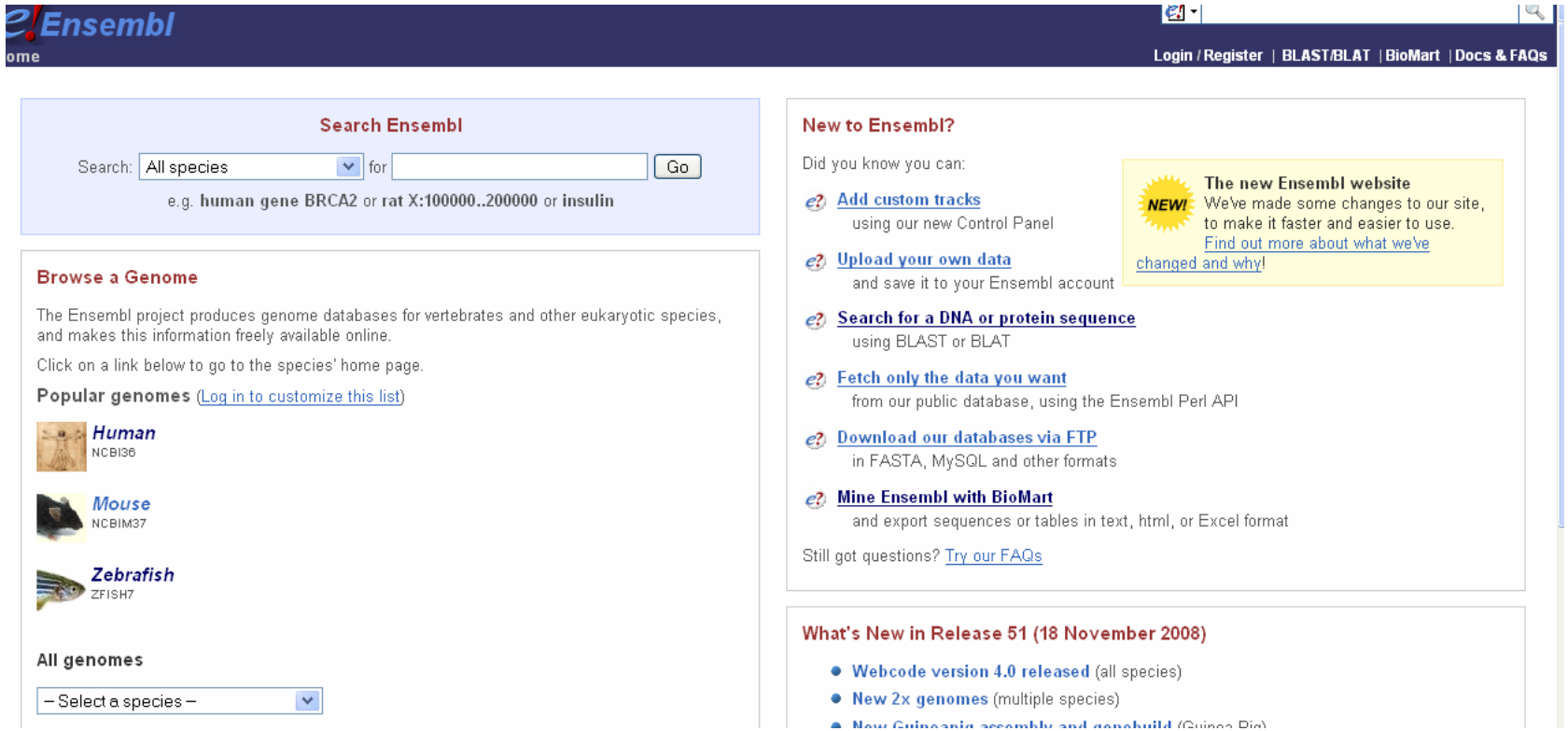
Genomes with Ensembl



Subjects

An introduction to genome browsers.

