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

http://training.ensembl.org/events
Browsing Genes and Genomes with Ensembl

Astrid Gall, Emily Perry and Erin Haskell
Ensembl Outreach

Webinar series
4, 6, 11 and 13 September 2018
Questions?

- We have muted all of your microphones
- Join our Slack workspace and ask questions (register to get a link sent to you)
- My Ensembl colleagues will respond during the talk
- Please type @username to reply to a specific person

Emily Perry  Erin Haskell
Course Exercises

All course materials including exercises and solutions are here:
http://www.ebi.ac.uk/training/online/course/ensembl-browser-webinar-series-2016

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

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

The “next page” will be the exercises.

http://training.ensembl.org/events
Get Help with the Exercises

- Use the exercise solutions in the online course
- Join our Slack workspace and discuss the exercises with everybody in dedicated channels (register to get a link sent to you)
- Email us: helpdesk@ensembl.org
Introduction to Ensembl

Astrid Gall
Ensembl Outreach Officer, EMBL-EBI

Webinar
4 September 2018
Objectives

You will learn:

• What Ensembl is
• What type of data you can explore in Ensembl
• How to navigate the Ensembl browser website
• Where to find help and documentation
# This Webinar Course

<table>
<thead>
<tr>
<th>Date</th>
<th>Webinar topic</th>
<th>Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>4th Sept</td>
<td><strong>Introduction to Ensembl</strong></td>
<td>Astrid Gall</td>
</tr>
<tr>
<td></td>
<td>Ensembl genes</td>
<td>Emily Perry</td>
</tr>
<tr>
<td>6th Sept</td>
<td><strong>Variation data in Ensembl and the Ensembl VEP</strong></td>
<td>Erin Haskell</td>
</tr>
<tr>
<td></td>
<td>Comparing genes and genomes with Ensembl Compara</td>
<td>Astrid Gall</td>
</tr>
<tr>
<td>11th Sept</td>
<td><strong>Finding features that regulate genes – the Ensembl Regulatory Build</strong></td>
<td>Emily Perry</td>
</tr>
<tr>
<td></td>
<td>Data export with BioMart</td>
<td>Erin Haskell</td>
</tr>
<tr>
<td>13th Sept</td>
<td><strong>Uploading your data to Ensembl</strong></td>
<td>Astrid Gall</td>
</tr>
<tr>
<td></td>
<td>Introduction to the Ensembl REST APIs</td>
<td>Emily Perry</td>
</tr>
</tbody>
</table>

[http://training.ensembl.org/events](http://training.ensembl.org/events)
Structure of this Session

Presentation:
Introduction to Ensembl

Demo:
Part 1: Homepage, assemblies and species
Part 2: Region in detail view

Exercises:
Available on the train online website

http://training.ensembl.org/events
Outline

• Genome browsers
• Ensembl features and access scales
• Ensembl and Ensembl Genomes
• Release Cycle
• Genome Assembly
Why do we need Genome Browsers?

1977: 1st genome sequenced (5 kb)
2004: Human genome sequence finished (3 Gb)
Why do we need Genome Browsers?

http://training.ensembl.org/events
Genome Browsers unlock the Code

https://www.ensembl.org


https://genome.ucsc.edu

http://training.ensembl.org/events
Ensembl Features

- Genomes and gene builds for >100 species
- Variation data
- Alignments, gene trees, homologues
- Regulatory build (ENCODE)

- BioMart (data export)
- Tools for data processing, e.g. VEP
- Display your own data
- Programmatic access via APIs
- Completely Open Source (FTP, GitHub)

http://training.ensembl.org/events
Access Scales

One by one

Main browser
Mobile site

BioMart
REST API
VEP

Perl API
MySQL

Groups

FTP

Whole genome

http://training.ensembl.org/events
Ensembl - 100+ Vertebrate and Model Organism Genomes

http://training.ensembl.org/events
Ensembl Genomes - Extending Ensembl across the Taxonomic Space

- Vertebrates
- Bacteria
- Fungi
- Metazoa
- Plants
- Protists
- Other representative species

http://training.ensembl.org/events
# Ensembl and Ensembl Genomes

<table>
<thead>
<tr>
<th>Released</th>
<th>Ensembl</th>
<th>Ensembl Genomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td></td>
<td>2009</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>Ensembl</th>
<th>Ensembl Genomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertebrates (fly, worm and yeast as outgroups)</td>
<td>By Ensembl</td>
<td>Non-vertebrates (protists, plants, fungi, metazoa, bacteria)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Annotation</th>
<th>Ensembl</th>
<th>Ensembl Genomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>In collaboration with the scientific communities</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>URL</th>
<th>Ensembl</th>
<th>Ensembl Genomes</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.ensembl.org">www.ensembl.org</a></td>
<td></td>
<td><a href="http://www.ensemblgenomes.org">www.ensemblgenomes.org</a></td>
</tr>
</tbody>
</table>

http://training.ensembl.org/events
Release Cycle

- New/updated interfaces
- Updated regulation data
- Updated variation data
- Comparative Genomics including new genes and genomes
- New genome assemblies
- Underlying software updates
- Updated gene sets
- ~3 months
How does a Genome Assembly work?

