

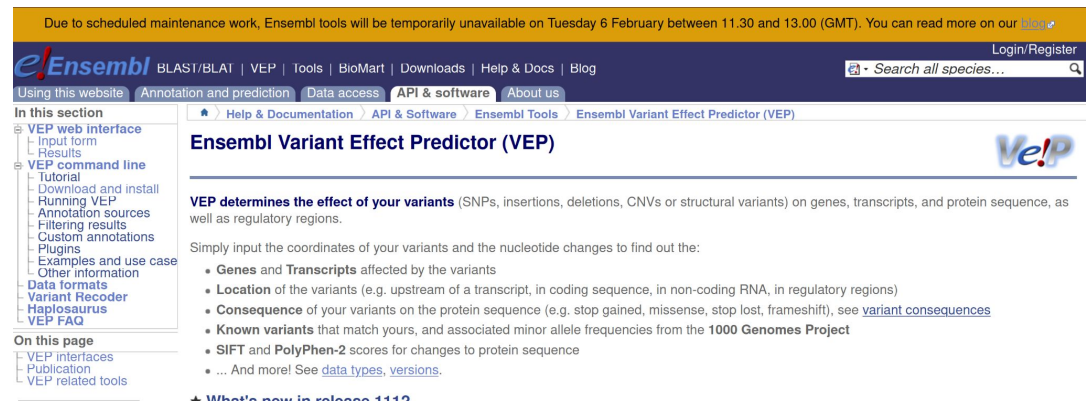


Anotação e Priorização de Variantes

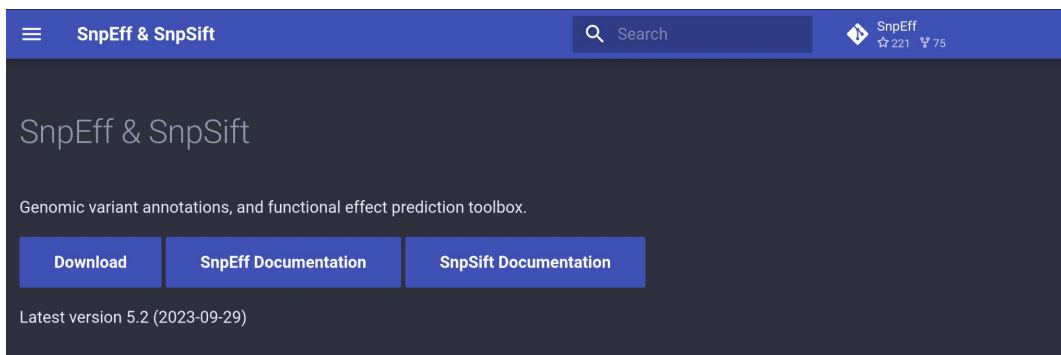
José Ferrão e Hugo Martiniano

Anotação

- Processo através do qual é feita a identificação do impacto das variantes e do seu significado funcional
- Existem várias ferramentas



<https://www.ensembl.org/info/docs/tools/vep/index.html>



<http://pcingola.github.io/SnpEff/>

VEP - Variant Effect Predictor

- Software para anotação de variantes
(<https://www.ensembl.org/info/docs/tools/vep/index.html>)



- Desenvolvido pelo Ensembl
(<https://www.ensembl.org/index.html>)

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Ensembl Release 111 (January 2024)

- MANE Select (v1.2) GRCh38.p14 patch annotation
- Human variation data updated to dbSNP156
- Updated genome assembly and annotation for sheep and cattle
- Regulatory annotation of open chromatin regions and promoters in common carp and rainbow trout (a collaboration with the AQUA-FAANG consortium)
- Updated regulatory annotation, including enhancers, for pig and chicken, Atlantic salmon, European seabass and turbot

Variant Effect Predictor

Web interface



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

 [Documentation](#)



Command line tool



- More options and flexibility
- For large volumes of data

 [Documentation](#)

 [Clone from GitHub](#)

 [Download \(zip\)](#)

