Ensembl: Quick tour

Emily Perry [1]

- DNA & RNA
- Beginner
- 0.5 hour

This course provides a very short introduction to the Ensembl website [2] for browsing chordate genomes. This is a quick overview to the browser, designed to give you an idea of what is possible, and will not give you a full tutorial on how to do specific tasks. Links are provided throughout to other tutorials and videos that may help you.

Learning objectives:

- Basic understanding of Ensembl and what you can do with the resource
- Know where to find out more about Ensembl

What is Ensembl?

Ensembl provides a genome browser that acts as a single point of access to annotated genomes, primarily for vertebrate species (Figure 1).

Information such as gene sequence, splice variants and further annotation [3] can be retrieved at the genome, gene and protein level. This includes information on protein domains, genetic variation, homology, syntenic regions and regulatory elements. Coupled with analyses such as whole genome alignments and the effects of sequence variation on proteins, this powerful tool aims to describe a gene or genomic region in detail.

Ensembl imports genome sequences from consortia, which is consistent with many other bioinformatics projects. Each species in Ensembl has its own homepage, where you can find out who provided the genome sequence and which version of the genome assembly is represented. To see an example, visit the Ensembl homepage for human [4].
Figure 1. Ensembl contains information about chordate genomes, including: human, mouse, rat, zebrafish, panda, and pufferfish.

What can you do with Ensembl?

Ensembl has many features. Some of the things you can do are:

- Examine the characteristics of a region, such as genes, regulatory features [5] or oligo probes.
- Retrieve genomic, cDNA and gene sequences.
- Align sequences against any genome in Ensembl.
- Study gene alignments between species.
- View gene transcripts and proteins.
- Export data such as sequences, tables and Single Nucleotide Polymorphism [6] (SNP [6]) data.
- Upload your own data to the browser.
- View sequence variation.

NB. This quick tour will not cover all of these features - the aim is to guide you so that you can explore the Ensembl website yourself. For a more detailed look at Ensembl, please see the full course: Ensembl: Browsing chordate genomes [7]

Access and navigate Ensembl

The Ensembl [8] homepage is shown in Figure 2, below.
Figure 2. The Ensembl homepage [2]. This particular screenshot was taken from Ensembl release 69; the page may change in future releases. Some of the key features are highlighted.

Searching and visualising data from Ensembl

You can search for a genomic region to view it in the genome browser. Figure 3 (below) highlights some of the features you may be interested in.
If you access the webpage above, you will not have the **gene**, **transcript**, **variation** and **regulation** tabs open, as in the diagram. These tabs will open automatically if you select the gene, transcript, variation or regulation features from this page or other pages. This allows easy navigation between different views.

Explore the left-hand menu and the different tabs (the menu will change within the different tabs) to find out about the region, genes, transcripts, variation and regulation. You can view sequence data, alignments, gene trees and ontology, and more.

We have a 'region in detail' video tutorial that will help you to explore this particular view, available from **YouTube** [10]. If you are in China and cannot see the video in YouTube, **watch it on YouKu**. [11]
Configuring (customising) a page

Click this button in the left-hand menu of the browser window to open the ‘Configure’ menu.

You can choose what data you want to see in the configure menu (Figure 4). Browse by track categories in the left-hand menu, or search for a track in the right-hand search box. Find out about tracks by clicking on the icon. Click on the check box beside the track names to turn the tracks on and off, or change the settings.

Figure 4. The Ensembl configure menu.
BLAST search with sequences

Input your sequence into BLAST [12] (Figure 5):

**Figure 5.** BLAST input menu - accessible from the blue banner at the top of every page.

And get some results (Figure 6):
Figure 6. Sample BLAST results.

Submitting and managing your own data

Click this button in the left-hand menu of the browser to open the 'Manage your data' menu (Figure 7).
Figure 7. The 'Manage your data' menu. This allows you to view your data in the Ensembl browser. Any data you submit will not be publicly visible.

