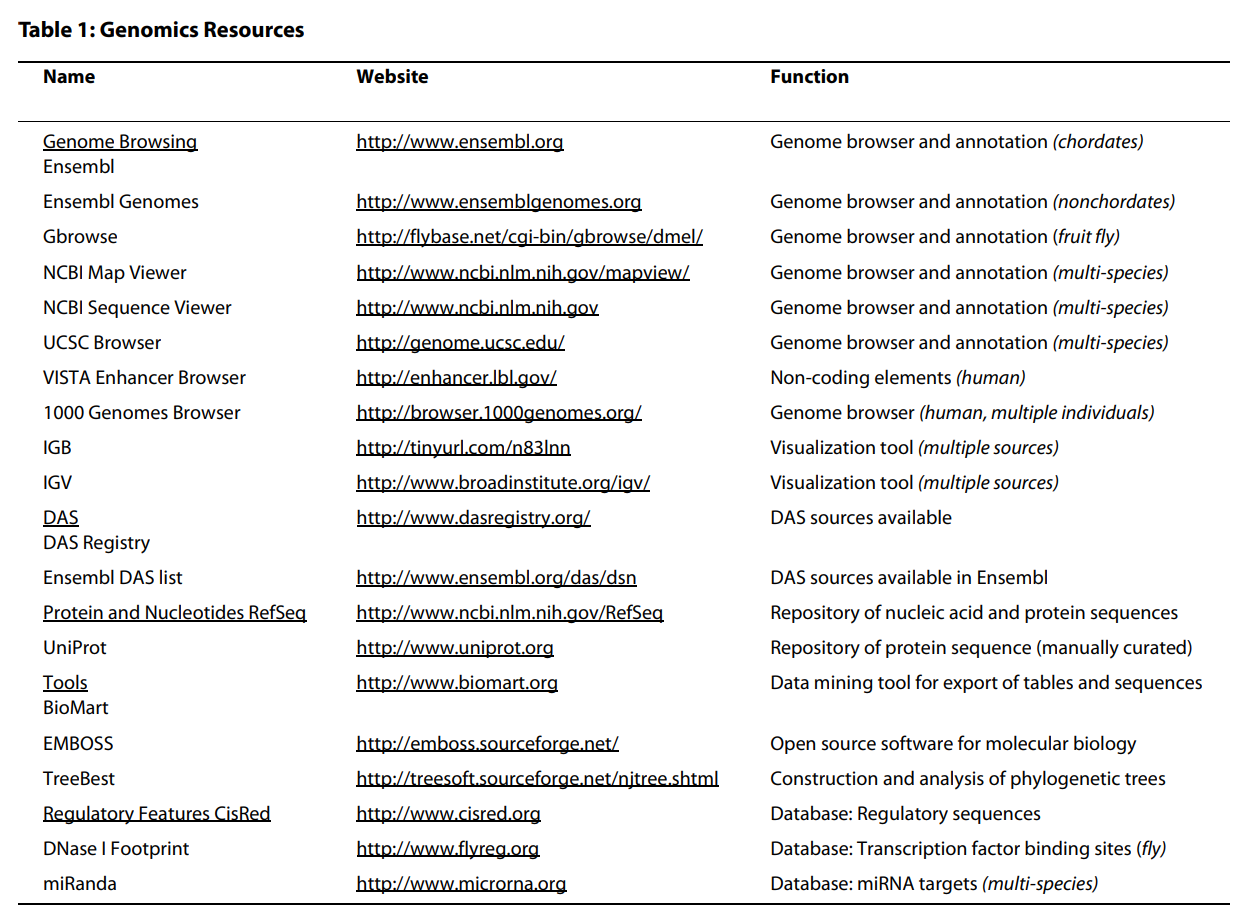
What is Genome Browser ?

