

Overview

The zebrafish genome, which consists of 25 linkage groups and is ~1.4Gb in size, is being sequenced, finished and analysed in its entirety at the Wellcome Trust Sanger Institute to provide an open source, high quality reference genome.

The manual annotation, which is compiled in close collaboration with the Zebrafish Information Network (<http://zfin.org>), is provided by the Human and Vertebrate Analysis and Annotation (HAVANA) group and is frequently released onto the Vertebrate Genome Annotation (Vega) database (<http://vega.sanger.ac.uk>) and may also be viewed as a DAS source in Ensembl (http://www.ensembl.org/Danio_rerio).

Annotation is based on the reference genome assembly sequence, which is derived from a minimal tile path composed of clones finished to a 99.9% sequence standard.

We approach annotation using two main strategies. Firstly, the generation of gene lists composed of ZFIN cDNA that map to our current finished assembly. And, secondly, by using whole chromosome clone by clone annotation, where we have annotated over 4200 clones.

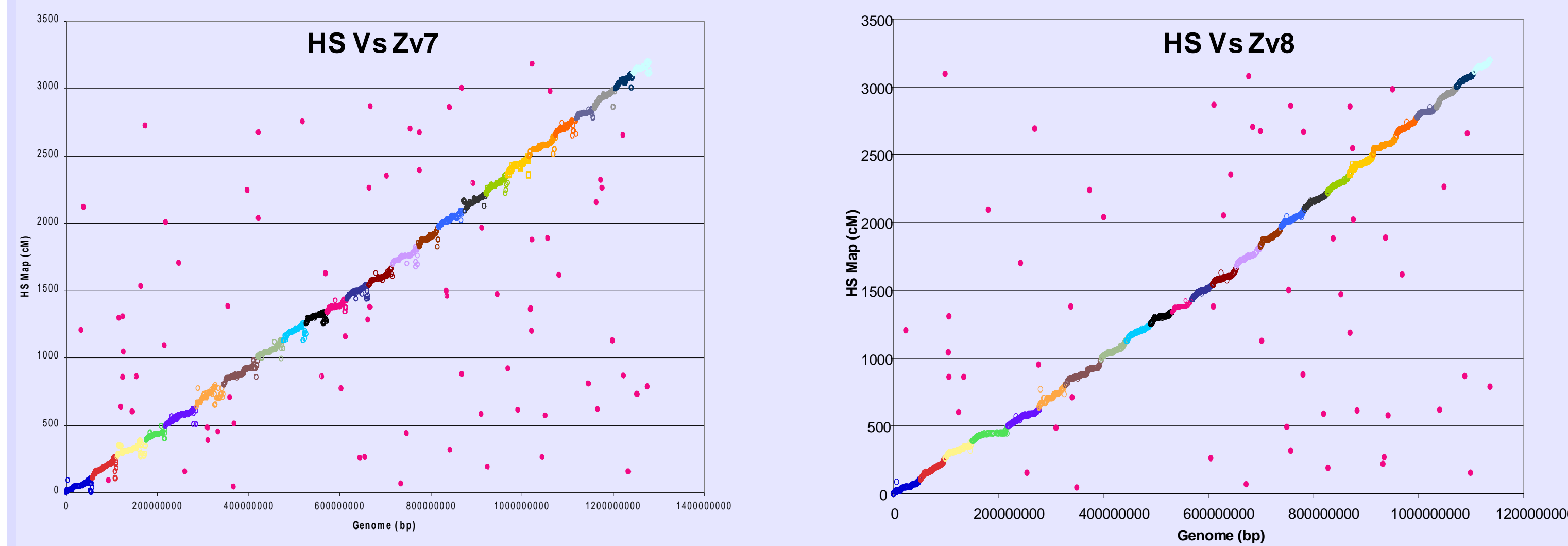
The latest zebrafish assembly (Zv8), which represents the most accurate assembly to date is available in Pre-Ensembl (http://pre.ensembl.org/Danio_rerio/Info/Index).

Zv8: The Latest Zebrafish Assembly

Our latest assembly was generated using 3 genetic maps: HS (meiotic), MGH (meiotic) & T51 (radiation hybrid). Each map was weighted based on the HS > MGH > T51 hierarchy, with one hit/marker/location.

