Module 2 – Maps and Genome Sequence

i. The Ensembl Genome Browser

Caveat: At the time of writing this tutorial, Zv6 had not been released with a full gene build yet. All following examples are therefore taken from the Zv5 Ensembl. If you're trying to work through the examples yourself, please be aware of the difference in the scaffold naming ('Zv5_...' versus 'Zv6_...').

Aims

- Explain the source for the data in Ensembl
- Introduce the Ensembl browser
- Show the different Ensembl views with examples

Introduction

Ensembl is a joint project of the European Bioinformatics Institute (EBI) and the Wellcome Trust Sanger Institute, funded mainly by the Wellcome Trust, with additional funding from EMBL and NIH-NIAID. Ensembl provides easy access to genomic information with a number of visualisation tools.

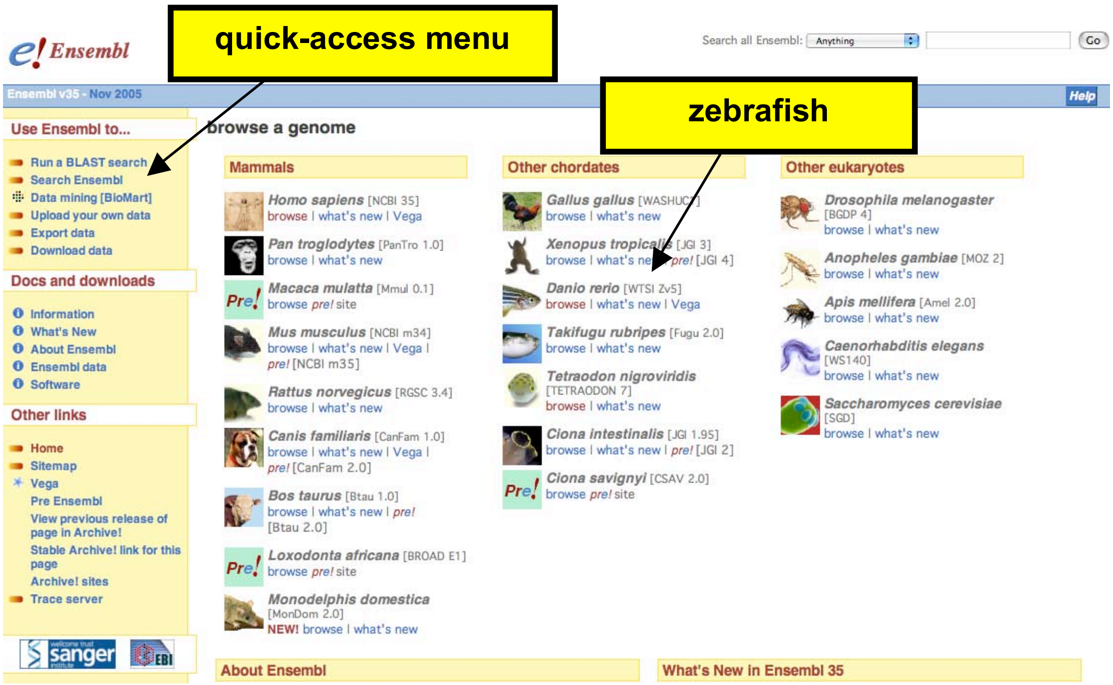
The Ensembl site provides automatic baseline annotation of the latest assembly sequence, including gene, transcript and protein predictions.The annotation is integrated with external data sources, such as ZFIN for the zebrafish site. The latest zebrafish assembly is Zv6, which was released on March 31st, 2006.

The key Ensembl web pages are called Views (e.g. GeneView, TextView, MapView, and ContigView). The Ensembl web site gives you the opportunity to directly download data, whether it is a DNA sequence of a genomic contig you are trying to identify novel genes in, or positions of SNPs in a gene you are working on. There is also an FTP site which you can use to download large amounts of data from the Ensembl database, as well as a data mining tool (BioMart, see section 6) which allows flexible and rapid retrieval of information from the databases. There are many ways you can access the data in Ensembl depending on your needs and these are explained here and in other sections.

The Ensembl site is at:

http://www.ensembl.org

On this page you will find links to all Ensembl species, documentation, search facilities, downloads and other related links. All Ensembl pages have a tool bar on the left-hand side with quick-access links to several resources and facilities.



browse a genome

Mammals

Homo sapiens [NCBI 35]
browse | what's new | Vega

Pan troglodytes [PanTro 1.0]
browse | what's new

Macaca mulatta [Mmul 0.1]
browse *pre!* site

Mus musculus [NCBI m34]
browse | what's new | Vega | *pre!* [NCBI m35]

Rattus norvegicus [RGSC 3.4]
browse | what's new

Canis familiaris [CanFam 1.0]
browse | what's new | Vega | *pre!* [CanFam 2.0]

Bos taurus [Btau 1.0]
browse | what's new | *pre!* [Btau 2.0]

Loxodonta africana [BROAD E1]
browse *pre!* site

Monodelphis domestica [MonDom 2.0]
NEW! browse | what's new

Other chordates

Gallus gallus [WASHUC1]
browse | what's new

Xenopus tropicalis [JGI 3]
browse | what's new | *pre!* [JGI 4]

Danio rerio [WTSI Zv5]
browse | what's new | Vega

Takifugu rubripes [Fugu 2.0]
browse | what's new

Tetraodon nigroviridis [TETRAODON 7]
browse | what's new

Ciona intestinalis [JGI 1.95]
browse | what's new | *pre!* [JGI 2]

Ciona savignyi [CSAV 2.0]
browse *pre!* site

Other eukaryotes

Drosophila melanogaster [BGDP 4]
browse | what's new

Anopheles gambiae [MOZ 2]
browse | what's new

Apis mellifera [Amel 2.0]
browse | what's new

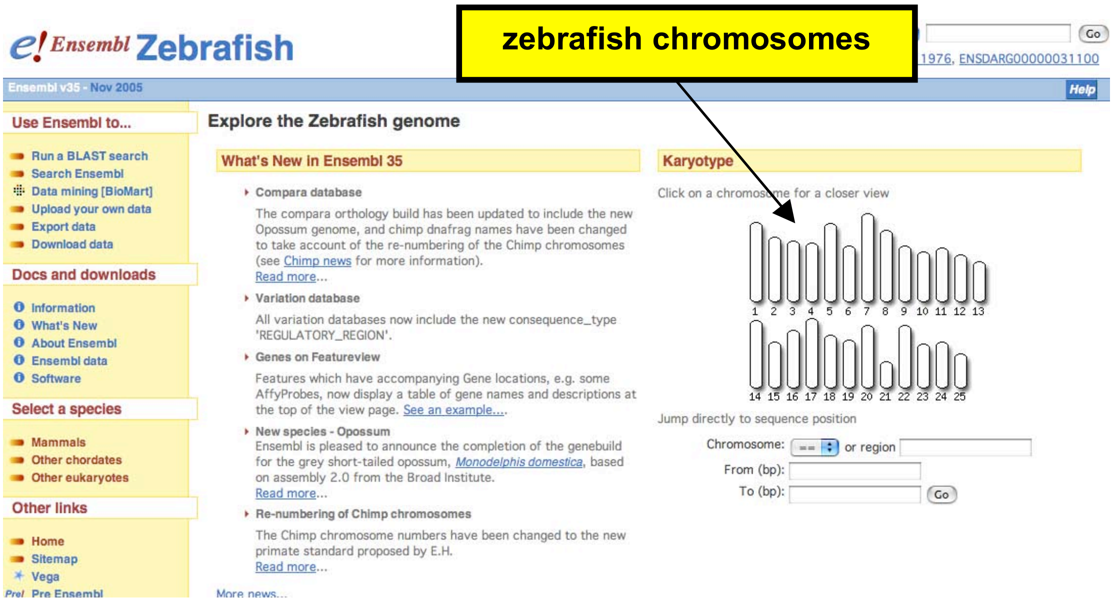
Caenorhabditis elegans [WS140]
browse | what's new

Saccharomyces cerevisiae [SGD]
browse | what's new

About Ensembl

What's New in Ensembl 35

From the main Ensembl site you can access the zebrafish site by clicking on the appropriate species button. As soon a new assembly is released the sequence is made available as a pre-Ensembl site. This includes valuable information such as EST and UniProt alignments and *ab initio* predictions. The main missing data are the Ensembl genes and Ensembl ESTgenes. A full Ensembl dataset for a new assembly is typically made public a couple of months after the assembly release date.



