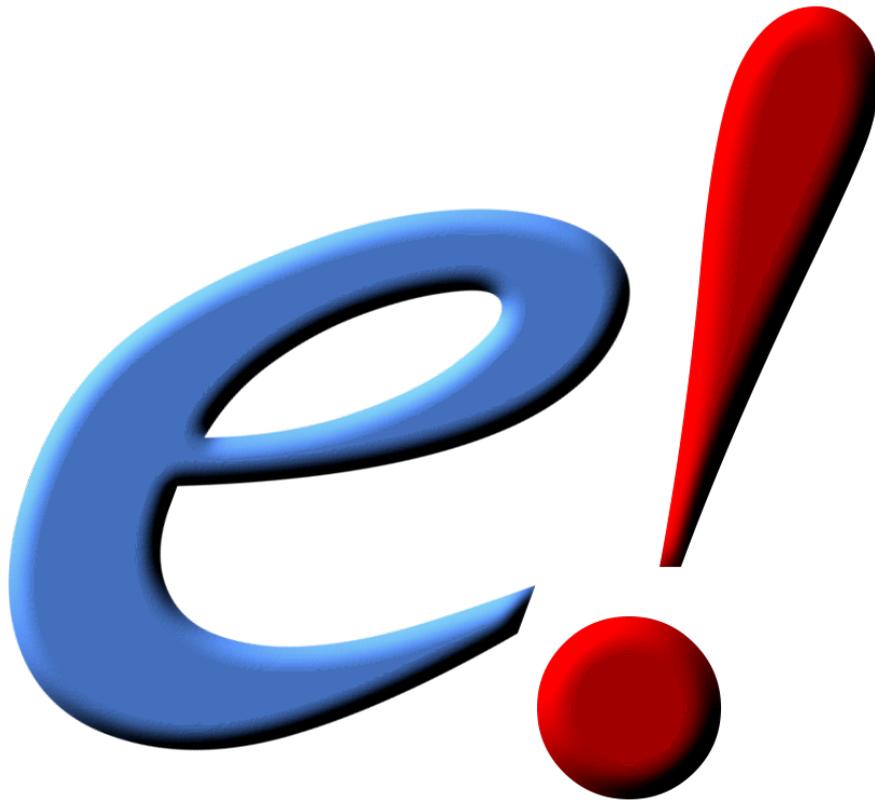**Using Bioinformatics Data and Tools to Engage
Students in Problem Solving:
A Curriculum Development Workshop**

Knoxville, Tennessee, United States

7-10 January 2013

ENSEMBL RESOURCES AT EBI

TUTORIAL ANSWERS

*Note: These exercises are based on Ensembl version 69 (October 2012).
After in future a new version has gone live, version 69 will still be available at
http://e69.ensembl.org.*

_____

BROWSER

_____

Exercise 1

⌁ Go to the Ensembl homepage (http://www.ensembl.org).
⌁ Select 'Search: Human' and type 'long wave sensitive opsin' in the 'for'
text box.
⌁ Click [Go].
⌁ Click on 'Gene' on the page with search results.
⌁ Click on 'Human'.
⌁ Click on 'OPN1LW' or 'ENSG00000102076'.

The official gene symbol for this gene, from the HGNC (Hugo Gene
Nomenclature Committee; http://www.genenames.org), is *OPN1LW*. Note that
'LW' in *'OPN1LW'* stands for 'long wave'.

It is located on the X chromosome, on the forward strand.

⌁ Click on the transcript models.

Ensembl/Havana merge transcripts, that are annotated by both Ensembl
(automatic annotation) and Havana (manual annotation), are shown in gold.
For the *OPN1LW* gene there is one golden transcript, i.e. ENST00000369951
(OPN1LW-001). Note that this is the longest transcript and that it has a CCDS
(Consensus CoDing Sequences; http://www.ncbi.nlm.nih.gov/projects/CCDS)
ID, which means that Ensembl, Havana, NCBI and UCSC all agree about the
CDS of this transcript.