The resulting ordered finger print contigs (FPC) will eventually represent our assembly on which our annotation is based.

The Ensembl release of Zv8 is due in late spring 2009.

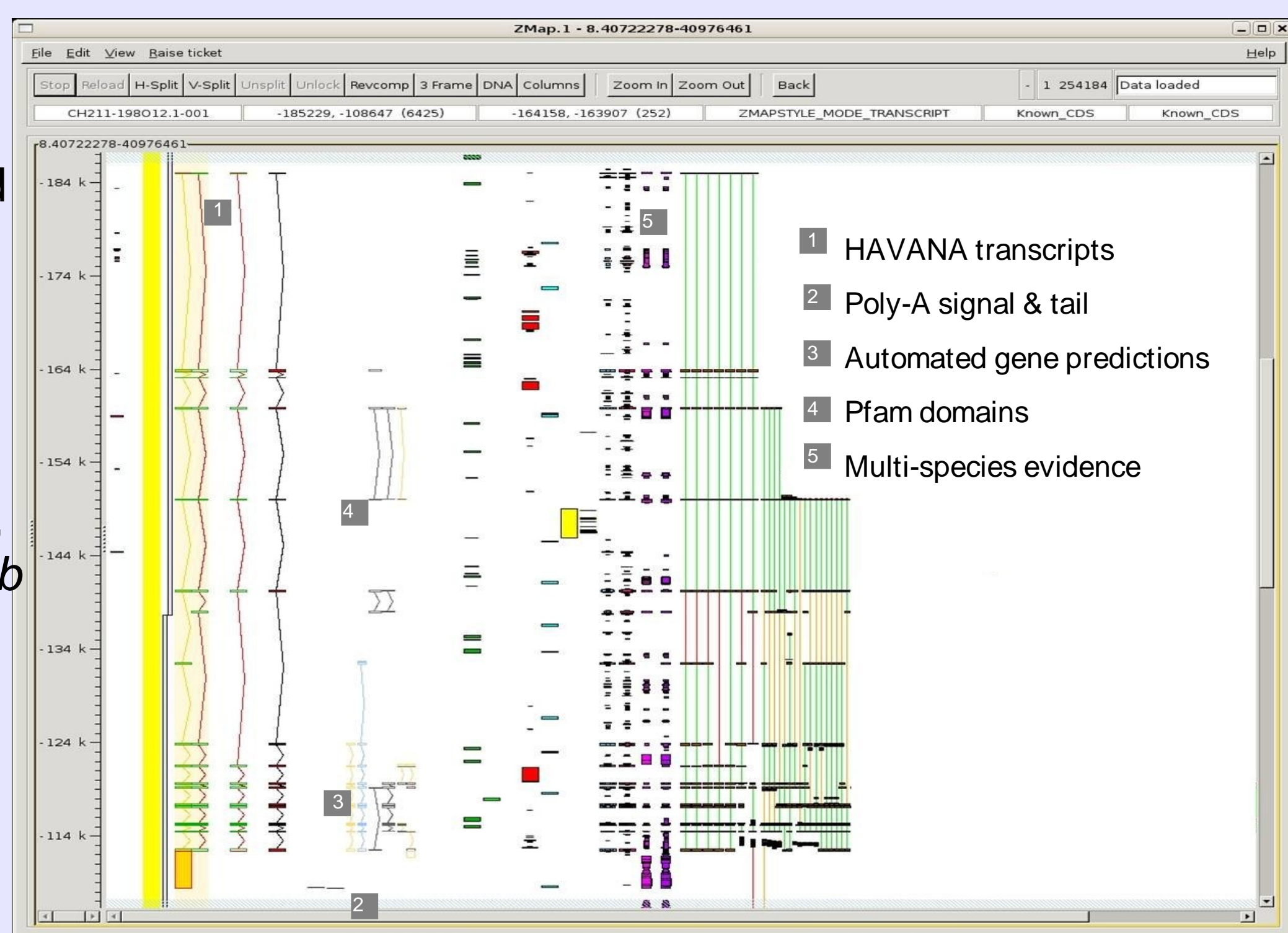


The colours represent linkage groups and the dots represent unplaced contigs that fall outside of the assembly. Zv8 is a smaller assembly than Zv7 due to the removal of over 260Mb of haplotypic sequence.

