

Training material

- <https://slegras.github.io/PhDprogram/>

Introduction to Ensembl/Biomart

PhD Program

Stéphanie Le Gras
[\(slegras@igbmc.fr\)](mailto:slegras@igbmc.fr)

Guidelines

- Introduction
 - Genomes
 - Genome Browser
 - Ensembl project
- Ensembl genome browser
 - Browse Ensembl
 - Tools
- Mining Ensembl
 - BioMart

Genomes

Genomes

```
...CGAGGGGCCTAGACATTGCCCTCCAGAGAGAGCACCAACACCCCTCCAGGC  
TTGACCGAGCCAGGGTGTCCCTTCCTACCTTGGAGAGAGCAGCCCCAGGGCA  
TCCTGCAGGGGGTGCATGGACACCAGCTGGCCTCAAGGTCTGCCTCCCT  
CCAGCCACCCCCTACACGCTGCATGGATCTGGATCTCAGCTCCCTGGCCG  
ACAACACTGGCAAACCTACTCATCCACGAAAGGCCCTCTGGCATGGTGG  
TCCTTCCCAGCCTGGCAGTCTGTTCTCACACACCTGTATAGTGCCAGCCCT  
GAGGTTGCAGCTGGGGTGTCTGAAGGGCTGTGAGCCCCCAGGAAAGCCC  
TGGGGAAAGTGCCTGCCTGCCTCCCCCGGCCAGCGCCTGGCTGCC  
CATCCTACCTGGCTCCCCCATCCAGCCTCCCTACACACTCCTCAAGG  
AGGCACCATGTCCTCCAGCTGCCGGCCTCAGAGCACTGTGGCGTCCTG  
GGCAGCCACCGCATAGTCTGCTGGCATGGCTCAGGGTGAAAGGGCG  
GAAGGGAGGGTCTGCAGATAGCTAGGTGCCACTACAAACCCGCTCGGG  
GCAGGAGAGCCAAGGCTGGGTGTGCAGAGCGAGCCCCGAGAGGTTCCG  
AGGCTGAGGCCAGGGTGGGACATAGGGATGCGAGGGGCCGGGCAACAGG  
ATACTCAACCTGCCTGCCCATGGTCTCATCCTCTGCTTCTGGACCTCTG  
AATCCTGCCCTGGTCTAAGAGGCAGGTAAAGGGCTGCAGGCAGGGCT  
CGGAGCCAAGTGGGGCCAGGACGGAGCTGGCCAGTGCACAGCTTCCC  
ACACCTGCCACCCCCAGAAGTCTGCCGCCACCCCCAGATCACAGGAAGAT  
GAGGTCCGAGTGGCTGCTGAGGACTTAGCTGCTGTCCCCAGGTCCCCAGGT  
CATGCCCTCTGGCCACCCCTGGGAGCTGAGGGACCTCAGCTGGGCTGCT  
GTCCTAAGGCAGGGTGGGAACTAGGCAGCCAGCAGGGAGGGAAACCCCTCC  
CTCACTCCCCTCTGGCCACCCCTGGGACAGGGTCCGGGACA  
GGGCATCGGGACTGGGGACAGGGTCTGGGGACAGGGTGAGCAGGGACAGGTCT  
GGCCACCGGGCCCTGGTTAACAC.....
```

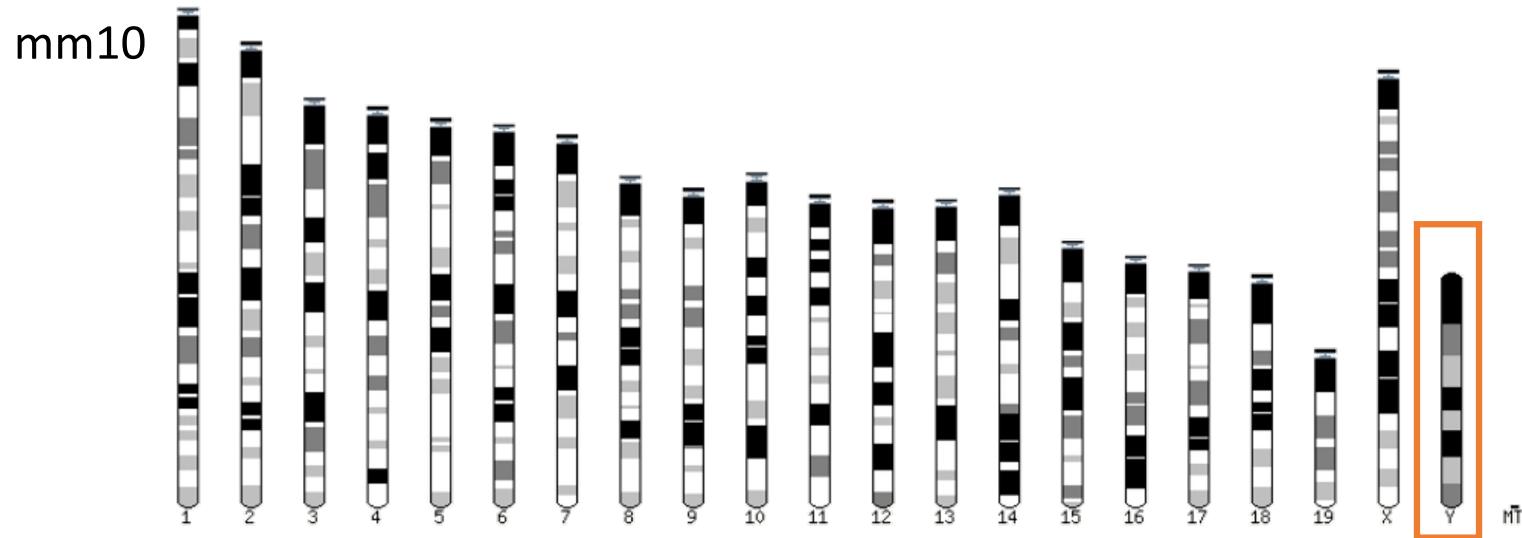
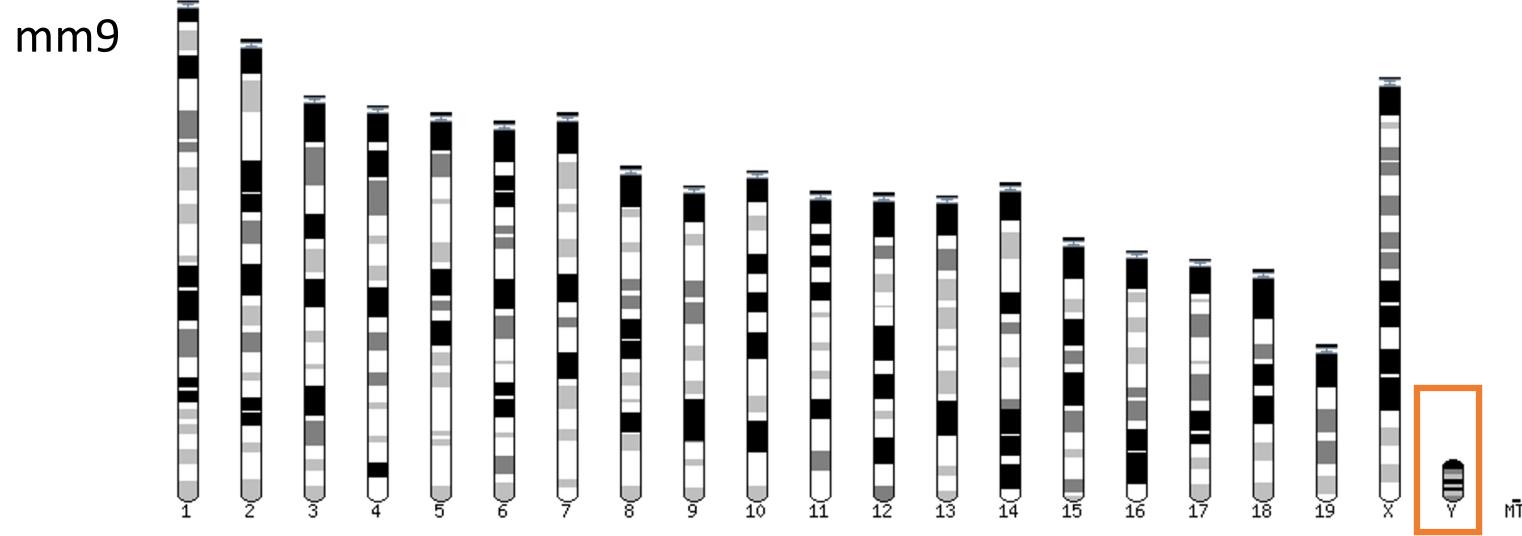
- 2000: First draft of the human genome
- 2003: Human genome sequencing complete

Genome builds

SPECIES	UCSC VERSION	RELEASE DATE	RELEASE NAME	STATUS
MAMMALS				
Human	hg38	Dec. 2013	Genome Reference Consortium GRCh38	Available
	hg19	Feb. 2009	Genome Reference Consortium GRCh37	Available
	hg18	Mar. 2006	NCBI Build 36.1	Available
	hg17	May 2004	NCBI Build 35	Available
	hg16	Jul. 2003	NCBI Build 34	Available
	hg15	Apr. 2003	NCBI Build 33	Archived
	hg13	Nov. 2002	NCBI Build 31	Archived
	hg12	Jun. 2002	NCBI Build 30	Archived
	hg11	Apr. 2002	NCBI Build 29	Archived (data only)
	hg10	Dec. 2001	NCBI Build 28	Archived (data only)
	hg8	Aug. 2001	UCSC-assembled	Archived (data only)
	hg7	Apr. 2001	UCSC-assembled	Archived (data only)
	hg6	Dec. 2000	UCSC-assembled	Archived (data only)
	hg5	Oct. 2000	UCSC-assembled	Archived (data only)
	hg4	Sep. 2000	UCSC-assembled	Archived (data only)
	hg3	Jul. 2000	UCSC-assembled	Archived (data only)
	hg2	Jun. 2000	UCSC-assembled	Archived (data only)
	hg1	May 2000	UCSC-assembled	Archived (data only)

