

BIOLOGICZNE BAZY DANYCH POPULARNE PRZEGLĄDARKI GENOMOWE

Podstawy bioinformatyki
Wykład 2

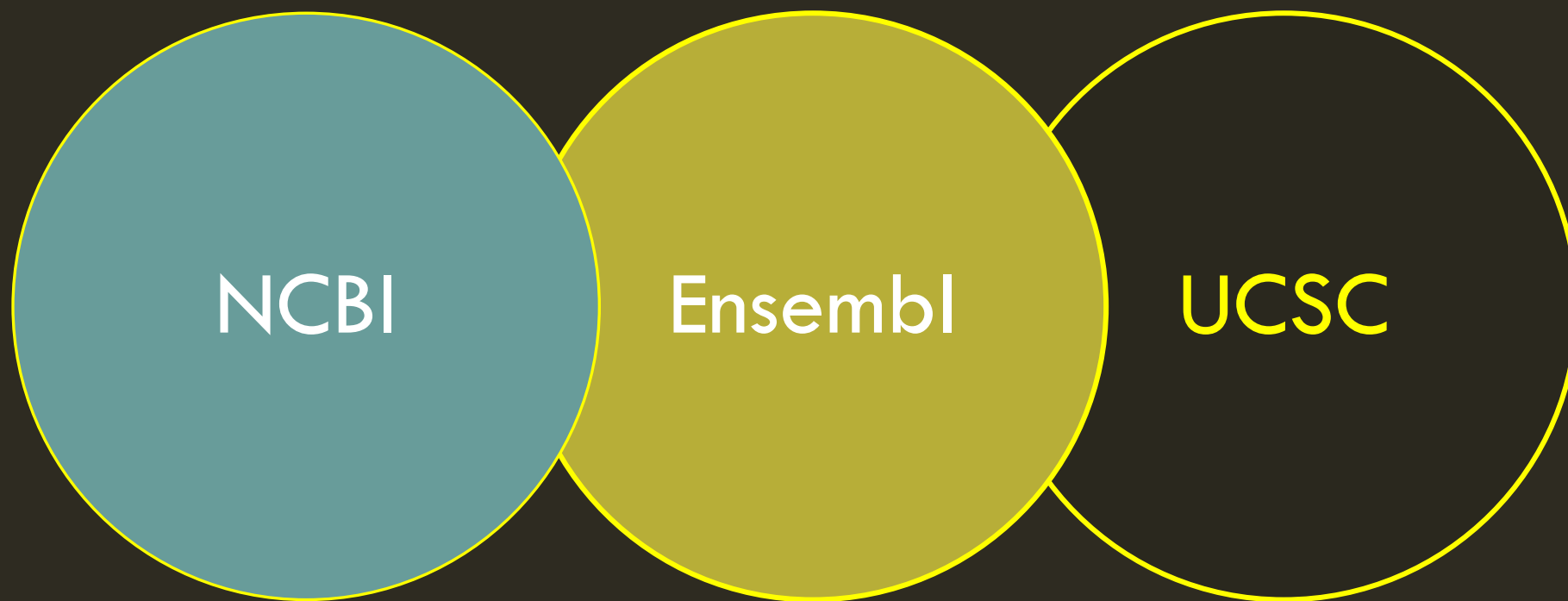
PRZEGLĄDARKA GENOMOWA

- Czym jest genom?
- Czym jest genom referencyjny?



<https://pl.freepik.com/>

WYBRANE BIOLOGICZNE BAZY DANYCH POPULARNE PRZEGLĄDARKI GENOMOWE

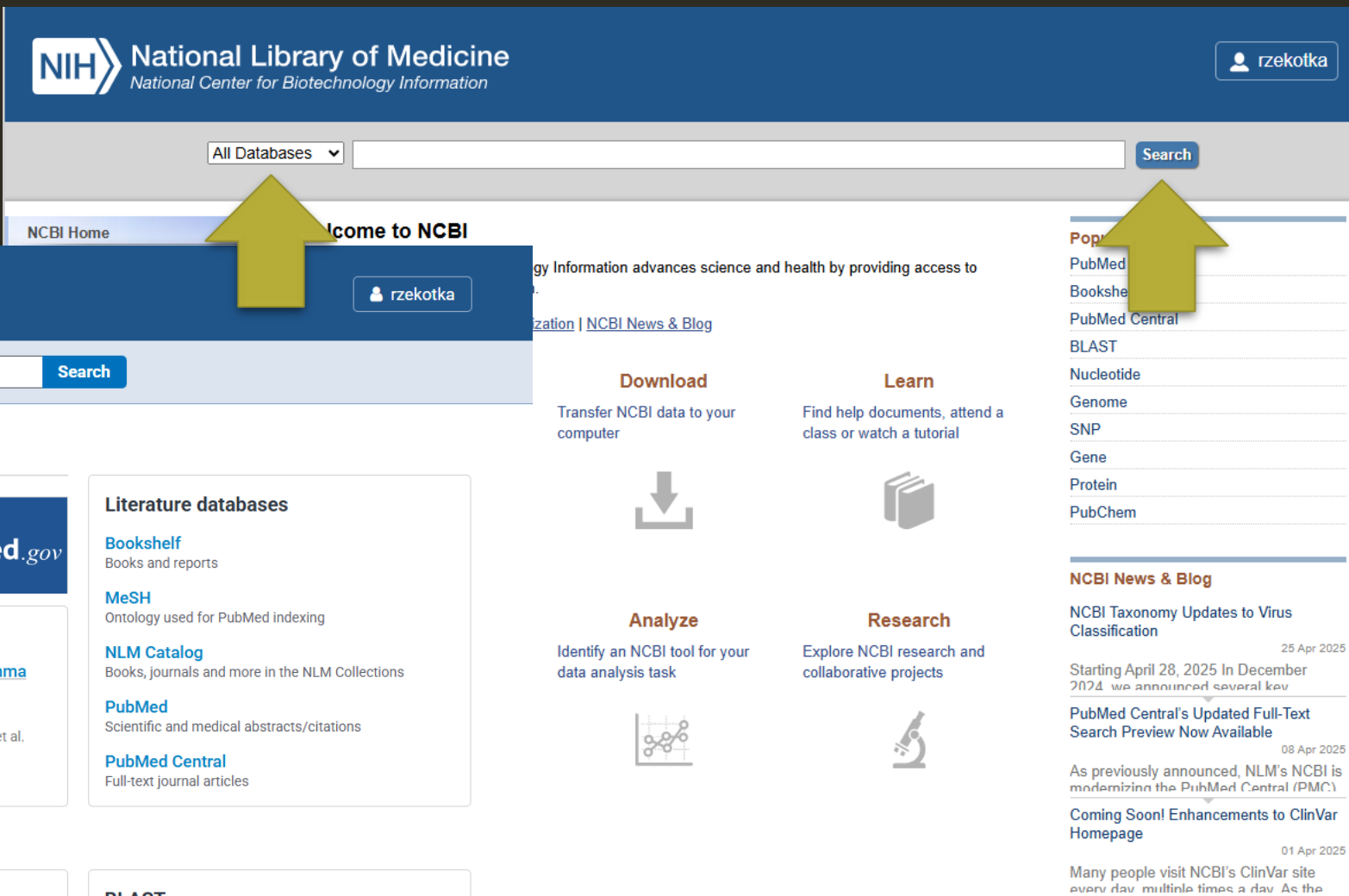


NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION



- Utworzone w 1988
- Dziedziny nauki: biologia molekularna, biochemia, genetyka
- Zadania
 - Przechowywanie → bazy danych
 - Analiza danych → oprogramowanie
 - Udostępnianie baz danych i oprogramowania do badań naukowych
 - Koordynacja gromadzenia danych na poziomie międzynarodowym

NCBI



Literature

PubMed

PubMed® comprises more than 38 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full text content from PubMed Central and publisher web sites.



Featured Bookshelf titles



[Drugs and Lactation Database \(LactMed®\)](#)



[Characteristics of Existing Asthma Self-Management Education Packages](#)

Leas BF, Tipton K, Bryant-Stephens T, et al.

[Browse the Bookshelf >](#)

Literature databases

Bookshelf

Books and reports

MeSH

Ontology used for PubMed indexing

NLM Catalog

Books, journals and more in the NLM Collections

PubMed

Scientific and medical abstracts/citations

PubMed Central

Full-text journal articles

Download

Transfer NCBI data to your computer



Learn

Find help documents, attend a class or watch a tutorial



Analyze

Identify an NCBI tool for your data analysis task



Research

Explore NCBI research and collaborative projects



Data

Genes

Gene sequences and annotations used as references for the study of orthologs structure, expression, and evolution

Gene

Collected information about gene loci

GEO DataSets

Functional genomics studies

GEO Profiles

Gene expression and molecular abundance profiles

Proteins

Protein sequences, 3-D structures, and tools for the study of functional protein domains and active sites

Conserved Domains

Conserved protein domains

Identical Protein Groups

Protein sequences grouped by identity

Protein

Protein sequences

BLAST

A tool to find regions of similarity between biological sequences

blastn

Search nucleotide sequence databases

blastp

Search protein sequence databases

blastx

Search protein databases using a translated nucleotide query

NCBI: GENOME

Genome

Search by taxonomic name or ID, Assembly name, BioProject, BioSample, WGS or Nucleotide accession

Search

Try examples: [Homo sapiens](#) [GCF_000001405.40](#) [PRJNA489243](#) [SAMN15960293](#) [WFKY01](#) [GRCh38.p14](#) [NC_000913.3](#)

Genomic data available from NCBI Datasets

Click below to learn more about the genomic data available from NCBI Datasets.



Eukaryota



Archaea



Bacteria



Viruses

All Genomes

2.98M

Total

40.31K

Reference

2.40M

Annotated

NCBI: GENOME

Genome

Download a genome data package including genome, transcript and protein sequence, annotation and a data report

Selected taxa: Enter common or scientific names

Filters

Download Select columns 67 Genomes Rows per page 20 1-20 of 67

☐	Assembly	GenBank	RefSeq	Scientific name	Modifier	Annotation	Action
☒	GRCm39	GCA_000001635.9	GCF_000001635.27	Mus musculus (house mouse)	C57BL/6J (strain)	NCBI RefSeq	
☐	C57BL_6J_T2T_v1	GCA_964188535.1		Mus musculus (house mouse)	C57BL_6J (strain)		
☐	DBA_2J_v3	GCA_921998315.2		Mus musculus (house mouse)			
☐	NEI_Mmus_1.0	GCA_030265425.1		Mus musculus (house mouse)	B10.RIII (strain)		
☐	mMusMuc1.1	GCA_949316315.1		Mus musculus (house mouse)			
☐	NOD_SCID	GCA_031763685.1		Mus musculus (house mouse)	NOD/SCID (strain)		
☐	BALB_cJ_v3	GCA_921997145.2		Mus musculus (house mouse)			
☐	NZO_HILtJ_v3	GCA_947593165.1		Mus musculus (house mouse)			

NCBI: GENOME

[NCBI Datasets](#) [Taxonomy](#) [Genome](#) [Gene](#) [Command-line tools](#) [Documentation](#)

Genome assembly GRCm39 reference

Download

 datasets

URL

FTP

Actions

NCBI RefSeq assembly	GCF_000001635.27	⋮
Submitted GenBank assembly	GCA_000001635.9	⋮
Taxon	Mus musculus (house mouse)	
Strain	C57BL/6J	
Synonym	mm39	
Assembly type	haploid	
Submitter	Genome Reference Consortium	
Date	Jun 24, 2020	



View annotated genes



See in Genome Data Viewer

Assembly statistics

	RefSeq	GenBank
Genome size	2.7 Gb	2.7 Gb
Total ungapped length	2.7 Gb	2.7 Gb
Gaps between scaffolds	143	143
Number of chromosomes	21	21
Number of organelles	1	1
Number of scaffolds	101	101
Scaffold N50	106.1 Mb	106.1 Mb
Scaffold L50	11	11
Number of contigs	305	305
Contig N50	59.5 Mb	59.5 Mb
Contig L50	15	15
GC percent	42	42
Assembly level	Chromosome	Chromosome
View sequences	view RefSeq sequences	view GenBank sequences

Addi

Browse a

BioP

PRJNA2

Genome

musculu

Genome

Publi

Showing 5

PLoS Biol

Moderni

DM Church

PLoS Biol

Lineage-specific biology revealed by a finished genome assembly of the mouse
DM Church, et al.

