

EXERCISES and ANSWERS (Ensemble r98)

<https://training.ensembl.org/exercises#questions-and-answers>

1. Exploring features related to a gene

(a) Find the human **MYH9** (myosin, heavy chain 9, non-muscle) gene, and go to the [Gene tab](#).

- On which chromosome and which strand of the genome is this gene located?
- How many transcripts (splice variants) are there and how many are protein coding?
- What is the longest transcript, and how long is the protein it encodes?
- Which transcript is the best quality?

(b) Click on [Phenotype](#) at the left side of the page. Are there any diseases associated with this gene, according to MIM (Mendelian Inheritance in Man)?

(c) What are some functions of MYH9 according to the Gene Ontology consortium? Have a look at the GO pages for this gene.

(d) In the transcript table, click on the [transcript ID](#) for MYH9-201, and go to the [Transcript tab](#).

- How many exons does it have?

- Are any of the exons completely or partially untranslated?
- Is there an associated sequence in UniProtKB/Swiss-Prot? Have a look at the [General identifiers](#) for this transcript.

(e) Are there microarray (oligo) probes that can be used to monitor ENST00000216181 expression?

2. Exploring a genomic region

(a) Go to the region from 31,937,000 to 32,633,000 bp on human chromosome 13. On which cytogenetic band is this region located? How many contigs make up this portion of the assembly (contigs are contiguous stretches of DNA sequence that have been assembled solely based on direct sequencing information)?

(b) Zoom in on the *BRCA2* gene.

(c) [Configure this page](#) to turn on the Tilepath track in this view. What is this track? Are there any Tilepath clones that contain the complete *BRCA2* gene?

(d) Create a [Share](#) link for this display. Email it to your neighbour. Open the link they sent you and compare. If there are differences, can you work out why?

(e) Export the genomic sequence of the region you are looking at in FASTA format.

(f) Turn off all tracks you added to the Region in detail page.

Exercise 3 – Exploring the pig (*Sus scrofa*) genome

- (a) How many coding and non-coding genes does pig have?
- (b) When was the current *Sus scrofa* genome assembly produced and by whom?

ANSWERS

Exercise 1 -Exploring features related to a gene

- (a) Go to the Ensembl homepage (<http://www.ensembl.org>).

Select Search: Human and type *MYH9*. Click [Go](#).

Click on either the Ensembl ID [ENSG00000100345](#) or the HGNC official gene name [MYH9](#).

- Chromosome 22 on the reverse strand.
 - Ensembl has 11 transcripts annotated for this gene, of which three are protein coding.
 - The longest transcript is MYH9-201 and it codes for a protein of 1,960 amino acids
 - MYH9-201 is the best quality transcript, as it has a CCDS associated with it, is TSL:1 and is Golden.
- (b) These are some of the phenotypes associated to *MYH9* according to MIM: autosomal dominant deafness, Epstein syndrome, and Fechtner syndrome. Click on the records for more information.
- (c) The Gene Ontology project

(<http://www.geneontology.org/>) maps terms to a protein in three classes: biological process, cellular component, and molecular function. Meiotic spindle organisation, cell morphogenesis, and cytokinesis are some of the roles associated with MYH9.

(d) Click on [ENST00000216181](#)

- It has 41 exons. This is shown in the Transcript summary or in the left hand side menu Exons.
- Click on the Exons link in this side menu. Exon 1 is completely untranslated, and exons 2 and 41 are partially untranslated (UTR sequence is shown in purple). You can also see this in the cDNA view if you click on the cDNA link in the left side menu.
- P35579 from UniProt/Swiss-Prot matches the translation of the Ensembl transcript. Click on P35579 to go to UniProtKB, or click align for the alignment.

(e) Click on Oligo probes in the side menu.

Probesets from Affymetrix, Agilent, Codelink, Illumina, and Phalanx match to this transcript sequence.

Expression analysis with any of these probesets would reveal information about the transcript. Hint: this information can sometimes be found in the ArrayExpress Atlas: www.ebi.ac.uk/arrayexpress/

Exercise 2- Exploring a genomic region

(a) Go to the Ensembl homepage (<http://www.ensembl.org/>).

Select [Search: Human](#) and type **13:31937000-32633000** in the text box (or alternatively leave the Search drop-down list like it is and type **human 13:31937000-32633000** in the text box).

Click [Go](#).

This genomic region is located on cytogenetic band q13.1. It is made up of ten contigs, indicated by the alternating light and dark blue coloured bars in the Contigs track. Note that some genes are located within different contigs.

(b) Draw with your mouse a box encompassing the *BRCA2* transcripts. Click on [Jump to region](#) in the pop-up menu.

(c) Click [Configure this page](#) in the side menu (or on the cog wheel icon in the top left hand side of the bottom image).

Type **tilepath** in the [Find a track](#) text box.

Select [Tilepath](#).

Click on the [\(i\)](#) button to find out more

The tilepath track shows the BAC clones that the assembly was based upon.

Save and close the new configuration by clicking on [✓](#) (or anywhere outside the pop-up window).

There is not just one clone that contains the complete *BRCA2* gene. The BAC clone RP11-37E23 contains most of the gene, but not its very 3' end (contained in RP11-298P3). This was reflected on the two contigs that make up the entire *BRCA2* gene (the Contigs track is on by default). You may find this easier to see if you highlight the 3' exon on *BRCA2*.

(d) Click [Share this page](#) in the side menu.

Select the link and copy.

Get your neighbour's email address and compose an email to them, paste the link in and send the message.

When you receive the link from them, open the email and click on your link. You should be able to view the page with the new configuration and data tracks they have added to in the Location tab. You might see differences where they specified a slightly different region to you, or where they have added different tracks.

(e) Click [Export data](#) in the side menu. Leave the default parameters as they are.

Click [Next>](#).

Click on [Text](#).

Note that the sequence has a header that provides information about the genome assembly (GRCh38), the chromosome, the start and end coordinates and the strand. For example:

```
>13 dna:chromosome  
chromosome:GRCh38:13:32311910:32405865:1
```

(f) Click [Configure this page](#) in the side menu.

Click [Reset configuration](#).

Click ✓.

Exercise 3 – Exploring the pig (*Sus scrofa*) genome

(a) Select [Pig](#) from the drop down species list, or click on [View full list of all Ensembl species](#), then choose [Pig](#) from the list to go to the species homepage. Click on [More information and statistics](#).

Pig has 21,301 coding genes, 8,971 non coding genes (of which 2,156 are small non coding genes, 36,798 are long non coding genes and 17 are misc non coding genes) and 1,626 pseudogenes.

(b) The Sscrofa11.1 assembly of the pig genome was produced in January 2017 by the Swine Genome Sequencing Consortium (SGSC).