Explore the Zebrafish genome

What's New in Ensembl 35

- **Compara database**
  The compara orthology build has been updated to include the new Opossum genome, and chimp dnafrag names have been changed to take account of the re-numbering of the Chimp chromosomes (see Chimp news for more information).
  Read more...
- **Variation database**
  All variation databases now include the new consequence_type 'REGULATORY_REGION'.
- **Genes on Featureview**
  Features which have accompanying Gene locations, e.g. some AffyProbes, now display a table of gene names and descriptions at the top of the view page. See an example....
- **New species - Opossum**
  Ensembl is pleased to announce the completion of the genebuild for the grey short-tailed opossum, *Monodelphis domestica*, based on assembly 2.0 from the Broad Institute.
  Read more...
- **Re-numbering of Chimp chromosomes**
  The Chimp chromosome numbers have been changed to the new primate standard proposed by E.H.
  Read more...

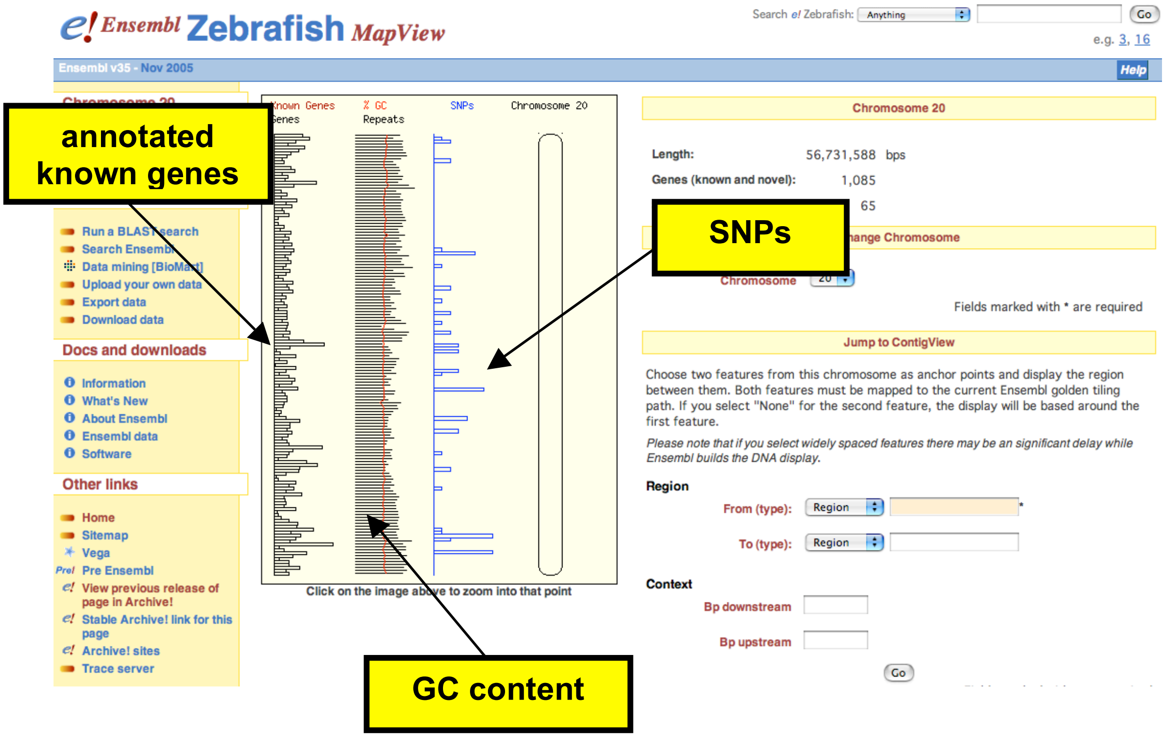
More news...

Click on a chromosome for a closer view

Jump directly to sequence position

MapView and ContigView

This zebrafish Ensembl page provides various access points to the assembly sequence. For example you can browse a particular chromosome. The chromosomes are linked to the **MapView** pages. The figure below shows the MapView for chromosome 20.



Chromosome 20

Length: 56,731,588 bps

Genes (known and novel): 1,085

65

Fields marked with * are required

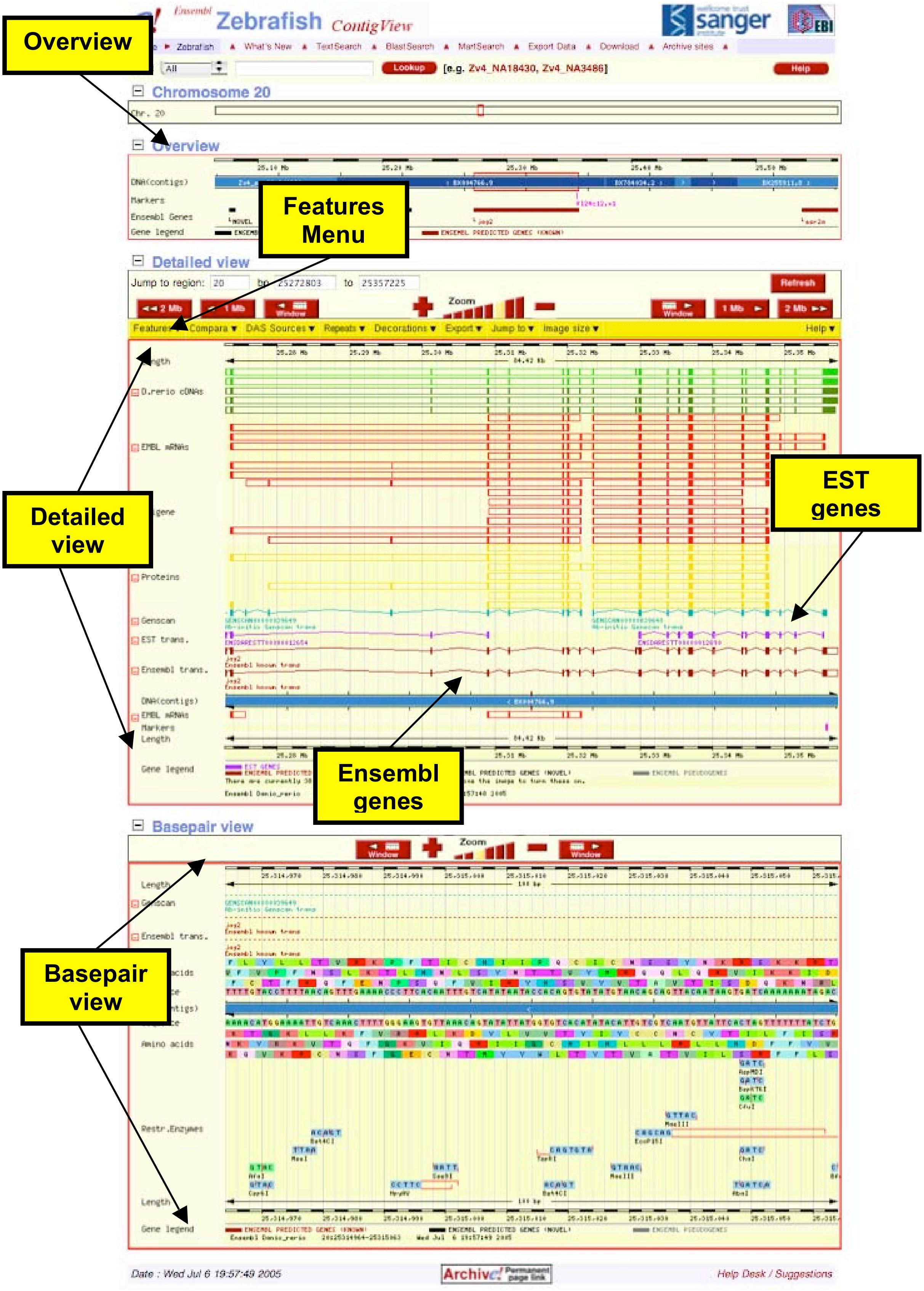
Choose two features from this chromosome as anchor points and display the region between them. Both features must be mapped to the current Ensembl golden tiling path. If you select "None" for the second feature, the display will be based around the first feature.

Please note that if you select widely spaced features there may be an significant delay while Ensembl builds the DNA display.

A MapView page plots the gene and SNP density and GC content. From this page you can zoom in to a more detailed display called ContigView by clicking on the schematic figure representing the chromosome.

ContigView can be considered the central view of the Ensembl web site. It shows the fragments (contigs, clones, etc) that make up a genome assembly. It allows you to scroll along entire chromosomes, whilst viewing the annotated features within a selected region in detail.

