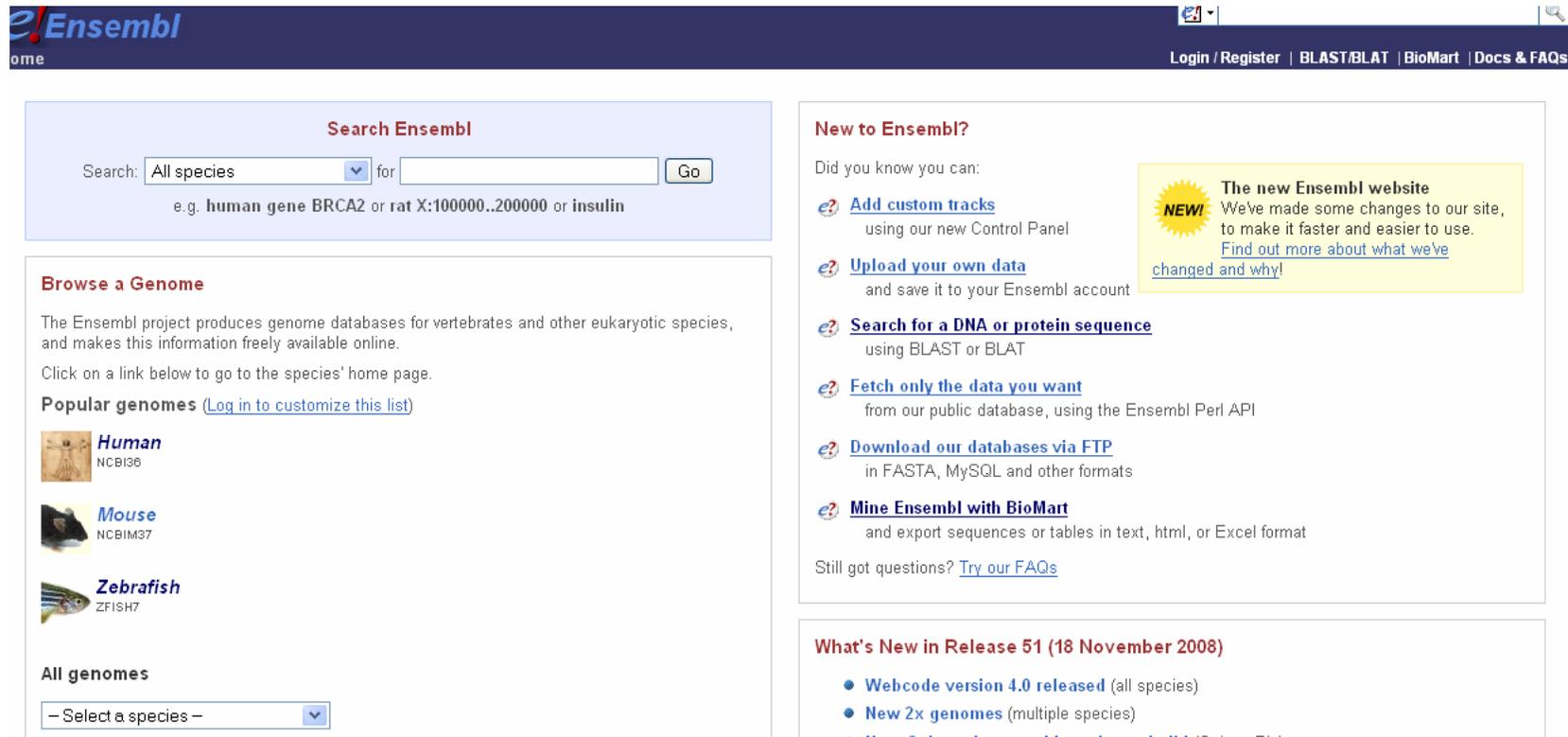


Genomes with Ensembl



The screenshot shows the Ensembl genome browser homepage. At the top left is the Ensembl logo. The main navigation bar includes links for 'Login / Register', 'BLAST/BLAT', 'BioMart', and 'Docs & FAQs'. The central search area is titled 'Search Ensembl' and contains a search box with a dropdown menu set to 'All species' and a 'Go' button. Below the search box, it provides an example: 'e.g. human gene BRCA2 or rat X:100000..200000 or insulin'. To the right of the search area is a 'New to Ensembl?' section with a list of links: 'Add custom tracks', 'Upload your own data', 'Search for a DNA or protein sequence', 'Fetch only the data you want', 'Download our databases via FTP', and 'Mine Ensembl with BioMart'. A yellow callout box highlights 'The new Ensembl website' with a 'NEW!' starburst icon. Below this is a 'What's New in Release 51 (18 November 2008)' section with a list of updates: 'Webcode version 4.0 released', 'New 2x genomes', and 'New Guinea assembly and rebuild'. On the left side, there is a 'Browse a Genome' section with a description of the project and a list of popular genomes: Human (NCBI36), Mouse (NCBIM37), and Zebrafish (ZFISH7). At the bottom left, there is a dropdown menu for 'All genomes' with the text '- Select a species -'.