**Sequence reads**

CGGCCTTTGGGCTCCGCCTTCAGCTCAAGA
CAGCTGTCCCAGATGAC
ACTTAACCTCCTCCAGCTGTCC
GGGCCTTCGCTTCAGCTC
AACTTCCCTCCCAGCT
GGGCTCCCCAGCTGTCCCAGATGACGCCAT
ACTTAACTTCCCTCCCAGCT
GGGCCTTTGGGCTCC
CAGATGACGCC

**Match up overlaps**

CGGCCTTTGGGCTCCGCCTTCAGCTCAAGA
AACTTCCCTCCCAGCT
CAGATGACGCC
TCCGCTGTCCAAGACTTAACCTTC
GGGCCTTCGCTTCAGCTC
AACTTCCCTCCCAGCT
GGGCTCCCCAGCTGTCCCAGATGACGCCAT
ACTTAACTTCCCTCCCAGCT
GGGCCTTTGGGCTCC
CAGATGACGCC

**Contig**

CGGCCTTTGGGCTCCGCCTTCAGCTCAAGA
ACTTAACCTCCTCCAGCTGTCC
GGGCCTTCGCTTCAGCTC
AACTTCCCTCCCAGCT
GGGCTCCCCAGCTGTCCCAGATGACGCCAT
CGGCCTTTGGGCTCC

http://training.ensembl.org/events
BACs from different individuals assembled, including overlaps: **Tilepath**

Overlaps trimmed to give **contigs**. A run of contigs with no gaps is a **scaffold**.

Genetic maps used to assemble scaffolds into a **chromosome**.

http://training.ensembl.org/events
A **contig** is a contiguous stretch of DNA sequence that has been assembled based on sequencing information.
Human Genome Assemblies

- **GRCh38 (aka hg38)**
  - No gaps. Many rare/private alleles replaced.
  - [www.ensembl.org](http://www.ensembl.org)
  - Most up-to-date and supported

- **GRCh37 (aka hg19)**
  - 250 gaps
  - [grch37.ensembl.org](http://grch37.ensembl.org)
  - Limited data and software updates

- **NCBI36 (aka hg18)**
  - 150,000 gaps
  - [ncbi36.ensembl.org](http://ncbi36.ensembl.org)
  - No longer updated

http://training.ensembl.org/events
Hands on

We will look at the Ensembl homepage and find information about the genome assemblies and species in Ensembl.

We will look at a region of the human genome, 4:122868000-122946000, and manipulate the view to see the data we are interested in.
Questions?

- We have muted all of your microphones
- Join our Slack workspace and ask questions (register to get a link sent to you)
- My Ensembl colleagues will respond during the talk
- Please type @username to reply to a specific person

Astrid Gall  Emily Perry  Erin Haskell

http://training.ensembl.org/events
Course Exercises

All course materials including exercises and solutions are here:
http://www.ebi.ac.uk/training/online/course/ensembl-browser-webinar-series-2016

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

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

The “next page” will be the exercises.
Get Help with the Exercises

• Use the exercise solutions in the online course

• Join our Slack workspace and discuss the exercises with everybody in dedicated channels (register to get a link sent to you)

• Email us: helpdesk@ensembl.org

http://training.ensembl.org/events
# Next Webinar

<table>
<thead>
<tr>
<th>Date</th>
<th>Webinar topic</th>
<th>Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>4th Sept</td>
<td>Introduction to Ensembl</td>
<td>Astrid Gall</td>
</tr>
<tr>
<td></td>
<td><strong>Ensembl genes</strong></td>
<td>Emily Perry</td>
</tr>
<tr>
<td>6th Sept</td>
<td>Variation data in Ensembl and the Ensembl VEP</td>
<td>Erin Haskell</td>
</tr>
<tr>
<td></td>
<td>Comparing genes and genomes with Ensembl Compara</td>
<td>Astrid Gall</td>
</tr>
<tr>
<td>11th Sept</td>
<td>Finding features that regulate genes – the Ensembl Regulatory Build</td>
<td>Emily Perry</td>
</tr>
<tr>
<td></td>
<td>Data export with BioMart</td>
<td>Erin Haskell</td>
</tr>
<tr>
<td>13th Sept</td>
<td>Uploading your data to Ensembl</td>
<td>Astrid Gall</td>
</tr>
<tr>
<td></td>
<td>Introduction to the Ensembl REST APIs</td>
<td>Emily Perry</td>
</tr>
</tbody>
</table>

http://training.ensembl.org/events
Next Webinar

Ensembl Genes

Ensembl provides gene annotation for a selection of vertebrate genomes and for some non-vertebrate model organisms for comparative purposes.