A ContigView page is divided into four panels: a chromosome overview, a zoomed-in **overview** of the region in the chromosome you are browsing, a **detailed view** showing features and a **basepair view** that goes down to individual bases. In order to continue with this module, jump to the region under the accession BX004766 (in chromosome 20) with start coordinate 1 and end coordinate 200000. (Use the text box provided to enter these coordinates.)

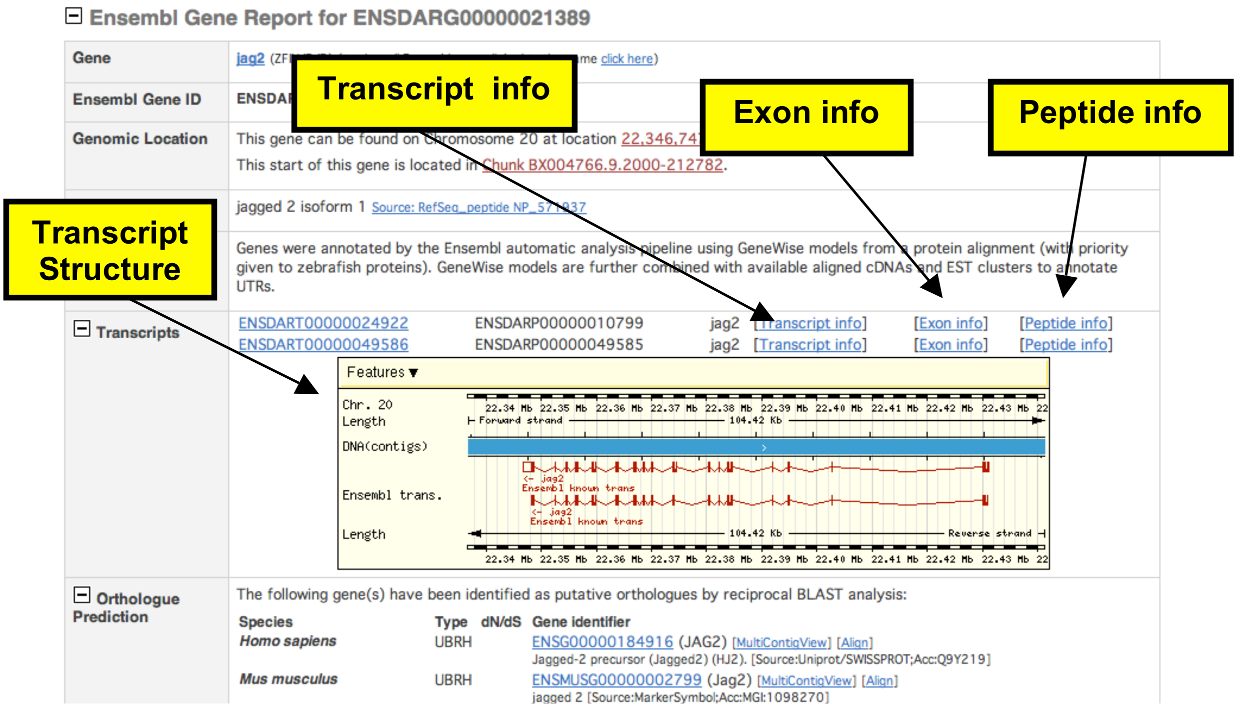

Overview
Features Menu
Detailed view
EST genes
Ensembl genes
Basepair view
Ensembl Zebrafish ContigView
sanger
EBI
Zebrafish
What's New
TextSearch
BlastSearch
MartSearch
Export Data
Download
Archive sites
All
Lookup
[e.g. Zv4_NA18430, Zv4_NA3486]
Help
Chromosome 20
Chr. 20
Overview
DNA(contigs)
Markers
Ensembl Genes
Gene legend
Detailed view
Jump to region: 20 bp 25272803 to 25357225
Refresh
2 Mb
1 Mb
Window
Zoom
Features
Compara
DAS Sources
Repeats
Decorations
Export
Jump to
Image size
Help
Length
D.rerio cDNAs
EMBL mRNAs
Proteins
Genscan
EST trans.
Ensembl trans.
DNA(contigs)
EMBL mRNAs
Markers
Length
Gene legend
Basepair view
Window
Zoom
Length
Genscan
Ensembl trans.
Amino acids
Restr.Enzymes
Length
Gene legend
Date : Wed Jul 6 19:57:49 2005
Archive! Permanent page link
Help Desk / Suggestions

The Features menu in the detailed view controls the tracks you can visualise in the panel. Tracks can be turned on and off and the features can be collapsed to simplify the view. Spend some time on this page trying the different menus and studying the displayed features. Observe that there are two tracks for predicted genes: Ensembl transcripts and EST transcripts. (If these features are not visible verify that the corresponding tracks are selected in the menu.)

GeneView, TransView, ExonView and ProtView

Another important view in Ensembl I are the **GeneView** pages with information about the Ensembl predicted genes. In the ContigView page above there is a predicted transcript on the forward strand called **jag2**. Clicking on this transcript displays a pop-up window with several options. Follow the ink labelled Ensembl Gene: ENSDARG00000021389. Below we only show the top of the GeneView page for jag2; scroll down to view all the information available.

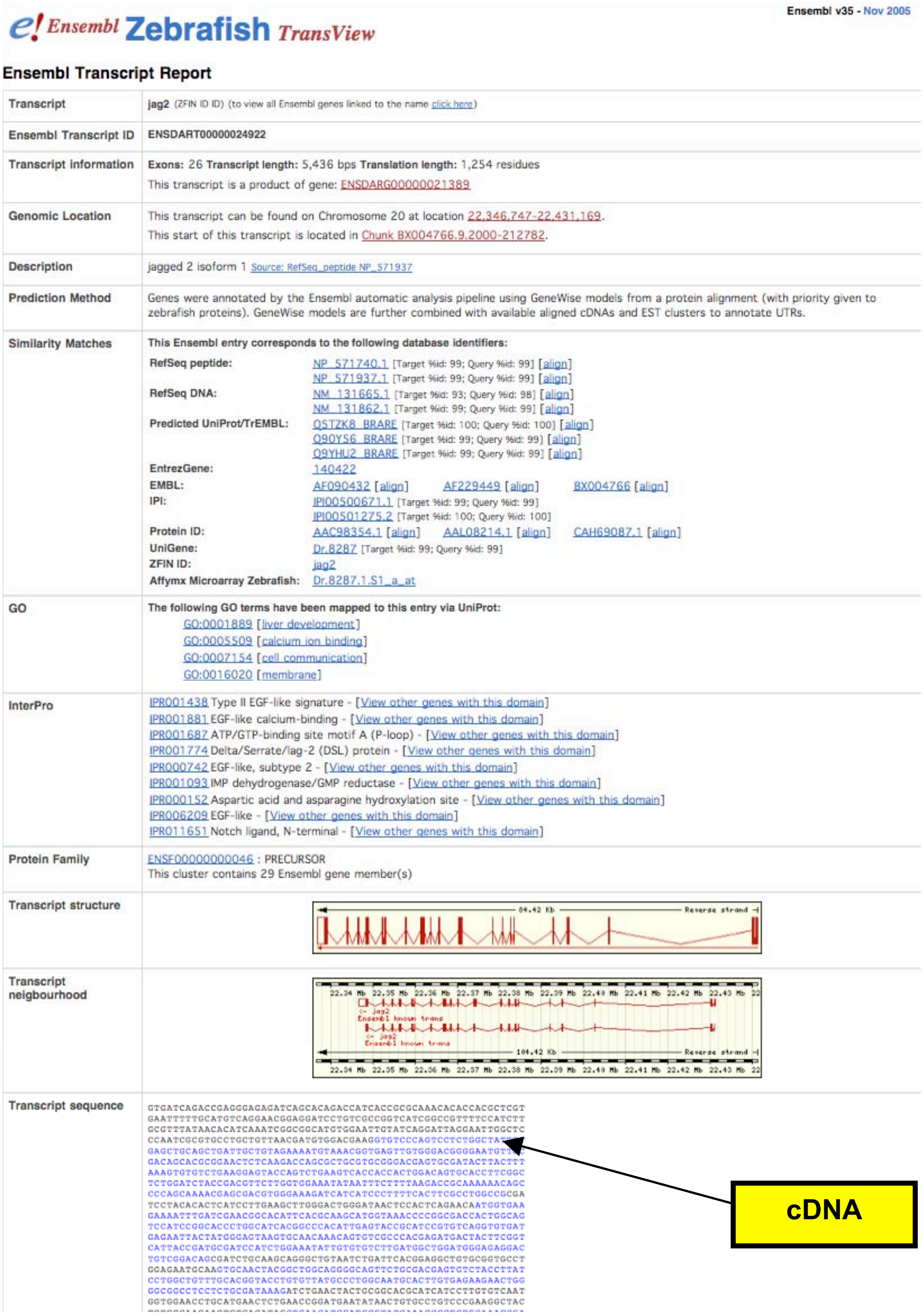
GeneView provides annotation and supporting evidence for the selected gene. The annotation consists of transcripts, homologies to other species, known and predicted proteins and domains, and links to external documentation. In this example, jag2 is a gene known to ZFIN and so a link to the corresponding external page is provided. The annotation for jag2 is based on 2 transcripts. In the Transcripts sections there are links to the corresponding TransView pages. Click on the link labelled “Transcript info” for the first one with identifier ENSDART00000024922.



