


NATIONAL HUMAN GENOME RESEARCH INSTITUTE Division of Intramural Research





***Current Topics in Genome Analysis 2012***

***Week 3: Genome Browsers***

***Tyra Wolfsberg, Ph.D.***

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES | NATIONAL INSTITUTES OF HEALTH | genome.gov/DIR




***Current Topics in Genome Analysis 2012***

***Tyra Wolfsberg, Ph.D.***

***No Relevant Financial Relationships with  
Commercial Interests***

NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
Division of Intramural Research



## Accessing the public genome sequence data

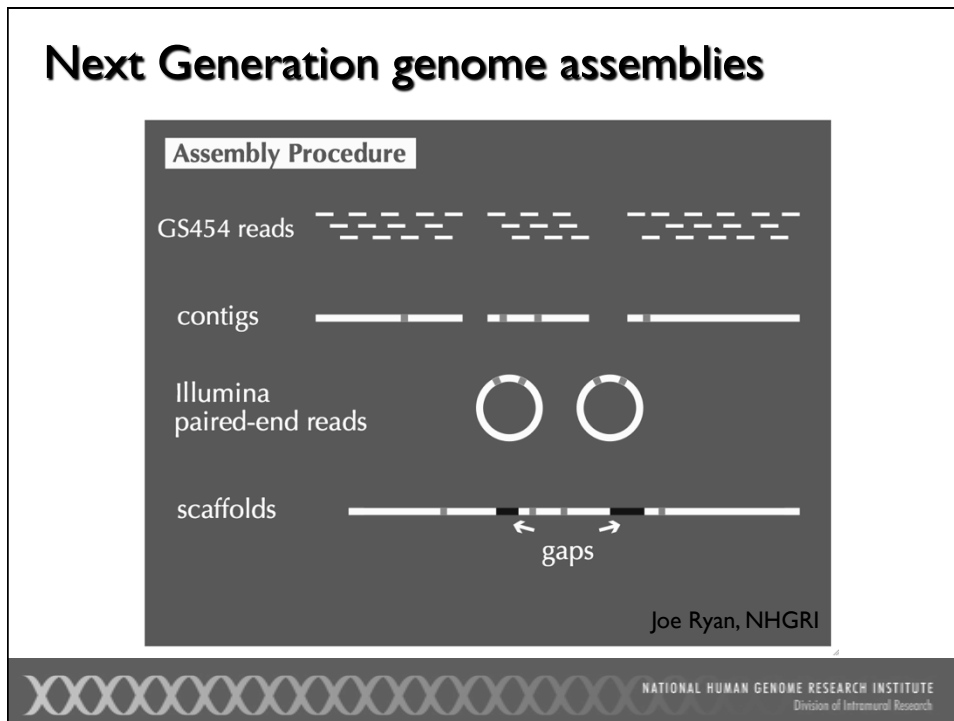
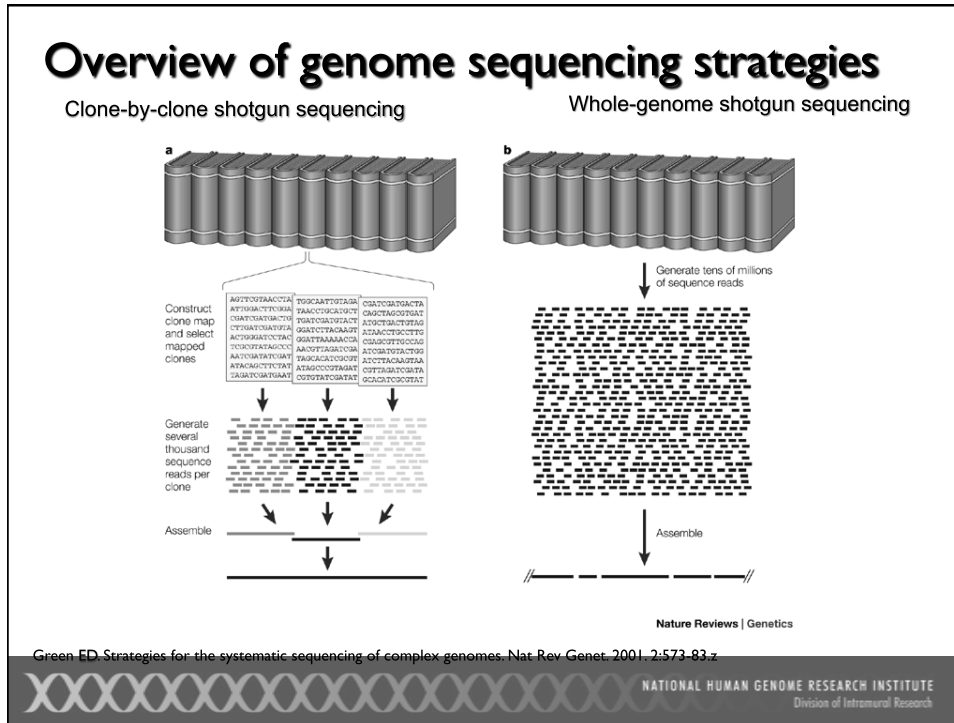
- UCSC's Genome Browser ("Golden Path")  
<http://genome.ucsc.edu>
- Ensembl  
<http://www.ensembl.org>
- NCBI's Map Viewer  
<http://www.ncbi.nlm.nih.gov/mapview/>



## Types of data integrated in genome browsers

- Same starting material for all genome browsers:  
genomic sequence
- Annotations calculated independently by each genome browser
  - Genes
    - RefSeq mRNAs (non-redundant)
    - GenBank mRNAs (redundant)
    - ESTs
    - Gene predictions
  - SNPs
  - Non-coding functional elements





## Genome Sequence Assemblies

- Complex algorithms needed to incorporate all sequence data
- Assemblies updated periodically as new sequence becomes available
  - Mouse, human, and zebrafish (future) genomes assembled by the Genome Reference Consortium (GRC)
  - Other genomes assembled by sequencing centers or consortia
- Assemblies not updated concurrently by the three Genome Browsers
  - “Pre-release” assemblies and annotations available at
    - UCSC: <http://genome-preview.cse.ucsc.edu/>
    - pre!Ensembl: <http://pre.ensembl.org/>
  - UCSC and Ensembl provide archive of all genome assemblies and annotations; NCBI provides only limited archive
- IF YOU ARE COMPARING DATA FROM DIFFERENT GENOME BROWSERS, MAKE SURE YOU ARE LOOKING AT THE SAME VERSION OF THE ASSEMBLY



## Genome Assembly Versions

	Same assembly?	UCSC	NCBI	Ensembl
Human	Yes	Feb 2009/GRCh37/hg19	Build 37.3	GRCh37
Mouse	Yes	July 2007 (NCBI37/mm9)	Build 37.2	NCBIM37
Dog	NO	May 2005 (Broad/canFam2) canFam3 at <a href="http://genome-preview.cse.ucsc.edu">genome-preview.cse.ucsc.edu</a>	Build 3.1/CanFam 3.1	BROAD2; CanFam 2.0
Zebrafish	Yes	Jul. 2010 (Zv9/danRer7)	Zv9	Zv9



## NCBI Reference Sequences (RefSeqs)

- Non-redundant collection of richly annotated DNA, RNA, and protein sequences from diverse taxa
- Each RefSeq represents a single, naturally occurring molecule from one organism

	derived from GenBank submissions	model reference sequences produced by NCBI's Genome Annotation project
mRNA	NM_123456	XM_123456
protein	NP_123456	XP_123456
non-coding transcripts	NR_123456	XR_123456