Source: <https://genome.ucsc.edu/FAQ/FAQreleases.html>

Genome builds



Need annotations

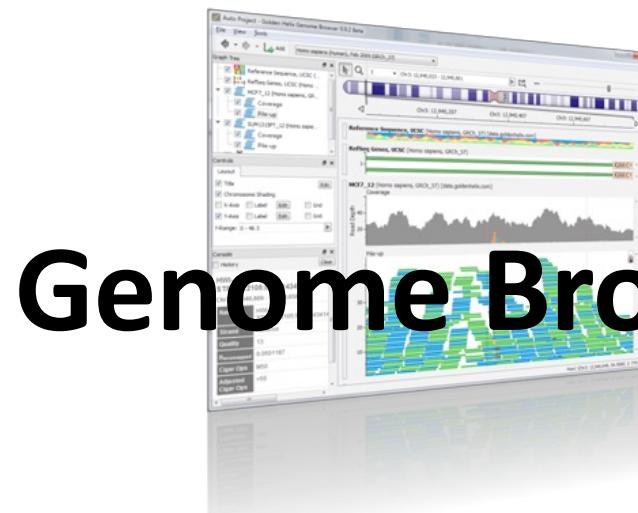
...CGAGGGGCCTAGACATTGCCCTCCAGAGAGAGCACCAACACCCCTCCAGGC
TTGACCGAGCCAGGGTGTCCCTTCACCTTGAGAGAGCAGCCCCAGGGCA
TCCTGCAGGGGGTGCATGGACACCAGCTGGCCTCAAGGTCTGCTCCCT
CCAGCCACCCCCTACACGCTGCATGGATCTGGATCTCAGCTCCCTGGCCG
ACAACACTGGCAAACCTACTCATCCACGAAAGGCCCTCTGGCATGGTGG
TCCTTCCAGCCTGGCAGTCTGTTCTCACACACCTGTATAGTGCCAGCCCT
GAGGTTGCAGCTGGGGTGTCTGAAGGGCTGTGAGCCCCCAGGAAAGCCC
TGGGGAAAGTGCCTGCCTGCCTCCCCCGGCCAGCAGCCTGGCTCTGCC
CATCCTACCTGGCTCCCCCATCCAGCCTCCCTCACACACTCCTCAAGG
AGGCACCATGTCCTCCAGCTGCCGGCCTCAGAGCACTGTGGCGTCCTG
GGCAGCCACCGCATAGTCTGCTGGCATGGCTCAGGGTGGAAAGGGCG
GAAGGGAGGGTCTGCAGATAGCTAGGTGCCACTACAAACCCGCTCGGG
GCAGGAGAGCCAAGGCTGGGTGTGCAGAGCAGCCCCGAGAGGTTCCG
AGGCTGAGGCCAGGGTGGGACATAGGGATGCGAGGGGCCGGGCAACAGG
ATACTCCAACCTGCCTGCCCATGGTCTCATCCTCTGCTTCTGGGACCTCTG
AATCCTGCCCTGGTCTAAGAGGCAGGTAAAGGGCTGCAGGCAGGGCT
CGGAGCCCAACTGGGGGCCAGGACGGAGCTGGCCAGTCAGCTTCC
ACACCTGCCACCCCCAGAAGTCCTGCCACCCCCAGATCACAGGAAGAT
GAGGTCCGAGTGGCTGCTGAGGACTTAGCTGCTGTCCCCAGGTCCCCAGGT
CATGCCCTCTGCCACCCCTGGGAGCTGAGGGACCTCAGCTGGGCTGCT
GTCTTAAGGCAGGGTGGGAACTAGGCAGCCAGCAGGGAGGGAAACCCCTCC
CTCACTCCCCTCTGCCACCCCTGGGAGCTGAGGGACAGGGTGGGACAGGGT
GGGCATCGGGACTGGGACAGGGTCTGGGACAGGGTCCGGGACA
GGGTCTGGAGGACAGGGTGTGGGACAGGGTGAGCAGGGACAGGTCT
GGCCACCGGGCCCTGGTTAACAC....



Biological
information

Get **access** to genomic data

- Need a way to gather all genomic information in one place
- Availability of the data
- Accessibility to the data



Genome Browser

Genome browsers

Genome browsers

- Graphical interface to display genomic data
- Visualize and browse entire genomes with annotated data
 - Gene prediction and structure
 - Proteins,
 - Expression,
 - Regulation,
 - Variation,
 - Comparative analysis...

There are Genome browsers...

EBI - Ensembl

The screenshot shows the EBI Ensembl genome browser interface for Human Chromosome 12. The top navigation bar includes links for BLAST/BLAT, Sanger, Tools, Downloads, Help & Documentation, Blog, and Mirror. The main content area displays a genomic track for chromosome 12, spanning from 76,738,254 to 76,742,222. The track includes various genomic features such as genes, transcripts, and regulatory elements. A detailed 'Region in detail' panel is open, showing specific gene structures and their associated genomic tracks. The bottom of the page includes a search bar and links for Help, Contact, and Log in.

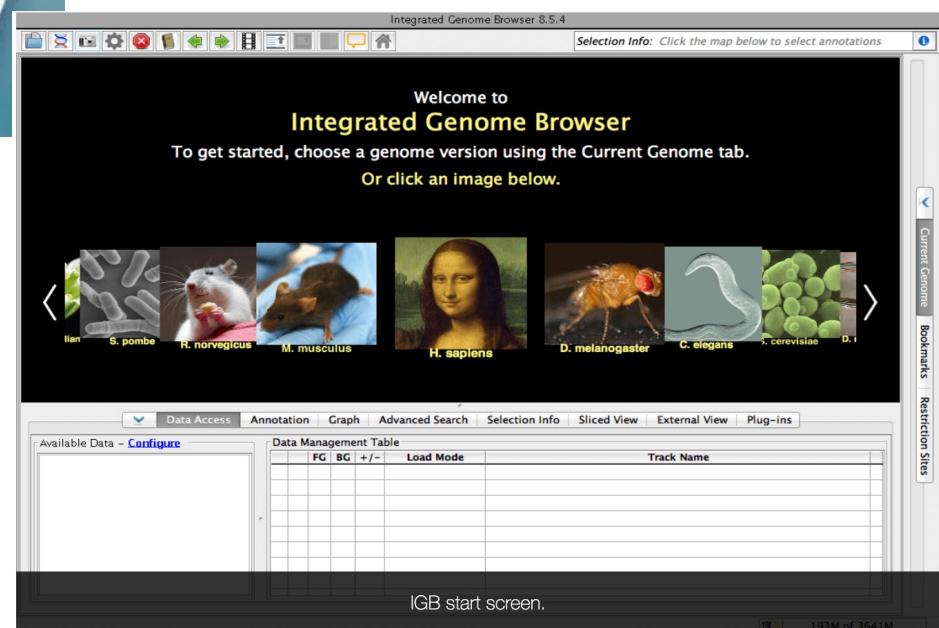
UCSC – Genome Browser

The screenshot shows the UCSC Genome Browser interface for Human Feb. 2009 (GRCh37/hg19) Assembly. The top navigation bar includes links for Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, About Us, View, and Help. The main content area displays a genomic track for chromosome 21, spanning from 33,021,623 to 33,051,544. The track includes various genomic features and tracks. A detailed 'Region in detail' panel is open, showing specific gene structures and their associated genomic tracks. The bottom of the page includes a search bar and links for Help, Contact, and Log in.

The screenshot shows the NCBI Map Viewer interface for Human chromosomes 76,070K to 77,410K. The top navigation bar includes links for PubMed, Entrez, BLAST, OMIM, Taxonomy, and Structure. The main content area displays a genomic map with various genomic features and tracks. A detailed 'Summary of Maps' panel is open, showing specific gene structures and their associated genomic tracks. The bottom of the page includes a search bar and links for Help, Contact, and Log in.

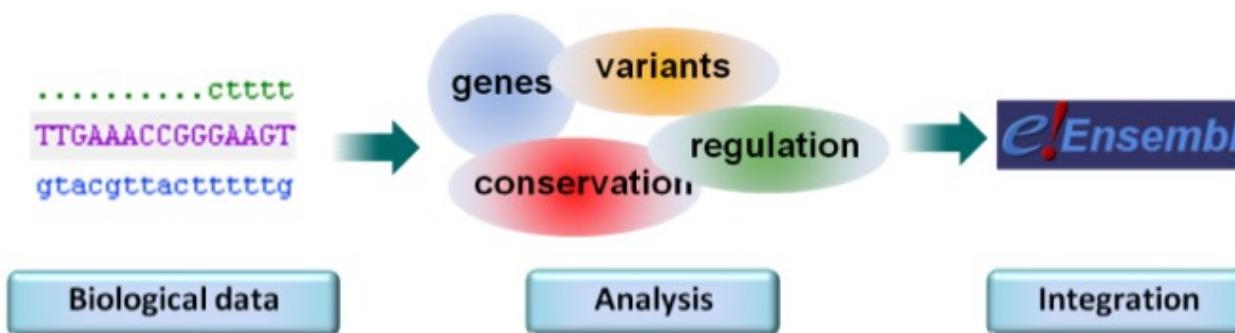
NCBI – Genome Data Viewer

And Genome **browsers**...



Ensembl project

- Available data :
 - Genomes
 - Comparative genomic data
 - Variations
 - Gene regulatory elements
 - External Annotations



- Website is launch in 2000 (at the beginning, only the human genome)

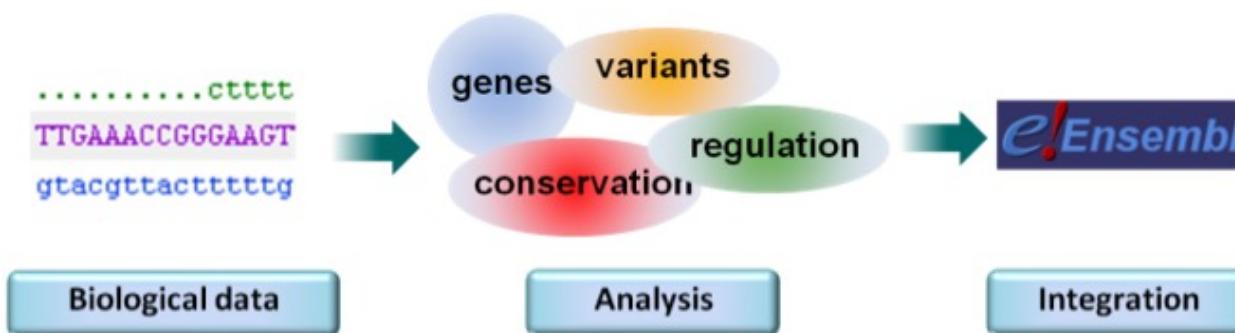
Ensembl project

Ensembl project

- Started in 1999 (before the draft human genome was completed)
- Joint project between European Bioinformatics Institute (EBI) and Wellcome Trust Sanger Institute (WTSI)
- Goal of Ensembl:
 - Automatically annotate genome
 - Integrate this annotation with other available biological data
 - Make all this publicly available via the web
- Ensembl do not produce any genome assembly data!