* Genome browsers provide an aid to the researcher by importing biological data from various sources and presenting these data in an integrated way
* Three multi-species genome browsers are widely used by the scientific community: the UCSC genome browser, NCBI Map Viewer, and Ensembl. Others include H-INvDB or the FlyBase genome browser, and focus on one or a few species



Ensembl

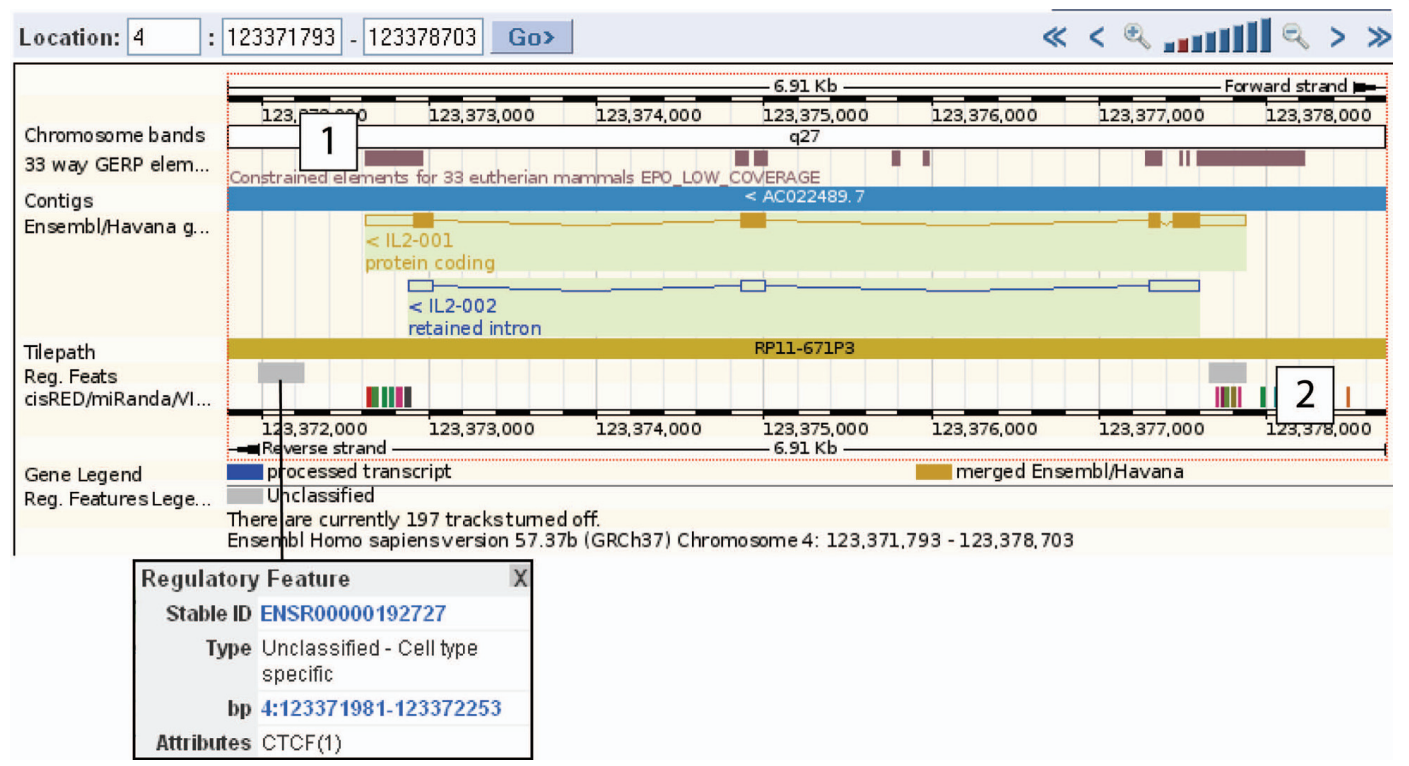
* The Ensembl project focuses on the chordate genomes, with the inclusion of additional model organisms that have been extensively studied in biological research and have a reliable, manually annotated gene set
* Ensembl includes annotation such as sequence variation, comparative associations, mRNA and protein from other databases, predicted features such as CpG islands , and repeats and motifs mapped along the genome.
* Homology relationships based on gene comparisons across all annotated species in Ensembl ,along with whole-genome alignments, such as alignments of 31 mammalian genomes, can be readily viewed in the browser.

