC=0,1,1;PHENO=0,1,1
rs10576	IMPACT=LOW;STRAND=-1
rs10576	IMPACT=LOW;STRAND=-1
rs10576	IMPACT=LOW;STRAND=-1;FLAGS=cds_end_NF

Variants from multiple sources

Basic flags

- Annotation source - choose one
 - `--cache` - use local data, use DB connections for certain functions
 - `--offline` - use local data, forbid external DB connections
 - `--database` - use remote DB (default `ensembl.db.ensembl.org`)
- Input/output
 - `--input_file (-i)` - will try to read from STDIN if absent
 - `--output_file (-o)` - defaults to “`variant_effect_output.txt`”
 - `--force_overwrite` - overwrite existing output file
 - `-o STDOUT` - write to terminal or pipe
 - `--tab`, `--vcf`, `--json` - different output formats, customise with `--fields`
- Go faster
 - `--fork 4`

<http://training.ensembl.org/events>

More annotation, more!

- Known variants
 - `--check_existing` - enable checking for known variants
 - `--gmaf`, `--maf_1kg`, `--maf_exac`, `--maf_esp` - frequency data
 - Can use to filter with `--filter_common`
- Pathogenicity predictions
 - `--sift b`, `--polyphen b` - score and prediction
- Regulatory data
 - `--regulatory`
- HGVS notations
 - `--hgvs`
- `--everything !!!`

<http://training.ensembl.org/events>

Filtering

- VEP comes with its own filter script
 - http://www.ensembl.org/info/docs/tools/vep/script/vep_filter.html
- Works with default and VCF output
- Simple query notation - [field] [operator] [value]

```
$ perl filter_vep.pl -i v.vcf -f "Consequence is missense_variant"
```

- Queries can be combined with "and" / "or"
- Nest queries with parentheses
- Resolve consequence types in ontology (--ontology / -y)

```
[...] -y -f "Consequence is coding_sequence_variant or (EXON is 1 and BIOTYPE is protein_coding)"
```

Annotation sources

- Limit gene set
 - `--gencode_basic` - high quality subset of GENCODE
- Use RefSeq transcripts in place of Ensembl transcripts
 - `--refseq` - must download cache first
 - `--merged` - both Ensembl and RefSeq
 - **WARNING:** RefSeq alignments to genome are imperfect, some annotations may be inaccurate, see flags for info
- Custom annotations
 - tabix indexed: BED, GFF, VCF
 - bigWig

Plugins

- Extend functionality of VEP
 - run algorithms - LD, GeneSplicer, MaxEntScan
 - fetch Ensembl data - Conservation, Phenotypes
 - fetch external data - dbNSFP, CADD, GXA
 - modify parameters - UpDownStream
- Set up with installer or download
 - https://github.com/Ensembl/VEP_plugins
- Simple to add to command

```
$ perl variant_effect_predictor.pl --cache -i example_GRCh38.vcf  
--plugin CADD,CADD.txt.gz --plugin UpDownStream,1000
```

The VEP REST API

VEP

Different GET and POST requests for submitting HGVS notation, variant IDs or variant coordinates to VEP

Resource	Description
GET vep/:species/hgvs/:hgvs_notation	Fetch variant consequences based on a HGVS notation
POST vep/:species/hgvs	Fetch variant consequences for multiple HGVS notations
GET vep/:species/id/:id	Fetch variant consequences based on a variant identifier
POST vep/:species/id	Fetch variant consequences for multiple ids
GET vep/:species/region/:region/:allele/	Fetch variant consequences
POST vep/:species/region	Fetch variant consequences for multiple regions

- Documentation:
 - <http://rest.ensembl.org>

<http://training.ensembl.org/events>

The VEP REST API

Required parameters

GET vep/:species/id/:id

Fetch variant consequences based on a variant identifier

Parameters

Required

Name	Type	Description	Default	Example Values
id	String	Query ID. Supports dbSNP, COSMIC and HGMD identifiers	-	rs56116432 COSM476
species	String	Species name/alias	-	homo_sapiens human