⊟ Ensembl Gene Report for ENSDARG00000021389

Gene	jag2 (ZF ... me click here)
Ensembl Gene ID	ENSDA
Genomic Location	This gene can be found on Chromosome 20 at location 22,346,74 This start of this gene is located in Chunk BX004766.9.2000-212782.
	jagged 2 isoform 1 Source: RefSeq_peptide NP_571937
	Genes were annotated by the Ensembl automatic analysis pipeline using GeneWise models from a protein alignment (with priority given to zebrafish proteins). GeneWise models are further combined with available aligned cDNAs and EST clusters to annotate UTRs.
⊟ Transcripts	ENSDART00000024922 ENSDARP00000010799 jag2 [Transcript info] [Exon info] [Peptide info] ENSDART00000049586 ENSDARP00000049585 jag2 [Transcript info] [Exon info] [Peptide info]
⊟ Orthologue Prediction	The following gene(s) have been identified as putative orthologues by reciprocal BLAST analysis:

Species	Type	dN/dS	Gene identifier
Homo sapiens	UBRH		ENSG00000184916 (JAG2) [MultiContigView] [Align] Jagged-2 precursor (Jagged2) (HJ2). [Source:Uniprot/SWISSPROT;Acc:Q9Y219]
Mus musculus	UBRH		ENSMUSG00000002799 (Jag2) [MultiContigView] [Align] jagged 2 [Source:MarkerSymbol;Acc:MGI:1098270]



e! Ensembl Zebrafish TransView

Ensembl v35 - Nov 2005

Ensembl Transcript Report

Transcript	**jag2** (ZFIN ID ID) (to view all Ensembl genes linked to the name click here)
Ensembl Transcript ID	**ENSDART00000024922**
Transcript information	**Exons:** 26 **Transcript length:** 5,436 bps **Translation length:** 1,254 residues This transcript is a product of gene: ENSDARG00000021389
Genomic Location	This transcript can be found on Chromosome 20 at location 22,346,747-22,431,169. This start of this transcript is located in Chunk BX004766.9.2000-212782.
Description	jagged 2 isoform 1 Source: RefSeq_peptide NP_571937
Prediction Method	Genes were annotated by the Ensembl automatic analysis pipeline using GeneWise models from a protein alignment (with priority given to zebrafish proteins). GeneWise models are further combined with available aligned cDNAs and EST clusters to annotate UTRs.
Similarity Matches	**This Ensembl entry corresponds to the following database identifiers:**

RefSeq peptide:	NP_571740.1 [Target %id: 99; Query %id: 99] [align]		
	NP_571937.1 [Target %id: 99; Query %id: 99] [align]		
RefSeq DNA:	NM_131665.1 [Target %id: 93; Query %id: 98] [align]		
	NM_131862.1 [Target %id: 99; Query %id: 99] [align]		
Predicted UniProt/TrEMBL:	Q5TZK8_BRARE [Target %id: 100; Query %id: 100] [align]		
	Q90Y56_BRARE [Target %id: 99; Query %id: 99] [align]		
	Q9YHU2_BRARE [Target %id: 99; Query %id: 99] [align]		
EntrezGene:	140422		
EMBL:	AF090432 [align]	AF229449 [align]	BX004766 [align]
IPI:	IPI00500671.1 [Target %id: 99; Query %id: 99]		
	IPI00501275.2 [Target %id: 100; Query %id: 100]		
Protein ID:	AAC98354.1 [align]	AAL08214.1 [align]	CAH69087.1 [align]
UniGene:	Dr.8287 [Target %id: 99; Query %id: 99]		
ZFIN ID:	jag2		
Affymx Microarray Zebrafish:	Dr.8287.1.S1_a_at		

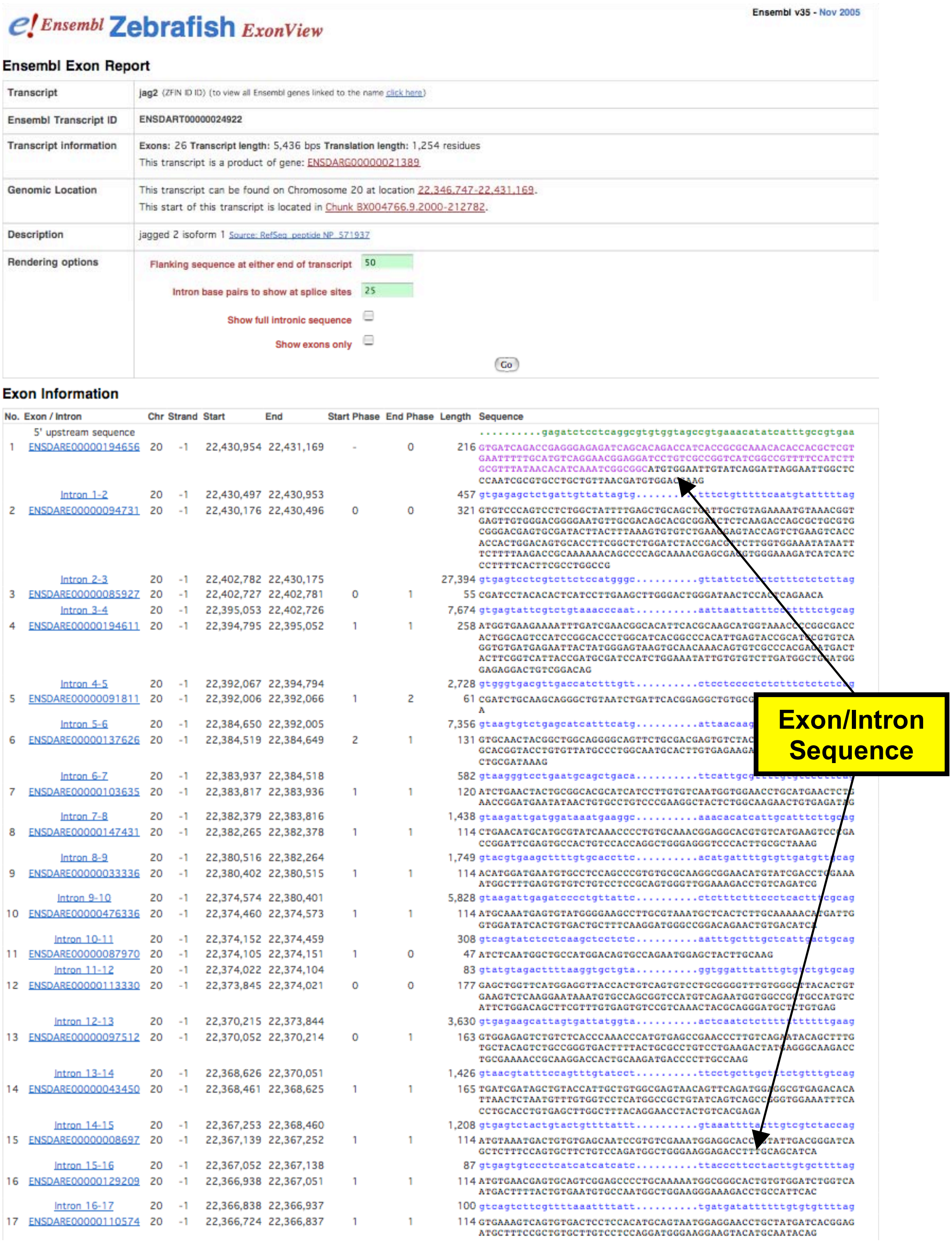
GO	**The following GO terms have been mapped to this entry via UniProt:** GO:0001889 [liver development] GO:0005509 [calcium ion binding] GO:0007154 [cell communication] GO:0016020 [membrane]
InterPro	IPR001438 Type II EGF-like signature - [View other genes with this domain] IPR001881 EGF-like calcium-binding - [View other genes with this domain] IPR001687 ATP/GTP-binding site motif A (P-loop) - [View other genes with this domain] IPR001774 Delta/Serrate/lag-2 (DSL) protein - [View other genes with this domain] IPR000742 EGF-like, subtype 2 - [View other genes with this domain] IPR001093 IMP dehydrogenase/GMP reductase - [View other genes with this domain] IPR000152 Aspartic acid and asparagine hydroxylation site - [View other genes with this domain] IPR006209 EGF-like - [View other genes with this domain] IPR011651 Notch ligand, N-terminal - [View other genes with this domain]
Protein Family	ENSF00000000046 : PRECURSOR This cluster contains 29 Ensembl gene member(s)
Transcript structure	
Transcript neigbourhood	
Transcript sequence	

```
GTGATCAGACCGAGGGAGAGATCAGCACAGACCATCACCGCGCAAACACACCACGCTCGT
GAATTTTTGCATGTCAGGAACGGAGGATCCTGTCGCCGGTCATCGGCCGTTTTCCATCTT
GCGTTTATAACACATCAAATCGGCGGCATGTGGAATTGTATCAGGATTAGGAATTGGCTC
CCAATCGCGTGCCTGCTGTTAACGATGTGGACGAAGGTGTCCCAGTCCTCTGGCTAT
GAGCTGCAGCTGATTGCTGTAGAAAATGTAAACGGTGAGTTGTGGACGGGGAATGTT
GACAGCACGCGGAACTCTCAAGACCAGCGCTGCGTGCGGACGAGTGCGATACTTACTTT
AAAGTGTGTCTGAAGGAGTACCAGTCTGAAGTCACCACCACTGGACAGTGCACCTTCGGC
TCTGGATCTACCGACGTTCTTGGTGGAAATATAATTTCTTTTAAGACCGCAAAAAACAGC
CCCAGCAAAACGAGCGACGTGGGAAAGATCATCATCCCTTTTCACTTCGCCTGGCCGCGA
TCCTACACACTCATCCTTGAAGCTTGGGACTGGGATAACTCCACTCAGAACAATGGTGAA
GAAAATTTGATCGAACGGCACATTCACGCAAGCATGGTAAACCCCGGCGACCACTGGCAG
TCCATCCGGCACCCTGGCATCACGGCCCACATTGAGTACCGCATCCGTGTCAGGTGTGAT
GAGAATTACTATGGGAGTAAGTGCAACAAACAGTGTCGCCCACGAGATGACTACTTCGGT
CATTACCGATGCGATCCATCTGGAAATATTGTGTGTCTTGATGGCTGGATGGGAGAGGAC
TGTCGGACAGCGATCTGCAAGCAGGGCTGTAATCTGATTCACGGAGGCTGTGCGGTGCCT
GGAGAATGCAAGTGCAACTACGGCTGGCAGGGGCAGTTCTGCGACGAGTGTCTACCTTAT
CCTGGCTGTTTGCACGGTACCTGTGTTATGCCCTGGCAATGCACTTGTGAGAAGAACTGG
GGCGGCCTCCTCTGCGATAAAGATCTGAACTACTGCGGCACGCATCATCCTTGTGTCAAT
GGTGGAACCTGCATGAACTCTGAACCGGATGAATATAACTGTGCCTGTCCCGAAGGCTAC
```