The ZMAP Annotation Interface

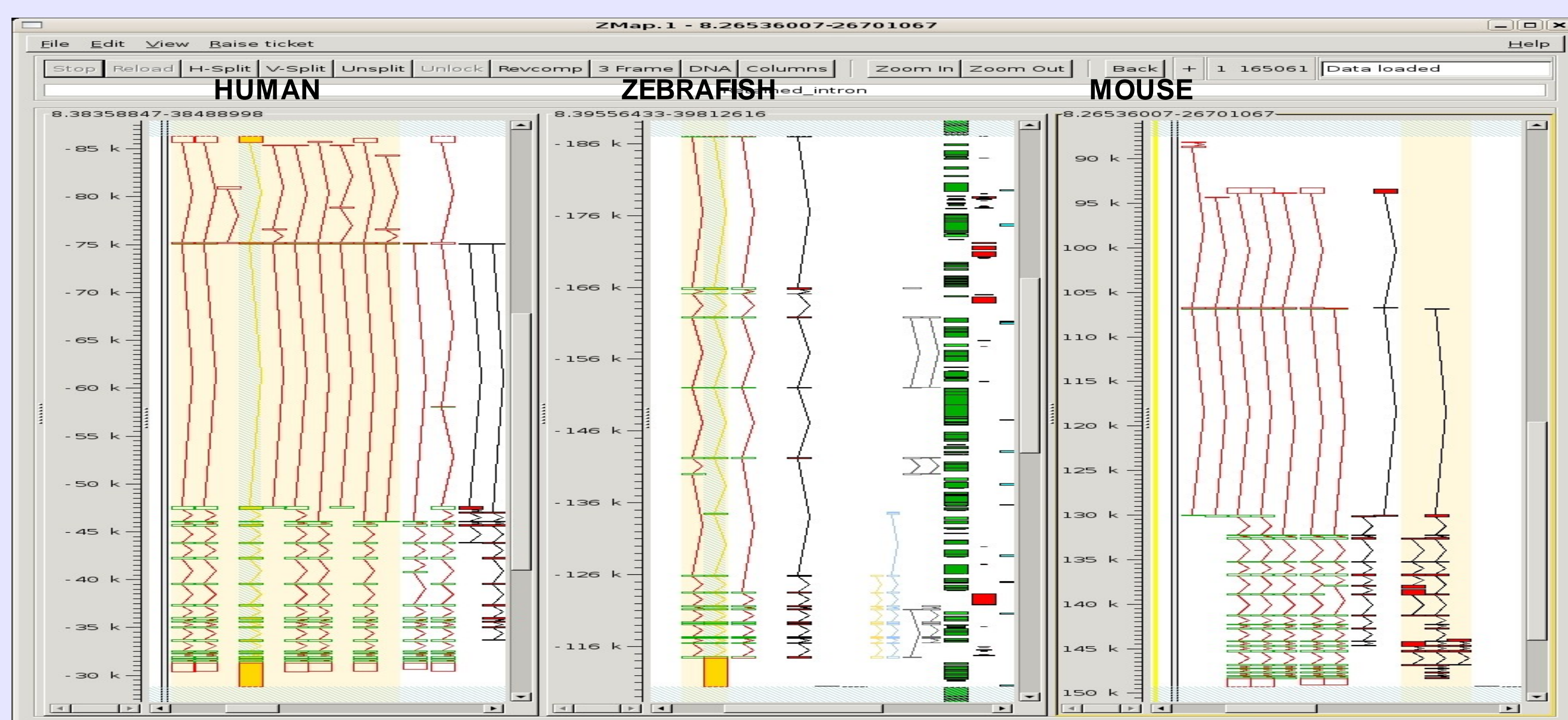
HAVANA annotation is based on same species and cross-species evidence (cDNA, EST, Protein) sourced from DNA databases and UniProt.

Refseq, Pfam, CpG islands, markers, tandem repeats, *ab initio* gene predictions are also displayed.