NCBI: GENOME

Chromosomes



Download

Chromosome	GenBank	RefSeq	Size (bp)	GC
1	CM000994.3	NC_000067.7	195 154 279	
2	CM000995.3	NC_000068.8	181 755 017	
3	CM000996.3	NC_000069.7	159 745 316	
4	CM000997.3	NC_000070.7	156 860 686	
5	CM000998.3	NC_000071.7	151 758 149	

Annotation details

[See full annotation report](#)

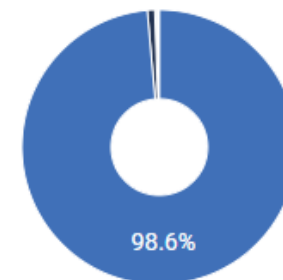
	RefSeq
Provider	NCBI RefSeq
Name	GCF_000001635.27-RS_2024_02
Date	Feb 1, 2024
Genes	50 763
Protein-coding	22 198
Software version	10.2

[View RefSeq annotation](#)

Quality analysis

BUSCO analysis (4.1.4)

- Single_copy 98.6%
- Duplicated 0.9%
- Fragmented 0.2%
- Missing 0.4%



NCBI: GENOME

Genome

Download a genome data package including genome, transcript and protein sequence, annotation and a data report

Selected taxa

Homo sapiens

Enter one or more taxonomic names

Filters

Download

Select columns

1596 Genomes

Rows per page

20

1-20 of 1596

<

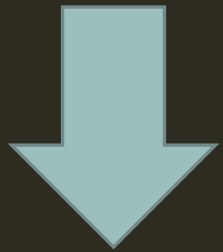
>

☐ Assembly	GenBank	RefSeq	Scientific name	Modifier	Annotation	Action
☐ GRCh38.p14	GCA_000001405.29	GCF_000001405.40	Homo sapiens (human)		NCBI RefSeq	
☐ T2T-CHM13v2.0	GCA_009914755.4	GCF_009914755.1	Homo sapiens (human)		NCBI RefSeq	
☐ hg002v1.1.mat	GCA_018852615.3		Homo sapiens (human)	NA24385 (isolate)		
☐ hg002v1.1.pat	GCA_018852605.3		Homo sapiens (human)	NA24385 (isolate)		
☐ Ash1_v2.2	GCA_011064465.2		Homo sapiens (human)	NA24385 (isolate)	Submitter	
☐ hg01243.v3.0	GCA_018873775.2		Homo sapiens (human)	HG01243 (isolate)	Submitter	

FORMAT FASTA

Prosty i popularny → czytelny dla wielu programów do analizy bioinformatycznej

Zapis sekwencji kwasów nukleinowych oraz białek



Identyfikator
sekwencji

Sekwencja

>gi|52693750|dbj|AB175071.1

ATGACCAACTTTCGAAAAACCCATCCATTAATAAAAATTCTTAACAACATTCATCGAC
CATCAAACATTTATCATGATGAAATTCGGGTCCCTTCTAGGATTGTGCCTAGTAATC

NCBI: REFSEQ

NCBI Reference Sequence Database:

kompleksowy, zintegrowany, niepowtarzalny dobrze opisany zestaw sekwencji referencyjnych

- genomy, transkrypty, białka
- baza niepowtarzalna (non-redundant)
- informacje sprawdzone (często weryfikowane manualnie)

NCBI: REFSEQ

RefSeq

RefSeq

mus musculus

Search



RefSeq: NCBI Reference Sequence Database

A comprehensive, integrated, non-redundant, well-annotated set of reference sequences including genomic, transcript, and protein.

Using RefSeq

- [About RefSeq](#)
- [Human Reference Genome](#)
- [Prokaryotic RefSeq Genomes](#)
- [FAQ](#)
- [NCBI Handbook](#)
- [Factsheet](#)

RefSeq Access

- [Human Genome Resources and Download](#)
- [RefSeq FTP](#)
- [RefSeq.genomes FTP](#)
- [New RefSeq.genomic \(last 30 days\)](#)
- [New RefSeq transcripts \(last 30 days\)](#)
- [New RefSeq proteins \(last 30 days\)](#)
- [Searching for RefSeq records \(Queries\)](#)

RefSeq projects

- [Consensus CDS \(CCDS\)](#)
- [RefSeq Functional Elements](#)
- [RefSeqGene](#)
- [Targeted Loci](#)
- [Virus Variation](#)

NCBI: REFSEQ

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide mus musculus AND srcdb_refseq[PROP] Search

Create alert Advanced Help

Species
Animals (116,577)
Plants (13)
Fungi (175)
Protists (256)
Bacteria (12,056)
Viruses (20)
Customize ...

Molecule types
genomic DNA/RNA (20,205)
mRNA (77,471)
rRNA (61)
Customize ...

Source databases
RefSeq (129,096)
Customize ...

Genetic

Items: 1 to 20 of 129096

1. [Mus musculus esterase D/formylglutathione hydrolase \(Esd\), transcript variant 2, mRNA](#)
1,988 bp linear mRNA
Accession: NM_016903.5 GI: 550544231
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

2. [Mus musculus esterase D/formylglutathione hydrolase \(Esd\), transcript variant 1, mRNA](#)
2,121 bp linear mRNA
Accession: NM_001285423.1 GI: 550544229
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

Send to: Filters: [Manage Filters](#)

Results by taxon

Top Organisms [Tree](#)

- Mus musculus (116395)
- Escherichia coli (2339)
- Lactobacillus reuteri (1328)
- Yersinia pestis PY-36 (592)
- Yersinia pestis PY-48 (585)
- All other taxa (7857)

More...

Find related data

Database: Select

Find items

NCBI: REFSEQ

 NCBI

Resources  How To 

Sign in to NCBI

Nucleotide

Nucleotide 

mus musculus AND srcdb_refseq[PROP] 

Search

Create alert Advanced Help

Species

Animals (116,577)

Plants (13)

Fungi (175)

Protists (256)

Bacteria (12,056)

Viruses (20)

Customize ...

Molecule types

genomic DNA/RNA (20,205)

mRNA (77,471)

rRNA (61)

Customize ...

Source databases

RefSeq (129,096)

Customize ...

Genetic

Summary  20 per page  Sort by Default order 

Send to: 

Filters: [Manage Filters](#)

Items: 1 to 20 of 129096

<< First < Prev Page 1 of 6455 Next > Last >>

☐ [Mus musculus esterase D/formylglutathione hydrolase \(Esd\), transcript variant 2, mRNA](#)

1. 1,988 bp linear mRNA

Accession: NM_016903.5 GI: 550544231

[Protein](#) [PubMed](#) [Taxonomy](#)

[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Mus musculus esterase D/formylglutathione hydrolase \(Esd\), transcript variant 1, mRNA](#)

2. 2,121 bp linear mRNA

Accession: NM_001285423.1 GI: 550544229

[Protein](#) [PubMed](#) [Taxonomy](#)

[GenBank](#) [FASTA](#) [Graphics](#)

Results by taxon 

Top Organisms [\[Tree\]](#)

Mus musculus (116395)

Escherichia coli (2339)

Lactobacillus reuteri (1328)

Yersinia pestis PY-36 (592)

Yersinia pestis PY-48 (585)

All other taxa (7857)

More...

Find related data 

Database: 

Find items

NCBI: REFSEQ

Species

Animals (116,577)

Plants (13)

Fungi (175)

Protists (256)

Bacteria (12,056)

Viruses (20)

Customize ...

Molecule types

genomic DNA/RNA (20,205)

mRNA (77,471)

rRNA (61)

Customize ...

Source database

RefSeq (129,056)

Customize ...

Genetic compartment

Mitochondrion

Plasmid (55)

Sequence length

Custom range...

Release date

Custom range...

Summary ▾

20 per page ▾

Sort by Default order ▾

Send to: ▾

Filters: [Manage Filters](#)

Items: 1 to 20 of 129096

<< First < Prev Page 1 of 6455 Next > Last >>

☐ [Mus musculus](#) [esterase D/formylglutathione hydrolase \(Esd\), transcript variant 2, mRNA](#)

1. 1,988 bp linear mRNA

Accession: NM_016903.5 GI: 550544231

[Protein](#) [PubMed](#) [Taxonomy](#)

[GenBank](#) [FASTA](#) [Graphics](#)

☒ genomic DNA/RNA

☒ mRNA

☐ cRNA

☐ ncRNA

☒ rRNA

☐ tRNA

☐ transcribed RNA

Show

[formylglutathione hydrolase \(Esd\), transcript variant 1, mRNA](#)

550544229

[ate isomerase 1 \(Tpi1\), mRNA](#)

958348

Results by taxon

Top Organisms [\[Tree\]](#)

[Mus musculus](#) (116395)

[Escherichia coli](#) (2339)

[Lactobacillus reuteri](#) (1328)

[Yersinia pestis](#) PY-36 (592)

[Yersinia pestis](#) PY-48 (585)

[All other taxa](#) (7857)

[More...](#)

Find related data

Database:

Select ▾

Find items

Search details

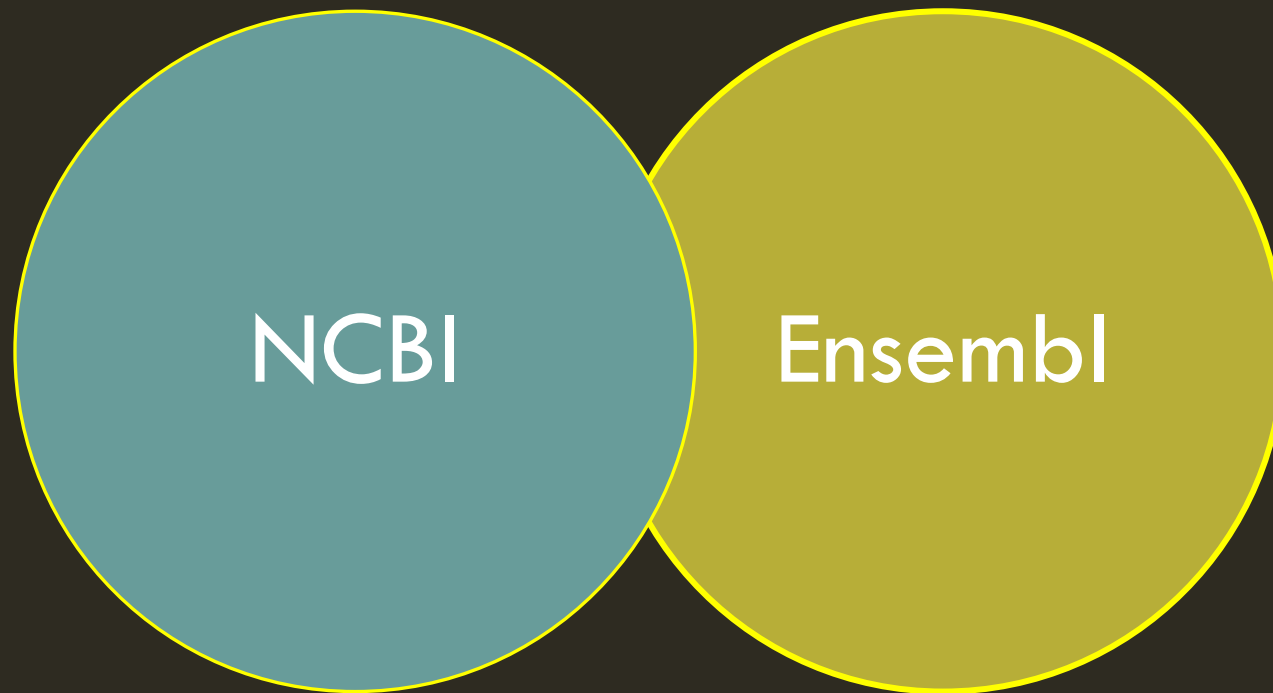
("Mus musculus"[Organism] OR mus musculus[All Fields]) AND srcdb_refseq[PROP]

NCBI

Identyfikatory:

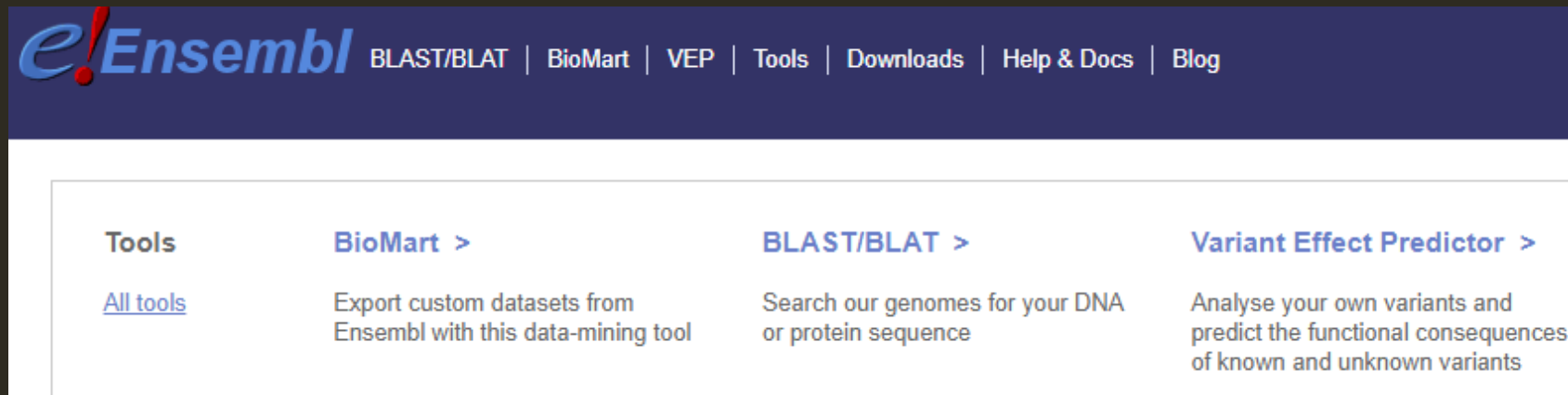
- **N**M_numer / **X**M_numer mRNA
 - **N**P_numer / **X**P_numer białka
 - **N**R_numer / **X**R_numer niekodujące RNA
 - **N**C_numer, **N**G_numer kontigi, sekwencje genomowe
-
- **N** – sekwencje uzyskane z wyników eksperymentów
 - **X** – sekwencje z adnotacji (sekwencja została przewidziana np. przez zmapowanie białka do genomu spokrewnionego organizmu)

GENOMY I ICH ADNOTACJE



ENSEMBL

- „Przeglądarka” genomów **kręgowców**
- Wspiera badania z zakresu genomiki porównawczej, ewolucji i regulacji ekspresji genów
- Dostępne narzędzia: **BioMart**, Variant Effect Predictor (**VEP**)



ENSEMBL

- Ensembl
- EnsemblGenomes
- EnsemblFungi
- EnsemblMetazoa
- EnsemblProtists
- EnsemblBacteria
- EnsemblPlants



[Caenorhabditis elegans](#)
WBcel235



[Drosophila melanogaster](#)
BDGP6.28



[Bombyx mori](#)
ASM15162v1

[Go to Ensembl Metazoa](#)



[Plasmodium falciparum 3D7](#)
ASM276v2



[Dictyostelium discoideum](#)
dicty_2.7



[Phytophthora infestans](#)
ASM14294v1

[Go to Ensembl Protists](#)

EnsemblFungi



[Magnaporthe oryzae](#)
MG8



[Saccharomyces cerevisiae](#)
R64-1-1



[Aspergillus nidulans](#)
ASM1142v1

[Go to Ensembl Fungi](#)

EnsemblBacteria

[Streptococcus pneumoniae](#)
ASM688v1

[Escherichia coli](#)
ASM584v2

[Bacillus subtilis](#)
ASM73511v1

[Go to Ensembl Bacteria](#)

ENSEMBL

- Ensembl
- EnsemblGenomes
- EnsemblFungi
- EnsemblMetazoa
- EnsemblProtists
- EnsemblBacteria
- EnsemblPlants



ENSEMBL

Identyfikatory (*Homo sapiens*):

- ENS**G** → Gene
- EN**S****T** → Transcript
- EN**S****P** → Protein
- EN**S****E** → Exon

Identyfikatory (*inne gatunki*):

- ENS**FCAT**00000032635 *Felis catus*
- ENS**RNOG**00000050313 *Rattus norvegicus*
- ENS**CAFG**00000022708 *Canis lupus familiaris*
- ENS**BTAT**00000064726 *Bos taurus*

Zagadka:

ENST00000471181.7

? ?

ENSRNOT00000075759.1

? ?

ENSSSCG00000018060

? ?

ENSEMBL

 [BLAST/BLAT](#) | [BioMart](#) | [VEP](#) | [Tools](#) | [Downloads](#) | [Help & Docs](#) | [Blog](#)

Tools
[All tools](#)

BioMart >
Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT >
Search our genomes for your DNA or protein sequence

Variant Effect Predictor >
Analyse your own variants and predict the functional consequences of known and unknown variants

Search

All species ▼

for

Go

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

All genomes

-- Select a species -- ▼

- [View full list of all Ensembl species](#)
- [Edit your favourites](#)

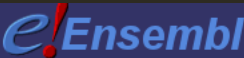
Favourite genomes

**Human**
GRCh38.p12
[Still using GRCh37?](#)

**Mouse**
GRCm38.p6

**Zebrafish**
GRCz11

ENSEMBL

 [BLAST/BLAT](#) | [BioMart](#) | [VEP](#) | [Tools](#) | [Downloads](#) | [Help & Docs](#) | [Blog](#)

Tools
[All tools](#)

BioMart >
Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT >
Search our genomes for your DNA or protein sequence

Variant Effect Predictor >
Analyse your own variants and predict the functional consequences of known and unknown variants



Search

Cow

for

19:12,571,690-12,571,788

Go

e.g. [BRCA2](#) or [rat 5:62797383-63627669](#) or [rs699](#) or [coronary heart disease](#)

All genomes

-- Select a species --

- [View full list of all Ensembl species](#)
- [Edit your favourites](#)

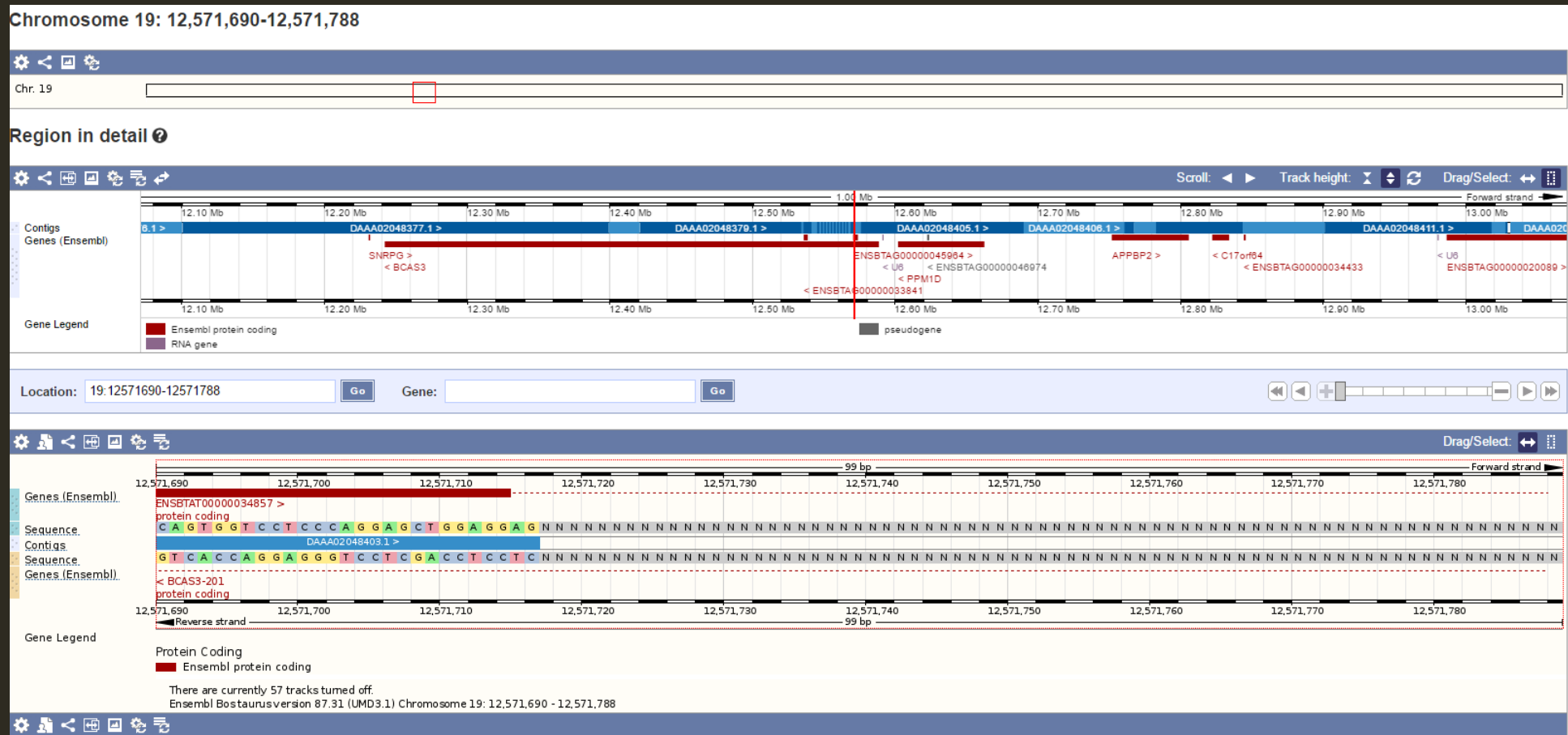
Favourite genomes

 **Human**
GRCh38.p12
[Still using GRCh37?](#)

 **Mouse**
GRCm38.p6

 **Zebrafish**
GRCz11

ENSEMBL



ENSEMBL

 [BLAST/BLAT](#) | [BioMart](#) | [VEP](#) | [Tools](#) | [Downloads](#) | [Help & Docs](#) | [Blog](#)

Tools
[All tools](#)

BioMart >
Export custom datasets from Ensembl with this data mining tool

BLAST/BLAT >
Search our genomes for your DNA or protein sequence

Variant Effect Predictor >
Analyse your own variants and predict the functional consequences of known and unknown variants

Search

Human ▾ for
BRCA1

e.g. [BRCA2](#) or [rat 5:62797383-63627669](#) or [rs699](#) or [coronary heart disease](#)

All genomes

-- Select a species -- ▾

- [View full list of all Ensembl species](#)
- [Edit your favourites](#)

Favourite genomes

 **Human**
GRCh38.p12
[Still using GRCh37?](#)

 **Mouse**
GRCm38.p6

 **Zebrafish**
GRCz11

ENSEMBL

Na którym chromosomie leży gen BRCA1?
Ile ma form splicingowych?
Który transkrypt jest najdłuższy?