Case Studies

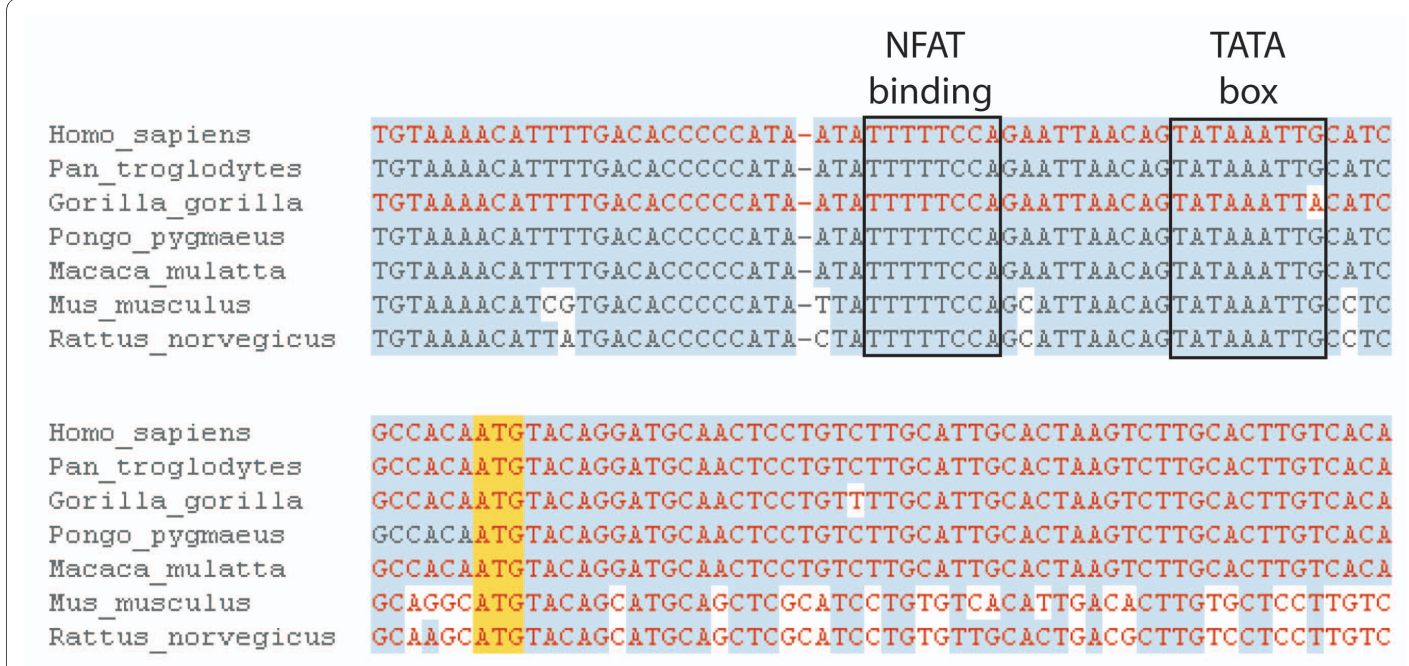
* In the following four case studies, we use the Ensembl genome browser to demonstrate how to view and predict functional regions in the genome
* First, we examine known regulatory features for the human IL2 gene and discuss how to display these features in Ensembl
* In study 2, we use human MYO6. Using comparative genomics, we show how the location of functional sequences may be predicted
* In case study 3, we demonstrate how the information in Ensembl can be extended through DAS (the Distributed Annotation System) to view data from external sources
* In study 4, we explore a variation associated with disease phenotypes.

