

# Coursebook: Browsing Genomes with Ensembl



# ENSEMBL

[www.ensembl.org](http://www.ensembl.org)

## Coursebook

<https://training.ensembl.org>

27<sup>th</sup> August 2026

## Table of contents

|                                                                          |          |
|--------------------------------------------------------------------------|----------|
| <b>Table of contents.....</b>                                            | <b>1</b> |
| <b>Introduction to Ensembl.....</b>                                      | <b>2</b> |
| Getting started with Ensembl.....                                        | 2        |
| Synopsis - What can I do with Ensembl?.....                              | 3        |
| Stay in touch!.....                                                      | 3        |
| Further reading.....                                                     | 3        |
| <b>EMBL-EBI Training Pre-learning.....</b>                               | <b>5</b> |
| <b>Exercise: Searching for genes across human genome assemblies.....</b> | <b>6</b> |
| Assessment.....                                                          | 7        |
| <b>Quick Guide to Databases and Projects.....</b>                        | <b>8</b> |
| General help.....                                                        | 8        |
| Sequences.....                                                           | 8        |
| Gene names.....                                                          | 9        |
| Protein signatures.....                                                  | 9        |
| Other projects.....                                                      | 9        |
| <b>Exercise Answers: Exploring genes with phenotypes.....</b>            | <b>1</b> |

# Introduction to Ensembl

## Getting started with Ensembl

Ensembl ([www.ensembl.org](http://www.ensembl.org)) is a project based at EMBL's European Bioinformatics Institute ([EMBL-EBI](http://EMBL-EBI)) that annotates genomes.

The Ensembl project's goal is to enable genomic science by providing high-quality, integrated and consistent annotation on all cellular genomes within a harmonious, scalable and accessible infrastructure.

Ensembl generates and imports genome annotation such as genes, variation, regulation and comparative genomics for vertebrates, non-vertebrates and key model organisms, and makes this information freely available online.

The website provides visualisations and functionality that enable the manipulation and analysis of this data. Information about genes, transcripts and further annotation can be retrieved at the genome, gene and protein level.

Ensembl provides genes and other **annotations** such as predicted regulatory regions, base pairs conserved across species, and observed sequence variations. The Ensembl gene set is based on protein and mRNA evidence in **UniProtKB** and **NCBI RefSeq** databases, along with manual annotation from the **Ensembl-Havana** group. All the data are freely available and can be accessed via the web browser at [www.ensembl.org](http://www.ensembl.org). Programmers can directly access Ensembl databases through [Ensembl GraphQL](#). Gene sequences can be downloaded from the Ensembl browser itself, or through programmatic use of [Ensembl refget](#).



## Synopsis - What can I do with Ensembl?

- View genes, with other annotations, along the chromosome.
- View alternative transcripts (such as splice variants) for a given gene.
- For any gene, explore homologues in a reference collection for that species
- Examine single nucleotide polymorphisms (SNPs) for a gene or chromosomal region.
- View SNPs across strains (rat, mouse), populations (human), or breeds (dog).
- Use BLAST against any Ensembl genome.
- Export sequences for genomic features.
- Determine how your variants affect genes and transcripts using the Variant Effect Predictor.

## Stay in touch!

- Email the Ensembl team with comments or questions at [helpdesk@ensembl.org](mailto:helpdesk@ensembl.org)
- Follow the [Ensembl Blog](#)
- Sign up to [Ensembl Mailing Lists](#)
- Follow us on social media:
  - [LinkedIn](#)
  - [Bluesky](#)
  - [X](#) (formerly Twitter)
  - [Facebook](#)

## Further reading

Yates A, et al. [Ensembl 2026](#). Nucleic Acids Research, Volume 54, Issue D1, 6 January 2026, Pages D1053–D1060

Dyer SC, Austine-Orimoloye O, Azov AG, et al. [Ensembl 2025](#). Nucleic Acids Research. 2025 Jan;53(D1):D948-D957. DOI: 10.1093/nar/gkae1071. PMID: 39656687; PMCID: PMC11701638.

Yates AD, Allen J, Amode RM, et al. [Ensembl Genomes 2022: an expanding genome resource for non-vertebrates](#). Nucleic Acids Research. 2022 Jan;50(D1):D996-D1003. DOI: 10.1093/nar/gkab1007. PMID: 34791415;



PMCID: PMC8728113.

Newman V, Moore B, Sparrow H, Perry E. [The Ensembl Genome Browser: Strategies for Accessing Eukaryotic Genome Data](#). Methods in Molecular Biology (Clifton, N.J.). 2018 ;1757:115-139. DOI: 10.1007/978-1-4939-7737-6\_6. PMID: 29761458.