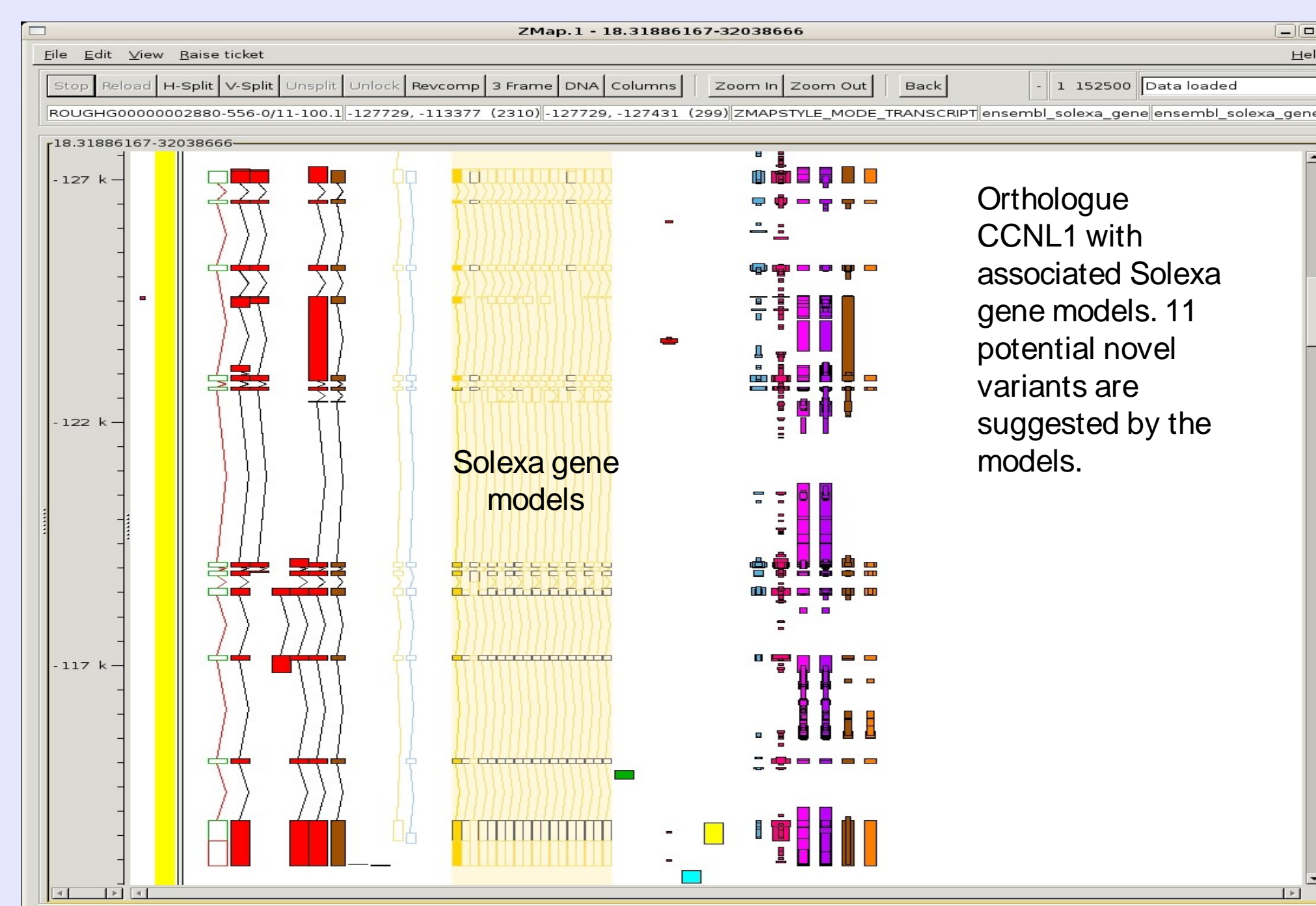
The fibroblast growth factor receptor 1 (fgfr1).

Compara View in Otterlace



Variants of fgfr1 in human, zebrafish & mouse.