<http://www.ncbi.nlm.nih.gov/RefSeq/key.html>



<p>LOCUS NM_001101 1852 bp mRNA linear PRI 27-DEC-2009                  DEFINITION Homo sapiens actin, beta (ACTB), mRNA.                  ACCESSION NM_001101                  VERSION NM_001101.3 GI:168480144                  KEYWORDS                  SOURCE Homo sapiens (human)                  ORGANISM Homo sapiens                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.                  REFERENCE 1 (bases 1 to 1852)                  AUTHORS Yamaguchi,H., Shiraishi,M., Fukami,K., Tanabe,A., Ikeda-Matsuo,Y., Naito,Y. and Sasaki,Y.                  TITLE MARCKS regulates lamellipodia formation induced by IGF-I via association with PIP2 and beta-actin at membrane microdomains                  JOURNAL J. Cell. Physiol. 220 (3), 748-755 (2009)                  PUBMED 19479567</p> <p>COMMENT REVIEWED REFSEQ: This record has been curated by NCBI staff. The reference sequence was derived from AK130157.1 and BC009636.1. On Feb 22, 2008 this sequence version replaced gi:5016088.</p> <p>Summary: This gene encodes one of six different actin proteins. Actins are highly conserved proteins that are involved in cell motility, structure, and integrity. This actin is a major constituent of the contractile apparatus and one of the two nonmuscle cytoskeletal actins. [provided by RefSeq].</p> <p>Publication Note: This RefSeq record includes a subset of the publications that are available for this gene. Please see the Entrez Gene record to access additional publications.                  COMPLETENESS: complete on the 3' end.</p> <p>CDS                  85..1212                  /gene="ACTB"                  /gene_synonym="PS1TP5BP1"                  /note="beta cytoskeletal actin; PS1TP5-binding protein 1; actin, cytoplasmic 1"                  /codon_start=1                  /product="beta actin"                  /protein_id="NP_001092.1"                  /db_xref="GI:4501885"                  /db_xref="CCDS:CCDS5341.1"                  /db_xref="GeneID:60"                  /db_xref="HGNC:132"                  /db_xref="HPRD:00032"                  /db_xref="MIM:102630"                  /translation="MDDIARLVVDNGSGMCKAGFAGDDAPRAVFPPIVCRPRHQGVV                  VGMGQKDSVVGDEAQSCKGILTLKYPIDHGLVTDNDMEKIWHHTFYNELRVAPPEHP                  VLLTEAPLNPKANREKMTQIMFETFTFAMYVAIQVLSLYASGRFTGIVMDSGQGVV                  HTVPIYEGVALPHAILRLDLAGRDLDLMLKILTERGYSFTTAEIRIVRDKELICY                  VALDFQEMATAAGSSLEKYEYLPQOVVIFGNFRFCPEALFPQFVGHSGCGIHE                  TTFNSIMKCDVDIRKDLVANTVLSGGTTPYPCIADRMQKEITLALPSTMKIKIIPPE                  RYKSVWIGGSILASLSTFQQMWLSKQEVDESGPSIVHRKCF"</p> <p>ORIGIN                  1 accgcgcgaga ccgcgtccgc ccgcgcgagca cagagcctcg cctttgcgca tcgcgcgcgc                  61 gtccaccacc gccgccagct caccatggat gatgatatcg ccgcgcctcg cgtcgacac                  121 gctccgcgca tgtgcaagc cggtctcgcg ggcgcgatg cccccgggc cgtcttcccc</p>	<h3>Beta actin mRNA RefSeq</h3>
--	---------------------------------

UCSC

View a region in the genome by querying with a gene symbol

<http://genome.ucsc.edu>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
Division of Intramural Research

UCSC Genome Browser Home

Back Forward Reload Stop Home <http://genome.ucsc.edu/> Google

## UCSC Genome Bioinformatics

Genomes - Blat - Tables - Gene Sorter - PCR - VisiGene - Proteome - Session - FAQ - Help

**click** about the UCSC Genome Bioinformatics Site

Welcome to the UCSC Genome Browser website. This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides portals to the [ENCODE](#) and [Neandertal](#) projects.

We encourage you to explore these sequences with our tools. The [Genome Browser](#) zooms and scrolls over chromosomes, showing the work of annotators worldwide. The [Gene Sorter](#) shows expression, homology and other information on groups of genes that can be related in many ways. [Blat](#) quickly maps your sequence to the genome. The [Table Browser](#) provides convenient access to the underlying database. [VisiGene](#) lets you browse through a large collection of *in situ* mouse and frog images to examine expression patterns. [Genome Graphs](#) allows you to upload and display genome-wide data sets.

The UCSC Genome Browser is developed and maintained by the Genome Bioinformatics Group, a cross-departmental team within the Center for Biomolecular Science and Engineering (CBSE) at the University of California Santa Cruz (UCSC). If you have feedback or questions concerning the tools or data on this website, feel free to contact us on our [public mailing list](#).

### News

To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the [genome-announce](#) mailing list.

#### 3 January 2012 - Roadmap Epigenomics Now Available through Data Hub at Washington University

We are pleased to announce the release of the Roadmap Epigenomics data on the UCSC Genome Browser through our Data Hub function. The Roadmap Epigenomics Project is part of The NIH Common Fund's Epigenomics Program. It was launched with the goal of producing a public resource of human epigenomic data to catalyze basic biology and disease-oriented research. The Consortium leverages experimental pipelines built around next-generation sequencing technologies to map DNA methylation, histone modifications, chromatin accessibility and small RNA transcripts in stem cells and primary ex vivo tissues selected to represent the normal counterparts of tissues and organ systems frequently involved in human disease. The Consortium expects to deliver a collection of normal epigenomes that will provide a framework or reference for comparison and integration within a broad array of future studies.

Human (Homo sapiens) Genome Browser Gateway

The UCSC Genome Browser was created by the [Genome Bioinformatics Group of UC Santa Cruz](#).  
 Software Copyright (c) The Regents of the University of California. All rights reserved.

clade: Mammal | genome: Human | assembly: Feb. 2009 (GRCh37/hg19) | position or search term: ADAM2 | gene:  | submit

Feb. 2009 (GRCh37/hg19)  
 Mar. 2006 (NCB36/hg18)  
 May 2004 (NCB35/hg17)  
 July 2003 (NCB34/hg16)

[Click here](#) interface settings to their defaults.  
[track search](#) [add custom tracks](#) [track menu](#) [configure tracks and display](#) [clear position](#)


**About the Human Feb. 2009 (GRCh37/hg19) assembly (sequences)**

The February 2009 human reference sequence (GRCh37) was produced by the [Genome Reference Consortium](#). For more information about this assembly, see [GRCh37](#) in the NCBI Assembly database.