 [Pull Docker image from](#)

REST API



- Language-independent API
- Simple URL-based queries

 [Documentation](#)

 [VEP REST API](#)



Variant Effect Predictor ⓘ

New job

Clear form

Species:

Homo_sapiens X

Assembly: GRCh38.p13

[Add/remove species](#)If you are looking for VEP for Human GRCh37, please go to [GRCh37 website](#).Reference species and
assembly selection

Name for this job (optional):

Input data:

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 >

Data input options

Examples: [Ensembl default](#), [VCF](#), [Variant identifiers](#), [HGVS notations](#), [SPOI](#)

Or upload file:

Choose file

No file chosen

Or provide file URL:

Transcript database to use:

- ☒ Ensembl/GENCODE transcripts
- ☐ Ensembl/GENCODE basic transcripts
- ☐ RefSeq transcripts
- ☐ Ensembl/GENCODE and RefSeq transcripts

Reference transcript
set options


Additional configurations:

- Identifiers** ⓘ Additional identifiers for genes, transcripts and variants
- Variants and frequency data** ⓘ Co-located variants and frequency data
- Additional annotations** ⓘ Additional transcript, protein and regulatory annotations
- Predictions** ⓘ Variant predictions, e.g. SIFT, PolyPhen
- Filtering options** ⓘ Pre-filter results by frequency or consequence type
- Advanced options** ⓘ Additional enhancements



Additional configuration
options

Run >


Variant Effect Predictor

Identifiers  Additional identifiers for genes, transcripts and variants


Identifiers

Gene symbol:	<input checked="" type="checkbox"/>	 Identifiers of overlapping/adjacent genes and transcripts selected by default
Transcript version:	<input checked="" type="checkbox"/>	
CCDS:	<input type="checkbox"/>	
Protein:	<input type="checkbox"/>	 Options for adding identifiers of overlapping/adjacent proteins to output
UniProt:	<input type="checkbox"/>	
HGVS:	<input type="checkbox"/>	

Variant Effect Predictor

[Variants and frequency data](#)  [Co-located variants and frequency data](#)

Variants and frequency data

Find co-located known variants: 
Option to find co-located known variants and report associated variant identifiers

Variant synonyms: ☐

Frequency data for co-located variants:

- ☒ [1000 Genomes global minor allele frequency](#)
- ☐ [1000 Genomes continental allele frequencies](#)
- ☐ [ESP allele frequencies](#)
- ☐ [gnomAD \(exomes\) allele frequencies](#)

Choose to retrieve allele frequencies for known variants from a range of projects

PubMed IDs for citations of co-located variants: ☒
Retrieve PubMed identifiers for co-located variants

Include flagged variants: ☐

Variant Effect Predictor

Additional annotations Additional transcript, protein and regulatory annotations

Transcript annotation

Transcript biotype: ☒

Exon and intron numbers: ☐

Transcript support level: ☒ Options for adding transcript attribute data to the output

APPRIS: ☒

MANE: ☒

Identify canonical transcripts: ☐

Upstream/Downstream distance (bp):

miRNA structure: ☐

Protein annotation

Protein domains: ☐ Option for adding identifiers of overlapping/ affected protein

Regulatory data

Get regulatory region consequences: Option to retrieve consequence predictions for regulatory features

Phenotype data and citations

Phenotypes: ☐

DisGeNET: ☐ Options for retrieving overlapping phenotype annotation and associated literature citations

Mastermind: ☐

Variant Effect Predictor

Predictions Variant predictions, e.g. SIFT, PolyPhen

Pathogenicity predictions

SIFT: Prediction and score

PolyPhen: Prediction and score

dbNSFP: ☒ Disabled ☐ Enabled Add pathogenicity predictions for variants to output

CADD: ☐

LoFtool: ☐

Splicing predictions

dbSNV: ☐

MaxEntScan: ☐ Add output from splicing prediction algorithms to output

SpliceAI: ☒ Disabled ☐ Enabled

Conservation

BLOSUM62: ☐

Ancestral allele: ☐ Add conservation scores and calculated ancestral alleles to output

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!