Case Study 1 : Regulatory Regions for the IL2 Gene

* We investigate IL2, the interleukin 2 gene, in human (ENSG00000109471). Gene regulation has been studied at the 5' end of the IL2 transcript and flanking sequence.
* The ENCODE pilot study mapped promoter regions and regulatory sequences in 1% of the human genome, and this approach is now being extended to the entire genome.
* Ensembl has made a first attempt at annotating these sequences genome-wide by producing a 'regulatory build' based on data from ChIP-Chip[18] and ChIP-Seq [19] experiments
* The ensuing data in the 'Regulatory regions' track in Ensembl are for specific cell types, and include DNase I Hypersensitive sites, CCCTC-binding factor (CTCF) sites, and Histone modification sites
* The IL2 gene possesses features from the regulatory build on the flanking regions to the IL2 transcript



* To look more closely at the nucleotide sequence itself, we can view an alignment of the upstream region of the IL2 gene across mammals at the base pair level

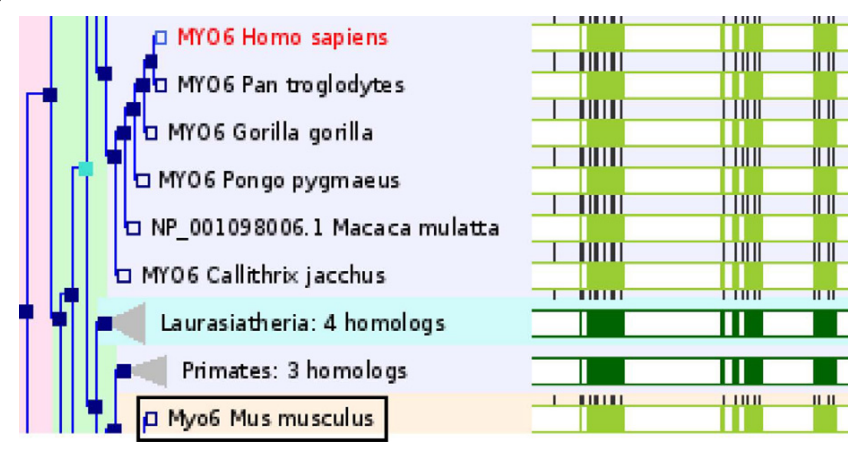


* The sequence in this region is highly conserved across the eutherian mammals shown. The presence of the NFAT (nuclear factor of activated T cells) binding site and TATA box (in the promoter region) for the IL2 gene are boxed, along with the translational start site (ATG)

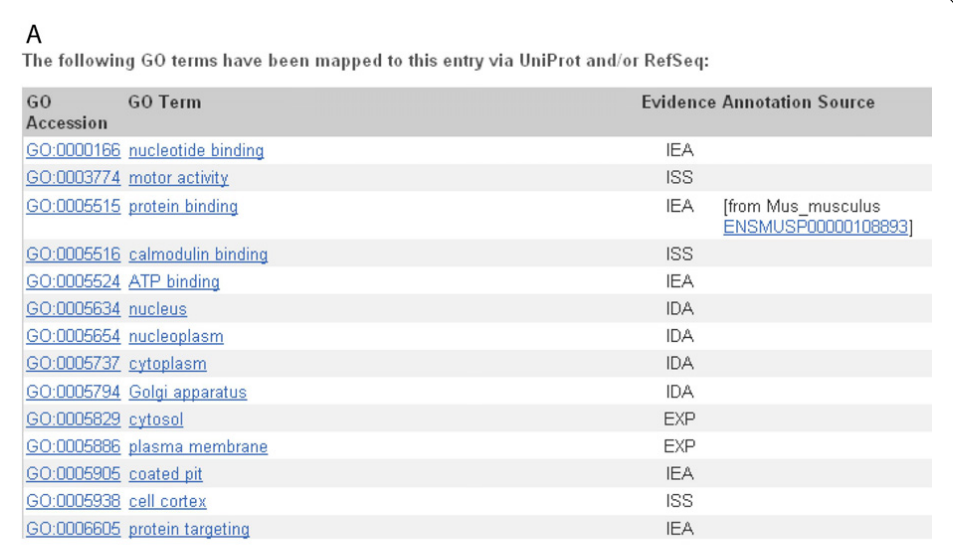
This is to illustrate how to view conserved regions in a sequence, and how rich the 5' sequence and flank can be in terms of binding sites and regulatory elements

