Training materials

- Ensembl training materials are protected by a CC BY license
- http://creativecommons.org/licenses/by/4.0/
- If you wish to re-use these materials, please credit Ensembl for their creation
- If you use Ensembl for your work, please cite our papers
- http://www.ensembl.org/info/about/publications.html
Browsing Genes and Genomes with Ensembl

Helen Sparrow
Ensembl Outreach Officer
EMBL-EBI
Objectives

- What is Ensembl?
- What type of data can you get in Ensembl?
- How to navigate the Ensembl browser website.
- How to use Ensembl tools
- Where to go for help and documentation.
## This webinar course

All webinars begin at 9am BST

<table>
<thead>
<tr>
<th>Date</th>
<th>Webinar topic</th>
<th>Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>6th April</td>
<td>Introduction to Ensembl</td>
<td>Helen Sparrow</td>
</tr>
<tr>
<td>13th April</td>
<td>Ensembl genes</td>
<td>Emily Perry</td>
</tr>
<tr>
<td>20th April</td>
<td>Data export with BioMart</td>
<td>Victoria Newman</td>
</tr>
<tr>
<td>27th April</td>
<td>Variation data in Ensembl and the Ensembl VEP</td>
<td>Victoria Newman</td>
</tr>
<tr>
<td>4th May</td>
<td>Comparing genes and genomes with Ensembl Compara</td>
<td>Ben Moore</td>
</tr>
<tr>
<td>11th May</td>
<td>Finding features that regulate genes – the Ensembl Regulatory Build</td>
<td>Ben Moore</td>
</tr>
<tr>
<td>18th May</td>
<td>Uploading your data to Ensembl and advanced ways to access Ensembl data</td>
<td>Emily Perry</td>
</tr>
</tbody>
</table>
Structure

Presentation:
What the data/tool is
How we produce/process the data

Demo:
Getting the data
Using the tool

Exercises:
On the train online course
Questions?

- Ask questions in the Chat box in the webinar interface
- My Ensembl colleagues will respond during
- There’s no threading so please respond with @username
Introduction to Ensembl

This webinar will be held on the 6th April 2017 at 9am BST and will be presented by Helen Sparrow.

This first webinar introduces you to the Ensembl project and provides a brief overview. Using the browser you will find out how to explore your species of interest and browse the genome.

A link to exercises and their solutions are in the page hierarchy

You can catch last year’s webinar on Ensembl, you can watch the video on YouKu.
You can also download the slides.

The “next page” is the exercises for this module

This text will be replaced by a YouTube (link to YouKu too) video of this webinar and a pdf of the slides
Get help with the exercises

- Use the exercise solutions in the online course
- Join our Facebook group and discuss the exercises with everybody (see the online course for the link)
- Email us: helpdesk@ensembl.org
Exploring the Ensembl genome browser
Introduction

Why do we need genome browsers?
Introduction

Why do we need genome browsers?

1977: 1st genome to be sequenced (5 kb)
2004: finished human sequence (3 Gb)
2015: completion of 1000 Genomes project
Buried in the genome sequence is information about:

- **Function** - such as transcripts and proteins
- **Expression** - regulatory elements
- **Variation** - SNPs, structural variants
We need to make the data mean something...

http://www.ensembl.org


http://genome.ucsc.edu
Ensembl Features

- Gene builds for ~70 species
- Gene trees
- Regulatory build (ENCODER)
- Variation display and VEP
- Display of user data
- BioMart (data export)
- Programmatic access via the APIs
- Completely Open Source
Ensembl Features

- Gene builds for ~70 species
- Gene trees
- Regulatory build (ENCODExE)
- Variation display and VEP
- Display of user data
- BioMart (data export)
- Programmatic access via the APIs
- Completely Open Source
Access scales

Main browser
Mobile site

One by one

BioMart
REST API
VEP

Perl API
MySQL

Groups

FTP

Whole genome
Ensembl is used worldwide

Top 10 countries

United Kingdom
United States
Spain
Germany
China
France
Italy
Japan
Netherlands
Singapore
Vertebrate species on Ensembl
Non-vertebrates on Ensembl genomes