**Sample position queries**

A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS marker, a chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples of valid position queries for the human genome. See the [User's Guide](#) for more information.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
chrUn_gi000212	Displays all of the unplaced contig gi000212
20p13	Displays region for band p13 on chr 20
chr3:1-1000000	Displays first million bases of chr 3, counting from p-arm telomere
chr3:1000000+2000	Displays a region of chr3 that spans 2000 bases, starting with position 1000000
RH18061;RH80175 15q11.1-15q13	Displays region between genome landmarks, such as the STS markers RH18061 and RH80175, or chromosome bands 15q11.1 to 15q13, or SNPs



U C S C  
 Homo sapiens  
 (Graphic courtesy of CRSE)

Human ADAM2 - UCSC Genome Browser v261

UCSC Genes

ADAM2 (uc011lck.1) at chr8:39601256-39695779 - ADAM metalloproteinase domain 2 preprotein  
 ADAM2 (uc003xm1.2) at chr8:39601256-39695779 - ADAM metalloproteinase domain 2 preprotein  
 ADAM2 (uc003xxk.2) at chr8:39601256-39695779 - ADAM metalloproteinase domain 2 preprotein  
 ADAM2 (uc003xxj.2) at chr8:39601256-39695779 - ADAM metalloproteinase domain 2 preprotein  
 ADAM28 (uc011laa.1) at chr8:24151580-24212725 - Homo sapiens metalloproteinase disintegrin cysteine-rich protein, transmembrane  
 ADAM28 (uc011kxz.1) at chr8:24151580-24193610 - SubName: Full=cDNA FLJ60418, highly similar to ADAM 28 (BC 3.4.24.-) (Adisin  
 ADAM29 (uc011cki.1) at chr4:175839509-175899330 - ADAM metalloproteinase domain 29 preprotein  
 adam23 (uc010tiv.1) at chr2:207310031-207482677 - Homo sapiens mRNA for MDC3, complete cds.  
 ADAM28 (uc010lta.2) at chr8:24184067-24212725 - ADAM metalloproteinase domain 28 isoform 1  
 ADAM29 (uc010lrr.2) at chr4:175839509-175899330 - ADAM metalloproteinase domain 29 preprotein  
 ADAM28 (uc003xdx.2) at chr8:24151580-24212725 - ADAM metalloproteinase domain 28 isoform 1  
 ADAM28 (uc003xj.2) at chr8:24151580-24193610 - ADAM metalloproteinase domain 28 isoform 3  
 ADAM22 (uc003ujp.1) at chr7:87564071-87811339 - ADAM metalloproteinase domain 22 isoform 4  
 ADAM22 (uc003ujo.2) at chr7:87563702-87826447 - ADAM metalloproteinase domain 22 isoform 3  
 ADAM22 (uc003ujn.2) at chr7:87563702-87826447 - ADAM metalloproteinase domain 22 isoform 1  
 ADAM22 (uc003ujm.2) at chr7:87563702-87826447 - ADAM metalloproteinase domain 22 isoform 2  
 ADAM22 (uc003ujl.1) at chr7:87563702-87811428 - ADAM metalloproteinase domain 22 isoform 5  
 ADAM22 (uc003ujk.1) at chr7:87563702-87811428 - ADAM metalloproteinase domain 22 isoform 4  
 ADAM22 (uc003ujj.1) at chr7:87563702-87762113 - ADAM metalloproteinase domain 22 isoform 5  
 ADAM22 (uc003uji.1) at chr7:87563702-87757991 - ADAM metalloproteinase domain 22 isoform 5  
 ADAM29 (uc003iud.2) at chr4:175839509-175899330 - ADAM metalloproteinase domain 29 preprotein  
 ADAM29 (uc003iuc.2) at chr4:175839509-175899330 - ADAM metalloproteinase domain 29 preprotein  
 ADAM23 (uc002vbg.2) at chr2:207308368-207482677 - ADAM metalloproteinase domain 23 preprotein  
 ADAM20 (uc001xme.2) at chr14:70989079-71001732 - ADAM metalloproteinase domain 20 preprotein  
 ADAM21 (uc001xmd.2) at chr14:70924217-70926622 - ADAM metalloproteinase domain 21 preprotein  
 YWHAB (uc002zxm.2) at chr20:43514344-43537160 - tyrosine 3-monooxygenase/tryptophan  
 YWHAQ (uc002zxx.2) at chr2:9724107-9771106 - tyrosine 3/tryptophan 5 -monooxygenase  
 YWHAQ (uc002zqw.2) at chr2:9724107-9770745 - tyrosine 3/tryptophan 5 -monooxygenase  
 YWHAZ (uc002zfy.2) at chr17:1247836-1303556 - tyrosine 3/tryptophan 5 -monooxygenase  
 YWHAH (uc003alr.2) at chr22:32340479-32353589 - tyrosine 3-monooxygenase/tryptophan  
 YWHAZ (uc011lhf.1) at chr8:101930804-101963560 - tyrosine 3/tryptophan 5 -monooxygenase  
 YWHAZ (uc011lhc.1) at chr8:101930804-101962799 - tyrosine 3/tryptophan 5 -monooxygenase  
 YWHAZ (uc010mnc.2) at chr8:101930804-101965221 - tyrosine 3/tryptophan 5 -monooxygenase  
 YWHAZ (uc003yix.2) at chr8:101930804-101965623 - tyrosine 3/tryptophan 5 -monooxygenase  
 YWHAZ (uc003yiw.2) at chr8:101930804-101964357 - tyrosine 3/tryptophan 5 -monooxygenase  
 YWHAZ (uc003yiv.2) at chr8:101930804-101963560 - tyrosine 3/tryptophan 5 -monooxygenase  
 ADAM21P (uc010ttg.1) at chr14:70712471-70714518 - SubName: Full=ADAM21-like protein;

**RefSeq Genes**

ADAM2 at chr8:39601255-39695779 - (NM\_001464) disintegrin and metalloproteinase domain-containing protein 2 preprotein  
 ADAM20 at chr14:70989078-71001732 - (NM\_003814) disintegrin and metalloproteinase domain-containing protein 20 preprotein  
 ADAM21 at chr14:70918874-70926622 - (NM\_003813) disintegrin and metalloproteinase domain-containing protein 21 preprotein

**click**








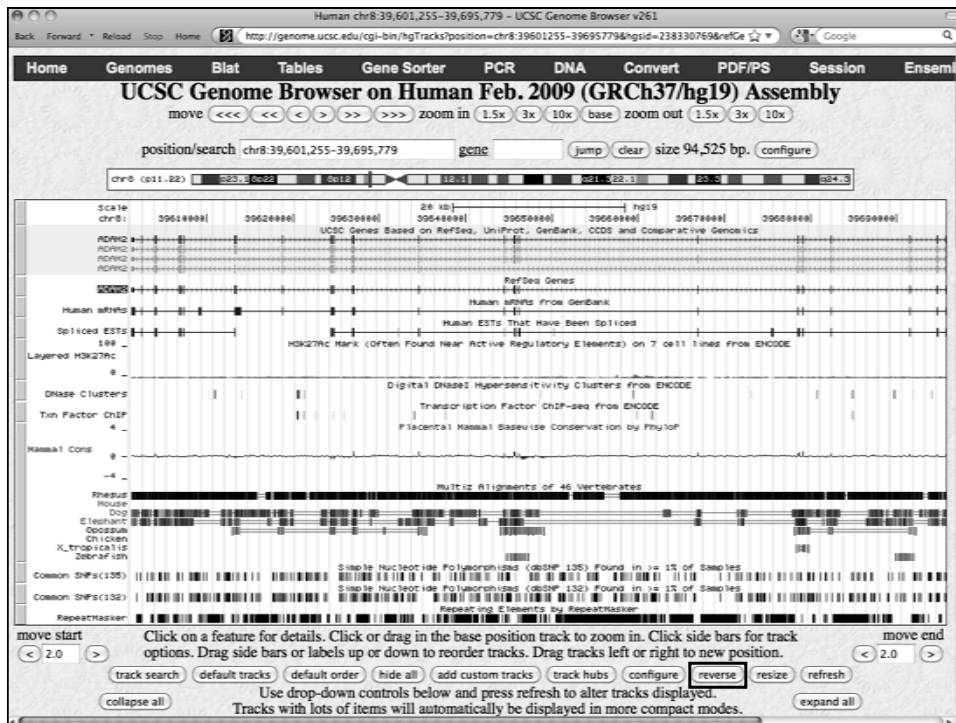
# UCSC

## Navigating around the Genome Browser

<http://genome.ucsc.edu>



NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
Division of Intramural Research