<http://training.ensembl.org/events>

The VEP REST API

Optional

Name	Type	Description
Blosum62	Boolean	Include BLOSUM62 amino acid conservation score (plugin details)
CSN	Boolean	Reports Clinical Sequencing Nomenclature (CSN) for variants (plugin details)
Conservation	Boolean	Retrieves a conservation score from the Ensembl Compara databases for variant positions (plugin details)
GeneSplicer	Boolean	Detects splice sites in genomic DNA (plugin details)
MaxEntScan	Boolean	Sequence motif and maximum entropy based splice site consensus predictions (plugin details)
appris	Boolean	Include APPRIS isoform annotation
callback	String	Name of the callback subroutine to be returned by the requested JSONP response. Required ONLY when using JSONP as the serialisation method. Please see the user guide .
canonical	Boolean	Include a flag indicating the canonical transcript for a gene
ccds	Boolean	Include CCDS transcript identifiers
dbNSFP	String	Include fields from dbNSFP, a database of pathogenicity predictions for missense variants. Multiple fields should be separated by commas. See dbNSFP README for field list. (plugin details)
dbscSNV	Boolean	Predictions for splicing variants from dbscSNV. (plugin details)
distance	Integer	Change the distance to transcript for which VEP assigns upstream and downstream consequences
domains	Boolean	Include names of overlapping protein domains
failed	Boolean	When checking for co-located variants, by default variants flagged as failed by Ensembl's QC pipeline will be excluded. Set this flag to 1 to include such variants
hgvs	Boolean	Include HGVS nomenclature based on Ensembl stable identifiers
merged	Boolean	Use merged Ensembl and RefSeq transcript set to report consequences (human only)
miRNA	Boolean	Determines where in the secondary structure of a miRNA a variant falls (plugin details)
minimal	Boolean	Convert alleles to their most minimal representation before consequence calculation i.e. sequence that is identical between each pair of reference and alternate alleles is trimmed off from both ends, with coordinates adjusted accordingly. Note this may lead to discrepancies between input coordinates and coordinates reported by VEP relative to transcript sequences
numbers	Boolean	Include affected exon and intron positions within the transcript
protein	Boolean	Include Ensembl protein identifiers
refseq	Boolean	Use RefSeq transcript set to report consequences (human only)
tsl	Boolean	Include transcript support level (TSL) annotation
uniprot	Boolean	Include best match accessions for translated protein products from three UniProt-related databases (SWISSPROT, TREMBL and UniParc)
variant_class	Boolean	Output the Sequence Ontology variant class for the input variant
xref_refseq	Boolean	Include aligned RefSeq mRNA identifiers for transcript. NB: theRefSeq and Ensembl transcripts aligned in this way MAY NOT, AND FREQUENTLY WILL NOT, match exactly in sequence, exon structure and protein product