Bacteria

Protists

Fungi

Metazoa

Plants

www.ensemblgenomes.org
# Ensembl and Ensembl Genomes

<table>
<thead>
<tr>
<th></th>
<th>Ensembl</th>
<th>Ensembl Genomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Released</td>
<td>2000</td>
<td>2009</td>
</tr>
<tr>
<td>Species</td>
<td>Vertebrates (fly, worm and yeast as outgroups)</td>
<td>Non-vertebrates (protists, plants, fungi, metazoa, bacteria)</td>
</tr>
<tr>
<td>Annotation</td>
<td>by Ensembl</td>
<td>in collaboration with the scientific communities</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://www.ensembl.org">www.ensembl.org</a></td>
<td><a href="http://www.ensemblgenomes.org">www.ensemblgenomes.org</a></td>
</tr>
</tbody>
</table>
What is a genome assembly?

**Sequence reads**

- CGGCCTTTGGGCTCCGCCTTCAGCTCAAGA
- CAGCTGTCCCAGATGAC
- ACTTAACTTCCCTCCCAGCTGTCC
- GGGCTCCGCCTTCAGCTC
- AACTTCCCTCCCAAGCT
- CGGCCTTTGGGCTCC
- TCCGCCTTCAGCTCAAGACTTAACTTC
- TCCCAGCTGTCCCAGATGACGACGCCATC
- CAGATGACGCC
- CAGCTGTCCCAGATGAC
- CGGCCTTTGGGCTCCGCCTTCAGCTCAAGA
- CAGCTGTCCCAGATGAC

**Match up overlaps**

- CGGCCTTTGGGCTCCGCCTTCAGCTCAAGA  AACTTCCCTCCCAAGCT  CAGATGACGCC
- TCCGCCTTCAGCTCAAGACTTAACTTC  TCCCAGCTGTCCCAGATGACGACGCCATC
- GGGCTCCGCCTTCAGCTC  AACTTAACTTCCCTCCCAGCTGTCC  CAGATGACGCC
- CGGCCTTTGGGCTCC

**Genome assembly**

- CGGCCTTTGGGCTCCGCCTTCAGCTCAAGA
- ACTTAACTTCCCTCCCAGCTGTCC
- TCCGCCTTCAGCTCAAGACTTAACTTC
- GGGCTCCGCCTTCAGCTC
- AACTTCCCTCCCAAGCT
- CGGCCTTTGGGCTCCGCCTTCAGCTCAAGA
- CAGCTGTCCCAGATGAC
- CAGATGACGCC
- CAGCTGTCCCAGATGAC
- CGGCCTTTGGGCTCCGCCTTCAGCTCAAGA
Genome contigs
Hands on

- We’re going to look at the Ensembl homepage and how to find information about the species and genome assemblies in Ensembl.

- We’re going to look at a region of the human genome, 4:122868000-122946000, and manipulate the view to see the data we’re interested in.
Ensembl: more than just a browser

- Import Genome Assemblies
- Annotate genomes
  - Genes and transcripts
  - Variants
  - Regulatory features
  - Comparative Genomics
- Display
- Tools
- APIs
# This webinar course

All webinars begin at 9am BST

<table>
<thead>
<tr>
<th>Date</th>
<th>Webinar topic</th>
<th>Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>6th April</td>
<td>Introduction to Ensembl</td>
<td>Helen Sparrow</td>
</tr>
<tr>
<td>13th April</td>
<td>Ensembl genes</td>
<td>Emily Perry</td>
</tr>
<tr>
<td>20th April</td>
<td>Data export with BioMart</td>
<td>Victoria Newman</td>
</tr>
<tr>
<td>27th April</td>
<td>Variation data in Ensembl and the Ensembl VEP</td>
<td>Victoria Newman</td>
</tr>
<tr>
<td>4th May</td>
<td>Comparing genes and genomes with Ensembl Compara</td>
<td>Ben Moore</td>
</tr>
<tr>
<td>11th May</td>
<td>Finding features that regulate genes – the Ensembl Regulatory Build</td>
<td>Ben Moore</td>
</tr>
<tr>
<td>18th May</td>
<td>Uploading your data to Ensembl and advanced ways to access Ensembl data</td>
<td>Emily Perry</td>
</tr>
</tbody>
</table>
Next webinar – Genes

Ensembl uses an automated pipeline to annotate genomic assemblies and this annotation is presented alongside Havana manual annotation in the browser.