Gene: BRCA1 ENSG00000012048

Description

BRCA1 DNA repair associated [Source:HGNC Symbol;Acc:[HGNC:1100](#)]

Gene Synonyms

BRCC1, FANCS, PPP1R53, RNF53

Location

[Chromosome 17: 43,044,295-43,170,245](#) reverse strand.
GRCh38:CM000679.2

About this gene

This gene has 40 transcripts ([splice variants](#)), [196 orthologues](#) and is associated with [111 phenotypes](#).

Transcripts

Hide transcript table

Show/hide columns (1 hidden)

Transcript ID	Name	bp	Protein	Biotype	CCDS	UniProt Match	RefSeq Match
ENST00000357654.9	BRCA1-203	7088	1863aa	Protein coding	CCDS11453	P38398-1	NM_007294.4
ENST00000618469.2	BRCA1-226	7613	1863aa	Protein coding	CCDS11453	G1UI37	-
ENST00000652672.2	BRCA1-232	7320	1816aa	Protein coding	CCDS11459	A0A494C182	-
ENST00000471181.7	BRCA1-209	7270	1884aa	Protein coding	CCDS11456	P38398-7	-
ENST00000494123.6	BRCA1-221	7168	1863aa	Protein coding	CCDS11453	Q3B891	-
ENST00000470026.6	BRCA1-208	7147	1863aa	Protein coding	CCDS11453	E7EWN5	-

ENSEMBL (DOWNLOADS → FTP DOWNLOAD)

Show 10 entries		Show/hide columns												Filter		
★	Species	DNA (FASTA)	cDNA (FASTA)	CDS (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Annotated sequence (EMBL)	Annotated sequence (GenBank)	Gene sets	Whole databases	Variation (GVF)	Variation (VCF)	Variation (VEP)	Regulation (GFF)	Data files	BAM/BigWig
Y	Human <i>Homo sapiens</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig
Y	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig
Y	Zebrafish <i>Danio rerio</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	-	-	BAM/BigWig
	Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Alpaca <i>Vicugna pacos</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Amazon molly <i>Poecilia formosa</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	BAM/BigWig
	Angola colobus <i>Colobus angolensis palliatus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Anole lizard <i>Anolis carolinensis</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	BAM/BigWig

ENSEMBL

Show 10 entries		Show/hide columns												mus		
★	Species	DNA (FASTA)	cDNA (FASTA)	CDS (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Annotated sequence (EMBL)	Annotated sequence (GenBank)	Gene sets	Whole databases	Variation (GVF)	Variation (VCF)	Variation (VEP)	Regulation (GFF)	Data files	BAM/BigWig
	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig
	Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Ferret <i>Mustela putorius furo</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	BAM/BigWig
	Mouse 129S1/SvImJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Mouse A/J <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Mouse AKR/J <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Mouse BALB/cJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Mouse C3H/HeJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-

Index of /pub/release-110/fasta/mus_musculus/dna

	<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
	Parent Directory		-	
	CHECKSUMS	2023-05-02 09:13	4.2K	
	Mus_musculus.GRCm39.dna.chromosome.1.fa.gz	2023-04-21 17:21	56M	
	Mus_musculus.GRCm39.dna.chromosome.2.fa.gz	2023-04-21 17:21	52M	
	156.17.191.216 - PuTTY			
	M			
	>1 dna:chromosome chromosome:GRCm38:1:1:195471971:1 REF			
	N			
	N			
	N			
	Mus_musculus.GRCm39.dna.chromosome.7.fa.gz	2023-04-21 17:21	41M	
	Mus_musculus.GRCm39.dna.chromosome.8.fa.gz	2023-04-21 17:21	36M	
	Mus_musculus.GRCm39.dna.chromosome.9.fa.gz	2023-04-21 17:21	35M	
	Mus_musculus.GRCm39.dna.chromosome.10.fa.gz	2023-04-21 17:21	37M	

ENSEMBL

Show	10	entries	Show/hide columns													mus	
★	Species	DNA (FASTA)	cDNA (FASTA)	CDS (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Annotated sequence (EMBL)	Annotated sequence (GenBank)	Gene sets	Whole databases	Variation (GVF)	Variation (VCF)	Variation (VEP)	Regulation (GFF)	Data files	BAM/BigWig	
Y	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig	
	Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	
	Ferret <i>Mustela putorius furo</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	BAM/BigWig	
	Mouse	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	-	VEP	-	-	-	



```

156.17.191.216 - PuTTY
>ENSMUST00000196221.1 cdna chromosome:GRCm38:14:54113468:54113476:1 gene:ENSMUSG00000096749.2 gene_biotype:TR_D_gene trans
cript_biotype:TR_D_gene gene_symbol:Trdd1 description:T cell receptor delta diversity 1 [Source:MGI Symbol;Acc:MGI:4439547
]
ATGGCATAT
>ENSMUST00000179664.1 cdna chromosome:GRCm38:14:54113468:54113478:1 gene:ENSMUSG00000096749.2 gene_biotype:TR_D_gene trans
cript_biotype:processed_transcript gene_symbol:Trdd1 description:T cell receptor delta diversity 1 [Source:MGI Symbol;Acc:
MGI:4439547]
ATGGCATATCA
>ENSMUST00000177564.1 cdna chromosome:GRCm38:14:54122226:54122241:1 gene:ENSMUSG00000096176.1 gene_biotype:TR_D_gene trans
cript_biotype:TR_D_gene gene_symbol:Trdd2 description:T cell receptor delta diversity 2 [Source:MGI Symbol;Acc:MGI:4439546
]
ATCGGAGGGATACGAG
  
```


ENSEMBL

Show	10	entries	Show/hide columns													mus	
★	Species	DNA (FASTA)	cDNA (FASTA)	CDS (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Annotated sequence (EMBL)	Annotated sequence (GenBank)	Gene sets	Whole databases	Variation (GVF)	Variation (VCF)	Variation (VEP)	Regulation (GFF)	Data files	BAM/BigWig	
Y	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig	
	Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	
	Ferret <i>Mustela putorius furo</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	BAM/BigWig	
	Mouse	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	-	VEP	-	-	-	



```

156.17.191.216 - PuTTY
>ENSMUST00000196221.1 cds chromosome:GRCm38:14:54113468:54113476:1 gene:ENSMUSG00000096749.2 gene_biotype:TR_D_gene transcript_biotype:TR_D_gene gene_symbol:Trdd1 description:T cell receptor delta diversity 1 [Source:MGI Symbol;Acc:MGI:4439547] ATGGCATAT
>ENSMUST00000177564.1 cds chromosome:GRCm38:14:54122226:54122241:1 gene:ENSMUSG00000096176.1 gene_biotype:TR_D_gene transcript_biotype:TR_D_gene gene_symbol:Trdd2 description:T cell receptor delta diversity 2 [Source:MGI Symbol;Acc:MGI:4439546] ATCGGAGGGGATACGAG
>ENSMUST00000178537.1 cds chromosome:GRCm38:6:41533201:41533212:1 gene:ENSMUSG00000095668.1 gene_biotype:TR_D_gene transcript_biotype:TR_D_gene gene_symbol:Trbd1 description:T cell receptor beta, D region 1 [Source:MGI Symbol;Acc:MGI:4439571] GGGACAGGGGGC
>ENSMUST00000178862.1 cds chromosome:GRCm38:6:41542163:41542176:1 gene:ENSMUSG00000094569.1 gene_biotype:TR_D_gene transcript_biotype:TR_D_gene gene_symbol:Trbd2 description:T cell receptor beta, D region 2 [Source:MGI Symbol;Acc:MGI:4439727] GGGACTGGGGGGGC
  
```


ENSEMBL

Show 10 entries		Show/hide columns												mus		
★	Species	DNA (FASTA)	cDNA (FASTA)	CDS (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Annotated sequence (EMBL)	Annotated sequence (GenBank)	Gene sets	Whole databases	Variation (GVF)	Variation (VCF)	Variation (VEP)	Regulation (GFF)	Data files	BAM/BigWig
Y	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig
	Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Ferret <i>Mustela putorius furo</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	BAM/BigWig
	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-



```

156.17.191.216 - PuTTY
>ENSMUST00000083969.1 ncRNA chromosome:GRCm38:X:64800511:64800616:1 gene:ENSMUSG00000065903.1 gene_biotype:snRNA transcript_biotype:snRNA gene_symbol:Gm26111 description:predicted gene, 26111 [Source:MGI Symbol;Acc:MGI:5455888]
GTGCTCACTTCCGCAGCACATACTAAAATTGGAACAAAACAGAGAAGGTTAGCATGGC
CCTTGTACAAGTATGACATGCACGTTTCATGAAGCTTCCATATTTTT
>ENSMUST000000157441.1 ncRNA chromosome:GRCm38:2:71495271:71495374:1 gene:ENSMUSG00000088066.1 gene_biotype:snRNA transcript_biotype:snRNA gene_symbol:Gm23253 description:predicted gene, 23253 [Source:MGI Symbol;Acc:MGI:5453030]
GCTGTGCTCACTTCAGCAGCAAACAGTAAAATCGGAATGAGCAGAGGTTAGTATGACCTT
TGTACAAGGGTGACACAGAAAATTGTGAGGTGCTCCATTAAAAA
>ENSMUST000000178877.1 ncRNA chromosome:GRCm38:16:14807163:14807269:-1 gene:ENSMUSG00000095165.1 gene_biotype:snRNA transcript_biotype:snRNA gene_symbol:Gm25547 description:predicted gene, 25547 [Source:MGI Symbol;Acc:MGI:5455324]
GTGCTCGCTTCGGCAGCACATATACTAAAATTGGAACGATACAGAGAAGATTAGCATGGC
CCCTGCGCAAGGATGACACGCAAATTCGTGAAGCGTTCATATTTTT
  