Ensembl project

- Available data :
 - Genomes
 - Comparative genomic data
 - Variations
 - Gene regulatory elements
 - External Annotations



- Website is launch in 2000 (at the beginning, only the human genome)

Ensembl Genomes

- Vertebrate species + some representative species -> <http://ensembl.org>
- EnsemblGenomes (since april 2009)
 - Metazoa: <http://metazoa.ensembl.org>
 - Bacteria: <http://bacteria.ensembl.org>
 - Plants: <http://plants.ensembl.org>
 - Fungi : <http://fungi.ensembl.org>
 - Protists : <http://protists.ensembl.org>

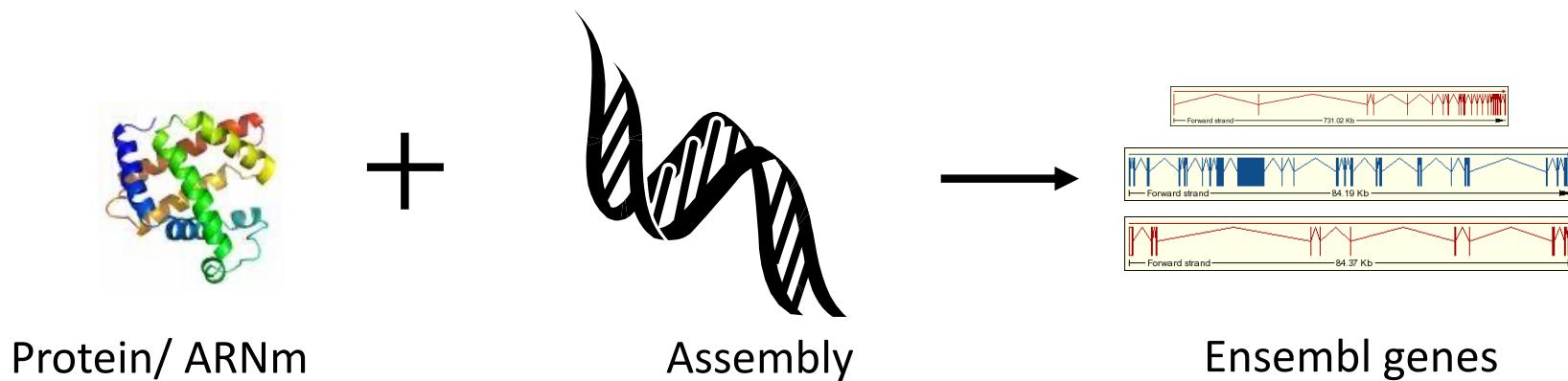
<http://ensemblgenomes.org/info/genomes>

Genome annotation

- Takes 3 to 6 months
- Gene annotation provided by Ensembl
 - Automatic annotation (Ensembl Genebuild) :
 - Genome-wide determination of transcripts
 - Based on mRNAs and protein sequences deposited into public databases

Genome annotation

- Ensembl transcripts are based on mRNA and protein from the following databases:
 - Uniprot/Swiss-Prot (manual curation)
 - Uniprot/TrEMBL
 - NCBI refSeq (manual curation)



Genome annotation

- Takes 3 to 6 months
- Gene annotation provided by Ensembl
 - Automatic annotation (Ensembl Genebuild) :
 - Genome-wide determination of transcripts
 - Based on mRNAs and protein sequences deposited into public databases
 - Manual curation : on a case by case basis. Only for Human, mouse, rat, Zebrafish + some other vertebrates (HAVANA/VEGA at the WTSI)
 - Annotations are also imported from FlyBase, WormBase, SGD (Saccharomyces)

GENCODE and CCDS

- GENCODE project
 - Sub-project of ENCODE (ENCyclopedia Of DNA Elements)
 - Aim : annotate all evidence-based gene features (genes, transcripts, coding sequences, ...) in the entire human and mouse genomes at a high accuracy
 - The default human and mouse gene sets in Ensembl is also the current version of GENCODE
- Consensus Coding Sequence
 - Collaborative effort between Ensembl, NCBI, UCSC, HAVANA and HGNC for human or MGI for mouse
 - Aim : identify a core set of human and mouse protein coding regions that are consistently annotated and of high quality

Ensembl and HAVANA merge

- For species with both gene annotation : merge of the 2 sets of gene models
- Where manual annotation is available for a transcript
 - Ensembl and HAVANA transcript models are merged when their splicing structure is identical (same internal exon-intron boundaries)
 - If the ends differ, the HAVANA annotated ends are used
- Updated manual annotation from Havana is merged into the Ensembl annotation every release
- For human and mouse, this combined Ensembl/HAVANA gene set is the gene set from the Gencode project
- For human and mouse, all transcripts from the CCDS* set are present in the Gencode gene set

Annotation of Non coding RNAs

- Non-Coding RNA gene types are annotated:
 - tRNA (transfer RNA)
 - Mt-rRNA (transfer RNA located in the mitochondrial genome)
 - rRNA (ribosomal RNA)
 - scRNA (small cytoplasmic RNA)
 - snRNA (mall nuclear RNA)
 - snoRNA (small nucleolar RNA)
 - miRNA (microRNA precursors)
 - misc_RNA (miscellaneous other RNA)
 - lincRNA (Long intergenic non-coding RNAs)

Annotation of Non coding RNAs

- **ncRNAs** are annotated by aligning genomic sequence against RFAM using BLASTN.
- **miRNAs** are predicted by BLASTN of genomic sequence slices against miRBase sequences.
- **tRNAs** are annotated as part of the raw compute process using tRNAscan-SE.
- **lincRNA** are annotated using Ensembl gene annotation, cDNA alignments and chromatin-state map data (H3K4me3 and H3K36me3) from the Ensembl regulatory build.

Ensembl identifiers

- Aim to be consistent across Ensembl releases, unlike gene names
- Format
 - ENSG### Ensembl Gene ID
 - ENST### Ensembl Transcript ID
 - ENSP### Ensembl Peptide ID
 - ENSE### Ensembl Exon ID
- Addition of a suffix for the other species
 - MUS (*Mus musculus*) for mouse: ENSMUSG###
 - DAR (*Danio rerio*) for zebrafish: ENSDARG###
 - etc.

Ensembl genome browser

<https://www.ensembl.org/index.html>

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

Tools [BioMart >](#) [BLAST/BLAT >](#) [Variant Effect Predictor >](#)

[All tools](#) Export custom datasets from Ensembl with this data-mining tool Search our genomes for your DNA or protein sequence 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 **Favourite genomes**  Human GRCh38.p13 [Still using GRCh37?](#)  Mouse GRCm39  Zebrafish GRCz11 [View full list of all species](#)

Ensembl Release 109 (Feb 2023)

- New gene sets for donkey and horse
- Updated SIFT and PolyPhen-2 missense variant pathogenicity
- New VEP plugins for UTR annotation
- New ATAC-seq tracks (peaks and signal) for fish species (Atlantic Salmon, European Seabass, Rainbow Trout and Turbot)

[More release news](#) on our blog

Ensembl Rapid Release

New assemblies with gene and protein annotation every two weeks.
Note: species that already exist on this site will continue to be updated with the full range of annotations.

The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-vertebrate genomes from biodiversity initiatives such as Darwin Tree of Life, the Vertebrate Genomes Project and the Earth BioGenome Project.

[Rapid Release news](#) on our blog

Compare genes across species 
Find SNPs and other variants for my gene 
Gene expression in different tissues 
Retrieve gene sequence 
Find a Data Display 
Use my own data in Ensembl 

EMBL-EBI  Ensembl creates, integrates and distributes reference datasets and analysis tools that enable genomics. We are based at [EMBL-EBI](#) and our software and data are freely available. Our [acknowledgements](#) page includes a list of current and previous funding bodies. [How to cite Ensembl](#) in your own publications.

GLOBAL CORE DATA RESOURCE 

Ensembl release 109 - Feb 2023 © EMBL-EBI [Permanent link](#) - [View in archive site](#)

<http://ensemblgenomes.org>

e! EnsemblGenomes Providing genome data for non-vertebrate species, with tools for the manipulation, analysis and visualisation of that data [Contact us](#)

 [Latest release notes, updates & news from our blog](#)

Ensembl COVID-19

 [SARS-CoV-2 Genome sequence & annotation data](#) [Go](#)

Ensembl Rapid Release

2-weekly releases of new assemblies with gene & protein feature annotation [Go](#)

EnsemblPlants

 [Triticum aestivum](#)
IWGSC

 [Oryza sativa Japonica Group](#)
IRGSP-1.0

 [Arabidopsis thaliana](#)
TAIR10

[Go to Ensembl Plants](#)

EnsemblMetazoa

 [Caenorhabditis elegans](#)
WBcel235

 [Drosophila melanogaster](#)
BDGP6.28

 [Bombyx mori](#)
ASM15162v1

[Go to Ensembl Metazoa](#)

EnsemblProtists

 [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](#)

Release

- ~ every 3 month
- Link to the latest version of Ensembl: <http://www.ensembl.org>



Ensembl creates, integrates and distributes reference datasets and analysis tools that enable genomics. We are based at [EMBL-EBI](#) and our software and data are freely available.

Our [acknowledgements page](#) includes a list of current and previous funding bodies. [How to cite Ensembl](#) in your own publications.



Ensembl release 109 - Feb 2023 © [EMBL-EBI](#)

[Permanent link](#) [View in archive site](#)

- Permanent link to a given release of Ensembl:
<http://Oct2024.archive.ensembl.org/index.html>

Ensembl Rapid Release

 **Ensembl**
Rapid Release

BLAST | Tools | Downloads | Help & Docs | Known Bugs | Blog

Tools [BLAST >](#)

[All tools](#) Search our genomes for your DNA or protein sequence

Search

Homo sapiens (Human) - GCA_ for [Go](#)

e.g. [Camarhynchus parvulus](#) 2:361680-384534 or [Clytia hemisphaerica](#) IPR001650

Ensembl Rapid Release is a new site designed to make our data available more quickly. Release of data occurs on a two-week cycle, meaning we can make our gene sets available with minimal delay once the annotation is complete. For each species we provide a gene set along with additional features such as protein feature annotation and BLAST functionality.

It is important to note that Ensembl Rapid Release is by nature not as fully featured as a typical data release on [www.ensembl.org](#). Currently we do not provide data archiving or programmatic access. We are gradually rolling out comparative analyses across all species, and are working on adding more functionality over the coming months to further improve usability.