Human chr8:39,601,255-39,695,779 - UCSC Genome Browser v261

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

position/search chr8:39,601,255-39,695,779 gene jump clear size 94,525 bp. configure

chr8 (p11.22) 39,600,000 39,610,000 39,620,000 39,630,000 39,640,000 39,650,000 39,660,000 39,670,000 39,680,000 39,690,000

Scale chr8: 39,610,000 39,620,000 39,630,000 39,640,000 39,650,000 39,660,000 39,670,000 39,680,000 39,690,000 hg19

UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics

RefSeq Genes

Human mRNAs from GenBank

Human mRNAs

Spliced ESTs

Human ESTs that have been spliced

Layered H3K27Ac

H3K27Ac Mark (Often Found Near Active Regulatory Elements) on 7 cell lines from ENCODE

DNase Clusters

Digital DNase Hypersensitivity Clusters from ENCODE

TF ChIP

Transcription Factor ChIP-seq from ENCODE

Repeatmasker

Placental Mammary Basepair Conservation by PhyloP

Mammary Cons

Multi-species Alignments of 46 Vertebrates

Rhesus Mouse Dog Elephant Gorilla Chimpanzee X\_Tropicalis Zebra Finch

Common SNPs (135)

Simple Nucleotide Polymorphisms (SNPs) found in 1% of samples

Common SNPs (132)

Simple Nucleotide Polymorphisms (SNPs) found in 1% of samples

Repeatmasker

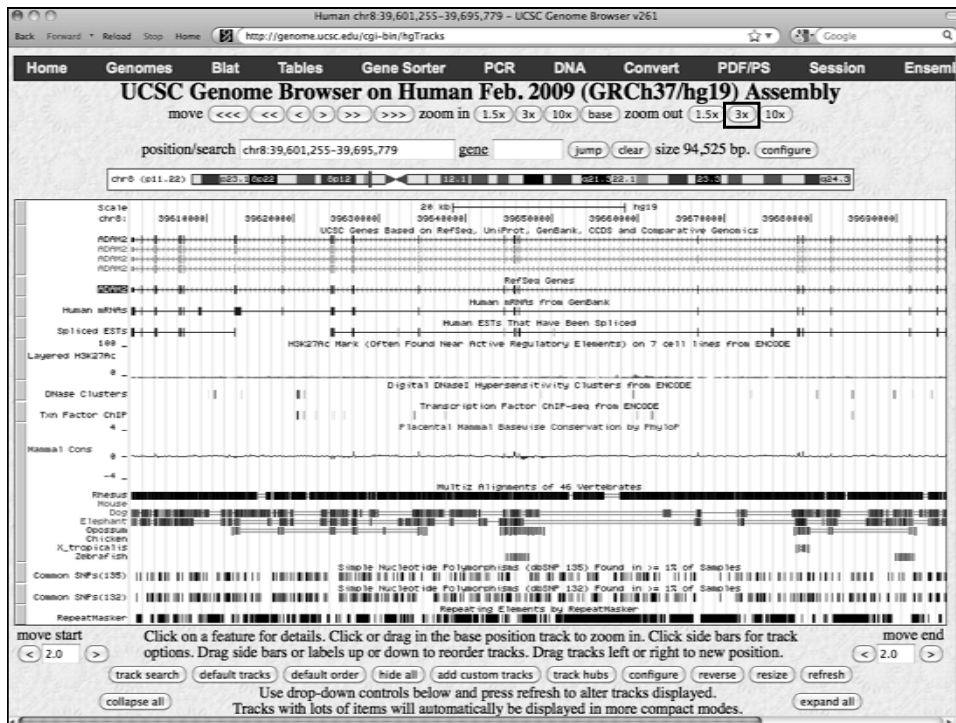
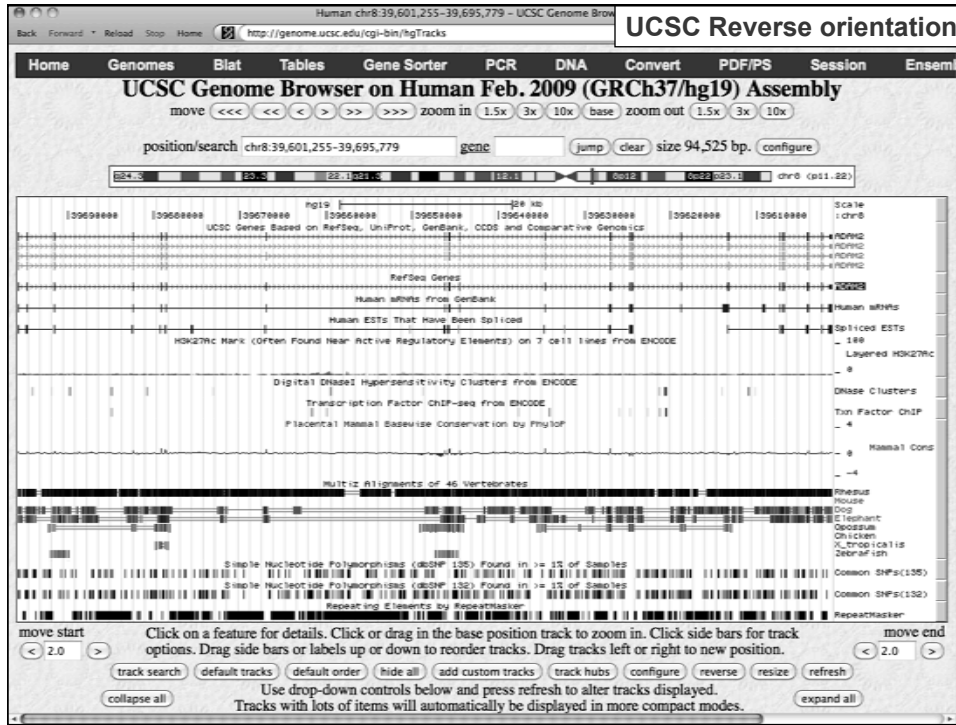
Repeatmasking Elements by Repeatmasker