```

ENSEMBL

Show	10	entries	Show/hide columns											mus		
★	Species	DNA (FASTA)	cDNA (FASTA)	CDS (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Annotated sequence (EMBL)	Annotated sequence (GenBank)	Gene sets	Whole databases	Variation (GVF)	Variation (VCF)	Variation (VEP)	Regulation (GFF)	Data files	BAM/BigWig
Y	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig
	Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Ferret <i>Mustela putorius furo</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	BAM/BigWig
	Mouse 129S1/SvJmJ	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-



```

156.17.191.216 - PuTTY
>ENSMUSP00000142546.1 pep chromosome:GRCm38:14:54113468:54113476:1 gene:ENSMUSG00000096749.2 transcript:ENSMUST00000196221
.1 gene_biotype:TR_D_gene transcript_biotype:TR_D_gene gene_symbol:Trdd1 description:T cell receptor delta diversity 1 [Source:MGI Symbol;Acc:MGI:4439547]
MAY
>ENSMUSP00000142955.1 pep chromosome:GRCm38:14:54122226:54122241:1 gene:ENSMUSG00000096176.1 transcript:ENSMUST00000177564
.1 gene_biotype:TR_D_gene transcript_biotype:TR_D_gene gene_symbol:Trdd2 description:T cell receptor delta diversity 2 [Source:MGI Symbol;Acc:MGI:4439546]
IGGIR
>ENSMUSP00000141764.1 pep chromosome:GRCm38:6:41533201:41533212:1 gene:ENSMUSG00000095668.1 transcript:ENSMUST00000178537.
1 gene_biotype:TR_D_gene transcript_biotype:TR_D_gene gene_symbol:Trbd1 description:T cell receptor beta, D region 1 [Source:MGI Symbol;Acc:MGI:4439571]
GTGG
  
```

ENSEMBL

More extensive sequence annotation by means of feature tables and contain thus the genome sequence as annotated by the automated Ensembl genome annotation pipeline. Each nucleotide sequence record in a flat file represents a 1 Mb slice of the genome sequence.

Show	10	entries	Show/hide columns	mus												
★	Species	DNA (FASTA)	cDNA (FASTA)	CDS (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Annotated sequence (EMBL)	Annotated sequence (GenBank)	Gene sets	Whole databases	Variation (GVF)	Variation (VCF)	Variation (VEP)	Regulation (GFF)	Data files	BAM/BigWig
Y	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig
	Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA		GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Ferret <i>Mustela putorius furo</i>	FASTA	FASTA	FASTA	FASTA	FASTA										
	Mouse 129S1/SvImJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA										
	Mouse A/J <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA										
	Mouse AKR/J <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA										
	Mouse BALB/cJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA										
	Mouse C3H/HeJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA										

156.17.191.216 - PuTTY
LOCUS 1 195471971 bp DNA HTG 9-MAR-2018
DEFINITION Mus musculus chromosome 1 GRCm38 full sequence 1..195471971 reannotated via Ensembl
ACCESSION chromosome:GRCm38:1:1:195471971:1
VERSION 1GRCm38
KEYWORDS .
SOURCE house mouse
ORGANISM Mus musculus
Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria;
Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata;
Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha;
Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria;
Euarchontoglires; Glires; Rodentia; Myomorpha; Muroidea; Muridae.
COMMENT This sequence was annotated by Ensembl(www.ensembl.org). Please visit the Ensembl or EnsemblGenomes web site, http://www.ensembl.org/ or http://www.ensemblgenomes.org/ for more information.

ENSEMBL

GTF → Gene sets for each species. These files include annotations of both coding and non-coding genes.

GFF3 → provides access to all annotated transcripts which make up an Ensembl gene set.

Show	10	entries	Show/hide columns													mus	
★	Species	DNA (FASTA)	cDNA (FASTA)	CDS (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Annotated sequence (EMBL)	Annotated sequence (GenBank)	Gene sets	Whole databases	Variation (GVF)	Variation (VCF)	Variation (VEP)	Regulation (GFF)	Data files	BAM/BigWig	
Y	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig	
	Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	
	Ferret <i>Mustela putorius furo</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	BAM/BigWig	
	Mouse 129S1/SvImJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	
	Mouse A/J <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	
	Mouse AKR/J <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	
	Mouse BALB/cJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	
	Mouse C3H/HeJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-	

GTF → Gene sets for each species. These files include annotations of both coding and non-coding genes

ENSEMBL

156.17.191.216 - PuTTY

```
#!/genome-build GRCm38.p6
#!/genome-version GRCm38
#!/genome-date 2012-01
#!/genome-build-accession NCBI:GCA_000001635.8
#!/genebuild-last-updated 2018-01
1      havana  gene      3073253 3074322 .      +      .      gene_id "ENSMUSG00000102693"; gene_version "1"; gene_name
"4933401J01Rik"; gene_source "havana"; gene_biotype "TEC";
1      havana  transcript 3073253 3074322 .      +      .      gene_id "ENSMUSG00000102693"; gene_version "1"; tr
anscript_id "ENSMUST00000193812"; transcript_version "1"; gene_name "4933401J01Rik"; gene_source "havana"; gene_biotype "T
EC"; transcript_name "4933401J01Rik-201"; transcript_source "havana"; transcript_biotype "TEC"; tag "basic"; transcript_su
pport_level "NA";
1      havana  exon      3073253 3074322 .      +      .      gene_id "ENSMUSG00000102693"; gene_version "1"; transcript
_id "ENSMUST00000193812"; transcript_version "1"; exon_number "1"; gene_name "4933401J01Rik"; gene_source "havana"; gene_b
iotype "TEC"; transcript_name "4933401J01Rik-201"; transcript_source "havana"; transcript_biotype "TEC"; exon_id "ENSMUSE0
0001343744"; exon_version "1"; tag "basic"; transcript_support_level "NA";
1      ensembl gene      3102016 3102125 .      +      .      gene_id "ENSMUSG00000064842"; gene_version "1"; gene_name
"Gm26206"; gene_source "ensembl"; gene_biotype "snRNA";
1      ensembl transcript 3102016 3102125 .      +      .      gene_id "ENSMUSG00000064842"; gene_version "1"; tr
anscript_id "ENSMUST00000082908"; transcript_version "1"; gene_name "Gm26206"; gene_source "ensembl"; gene_biotype "snRNA"
; transcript_name "Gm26206-201"; transcript_source "ensembl"; transcript_biotype "snRNA"; tag "basic"; transcript_support_
level "NA";
1      ensembl exon      3102016 3102125 .      +      .      gene_id "ENSMUSG00000064842"; gene_version "1"; transcript
_id "ENSMUST00000082908"; transcript_version "1"; exon_number "1"; gene_name "Gm26206"; gene_source "ensembl"; gene_biotyp
e "snRNA"; transcript_name "Gm26206-201"; transcript_source "ensembl"; transcript_biotype "snRNA"; exon_id "ENSMUSE0000052
2066"; exon_version "1"; tag "basic"; transcript_support_level "NA";
```

GFF3 → provides access to all annotated transcripts which make up an Ensembl gene set

ENSEMBL

```
156.17.191.216 - PuTTY
##gff-version 3
##sequence-region 1 1 195471971
#!genome-build Ensembl GRCm38.p6
#!genome-version GRCm38
#!genome-date 2012-01
#!genome-build-accession NCBI:GCA_000001635.8
#!genebuild-last-updated 2018-01

1 ensembl_havana gene 3205901 3671498 . - . ID=gene:ENSMUSG00000051951;Name=Xkr4;biotype=protein;description=X-linked Kx blood group related 4 [Source:MGI Symbol%3BAcc:MGI:3528744];gene_id=ENSMUSG00000051951;logically_name=ensembl_havana_gene;version=5
1 havana lnc_RNA 3205901 3216344 . - . ID=transcript:ENSMUST00000162897;Parent=gene:ENSMUSG00000051951;Name=Xkr4-203;biotype=processed_transcript;transcript_id=ENSMUST00000162897;transcript_support_level=1;version=1
1 havana exon 3205901 3207317 . - . Parent=transcript:ENSMUST00000162897;Name=ENSMUSE00000866652;constitutive=0;ensembl_end_phase=-1;ensembl_phase=-1;exon_id=ENSMUSE00000866652;rank=2;version=1
1 havana exon 3213609 3216344 . - . Parent=transcript:ENSMUST00000162897;Name=ENSMUSE00000858910;constitutive=0;ensembl_end_phase=-1;ensembl_phase=-1;exon_id=ENSMUSE00000858910;rank=1;version=1
1 havana lnc_RNA 3206523 3215632 . - . ID=transcript:ENSMUST00000159265;Parent=gene:ENSMUSG00000051951;Name=Xkr4-202;biotype=processed_transcript;transcript_id=ENSMUST00000159265;transcript_support_level=1;version=1
1 havana exon 3206523 3207317 . - . Parent=transcript:ENSMUST00000159265;Name=ENSMUSE00000867897;constitutive=0;ensembl_end_phase=-1;ensembl_phase=-1;exon_id=ENSMUSE00000867897;rank=2;version=1
1 havana exon 3213439 3215632 . - . Parent=transcript:ENSMUST00000159265;Name=ENSMUSE00000863980;constitutive=0;ensembl_end_phase=-1;ensembl_phase=-1;exon_id=ENSMUSE00000863980;rank=1;version=1
1 ensembl_havana mRNA 3214482 3671498 . - . ID=transcript:ENSMUST00000070533;Parent=gene:ENSMUSG00000051951;Name=Xkr4;biotype=mRNA;transcript_id=ENSMUST00000070533;transcript_support_level=1;version=1
```

ENSEMBL

Show 10 entries																
Show/hide columns																
★	Species	DNA (FASTA)	cDNA (FASTA)	CDS (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Annotated sequence (EMBL)	Annotated sequence (GenBank)	Gene sets	Whole databases	Variation (GVF)	Variation (VCF)	Variation (VEP)	Regulation (GFF)	Data files	BAM/BigWig
Y	Mouse <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	GVF	VCF	VEP	Regulation (GFF)	Regulation data files	BAM/BigWig
	Algerian mouse <i>Mus spretus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-		VEP	-	-	-
	Ferret <i>Mustela putorius furo</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	BAM/BigWig
	Mouse 129S1/SvImJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Mouse A/J <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Mouse AKR/J <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Mouse BALB/cJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-
	Mouse C3H/HeJ <i>Mus musculus</i>	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF GFF3	MySQL	-	-	VEP	-	-	-

VEP → Analyse your own variants and predict the functional consequences of known and unknown variants

ENSEMBL TOOLS

The screenshot shows the Ensembl website interface. At the top is a dark blue navigation bar with the Ensembl logo and links for BLAST/BLAT, BioMart, VEP, Tools, Downloads, Help & Docs, and Blog. Below this is a white navigation bar with four main sections: Tools, BioMart, BLAST/BLAT, and Variant Effect Predictor. The Variant Effect Predictor section is highlighted with a yellow arrow. Below the navigation bar is a light blue search bar with a dropdown menu for 'All species', a text input field, and a 'Go' button. Below the search bar is a white section with two columns: 'All genomes' and 'Favourite genomes'. The 'All genomes' column has a dropdown menu for 'Select a species' and links for 'View full list of all Ensembl species' and 'Edit your favourites'. The 'Favourite genomes' column lists three genomes: Human (GRCh38.p12), Mouse (GRCm38.p6), and Zebrafish (GRCz11).