[More details](#) about Ensembl Rapid Release and the current and planned features.

All genomes

-- Select a species --

[View and download available data for all species](#)

Highlighted genomes

 **Homo sapiens** (Human)
T2T-CHM13v2.0

 **Camarhynchus parvulus**
GCA_902806625.1
Camarhynchus_parvulus

Latest Genomes

We have 6 new genomes this release:

- [Bernisia tabaci](#) (Silverleaf whitefly) - GCA_918797505.1 [Community annotation]
- [Fragrum fragum](#) (Bivalves) - GCA_946902895.1
- [Orius laevigatus](#) (Minute pirate bug) - GCA_018703685.1
- [Spodoptera exigua](#) (Beet armyworm) - GCA_902829305.1
- [Tridacna crocea](#) (bivalves) - GCA_943736015.1
- [Tridacna gigas](#) (Giant clam) - GCA_945859785.2

[View all species and download data](#)

Comparative analyses

- All species now have homologues with an appropriate set of reference species, allowing us to assign gene symbols where possible
- We are adding [HAL multiple alignments](#) for selected clades, available to download from our FTP site

Variation

[Variation data](#) is displayed for the human pangenome, [Drosophila melanogaster](#) (GCA_000001215.4), and [Cajanus cajan](#) (GCA_000340665.1) assemblies

- Fast access to annotation data (2-4 weeks)
- Less features than on [www.ensembl.org](#) (focuses on gene sets)
- No programmatic access

Ensembl: Archives

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

[Login/Register](#)



Using this website [Annotation and prediction](#) [Data access](#) [API & software](#) [About us](#)

In this section [Help & Documentation](#) > [Using this website](#) > Archives

[Archives: Table of assemblies](#)

[Search documentation](#) 

Ensembl Archives

About Archive Ensembl

The main Ensembl site (www.ensembl.org) and the mirror sites are updated with the latest data approximately every three months. We maintain the Ensembl Archive sites so that there are stable links to data from a particular release. As of December 2016 these will be available for five years, together with the following longer term archives:

- Annotation on the **human NCBI36 assembly** is available at our [Ensembl 54 archive](#) site.
- Annotation on the **mouse NCBIm37 assembly** is available at our [Ensembl 67 archive](#) site.
- As from August 2014 we are supporting the **human GRCh37 assembly** at our dedicated [GRCh37 human](#) site. Unlike the other Ensembl archive sites, this will be updated to the latest web interface every Ensembl release and there may be occasional data updates to human.

Archived databases are also maintained for at least 10 years. More information is available from our [MySQL database documentation](#). We also maintain data archives from 2004 available from our [FTP site](#).

For all enquiries, please [contact the Ensembl HelpDesk](#).

Notes

- Ensembl aims to maintain stable identifiers for genes (ENSG), transcripts (ENST), proteins (ENSP) and exons (ENSE) as long as possible. Changes within the genome sequence assembly or an updated genome annotation may dramatically change a gene model. In these cases, the old set of stable IDs is retired and a new one assigned. Gene and transcript pages both have an ID History view which maps changes in the ID from the earliest version in Ensembl.
- Protein family identifiers (fam), Ensembl EST gene identifiers (ENSESTG) and Genscan identifiers (GENSCAN) are currently not stable.
- With the exception of the GRCh37 human site **BLAST**, **BLAT** and **other tools** are not available from the archive sites.
- Accounts** are shared between the current site and almost all archives. The exceptions are the older human NCBI36 and the mouse GRCh37 sites where changes in architecture and code make sharing logins impractical.

List of currently available archives

- [Ensembl GRCh37](#) - Full Feb 2014 archive with BLAST, VEP and BioMart
- [Ensembl 109: Feb 2023](#) - this site
- [Ensembl 108: Oct 2022](#)
- [Ensembl 107: Jul 2022](#)
- [Ensembl 106: Apr 2022](#)
- [Ensembl 105: Dec 2021](#)
- [Ensembl 104: May 2021](#)
- [Ensembl 103: Feb 2021](#)
- [Ensembl 102: Nov 2020](#)
- [Ensembl 101: Aug 2020](#)
- [Ensembl 100: Apr 2020](#)
- [Ensembl 99: Jan 2020](#)
- [Ensembl 98: Sep 2019](#)
- [Ensembl 97: Jul 2019](#)
- [Ensembl 96: Apr 2019](#)
- [Ensembl 95: Jan 2019](#)
- [Ensembl 94: Oct 2018](#)
- [Ensembl 93: Jul 2018](#)
- [Ensembl 92: Apr 2018](#)
- [Ensembl 80: May 2015](#)
- [Ensembl 77: Oct 2014](#)
- [Ensembl 75: Feb 2014](#)
- [Ensembl 54: May 2009](#)

[Table of archives showing assemblies present in each one.](#)

[Linking to the Archive Ensembl sites](#)

<http://www.ensembl.org/info/website/archives/index.html>

Ensembl: Archives

Archive! Ensembl BioMart | Tools | Downloads | Help & Documentation | Blog [Login/Register](#)

Search: for

e.g. [BRCA2](#) or [rat X:100000..200000](#) or [coronary heart disease](#)

Browse a Genome
The Ensembl project produces genome databases for vertebrates and other eukaryotic species, and makes this information freely available online.

Popular genomes

Human GRCh37	Mouse GRCm38
Zebrafish Zv9	

[★ Log in to customize this list](#)

All genomes

-- Select a species --

[View full list of all Archive EnsEMBL species](#)

Other species are available in [Ensembl Pre!](#) and [EnsemblGenomes](#)

ENCODE data in Ensembl

Variant Effect Predictor

Gene expression in different tissues

Find SNPs and other variants for my gene

Retrieve gene sequence
`GGCTGACTTCGGTGC
GGGCTTGCGCGAGGC
GGGCCCTGCTCGCGCCCT
AGGGGACAGATTGTGA
CACCTCTGGAGCGGGTT
CCCAGTCCAGCGTGGCG`

Compare genes across species

Use my own data in Ensembl

Learn about a disease or phenotype

Looking for BLAST, BLAT, VEP or Assembly Converter?
Visit our [dedicated GRCh37 archive](#) to enjoy the latest web features on release 75 databases.

What's New in Release 75 (February 2014)

- [New VEP interface](#)
- [New 'Age of base' track for human](#)
- [New GENCODE basic renderer for human and mouse](#)

[Full details of this release](#)

[More release news on our blog →](#)

Latest blog posts

- 10 Mar 2016: [Ensembl 84 has been released!](#)
- 16 Feb 2016: [Learn about Ensembl – online, live and free](#)
- 25 Jan 2016: [Sharing feature on the new mobile site \(\[m.ensembl.org\]\(#\)\)](#)

[Go to Ensembl blog →](#)

Did you know...?

Major genomics projects, Quantomics, HEROIC and out more on [our projects](#)

Old Ensembl releases are kept for 5 years unless they include the last release of an assembly for a key genome.

Ensembl: Archives

<http://www.ensembl.org/info/website/archives/assembly.html>

The screenshot shows a grid-based interface for viewing genome assembly versions across different species. The columns represent assembly versions from Oct 2022 (v108) down to Sep 2015 (v82). The rows list various species, each with a corresponding assembly identifier. A legend at the top indicates three categories: New species (yellow), Species present in archive (light blue), and Species not in this version of Ensembl (white).

	Oct 2022 v108	Jul 2022 v107	Apr 2022 v106	Dec 2021 v105	May 2021 v104	Feb 2021 v103	Nov 2020 v102	Aug 2020 v101	Apr 2020 v100	Jan 2020 v99	Sep 2019 v98	Jul 2019 v97	Apr 2019 v96	Jan 2019 v95	Oct 2018 v94	Jul 2018 v93	Apr 2018 v92	Dec 2017 v91	Aug 2017 v90	May 2017 v89	Mar 2017 v88	Dec 2016 v87	Oct 2016 v86	Jul 2016 v85	Mar 2016 v84	Dec 2015 v83	Sep 2015 v82	
Abingdon island giant tortoise	ASM359739v1																											
African ostrich	ASM69896v1																											
Agassiz's desert tortoise	ASM289941v1																											
Algerian mouse	SPRET_EiU_v1																											
Alpaca	vicPac1																											
Alpine marmot	marMar2.1																											
Amazon molly	Poecilia_formosa-5.1.2																											
American beaver	C.can_genome_v1.0																											
American bison	Bison_UMD1.0																											
American black bear	ASM34442v1																											
American mink	NNGGG.v01																											
Angola colobus	Cang.pa_1.0																											
Arabian camel	CamDro2																											
Arctic ground squirrel	ASM342692v1																											
Argentine black and white tegu	HLtpMer3																											
	Oct 2022 v108	Jul 2022 v107	Apr 2022 v106	Dec 2021 v105	May 2021 v104	Feb 2021 v103	Nov 2020 v102	Aug 2020 v101	Apr 2020 v100	Jan 2020 v99	Sep 2019 v98	Jul 2019 v97	Apr 2019 v96	Jan 2019 v95	Oct 2018 v94	Jul 2018 v93	Apr 2018 v92	Dec 2017 v91	Aug 2017 v90	May 2017 v89	Mar 2017 v88	Dec 2016 v87	Oct 2016 v86	Jul 2016 v85	Mar 2016 v84	Dec 2015 v83	Sep 2015 v82	
Armadillo	Dasnov3.0																											
Asian bonytongue	fSciFor1.1														ASM162426v1													
Asiatic black bear	ASM966005v1																											
Atlantic cod	gadMor3.0													gadMor1														
Atlantic herring	Ch_v2.0.2																											
Atlantic salmon	Ssal_v3.1													ICSASG_v2														
Australian saltwater crocodile	CroPor_comp1																											
Ballan wrasse	BallGen_V1																											
Barramundi perch	ASB_HGAPassembly_v1																											
Beluga whale	ASM228892v3																											
Bengalese finch	LonStrDom1																											
Bicolor damselfish	Stegastes_partitus-1.0.2																											

Browse through genomes with
Ensembl

Ensembl home page

Link back to home page

Tools

Search



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

Login/Register

Search all species...

Tools

BioMart >

[All tools](#)

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

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 109 (Feb 2023)

- New gene sets for donkey and horse
- Updated SIFT and PolyPhen-2 missense variant pathogenicity
- New VEP plugins for UTR annotation
- New ATAC-seq tracks (peaks and signal) for fish species (Atlantic Seabass, Rainbow Trout and Turbot)

News

[More release](#)

Search

Search

All species for Go

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

All genomes

-- Select a species --



Pig breeds

Pig reference genome

[View full list of all species](#)

Favourite genomes



Human

GRCh38.p13



Mouse

GRCm39



Zebrafish

GRCz11

Ensembl Rapid Release

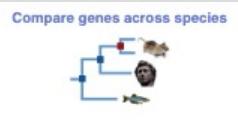
New assemblies with gene and protein annotation every two weeks.

