Module 2 - Maps and Genome Sequence

**Aims**

- Introduce the Vega genome browser
- Explain the source of the data in Vega
- Show the different Vega views stressing the differences to the Ensembl views

**Introduction**

The Vertebrate Genome Annotation (Vega) database is a central repository for high quality, frequently updated, manual annotation of vertebrate finished genome sequence. The *Danio rerio* Vega database contains all the finished clones. Unlike the *Danio rerio* Ensembl database, the Vega database only contains high-quality sequence with high-quality manual annotation. The annotation is undertaken in collaboration and synchronisation with the central zebrafish database ZFIN. The implementation of the Vega browser is based on the Ensembl code and so they share many features and functionality. This section gives a brief introduction to the Vega views emphasising the differences with Ensembl. Refer to section 2 for more details on the Ensembl views.

The main Vega page is [http://vega.sanger.ac.uk](http://vega.sanger.ac.uk). One obvious difference between Ensembl and Vega is the background colour. In Ensembl it is yellow whereas in Vega is blue.
Follow the link to the zebrafish database.

This zebrafish page displays all the chromosomes plus two ‘artificial’ ones called AB and U. Chromosome AB groups all the sequenced clones for a PAC library made from the AB strain. Some of these clones have been manually annotated. Chromosome U contains finished clones which have not been placed in the physical map. The lengths do not represent an estimation of the size of the real chromosomes, but the amount of the current finished sequence. Chromosome 20 is top priority and that is reflected by the fact that it is the longest, i.e., the one with the most finished sequence.

MapView and ContigView

Clicking on a chromosome links to the corresponding MapView page.
The regions shaded in grey in the MapView pages indicate segments of the chromosome that have not been annotated yet. Check the difference between chromosomes 20 and 7 in terms of how much sequence has been annotated. As chromosome 20 is a priority one most of the sequence is not greyed, for chromosome 7 the situation is just the opposite.

The ContigView is, like in Ensembl, one of the main pages in Vega. The contents of the ContigView include some data specific to the manual annotation, for example there is a track for polyA signals. In order to facilitate the task of the annotators, the alignments of protein and ESTs is done more aggressively over finished clones than over the assembly. The most important track in the Vega ContigView page is the ‘Zfish transcripts’, the manually annotated transcripts. Observe that shaded regions do not contain this kind of transcript since they have not been annotated yet.

Jump to the ContigView for the region in chromosome 20 from 17925000 to 18135782.
This region contains a transcript labelled jag2. Look for this name in the ‘Overview’ panel.

**GeneView, TransView, ExonView and ProteinView**

Follow the link to the GeneView page from the jag2 transcript jag2-001.
Every annotated gene in Vega has a ZFIN gene entry. Follow the link in the example by clicking on the name of the gene (jag2). Another special feature of the Vega GeneView page is the fields for the authors of the annotation, and
the type of gene. The gene type gives an indication of the confidence of the annotation based on the available evidence, for example:

- a gene has type **known** if it was listed by ZFIN at the moment of the annotation (eventually every annotated gene will have an entry in ZFIN), and
- a gene has type **novel CDS** if its product was similar to, but not identical to, a known protein from zebrafish or another organism.

Transcripts are also classified in several categories as well.

The gene jag2 has been annotated with two transcripts. Follow the link labelled ‘Transcript information’ for the transcript OTTDART00000005844 to open the **TransView** page.

The **ExonView** page for this transcript gives more information about the sequence of the exons and introns and the supporting evidence used in the annotation. Follow the link labelled ‘Exon information & supporting evidence’ to open the ExonView page.
At the bottom of this page there is a diagram showing the supporting evidence for this annotation.

Supporting evidence

In this example there is no evidence for exon 19, indicating that the annotator has 'built' this exon from other evidences such as splice sites, codon bias and ORFs. Compare this situation and what you see in the Ensembl predictions.

The link labelled 'Peptide Information' opens the ProteinView page. This data is generated automatically using the predicted translation in very much the same fashion as done for the Ensembl annotation.

Other views that we discussed in the module for the Ensembl browser are also present in Vega, for example, ExportView to download the data in files.

**Exercises**

1. Open the GeneView page for jag2 and visit the associated links. In particular open the ContigView page showing this gene.
2. Study the differences between the manual annotation for jag2 in Vega and the automatic annotation in Ensembl (see the Ensembl section for an example of how to open the GeneView page for jag2 in Ensembl).

3. One of the differences between the automatic prediction and the manually annotated jag2 is the number of exons and the UTR. Why do you think these data are different?

4. Customise the ContigView page to turn on the track for poly-A signals.

5. Another special track in Vega ContigView is ‘Assembly tags’. This features information on special regions of the clones. These data are entered by the person in charge of finishing the sequence. A region that contains one of these tags is the finished clone ‘AL928990’. Open a ContigView page for this clone in Vega and check the text for the assembly tag.

6. Many clones present in Vega are also placed in the assembly and therefore can be browsed in Ensembl. As Vega is updated more often there might be a difference in the versions. A sequenced clone may go through several updates from its first to its final submission, these are recorded via the version numbers. Check for clone present in the assembly and compare it to one in Vega.