Tools
[All tools](#)

BioMart >
Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT >
Search our genomes for your DNA or protein sequence

Variant Effect Predictor >
Analyse your own variants and predict the functional consequences of known and unknown variants

Search
All species for
Go
e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

All genomes
-- Select a species --
[View full list of all Ensembl species](#)
[Edit your favourites](#)

Favourite genomes
 Human
GRCh38.p12
[Still using GRCh37?](#)
 Mouse
GRCm38.p6
 Zebrafish
GRCz11

VEP

VEP → Analyse your own variants and predict the functional consequences of known and unknown variants

The screenshot shows the Ensembl website's Variant Effect Predictor (VEP) page. The top navigation bar includes links for BLAST/BLAT, BioMart, VEP, Tools, Downloads, Help & Docs, and Blog. A search bar on the right allows users to search for species. Below the navigation bar, a series of tabs categorize the site's content: Using this website, Annotation and prediction, Data access, API & software, and About us. The 'API & software' tab is active, leading to a breadcrumb trail: Home > Help & Documentation > API & Software > Ensembl Tools > Variant Effect Predictor. On the left, a sidebar titled 'In this section' lists various resources, including 'Web interface' (Input form, Results), 'VEP script' (Tutorial, Download and install, Running VEP, Annotation sources, Filtering results, Custom annotations, Plugins, Examples and use cases, Other information), 'Data formats', and 'FAQ'. The main content area is titled 'Variant Effect Predictor' and features the 'Ve!P' logo. It explains that VEP determines the effect of variants (SNPs, insertions, deletions, CNVs, or structural variants) on genes, transcripts, and protein sequences. It prompts users to input variant coordinates and lists the types of information VEP can provide: Genes and Transcripts affected, Location of variants, Consequence of variants on protein sequence, Known variants from the 1000 Genomes Project, SIFT and PolyPhen scores, and more. A large blue button with a play icon and the text 'Launch Ve!P' is positioned in the bottom right corner of the main content area.

Ensembl BLAST/BLAT | BioMart | VEP | Tools | Downloads | Help & Docs | Blog Login/Register

Search all species...

Using this website | Annotation and prediction | Data access | **API & software** | About us

[Home](#) > [Help & Documentation](#) > [API & Software](#) > [Ensembl Tools](#) > **Variant Effect Predictor**

Variant Effect Predictor

Ve!P

VEP determines the effect of your variants (SNPs, insertions, deletions, CNVs or structural variants) on genes, transcripts, and protein sequence, as well as regulatory regions.

Simply input the coordinates of your variants and the nucleotide changes to find out the:

- Genes and Transcripts affected by the variants
- Location of the variants (e.g. upstream of a transcript, in coding sequence, in non-coding RNA, in regulatory regions)
- Consequence of your variants on the protein sequence (e.g. stop gained, missense, stop lost, frameshift)
- Known variants that match yours, and associated minor allele frequencies from the 1000 Genomes Project
- SIFT and PolyPhen scores for changes to protein sequence
- ... And more! See [data types](#), [versions](#).

Search documentation... **Go**

Launch Ve!P

VEP

VEP → Analyse your own variants and predict the functional consequences of known and unknown variants

e!Ensembl BLAST/BLAT | BioMart | Tools | Downloads

Species ▾ VEP ▾

Web Tools

- Web Tools
 - BLAST/BLAT
 - Variant Effect Predictor**
 - File Chameleon
 - Assembly Converter
 - ID History Converter

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Variant Effect Predictor

VEP for Human GRCh37

If you are looking for VEP for Human

Species:

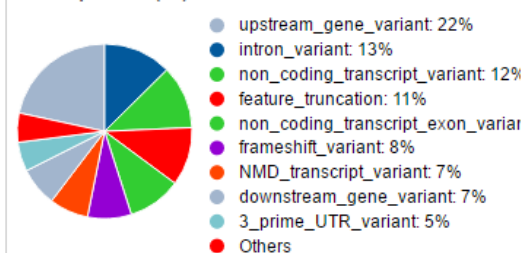
Name for this job (optional):

Either paste data:

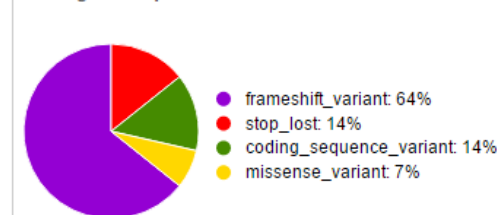
Or upload file:

Category	Count
Variants processed	6
Variants filtered out	0
Novel / existing variants	5 (83.3) / 1 (16.7)
Overlapped genes	10
Overlapped transcripts	70
Overlapped regulatory features	1

Consequences (all)



Coding consequences



Results preview

Navigation

Page: 1 of 1 | Show: All variants

Filters

Uploaded variant is defined

Download

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

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

Show/hide columns

Uploaded variant	Location	Allele	Consequence	Impact	Symbol	Gene	Feature type	Feature	Biotype
1_160283_duplication	1:160282-160282	duplication	upstream_gene_variant	MODIFIER	RNU6-1100P	ENSG00000222623	Transcript	ENST00000410691	snRNA
1_160283_duplication	1:160282-160282	duplication	non_coding_transcript_exon_variant, intron_variant, non_coding_transcript_variant	MODIFIER	RP11-34P13.13	ENSG00000241860	Transcript	ENST00000466557	lincRNA
1_160283_duplication	1:160282-160282	duplication	non_coding_transcript_exon_variant, intron_variant, non_coding_transcript_variant	MODIFIER	RP11-34P13.13	ENSG00000241860	Transcript	ENST00000491962	lincRNA
1_160283_duplication	1:160282-160282	duplication	non_coding_transcript_exon_variant, intron_variant, non_coding_transcript_variant	MODIFIER	RP11-34P13.9	ENSG00000241599	Transcript	ENST00000496488	lincRNA
1_182712_A/C	1:182712-182712	C	downstream_gene_variant	MODIFIER	FO538757.1	ENSG00000279457	Transcript	ENST00000623083	protein_coding
1_182712_A/C	1:182712-182712	C	downstream_gene_variant	MODIFIER	FO538757.1	ENSG00000279457	Transcript	ENST00000623834	protein_coding
1_182712_A/C	1:182712-182712	C	missense_variant	MODERATE	FO538757.2	ENSG00000279928	Transcript	ENST00000624431	protein_coding
1_182712_A/C	1:182712-182712	C	downstream_gene_variant	MODIFIER	FO538757.1	ENSG00000279457	Transcript	ENST00000624735	protein_coding
1_1385015_deletion	1:1385014-1385014	deletion	stop_lost, coding_sequence_variant, 3_prime_UTR_variant, feature_truncation	HIGH	CCNL2	ENSG00000221978	Transcript	ENST00000400809	protein_coding

BIOMART

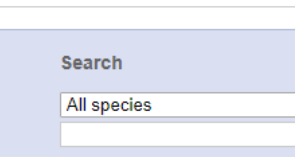
 [BLAST/BLAT](#) | [BioMart](#) | [VEP](#) | [Tools](#) | [Downloads](#) | [Help & Docs](#) | [Blog](#)

Tools
[All tools](#)

BioMart >
Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT >
Search our genomes for your DNA or protein sequence

Variant Effect Predictor >
Analyse your own variants and predict the functional consequences of known and unknown variants



Search

All species ▼

for

Go

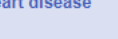
e.g. [BRCA2](#) or [rat 5:62797383-63627669](#) or [rs699](#) or [coronary heart disease](#)

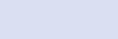
All genomes

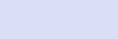
-- Select a species -- ▼

- [View full list of all Ensembl species](#)
- [Edit your favourites](#)

Favourite genomes

**Human**
GRCh38.p12
[Still using GRCh37?](#)

**Mouse**
GRCm38.p6

**Zebrafish**
GRCz11

BioMart → Export custom datasets from Ensembl with this data-mining tool

BIOMART

Dataset
[None selected]

- CHOOSE DATABASE - ▼

↑

Dataset:
Ensembl genes
Mouse strains
Ensembl Variation
Ensembl Regulation

BioMart → Export custom datasets from Ensembl with this data-mining tool

BIOMART

The screenshot shows the Ensembl BioMart web interface. At the top, the Ensembl logo is followed by navigation links: BLAST/BLAT, BioMart, BLAST, VEP, Tools, Downloads, Help & Docs, and Blog. A search bar on the right says "Search all species...". Below the navigation bar is a toolbar with buttons for "New", "Count", "Results", "URL", "XML", "Perl", and "Help". The main content area has two dropdown menus: "Ensembl Genes 92" and "Cow genes (UMD3.1)". On the left, a sidebar contains sections for "Dataset" (Cow genes (UMD3.1)), "Filters" ([None selected]), "Attributes" (Gene stable ID, Transcript stable ID), and another "Dataset" section ([None Selected]). A yellow box on the right contains a legend for the sidebar items:

Filters	→ specyfikacja przeszukiwań
Attributes	→ określenie formatowania danych wyjściowe
Count	→ dostępne rekordy o określonych parametrach
Results	→ wyniki, eksport plików

BioMart → Export custom datasets from Ensembl with this data-mining tool

BIOMART

The screenshot shows the Ensembl BioMart interface. The top navigation bar includes links for BLAST/BLAT, BioMart, BLAST, VEP, Tools, Downloads, Help & Docs, and Blog. A search bar on the right says "Search all species...". Below the navigation bar, there are buttons for "New", "Count", and "Results". On the left sidebar, under the "Dataset" section, "Cow Structural Variants (UMD3.1)" is selected. Under the "Filters" section, "Limit to variants from this source: DGVa" is selected. Under the "Attributes" section, "Study accession", "Structural variant name", "Chromosome/scaffold name", "Chromosome/scaffold position start (bp)", and "Chromosome/scaffold position end (bp)" are listed. In the main content area, "Ensembl Variation 92" is selected in the first dropdown, and "Cow Structural Variants (UMD3.1)" is selected in the second dropdown. A large yellow arrow points from the "BioMart" link in the top navigation bar to the "Results" button. Another yellow arrow points from the "Results" button to the "Ensembl Variation 92" dropdown. A third yellow arrow points from the "Ensembl Variation 92" dropdown to the "Cow Structural Variants (UMD3.1)" dropdown. A fourth yellow arrow points from the "Cow Structural Variants (UMD3.1)" dropdown to the left sidebar.

Ensembl BLAST/BLAT | BioMart | BLAST | VEP | Tools | Downloads | Help & Docs | Blog

Search all species...

New Count Results

★ URL XML Perl Help

Dataset
Cow Structural Variants (UMD3.1)

Filters
Limit to variants from this source: DGVa

Attributes
Study accession
Structural variant name
Chromosome/scaffold name
Chromosome/scaffold position start (bp)
Chromosome/scaffold position end (bp)

Dataset
[None Selected]

Ensembl Variation 92

Cow Structural Variants (UMD3.1)

BioMart → Export custom datasets from Ensembl with this data-mining tool

BIOMART

e!Ensembl BLAST/BLAT | BioMart | BLAST | VEP | Tools | Downloads | Help & Docs | Blog Search all species...