Note: species that already exist on this site will continue to be updated with the full range of annotations.

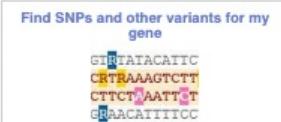
Go

The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-vertebrate genomes from biodiversity initiatives such as Darwin Tree of Life, the Vertebrate Genomes Project and the Earth BioGenome Project.

[Rapid Release news](#) on our blog

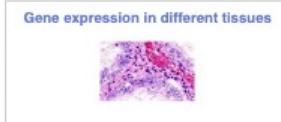


Compare genes across species



Find SNPs and other variants for my gene

GTGATAACATTC
CTTAAAGTCTT
CTTCATTATTT
GTAACATTTCC



Gene expression in different tissues



Retrieve gene sequence

GGCTGACTTCGGGTTGG
GGGGCTTGCGGGGGGGGGGG
GGGGCTCTCTGCTGGGGGG
AAGGGGGGGGGGGGGGGGG
CAGCTCTGGGGGGGGGGGG
CCCCGGGGGGGGGGGGGGGG

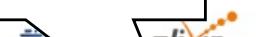


Access to Ensembl Archives



Ensembl creates, integrates and distributes reference datasets and analysis tools that enable genomics. We are based at [EMBL-EBI](#) and our software and data are freely available.

Our [acknowledgements page](#) includes a list of current and previous funding bodies. [How to cite Ensembl](#) in your own publications.



Core Data Resource

Search

e!Ensembl [Login/Register](#)

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

Search all species... 

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

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

Search: gene name, region, disease...

-- Select a species --

Pig breeds Pig reference genome and 12 additional breeds 

[View full list of all species](#)

Human GRCh38.p13 
[Still using GRCh37?](#)

Mouse GRCm39 

Zebrafish GRCz11 

Compare genes across species 

Find SNPs and other variants for my gene

GTTAATACATTC
C T RAAGTCTT
CTTC T AATT T
GTAACTTTC

Gene expression in different tissues 

Retrieve gene sequence

```
GGCTGACTTCGCGGTTGG  
GGAGCTTGAGGGGGGGGGGG  
GGCGCTCTGCTGCTGCGCT  
AAGGGAGCAATTGGTGTG  
CAGCTCTGGAGCGCGGTG  
CCCAAGTCCAGCGTGGCGG
```

Find a Data Display 

Use my own data in Ensembl 

Ensembl creates, integrates and distributes reference datasets and analysis tools that enable genomics. We are based at [EMBL-EBI](#) and our software and data are freely available. Our [acknowledgements page](#) includes a list of current and previous funding bodies. [How to cite Ensembl](#) in your own publications.



Ensembl creates, integrates and distributes reference datasets and analysis tools that enable genomics. We are based at [EMBL-EBI](#) and our software and data are freely available. Our [acknowledgements page](#) includes a list of current and previous funding bodies. [How to cite Ensembl](#) in your own publications.



Genome page

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

Human (GRCh38.p13) ▾

Search Human (Homo sapiens)

Search all categories ▾ Search... **Search**

e.g. BRCA2 or 17:63992802-64038237 or rs699 or osteoarthritis

Information, statistics

Genome assembly: GRCh38.p13 (GCA_000001405.28)

- More information and statistics
- Download DNA sequence (FASTA)
- Convert your data to GRCh38 coordinates
- Display your data in Ensembl

Other assemblies

GRCh37 Full Feb 2014 archive with BLAST, VEP and BioMart ▾ Go

Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs

- More about this genebuild
- Download FASTA files for genes, cDNAs, ncRNA, proteins
- Download GTF or GFF3 files for genes, cDNAs, ncRNA, proteins
- Update your old Ensembl IDs

Link to examples

Pax6 INS FUSP2 BRCA2 DMD ssh Example gene

Example transcript

Comparative genomics

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

- More about comparative analysis
- Download alignments (EMF)

Regulation

What can I find? DNA methylation, transcription factor binding sites, histone modifications, and regulatory features such as enhancers and repressors, and microarray annotations.

- More about the Ensembl regulatory build and microarray annotation
- Experimental data sources
- Download all regulatory features (GFF)

Variation

What can I find? Short sequence variants and longer structural variants; disease and other phenotypes

- More about variation in Ensembl
- Download all variants (GVF)
- Variant Effect Predictor

Ve!P

ATCGAGCT ATCCAGCT ATCGAGAT Example variant

Example phenotype

Example structural variant

Genome statistics

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

Human (GRCh38.p13) ▾

Human assembly and gene annotation

Assembly

This site provides a data set based on the December 2013 *Homo sapiens* high coverage assembly GRCh38 from the [Genome Reference Consortium](#). This assembly was used by UCSC to create their hg38 database. The data set consists of gene models built from the genewise alignments of the human proteome as well as from alignments of human cDNAs using the cDNA2genome model of exonerate.

This release of the assembly has the following properties:

- contig length total 3.4 Gb.
- chromosome length total 3.1 Gb (excluding haplotypes).

It also includes 261 alt loci scaffolds, mainly in the LRC/KIR complex region on chromosome 6 (7 alternate sequence representations) and the MHC region on chromosome 6 (7 alternate sequence representation).

Watch a video on YouTube about patches and haplotypes in the Human genome.

Patches

As the GRC maintains and improves the assembly, patches are being introduced. Currently, assembly patches are of two types:

- Novel patch: new sequences that add alternative sequence at a loci and will remain as haplotypes in the next major assembly release by GRC.
- Fix patch: sequences that correct the reference sequence and will replace the given region of the reference assembly at the next major assembly release by GRC.

Other assemblies

GRCh37 Full Feb 2014 archive with BLAST, VEP and BioMart ▾ Go

Gene annotation

The Ensembl human gene annotations have been updated using Ensembl's automatic annotation pipeline. The updated annotation incorporates new protein and cDNA sequences which have become publicly available since the last GRCh38 genebuild (December 2013).

In the current release, we continue to display a joint gene set based on the merge between the automatic annotation from Ensembl and the manually curated annotation from Havana. See the statistics table, right, for the corresponding GENCODE version number. The Consensus Coding Sequence (CCDS) identifiers have also been mapped to the annotations. More information about the [CCDS project](#).

Updated manual annotation from Havana is merged into the Ensembl annotation every release. Transcripts from the two annotation sources are merged if they share the same internal exon-intron boundaries (i.e. have identical splicing pattern) with slight differences in the terminal exons allowed. Importantly, all Havana transcripts are included in the final Ensembl/Havana merged (GENCODE) gene set.

- [Detailed information on genebuild \(PDF\)](#)

Neanderthal genome

A preliminary assembly of the Neanderthal (*Homo sapiens neanderthalensis*) genome is available via the [Neanderthal Genome Browser](#), an Ensembl-powered project based at the Max Planck Institute.

More information

General information about this species can be found in [Wikipedia](#).

Search all species... ▾

Login/Register

Information

Statistics

Statistiques

Assembly	GRCh38.p13 (Genome Reference Consortium Human Build 38), INSDC Assembly GCA_000001405.28, Dec 2013
Base Pairs	3,096,649,726
Golden Path Length	3,096,649,726
Assembly provider	Genome Reference Consortium
Annotation provider	Ensembl
Annotation method	Full genebuild
Genebuild started	Jan 2014
Genebuild released	Jul 2014
Genebuild last updated/patched	Nov 2022
Database version	109.38
Gencode version	GENCODE 43

Gene counts (Primary assembly)	
Coding genes	19,827 (excl 649 readthrough)
Non coding genes	25,967
Small non coding genes	4,864
Long non coding genes	18,882
Misc non coding genes	2,221
Pseudogenes	15,241
Gene transcripts	252,974

Gene counts (Alternative sequence)	
Coding genes	3,028 (excl 26 readthrough)
Non coding genes	1,682
Small non coding genes	297
Long non coding genes	1,198
Misc non coding genes	187
Pseudogenes	1,796
Gene transcripts	21,630

Other	
Genscan gene predictions	51,756
Short Variants	715,081,111
Structural variants	7,549,196

Genome page

e!Ensembl [BLAST/BLAT](#) | [VEP](#) | [Tools](#) | [BioMart](#) | [Downloads](#) | [Help & Docs](#) | [Blog](#) [Login/Register](#)

 **Human (GRCh38.p13)** ▾ 

Search Human (Homo sapiens)

Search all categories ▾ Search...

e.g. [BRCA2](#) or [17:63992802-64038237](#) or [rs699](#) or [osteoarthritis](#)

Genome assembly: GRCh38.p13 (GCA_000001405.28)

 [More information and statistics](#)

 [Download DNA sequence \(FASTA\)](#)

 [Convert your data to GRCh38 coordinates](#)

 [Display your data in Ensembl](#)

Other assemblies

[GRCh37 Full Feb 2014 archive with BLAST, VEP and BioMart](#)

Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

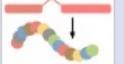
 [More about this genebuild](#)

 [Download FASTA files for genes, cDNAs, ncRNA, proteins](#)

 [Download GTF or GFF3 files for genes, cDNAs, ncRNA, proteins](#)

 [Update your old Ensembl IDs](#)


Pax6 **INS**
Foxp2 **BRCA2**
DMD **ssh**
Example gene


Example transcript

Comparative genomics

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

 [More about comparative analysis](#)

 [Download alignments \(EMF\)](#)

Regulation

What can I find? DNA methylation, transcription factor binding sites, histone modifications, and regulatory features such as enhancers and repressors, and microarray annotations.

 [More about the Ensembl regulatory build and microarray annotation](#)

 [Experimental data sources](#)

 [Download all regulatory features \(GFF\)](#)

Example gene tree 

Example regulatory feature 

Variation

What can I find? Short sequence variants and longer structural variants; disease and other phenotypes

 [More about variation in Ensembl](#)

 [Download all variants \(GVF\)](#)

 [Variant Effect Predictor](#) 


ATCGAGCT
ATCCAGCT
ATCGAGAT
Example variant


Example phenotype


Example structural variant

Karyotype

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

Human (GRCh38.p13) ▾

Login/Register

Search all species...