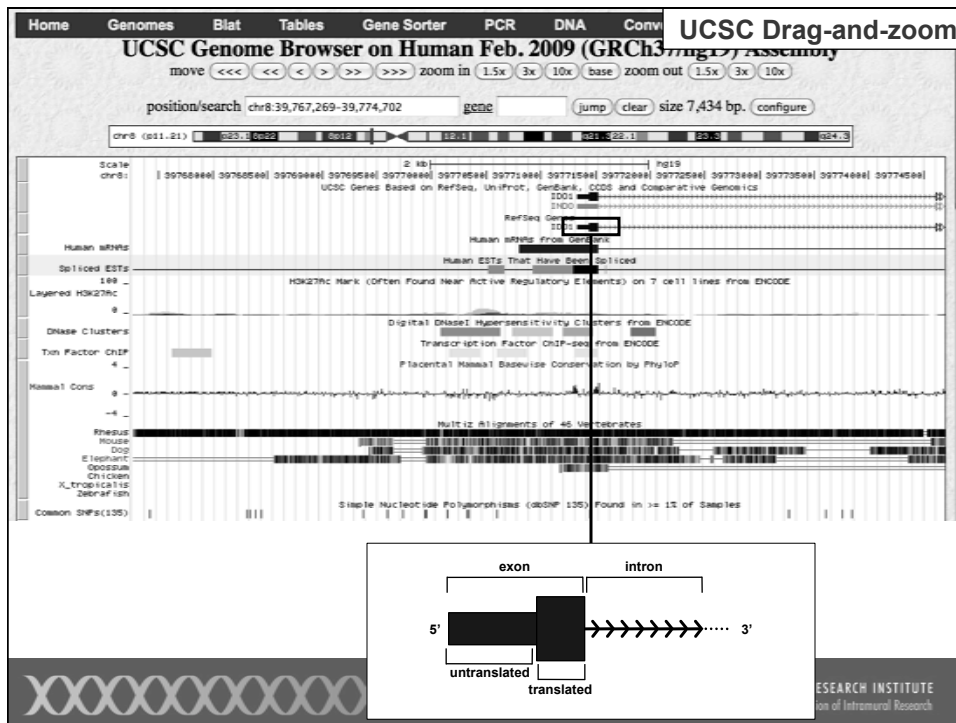
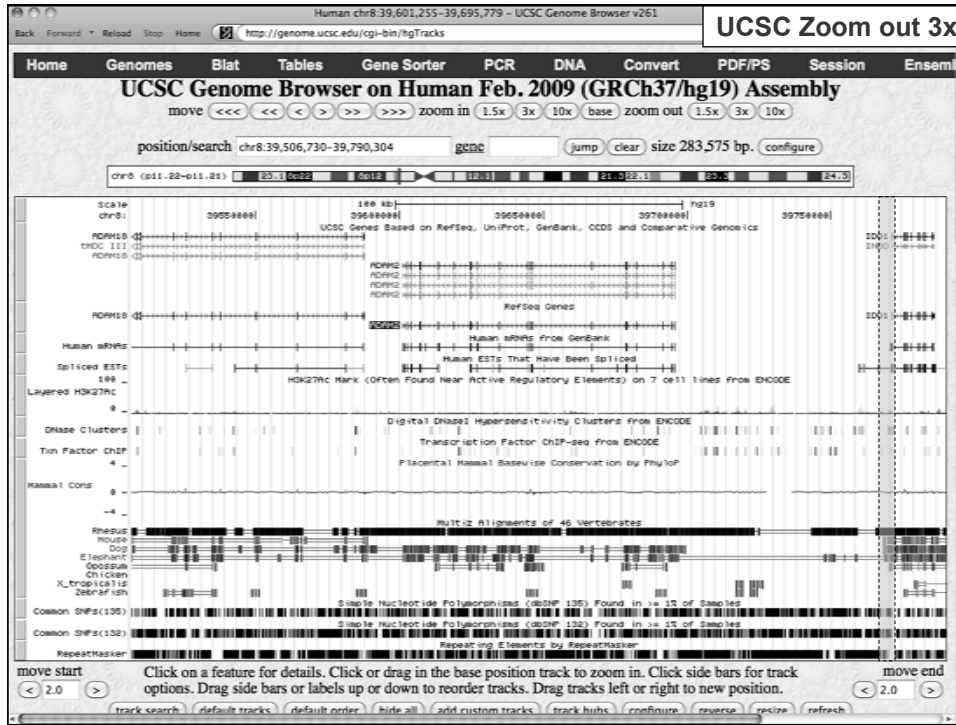
move start < 2.0 > Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position. move end < 2.0 >

track search default tracks default order hide all add custom tracks track hubs configure reverse resize refresh

collapse all Use drop-down controls below and press refresh to alter tracks displayed. expand all

Tracks with lots of items will automatically be displayed in more compact modes.





# UCSC

## Add a track to the Genome Browser

<http://genome.ucsc.edu>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
 Division of Intramural Research

The screenshot shows the UCSC Genome Browser interface with the following tracks visible:

- Affy Exon Array
- Affy GNF1H
- Affy RNA Loc
- Affy U133
- Affy U133Plus2
- Affy U95
- Allen Brain
- Burge RNA-seq
- ENC Exon Array...
- ENC ProtGeno...
- ENC RNA-seq...
- GIS RNA PET
- GNF Atlas 2
- Illumina WG-6
- Sestan Brain

**Regulation**

- ENCODE Regulation...
- ENC RNA Binding...
- TFBS Conserved
- CD34 DnaseI
- ENC TF Binding...
- TS miRNA sites
- CpG Islands
- OREgAnno
- UMMS Brain Hist
- ENC DNA Methy
- ENC
- ENC Histone
- Sta Nucle
- Vista

**Comparative Genom**

- Conservation
- hg19Patch2 Chain/Net
- Cons Indels
- MmCf
- Vertebrate Chain/Net
- GERP
- Evo

**Neanderthal Assembly and Analysis**

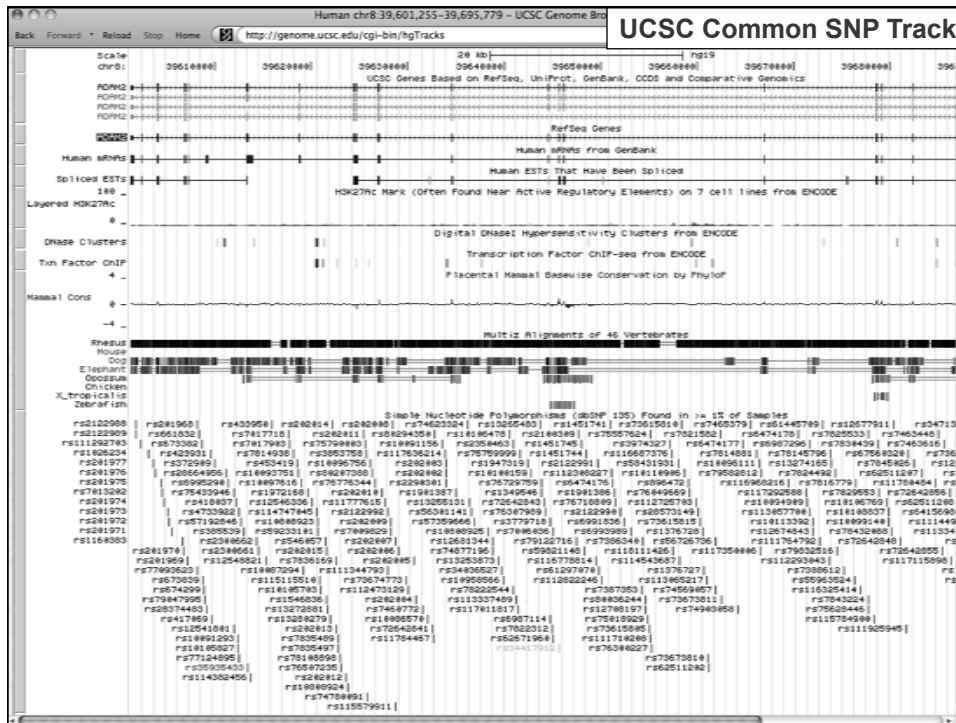
**Variation and Repeats**

- Common SNPs(135)
- Flagged SNPs(135)
- Mult. SNPs(135)
- All SNPs(135)
- Common SNPs(132)
- Flagged SNPs(132)
- All SNPs(132)
- SNPs (131)
- Arrays
- GIS DNA PET
- HAIB Genotype
- HGDP Allele

A callout box on the right side of the screenshot lists the following definitions:

- Common SNPs(135):** SNPs with  $\geq 1\%$  minor allele frequency (MAF), mapping only once to reference assembly.
- Flagged SNPs(135):** SNPs  $< 1\%$  MAF (or unknown), mapping only once to reference assembly, flagged in dbSNP as "clinically associated" -- not necessarily a risk allele.
- Mult. SNPs(135):** SNPs mapping in more than one place on reference assembly.
- All SNPs(135):** all SNPs from dbSNP mapping to reference assembly.