[New](#) [Count](#) [Results](#) [★ URL](#) [XML](#) [Perl](#) [Help](#)

Dataset

Cow Structural Variants (UMD3.1)

Filters

Limit to variants from this source: DGVa

Chromosome/scaffold: 29

Attributes

Study accession

Structural variant name

Chromosome/scaffold name

Chromosome/scaffold position start (bp)

Chromosome/scaffold position end (bp)

Dataset

[None Selected]

Please restrict your query using criteria below
(If filter values are truncated in any lists, hover over the list item to see the full text)

☒ **REGION**

☒ Chromosome/scaffold

25
26
27
28
29
X

☐ **Coordinates**

Start

End

☐ **Marker**

Marker Start

Marker End

BioMart → Export custom datasets from Ensembl with this data-mining tool

BIOMART



e!Ensembl BLAST/BLAT | **BioMart** | BLAST | VEP | Tools | Downloads | Help & Docs | Blog Search all species... Login/Register

New **Count** **Results** **★ URL** **XML** **Perl** **Help**

Dataset 365 / 10462 Entries
Cow Structural Variants (UMD3.1)

Filters
Limit to variants from this source: DGVa
Chromosome/scaffold: 29

Attributes
Study accession
Structural variant name
Chromosome/scaffold name
Chromosome/scaffold position start (bp)
Chromosome/scaffold position end (bp)

Dataset
[None Selected]

Please restrict your query using criteria below
(If filter values are truncated in any lists, hover over the list item to see the full text)

☒ **REGION**

☒ Chromosome/scaffold

25 ▲
26
27
28
29
X ▼

☐ **Coordinates**

Start

End

☐ **Marker**

Marker Start

Marker End

BioMart → Export custom datasets from Ensembl with this data-mining tool

BIOMART

e!Ensembl | [FAST/BLAT](#) | [BioMart](#) | [BLAST](#) | [VEP](#) | [Tools](#) | [Downloads](#) | [Help & Docs](#) | [Blog](#) Search all species...

[New](#) [Count](#) [Results](#) [★ URL](#) [XML](#) [Perl](#) [Help](#)

Dataset 365 / 10462 Entries
Cow Structural Variants (UMD3.1)

Filters
Limit to variants from this source: DGVa
Chromosome/scaffold: 29

Attributes
Study accession
Structural variant name
Chromosome/scaffold name
Chromosome/scaffold position start (bp)
Chromosome/scaffold position end (bp)

Dataset
[None Selected]

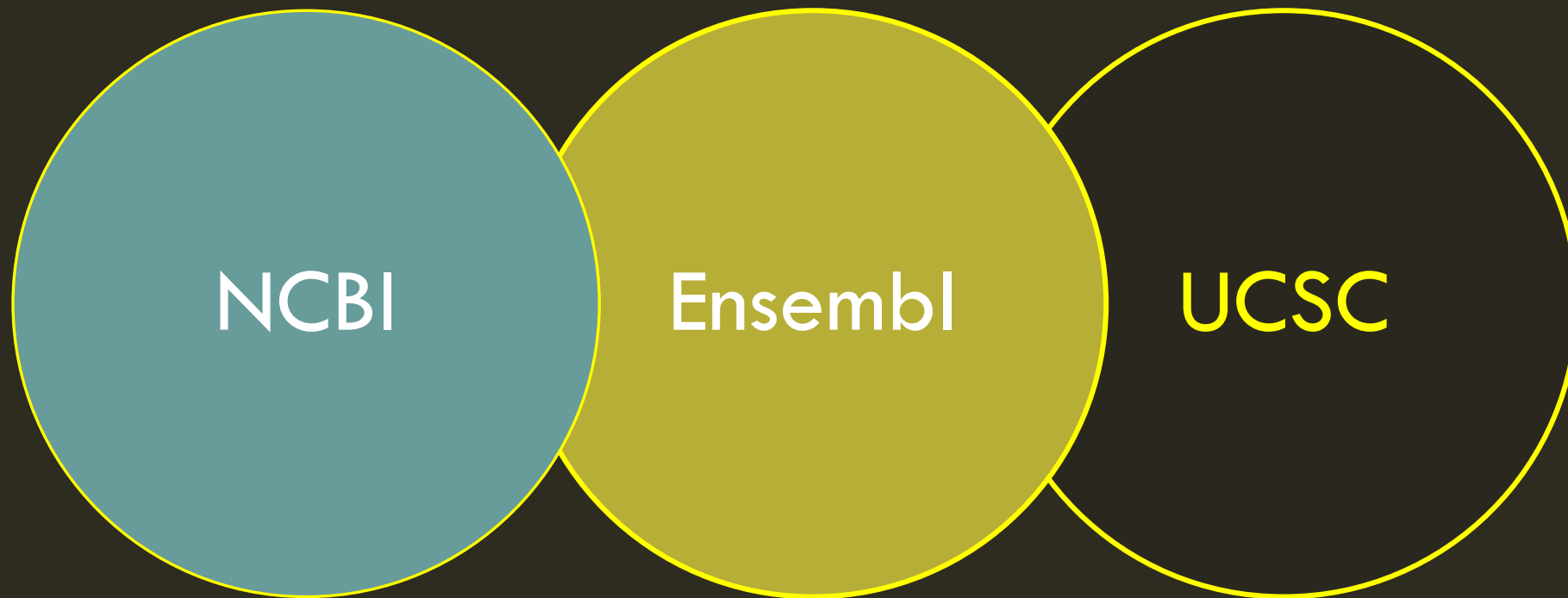
Export all results to ☐ Unique results only [Go](#)

Email notification to

View rows as ☐ Unique results only

Study accession	Structural variant name	Chromosome/scaffold name	Chromosome/scaffold position start (bp)	Chromosome/scaffold position end (bp)
nstd69	nsv834954	29	68826	191656
nstd69	nsv834925	29	499056	534829
nstd69	nsv836026	29	535750	577894
nstd60	nsv810457	29	630088	724199
estd223	esv3895277	29	903671	903752
estd223	esv3899787	29	1391865	1492766
nstd131	nsv2727400	29	1675001	1685000
estd223	esv3895278	29	1801340	1801393
nstd131	nsv2727427	29	1802501	1820000
nstd69	nsv834903	29	2072222	2082245
nstd135	nsv2779379	29	2371184	2378383
nstd56	nsv616135	29	2530804	2656091
nstd131	nsv2727401	29	3570001	3577500
estd223	esv3895280	29	4062992	4063066

GENOMY I ICH ADNOTACJE





UNIVERSITY OF CALIFORNIA
SANTA CRUZ



Genome Browser

[Home](#)
[Genomes](#)
[Genome Browser](#)
[Tools](#)
[Mirrors](#)
[Downloads](#)
[My Data](#)
[Help](#)
[About Us](#)



Our tools

- **Genome Browser**
interactively visualize genomic data
- **BLAT**
rapidly align sequences to the genome
- **Table Browser**
download data from the Genome Browser database
- **Variant Annotation Integrator**
get functional effect predictions for variant calls
- **Data Integrator**
combine data sources from the Genome Browser database
- **Gene Sorter**
find genes that are similar by expression and other metrics
- **Genome Browser in a Box (GBiB)**
run the Genome Browser on your laptop or server
- **In-Silico PCR**
rapidly align PCR primer pairs to the genome
- **LiftOver**
convert genome coordinates between assemblies
- **VisiGene**
interactively view in situ images of mouse and frog

[More tools...](#)

Our story

On June 22, 2000, UCSC and the other members of the International Human Genome Project consortium completed the first working draft of the human genome assembly, forever ensuring free public access to the genome and the information it contains. A few weeks later, on July 7, 2000, the newly assembled genome was released on the web at <http://genome.ucsc.edu>, along with the initial prototype of a graphical viewing tool, the UCSC Genome Browser. In the ensuing years, the website has grown to include a broad collection of vertebrate and model organism assemblies and annotations, along with a large suite of tools for viewing, analyzing and downloading data.

What's new

Apr. 04, 2018 - [New NCBI RefSeq tracks for model organisms](#)

Apr. 03, 2018 - [New tool: Track Collection Builder](#)

Feb. 20, 2018 - [New video: Visibility control in the Browser](#)

[More news...](#)

[Subscribe](#)

*The UCSC Genome Browser is developed and maintained by the **Genome Bioinformatics Group**, a cross-departmental team within the **UCSC Genomics Institute**.*



UNIVERSITY OF CALIFORNIA
UCSC

Genome Browser

Home Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us



Our tools

- **Genome Browser**
interactively visualize genomic data
- **BLAT**
rapidly align sequences to the genome
- **Table Browser**
download data from the Genome Browser database
- **Variant Annotation Integrator**
get functional effect predictions for variant calls
- **Data Integrator**
combine data sources from the Genome Browser database
- **Gene Sorter**
find genes that are similar by expression and other metrics
- **Genome Browser in a Box (GBiB)**
run the Genome Browser on your laptop or server
- **In-Silico PCR**
rapidly align PCR primer pairs to the genome
- **LiftOver**
convert genome coordinates between assemblies
- **VisiGene**
interactively view in situ images of mouse and frog

[More tools...](#)

Our story

On June 22, 2000, UCSC and the other members of the International Human Genome Project consortium completed the first working draft of the human genome assembly, forever ensuring free public access to the genome and the information it contains. A few weeks later, on July 7, 2000, the newly assembled genome was released on the web at <http://genome.ucsc.edu>, along with the initial prototype of a graphical viewing tool, the UCSC Genome Browser. In the ensuing years, the website has grown to include a broad collection of vertebrate and model organism assemblies and annotations, along with a large suite of tools for viewing, analyzing and downloading data.

What's new

Apr. 04, 2018 - [New NCBI RefSeq tracks for model organisms](#)

Apr. 03, 2018 - [New tool: Track Collection Builder](#)

Feb. 20, 2018 - [New video: Visibility control in the Browser](#)

[More news...](#)

[Subscribe](#)

The UCSC Genome Browser is developed and maintained by the Genome Bioinformatics Group, a cross-departmental team within the UCSC Genomics Institute.

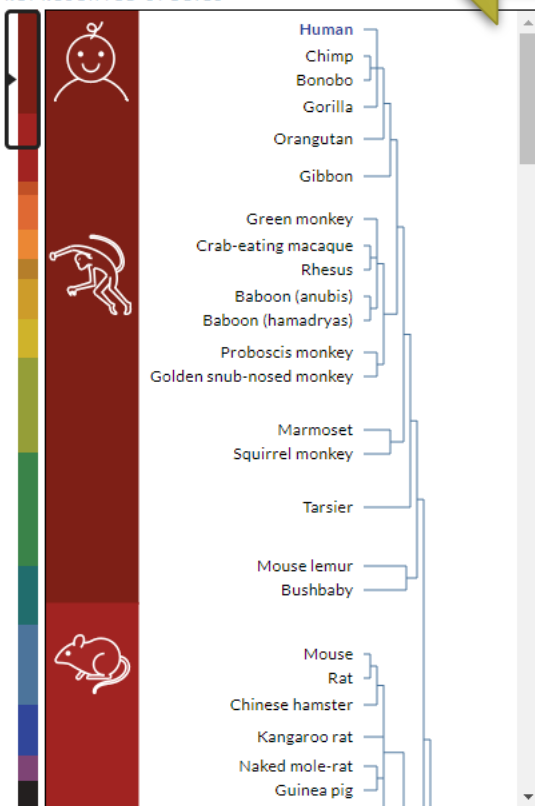
Browse/Select Species

POPULAR SPECIES



Enter species or common name

REPRESENTED SPECIES



Find Position

Human Assembly

Dec. 2013 (GRCh38/hg38)

Position/Search Term

Enter position, gene symbol or search terms

Current position: chr1:11,102,837-11,267,747

Human Genome Browser - hg38 assembly

[view sequences](#)

UCSC Genome Browser assembly ID: hg38

Sequencing/Assembly provider ID: Genome Reference Consortium Human GRCh38 (GCA_000001405.15)

Assembly date: Dec. 2013

Accession ID: [GCA_000001405.15](#)

NCBI Genome ID: 51 (Homo sapiens (human))

NCBI Assembly ID: 883148 (GRCh38, GCA_000001405.15)

BioProject ID: 31257

Search the assembly:

- **By position or search term:** Use the "position or search term" box to find areas of the genome associated with many different attributes, such as a specific chromosomal coordinate range; mRNA, EST, or STS marker names; or keywords from the GenBank description of an mRNA. [More information](#), including sample queries.
- **By gene name:** Type a gene name into the "search term" box, choose your gene from the drop-down list, then press "submit" to go directly to the assembly location associated with that gene. [More information](#).
- **By track type:** Click the "track search" button to find Genome Browser tracks that match specific selection criteria. [More information](#).