⌁ Click [*help*].

The empty boxes, filled boxes and lines in the transcript models represent the
UTRs (UnTranslated Regions), CDS (CoDing Sequence) and introns,
respectively.

⌁ Click on 'Phenotype' in the side menu.

Yes, according to the cross-references to the Orphanet (http://www.orpha.net)
and OMIM (Online Mendelian Inheritance in Man; http://www.omim.org)
databases, the *OPN1LW* gene has been associated with blue cone

monochromatism / monochromacy, cone rod dystrophy and partial protan colorblindness.

✋ Click on 'Show' in front of 'BLUE CONE MONOCHROMACY' and 'COLORBLINDNESS, PROTAN' in the 'Phenotypes associated with the gene from variation annotations' table.

Yes, blue cone monochromacy and partial protan colorblindness have been associated with variant rs104894912 and rs104894913, respectively.

Exercise 2

✋ Click on 'Comparative Genomics - Paralogues' in the side menu.

There have been nine paralogues identified for the *OPN1LW* gene. *OPN1MW* (ENSG00000147380) and *OPN1MW2* (ENSG00000166160), the genes encoding the medium-wave-sensitive (green) opsins, show the highest Target %id and Query %id. The green opsins thus have the highest sequence similarity to red opsin (Target %id indicates the percentage of the sequence of red opsin matching the sequence of the paralog protein. Query %id indicates the percentage of the sequence of the paralog protein matching the sequence of red opsin).

The gene encoding red opsin as well as the genes encoding the green opsins are located on the X chromosome. As females have two X chromosomes a normal gene on one chromosome can often make up for a defective one on the other, whereas males cannot make up for a defective gene. Thus, red-green colour blindness is much more prevalent in males than in females. Variations in the red and green opsin genes can cause subtle differences in colour perception as well as red-green colourblindness (see Exercise 5). Tandem rearrangements due to unequal crossing-over between these genes also can cause serious defects in colour vision.

✋ Click on 'Gene Tree (image)' in the side menu.
✋ Click on 'View option - View paralogs of current gene' below the Gene Tree display.
✋ Click on the various duplication nodes (shown in red).

A duplication event on the level of the Hominines / Homininae (humans, gorillas and chimpanzees; http://en.wikipedia.org/wiki/Homininae) has given rise to the red and green opsin genes. The other paralogs are due to earlier duplication events. This agrees with the fact that the green opsins show the highest sequence similarity with the red opsin, as seen on the 'Paralogues' page. The two green opsin genes themselves are the result of a very recent duplication in human.

Exercise 3

✆ Click on 'ENST00000369951' in the list of transcripts.

Transcript ENST00000369951 has six exons. Both the first and the last exon are partially untranslated, as can be seen from the partially open boxes at the 5' and 3' ends. The nucleotide sequence of the transcript is shown on the 'Sequence - Exons' page. On this page the UTRs are shown in purple.

✆ Click on 'External References – General identifiers' in the side menu.

Yes, ENST00000369951 is cross-referenced to the UniProtKB/Swiss-Prot database. The UniProtKB/Swiss-Prot ID for the cross-referenced protein is OPSR_HUMAN and the recommended name 'Long-wave-sensitive opsin 1'. Note that the 'R' in OPSR_HUMAN stands for 'Red'. Obviously, there are also an OPSG_HUMAN (green opsin) and an OPSB_HUMAN (blue opsin).

✆ Click on 'Ontology – Ontology table' in the side menu.

Yes, quite some of the 'biological process' and 'molecular function' GO terms hint at a role of red opsin in vision, e.g. 'visual perception', 'phototransduction', 'protein-chromophore linkage' and 'photoreceptor activity'.

According to the 'cellular component' GO terms, red opsin is located in the plasma membrane.

✆ Click on 'Protein Information – Protein summary' or 'Protein Information – Domains & features' in the side menu.