Case Study 2 : Function for a Gene

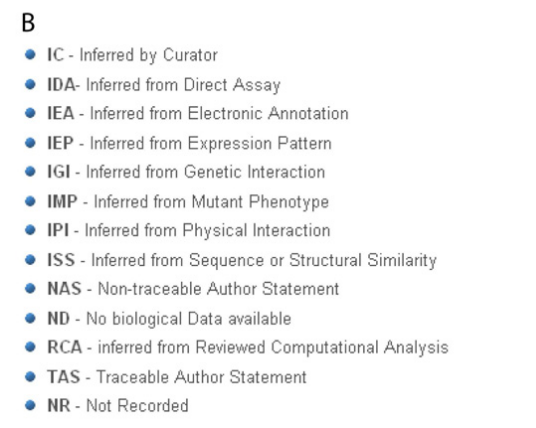
* In this example we consider human MYO6, ENSG00000196586, which has been studied in the mouse model to understand its role in endocytosis and inner-ear development
* Orthologues and paralogues in Ensembl are determined using phylogenetic gene trees across all available species



* Protein relationships are clustered into a tree diagram with clickable nodes depicting taxonomic clades, evolutionary events, and links to protein alignments using JalView.
* Red nodes correspond to duplication events, dark blue nodes show speciation events, and light blue nodes are ambiguous duplications The filled green rectangles at the right demonstrate protein alignments. Light green alignments represent one protein, dark green shading shows a consensus alignment for a collapsed node in the tree. Black ticks in the green bars show positions of introns
* The tree in the figure shows the human MYO6 protein in red. The mouse orthologue is boxed.

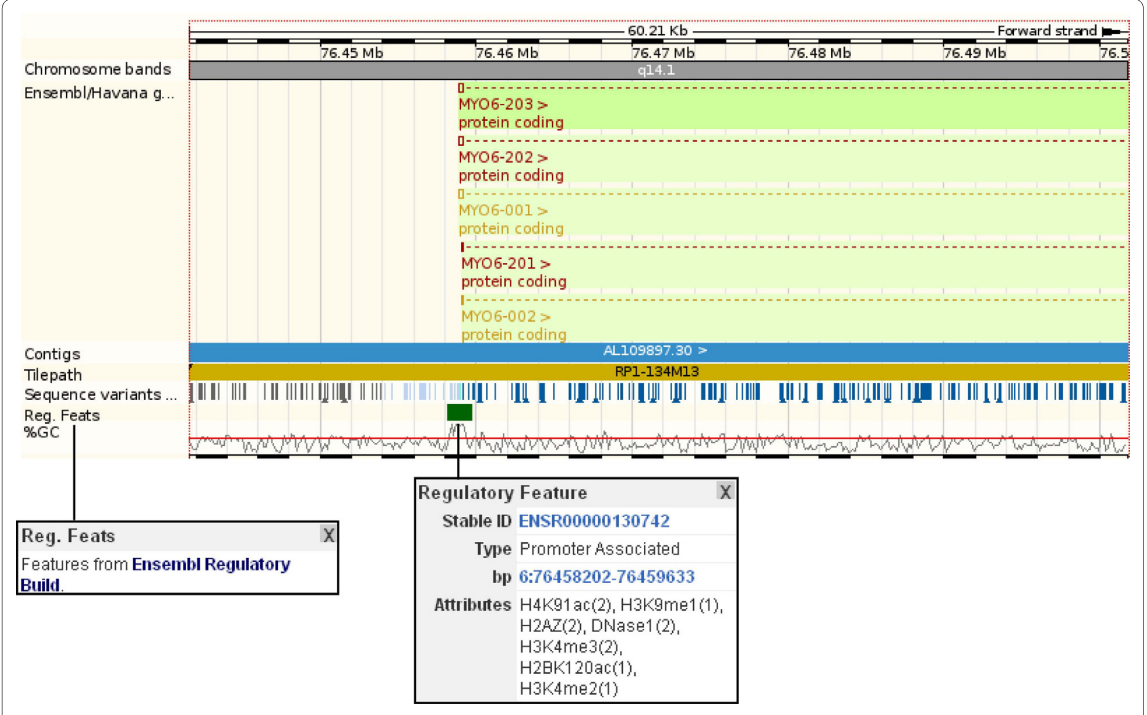


GO terms listed for Homo sapiens MYO6 transcript ENST00000428345. Classifications inferred by comparison to the mouse homologue have evidence code IEA.



