Introduction to the Ensembl Genome Browser - Virtual Workshop

26th January 2021 - 28th January 2021

https://tinyurl.com/EnsemblVirtualWorkshop

helpdesk@ensembl.org

Welcome to the ‘Living Document’ for this webinar series. Everyone who is registered for the course has access to edit this document. Each of the six modules has a page where you can make notes and record any questions. Please feel free to use this document to ask questions to the Ensembl team between the webinars. The Ensembl team will check the document periodically throughout the course to answer the questions within the document. If you wish to ask questions privately, please do not hesitate to contact the Ensembl Helpdesk.

The ‘Living Document’ is a great way of capturing the knowledge exchanged during the course and saving it for future use by yourselves and those who can’t attend this course. Remember - you don’t have to contribute, but any additions will be welcomed!

The course learning outcomes are:

- view genomic regions and manipulate the view to add features.
- explore information about genes and their sequences.
- analyse genomic variants and associated phenotypes and your own variation data using the VEP.
- view homologous genes and genomic regions.
- view functional elements involved in gene regulation and their activity in different cell types.
- export gene data in bulk using BioMart.

Course Overview

Workshop Slides and Demo available to download:
http://training.ensembl.org/events/2021/2021-01-26-VirtualBrowser

Feedback survey:
Ensembl virtual browser workshop, 26th-28th January 2021 Course Survey

We would really appreciate it if you could share your thoughts with us regarding these sessions. We are interested in your opinions, how you feel the experience has benefited you and how it could be improved. If you could find a few minutes to complete a short survey at the end of the last session it would really help us in improving the training we can deliver.

The golden rules for efficient online training
We suggest everyone follows these few simple rules for the course to run as smoothly as possible:

- Mute all your microphones - only unmute when tutor asks you to do that
- If you have questions during the course:
  - If you would like to ask a question to all the class: type your question in this shared reference document. Avoid typing in the Zoom chat as it is difficult to keep track there

When you are ready to move on from a practical/exercises please click the yes button ✋ Raise Hand so that the tutor knows you are ready and can proceed with the course. Please remember there are more exercises than we have time for. You are not expected to complete all of them, but rather pick and choose the ones most relevant to your work and you are welcome to finish them in your own time.

Module 1: Introduction to Ensembl

Resources:
www.ensembl.org - Ensembl genome browser (chordates)
grch37.ensembl.org - Ensembl archive for browsing data associated with the human GRCh37 genome assembly
www.ensemblgenomes.org - Ensembl Genomes genome browser (non-chordates)
www.ebi.ac.uk - EMBL-EBI website
www.ebi.ac.uk/services - EMBL-EBI databases and tools
Train Online - EMBL-EBI e-learning for our databases and tools

Recording available:
https://www.youtube.com/watch?v=musm97ACrVs&list=PLqB8Yx1tGBMb8jn57vpEGIPh1RfTRqQko&index=1

Notes:

Genomes outside vertebrates can be accessed by visiting www.ensemblgenomes.org

Valya Vassileva: Panda Genome ailMel1; Base Pairs2,299,509,015; Coding genes19,343

Questions:
Lidia Feliubadaló: Which version of gnomAD is shown by default?
RE Lidia. The population genetic tables display data from gnomAD r3.0, e.g. https://www.ensembl.org/Homo_sapiens/Variation/Population?db=core;r=1:230709548-230710548;v=rs699;vdb=variation;vf=179
However, the VEP web uses gnomAD r2.1 exomes: https://www.ensembl.org/info/docs/tools/vep/script/vep_cache.html
Lidia: Thank you!

William Michels, Ph.D. (question posted in Zoom Chat): The rat genome assembly has recently been updated to mRatBN7.1 (generated by the Darwin Tree of Life Project at the Wellcome Sanger Institute). When might Ensembl update rat to mRatBN7.1 ? Thank you!
[**https://rgd.mcw.edu/wg/news/12-10-announcing-the-release-of-mratbn7/**]
Hi William - the mRatBN7.1 assembly and full annotation will be included in the Ensembl 105 release (predicted for around Summer/Autumn time). However, we hope to have the mRatBN7.1 assembly and gene annotation in the Ensembl rapid release site by the end of February:

https://rapid.ensembl.org/index.html

Thank you! (reply from W. Michels)

The Rapid Release site is a new, lightweight Ensembl genome browser that allows us to make basic genome assembly and annotation data available as quickly as possible. More information can be found here:

https://rapid.ensembl.org/info/about/release_cycle.html

Module 2: Genes and Transcripts

Recording available:
https://www.youtube.com/watch?v=hC7mUY6cFVw&list=PLqB8Yx1tGBMb8jn57vpEGIPh1RfTRqQko&index=2

Notes:

Questions:

I copied and emailed the link in RD1-d to myself and the copied version did not have uniProt or repeats tracks. Not sure why?