cDNA

TransView provides annotation and supporting evidence for the selected transcript (structure, transcribed proteins, Gene Ontology and InterPro associated entries). The Transcript report panel provides a top-level summary of the transcript, with links to its genomic location, alignments to sequences in external databases, and export options. Underneath the report, the cDNA sequence of the transcript can be shown with codons, peptide sequence and/or SNPs highlighted.

From the GeneView page there are also links to the ExonViews labelled as “Exon info”.



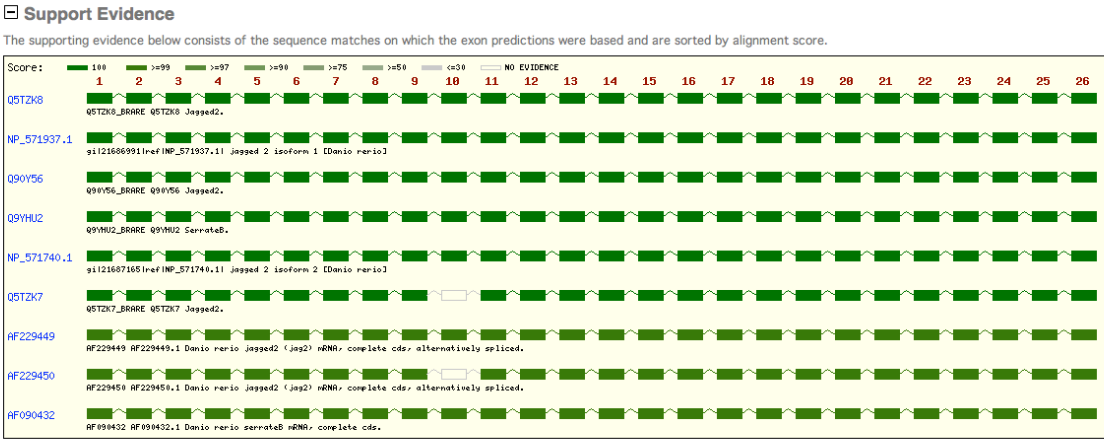
Ensembl Exon Report

Transcript	jag2 (ZFIN ID ID) (to view all Ensembl genes linked to the name click here)
Ensembl Transcript ID	ENSDART00000024922
Transcript information	Exons: 26 Transcript length: 5,436 bps Translation length: 1,254 residues This transcript is a product of gene: ENSDARG00000021389
Genomic Location	This transcript can be found on Chromosome 20 at location 22,346,747-22,431,169. This start of this transcript is located in Chunk BX004766.9.2000-212782.
Description	jagged 2 isoform 1 Source: RefSeq_peptide NP_571937
Rendering options	Flanking sequence at either end of transcript: 50 Intron base pairs to show at splice sites: 25 Show full intronic sequence ☐ Show exons only ☐ Go