Subjects

An introduction to genome browsers.

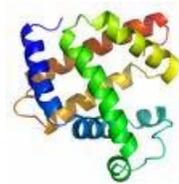
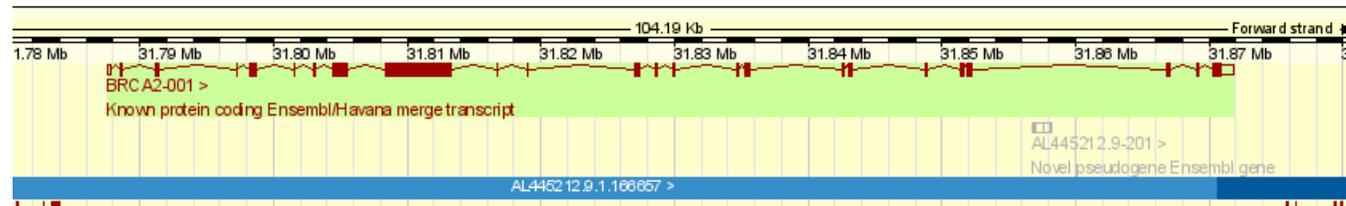
What is an Ensembl gene?

How can we extract data from Ensembl?

Genome Browsers

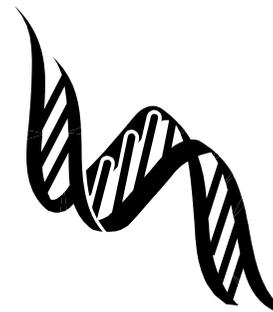
- Ensembl Genome browser
<http://www.ensembl.org>
- NCBI Map Viewer
<http://www.ncbi.nlm.nih.gov/mapview/>
- UCSC Genome Browser
<http://genome.ucsc.edu>

Ensembl Genome Browser



Protein/ mRNA

+



Sequence Assembly



Ensembl Genes



NCBI Map Viewer



NCBI  NCBI Map Viewer

PubMed Entrez BLAST OMIM Taxonomy Structure

Search Find Find in This View Advanced Search [BLAST The Human Genome](#)

Homo sapiens (human) Build 36.2 (Current)

Chromosome: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [**13**] [14](#) [15](#) [16](#) [17](#) [18](#) [19](#) [20](#) [21](#) [22](#) [X](#) [Y](#) [MT](#)

Master Map: **Genes On Sequence** [Summary of Maps](#) [Maps & Options](#)

Region Displayed: 31,788K-31,872K bp [Download/View Sequence/Evidence](#)

Ideogram Contig Hs UniG Genes_seq Symbol Links E Cyto Description

31790K
31800K
31810K
31820K
31830K
31840K
31850K
31860K

NT_024524.1
Unknown
Hs_34612
Hs_161228

BRCA2 + [OMIM](#) [HGNC](#) [sv](#) [pr](#) [dl](#) [ev](#) [mm](#) [hm](#) [sts](#) [CCDS](#) [SNP](#) best RefSeq 13q12.3 breast cancer 2, early onset