<http://training.ensembl.org/events>



The VEP REST API

</vep/human/id/rs56116432?content-type=application/json>

Example request

Example output

Perl

Python2

Python3

Ruby

Java

R

Curl

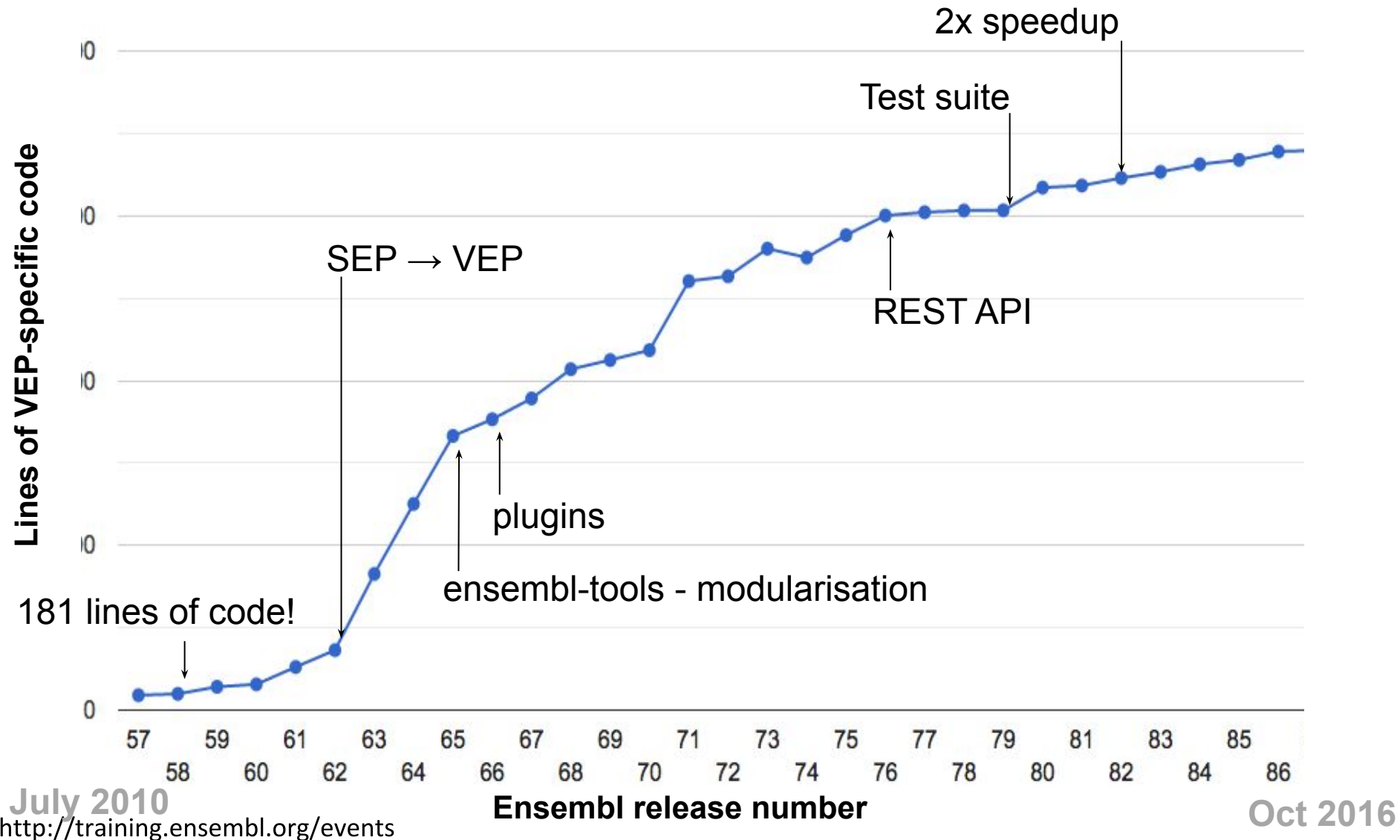
Wget

Example output

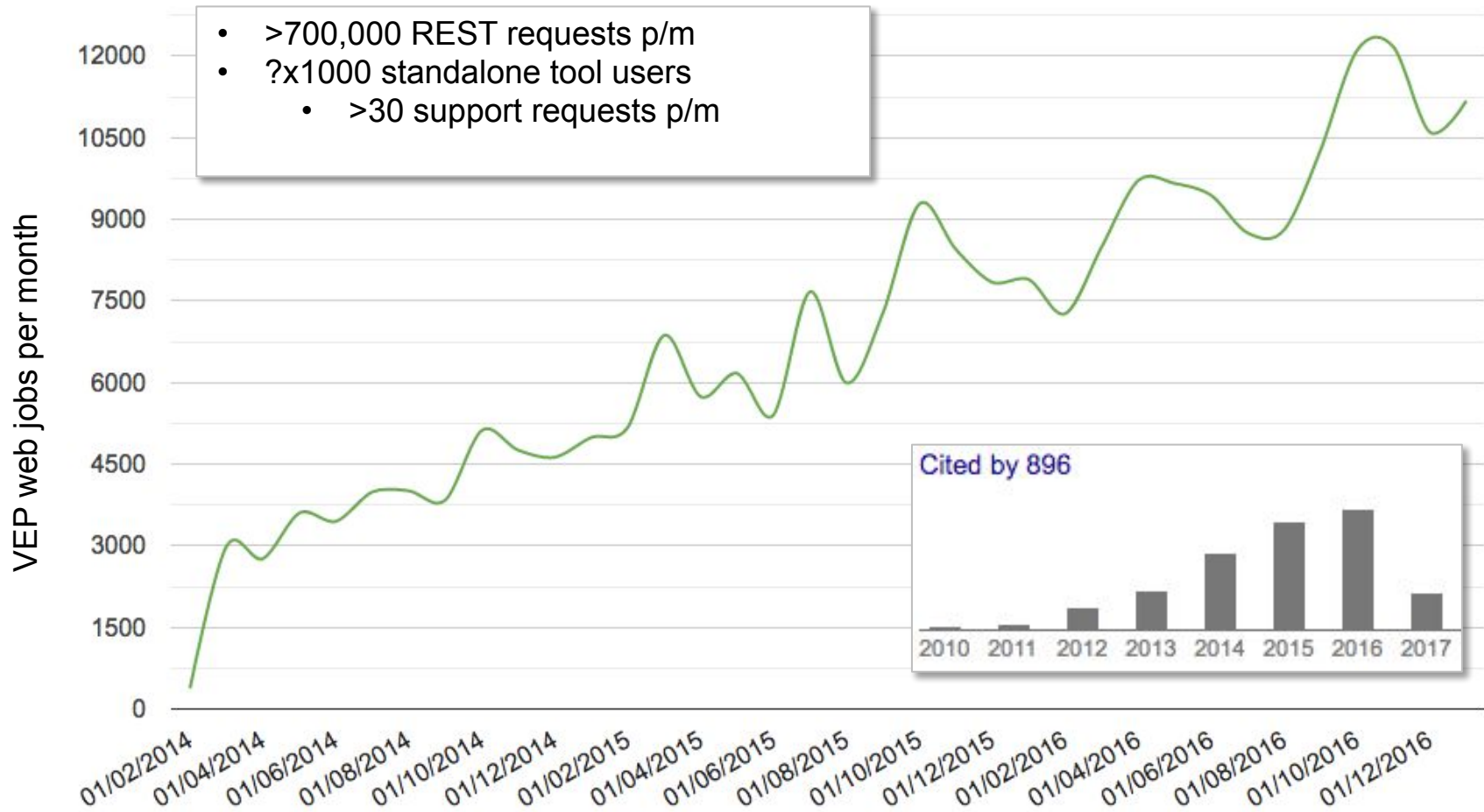
Example clients

```
[
  {
    "input": "rs56116432",
    "colocated_variants": [
      {
        "aa_maf": 0.0007102,
        "gnomad_nfe_maf": 0.003593,
        "ea_maf": 0.003809,
        "gnomad_nfe_allele": "T",
        "gnomad_afr_maf": 0.0006606,
        "eas_allele": "T",
        "amr_maf": 0.0014,
        "id": "rs56116432",
        "sas_allele": "T",
        "sas_maf": 0.001,
        "amr_allele": "T",
        "minor_allele_freq": 0.0026,
```