A "click" arrow points to the "refresh" button next to the "Common SNPs(135)" track.



UCSC SNP Track details

Common SNPs(135) Track Settings

Home Genomes Genome Browser Blat Tables Gene Sorter PCR Session FAQ Help

Common SNPs(135) Track Settings

Simple Nucleotide Polymorphisms (dbSNP 135) Found in >= 1% of Samples (▲ All Variation and Repeats tracks)

Display mode:

Include Chimp state and observed human alleles in name:   
 (If enabled, chimp allele is displayed first, then '>', then human alleles.)

Use Gene Tracks for Functional Annotation

Filtering Options

Coloring Options

SNP Feature for Color Specification:

The selected "Feature for Color Specification" above has the selection of colors below for each attribute. Only the color options for the feature selected above will be used to color items; color options for other features will not be shown. If a SNP has more than one of these attributes, the stronger color will override the weaker color. The order of colors, from strongest to weakest, is red, green, blue, gray, and black.

Unknown  Locus  Coding - Synonymous  Coding - Non-Synonymous

Untranslated  Intron  Splice Site

View table schema  
 Data last updated: 2011-11-14

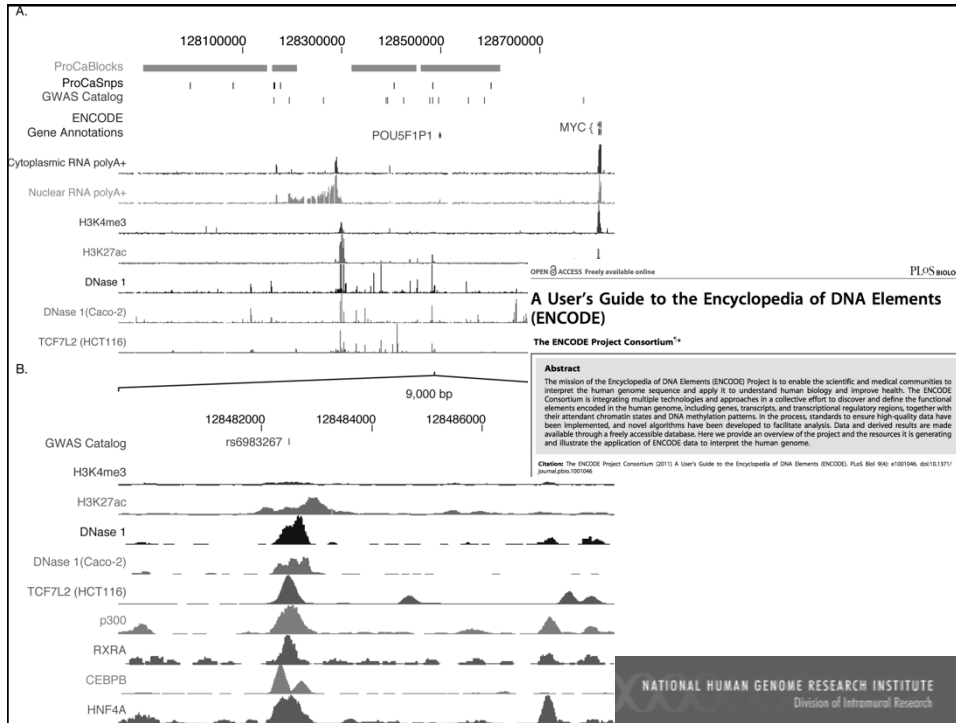
Description

This track contains information about a subset of the single nucleotide polymorphisms and small insertions and deletions (indels) — collectively Simple Nucleotide Polymorphisms — from dbSNP build 135, available from [ftp.ncbi.nih.gov/snp](http://ftp.ncbi.nih.gov/snp). Only SNPs that have a minor allele frequency of at least 1% and are mapped to a single location in the reference genome assembly are included in this subset. Frequency data are not available for all SNPs, so this subset is incomplete.









**UCSC**

Find a chicken homolog of a human protein

<http://genome.ucsc.edu>

**NATIONAL HUMAN GENOME RESEARCH INSTITUTE**  
 Division of Intramural Research

The screenshot shows the NCBI Entrez Protein interface. The browser address bar displays the URL: <http://www.ncbi.nlm.nih.gov/protein/55743080?report=fasta>. The page title is "disintegrin and metalloproteinase domain-containing protein 2 preproprotein [Homo sapiens]". The NCBI Reference Sequence is NP\_001455.3. The protein sequence is displayed in FASTA format, starting with >gi|55743080|ref|NP\_001455.3| disintegrin and metalloproteinase domain-containing protein 2 preproprotein [Homo sapiens]. The sequence is: MNRVLLSGLGQLRMDNFDSLPTQITVPEKIRSIIEKGISSQASYKIVIEGKPTTVNLMQKNLPHNF RYVYSOTGIMKPLDQDPONFCHYQYIEGYPKSYVMVSTCTGLAGVLQFEMVSYGIEPLESSVGFPHV YQVKKKADVSLYNEKDISSRDLSPKLSQVPEQDFAKYIEHVIIEKQLYHNGSDTTVAQKVFQLIG LTNALFVSNITIISSLELWIDENKIATTEANELLHTFLRWKTSYLVLRPHDVAFLVYREKSNYVGA TFQGMCDANYAGGVVLRHPTISLES LAVILAQLLSLWGIITDINKQCSGAVCINMPEAIHFGVGI FSNCSFEDFAHFISKQSQCLNQPRLDFFKQAVCGNAKLEAGEECDCGTEQDCALIGETCCDIATCR FKAGSNACBPCCENCLFMSKERKCPSEFCLEPFCNGSSASCENHVVQVPGQLGQWICIDGVCMSGDWQCTDTDFGKEVEFGRESECTSELNSKTDVSGWCGISDSGYTQCADNLQCKLICKYVGRFLQIPRA TIIYANISGHLCAVEFASDHADSQKMKIKDGTSCGSKVCRNQRQCVSSSYLGYDCTTDCNDRQVCMNK KHCRCASLYLPDCVQSDLWPGGSDSGNFPFVAIPARLPERRYIENIYBSKPMRWFPLIFPFIIFC VLIALMVKVNFQRKKWRTEYSSDQPESESEPKG. On the right side, there are several sections: "Analyze this sequence" with options like Run BLAST, Identify Conserved Domains, Highlight Sequence Features, and Find in this Sequence; "Articles about the ADAM2 gene" with links to "Mapping, sequence, and expression analysis of the human fert [Genomics. 1997]" and "Role of the integrin-associated protein CD9 in binding t [Proc Natl Acad Sci U S A. 1999]"; "Identical proteins for NP\_001455.3" listing "unnamed protein product [Horn][CBH30599]", "ADAM metalloproteinase domain [EAW63273]", and "RecName: Full=Disintegrin and [Q69965]"; and "Reference sequence information" with a link to "RefSeq mRNA" and a note: "See reference mRNA sequence for the ADAM2 gene (NM\_001464.3)."