Genome Jobs

Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- Region in detail

Comparative Genomics

- Synteny
- Alignments (image)
- Alignments (text)
- Region Comparison

Genetic Variation

- Variant table
- Resequencing
- Strain table
- Linkage Data

Markers

Other genome browsers

- UCSC
- NCBI
- Ensembl GRCh37

Add features

Add/remove tracks | Custom tracks | Share | Export Image | Reset configuration

Click on the image above to jump to a chromosome, or click and drag to select a region

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

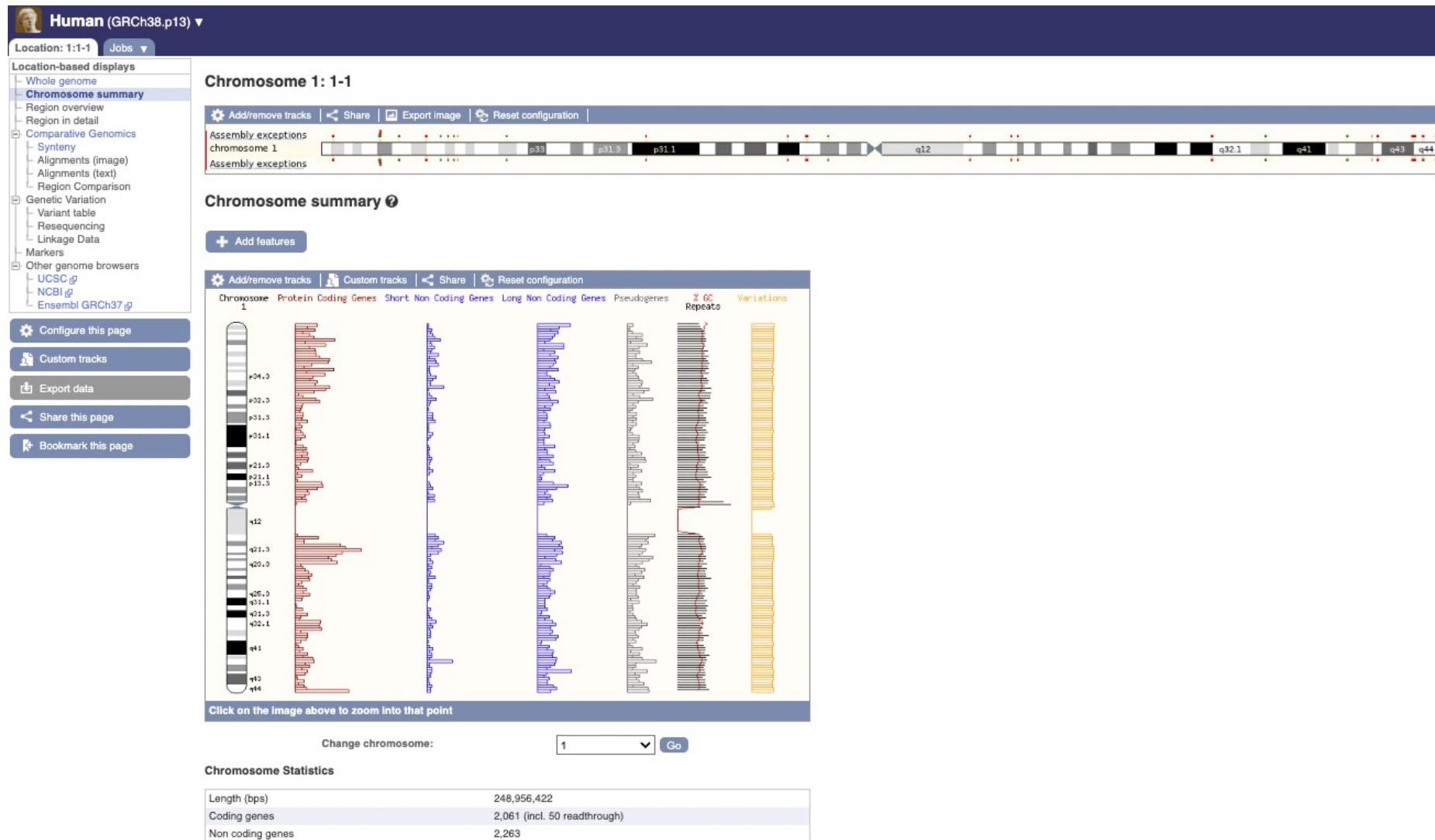
Summary

Assembly	GRCh38.p13 (Genome Reference Consortium Human Build 38), INSDC Assembly GCA_000001405.28 , Dec 2013
Base Pairs	3,096,649,726
Golden Path Length	3,096,649,726
Assembly provider	Genome Reference Consortium
Annotation provider	Ensembl
Annotation method	Full genebuild
Genebuild started	Jan 2014
Genebuild released	Jul 2014
Genebuild last updated/patched	Nov 2022
Database version	109.38
Gencode version	GENCODE 43

Gene counts (Primary assembly)

Coding genes	19,827 (excl 649 readthrough)
Non coding genes	25,967
Small non coding genes	4,864
Long non coding genes	18,882
Misc non coding genes	2,221

Statistics per chromosome



Genome browser view

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

Login/Register

Human (GRCh38.p12) ▾

Location: 17:63,992,802-64,038,237

Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- Region in detail

Comparative Genomics

- Synteny
- Alignments (image)
- Alignments (text)
- Region Comparison

Genetic Variation

- Variant table
- Rerescoring
- Linkage Data
- Markers

Other genome browsers

- UCSC
- NCBI
- Ensembl GRCh37

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Chromosome 17: 63,992,802-64,038,237

Assembly exceptions: Chr. 17 p13.3 p13.2 p13.1 p12 p11.2 q11.2 q12 q21.2 q21.31 q21.32 q21.33 q22 q23.2 q24.2 q24.3 q25.1 q25.3

Region in detail

Forward strand

Chromosome bands

Contigs

Genes (Comprehensive set from GENCODE 29)

Regulatory Build

Gene Legend

Regulation Legend

Location: 17:63992802-64038237 Go Gene: Go

Drag/Select: ↪

Chromosome bands.

75 way GERP elements

Human cDNAs (RefSeq/ENAS). CCDS.set

Constrained elements for 75 eutherian mammals EPO-Low-Coverage

CCDS54158.1 > protein coding

CCDS54159.1 > protein coding

CCDS54157.1 > protein coding

45.44 kb

64.00Mb 64.01Mb 64.02Mb 64.03Mb

q23.3

Forward strand

Gene view

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

Human (GRCh38.p12) ▾

Location: 13:315,474-32,400,266

Gene-based displays

- Summary
 - Splice variants
 - Transcript comparison
 - Gene alleles
- Sequence
 - Secondary Structure
- Comparative Genomics
 - Genomic alignments
 - Gene tree
 - Gene gain/loss tree
 - Orthologues
 - Paralogues
- Ensembl protein families
- Ontologies
 - GO: Biological process
 - GO: Cellular component
 - GO: Molecular function
- Phenotypes
- Genetic Variation
 - Variant table
 - Variant image
 - Structural variants
- Gene expression
- Pathway
- Regulation
- External references
- Supporting evidence
- ID History
- Gene history

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Gene: BRCA2 ENSG00000139618

Description: BRCA2, DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]

Gene Synonyms: BRCC2, FACD, FAD, FAD1, FANCD, FANCD1, XRCC11

Location: Chromosome 13: 32,315,474-32,400,266 forward strand. GRCh38:CM000675.2

About this gene: This gene has 7 transcripts (splice variants), 132 orthologues, is a member of 1 Ensembl protein family and is associated with 128 phenotypes.

Transcripts: Hide transcript table

Name	Transcript ID	bp	Protein	Biotype	CCDS	UniProt	RefSeq	Flags
BRCA2-201	ENST00000380152.7	11986	3418aa	Protein coding	CCDS9344.1	P51587	-	TSL:5 GENCODE basic APPRIS P1
BRCA2-206	ENST00000544455.5	10984	3418aa	Protein coding	CCDS9344.1	P51587	NM_000059 NP_000050	TSL:1 GENCODE basic APPRIS P1
BRCA2-202	ENST00000470094.1	842	186aa	Nonsense mediated decay	-	HOYE37	-	CDS 5' incomplete TSL:5
BRCA2-203	ENST00000528762.1	495	64aa	Nonsense mediated decay	-	HOYD86	-	CDS 5' incomplete TSL:4
BRCA2-207	ENST00000614259.1	7950	No protein	Processed transcript	-	-	-	TSL:2
BRCA2-204	ENST00000530893.6	2011	No protein	Processed transcript	-	-	-	TSL:1
BRCA2-205	ENST00000533776.1	523	No protein	Retained intron	-	-	-	TSL:3

Summary: BRCA2 (HGNC Symbol)

CCDS: This gene is a member of the Human CCDS set: CCDS9344.1

UniProtKB: This gene has proteins that correspond to the following UniProtKB identifiers: P51587

RefSeq: Overlapping RefSeq Gene ID 675 matches and has similar biotype of protein_coding

LRG: LRG_293 provides a stable genomic reference framework for describing sequence variants for this gene

Ensembl version: ENSG00000139618.14

Other assemblies: This gene maps to 32,889,611-32,974,403 in GRCh37 coordinates. View this locus in the GRCh37 archive: ENSG00000139618

Gene type: Protein coding

Annotation method: Annotation for this gene includes both automatic annotation from Ensembl and Havana manual curation, see article.

Annotation Attributes: overlapping locus [Definitions]

Go to Region in Detail for more tracks and navigation options (e.g. zooming)

Drag>Select: ↪

Forward strand

Genes (Comprehensive set...)