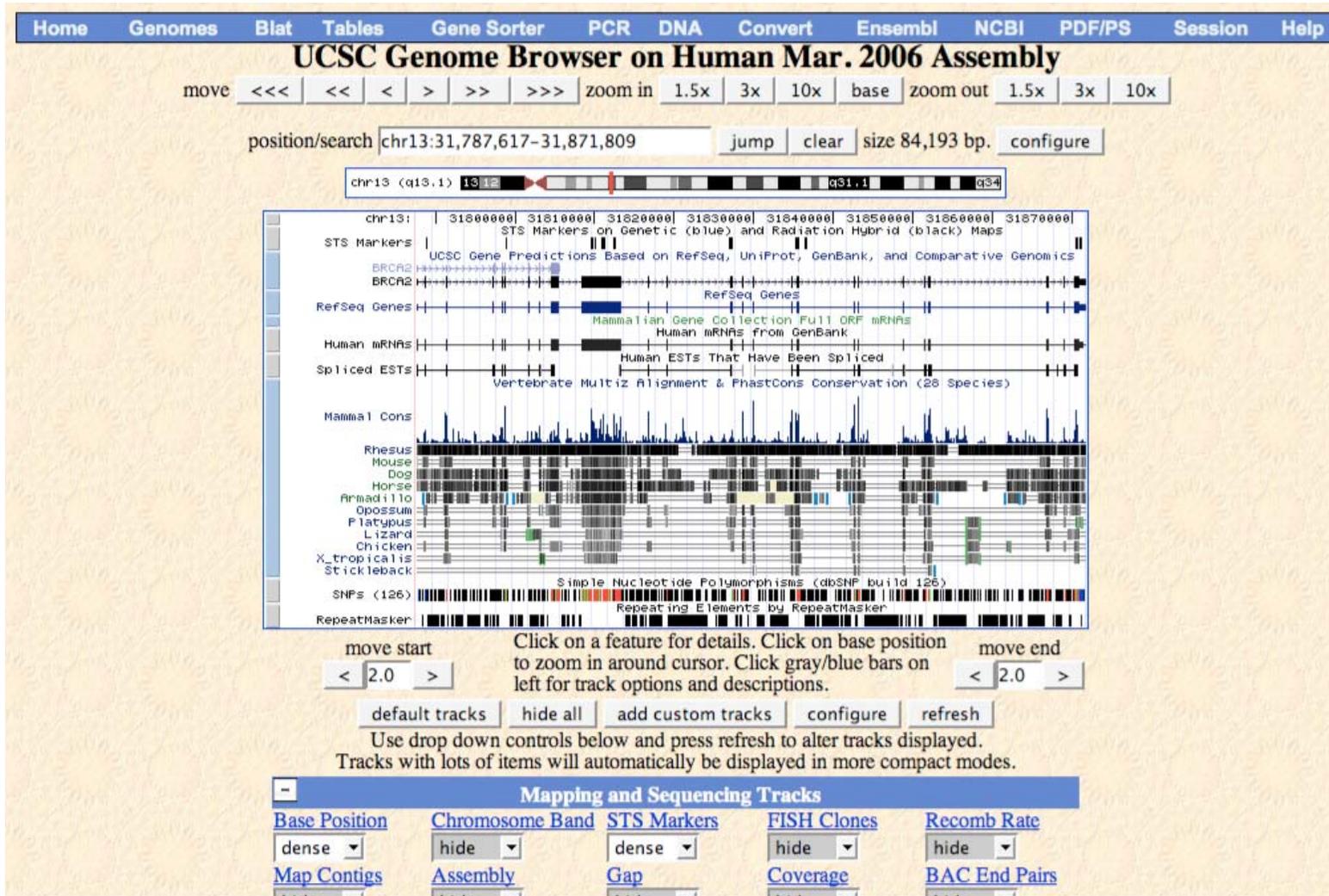
IFT1P + [HGNC](#) [sv](#) [dl](#) [ev](#) [mm](#) best RefSeq 13q12-q13 interferon-induced protein wi

Map Viewer Help
Human Maps Help
FTP
Data As Table View
[Maps & Options](#)
Compress Map
Region Shown:

out
zoom
in
You are here:
Ideogram
13q13
13q12
13q11.2
13q11.1
13q11
13q12
13q13
13q14
13q21
13q22
13q21
13q22
13q23
13q33
13q34
default
master



UCSC Genome Browser



The screenshot displays the UCSC Genome Browser interface for the Human Mar. 2006 Assembly. The main track shows the BCR2 gene on chromosome 13 (q13.1). The interface includes navigation controls (move, zoom in/out), a search bar with the coordinates chr13:31,787,617-31,871,809, and a detailed track view. The track view includes STS Markers, RefSeq Genes, Human mRNAs, Spliced ESTs, Mammal Cons, Rhesus, Mouse, Dog, Horse, Armadillo, Opossum, Platypus, Lizard, Chicken, X_tropicalis, Stickleback, SNPs (126), and RepeatMasker. Below the track view are controls for zooming and track configuration.