For a complete list of publications, visit:

<https://www.ensembl.org/about/articles/citing-ensembl>

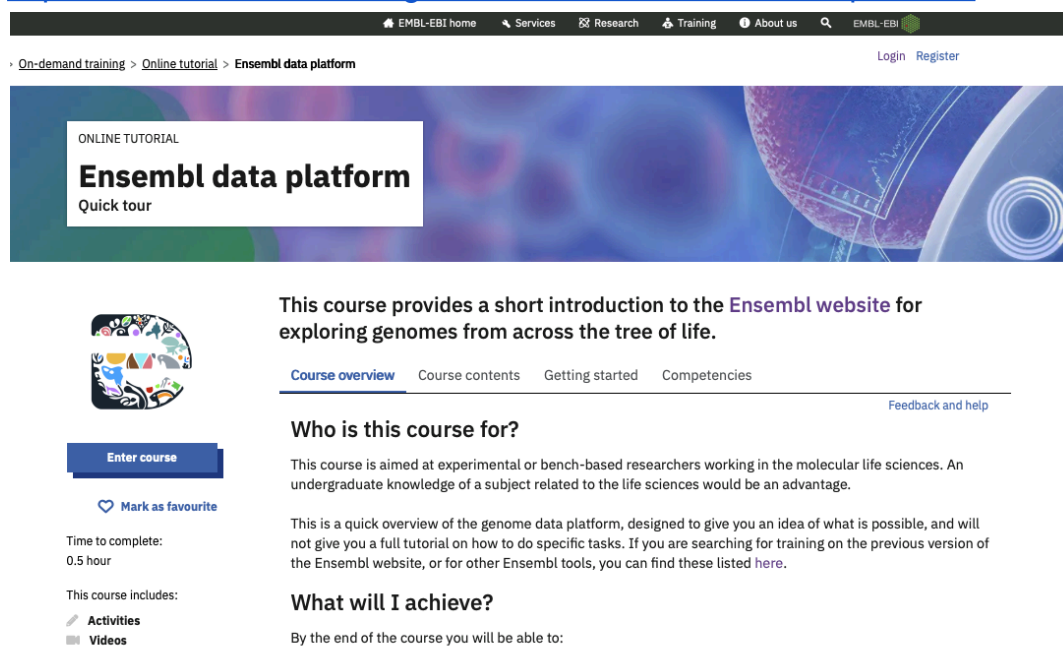


## EMBL-EBI Training Pre-learning

Please complete the following work before the session. There is a quick tour on the Ensembl data platform on the EBI Training site -

The course can be found on the following page:

<https://www.ebi.ac.uk/training/online/courses/ensembl-data-platform/>



The screenshot shows the EMBL-EBI Training website. The top navigation bar includes links for EMBL-EBI home, Services, Research, Training, and About us. Below the navigation bar, a breadcrumb trail reads: On-demand training > Online tutorial > Ensembl data platform. The main header area features a large image with the text "ONLINE TUTORIAL" and "Ensembl data platform Quick tour". Below this, a circular icon with a globe and DNA helix is shown. The course description states: "This course provides a short introduction to the Ensembl website for exploring genomes from across the tree of life." The course overview section includes a list of topics: Course overview, Course contents, Getting started, and Competencies. The "Who is this course for?" section explains that the course is aimed at experimental or bench-based researchers working in the molecular life sciences. The "What will I achieve?" section lists the outcome: "Outline the Ensembl Data Platform and what you can do with the resource".

ONLINE TUTORIAL

### Ensembl data platform

Quick tour

EMBL-EBI home Services Research Training About us

On-demand training > Online tutorial > Ensembl data platform

Login Register

Enter course

Mark as favourite

Time to complete: 0.5 hour

This course includes:

- Activities
- Videos

This course provides a short introduction to the **Ensembl website** for exploring genomes from across the tree of life.

[Course overview](#) [Course contents](#) [Getting started](#) [Competencies](#) [Feedback and help](#)

#### Who is this course for?

This course is aimed at experimental or bench-based researchers working in the molecular life sciences. An undergraduate knowledge of a subject related to the life sciences would be an advantage.

This is a quick overview of the genome data platform, designed to give you an idea of what is possible, and will not give you a full tutorial on how to do specific tasks. If you are searching for training on the previous version of the Ensembl website, or for other Ensembl tools, you can find these listed [here](#).

#### What will I achieve?

By the end of the course you will be able to:

- Outline the Ensembl Data Platform and what you can do with the resource



## Exercise: Searching for genes across new human genome assemblies

There are many new human assemblies available to explore in the new Ensembl site, including the new Telomere to telomere assembly of the human genome - T2T-CHM13

*BRCA2* is a well researched gene linked to hereditary breast cancer. Explore the gene and available transcripts in the human reference genome (GRCh38) and the new complete genome - T2T-CHM13.

- 1.) Do you find that genes lie on the same chromosome?
- 2.) Are their chromosomal coordinates identical?
- 3.) Are the number of transcripts available the same?
- 4.) Are there any differences in how these genomes were annotated? - (hint- check the right hand side of genome homepage in Genome selector)
- 5.) Bonus - Explore the metadata for the samples on the Human Pangenome Reference Consortium (HPRC) data explorer - <https://data.humanpangenome.org/samples> - and choose a sample from the population closest to your home country. See if it's available in Ensembl and explore as above!

