

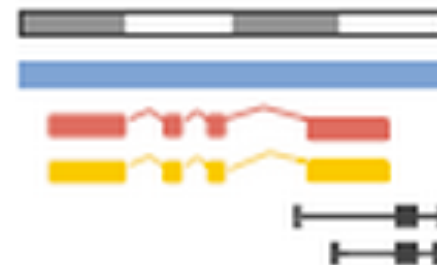
Module 5

Advanced Genome Browsing

Using Ensembl BioMart and
uploading your own data to view in
Ensembl



Emily Perry
Ensembl Outreach Team
EMBL-EBI



This session

- Data mining with Ensembl BioMart
- Uploading/attaching files for viewing in Ensembl
- Exercises

Course materials

<http://www.sanger.ac.uk/resources/talksandtraining/opendoor/hinxton.html>

- Presentations
- Coursebook
- Coursebook page 135-148

Data Mining with BioMart

Outline of this session

- What is BioMart?
- What can I use it for?
- The principle: 4 steps

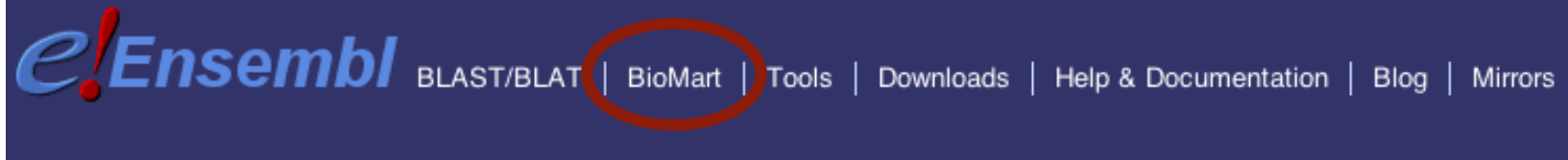
What is BioMart?

- Data export tool with no programming required
- Quick table generator (eg Excel)
- Web interface to mine Ensembl data



Where to find BioMart

- central.biomart.org
- www.ensembl.org/biomart/martview



What can I do with BioMart?

Define a set of genes/variants/regulatory features by combinations of parameters, eg:

- A region
- A list of IDs
- Function
- Phenotypes

Export data for that set, eg:

- IDs
- Sequences
- Homologues

4 steps

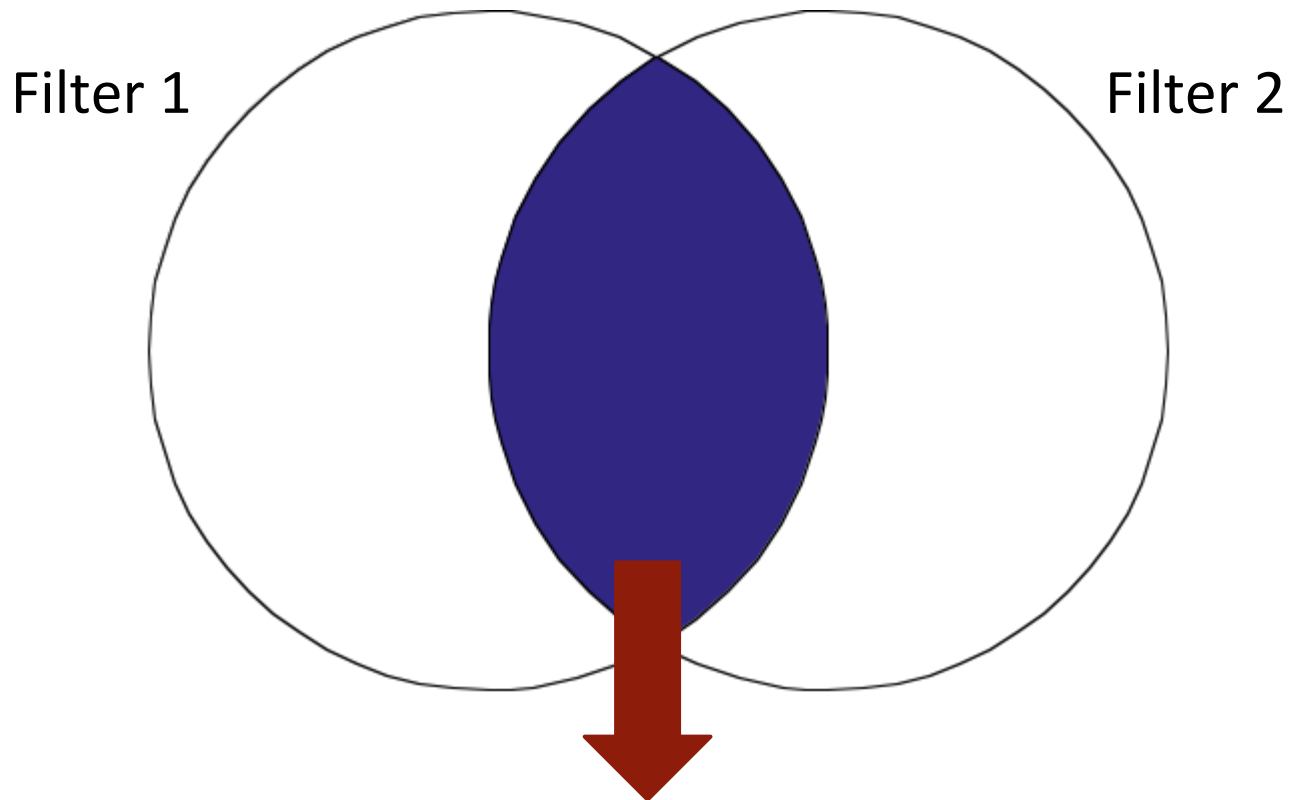
Dataset
choose
database
& species

Filters
what you
know

Attributes
what you want
to know

Results
table/
sequences

Filters



Get Attributes
for these

Hands on

- We're going to look at a set of six genes: *ESPN*, *MYH9*, *USH1C*, *CISD2*, *THRB* and *DFNB31* and find out:
 - Their EntrezGene IDs
 - Their function via GO terms
 - Their cDNA sequences
- Demo: page 136-139

Visualise your own data

Visualise your own data

- Upload data
 - Data saved by Ensembl
 - 5 MB limit (so not possible for large file formats)
- Attach remote file
 - URL-based (http or ftp)
 - No size limit

<http://www.ensembl.org/info/website/upload/index.html>

<http://www.sanger.ac.uk/resources/talksandtraining/opendoor/hinxton.html>

Visualise your own data – Pros and Cons

- Tracks can be saved (to your account)
- Tracks can be shared with other users
- Only trivial security, do not use for sensitive data

Data visualisation formats

BedGraph	continuous	values
GBrowse	genes/features	
GFF/GTF	genes/features	
PSL	sequence alignments	
WIG	continuous	values
BAM	sequence alignments	
BigBed	genes/features	
BigWig	continuous	values
VCF	variants	

<http://www.ensembl.org/info/website/upload/index.html#formats>

<http://www.sanger.ac.uk/resources/talksandtraining/opendoor/hinxton.html>

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



Facebook www.facebook.com/Ensembl.org



Twitter <https://twitter.com/Ensembl>



Blog www.ensembl.info

Publications

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

- Flicek, P. *et. al.*
Ensembl 2013
Nucleic Acids Res **41**:D48-55(2013)
<http://nar.oxfordjournals.org/content/41/D1/D48.long>
- 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
BMC Genomics 11:295 (2010)
www.biomedcentral.com/1471-2164/11/295

<http://www.sanger.ac.uk/resources/talksandtraining/opendoor/hinxton.html>

Ensembl 2014



Acknowledgements

Ensembl 2015

Fiona Cunningham¹, M. Ridwan Amode¹, Daniel Barrell^{1,2}, Kathryn Beal¹, Konstantinos Billis¹, Simon Brent², Denise Carvalho-Silva¹, Peter Clapham², Guy Coates², Stephen Fitzgerald¹, Laurent Gil¹, Carlos García Girón¹, Leo Gordon¹, Thibaut Hourlier¹, Sarah E. Hunt¹, Sophie H. Janacek¹, Nathan Johnson¹, Thomas Juettemann¹, Andreas K. Kähäri², Stephen Keenan¹, Fergal J. Martin¹, Thomas Maurel¹, William McLaren¹, Daniel N. Murphy^{1,2}, Rishi Nag¹, Bert Overduin¹, Anne Parker¹, Mateus Patricio¹, Emily Perry¹, Miguel Pignatelli¹, Harpreet Singh Riat¹, Daniel Sheppard¹, Kieron Taylor¹, Anja Thormann¹, Alessandro Vullo¹, Steven P. Wilder¹, Amonida Zadissa¹, Bronwen L. Aken¹, Ewan Birney¹, Jennifer Harrow², Rhoda Kinsella¹, Matthieu Muffato¹, Magali Ruffier¹, Stephen M.J. Searle², Giulietta Spudich¹, Stephen J. Trevanion¹, Andy Yates¹, Daniel R. Zerbino¹ and Paul Flicek^{1,2,*}

Funding

welcometrust

European Commission
Framework Programme 7



EMBL



National
Human Genome
Research Institute



BBSRC
bioscience for the future

EURATRANS

**HOLIX
NOBULA**
THE SCIENCE CLOUD

Hands on

- We're going to attach a BAM file from an RNASeq experiment to view in Ensembl
- Demo: page 140-142
- Exercises: page 142-144
 - Answers: page 144-148