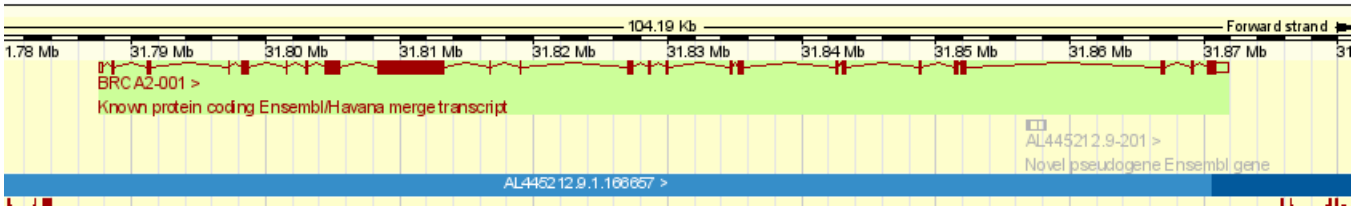
What is an Ensembl gene?

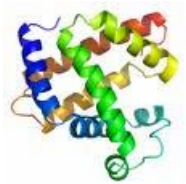
How can we extract data from Ensembl?

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





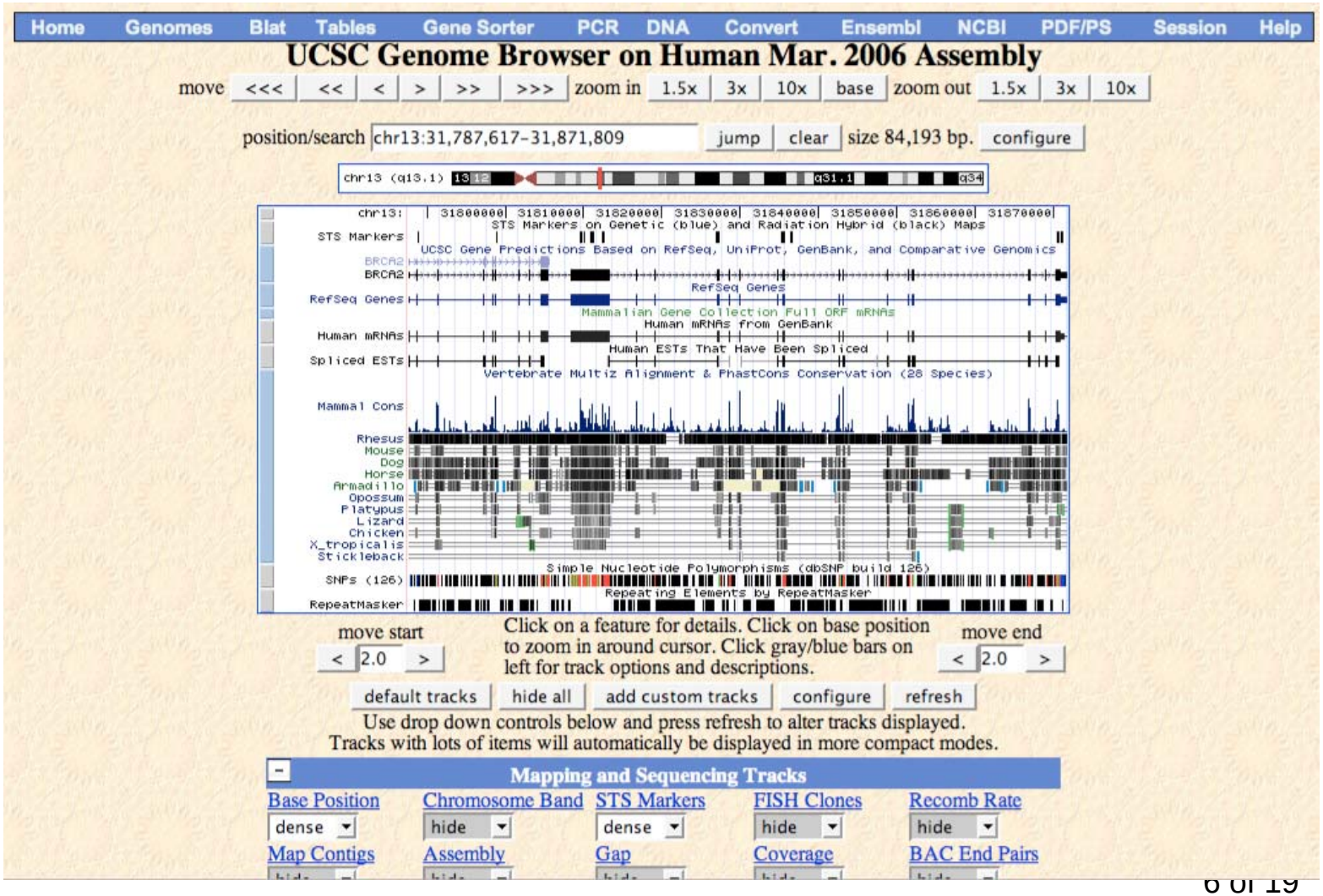
+