## Assessment

- A. Search Ensembl for the human GRCh38 gene *PCNA*. On which chromosome, and on which strand of the genome, is this gene located?
- B. Are there any other features (genes, lncRNAs etc) which overlap the start site of the *PCNA* gene?
- C. For the default gene in the view, there is a label - MANE select. What does this label mean?
- D. Open the gene in Feature explorer. Use the external reference for MIM-morbid to explore any linked disorders on the [MIM-morbid](#) link. What is the disorder and the reported inheritance pattern?
- E. What is the MIM Gene ID for *PCNA*?
- F. How many transcripts (splice variants) does the *PCNA* gene have and how many are protein-coding?
- G. What is the longest transcript? How long is the protein it encodes for?
- H. Can you decide which transcripts are of better quality/has higher support?
- I. Click on the [transcript ID](#) for ENST00000379160.4 to expand its information. How many coding exons does it have?

## Quick Guide to Databases and Projects

Here is a list of databases and projects you will come across in these exercises. Google any of these to learn more. Projects include many species, unless otherwise noted.

### General help

[Ensembl Glossary](#) (hasn't moved to new site as yet)  
[Ensembl FAQs](#)

### Sequences

EMBL-Bank, NCBI GenBank, DDBJ: Contain nucleic acid sequences deposited by submitters such as wet-lab biologists and gene sequencing projects. These three databases are synchronised with each other every day, so the same sequences should be found in each.

CCDS: coding sequences that are agreed upon by Ensembl, Ensembl-Havana, NCBI, HGNC and MGI. (*Human and mouse*)

Ensembl HAVANA: The team responsible for the manual annotation of human and mouse genes, located at the European Bioinformatics Institute (EBI), Hinxton, UK.

MANE: Matched Annotation from NCBI and EMBL-EBI, an initiative to agree on a well-supported and biologically relevant minimal transcript set for each protein-coding gene that is identical in Ensembl-and NCBI RefSeq (*Human*).

NCBI Entrez Gene: NCBI's gene collection.

NCBI RefSeq: NCBI's collection of "reference sequences", includes genomic DNA, transcripts, and proteins. NM stands for "Known mRNA" (e.g. NM\_005476) and NP (e.g. NP\_005467) are "Known proteins".

UniProtKB: the "Protein knowledgebase", a comprehensive set of protein sequences. Divided into two parts: Swiss-Prot and TrEMBL.



UniProt Swiss-Prot: the manually annotated, reviewed protein sequences in the UniProtKB. High quality.

UniProt TrEMBL: the automatically annotated, unreviewed set of proteins (EMBL-Bank translated). Varying quality.

VEGA: Vertebrate Genome Annotation, a selection of manually-curated genes, transcripts, and proteins. (*Human, mouse, zebrafish, gorilla, wallaby, pig, and dog*). The VEGA project has now been archived, and will not be updated, but the manual annotation will continue by the Ensembl-HAVANA team (see below).

## Gene names

HGNC: HUGO Gene Nomenclature Committee, a project assigning a unique and meaningful name and symbol to every human gene.

ZFIN: The Zebrafish Model Organism Database. Gene names are only one part of this project.

## Protein signatures

InterPro: A collection of domains, motifs, and other protein signatures. Protein signature records are extensive, and combine information from individual projects such as UniProt, along with other databases such as SMART, PFAM, and PROSITE (explained below).

PFAM: A collection of protein families.

PROSITE: A collection of protein domains, families, and functional sites.

SMART: A collection of evolutionarily conserved protein domains.

## Other projects

NCBI dbSNP: A collection of sequence polymorphisms, mainly single nucleotide polymorphisms, along with insertion-deletions.

NCBI OMIM: Online Mendelian Inheritance in Man - a resource showing phenotypes and diseases related to human genes.



# Exercise Answer

## Exercise Answers: Exploring genes across human assemblies

### 1. Do you find that genes lie on the same chromosome?

First select the GRCh38 and CHM13 genomes using the Genome selector <https://www.ensembl.org/genome-selector>. You can search for each genome by using the genome accession name "GRCh38"/"CHM13". You can also click on the human handprint icon to show all human assemblies, and select the two you need from the list and add.

### 2. Are there chromosomal coordinates identical?

Use the search function from the Genome selector or another app to search for the gene feature "BRCA2". Explore the result in the Genome Browser app or the Feature explorer app. The coordinates for features are on the top of the display. Both genes are on chromosome 13 but at different coordinates on the chromosome.

### 3. Are the number of transcripts available the same?

Transcript count is shown as a number on the left hand side of the gene on the Genome browser. You can click this to expand and see the transcripts. You can close this again with an "X" to reduce it to the best evidence transcript. The number of transcripts in the Feature explorer is shown to the right of the visual summary image of all transcripts for that gene. The transcript counts between these two genomes are not the same.

### 4. Are there any differences in how these genomes were annotated? - (hint- check the right hand side of genome homepage in Genome selector)

Click on the Genome selector app and then click the "lozenge" or tab with for the genome you want to explore the assembly details of. On the right hand side panel, you can see annotation information. GRCh38 is annotated manually. The CHM13 genome was annotated by mapping over annotations from GRCh38 with a customised [computational pipeline](#).

### 5. Bonus - you may find interesting results based on which sample you choose. This is the first step to exploring human pangenomes.