The organic growth of VEP



VEP usage continues to increase

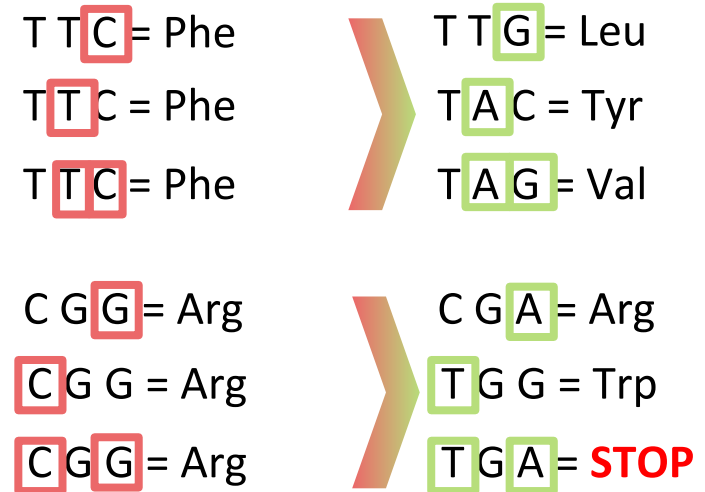


<http://training.ensembl.org/events>

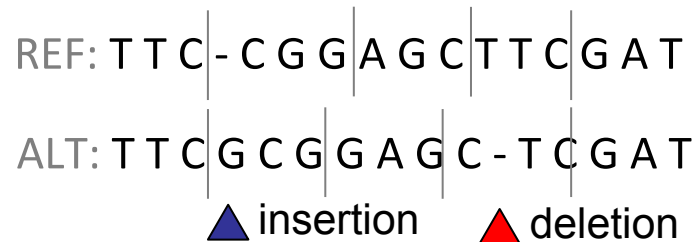
Beyond single variant analysis

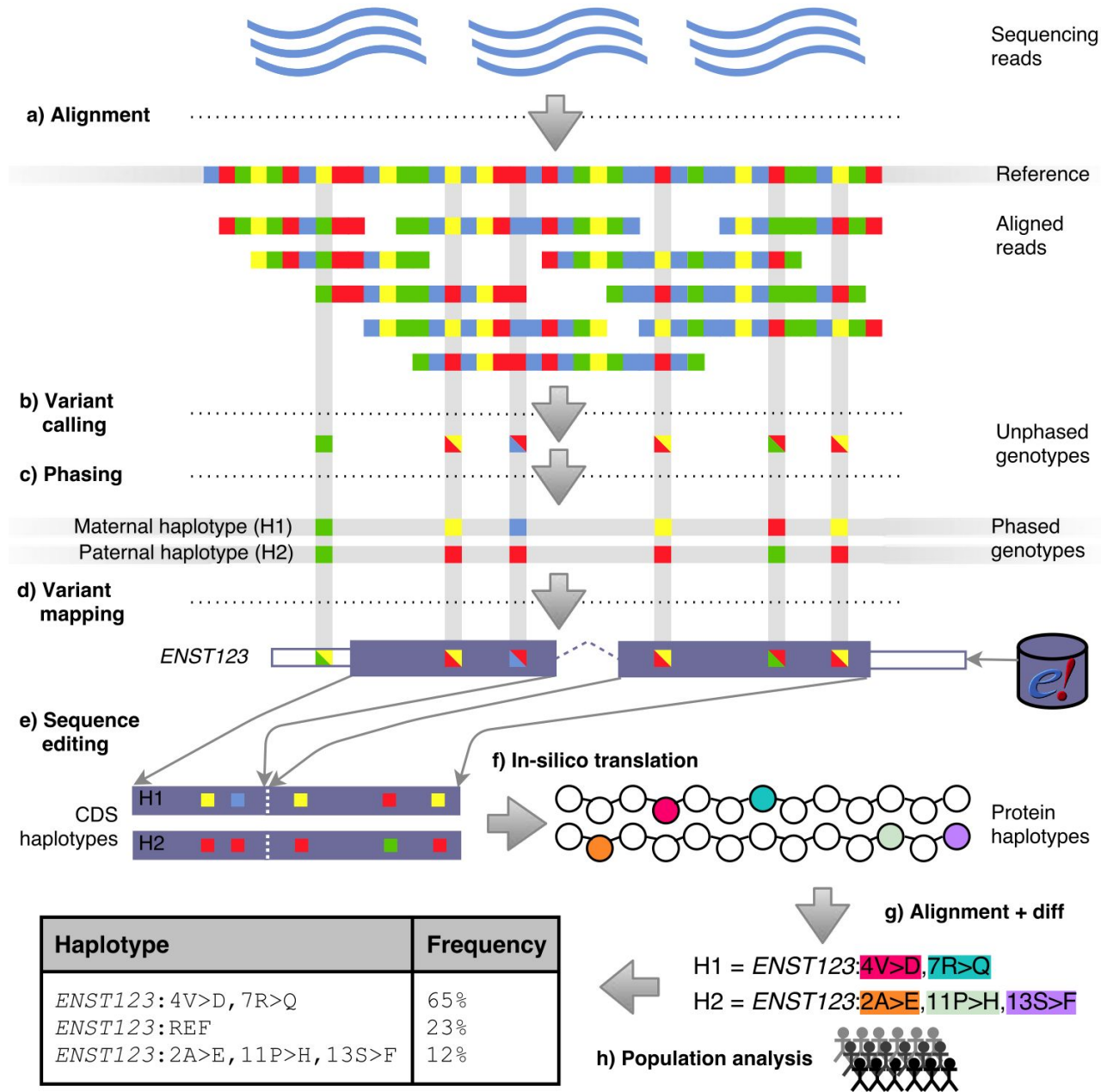
- Almost all current variant annotation is done on a per-variant basis, which misses:

- variants located in the same codon
 - incorrect missense prediction
 - cryptic stop gained or protein-truncating variant (PTV)
 - “rescued” PTV



- frame-rescuing in/del pairs





<http://training.ensembl.org/events>

Ensembl transcript haplotypes view

Haplotypes ?

