Browsing Genes and Genomes with Ensembl

Victoria Newman
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.
- Where to go for help and documentation.
# This webinar course

<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 BioMart is, and how to set up a query

Demo:
Using Genemart

Exercises:
On the train online course
Questions?

- We’ve muted all the mics
- Ask questions in the Chat box in the webinar interface
- My Ensembl colleagues will respond during the talk
- There’s no threading so please respond with @name
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 project and will provide an 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
Data Mining with BioMart
What is BioMart?

- A tool in your browser to allow

  - Direct export of Ensembl data from our databases, no programming required
  - Customised query-building with a few mouse clicks
  - Generation of datatables and files tailored to downstream application
Why use BioMart?

For things that would be time consuming/difficult with the Ensembl browser

- Query multiple things (genes / variants / regulatory features) at once
- Export large amounts of data
Availability

Ensembl:
main site, GRCh37, archived browser versions

Ensembl Plants:
main site, archive

Ensembl Fungi (some exceptions)

Ensembl Metazoa

Ensembl Protists (some exceptions)
Where to find BioMart

- www.ensembl.org/biomart/martview

- plants.ensembl.org/biomart/martview
- fungi.ensembl.org/biomart/martview
- metazoa.ensembl.org/biomart/martview
- protists.ensembl.org/biomart/martview
How do I use BioMart?

The 4 steps

1. Dataset
   - choose database & species

2. Filters
   - narrow down the dataset

3. Attributes
   - what to print in your table

4. Results
   - table/sequences
Step 1: Dataset

- Define the database that you want to search with your filters
  - Genes, Variation, Regulation

- Define the species
Step 2: Filters

Define the parameters of your query, e.g.:
- A genomic region
- A phenotype
- A set of genes or transcripts
- Function (GO term)
Step 2: Filters

Dataset → Filters → Attributes → Results

Get Attributes for these
Step 3: Attributes

Define the data you want for that set, e.g.:

- IDs
- Genomic coordinates for exons, transcripts, genes
- Sequences
- Orthologues/Paralogues
Step 4: Results

View and download the datatable in a number of formats:

- html
- tsv
- csv
- xls

Sequences can be viewed and downloaded in fasta format
Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data using the statistical programming language R.

Package for Biomart called BiomaRt:

Easy to install in R:
- `source("http://bioconductor.org/biocLite.R")`
- `biocLite("biomaRt")`

Documentation:
- We’re going to look at the set of six *Homo sapiens* genes *ESPN*, *MYH9*, *USH1C*, *CISD2*, *THRB* and *GIPC3*, and find out:
  - Their function via GO terms
  - The length of the associated transcripts
  - Their cDNA sequences
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

http://www.ebi.ac.uk/bmoore/workshops/
Follow us

www.facebook.com/Ensembl.org

@Ensembl

www.ensembl.info

http://www.ebi.ac.uk/bmoore/workshops/
Publications

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

Aken, B. et al.
Ensembl 2017
Nucleic Acids Research
http://europepmc.org/abstract/med/27899575

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

http://www.ebi.ac.uk/bmoore/workshops/
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 Juettmann¹, 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 Thomann¹, 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,*
Training materials

- Ensembl training materials are protected by a CC BY license
  ![CC BY license](https://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)