* Description of evidence codes showing how a transcript was assigned to a GO term

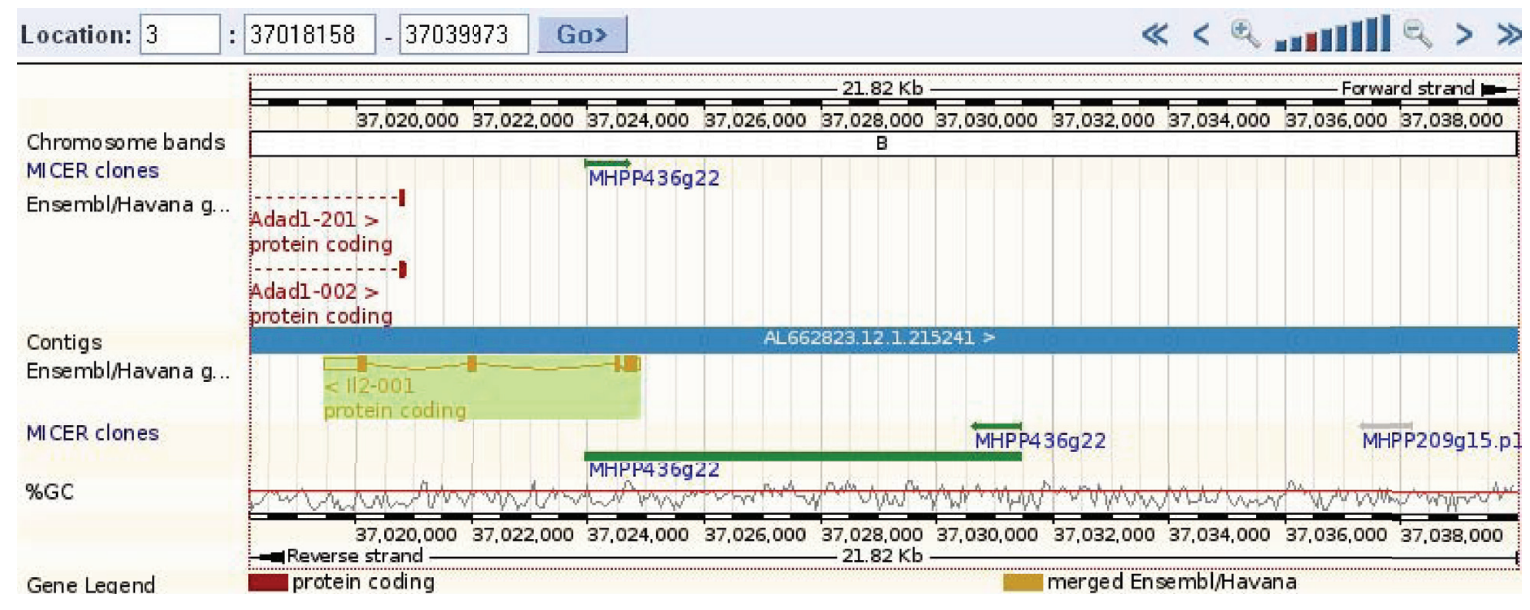
Predicted Regulatory Features for Myosin 6



* The 'regulatory features' track in the 'region in detail' view reveal DNase I hypersensitive sites and numerous histone modification and methylation sites aligning to the 5'UTR and upstream region of MYO6 transcripts

