

## Module 4 – Region in Detail

### You will learn about

- The **Region in Detail** layout
- How to view genes and other information on a chromosomal region

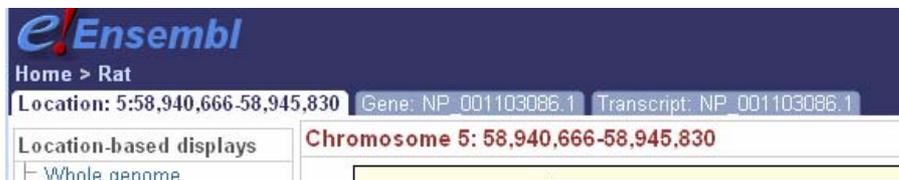
The **Location** pages hold a wealth of information and allow the customised display of genes and other information along the genome. In this example, let's use the rat assembly in order to explore a species other than human (which was addressed in other modules).

Note, the examples in this view were taken from Ensembl release 52.

To find the **Region in Detail** page for rat gene 'ENSRNOG00000039559', search for this identifier in the main page. Then click on the link circled below from the search result.

Your query matched 1 entries in the search database  
 Ensembl protein\_coding Gene: [ENSRNOG00000039559](#) (RefSeq peptide: NP\_001103086.1) [Region in detail](#)  
 Ensembl protein\_coding gene [ENSRNOG00000039559](#) has 2 transcripts: ENSRNOT00000018338, ENSRNO100000060642  
 ENSRNOP000000057375 and 7 exons: ENSRNOE00000127862, ENSRNOE00000128010, ENSRNOE00000128170, ENSRN  
 ENSRNOE00000294793, ENSRNOE00000411029

Alternatively, from the **Gene** or transcript page, click on the **Location** tab.



The '**Region in Detail**' page should now be shown, depicting a region of rat chromosome 5 centred on our gene of interest.

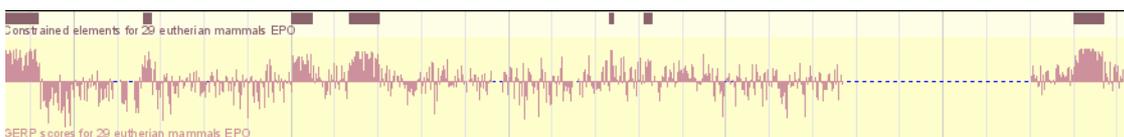
[http://Dec2008.archive.ensembl.org/Rattus\\_norvegicus/Location/View?db=core:g=ENSRNOG00000039559;r=5:58940666-58945830;t=ENSRNOT00000018338](http://Dec2008.archive.ensembl.org/Rattus_norvegicus/Location/View?db=core:g=ENSRNOG00000039559;r=5:58940666-58945830;t=ENSRNOT00000018338)

This page is separated into 3 panels. The first, labeled '1' in the figure below, shows chromosome 5 with the position of rat gene ENSRNOG00000039559 marked in red. The little red box on the chromosome is expanded into the 'Top' panel, showing a 1Mb region of the chromosome, centred on ENSRNOG00000039559. Two gaps in the sequence assembly are present, indicated by interruptions in the blue bar. Individual **contigs** that make up the area of the sequence assembly are shown in alternating dark and light blue colours. Click on any of the genes shown in this panel for links to their gene summary pages.



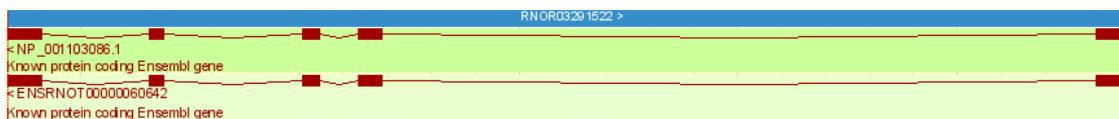
Panel 3 is the main panel, and is highly customisable. By default, it is zoomed in to our transcripts of interest. Use the zoom controls marked by the '2' in the figure above to explore a larger region. For larger regions, you may have to use the 'Region overview' link in the left hand menu.

The 'conservation' or 'constrained elements' track is shown. This track shows a conservation score for every nucleotide calculated across 29 species.



The plot above shows the 'Gerp' score for each nucleotide in the 29 species alignment. If the nucleotide is identical across all 29 species, the conservation score will be high, and if several adjacent nucleotides are highly conserved, a dark module is drawn above the plot. Customise this display using 'Configure this page'.

Below the chromosome, two Ensembl transcripts are drawn. They are on the reverse strand of the chromosome. The transcripts are labelled 'NP\_001103086.1' and 'ENSRNOT0000060642', respectively. Click on a transcript to jump to more information about it.