 Export data as JSON

Gene: **FPR1**

www.ensembl.org/Homo_sapiens/Transcript/Haplotypes?t=ENST00000304748

[Switch to CDS view](#) 

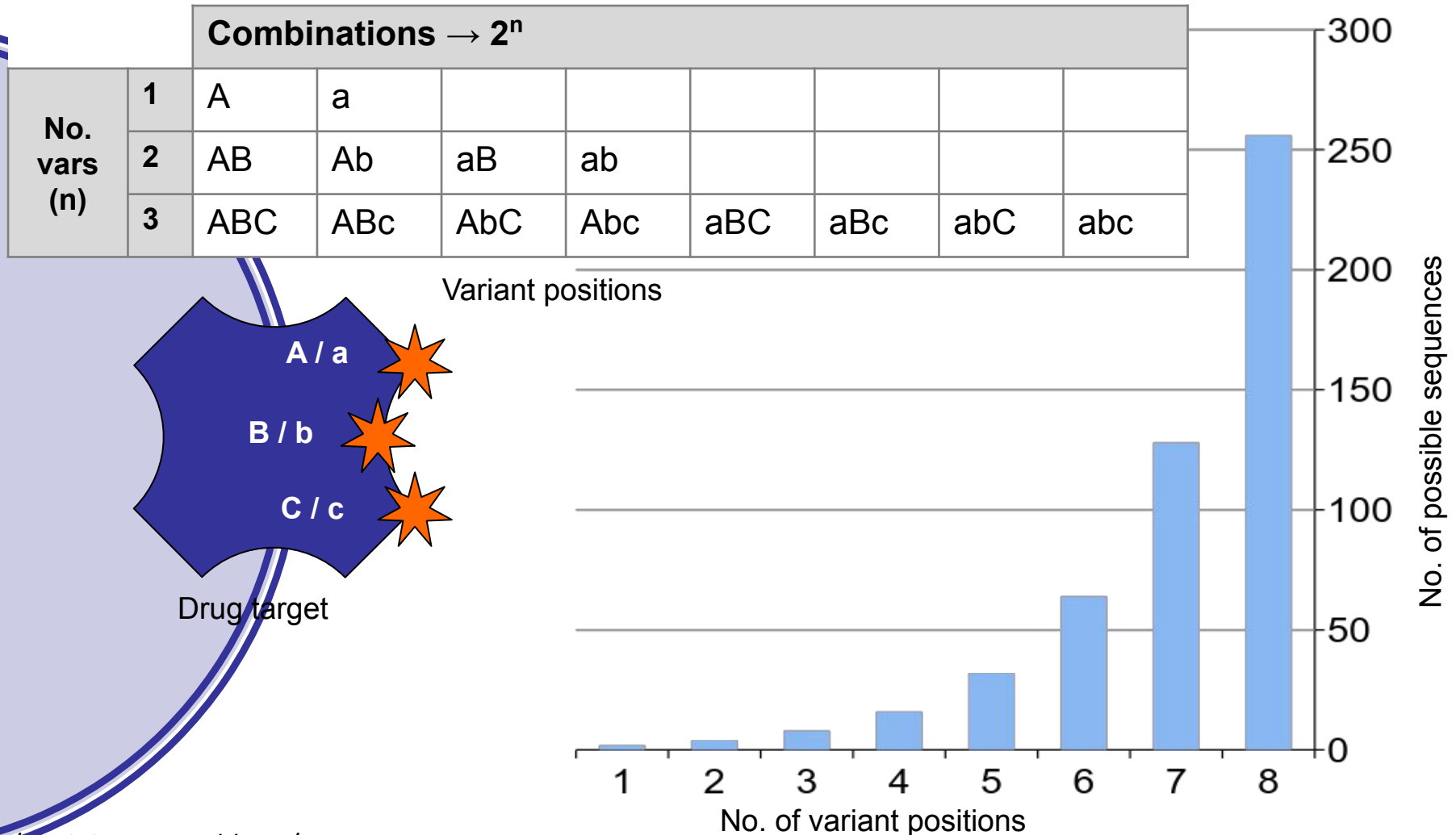
Show 10 entries		Show/hide columns		Filter			
Protein haplotype	Flags	Frequency (count)	AFR	AMR	EAS	EUR	SAS
11I>T,192N>K,346E>A	D	0.2 (1002)	0.226 (299)	0.15 (104)	0.121 (122)	0.261 (263)	0.219 (214)
11I>T,101V>L,346E>A	D	0.119 (595)	0.0454 (60)	0.235 (163)	0.115 (116)	0.128 (129)	0.13 (127)
11I>T,346E>A	D	0.115 (574)	0.168 (222)	0.117 (81)	0.0933 (94)	0.0676 (68)	0.111 (109)
11I>T		0.097 (486)	0.12 (158)	0.0576 (40)	0.124 (125)	0.0726 (73)	0.092 (90)
11I>T,101V>L,192N>K,346E>A	D	0.0883 (442)	0.116 (153)	0.0908 (63)	0.0823 (83)	0.0527 (53)	0.092 (90)
11I>T,190R>W,346E>A	D	0.0811 (406)	0.025 (33)	0.0706 (49)	0.0615 (62)	0.106 (107)	0.158 (155)
101V>L,346E>A	D	0.0623 (312)	0.00832 (11)	0.101 (70)	0.0198 (20)	0.127 (128)	0.0849 (83)
11I>T,101V>L		0.0381 (191)	0.0144 (19)	0.0346 (24)	0.0794 (80)	0.0388 (39)	0.0297 (29)
REF		0.0323 (162)	0.0136 (18)	0.0692 (48)	0.00397 (4)	0.0785 (79)	0.0133 (13)
101V>L,192N>K,346E>A	D	0.0292 (146)	0.0983 (130)	0.0159 (11)	0.000992 (1)		0.00409 (4)