Home Genomes Blat Tables Gene Sorter PCR DNA Convert Ensembl NCBI PDF/PS Session Help

UCSC Genome Browser on Human Mar. 2006 Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

position/search chr13:31,787,617-31,871,809 jump clear size 84,193 bp. configure

chr13 (q13.1) 13 12 q31.1 q34

chr13: 31800000 31810000 31820000 31830000 31840000 31850000 31860000 31870000

STS Markers STS Markers on Genetic (blue) and Radiation Hybrid (black) Maps

UCSC Gene Predictions Based on RefSeq, UniProt, GenBank, and Comparative Genomics

RefSeq Genes BCR2 RefSeq Genes

Human mRNAs Mammalian Gene Collection Full ORF mRNAs Human mRNAs from GenBank

Spliced ESTs Human ESTs That Have Been Spliced

Mammal Cons Vertebrate Multiz Alignment & PhastCons Conservation (28 Species)

Rhesus Mouse Dog Horse Armadillo Opossum Platypus Lizard Chicken X_tropicalis Stickleback

SNPs (126) Simple Nucleotide Polymorphisms (dbSNP build 126)

RepeatMasker Repeating Elements by RepeatMasker

move start Click on a feature for details. Click on base position to zoom in around cursor. Click gray/blue bars on left for track options and descriptions. move end

< 2.0 >

default tracks hide all add custom tracks configure refresh

Use drop down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes.

Mapping and Sequencing Tracks				
Base Position	Chromosome Band	STS Markers	FISH Clones	Recomb Rate
dense ▾	hide ▾	dense ▾	hide ▾	hide ▾
Map Contigs	Assembly	Gap	Coverage	BAC End Pairs
hide ▾	hide ▾	hide ▾	hide ▾	hide ▾

To meet a challenge...

Ensembl's AIM: To provide annotation for the biological community that is freely available and of high quality

- **Started in 2000**
- **Joint project between EBI and Sanger**
- **Funded primarily by the Wellcome Trust, additional funding by EMBL, NIH-NIAID, EU, BBSRC and MRC**

What Distinguishes Ensembl from the UCSC and NCBI Browsers?

- **The gene set. Automatic annotation based on mRNA and protein information.**
- **Programmatic access via the Perl API (open source)**
- **BioMart**
- **Integration with other databases (DAS)**
- **Comparative analysis (gene trees)**

Ensembl Genes – biological basis

All Ensembl transcripts are based on proteins and mRNAs in:

- UniProt/Swiss-Prot (*manually curated*)
- UniProt/TrEMBL

www.uniprot.org

- NCBI RefSeq (*manually curated*)

www.ncbi.nlm.nih.gov/RefSeq

Ensembl and VEGA/Havana

- **Automatic annotation pipeline: Gene building all at once (whole genome)**

Ensembl

- **Manual curation: case-by-case basis**

VEGA: Vertebrate Genome Annotation

Havana

Genes and Transcripts in Ensembl

- **Ensembl known transcripts**
- **Ensembl novel transcripts**
- **Ensembl merged transcripts (Havana)**

- **EST clusters**

- **More manual curation (SGD, WormBase, FlyBase)**

What annotation is available?

- Gene/transcript/peptide models (coding and noncoding (ncRNAs))
- IDs in other databases
- Mapped cDNAs, peptides, micro array probes, BAC clones etc.
- Other features of the genome:
cytogenetic bands, markers, repeats etc.
- Comparative data:
orthologues and paralogues, protein families, whole genome alignments, syntenic regions
- Variation data:
SNPs
- Regulatory data:
“best guess” set of regulatory elements from ENCODE
- Data from external sources (DAS)



What species are available (v52)? Chordate focus.