Learn about the methods for gene annotation we use and how you can assess the quality of different transcripts.

We will introduce the system of Ensembl stable identifiers, as well as Gene Ontology to describe protein function.

Starts in 5 minutes.

http://training.ensembl.org/events

Emily Perry
Help and Documentation

Courses https://www.ebi.ac.uk/training/online/
Tutorials www.ensembl.org/info/website/tutorials

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

Email us helpdesk@ensembl.org
Public mailing lists dev@ensembl.org
announce@ensembl.org

http://training.ensembl.org/events
Follow us

www.facebook.com/Ensembl.org

@Ensembl

www.ensembl.info

http://training.ensembl.org/events
Publications

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

Zerbino D. et al
**Ensembl 2018**

Xosé M. Fernández-Suárez and Michael K. Schuster
**Using the Ensembl Genome Server to Browse Genomic Sequence Data.**
*Current Protocols in Bioinformatics* (2010) 30:1.15.1-1.15.48
http://europepmc.org/abstract/MED/20521244

Giulietta M. Spudich and Xosé M. Fernández-Suárez
**Touring Ensembl: A practical guide to genome browsing**
http://europepmc.org/articles/PMC2894802

and topic-specific publications mentioned throughout the workshop

http://training.ensembl.org/events
Ensembl Acknowledgements
The Entire Ensembl Team

Daniel R. Zerbino\textsuperscript{1}, Premanand Achuthan\textsuperscript{1}, Wasiu Akanni\textsuperscript{1}, M. Ridwan Amode\textsuperscript{1}, Daniel Barrell\textsuperscript{1,2}, Jyothish Bhai\textsuperscript{1}, Konstantinos Billis\textsuperscript{1}, Carla Cummins\textsuperscript{1}, Astrid Gall\textsuperscript{1}, Carlos García Giro\textsuperscript{1}, Laurent Gil\textsuperscript{1}, Leo Gordon\textsuperscript{1}, Leanne Haggerty\textsuperscript{1}, Erin Haskell\textsuperscript{1}, Thibaut Hourlier\textsuperscript{1}, Osagie G. Izuogu\textsuperscript{1}, Sophie H. Janacek\textsuperscript{1}, Thomas Juettemann\textsuperscript{1}, Jimmy Kiang To\textsuperscript{1}, Matthew R. Laird\textsuperscript{1}, Ilias Lavidas\textsuperscript{1}, Zhicheng Liu\textsuperscript{1}, Jane E. Loveland\textsuperscript{1}, Thomas Maurel\textsuperscript{1}, William McLaren\textsuperscript{1}, Benjamin Moore\textsuperscript{1}, Jonathan Mudge\textsuperscript{1}, Daniel N. Murphy\textsuperscript{1}, Victoria Newman\textsuperscript{1}, Michael Nuhn\textsuperscript{1}, Denye Ogeh\textsuperscript{1}, Chuang Kee Ong\textsuperscript{1}, Anne Parker\textsuperscript{1}, Mateus Patricio\textsuperscript{1}, Harpreet Singh Riat\textsuperscript{1}, Helen Schuilenburg\textsuperscript{1}, Dan Sheppard\textsuperscript{1}, Helen Sparrow\textsuperscript{1}, Kieron Taylor\textsuperscript{1}, Anja Thormann\textsuperscript{1}, Alessandro Vullo\textsuperscript{1}, Brandon Walts\textsuperscript{1}, Amonida Zadissa\textsuperscript{1}, Adam Frankish\textsuperscript{1}, Sarah E. Hunt\textsuperscript{1}, Myrto Kostadima\textsuperscript{1}, Nicholas Langridge\textsuperscript{1}, Fergal J. Martin\textsuperscript{1}, Matthieu Muffato\textsuperscript{1}, Emily Perry\textsuperscript{1}, Magali Ruffier\textsuperscript{1}, Dan M. Staines\textsuperscript{1}, Stephen J. Trevanion\textsuperscript{1}, Bronwen L. Aken\textsuperscript{1}, Fiona Cunningham\textsuperscript{1}, Andrew Yates\textsuperscript{1} and Paul Flicek\textsuperscript{1,3}

\textsuperscript{1}European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK, \textsuperscript{2}Eagle Genomics Ltd., Wellcome Genome Campus, Hinxton, Cambridge CB10 1DR, UK and \textsuperscript{3}Wellcome 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/
- 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