Advanced options
Additional enhancements

Advanced options

Buffer size:

5000

NB: When the Regulatory data option is selected then due to the large amount of regulatory data available, the maximum buffer size is automatically reduced from the default value of 5000 to 500. This reduces the memory requirement but might increase the run time. If you find that your jobs are still failing due to memory limitations then you can select a value lower than 500.

Right align variants prior to consequence calculation:

No

Options for optimising running speed and alignment behaviour

Click 'Run' to submit the query

Run

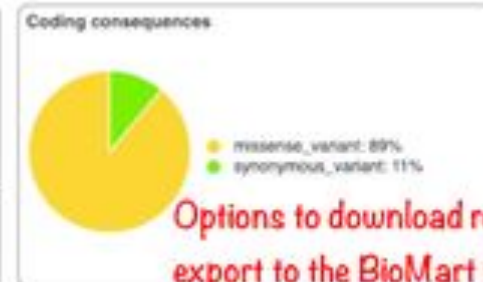
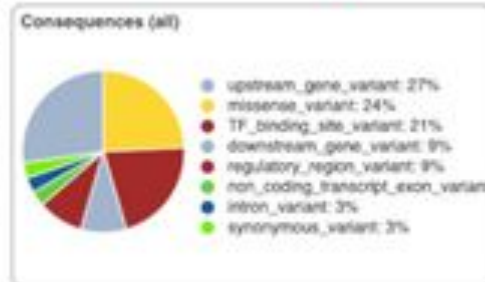
Variant Effect Predictor

Variant Effect Predictor results

Job details

Summary statistics

Category	Count
Variants processed	4
Variants filtered out	0
Novel / existing variants	1 (25.0) / 3 (75.0)
Overlapped genes	2
Overlapped transcripts	7
Overlapped regulatory features	1



Results preview

Navigation (per variant)
Page: 1 of 1 | Show: 1 All variants

Filters
Uploaded variant is defined Add

Download
All: VCF VEP TXT
BioMart: Variants & Genes

New job

Variant Effect Predictor

Link to genomic location in the Ensembl genome browser

Link to gene tab in Ensembl

Consequence prediction

Link to variant tab in Ensembl

Location	Alt	Consequence	Symbol	Gene	Feature	Biotype	cDNA position	CDS position	Amino acids	Existing variant	WAVE SELECT	gPT	PolyPhen	AP
chr10:111,111,111-111,111,112	G	missense_variant	COG4	ENSG00000111111	ENST00000111111.2	protein_coding	104	104	GAC	GGTTTGGG	NAI_010000-0	G	0.875	0.0001
chr10:111,111,111-111,111,112	G	missense_variant	COG4	ENSG00000111111	ENST00000111111.2	protein_coding	104	104	GAC	GGTTTGGG	-	G	0.875	0.0001
chr10:111,111,111-111,111,112	G	Ensembl_gene_variant	TRU80	ENSG00000111111	ENST00000111111.2	protein_coding	-	-	-	GGTTTGGG	NAI_010000-0	-	-	0.0001
chr10:111,111,111-111,111,112	G	Ensembl_gene_variant	TRU80	ENSG00000111111	ENST00000111111.2	processed_transcript	-	-	-	GGTTTGGG	-	-	-	0.0001
chr10:111,111,111-111,111,112	G	missense_variant	COG4	ENSG00000111111	ENST00000111111.2	protein_coding	401	104	GAC	GGTTTGGG	-	G	0.8	0.0001
chr10:111,111,111-111,111,112	G	missense_variant	COG4	ENSG00000111111	ENST00000111111.2	protein_coding	401	104	GAC	GGTTTGGG	-	G	0.8	0.0001
chr10:111,111,111-111,111,112	G	Ensembl_gene_variant	-	-	ENST00000111111.2	promoter	-	-	-	GGTTTGGG	-	-	-	0.0001