32.31Mb 32.32Mb 32.33Mb 32.34Mb 32.35Mb 104.79 kb 32.36Mb 32.37Mb 32.38Mb 32.39Mb 32.40Mb 32.41Mb



Transcript view

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

Human (GRCh38.p12) ▾

Location: 13:32,315,474-32,400,266 Gene: BRCA2 Transcript: BRCA2-201

Transcript-based displays

- Summary
- Sequence
 - Exons
 - cDNA
 - Protein
- Protein Information
 - Protein summary
 - Domains & features
 - Variants
 - 3D Protein model
- Genetic Variation
 - Variant table
 - Variant image
 - Haplotypes
 - Population comparison
 - Comparison image
- External References
 - General identifiers
 - Oligo probes
 - Supporting evidence
- ID History
 - Transcript history
 - Protein history

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Transcript: BRCA2-201 ENST00000380152.7

Description: BRCA2, DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]. Gene Synonyms: BRCC2, FACD, FAD1, FANCD, FANCD1, XRCC11. Location: Chromosome 13: 32,315,474-32,400,266 forward strand. About this transcript: This transcript has 27 exons, is annotated with 51 domains and features, is associated with 29608 variations and maps to 1004 oligo probes. Gene: This transcript is a product of gene ENSG00000139618. Hide transcript table

Show/hide columns (1 hidden)

Name	Transcript ID	bp	Protein	Biotype	CCDS	UniProt	RefSeq	Flags
BRCA2-201	ENST00000380152.7	11986	3418aa	Protein coding	CCDS9344	P51587	-	TSL:5 GENCODE basic APPRIS P1
BRCA2-206	ENST00000544455.5	10984	3418aa	Protein coding	CCDS9344	P51587	NM_000059 NP_000050	TSL:1 GENCODE basic APPRIS P1
BRCA2-202	ENST00000470094.1	842	186aa	Nonsense mediated decay	-	HOYE37	-	CDS 5' incomplete TSL:5
BRCA2-203	ENST00000528762.1	495	64aa	Nonsense mediated decay	-	HOYD86	-	CDS 5' incomplete TSL:4
BRCA2-207	ENST00000614259.1	7950	No protein	Processed transcript	-	-	-	TSL:2
BRCA2-204	ENST00000530893.6	2011	No protein	Processed transcript	-	-	-	TSL:1
BRCA2-205	ENST00000533776.1	523	No protein	Retained intron	-	-	-	TSL:3

Summary

Statistics: Exons: 27, Coding exons: 26, Transcript length: 11,986 bps, Translation length: 3,418 residues. CCDS: This transcript is a member of the Human CCDS set: CCDS9344. UniProt: This transcript corresponds to the following UniProt identifiers: P51587. Transcript Support Level (TSL): TSL:5. Version: ENST00000380152.7. Type: Protein coding. Annotation Method: Transcript where the Ensembl genebuild transcript and the Vega manual annotation have the same sequence, for every base pair. See article. GENCODE basic gene: This transcript is a member of the Gencode basic gene set.

Browse Ensembl (part 1)

hands-on

Ensembl tools

Tools

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

BLAST/BLAT Search datasets from Ensembl with this BLAST/BLAT 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
e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

All genomes -- Select a species --
Pig breeds Pig reference genome and 12 additional breeds
[View full list of all species](#)

Favourite genomes Human GRCh38.p13
Still using GRCh37?
Mouse GRCm39
Zebrafish GRCz11

Compare genes across species 
Find SNPs and other variants for my gene
GTTAATACATTG
CTTAAAGTCCTT
CTTCTTAATTGT
GGAACATTTCC
Gene expression in different tissues 
Retrieve gene sequence
GCCTGACTTCGGGTTGGC
GGGGTTGGGGGGGGGGGGGG
GGGGCTCTGCTGGGGGGGG
AAGGGGAGGAATTGGGGGG
CAGCTCTGGGGGGGGGGGG
CCCCGGGGGGGGGGGGGGGG
Find a Data Display TABLE HEATMAP SEQUENCE PIE CHART
Use my own data in Ensembl 

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 109 (Feb 2023)

- New gene sets for donkey and horse
- Updated SIFT and PolyPhen-2 missense variant pathogenicity
- New VEP plugins for UTR annotation
- New ATAC-seq tracks (peaks and signal) for fish species (Atlantic Salmon, European Seabass, Rainbow Trout and Turbot)

[More release news](#) on our blog

Ensembl Rapid Release

New assemblies with gene and protein annotation every two weeks.
Note: species that already exist on this site will continue to be updated with the full range of annotations.

Go

The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-vertebrate genomes from biodiversity initiatives such as Darwin Tree of Life, the Vertebrate Genomes Project and the Earth BioGenome Project.

[Rapid Release news](#) on our blog



Ensembl creates, integrates and distributes reference datasets and analysis tools that enable genomics. We are based at [EMBL-EBI](#) and our software and data are freely available. Our [acknowledgements page](#) includes a list of current and previous funding bodies. [How to cite Ensembl](#) in your own publications.



Tools

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

Search all species... 🔍

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

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

All genomes

-- Select a species --

 **Pig breeds**
Pig reference genome and 12 additional breeds
[View full list of all species](#)

Favourite genomes

 **Human**
GRCh38.p13
[Still using GRCh37?](#)

 **Mouse**
GRCm39

 **Zebrafish**
GRCz11

Ensembl Rapid Release

New assemblies with gene and protein annotation every two weeks.

Note: species that already exist on this site will continue to be updated with the full range of annotations.

The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-vertebrate genomes from biodiversity initiatives such as Darwin Tree of Life, the Vertebrate Genomes Project and the Earth BioGenome Project.

[Rapid Release news](#)

Compare genes across species



Find SNPs and other variants for my gene



GTATACATTTC
CTRAAGTCTT
CTTCATAATTAT
GAAACATTTCC

Gene expression in different tissues



Retrieve gene sequence

```
GGCTTACCTTCGGGTTGG  
GGGGTTGGGGGGGGGGGGGG  
GGGGCTCTCTGGGGGGGGGG  
AAGGGGGGGGGGGGGGGGGGG  
CAGCTCTGGGGGGGGGGGGGG  
CCCCGGGGGGGGGGGGGGGG
```

Find a Data Display



Use my own data in Ensembl



EMBL-EBI  Ensembl creates, integrates and distributes reference datasets and analysis tools that enable genomics. We are based at [EMBL-EBI](#) and our software and data are freely available. Our [acknowledgements page](#) includes a list of current and previous funding bodies. [How to cite Ensembl](#) in your own publications.

 **GLOBAL CORE BIODATA RESOURCE**

 **elixir Core Data Resource**

Fetch data

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

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

Help & Documentation Accessing Ensembl Data

In this section

- Exporting data via website
- API data access
- Public MySQL Server
- FTP Download
 - Downloading with rsync
- BioMart
 - Biomart Bioc R package
 - Biomart Perl API
 - Biomart RESTful access
 - Combining species datasets
 - How to use BioMart
 - Virtual Machine

Search documentation | Go

Accessing Ensembl Data

Ensembl data is available through a number of routes - which you choose depends on the amount and type of data you wish to fetch. Please note that Ensembl coordinates always have a one-based start.

Small quantities of data

Many of the pages displaying Ensembl genomic data offer an [export](#) option, suitable for small amounts of data, e.g. a single gene sequence.

Click on the 'Export data' button in the lefthand menu of most pages to export:

- FASTA sequence
- GTF or GFF features

...and more!



Fast programmatic access

For fast access in any programming language, we recommend using our [REST server](#). Various REST endpoints provide access to vast amounts of Ensembl data.



Complete datasets and databases

Many datasets, e.g. all genes for a species, are available to download in a variety of formats from our [FTP site](#).

Entire databases are also available via FTP as MySQL dumps.



Complex cross-database queries

More complex datasets can be retrieved using the [BioMart](#) data-mining tool.



All data produced by the Ensembl project is [freely available](#) for your own use.

Ensembl release 109 - Feb 2023 © EMBL-EBI

[Permanent link](#)

About Us

[About us](#)

[Contact us](#)

[Citing Ensembl](#)

[Privacy policy](#)

[Disclaimer](#)

Get help

[Using this website](#)

[Adding custom tracks](#)

[Downloading data](#)

[Video tutorials](#)

[Variant Effect Predictor \(VEP\)](#)

Our sister sites

[Ensembl Bacteria](#)

[Ensembl Fungi](#)

[Ensembl Plants](#)

[Ensembl Protists](#)

[Ensembl Metazoa](#)

Follow us

[Blog](#)

[Twitter](#)

[Facebook](#)

Use your own data

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

Search all species... 

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

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 109 (Feb 2023)

- New gene sets for donkey and horse
- Updated SIFT and PolyPhen-2 missense variant pathogenicity
- New VEP plugins for UTR annotation
- New ATAC-seq tracks (peaks and signal) for fish species (Atlantic Salmon, European Seabass, Rainbow Trout and Turbot)

[More release news](#) on our blog

Search

All species for

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

All genomes

-- Select a species --

Pig breeds
Pig reference genome and 12 additional breeds

[View full list of all species](#)

Favourite genomes 

Human
GRCh38.p13
[Still using GRCh37?](#)

Mouse
GRCm39


Zebrafish
GRCz11


Ensembl Rapid Release

New assemblies with gene and protein annotation every two weeks.
Note: species that already exist on this site will continue to be updated with the full range of annotations.

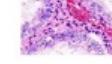
The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate and non-vertebrate genomes from biodiversity initiatives such as Darwin Tree of Life, the Vertebrate Genomes Project and the Earth BioGenome Project.

[Rapid Release news](#) on our blog

Compare genes across species 

Find SNPs and other variants for my gene

GTGATAACATTC
CTTAAAGTCTT
CTTC ^A AATT ^T
GTAAACATTTCC

Gene expression in different tissues 

Retrieve gene

GGCTTACG
GGCGCTTG
GGCGCTTG
AGGGGAGC
CACCTTG
CCCGATG

Use my own data in Ensembl 

Use your own data in Ensembl

Ensembl creates, integrates and distributes reference datasets and analysis tools that enable genomics. We are based at [EMBL-EBI](#) and our software and data are freely available. Our [acknowledgements page](#) includes a list of current and previous funding bodies. [How to cite Ensembl](#) in your own publications.