Yes, red opsin contains seven transmembrane domains.

Exercise 4

✆ Click on the 'Location: X:153,409,698-153,424,507' tab.

The *OPN1LW* gene is located on cytogenetic band q28.

✆ Zoom out using the zoom tool at the upper right top of the main panel until both the *OPN1LW* and the *OPN1MW* and *OPN1MW2* genes are shown.

… or …

✆ Draw with the cursor a box around the *OPN1LW* and the *OPN1MW* and *OPN1MW2* genes in the upper panel.
✆ Click on 'Jump to region' in the pop-up menu.

🖰 Click [Configure this page] in the side menu.
🖰 Type 'clone' in the 'Search display' text box on the pop-up screen.
🖰 Select 'the 'Tilepath', '1Mb clone set', '30k clone set' and '32k clone set'.
🖰 Click (✓).

Yes, there is one clone, i.e. RP11-330B02, that contains the *OPN1LW* and the *OPN1MW* and *OPN1MW2* genes.

🖰 Click [Manage your data] in the side menu.
🖰 Click on 'Attach Remote File'.
🖰 Enter the URL of the file in the 'File URL' text box.
🖰 Select 'Data format: GTF'.
🖰 Enter 'segdups in the 'Name for this track' text box.
🖰 Click [Next›].
🖰 Click (✓).

A new track, named 'segdups', has now been added to the display.

Yes, the region containing the *OPN1MW* and *OPN1MW2* genes has been (recently) duplicated.

Exercise 5

🖰 Click on the 'Transcript: OPN1LW-001' tab.
🖰 Click on 'Protein Information - Variations' in the side menu.

The dbSNP reference cluster IDs (rs#) of the Gly338Glu (G338E) and the Ser180Ala (S180A) variants are rs104894913 and rs949431, respectively.

The alleles of rs104894913 are G/A, the alleles of rs949431 are T/G.

🖰 Hover over the SIFT and PolyPhen scores.

According to SIFT the Gly/Glu change is predicted to be 'deleterious' (score 0) and the Ser/Ala change 'tolerated' (score 1). According to PolyPhen the Gly/Glu change is predicted to be 'probably damaging' (score 1) and the Ser/Ala change 'benign' (score 0.011). This agrees with the fact that the Gly/Glu change abolishes absorbance of the red opsin, resulting in red-green colourblindness, while the Ser/Ala change only slightly seems to affect its absorption maximum, resulting in a subtle perceptual difference in red-green colour vision.

Exercise 6

🖰 Click the 'BioMart' link on the toolbar.
🖰 Choose the 'Ensembl Genes 69' database.
🖰 Choose the 'Homo sapiens genes (GRCh37.p8)' dataset.

🖰 Click on 'Filters' in the left panel.
🖰 Expand the 'REGION section by clicking on the + box.
🖰 Select 'Chromosome – X'.
🖰 Expand the 'GENE ONTOLOGY section by clicking on the + box.
🖰 Enter 'visual perception' in the 'GO Term Name' text box. Make sure the check box in front of the filter is ticked, otherwise the filter won't work.
🖰 Click the [Count] button on the toolbar.

🖰 Click on 'Attributes' in the left panel.
🖰 Expand the 'GENE' section by clicking on the + box.
🖰 Deselect 'Ensembl Transcript ID'.
🖰 Select 'Description'.
🖰 Expand the 'EXTERNAL' section by clicking on the + box.
🖰 Select 'HGNC symbol'.

🖰 Click the [Results] button on the toolbar.
🖰 Select 'View All rows as HTML.
🖰 Check 'Unique results only'.

… or …

🖰 Select 'Export all results to File – XLS'.
🖰 Check 'Unique results only'.
🖰 Click [Go].

Your query should give you 13 out of a total of 60684 genes.

Yes, the *OPN1LW*, *OPN1MW* and *OPN1MW2* genes are all amongst the results.