The screenshot shows the UCSC BLAT search interface. The browser address bar displays the URL: <http://genome-preview.cse.ucsc.edu/cgi-bin/hgBlat>. The page title is "Chicken BLAT Search". The main heading is "BLAT Search Genome". Below the heading, there are input fields for "Genome:" (set to "Chicken"), "Assembly:" (set to "Feb. 2004 (MUCSC 1.0/galGal2)"), "Query type:" (set to "BLAT's guess"), "Sort output:" (set to "query.score"), and "Output type:" (set to "hyperlink"). A text area contains the same protein sequence as in the first screenshot. Below the text area are buttons for "submit", "I'm feeling lucky", and "clear". Below the buttons, there is a paragraph of instructions: "Paste in a query sequence to find its location in the the genome. Multiple sequences may be searched if separated by lines starting with > followed by the sequence name." Below this is a section for "File Upload:" with a "Browse..." button and a "submit file" button. At the bottom, there is a paragraph: "Only DNA sequences of 25,000 or fewer bases and protein or translated sequence of 10000 or fewer letters will be processed. Up to 25 sequences can be submitted at the same time. The total limit for multiple sequence submissions is 50,000 bases or 25,000 letters." and a note: "For locating PCR primers, use In-Silico PCR for best results instead of BLAT."

Chicken BLAT Results

UCSC BLAT search

Home Genomes Tables PCR Session FAQ Help

### BLAT Search Results

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
<a href="#">browser</a>	NP_001455.3	44	539	600	735	71.6%	Un	++	635370	635555	186
<a href="#">details</a>	NP_001455.3	12	301	304	735	100.0%	1	++	67659709	67659720	12
<a href="#">browser</a>	NP_001455.3	12	437	440	735	100.0%	1	++	67660117	67660128	12
<a href="#">details</a>	NP_001455.3	12	385	390	735	83.4%	1	++	67659961	67659978	18

UCSC Preview Genome Browser on Chicken Feb. 2004 (WUGSC 1.0/galGal2) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x (base) zoom out 1.5x 3x 10x preview

position/search chrUn:635,370-635,555 gene  jump clear size 186 bp. configure

Chicken BLAT Results

UCSC BLAT search

Home Genomes Tables PCR Session FAQ Help

### BLAT Search Results

#### Alignment of NP\_001455.3 and chrUn:635370-635555

Click on links in the frame to the left to navigate through the alignment. Matching bases are colored blue and capitalized. Light blue bases mark the boundaries of gaps in either sequence.

NP\_001455.3

```

mvrvfllsg lgglrmdanf dslpvqitvp ekirsiikeg iesqasykiv iegkpytvnl 60
mqknflphnf rvyysygtgi mkpldqdfgn fohygyieg ypksvmvvat ctglrgvlqf 120
enbvyliepl esvvgfshvi ygvkhkadv slynekies rdlsfklsqv epqgdifakyi 180
emhiviekql ynhmsdttv vagkvfqlig ltnaifvsfn itilslsel widenkiatt 240
geanelhft lrwktaylvi rphdvafllv yreksnyvga tfggkmedan yagvvlhpr 300
tisleslavi laqlslsmg ityddinkq csgavcimp eaihfgvki fscnscfedfa 360
hfiskqkscq lhnqprldpf fkqgavcna kleageecd gteqdcailg etccdiatr 420
fkagsncaeg pccencifms kermcrpsfe ecdlpeycng ssascpenhy vqtghpcgln 480
qvicidvcvm sqdqeetft gkevefsgse cyhlnsktd vsnqcgidsa gytqceadL 540
qCGKLiCKv gkflqipra IIYAnisgH L'lavefasd hadsqkmiX DGTsCGenKv 600
crngrcvsss ylygdcttdk cndrgvcnk khocsaayl pdcsvqsd1 wpggsidsqn 660
fppvaiparl perryieniy hskpmrwpff lfipffiiic vliaimvkn fgrkkwrted 720
yssdeqese sepgk
    
```

Chicken.chrUn :

```

:ATCTGggcT GTGGAAACT CATCTGcaca TA:ccaaaac gagttccctt caccaaat 635429
aaggg:CCA TCATCTATGC Tcaagtcaa gaacATCTGT C:gtgtett t gatgtaatg 635489
catgcacct ceggacaga toctctctg gtt:AGGATG GCACGaaATG CGGTccgga 635549
AAGGT:
    
```

Side by Side Alignment\*


```

001615 N L Q C G K L I C K Y 001647
>>>>> | | G | | | | | T | >>>>>
635370 aatctgggctgtgaaactcatctgcacatac 635402
    
```

# UCSC

Add your own custom tracks

<http://genome.ucsc.edu>




NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
Division of Intramural Research

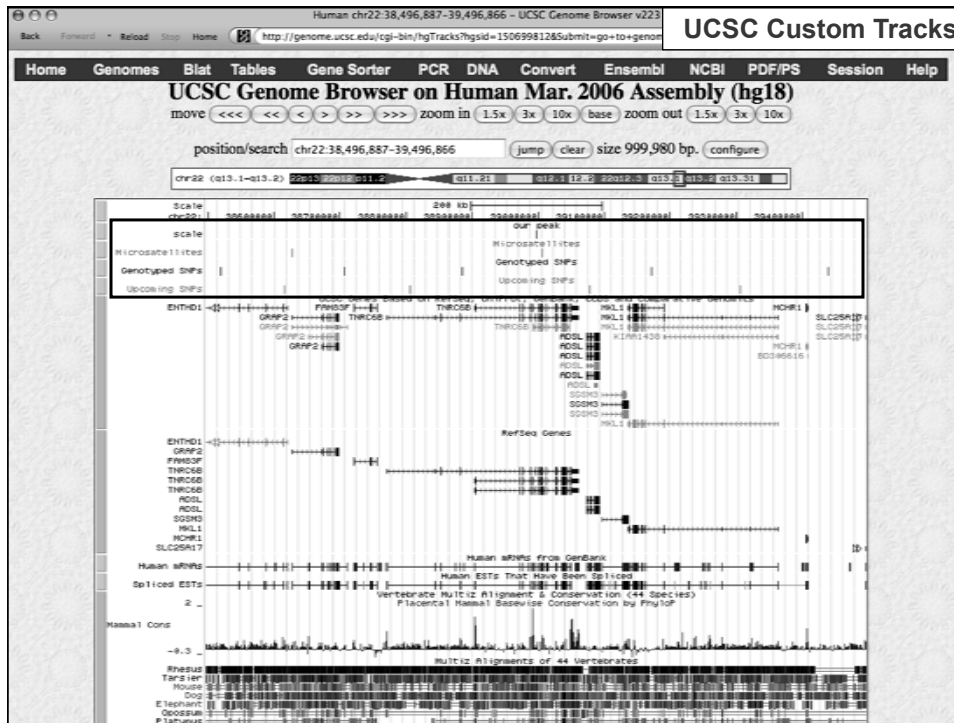
## UCSC custom track format

```
browser position chr22:38496887-39496866
browser hide cytoBand
browser hide stsMap
browser hide gap
browser hide clonePos
browser full refGene
browser dense mrna
track name="scale" description="our peak"
chr22 38996887 38996888 peak
track name="Microsatellites" description="Microsatellites" color=0,128,0
chr22 38627059 38627060 D22S276
chr22 39005417 39005418 D22S307
track name="Genotyped SNPs" description="Genotyped SNPs" color=0,0,255
chr22 38518342 38518343 ss146131
chr22 38705963 38705964 ss2941443
chr22 38884157 38884158 ss141110
chr22 39171390 39171391 ss22916
chr22 39438769 39438770 ss1479794
track name="Upcoming SNPs" description="Upcoming SNPs" color=0,128,192
chr22 38615712 38615713 ss86855
chr22 38804838 38804839 ss85533
chr22 39077895 39077896 ss141190
chr22 39305065 39305066 ss137027
```