Building Genes Using Solexa Data

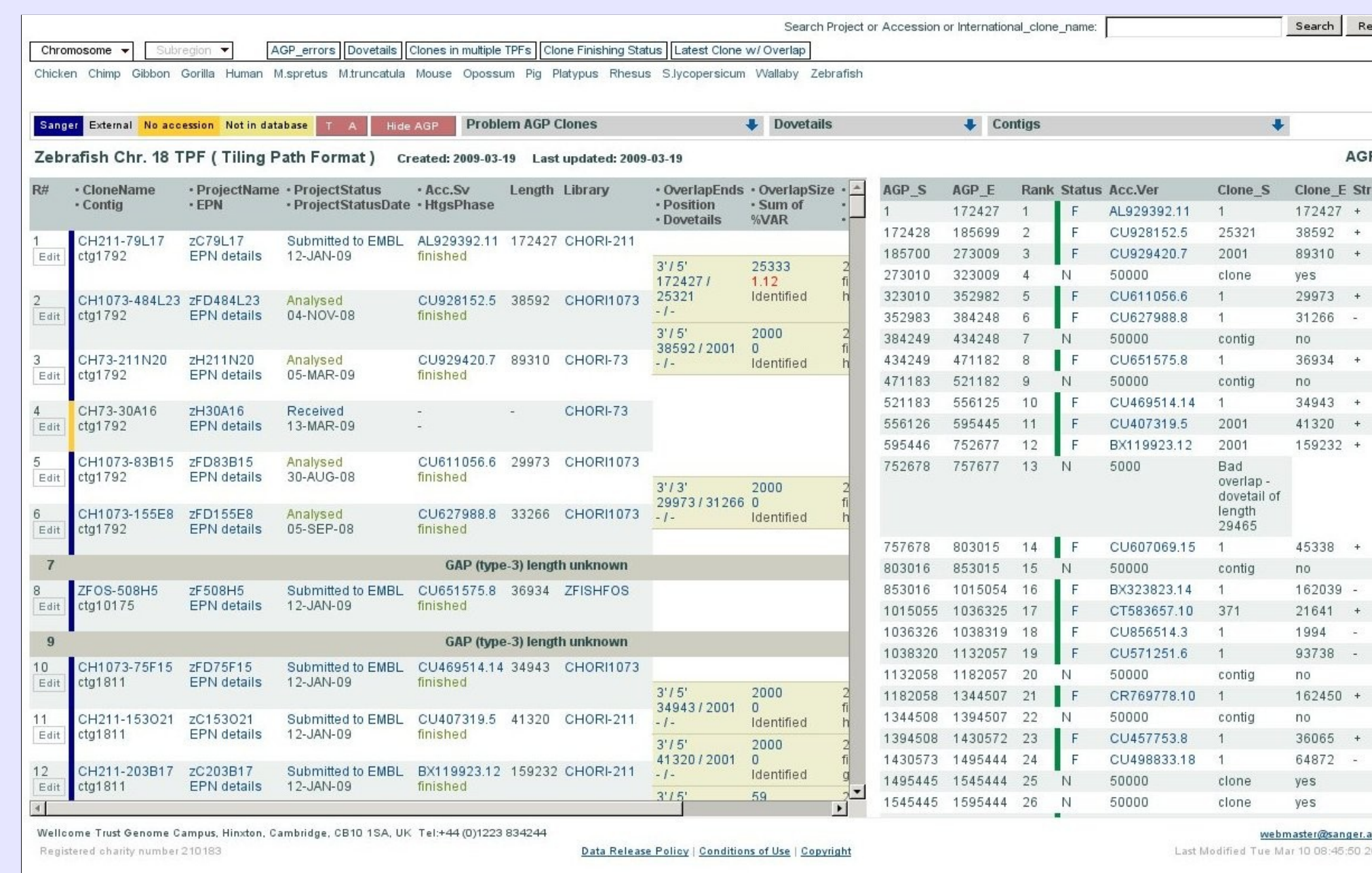


Solexa gene models are now incorporated alongside our dataset and are used as a guide by annotators.

They are derived from over 32Gb of short read pairs obtained by deep sequencing 5 tissue types at 6 stages of development.

Access To Our Sequencing Pipeline

<http://www.sanger.ac.uk/cgi-bin/humpub/chromoview>



Chromoview is a resource tool that enables users to view our Tiling Path Format (TPF) alongside the 'A Golden Path' (AGP) or assembly, which is derived from the TPF.