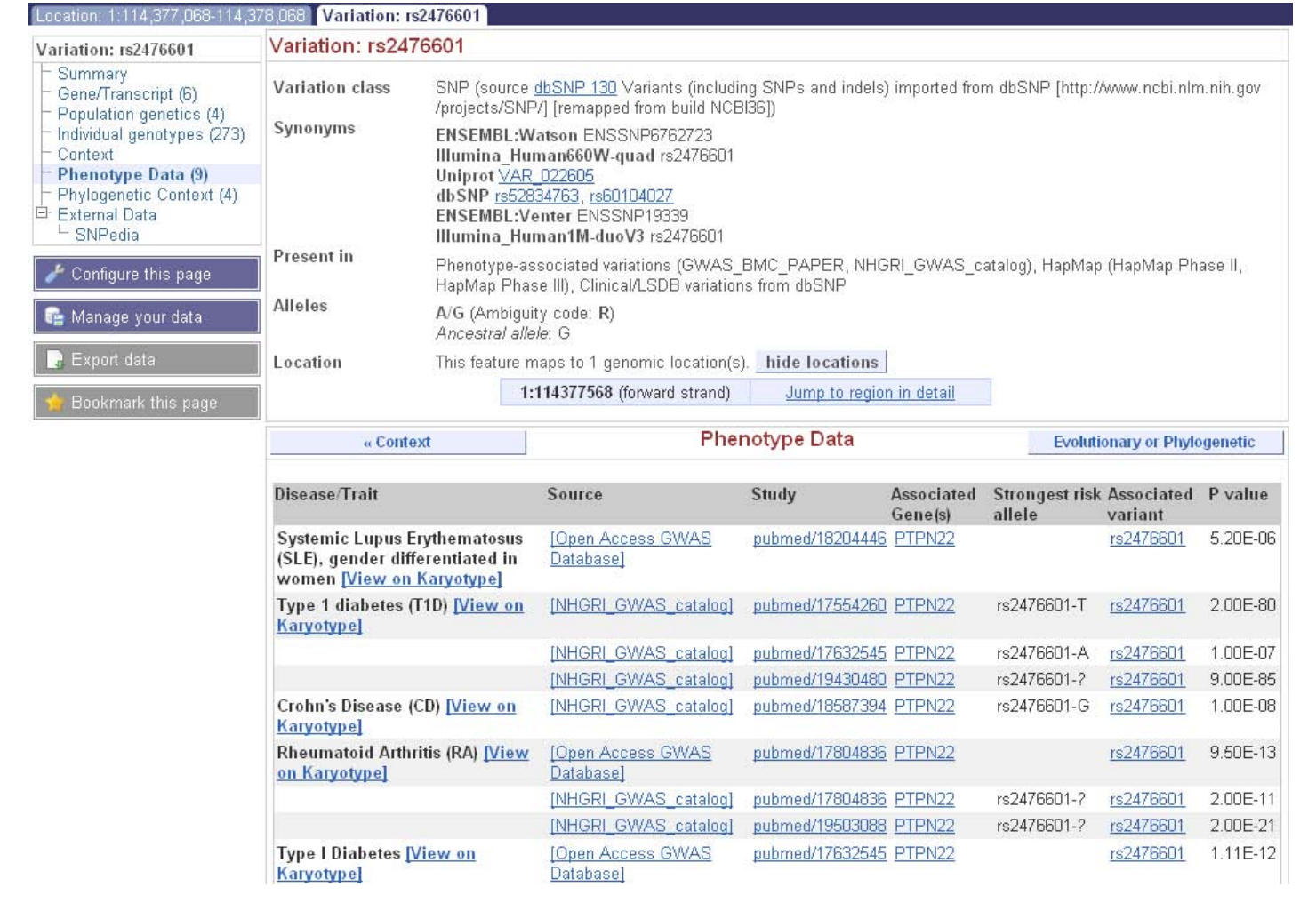
Case Study 3: Viewing information outside Ensembl databases

* The Distributed Annotation System (DAS) allows Ensembl to link out to and display information from external databases in supported formats. DAS transforms Ensembl into a framework where third party annotation can be added and viewed alongside Ensembl annotation.