You can upload data from a file on your computer, by copying and pasting, or just link to a webpage. Manage data you have already uploaded by clicking on the button.


Getting data from Ensembl using Biomart

Our Biomart tool [14] allows you to extract data from our databases. Select your data type (genes, variation or regulation) and species. Open 'Filters' to filter your data. Open 'Attributes' to select the output you want to recover. Click 'Results'.
Figure 8. The BioMart tool enables you to extract data from our databases.

If you want to learn more about BioMart, here is a short (5 minute) video introduction. If you cannot see the YouTube video above, watch it on YouKu [15].

Access our data directly

You can access our data directly in many ways:

- Access our databases using our Perl API [16].
- Access our databases via our public MySQL client [17].
- Download complete databases from our FTP site [18].
The Perl API is a flexible tool that can handle large data sets.

Your feedback

Please tell us what you thought about this course. Your feedback is invaluable and helps us to improve our courses and thus enhance your learning experience.

Get help and support on Ensembl

Tutorials

- Visit the full 3 hour course on Train online: Ensembl: Browsing chordate genomes [7].
- Try out an Ensembl tutorial [19]. There are videos and coursebooks on a variety of subjects.
- See if your question has already been answered in the Ensembl FAQs [20].
- Have a look at the Ensembl Glossary [21].
- Check out our YouTube [22] or YouKu [23] channel.
- View technical documentation on Ensembl [24].

Support

- If you cannot find the answer to your question, contact the Ensembl helpdesk [25].
- If you are using Ensembl programmatically, our dev list [26] is a community of Ensembl developers where you can ask and answer questions.
- Host a workshop [27] at your institution to train you and your colleagues to use Ensembl.

Contributors

Emily Perry [1]
EMBL-EBI
Ensembl Outreach Project Leader

Emily is the Outreach Project Leader for Ensembl: she is responsible for the team that teaches workshops, creates training materials and help pages, manages social media, answers helpdesk queries and aids development of new tools for the resource. Emily started at EMBL-EBI as an Ensembl Outreach Officer in September 2012 and became the Project Leader in March 2015. Before working at EMBL-EBI, Emily did her PhD in molecular biology at the MRC Human Genetics Unit in Edinburgh, then worked for the University of Edinburgh's SCI-FUN group, touring Scottish secondary schools with an interactive science roadshow.

Giulietta M. Spudich [28]
EMBL-EBI
Outreach project leader, Ensembl

Giulietta was the outreach project leader for the Ensembl at EMBL’s European Bioinformatics Institute (EBI). The Ensembl project freely provides high quality annotation such as genes, sequence variation, and whole genome alignments across mainly vertebrate genomes. She leads a small team that organises and delivers training courses worldwide, and supports scientific communication about the project.

Before she started working with the Ensembl project in 2006, she obtained her PhD in Susan Marqusee's lab at UC Berkeley in 2002. She 'hopped the pond' to carry out postdoctoral research in biochemical studies of Myosin VI at the MRC-LMB in Cambridge, UK (2002-2006).

Funding

Ensembl is a joint project of the Wellcome Trust Sanger Institute and the European Bioinformatics Institute. The Wellcome Trust [29] provides majority funding for the project (grant numbers WT062023 and WT079643) with additional funding for specific project components from the National Human Genome Research Institute [30] (U01HG004695, U54HG004563 and U41HG006104), the BBSRC [31] (BB/I025506/1) and the European Molecular Biology Laboratory [32]. Additional support for specific project components as specified: "The research leading to these results has received funding from the European Community's Seventh Framework Programme [33](FP7/2007-2013) under grant agreement n° 222664. ("Quantomics" [34]). This Publication reflects only the author's views and the European Community is not liable for any use that may be made of the information contained herein"; "The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement number 200754 - the GEN2PHEN [35] project."; "The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 282510 - BLUEPRINT [36]." Rat genomics resources receive additional support as specified: "The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement No HEALTH-F4-2010-241504 (EURATRANS)."