Focusing on the gene tab within the Ensembl browser, we will explain the process of genome annotation and explore the range of data available for genes and transcripts.

Note the time: 9am BST (not GMT!)

Emily Perry
Questions?

- You can continue to use the chat box
Course exercises

http://www.ebi.ac.uk/training/online/course/ensembl-browser-webinar-series-2016

Introduction to Ensembl

This webinar will be held on the 24th March.

This first webinar will introduce you to the Ensembl overview of the resources it provides. You will then be taken to the Ensembl browser.

Using the browser you will find out how to explore your species of interest, look at the region in detail view and how to browse the genome.

A link to exercises and their solutions will appear in the page hierarchy

The “next page” will be the exercises

This text will be replaced by a YouTube (link to YouKu too) video of the webinar and a pdf of the slides.
Get help with the exercises

- Use the exercise solutions in the online course

- Join our Facebook group and discuss the exercises with everybody (see the online course for the link)

- Email us: helpdesk@ensembl.org
Help and documentation

Course online
http://www.ebi.ac.uk/training/online/subjects/11
Tutorials www.ensembl.org/info/website/tutorials

Flash animations
www.youtube.com/user/EnsemblHelpdesk
http://u.youku.com/Ensemblhelpdesk

Email us helpdesk@ensembl.org
Ensembl public mailing lists dev@ensembl.org, announce@ensembl.org
Follow us

www.facebook.com/Ensembl.org

@Ensembl

www.ensembl.info
Publications

http://www.ensembl.org/info/about/publications.html

Aken, BL. et al
Ensembl 2017
Nucleic Acids Research
http://europepmc.org/articles/PMC5210575

Xosé M. Fernández-Suárez and Michael K. Schuster
Using the Ensembl Genome Server to Browse Genomic Sequence Data.
Current Protocols in Bioinformatics 1.15.1-1.15.48 (2010)
www.ncbi.nlm.nih.gov/pubmed/20521244

Giulietta M Spudich and Xosé M Fernández-Suárez
Touring Ensembl: A practical guide to genome browsing
www.biomedcentral.com/1471-2164/11/295
Ensembl Acknowledgements

The Entire Ensembl Team

Bronwen L. Aken¹, Premanand Achuthan¹, Wasiu Akanni¹, M. Ridwan Amode¹, Friederike Bernsdorff¹, Jyothish Bhai¹, Konstantinos Billis¹, Denise Carvalho-Silva¹, Carla Cummins¹, Peter Clapham², Laurent Gil¹, Carlos García Girón¹, Leo Gordon¹, Thibaut Hourlier¹, Sarah E. Hunt¹, Sophie H. Janacek¹, Thomas Juettemann¹, Stephen Keenan¹, Matthew R. Laird¹, Ilias Lavidas¹, Thomas Maurel¹, William McLaren¹, Benjamin Moore¹, Daniel N. Murphy¹, Rishi Nag¹, Victoria Newman¹, Michael Nuhn¹, Chuang Kee Ong¹, Anne Parker¹, Mateus Patricio¹, Harpreet Singh Riat¹, Daniel Sheppard¹, Helen Sparrow¹, Kieron Taylor¹, Anja Thormann¹, Alessandro Vullo¹, Brandon Walts¹, Steven P. Wilder¹, Amonida Zadissa¹, Myrto Kostadima¹, Fergal J. Martin¹, Matthieu Muffato¹, Emily Perry¹, Magali Ruffier¹, Daniel M. Staines¹, Stephen J. Trevanion¹, Fiona Cunningham¹, Andrew Yates¹, Daniel R. Zerbino¹ and Paul Flicek¹,²,*

¹European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK and ²Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

Funding
Training materials

- Ensembl training materials are protected by a CC BY license
- [http://creativecommons.org/licenses/by/4.0/](http://creativecommons.org/licenses/by/4.0/)
- If you wish to re-use these materials, please credit Ensembl for their creation
- If you use Ensembl for your work, please cite our papers
- [http://www.ensembl.org/info/about/publications.html](http://www.ensembl.org/info/about/publications.html)