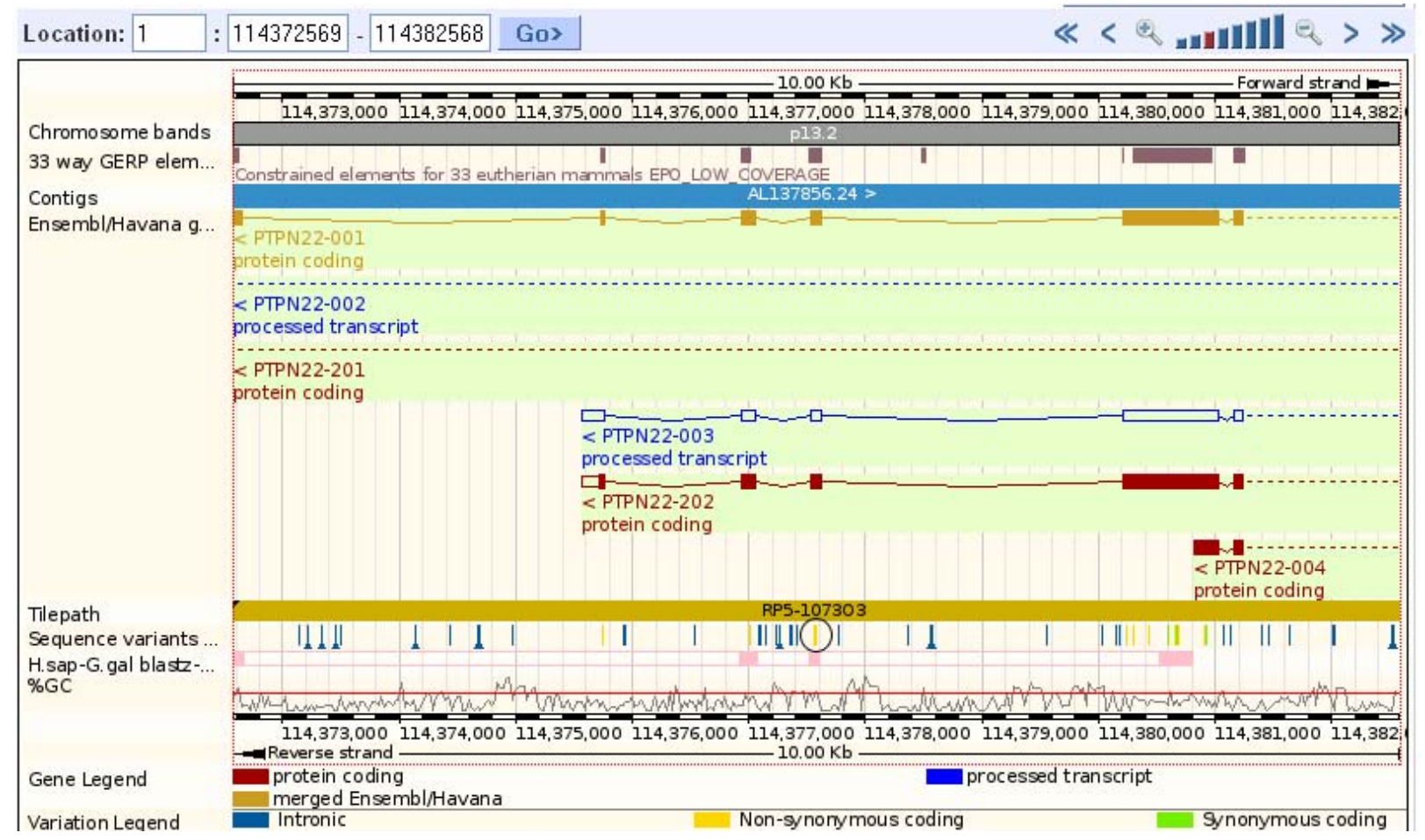


This figure demonstrates how to view external data using DAS along the genome. Data from the MICER project (a resource containing vectors and information to generate knock-out mice) is drawn for a region of the mouse genome

Case Study 4 : From phenotype to SNP- exploring variation



* The phenotype data link at the left reveals this variation is implicated in several diseases, including Rheumatoid Arthritis and Crohn's Disease



* The region around rs2476601 is shown, with the non-synonymous SNPs in yellow. rs2476601 is circled, and it aligns with an exon found in two coding, and one non-coding, splice variants of the PTPN22 gene