Variant Effect Predictor

Ensembl stable ID and symbol of overlapping genes

Location	Allele	Consequence	Symbol	Gene	Feature	Existing variant	SIFT	PolyPhen	AF	AFR AF	AMR AF	EAS AF	EUR AF	SAS AF
11:5227002-5227002	A	missense_variant	HBB	ENSG00000244734	ENST00000336295.4	rs334, CD830010, CM097155, CM000038	0.01	0.007	0.0274	0.0998	0.0072	0	0	0
11:5227002-5227002	C	missense_variant	HBB	ENSG00000244734	ENST00000336295.4	rs334, CD830010, CM097155, CM000038	0.05	0.008	-	-	-	-	-	-
11:5227002-5227002	G	missense_variant	HBB	ENSG00000244734	ENST00000336295.4	rs334, CD830010, CM097155, CM000038	0.29	0.006	-	-	-	-	-	-
11:5227002-5227002	A	missense_variant	HBB	ENSG00000244734	ENST00000380315.2	rs334, CD830010, CM097155, CM000038	0.01	0.007	0.0274	0.0998	0.0072	0	0	0
11:5227002-5227002	C	missense_variant	HBB	ENSG00000244734	ENST00000380315.2	rs334, CD830010, CM097155, CM000038	0.07	0.008	-	-	-	-	-	-
11:5227002-5227002	G	missense_variant	HBB	ENSG00000244734	ENST00000380315.2	rs334, CD830010, CM097155, CM000038	0.34	0.006	-	-	-	-	-	-

Allele frequencies reported for continental populations from the 1000 Genomes project

IDs of existing co-located variants

Variant Effect Predictor

Clinical significance

Clinical significance	Phenotype or disease	Pubmed	Associated phenotypes
protective, pathogenic, other, likely_benign, conflicting_interpretations_of_pathogenicity	1, 1, 1, 1	119 PubMed IDs	39 Phenotype associations
-	1, 1, 1, 1	119 PubMed IDs	39 Phenotype associations
other	1, 1, 1, 1	119 PubMed IDs	39 Phenotype associations
protective, pathogenic, other, likely_benign, conflicting_interpretations_of_pathogenicity	1, 1, 1, 1	119 PubMed IDs	39 Phenotype associations
-	1, 1, 1, 1	119 PubMed IDs	39 Phenotype associations
other	1, 1, 1, 1	119 PubMed IDs	39 Phenotype associations
protective, pathogenic, other, likely_benign, conflicting_interpretations_of_pathogenicity	1, 1, 1, 1	119 PubMed IDs	39 Phenotype associations

Associated phenotypes -
click to expand


Priorização

- Processo através do qual é feito um ranking de variantes, de acordo com a sua patogenicidade e/ou relevância para o fenótipo em causa.

Exomiser

Published: 12 November 2015

Next-generation diagnostics and disease-gene discovery with the Exomiser

Damian Smedley, Julius O B Jacobsen, Marten Jäger, Sebastian Kölsch, Enrico Siragusa, Tomasz Zemojtel, Orion J Buske, Nicole L Washinowicz & Peter N Robinson 

Nature Protocols **10**, 2004–2015(2015) | [Cite this article](#)

GCAT
TACG
GCAT

genes



Article

An Improved Phenotype-Driven Tool for Rare Mendelian Variant Prioritization: Benchmarking Exomiser on Real Patient Whole-Exome Data

- **Priorização de variantes:**

- Fenótipo/Termos HPO – Específicos; Representativos do diagnóstico clínico

- **Variantes + Termos HPO:**

- **74% dos casos** com variante causal na **1ª posição** do ranking
- **94% no top 5**
- Posição mais baixa foi a 42ª

- **Só Variantes (sem termos HPO):**

- **3% dos casos** com variante causal na **1ª posição**
- **27% no top 5**

The Exomiser: A Tool to Annotate and Filter Variants

The Exomiser is a Java program that functionally annotates variants from whole-exome sequencing data in VCF 4 format. The functional annotation is performed with [Jannovar](#) and uses [UCSC](#) KnownGene transcript definitions and hg19 genomic coordinates.