Download sequence and annotation data:

- [Using rsync](#) (recommended)
- [Using FTP](#)
- [Using HTTP](#)
- [Data use conditions and restrictions](#)
- [Acknowledgments](#)

Assembly Details

The GRCh38 assembly is the first major revision of the human genome released in more than four years. As with the previous assembly, the [Genome Reference Consortium](#) (GRC) is now the primary source for human genome assembly data submitted to GenBank. Beginning with this release, the UCSC Genome Browser version of the human assemblies now match those of the GRC to minimize version confusion. Hence, the GRCh38 assembly is referred to as "hg38" in the Genome Browser datasets and documentation. For more assembly-related terms, see the [GRC Assembly Terminology](#) page.



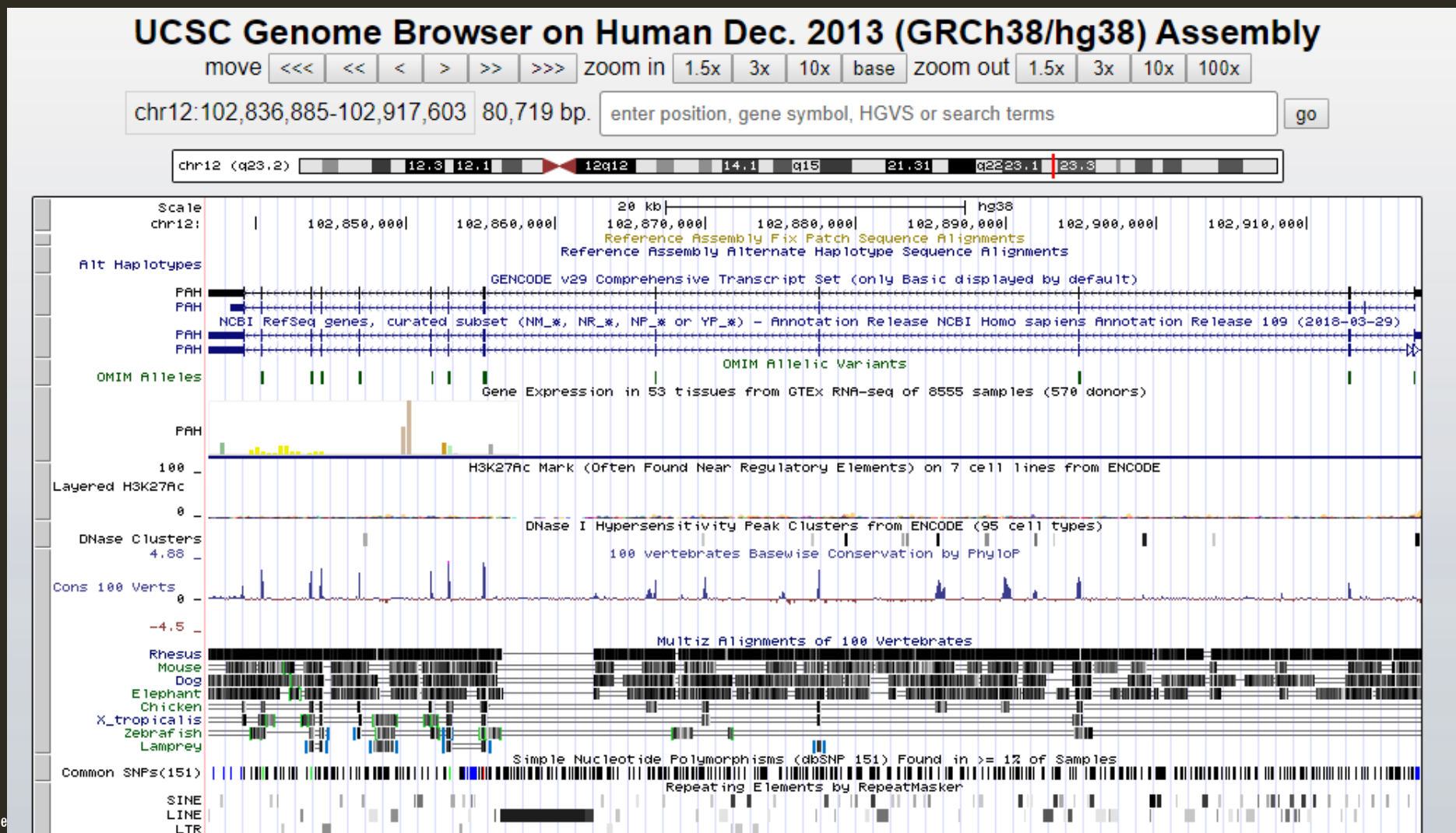
Homo sapiens
(Graphic courtesy of CBSE)

UCSC — GEN PAH

Gen PAH (enzym hydroksylazy fenyloalaninowej)

Na którym chromosomie jest zlokalizowany?

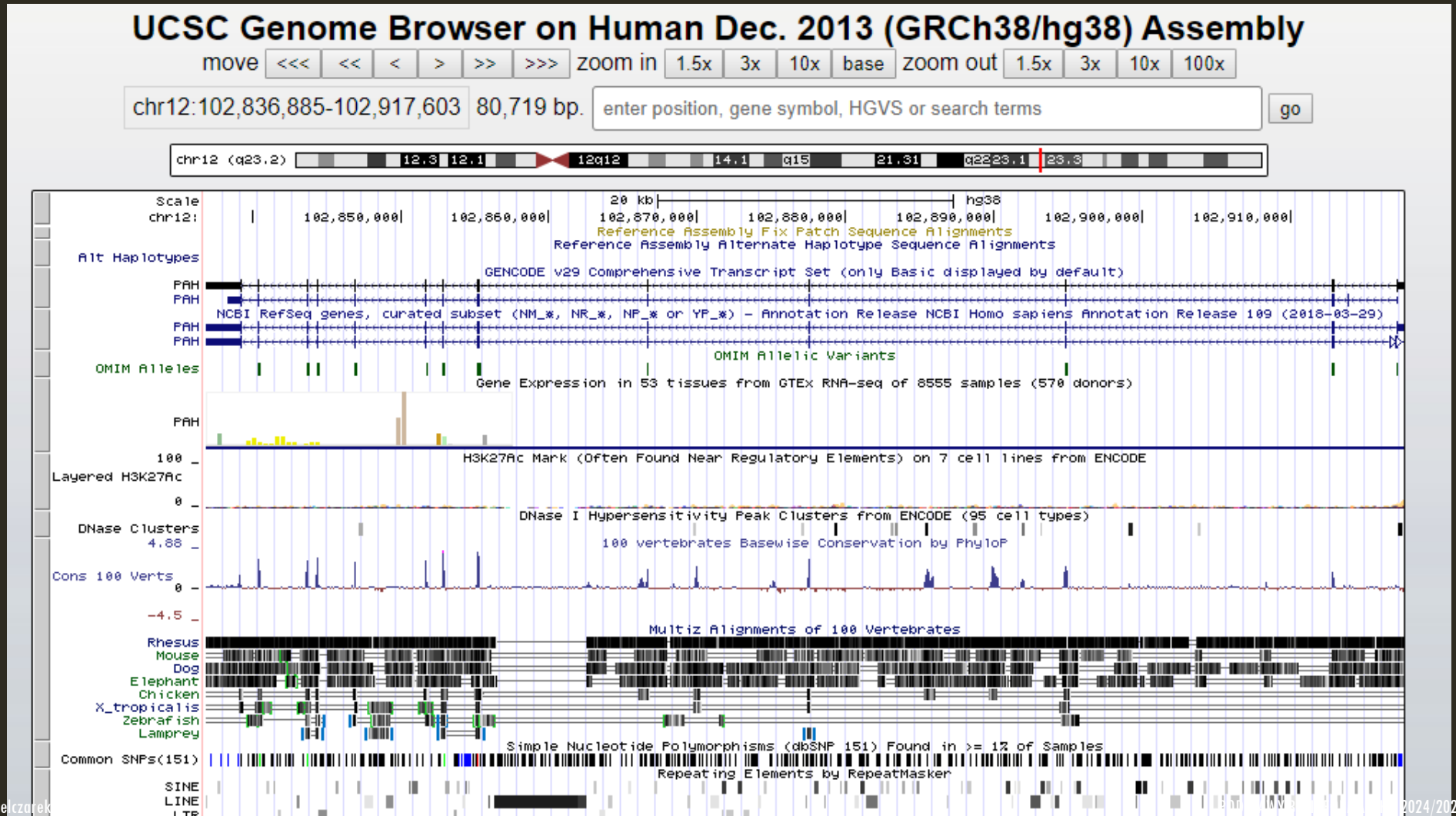
Jakiej długości jest sekwencja?



Gen PAH (enzym hydroksylazy fenyloalaninowej)

Ile ma egzonów?

UCSC — GEN PAH



UCSC — GEN PAH



UCSC



Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the OpenHelix Table Browser [tutorial](#) for a narrated presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our [public MySQL server](#). To examine the biological function of your set through annotation enrichments, send the data to [GREAT](#). Send data to [GenomeSpace](#) for use with diverse computational tools. Refer to the [Credits](#) page for the list of contributors and usage restrictions associated with these data. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

clade: genome: assembly:

group: track:

table:

region: ☒ genome ☐ position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format: Send output to ☐ [Galaxy](#) ☐ [GREAT](#) ☐ [GenomeSpace](#)

output file: (leave blank to keep output in browser)

file type returned: ☒ plain text ☐ gzip compressed

the first working draft of the human genome assembly, forever ensuring free public access to the genome and the information it contains. A few weeks later, on July 7, 2000, the newly assembled genome was released on the web at <http://genome.ucsc.edu>, along with the initial prototype of a graphical viewing tool, the UCSC Genome Browser. In the ensuing years, the website has grown to include a broad collection of vertebrate and model organism assemblies and annotations, along with a large suite of tools for viewing, analyzing and downloading data.

Apr. 03, 2018 - New tool: [Track Collection Builder](#)

Feb. 20, 2018 - New video: [Visibility control in the Browser](#)

[More news...](#)

[Subscribe](#)

The UCSC Genome Browser is developed and maintained by the [Genome Bioinformatics Group](#), a cross-departmental team within the [UCSC Genomics Institute](#).