RE: Hi. That's strange. Can you see extra Uniprot and repeats tracks here: https://nov2020.archive.ensembl.org/Homo_sapiens/Share/7070a3070efba240c8f2cc20f5f1fc20?redirect=no;mobileredirect=no

Yep both there. Not a problem, I will sort it out but may come back to you if there is still a problem on Friday.
Thanks

Module 3: Variation and the Variant Effect Predictor (VEP)

Recording, Slides and Demo available to download:
https://www.youtube.com/watch?v=6MltwC-ydN0&list=PLqB8Yx1tGBMb8jn57vpEGIPh1RfTRqQko&index=3

Notes:

Questions:

I am running a relatively large VEP analysis on data retrieved from TGex. 950 variants in a single patient. When I get the analysis I download as VCF, VEP and txt but the VEP and txt when opened in excel stop at column M and just put all the remaining columns in column N. I think it is just too large a file to open in excel. If I filter in Ensembl the smaller file opens fine. Is this something you have come across previously? Andy Whatmore

Emily: VCFs are really not intended to be opened in Excel
Andy: vcf does not open well in excel but the txt file should open fine?

Module 4: Comparative Genomics

Recording available:
https://www.youtube.com/watch?v=_F2AqireuDg&list=PLqB8Yx1tGBMb8jn57vpEGIPh1RfTRqQko&index=4

Notes:

Questions:

Arun: Why are pan-taxomonic compara options greyed out for some genes in Ensembl Metazoa?
RE: It means the data for those genes is unavailable. It might be that those genes do not have any homologues.

Arun: In (vertebrate) Ensembl, in the gene tree and orthologs results for a gene, is there a way to include orthologs from Ensembl Metazoa species as well? (More generally, I often have to compare genes for species across vertebrate Ensembl and Ensembl Metazoa, e.g. find orthologs across bilaterian animals for a gene, but I haven't found a way to do that yet with Ensembl.)
RE: I'm afraid it's not possible as we only perform comparative analyses within divisions. There are however few metazoan model organisms in Ensembl for comparison.

Arun: Do you mean species like *Ciona intestinalis*, for example, which appears in Ensembl rather than Ensembl Metazoa?
RE: Yes, Ciona, Drosophila and C. elegans.

Module 5: Regulation

Recording available:
https://www.youtube.com/watch?v=sxk_ayNE-XI&list=PLqB8Yx1tGBMb8jn57vpEGIPh1RfTRqQko&index=5

Notes:

Questions:

From W. Michels, PhD: The left side of the BPTF search page says "Restrict category to... Gene: 8". Are there 8 BPTF gene links?
http://uswest.ensembl.org/Homo_sapiens/Search/Results?q=BPTF;site=ensembl;facet_species=Human
Emily: the search is text-based and searches the descriptions as well as the names, so it will find things like BPTF-antisense too. Limit to Genes too to see the 8
W. Michels, PhD: Thank you Emily! I should have clicked that link:
http://uswest.ensembl.org/Human/Search/Results?q=BPTF;site=ensembl;facet_species=Human;page=1;facet_feature_type=Gene

Nechama: HI is there a way to download regulatory elements for a gene to be able to then compare to a disease set I have?
Emily: a set of genes? A set of variants?
Nechama: for example, i want all the regulation elements for BRCA1 gene as a reference for 'healthy normal' BRCA1. I have my disease cohort I want to check against all kinds of regulatory elements
Emily: The next module is BioMart, and there are options in BioMart to download all the features, motifs or binding peaks for a locus (perhaps your gene +/- 5kb). Another option is that you can get the reg-feats and TF-motifs in your VEP output, so if you want to input your patient variants into the VEP, you can find out if they affect reg-feats and motifs in your output.
Nechama: awesome thank you Emily!

Module 6: BioMart and Custom Data Visualisation

Recording available:
https://www.youtube.com/watch?v=VU4bnFwZIM0&list=PLqB8Yx1tGBMb8jn57vpEGIPh1RfTRqQko&index=6

Notes:

Questions:

From W. Michels, PhD: Is there programmatic access to the BioMart API using programming languages such as Ruby and Raku (a.k.a. Perl_6)? Thank you!
As well as the Bioconductor R biomaRt package, RESTful BioMart access is possible using Perl and wget:
https://www.ensembl.org/info/data/biomart/biomart_restful.html

From Sibel Karchner: I have a general question about annotations. Would it be possible to merge the annotation info from the NCBI with the Ensembl data? For example, lots of zebrafish genes are annotated in the NCBI, but they don't have a gene name or description in Ensembl. I mean mostly for gene naming.
Ben: Hi Sibel - Ensembl cross references our annotations with NCBI and will display NCBI gene names and links to NCBI in the external references section in the gene tab:
https://www.ensembl.org/Danio_rerio/Gene/Matches?db=core;g=ENSDARG00000024771;r=18:5213338-5227420;t=ENSDART00000033574