 Aedes	 Fruitfly	 Opossum
 Alpaca	 Fugu	 Orangutan
 Anole lizard (preview - assembly only)	 Guinea Pig	 Pig (preview - assembly only)
 Anopheles	 Hedgehog	 Pika
 Armadillo	 Horse	 Platypus
 Bushbaby	 Human	 Rabbit
 C.elegans	 Hyrax	 Rat
 C.intestinalis	 Kangaroo rat	 S.cerevisiae
 C.savianvi	 Lamprev (preview - assembly only)	 Shrew
 Cat	 Lesser hedgehog tenrec	 Squirrel
 Chicken	 Macaque	 Stickleback
 Chimpanzee	 Medaka	 Tarsier
 Cow	 Megabat	 Tetraodon
 Dog	 Microbat	 Tree Shrew
 Dolphin	 Mouse	 X.tropicalis
 Elephant	 Mouse Lemur	 Zebrafish

How is this information organised?

- **Ensembl Views (Website)**



- **Ensembl Database (open source)**

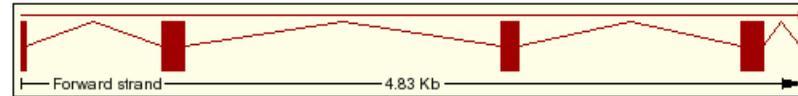
- **BioMart 'DataMining tool'**

Names in Ensembl

- **ENSG###** Ensembl Gene ID
 - **ENST###** Ensembl Transcript ID
 - **ENSP###** Ensembl Peptide ID
 - **ENSE###** Ensembl Exon ID
-
- For other species than human a suffix is added:
MUS (*Mus musculus*) for mouse: **ENSMUSG###**
DAR (*Danio rerio*) for zebrafish: **ENSDARG###**, etc.

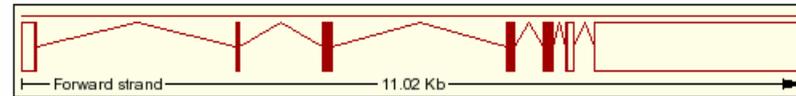
Gene Structure in Ensembl

No UTRs



Calmodulin Chicken

UTRs annotated



Calmodulin Human



Help and Information

- Comments and questions?
helpdesk@ensembl.org
- View our help videos
- Mailing lists:
ensembl-dev@ebi.ac.uk
ensembl-announce@ebi.ac.uk
- Come visit our blog! <http://ensembl.blogspot.com/>



Cambridge



Ensembl Team

Ensembl	<u>Paul Flicek (EBI)</u> , <u>Steve Searle (Sanger Institute)</u>
Vertebrate Genomics	Mario Caccamo, Laura Clark, Jonathan Hinton, Zam Iqbal, Vasudev Kumanduri, Ilkka Lappalainen
Software	<u>Glenn Proctor</u> , Syed Haider, Andrew Jenkinson, Andreas Kähäri, Stephen Keenan, Rhoda Kinsella, Eugene Kulesha, Ian Longden, Daniel Rios
Comparative Genomics	<u>Javier Herrero</u> , Kathryn Beal, Benoît Ballester, Stephen Fitzgerald, Leo Gordon, Albert Vilella
Functional Genomics	Nathan Johnson, Stefan Gräf, Steven Wilder
Variation	<u>Fiona Cunningham</u> , Yuan Chen
Analysis and Annotation	Bronwen Aken, Julio Banet, Susan Fairley, Jan-Hinnerck Vogel, Simon White, Amonida Zadissa
Web Team	<u>James Smith</u> , Eugene Bragin, Anne Parker, Bethan Pritchard, Steve Trevanion (VEGA)
Zebrafish	<u>Kerstin Howe</u> , Britt Reimholz, James Torrance
VectorBase	<u>Dan Lawson</u> , Martin Hammond, Karyn Megy
Outreach	<u>Xosé M Fernández</u> , Bert Overduin, Michael Schuster (QC), Giulietta Spudich
Systems & Support	Guy Coates, Tim Cutts, Shelley Goddard
Research	Ian Dunham, Damian Keefe, Alison Meynert, Dace Ruklisa, Guy Slater, Daniel Zerbino
Ensembl Strategy	Ewan Birney, Richard Durbin, Tim Hubbard