

EXERCISES and ANSWERS

(NOTE: please use release 42 (from the archive site) as the answers may have changed in newer releases. Thank you.)

1. Exploring features related to a gene

(a) Search for the human TAC1 gene by typing 'human TAC1 gene' in the search window.

(b) How many transcripts are predicted for this gene? What is the size of the longest predicted mRNA? How many exons does it have? How many amino acids does it code for?

(c) Look up 'Similarity Match' in the glossary:

Follow the 'Help and Documentation' link, 'HelpDesk section' to this url:

http://www.ensembl.org/Homo_sapiens/glossaryview

Follow some of the links in the 'Similarity Matches' section of GeneView. What is a possible function of TAC1?

(d) Which InterPro domains does the protein product contain?

(e) Find the GO section of GeneView and follow some of the links to explore the 'Gene ontology' terms (describing gene and protein function) in Ensembl GOView.

(f) In which chromosomal band and on which clone and contig in the genomic sequence assembly is the TAC1 gene located?

(g) Go back to GeneView by clicking on 'TAC1' in the Overview panel and following the link for the gene. Is there a putative mouse orthologue? If so, where is it in the mouse genome?

2. Exploring a region

(a) Display the region between markers D12S764 and D12S1871 in ContigView. Start on the human homepage, and click on chromosome 12.

(b) How many contigs are used to make this portion of the assembly? View the human tile path clones. Do they correspond to the assembly?

(c) What is the closest marker to the TENC1 gene? How many synonyms does this marker have?

(e) Zoom in (towards the '+') three steps on the zoom triangle and turn on the SNP track. Identify an intronic SNP and look at the corresponding SNPView page.

3. Exploring the zebrafish (*Danio rerio*) genome with Ensembl

- (a) Bring up a ContigView display of zebrafish (*Danio rerio*) chromosome 1 between 64.0 Mb and 64.5 Mb.
- (b) How many 'known' and 'novel' genes are predicted in this region? For one of the known genes, find some information about its function, and look at an entry for it in EntrezGene, UniProt/Swiss-Prot or the ZFIN site.
- (c) Can you find out anything about the possible functions of one of the novel genes? For this, try looking at homologues in other species, at other members of protein families and InterPro domains.

Answers (Browsing Ensembl)

1. Exploring features related to a gene

- (a) A 'Vega' gene and 'Ensembl' gene will be shown. VEGA (Vertebrate Genome Annotation) is a consortium of manual curators for certain chromosomes in human, mouse, zebrafish, pig and dog. However, we would like to explore the 'Ensembl Gene: ENSG00000006128'. To ascertain it is indeed the TAC1 gene check that the HGNC symbol (the 'official' gene name given by the HUGO Gene Nomenclature Committee) is 'TAC1'. Click on the 'Ensembl Gene: ENSG00000006128' link to go to the GeneView page for this gene.
- (b) The TAC1 gene (ENSG00000006128) has 3 predicted transcripts, ENST00000319273, ENST00000346867 and ENST00000350485. Scroll down to the 'Transcript' sections for more information about these transcripts. The longest transcript is ENST00000319273. The length of this transcript is 1060 bp. It has 7 exons and codes for 129 aa.
- (c) The TAC1 gene is Protachykinin 1 precursor. Follow the links to MIM and EntrezGene or UniProt/Swiss-Prot in the 'Similarity Matches' section to learn more. Choose 'UniProt' under 'DAS Sources' to see references in the literature (click 'Update' after making the selection). Also the GO (Gene Ontology) and InterPro sections can give you clues about the biological and molecular function of the TAC1 protein. Tachikinins are neuropeptides. These hormones are thought to function as neurotransmitters which interact with nerve receptors and smooth muscle cells. They are known to induce behavioral responses and function as vasodilators and secretagogues.
- (d) Check the 'InterPro' section in GeneView. The domains include IPR013055 (Tachykinin/Neurokinin like), IPR002040 (Tachykinin/Neurokinin), IPR008215 (Tachykinin) and IPR008216 (Protachykinin).
- (e) Clicking on a GO identifier gives you a GOView page (loading of the page can take a while) showing the position of that term in the GO structure (note the number of Ensembl genes mapped to each term). Click [Help] to find out more about GOView.

(f) Go back to GeneView and click the 'Graphical View' link in the side menu to go to ContigView. In the 'Overview' panel you can see that TAC1 is located on band 7q21.3 ('Chr.7 band' track). In the 'Detailed view' panel you can see that it is located on contig AC004140.2.1.74918 ('DNA(contigs)' track). If you click on the contig and follow the link to the EMBL source (or if you turn on the 'Human tile path clones' track from the 'Decorations' menu of ContigView) you can see that this sequence is derived from clone RP5-841B21.

(g) In GeneView, ENSMUSG00000061762 (Tac1) is named in the 'Orthologue Prediction' section. Click on it to go to its GeneView page to find that it is located on mouse chromosome 6.

2. Exploring a region

(a) Start on the homepage for human and click on chromosome 12 to go MapView. In the 'Jump to ContigView' section choose 'From (type): Marker D12S764 To (type): Marker D12S1871' and click [Go]. This leads you to ContigView.

(b) The displayed region in the Overview panel is larger than the area between the two markers. The red line or small box is drawn over the first marker (D12S764). Zoom into the region between the two markers by drawing a box with the mouse around it.

The region will be displayed below, in 'Detailed View'. This region includes sequence from 4 different contigs (one is quite small), displayed in light blue and dark blue in the 'DNA(contigs)' track. To see also the 4 clones that make up this region, select the 'Human tilepath clones' track from the 'Decorations' menu. Clones are shown in gold and pink. Portions of the 'Tile path clones' were used to form the assembly and correspond to 'contigs'. The clones overlap each other whereas the contigs don't.

(c) Marker D12S2110 is closest to the TENC1 gene. Click on the marker (i.e. on the vertical bar representing it, not on its name) and follow the link 'Marker info' to the MarkerView page. There are 2 synonyms listed.

(d) SNPs can be turned on using the 'Features' menu. Coding SNPs are shown in yellow (non-synonymous) and green (synonymous), intronic SNPs are dark blue. Click on a SNP. Be careful to click exactly on the vertical bar representing the SNP, otherwise you will get the wrong pop-up menu. Follow the link 'SNP properties' to the SNPView page. Note the 'SNP Context' display in SNPView.

3. Exploring the zebrafish (*Danio rerio*) genome with Ensembl

(a) Start on the homepage for zebrafish (*Danio rerio*). In the 'Karyotype' section choose 'Chromosome: 1', 'From (type): Base pair: 64000000 To (type) Base pair: 64500000' and click [Go]. This leads you to ContigView for a larger region. Type in the base pairs in Detailed View (change the second number to 64500000).

(b) On the 'Overview' panel of ContigView Ensembl known and novel genes are displayed in the 'Ensembl Genes' track in reddish brown and black, respectively. Known genes are Ensembl gene predictions that match species specific entries in the UniProt and/or RefSeq database, while novel genes map back to entries from other species. There are 12 known and 9 novel genes. Click on one of the known genes to go to its GeneView page and explore the links in the 'Similarity Matches' section.

(c) To find out more about the possible function of a novel gene there are many options. Click on the gene in ContigView to go to its GeneView page. If the gene has orthologues in other species (shown in the 'Orthologue Prediction' section) and these orthologues are better characterized than the novel zebrafish gene this can give a clue about the possible function of this gene. If the gene belongs to a family (shown in the 'Protein Family' section) other family members may provide information. InterPro domains (shown in the 'InterPro' section) may also provide clues.