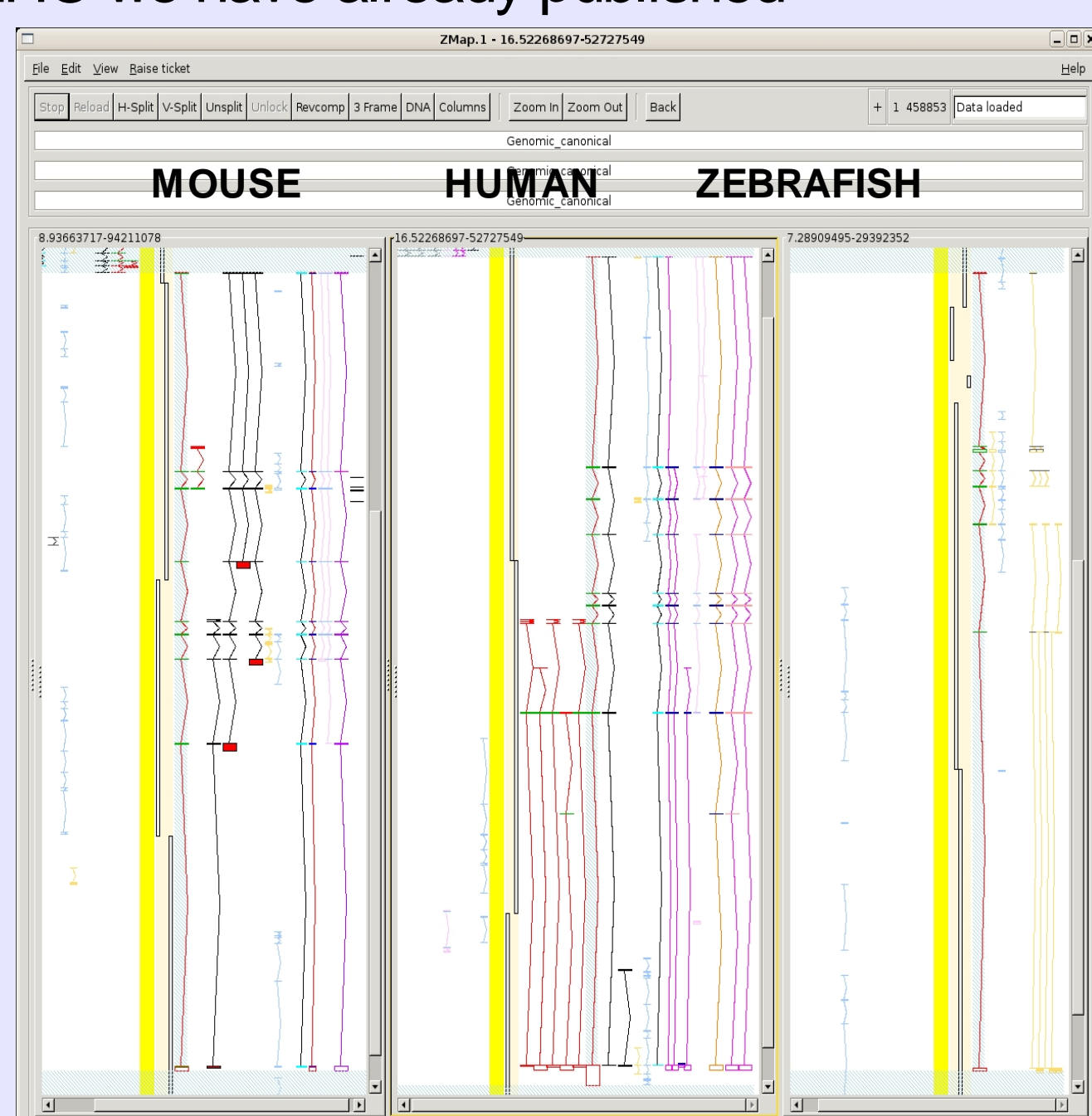
Clones may be queried, providing information on assembly location and status in the pipeline.