Showing 1 to 10 of 82 entries

<< < 1 2 3 4 5 > >>

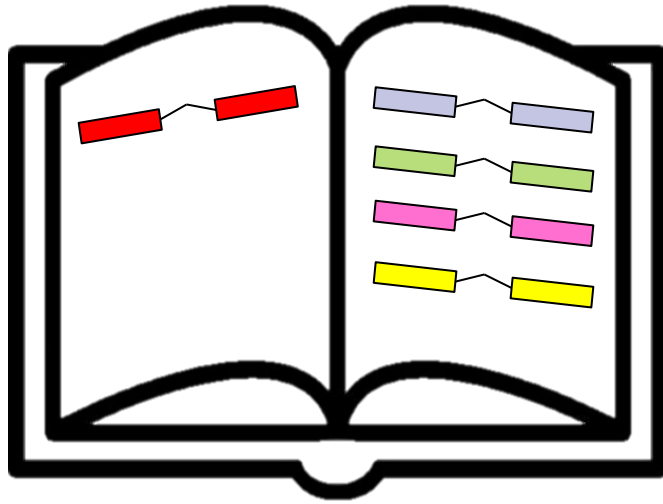
<http://training.ensembl.org/events>

Background: drug development

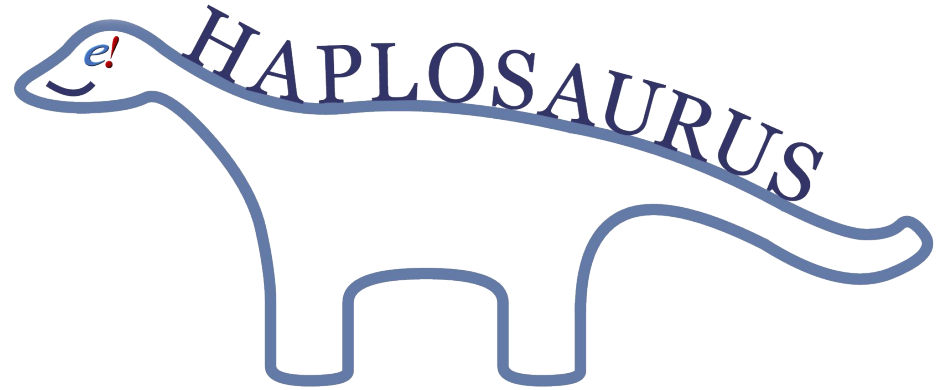


<http://training.ensembl.org/events>

Introducing: Haplosaurus



~~The~~sauros
Haplo

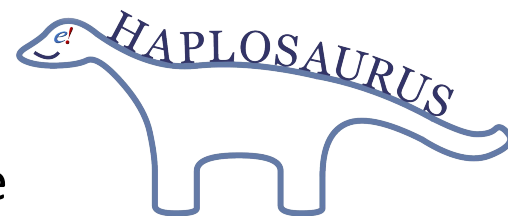


~~Stego~~saurus
Haplo

- VEP-like tool for haplotype analysis of phased data

<http://training.ensembl.org/events>

Haplosaurus examples



- cryptic LoF mutation in Von Willebrand Disease
- VEP:

rs1800380	12:6138595	T	ENST00000261405	synonymous_variant	2880	cgG/cgA
rs370984712	12:6138597	A	ENST00000261405	missense_variant	2878	Cgg/Tgg