Exon Information

No.	Exon / Intron	Chr	Strand	Start	End	Start Phase	End Phase	Length	Sequence
	5' upstream sequence								..........gagatctcctcaggcgtgtggtagccgtgaaacatatcatttgccgtgaa
1	ENSDARE00000194656	20	-1	22,430,954	22,431,169	-	0	216	GTGATCAGACCGAGGGAGAGATCAGCACAGACCATCACCGCGCAAACACACCACGCTCGT GAATTTTTGCATGTCAGGAACGGAGGATCCTGTCGCCGGTCATCGGCCGTTTTCCATCTT GCGTTTATAACACATCAAATCGGCGGCATGTGGAATTGTATCAGGATTAGGAATTGGCTC CCAATCGCGTGCCTGCTGTTAACGATGTGGACGAAG
	Intron 1-2	20	-1	22,430,497	22,430,953			457	gtgagagctctgattgttattagtg..........tttctgttttcaatgtattttttag
2	ENSDARE00000094731	20	-1	22,430,176	22,430,496	0	0	321	GTGTCCCAGTCCTCTGGCTATTTTGAGCTGCAGCTGATTGCTGTAGAAAATGTAAACGGT GAGTTGTGGGACGGGGAATGTTGCGACAGCACGCGGAACTCTCAAGACCAGCGCTGCGTG CGGGACGAGTGCGATACTTACTTTAAAGTGTGTCTGAAGGAGTACCAGTCTGAAGTCACC ACCACTGGACAGTGCACCTTCGGCTCTGGATCTACCGACGTTCTTGGTGGAAATATAATT TCTTTTAAGACCGCAAAAAACAGCCCCAGCAAAACGAGCGAAGTGGGAAAGATCATCATC CCTTTTCACTTCGCCTGGCCG
	Intron 2-3	20	-1	22,402,782	22,430,175			27,394	gtgagtcctcgtcttctccatgggc..........gttattctctctctttctctcttag
3	ENSDARE00000085927	20	-1	22,402,727	22,402,781	0	1	55	CGATCCTACACACTCATCCTTGAAGCTTGGGACTGGGATAACTCCACTCAGAACA
	Intron 3-4	20	-1	22,395,053	22,402,726			7,674	gtgagtattcgtctgtaaacccaat..........aattaattatttcctttttctgcag
4	ENSDARE00000194611	20	-1	22,394,795	22,395,052	1	1	258	ATGGTGAAGAAAATTTGATCGAACGGCACATTCACGCAAGCATGGTAAACCCCGGCGACC ACTGGCAGTCCATCCGGCACCCTGGCATCACGGCCCACATTGAGTACCGCATCCGTGTCA GGTGTGATGAGAATTACTATGGGAGTAAGTGCAACAAACAGTGTCGCCCACGAGATGACT ACTTCGGTCATTACCGATGCGATCCATCTGGAAATATTGTGTGTCTTGATGGCTGGATGG GAGAGGACTGTCGGACAG
	Intron 4-5	20	-1	22,392,067	22,394,794			2,728	gtgggtgacgttgaccatctttgtt..........ctcctcccctctctttctctctcag
5	ENSDARE00000091811	20	-1	22,392,006	22,392,066	1	2	61	CGATCTGCAAGCAGGGCTGTAATCTGATTCACGGAGGCTGTGCG A
	Intron 5-6	20	-1	22,384,650	22,392,005			7,356	gtaagtgtctgagcatcatttcatg..........attaacaag
6	ENSDARE00000137626	20	-1	22,384,519	22,384,649	2	1	131	GTGCAACTACGGCTGGCAGGGGCAGTTCTGCGACGAGTGTCTAC GCACGGTACCTGTGTTATGCCCTGGCAATGCACTTGTGAGAAGA CTGCGATAAAG
	Intron 6-7	20	-1	22,383,937	22,384,518			582	gtaagggtcctgaatgcagctgaca..........ttcattgcg
7	ENSDARE00000103635	20	-1	22,383,817	22,383,936	1	1	120	ATCTGAACTACTGCGGCACGCATCATCCTTGTGTCAATGGTGGAACCTGCATGAACTCTG AACCGGATGAATATAACTGTGCCTGTCCCGAAGGCTACTCTGGCAAGAACTGTGAGATAG
	Intron 7-8	20	-1	22,382,379	22,383,816			1,438	gtaagattgatggataaatgaaggc..........aaacacatcattgcatttcttgcag
8	ENSDARE00000147431	20	-1	22,382,265	22,382,378	1	1	114	CTGAACATGCATGCGTATCAAACCCCTGTGCAAACGGAGGCACGTGTCATGAAGTCCTGA CCGGATTCGAGTGCCACTGTCCACCAGGCTGGGAGGGTCCCACTTGCGCTAAAG
	Intron 8-9	20	-1	22,380,516	22,382,264			1,749	gtacgtgaagcttttgtgcaccttc..........acatgatttgtgttgatgttgcag
9	ENSDARE00000033336	20	-1	22,380,402	22,380,515	1	1	114	ACATGGATGAATGTGCCTCCAGCCCGTGTGCGCAAGGCGGAACATGTATCGACCTGGAAA ATGGCTTTGAGTGTGTCTGTCCTCCGCAGTGGGTTGGAAAGACCTGTCAGATCG
	Intron 9-10	20	-1	22,374,574	22,380,401			5,828	gtaagattgagatcccctgttattc..........ctctttctttccctcactttcgcag
10	ENSDARE00000476336	20	-1	22,374,460	22,374,573	1	1	114	ATGCAAATGAGTGTATGGGGAAGCCTTGCGTAAATGCTCACTCTTGCAAAAACATGATTG GTGGATATCACTGTGACTGCTTTCAAGGATGGGCCGGACAGAACTGTGACATCA
	Intron 10-11	20	-1	22,374,152	22,374,459			308	gtcagtatctcctcaagctcctctc..........aatttgctttgctcattgactgcag
11	ENSDARE00000087970	20	-1	22,374,105	22,374,151	1	0	47	ATCTCAATGGCTGCCATGGACAGTGCCAGAATGGAGCTACTTGCAAG
	Intron 11-12	20	-1	22,374,022	22,374,104			83	gtatgtagacttttaaggtgctgta..........ggtggatttatttgtgtctgtgcag
12	ENSDARE00000113330	20	-1	22,373,845	22,374,021	0	0	177	GAGCTGGTTCATGGAGGTTACCACTGTCAGTGTCCTGCGGGGTTTGTGGGCTTACACTGT GAAGTCTCAAGGAATAAATGTGCCAGCGGTCCATGTCAGAATGGTGGCCGCTGCCATGTC ATTCTGGACAGCTTCGTTTGTGAGTGTCCGTCAAACTACGCAGGGATGCTCTGTGAG
	Intron 12-13	20	-1	22,370,215	22,373,844			3,630	gtgagaagcattagtgattatggta..........actcaatctcttttttttttgaag
13	ENSDARE00000097512	20	-1	22,370,052	22,370,214	0	1	163	GTGGAGAGTCTGTCTCACCCAAACCCATGTGAGCCGAACCCTTGTCAGAATACAGCTTTG TGCTACAGTCTGCCGGGTGACTTTTACTGCGCCTGTCCTGAAGACTATGAGGGCAAGACC TGCGAAAACCGCAAGGACCACTGCAAGATGACCCCTTGCCAAG
	Intron 13-14	20	-1	22,368,626	22,370,051			1,426	gtaacgtatttccagtttgtatcct..........ttcctgcttgcttctgtttgtcag
14	ENSDARE00000043450	20	-1	22,368,461	22,368,625	1	1	165	TGATCGATAGCTGTACCATTGCTGTGGCGAGTAACAGTTCAGATGGAGGCGTGAGACACA TTAACTCTAATGTTTGTGGTCCTCATGGCCGCTGTATCAGTCAGCCGGGTGGAAATTTCA CCTGCACCTGTGAGCTTGGCTTTACAGGAACCTACTGTCACGAGA
	Intron 14-15	20	-1	22,367,253	22,368,460			1,208	gtgagtctactgtactgttttattt..........gtaaatttacttgtcgtctaccag
15	ENSDARE00000008697	20	-1	22,367,139	22,367,252	1	1	114	ATGTAAATGACTGTGTGAGCAATCCGTGTCGAAATGGAGGCACCTGTATTGACGGATCA GCTCTTTCCAGTGCTTCTGTCCAGATGGCTGGGAAGGAGACCTTTGCAGCATCA
	Intron 15-16	20	-1	22,367,052	22,367,138			87	gtgagtgtccctcatcatcatcatc..........ttacccttcctacttgtgcttttag
16	ENSDARE00000129209	20	-1	22,366,938	22,367,051	1	1	114	ATGTGAACGAGTGCAGTCGGAGCCCCTGCAAAAATGGCGGGCACTGTGTGGATCTGGTCA ATGACTTTTACTGTGAATGTGCCAATGGCTGGAAGGGAAAGACCTGCCATTCAC
	Intron 16-17	20	-1	22,366,838	22,366,937			100	gtcagtcttcgttttaaattttatt..........tgatgatattttttgtgtgttttag
17	ENSDARE00000110574	20	-1	22,366,724	22,366,837	1	1	114	GTGAAAGTCAGTGTGACTCCTCCACATGCAGTAATGGAGGAACCTGCTATGATCACGGAG ATGCTTTCCGCTGTGCTTGTCCTCCAGGATGGGAAGGAAGTACATGCAATACAG

ExonView provides annotation and supporting evidence for the exons of a selected transcript. Ensembl gene predictions are based on aligned evidence from external databases like UniProt and RefSeq. At the bottom of an ExonView page you can find the evidence linked to this prediction.