Vertebrate cDNA sequences in EMBL-Bank are drawn below the transcripts. These are cDNAs that align to this region of the rat genome, based on sequence matching. Click on any cDNA to find its identifier. Or, click on the track name 'EMBL vertebrate cDNA' for more information about the set.

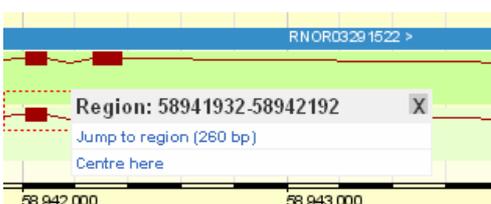
Let's customise this display. Click on '**Configure this page**' at the left. Choose '**cDNA/mRNA alignments**' at the left. Turn off these alignments by clicking on the box to the left of each transcript type and choosing '**off**'. Now click '**Variations**' and select '**All variations**'. '**SAVE and close**' to find a new track drawn on the **Region in detail** page. Click on any vertical line (variation) for a menu about the SNP (single nucleotide polymorphism) or InDel (insertion deletion mutation). Clicking on '**Variation properties**' in the box shown below will bring you to an information page for the genetic variation, including links to population frequencies, if known.

**Variation: ENSRNOSNP2008817**

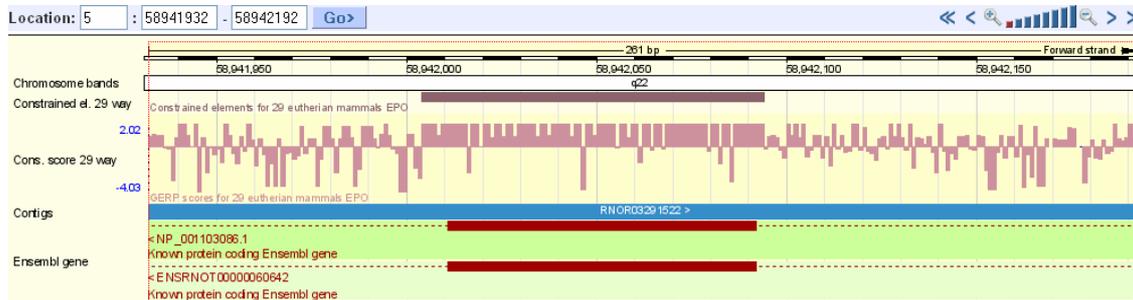
Variation Properties

bp:	58943989
status:	-
class:	snp
ambiguity code:	R
mapweight:	1
alleles:	A/G
source:	ENSEMBL:ceiera
type:	INTRONIC

How can we zoom in and out, or change the region on the **Detailed view** panel? Either use the ladder (or triangle) labelled **2** in the figure showing the whole page, or type in a new region (chromosome and/or base pairs). Or, zoom closer in to the transcript by clicking and dragging the mouse around an exon.



The link '**Jump to region**' shows a zoomed-in view. (Note: if you turn on '**Sequence**' in '**Configure this page**', the genomic sequence will be revealed when you zoom in.)



We have zoomed in to a highly conserved exon (note the conservation scores and corresponding module of constrained elements.) The dotted lines on either side of the exon signify that the transcript extends beyond our view.

## Glossary

**Contig** A continuous sequence module forming part of the genomic assembly

**DAS The Distributed Annotation System**- a platform for viewing and integrating external data

**Constrained elements (In Ensembl)** Sequence conservation on the nucleotide level as calculated across multiple species in whole genome alignments.

**Gerp score** A score assigned to sequence within an alignment in order to determine how strongly it is conserved., to determine the constrained elements. See **reference**.

**Known Ensembl gene** Transcripts determined by the Ensembl genebuild that match to a gene sequence found in databases such as UniProt and RefSeq for the same species.

**Merged Known Ensembl/Havana gene** An Ensembl known gene that possesses at least one transcript that has been agreed upon by the Havana team.

**Novel Ensembl gene** Transcripts determined by the Ensembl genebuild that match to a gene sequence found in databases such as UniProt and RefSeq for a similar species, but not the same species.

**QTL- Quantitative Trait Locus** Stretches of sequences that correlate to phenotypes.

**SNPs Single Nucleotide Polymorphisms** - base pair changes that occur at a fixed frequency in a population (usually 1% or higher). Natural variation.

**Unigene** Database of sequences that appear to come from the same transcription locus. Information is housed for a large variety of species.

<http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene>

## ***Further reading***

Cooper et. al., Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* 15:901-913, 2005

[www.ensembl.org/info/docs/compara/index.html](http://www.ensembl.org/info/docs/compara/index.html)

## ***What to do next***

Learn how to use our BioMart tool in **module 5!**

**Watch the related video!**

[www.ensembl.org/info/website/tutorials/index.html](http://www.ensembl.org/info/website/tutorials/index.html)

**'Region in Detail'**

---