→ **Ensembl Genes**

Protein/ mRNA Sequence Assembly

NCBI Map Viewer

UCSC Genome Browser



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:

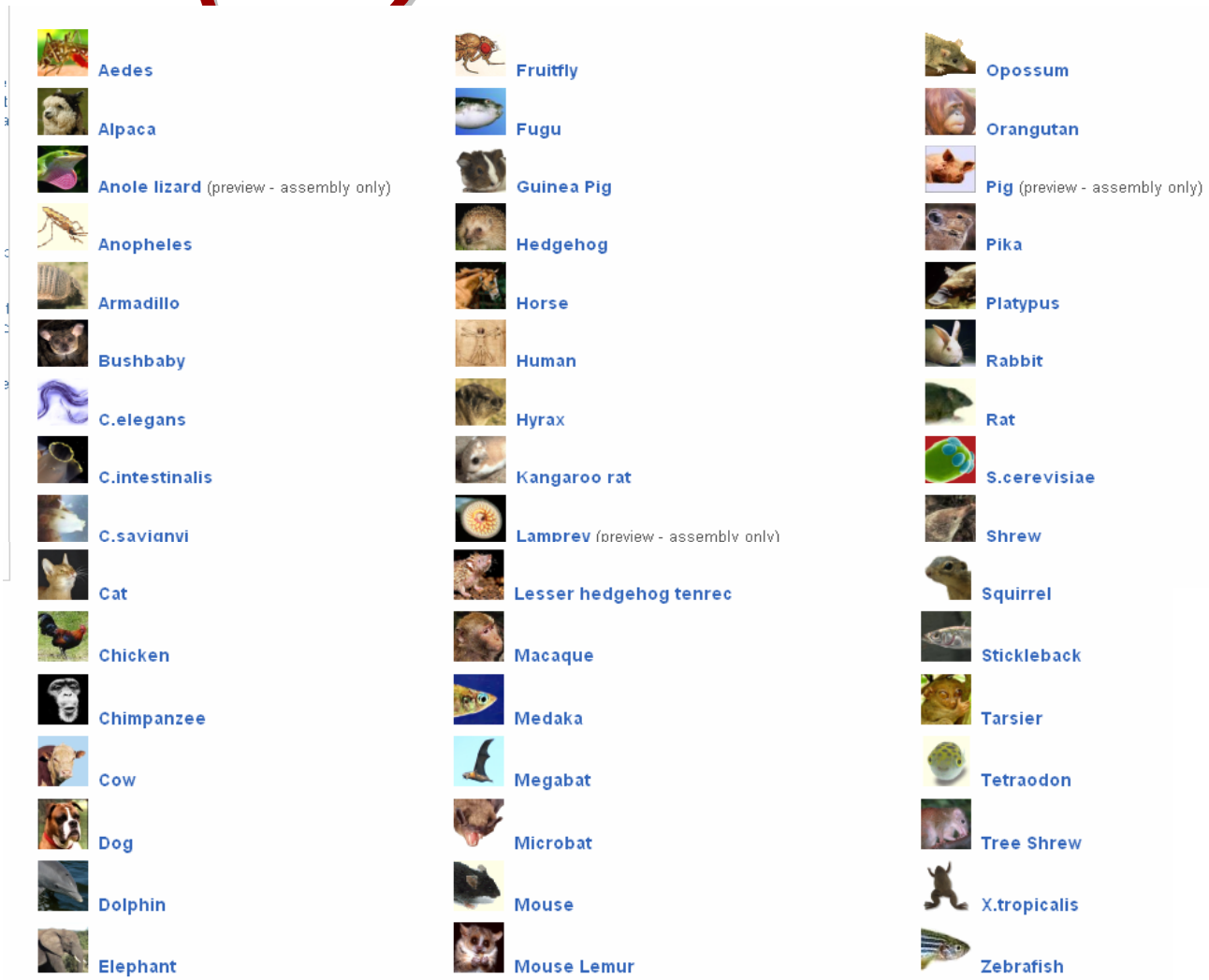
  cytogenetic bands, markers, repeats etc.
- Comparative data:

  orthologues and paralogues, protein families, whole genome alignments, syntenic regions
- Variation data:

  SNPs
- Regulatory data:

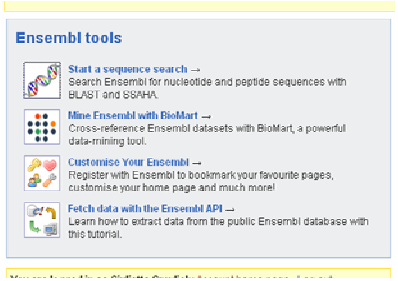
  “best guess” set of regulatory elements from ENCODE
- Data from external sources (DAS)

What species are available (v52)? Chordate focus.



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**

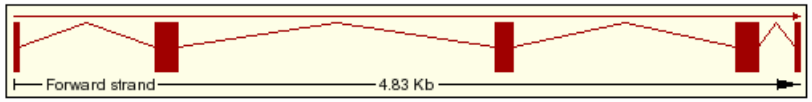
- **For other species than human a suffix is added:**

MUS (*Mus musculus*) for mouse: ENSMUSG###

DAR (*Danio rerio*) for zebrafish: ENSDARG###, etc.

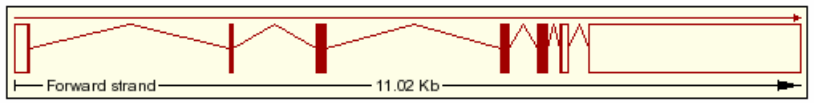
Gene Structure in Ensembl

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

© Sean T. McHugh (http://www.cambridgeincolour.com)

Ensembl Team

Ensembl	Paul Flicek **(*EBI*)**, Steve Searle **(Sanger Institute)**
Vertebrate Genomics	Mario Caccamo, Laura Clark, Jonathan Hinton, Zam Iqbal, Vasudev Kumanduri, Ilkka Lappalainen
Software	Glenn Proctor, Syed Haider, Andrew Jenkinson, Andreas Kähäri, Stephen Keenan, Rhoda Kinsella, Eugene Kulesha, Ian Longden, Daniel Rios
Comparative Genomics	Javier Herrero, Kathryn Beal, Benoît Ballester, Stephen Fitzgerald, Leo Gordon, Albert Vilella
Functional Genomics	Nathan Johnson, Stefan Gräf, Steven Wilder
Variation	Fiona Cunningham, Yuan Chen
Analysis and Annotation	Bronwen Aken, Julio Banet, Susan Fairley, Jan-Hinnerck Vogel, Simon White, Amonida Zadissa
Web Team	James Smith, Eugene Bragin, Anne Parker, Bethan Pritchard, Steve Trevanion (VEGA)
Zebrafish	Kerstin Howe, Britt Reimholz, James Torrance
VectorBase	Dan Lawson, Martin Hammond, Karyn Megy
Outreach	Xosé M Fernández, 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