Community Directed Annotation

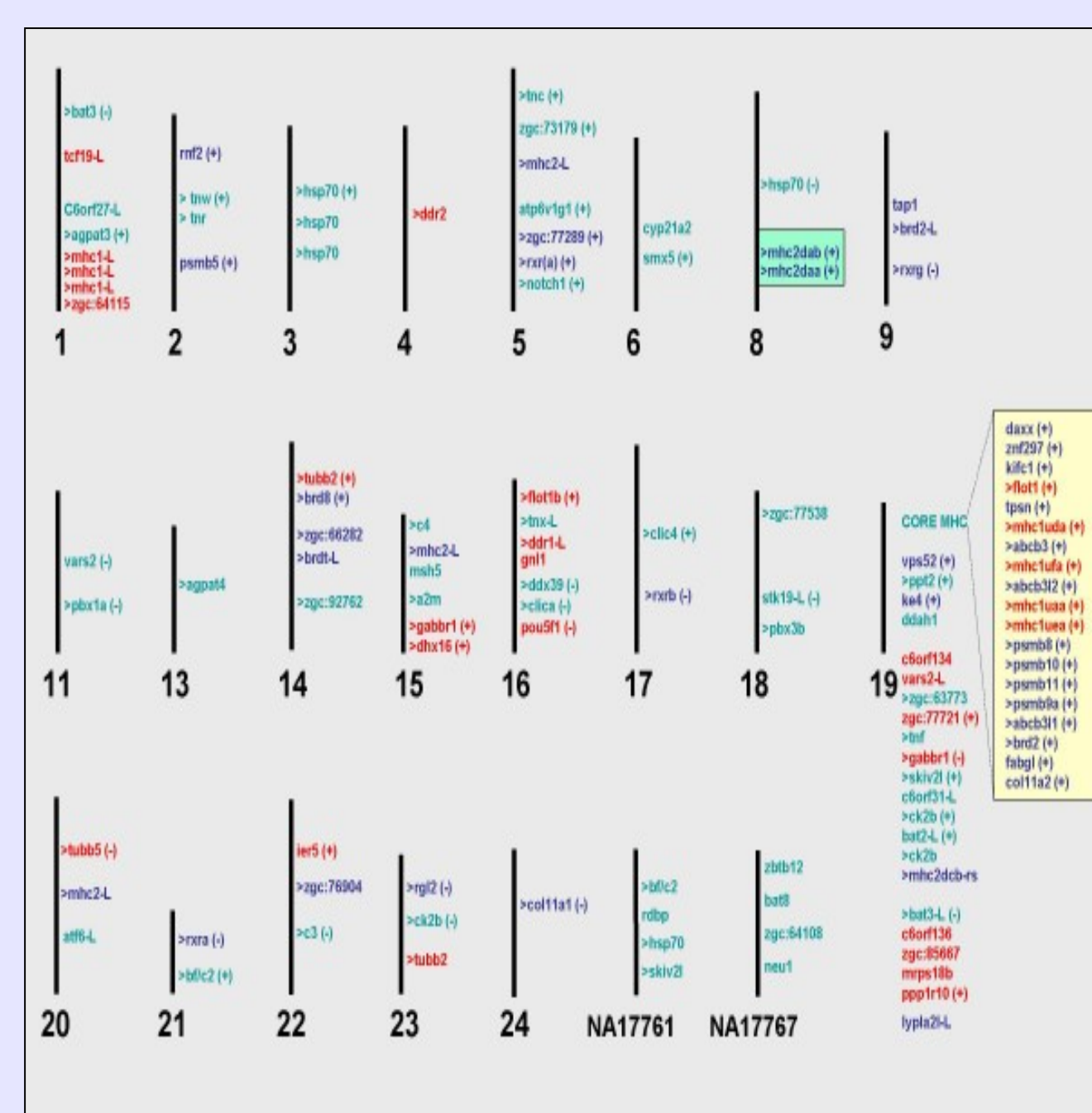
As well as our on-going genome annotation we also welcome feedback and annotation requests for genes and regions of interest (zfsh-help@sanger.ac.uk).

We have recently annotated 93 genes to facilitate the production of gene knock-out mutants, with the aim of studying the effect on phenotype in relation to human obesity.

Annotation of the core MHC is also planned, which will complement the human and mouse MHC we have already published



The FTO gene was annotated in human, mouse and zebrafish as a direct response to a request from the community.