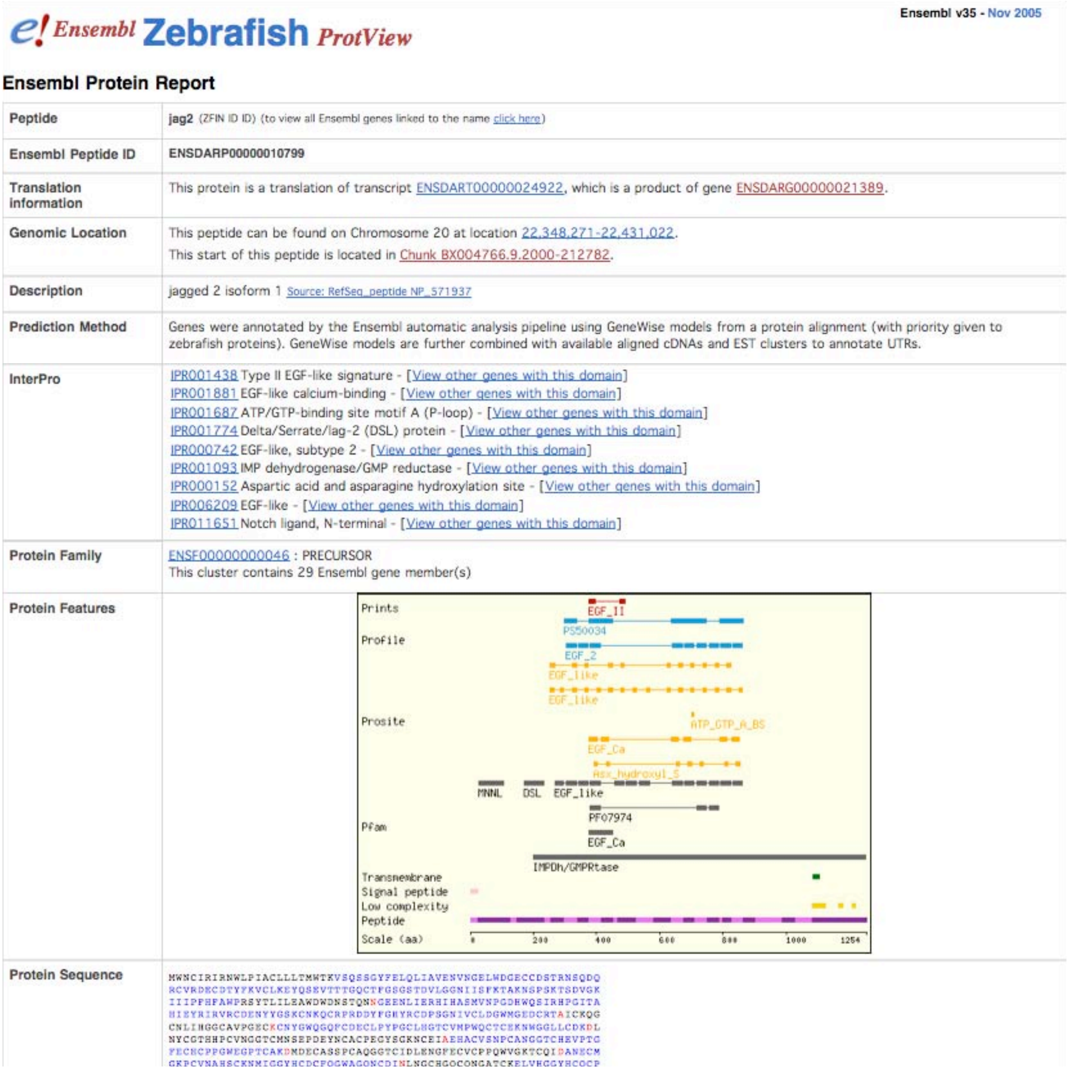
⊟ Support Evidence

The supporting evidence below consists of the sequence matches on which the exon predictions were based and are sorted by alignment score.

Score: 🟩 100 🟩 >=99 🟩 >=97 🟩 >=90 🟩 >=75 🟩 >=50 ▭ <=30 ▭ NO EVIDENCE

Q5TZK8	Q5TZK8_BRARE Q5TZK8 Jagged2.
NP_571937.1	gi|21686991|ref|NP_571937.1| jagged 2 isoform 1 [Danio rerio]
Q90Y56	Q90Y56_BRARE Q90Y56 Jagged2.
Q9YHU2	Q9YHU2_BRARE Q9YHU2 SerrateB.
NP_571740.1	gi|21687165|ref|NP_571740.1| jagged 2 isoform 2 [Danio rerio]
Q5TZK7	Q5TZK7_BRARE Q5TZK7 Jagged2.
AF229449	AF229449 AF229449.1 Danio rerio jagged2 (jag2) mRNA, complete cds, alternatively spliced.
AF229450	AF229450 AF229450.1 Danio rerio jagged2 (jag2) mRNA, complete cds, alternatively spliced.
AF090432	AF090432 AF090432.1 Danio rerio serrateB mRNA, complete cds.

Finally from the links labelled “Peptide info” in the GeneView page we can visit the ProtView page for the associated translation.



e! Ensembl Zebrafish ProtView

Ensembl v35 - Nov 2005

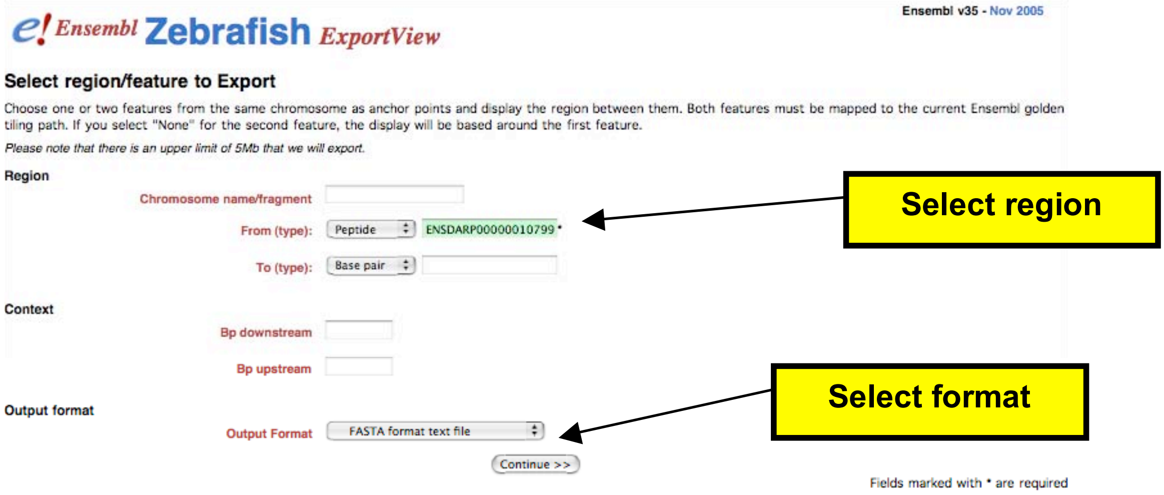
Ensembl Protein Report

Peptide	jag2 (ZFIN ID ID) (to view all Ensembl genes linked to the name click here)
Ensembl Peptide ID	ENSDARP00000010799
Translation information	This protein is a translation of transcript ENSDART00000024922, which is a product of gene ENSDARG00000021389.
Genomic Location	This peptide can be found on Chromosome 20 at location 22,348,271-22,431,022. This start of this peptide is located in Chunk BX004766.9.2000-212782.
Description	jagged 2 isoform 1 Source: RefSeq_peptide NP_571937
Prediction Method	Genes were annotated by the Ensembl automatic analysis pipeline using GeneWise models from a protein alignment (with priority given to zebrafish proteins). GeneWise models are further combined with available aligned cDNAs and EST clusters to annotate UTRs.
InterPro	IPR001438 Type II EGF-like signature - [View other genes with this domain] IPR001881 EGF-like calcium-binding - [View other genes with this domain] IPR001687 ATP/GTP-binding site motif A (P-loop) - [View other genes with this domain] IPR001774 Delta/Serrate/lag-2 (DSL) protein - [View other genes with this domain] IPR000742 EGF-like, subtype 2 - [View other genes with this domain] IPR001093 IMP dehydrogenase/GMP reductase - [View other genes with this domain] IPR000152 Aspartic acid and asparagine hydroxylation site - [View other genes with this domain] IPR006209 EGF-like - [View other genes with this domain] IPR011651 Notch ligand, N-terminal - [View other genes with this domain]
Protein Family	ENSF00000000046 : PRECURSOR This cluster contains 29 Ensembl gene member(s)
Protein Features	
Protein Sequence	MWNCIRIRNWLPIACLLLTMWTKVSQSSGYFELQLIAVENVNGELWDGECCDSTRNSQDQ RCVRDECDTYYKVCLKEYQSEVTTTGQCTFGSGSTDVLGGNIISFKTAKNSPSKTSDVGK IIIPFHFAWPRSYTLILEAWDWDNSTQNNGEENLIERHIHASMVNPGDHWQSIRHPGITA HIEYRIRVRCDENYYGSKCNKQCRPRDDYFGHYRCDPSGNIVCLDGWMGEDCRTAICKQG CNLIHGGCAVPGECKCNYGWQGQFCDECLPYPGCLHGTCVMPWQCTCEKNWGGLLCDKDL NYCGTHHPCVNGGTCMNSEPDEYNCACPEGYSGKNCEIAEHACVSNPCANGGTCHEVPTG FECHCPPGWEGPTCAKDMDECASSPCAQGGTCIDLENGFECVCPPQWVGKTCQIDANECM GKPCVNAHSCKNMIGGYHCDCFQGWAGQNCDINLNGCHGQCQNGATCKELVHGGYHCQCP

ProtView shows information about the structure and function of the encoded protein in the transcript's report with external links to various databases like Pfam, Prosite, etc…

ExportView

ExportView lets you download/dump data. All the features for a genomic region may be downloaded or exported to several formats (for example, FASTA, GenBank or EMBL-style flat file, as a feature list or an image). The ExportView pages are accessible from the link ‘Export data’ in the left-hand side menu from any of the pages above.