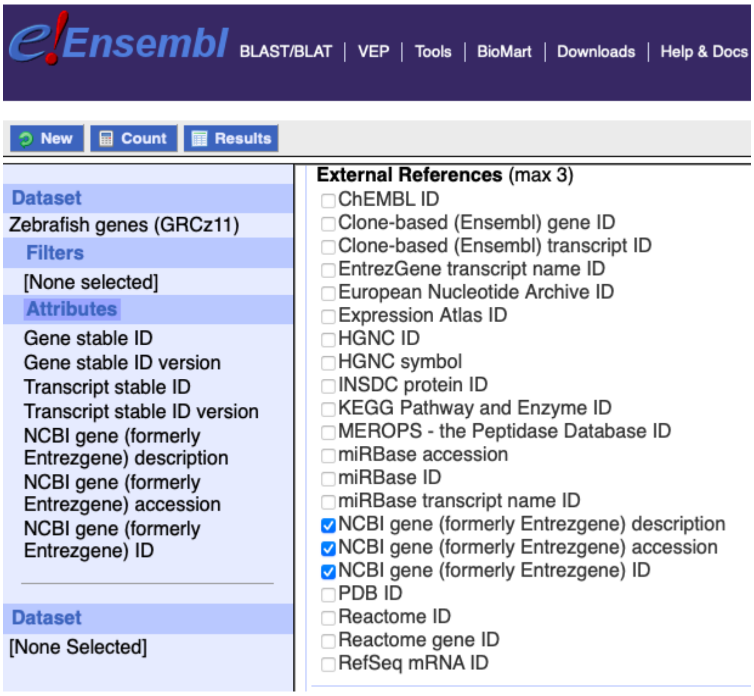
You can also retrieve this data using BioMart. However, NCBI and Ensembl annotations are not always identical, and so there may be genes annotated in Ensembl that do not have a direct cross reference with an NCBI gene. E.g:
https://www.ensembl.org/Danio_rerio/Gene/Matches?db=core;g=ENSDARG00000100084;r=18:5357216-5405137;t=ENSDART00000169896

You can also use ZFIN to find information from both the NCBI and Ensembl from a gene feature:

http://zfin.org/ZDB-GENE-060510-3#summary

From Sibel Karchner: Can the NCBI gene name and description be added to the Attributes list in BioMart (in addition to the NCBI gene ID)? This would enable one to see if the gene name was annotated by NCBI. Actually I just realized that this is already available! This answers my question. If someone else asks the same question, you could point them to BioMart Attributes - NCBI Gene Description.

Ben: Hi Sibel - I'm glad you found the NCBI description, accession and ID attributes in BioMart. I'll leave your query here for others to see as well a screenshot of where to find the attributes:



Dataset
Zebrafish genes (GRCz11)
Filters
[None selected]
Attributes
Gene stable ID
Gene stable ID version
Transcript stable ID
Transcript stable ID version
NCBI gene (formerly Entrezgene) description
NCBI gene (formerly Entrezgene) accession
NCBI gene (formerly Entrezgene) ID

Dataset
[None Selected]

External References (max 3)
- ☐ ChEMBL ID
- ☐ Clone-based (Ensembl) gene ID
- ☐ Clone-based (Ensembl) transcript ID
- ☐ EntrezGene transcript name ID
- ☐ European Nucleotide Archive ID
- ☐ Expression Atlas ID
- ☐ HGNC ID
- ☐ HGNC symbol
- ☐ INSDC protein ID
- ☐ KEGG Pathway and Enzyme ID
- ☐ MEROPS - the Peptidase Database ID
- ☐ miRBase accession
- ☐ miRBase ID
- ☐ miRBase transcript name ID
- ☑ NCBI gene (formerly Entrezgene) description
- ☑ NCBI gene (formerly Entrezgene) accession
- ☑ NCBI gene (formerly Entrezgene) ID
- ☐ PDB ID
- ☐ Reactome ID
- ☐ Reactome gene ID
- ☐ RefSeq mRNA ID

Other resources

This section of the Living Document provides additional resources that might be useful to you in developing skills and knowledge in the course topic area.

Train online

EMBL-EBI provides an e-learning platform called Train online. Train online provides free courses on Europe's most widely used data resources, created by experts at EMBL-EBI and collaborating institutes. You do not need to have any previous experience of bioinformatics to benefit from this training. We want to help you to become a confident user of our data resources; we are not trying to train you to become a bioinformatician.

Tutorials of interest might include:

- **Ensembl: Quick Tour**
- **Ensembl: Browsing Genomes**
- **Ensembl Genomes (non-chordates): Quick Tour**
- **Ensembl REST API**

Webinar series

The EMBL-EBI training team also run regular webinar series featuring the EBI resources. See the Training pages for more information. You can also catch up on any webinars that you might have missed in Train online or on Youtube.