Variants are prioritized according to user-defined criteria on variant frequency, pathogenicity, quality, inheritance pattern, and model organism phenotype data. Predicted pathogenicity data was extracted from the [dbNSFP](#) resource. Cross-species phenotype comparisons come from our [PhenoDigm](#) tool powered by the [OWLSim](#) algorithm.

The Exomiser was developed by the Computational Biology and Bioinformatics group at the Institute for Medical Genetics and Human Genetics of the Charité - Universitätsmedizin Berlin, the Mouse Informatics Group at the Sanger Institute and other members of the Monarch initiative.

[→ Go to data submission and analysis](#)

CAUTION! This instance of the Exomiser is not running in a clinical-grade data centre. Under no circumstances should this be used for analysing real patient data as there are no safeguards in place for protecting patient privacy.

Analysis Options

Upload Sample Files

VCF file:

Nenhum ficheiro selecionado

Required. Upload exome sequencing results in VCF format. We can only accept files containing up to 100000 variants. [Example file](#) with causative FGFR2 variant for the autosomal dominant Pfeiffer syndrome added to exome of a [healthy individual](#)

PED file:

Nenhum ficheiro selecionado

Only required for multi-sample VCF files

Proband:

Sample identifier for the proband as in the VCF and PED files. Only required for multi-sample VCF files

Enter Sample Phenotypes

Phenotypes associated with Mendelian disease:

Clinical phenotypes:

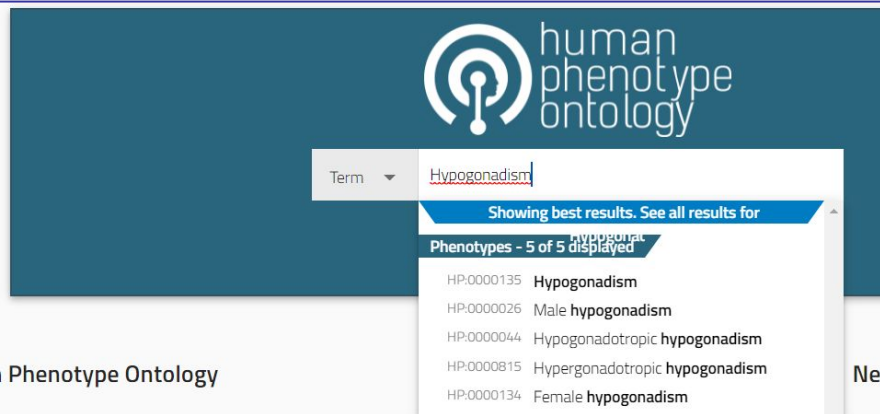
Input terms from the [HPO](#). These will override any phenotypes derived from the specified disease!