Select region/feature to Export

Choose one or two features from the same chromosome as anchor points and display the region between them. Both features must be mapped to the current Ensembl golden tiling path. If you select "None" for the second feature, the display will be based around the first feature.

Please note that there is an upper limit of 5Mb that we will export.

Region

Chromosome name/fragment

From (type): Peptide ENSDARP00000010799 *

To (type): Base pair

Context

Bp downstream

Bp upstream

Output format

Output Format FASTA format text file

Fields marked with * are required

Select region

Select format

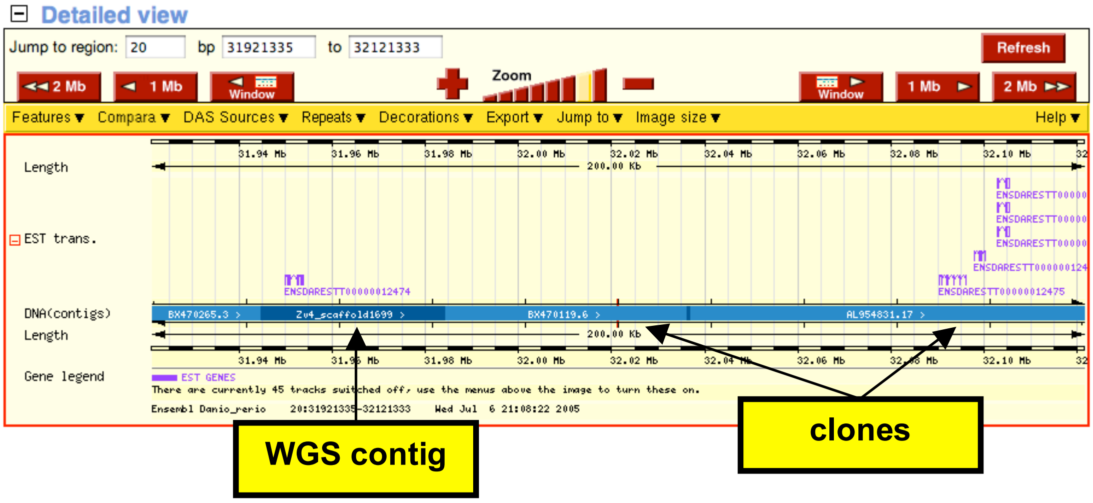
Zebrafish assembly in Ensembl

The sequence in the *Danio rerio* Ensembl database is the latest assembly release with automatic annotation. The genomic sequence released is based on all the sequenced clones with remaining gaps covered by contigs from a whole genome shotgun (WGS) assembly. The WGS fragments are placed in those gaps using a mixed strategy that looks at sequence similarity and other anchors as BAC-ends and markers. This placement is hard to perform without errors - mainly due to the presence of mis-joins in the WGS assembly and duplicates. It is even more difficult to place sequence where there is no sequenced clone or marker to use as an anchor.

In this context the user has to evaluate the data with a critical eye. In particular when the sequence of interest is known to the community but it is wrong in the assembly. There are three kinds of scaffolds and these are, in order of quality from best to worst:

1. scaffolds that have been attached to chromosomes (they may contain sequenced clones),
2. scaffolds that can be aligned to clones but the physical map cannot assign a chromosome yet (they may contain sequenced clones), and
3. NA (non-attached) scaffolds that corresponds to WGS contigs that could not be placed in the map (they don not contain sequenced clones).

Zv5_scaffold1699 is an example of category 1 above. This scaffold is placed in chromosome 20.

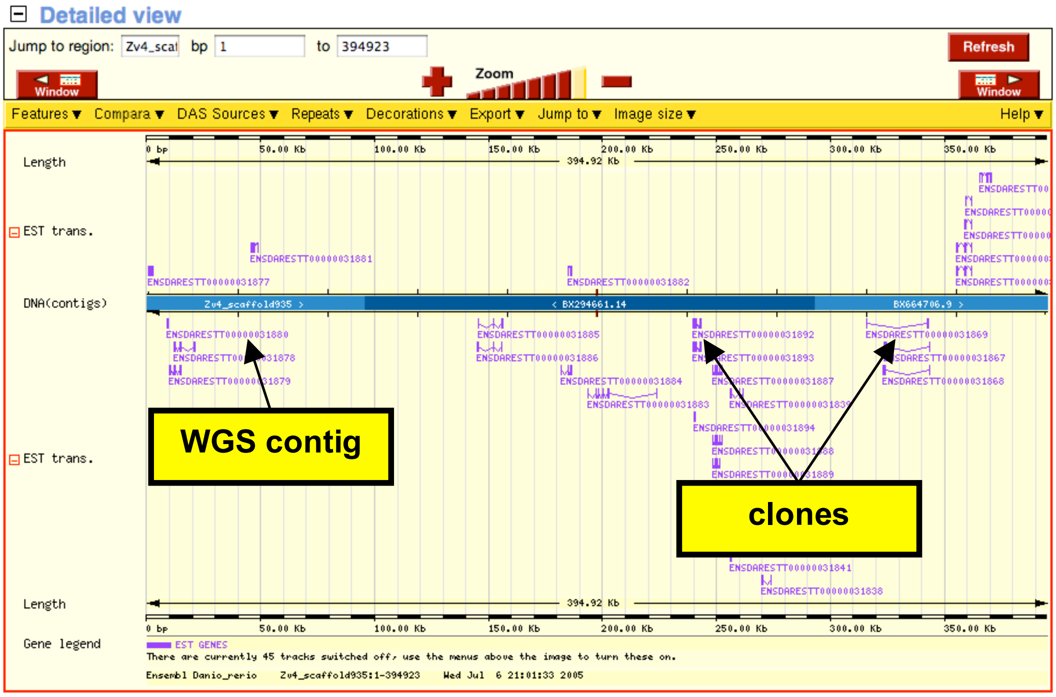


There are currently 45 tracks switched off, use the menus above the image to turn these on.

Ensembl Danio_rerio 20:31921335-32121333 Wed Jul 6 21:08:22 2005

In the detailed view for this page there is a genomic region labelled BX470265.3. This is the accession number of a sequenced clone. The region labelled Zv5_scaffold1699 is a WGS supercontig. This is of lower quality than the sequenced clone (and may contain gaps represented by a sequence of Ns).

Zv5_scaffold935 is an example of a region that is part of the map but, when the assembly was built, did not have a placement in a chromosome (category 2). This example shows that the region contains some sequenced clones as shown by the presence of their accession numbers.



There are currently 45 tracks switched off, use the menus above the image to turn these on.

Ensembl Danio_rerio Zv4_scaffold935:1-394923 Wed Jul 6 21:01:33 2005

Finally a scaffold from category 3 is Zv5_NA10. This region does not contain any finished clones.

Exercises

This section introduces the Ensembl browser and some of its basic views. In other section we will study more advanced features like the compara database and Blast/SSAHA search facilities. The user is encouraged to navigate the site and experiment with the different views discussed above.

1. Find the GeneView page for jag2 (Ensembl gene), and scroll down to the first 'Transcript/Translation Summary'. As jag2 has been identified in Zv6 you can use this gene name in a text search box.

2. Examine the genomic context. From GeneView, follow the link 'View gene in genomic location' to ContigView.

3. Customise the display of ContigView selecting different tracks and comparing the data from different tracks.

4. In ContigView zoom in to examine the data in more detail.

5. Export a file containing the cDNA of one of the predicted transcripts for jag2.

6. One of the Ensembl tracks displays probes for which ZFIN has a expression pattern page. Search for the mapping of the EST with accession CK685476 and open the corresponding ContigView page. Make sure that the 'expression pattern' track is selected in the 'features' menu. The ContigView page displays a link to the expression pattern page in ZFIN, try it.