- Haplosaurus:

ENST00000261405:2878C>T,2880G>A ENSP00000261405:960R>*,961del{1854} stop_change

- Frame rescuing indel pair (f = 10% in 1000 Genomes)
- VEP:

rs201941751	12:12630676-12630680	-	frameshift_variant	aGCGTG/a
rs200271649	12:12630682-12630685	-	frameshift_variant	gtGCCC/gt

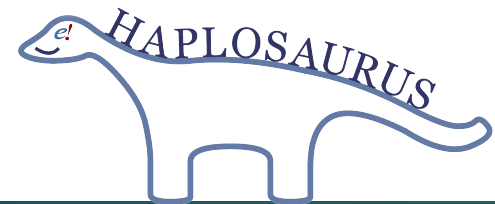
- Haplosaurus:

ENST00000228862:1080del{4},1085del{5} resolved_frameshift,indel
ENSP00000228862:361delPSV indel

<http://training.ensembl.org/events>

Want to use the VEP and Haplosaurus?

- VEP web interface, standalone tool, Ensembl REST and Perl APIs
- A new rewrite of the VEP codebase available! (beta)
 - <https://github.com/Ensembl/ensembl-vep>
 - Faster, more robust, more features, backward compatible
- Also includes Haplosaurus
 - generate whole-transcript haplotypes per individual from phased genotype data
 - overcomes limitations of current per-variant annotation



<http://training.ensembl.org/events>



Will McLaren

The Ensembl Variation Team

William Spooner

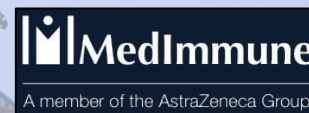


Robin Butler

Donna Finch

Catherine Huntington

Tim Slidel



EMBL-EBI



Ensembl Acknowledgements

The Entire Ensembl Team

Daniel R. Zerbino¹, Premanand Achuthan¹, Wasiu Akanni¹, M. Ridwan Amode¹, Daniel Barrell^{1,2}, Jyothish Bhai¹, Konstantinos Billis¹, Carla Cummins¹, Astrid Gall¹, Carlos García Giroñ¹, Laurent Gil¹, Leo Gordon¹, Leanne Haggerty¹, Erin Haskell¹, Thibaut Hourlier¹, Osagie G. Izuogu¹, Sophie H. Janacek¹, Thomas Juettemann¹, Jimmy Kiang To¹, Matthew R. Laird¹, Ilias Lavidas¹, Zhicheng Liu¹, Jane E. Loveland¹, Thomas Maurel¹, William McLaren¹, Benjamin Moore¹, Jonathan Mudge¹, Daniel N. Murphy¹, Victoria Newman¹, Michael Nuhn¹, Denye Ogeh¹, Chuang Kee Ong¹, Anne Parker¹, Mateus Patricio¹, Harpreet Singh Riat¹, Helen Schuilenburg¹, Dan Sheppard¹, Helen Sparrow¹, Kieron Taylor¹, Anja Thormann¹, Alessandro Vullo¹, Brandon Walts¹, Amonida Zadissa¹, Adam Frankish¹, Sarah E. Hunt¹, Myrto Kostadima¹, Nicholas Langridge¹, Fergal J. Martin¹, Matthieu Muffato¹, Emily Perry¹, Magali Ruffier¹, Dan M. Staines¹, Stephen J. Trevanion¹, Bronwen L. Aken¹, Fiona Cunningham¹, Andrew Yates¹ and Paul Flicek^{1,3}

¹European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK, ²Eagle Genomics Ltd., Wellcome Genome Campus, Hinxton, Cambridge CB10 1DR, UK and ³Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Funding



Co-funded by
the European
Union

Training materials

- Ensembl training materials are protected by a CC BY license
- <http://creativecommons.org/licenses/by/4.0/>
- If you wish to re-use these materials, please credit Ensembl for their creation
- If you use Ensembl for your work, please cite our papers
- <http://www.ensembl.org/info/about/publications.html>



<http://training.ensembl.org/events>