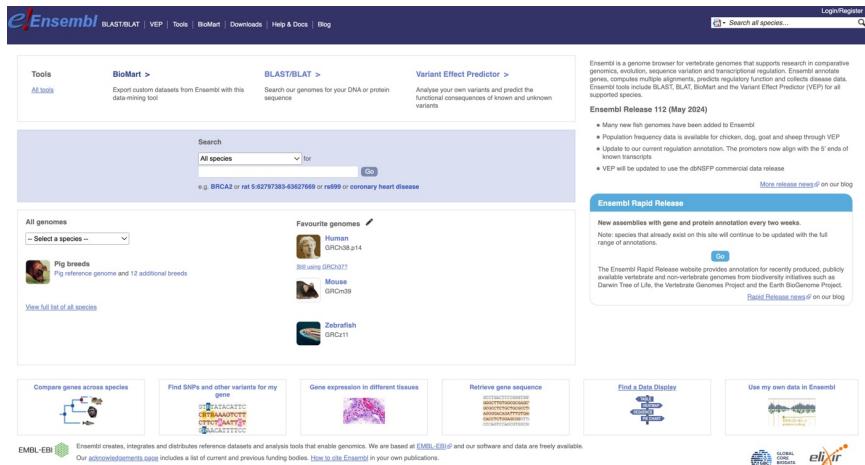
Help and documentations

- Youtube videos (workshop...)
- FAQ
- Exercises
- Online courses
- Publications :
 - Flicek, P. et al. **Ensembl 2013**. Nucleic Acids Res. Advanced Access (Database Issue). <http://www.ncbi.nlm.nih.gov/pubmed/23203987>
 - Xosé M. Fernández-Suárez and Michael K. Schuster. **Using the Ensembl Genome Server to Browse Genomic Sequence Data**. UNIT 1.15 in Current Protocols in Bioinformatics, Jun 2010
 - Giulietta M Spudich and Xosé M Fernández Suárez. **Touring Ensembl: A practical guide to genome browsing**. BMC Genomics 2010, 11:295 (11 May 2010)

Getting access to genomic data: Ensembl/BIOmart

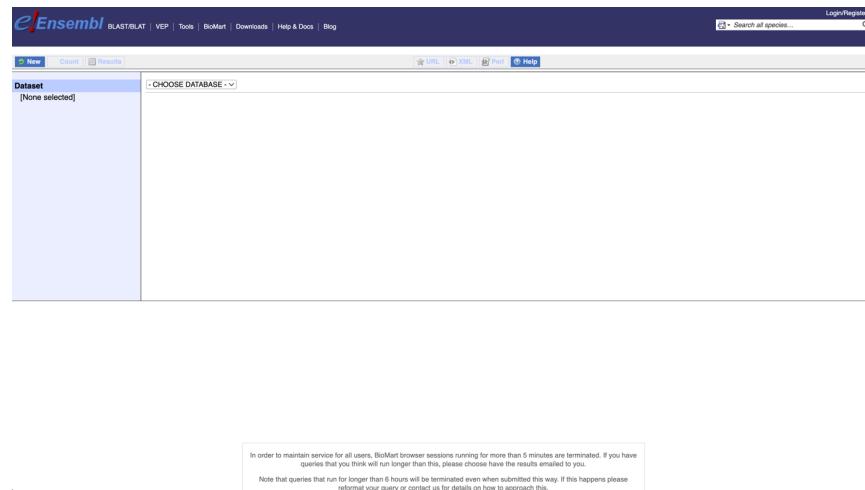
Access Ensembl's data

Web site



The screenshot shows the Ensembl homepage with a dark blue header containing the Ensembl logo, navigation links (BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, Blog), and a search bar ('Search all species...'). Below the header are three main tool sections: 'BioMart' (with a sub-link 'All tools'), 'BLAST/BLAT' (with a sub-link 'Search genomes for your DNA or protein sequence'), and 'Variant Effect Predictor' (with a sub-link 'Analyse your own variants and predict the functional consequences of known and unknown variants'). The main content area features a search bar, a 'Dataset' dropdown menu, and several cards for different species: Human (GRCm38.p14), Bilt more (GRCh37), Mouse (GRCn38), and Zebrafish (GRCz11). Below these are links for 'Compare genes across species', 'Find SNPs and other variants for my gene', 'Gene expression in different tissues', 'Retrieve gene sequence', 'Find a Data Display', and 'Use my own data in Ensembl'. The footer includes the EMBL-EBI logo and a note about funding.

Mining tool: BioMart



The screenshot shows the BioMart interface with a dark blue header and a search bar ('Search all species...'). Below the header is a toolbar with buttons for 'New', 'Count', 'Results', 'URL', 'XML', 'Fast', and 'Help'. A 'Dataset' dropdown menu is open, showing '[None selected]'. To the right is a 'CHOOSE DATABASE' dropdown menu. A note at the bottom left states: 'In order to maintain service for all users, BioMart browser sessions running for more than 5 minutes are terminated. If you have queries that you think will run longer than this, please choose have the results emailed to you.' Another note below it says: 'Note that queries that run for longer than 6 hours will be terminated even when submitted this way. If this happens please reformat your query or contact us for details on how to approach this.'



User friendly



Straightforward



Only one request at once



Get answers to complex queries



Very fast



Need training

BioMart

- <http://www.biomart.org/>
- Joint development between EBI and Cold Spring Harbor Laboratory (CSHL)
- Open source project
- BioMart can access diverse databases from a single interface
- It is a search engine that can find multiple terms and put them into a table format
- No programming required!

BioMart/Ensembl



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

Login/Register

Biomart

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 112 (May 2024)

- Many new fish genomes have been added to Ensembl
- Population frequency data is available for chicken, dog, goat and sheep through VEP
- Update to our current regulation annotation. The promoters now align with the 5' ends of known transcripts
- VEP will be updated to use the dbNSFP commercial data release

[More release news](#) on our blog

Search

All species for

Go

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

All genomes

-- Select a species --

Pig breeds
Pig reference genome and 12 additional breeds

Favourite genomes

Human
GRCh38.p14

Still using GRCh37?

Mouse
GRCm39

Ensembl Rapid Release

New assemblies with gene and protein annotation every two weeks.

Note: species that already exist on this site will continue to be updated with the full range of annotations.

Go

The Ensembl Rapid Release website provides annotation for

- Get access to :
 - Genomic annotation (genes, SNPs)
 - Functional annotation
 - Expression data

Example: Step 1 (Select datasets)

The screenshot shows the Ensembl BioMart interface. At the top, there is a navigation bar with links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. On the right side of the bar are 'Login/Register' and a search bar labeled 'Search all species...' with a magnifying glass icon. Below the navigation bar, there is a toolbar with buttons for 'New', 'Count', and 'Results'. To the right of the toolbar are links for 'URL', 'XML', 'Perl', and 'Help'. The main area has a header 'Dataset' with the subtext '[None selected]'. A dropdown menu is open, showing a list of datasets. The first item in the list, 'Ensembl Genes 112', is highlighted with a red arrow pointing to it. The dropdown menu title is 'CHOOSE DATASET -'. The list includes the following entries:

- Ensembl Genes 112
- Chicken genes (bGalGal1.mat.broiler.GRCg7b)
- Human genes (GRCh38.p14)
- Mouse genes (GRCm39)
- Rat genes (mRatBN7.2)
- Zebrafish genes (GRCz11)
- Abingdon island giant tortoise genes (ASM359739v1)
- African ostrich genes (ASM69896v1)
- Algerian mouse genes (SPRET_EiJ_v1)
- Alpaca genes (vicPac1)
- Alpine marmot genes (marMar2.1)
- Amazon molly genes (Poecilia_formosa-5.1.2)
- American bison genes (Bison_UMD1.0)
- American black bear genes (ASM334442v1)
- American mink genes (NNQGG.v01)
- Arabian camel genes (CamDro2)
- Arctic ground squirrel genes (ASM342692v1)
- Argentine black and white tegu genes (HLtupMer3)
- Armadillo genes (Dasnov3.0)

On the right side of the dropdown menu, there is a large red arrow pointing downwards, and next to it is the text 'First choose the database and dataset'.

Example: Step 2 (Filter)

The screenshot shows the Ensembl BioMart interface. In the top left, the Ensembl logo is visible. The top navigation bar includes links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. On the right, there are links for Login/Register and a search bar labeled "Search all species...".

The main interface has tabs for New, Count, and Results. Below these are sections for Dataset and Filters.

Dataset: Human genome (GRCh38.p14)

Filters:

- Start: 3653642
- End: 4543654
- Chromosome/scaffold: 1

Attributes:

- Gene stable ID
- Gene stable ID version
- Transcript stable ID
- Transcript stable ID version

Dataset: [None Selected]

Coordinates:

- Start: 3653642
- End: 4543654

An orange callout box points to the "Chromosome/scaffold: 1" filter with the text "Limit to chromosome 1". Another orange callout box points to the "Start" and "End" coordinates with the text "Limit to given coordinates".

Example: Step 3 (Count results)

Compute match count

New Count Results

Dataset 22 / 70611 Genes

Human genes (GRCn38.p14)

Filters

Start: 3653642
End: 4543654
Chromosome/scaffold: 1

Attributes

Gene stable ID
Gene stable ID version
Transcript stable ID
Transcript stable ID version

Dataset

[None Selected]

URL XML Perl Help

Search all species...

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

Coordinates
Start: 3653642
End: 4543654

Example: Step 4 (Select attributes)

Screenshot of the Ensembl BioMart interface showing the 'Select attributes' step.

The interface includes a navigation bar with links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar at the top right allows searching across all species. The main content area shows a dataset summary for Human genes (GRCh38.p14) and a list of filters and attributes selected. A large orange callout box highlights the 'Select attributes to be output' section, which contains a list of checkboxes for selecting specific gene and transcript attributes.

Dataset 22 / 70611 Genes
Human genes (GRCh38.p14)

Filters
Start: 3653642
End: 4543654
Chromosome/scaffold: 1

Attributes
Gene stable ID
Transcript stable ID

Dataset
[None Selected]

Please select columns to be included in the output and hit 'Results' when ready
Missing non coding genes in your mart query output, please check the following [FAQ](#)

Features Variant (Germline)
 Structures Variant (Somatic)
 Homologues (Max select 6 orthologues) Sequences

GENE:
Ensembl
 Gene stable ID
 Gene stable ID version
 Transcript stable ID
 Transcript stable ID version
 Protein stable ID
 Protein stable ID version
 Exon stable ID
 Gene description
 Chromosome/scaffold name
 Gene start (bp)
 Gene end (bp)
 Strand
 Karyotype band
 Transcript start (bp)

Select attributes to be output

GENCODE basic annotation
 APPRIS annotation
 Ensembl Canonical
 RefSeq match transcript (MANE Select)
 RefSeq match transcript (MANE Plus Clinical)
 Gene name
 Source of gene name
 Transcript name
 Source of transcript name
 Gene % GC content
 Gene type
 Transcript type
 Source (gene)
 Source (transcript)

Example: Step 5 (get results)

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

Login/Register

New Count Results

URL XML Perl Help

Dataset 22 / 70611 Genes
Human genes (GRCh38.p14)

Filters

Start: 3653642
End: 4543654
Chromosome/scaffold: 1

Attributes

Gene stable ID
Transcript stable ID

Dataset

[None Selected]

Export all results to File TSV Unique results only Go

Email notification to

View 10 rows as HTML Unique results only

Gene stable ID	Transcript stable ID
ENSG00000078900	ENST00000378295
ENSG00000078900	ENST00000604074
ENSG00000078900	ENST00000603362
ENSG00000078900	ENST00000604479
ENSG00000078900	ENST00000354437
ENSG00000078900	ENST00000713570
ENSG00000078900	ENST00000713571
ENSG00000078900	ENST00000713572
ENSG00000078900	ENST00000378280
ENSG00000078900	ENST00000378285

Ensembl/Biomart (part 2) hands-on