HPO

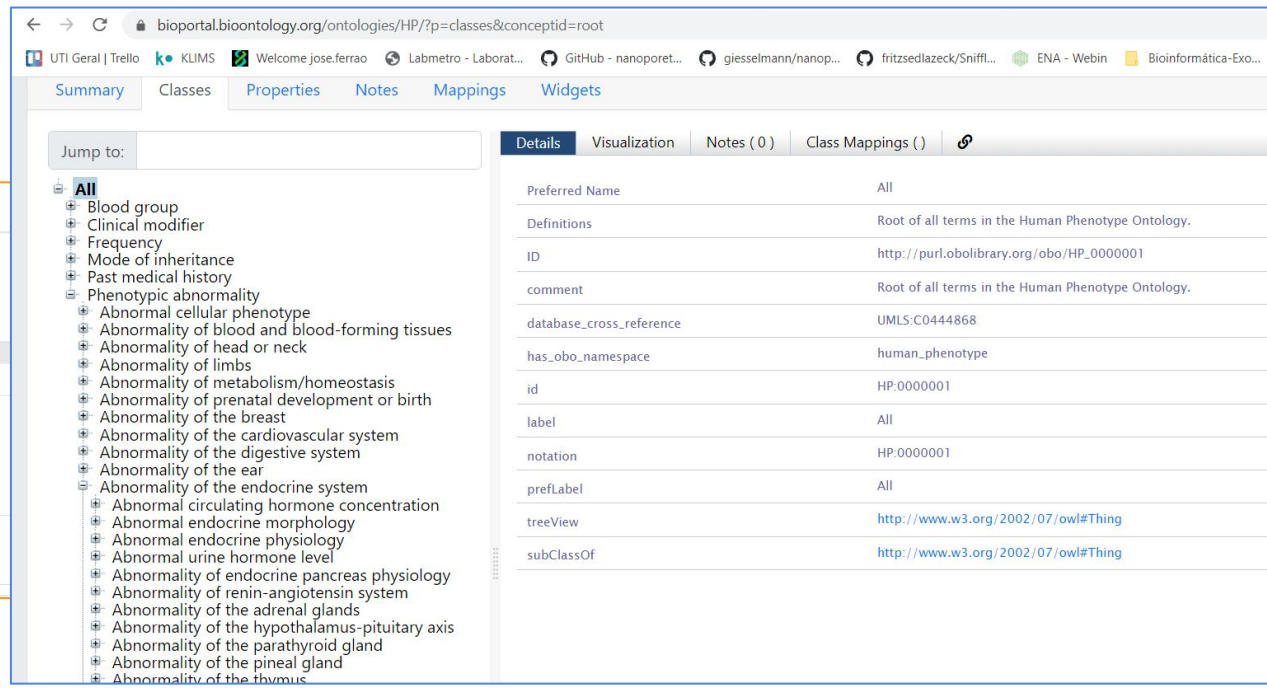
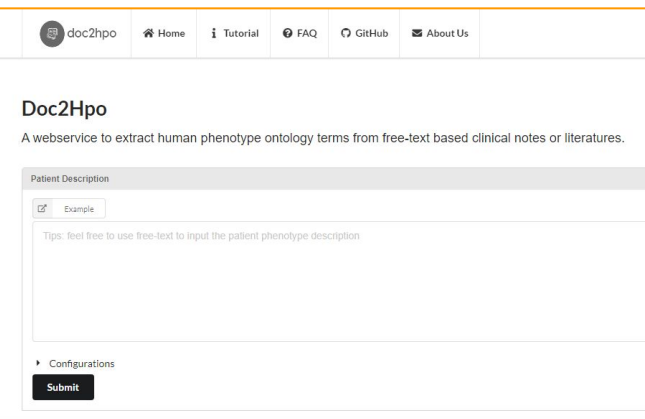
<https://hpo.jax.org/app/>

<https://doc2hpo.wglab.org/>

<https://bioportal.bioontology.org/ontologies/HP/?p=classes&conceptid=root>

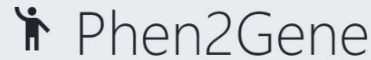


The Human Phenotype Ontology



Phen2Gene - HPO's to gene list

[Phen2Gene](#) [Home](#) [Phen2Gene GitHub Code](#) [HPO2Gene KnowledgeBase](#) [API](#) [Contact](#)



Phen2Gene is a real-time phenotype-based gene prioritization tool from HPO IDs or clinical notes on patients.

[HPO IDs](#) [Patient notes](#)

Human Phenotype Ontology (HPO) Term IDs

HP:0000707;HP:0007598;HP:0001156;HP:0012446;HP:0004209;HP:0000405;HP:0001627;HP:0002750;HP:0003235;HP:0000239;HP:0001572;HP:0002123;HP:0011220;HP:0004482;HP:0002069;HP:0012471;HP:0001566;HP:0001250

Please enter your focused HPO term IDs, in the format of HP:digits, e.g. HP:0000707, separated by semicolons, newlines, commas, or spaces.

Weight Model

sk - weights HPO terms by skewness (default)

 Reset

 Submit