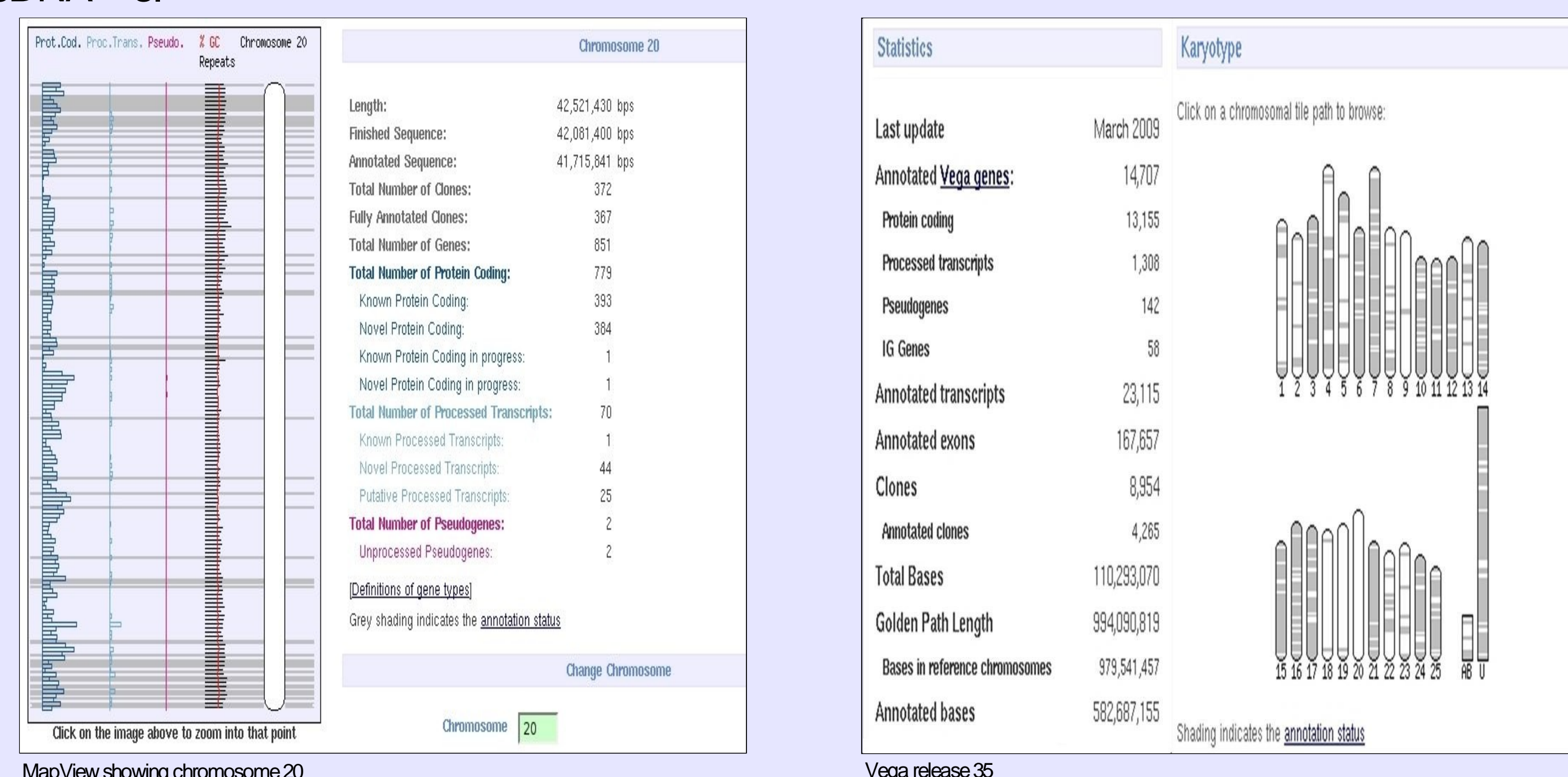
Annotation of the zebrafish MHC will focus on the core region on chromosome 19. Picture obtained from BMC Genomics. 2005; 6: 152.

Vega: The Manual Annotation Database

The Vega database is a unique central repository for high quality, frequently updated, manual annotation of vertebrate finished genomic sequence.

Vega gives an accurate record of your gene of interest, including details of coding and non-coding variants and pseudogenes.

Vega currently includes the full clone annotation of 12 linkage groups and all mapped ZFIN cDNA's.



MapView showing chromosome 20

Vega release 35