[http://research.nhgri.nih.gov/teaching/custom\\_tracks.shtml](http://research.nhgri.nih.gov/teaching/custom_tracks.shtml)



NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
Division of Intramural Research

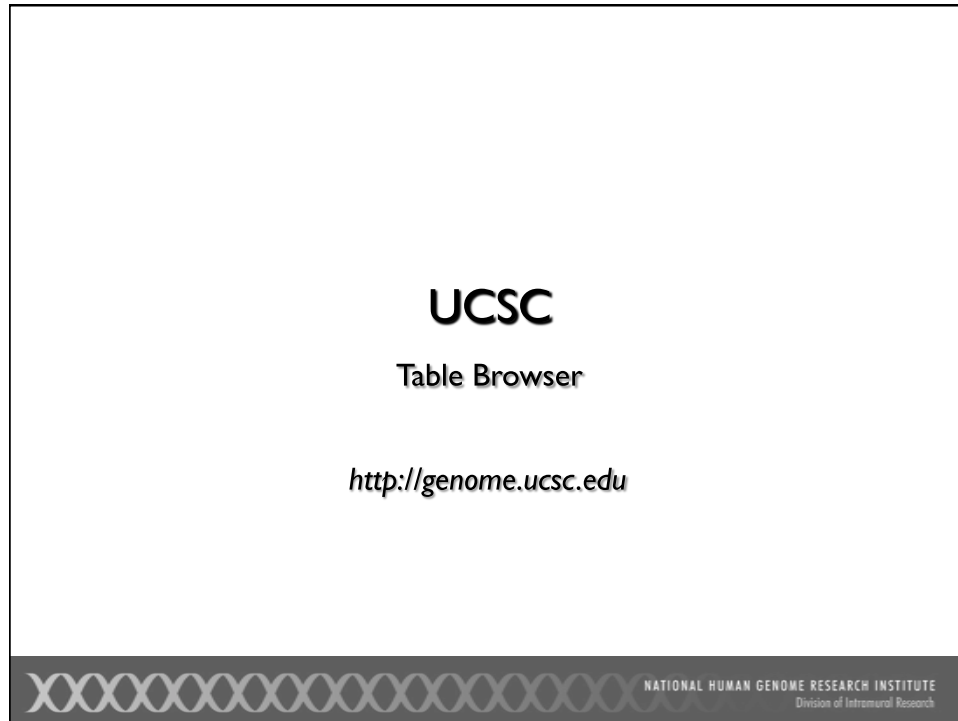


## Types of UCSC custom tracks

- Upload annotation data from your computer
  - Only viewable on the machine from which they were uploaded
  - Discarded after 48 hours
- Post annotation data to your Web site
  - URL, with link to Genome Browser, can be shared with anyone
  - Never discarded
- Create a Session to configure your browser with specific track combinations, including custom tracks
  - Can be shared or non-shared
  - Session persists for 4 months; custom tracks for 48 hours
- Contribute your tracks to the UCSC Genome Browser

<http://genome.ucsc.edu/goldenPath/help/hgTracksHelp.htm#CustomTracks>

<http://genome.ucsc.edu/goldenPath/help/hgSessionHelp.html>



UCSC  
Table Browser  
<http://genome.ucsc.edu>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
Division of Intramural Research

## UCSC Table Browser

- Download track in text format
- Retrieve DNA sequence covered by a track
  - Get sequence 1 Kb upstream of each RefSeq gene
- Calculate intersections between tracks and view in the Genome Browser.
  - List all SNPs in a gene
- Filter track data based on certain criteria
  - Show all RefSeq genes that contain only one exon



NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
Division of Intramural Research

**clade:** Mammal | **genome:** Human | **assembly:** Feb. 2009 (GRCh37/hg19) |  
**group:** Genes and Gene Prediction Tracks | **track:** RefSeq Genes | add custom tracks | track hubs  
**table:** refGene | describe table schema  
**region:** genome | position: chr21:33031597-33041570 | lookup | define regions  
**identifiers (names/accessions):** paste list | upload list  
**filter:** create  
**intersection:** create  
**correlation:** create  
**output format:** sequence | end output to: Galaxy | GREAT  
**output file:** (leave blank to keep output in browser)  
**file type returned:** plain text | gzip compressed  
 get output | summary/statistics

**UCSC Table Browser:**  
 200 nt upstream of each RefSeq gene

Select sequence type for RefSeq Genes

genomic  
 protein  
 mRNA  
 submit | cancel

**Sequence Retrieval Region Options:**

Promoter/Upstream by 200 bases  
 5' UTR Exons  
 CDS Exons  
 3' UTR Exons  
 Introns  
 Downstream by 1000 bases  
 One FASTA record per gene.  
 One FASTA record per region (exon, intron, etc.)  
 Split UTR and CDS parts of an exon into sep

Note: if a feature is close to the beginning or end of past the edge of the chromosome.

**Sequence Formatting Options:**

Exons in upper case, everything else in lower ca  
 CDS in  
 All up  
 All low  
 Mask

get sequen

```

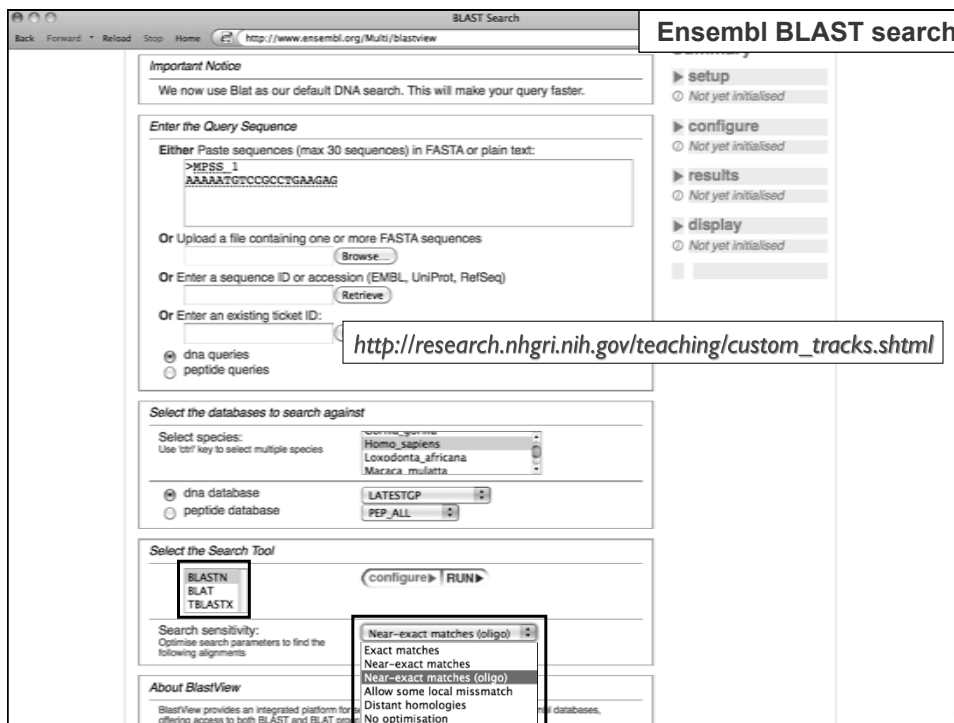
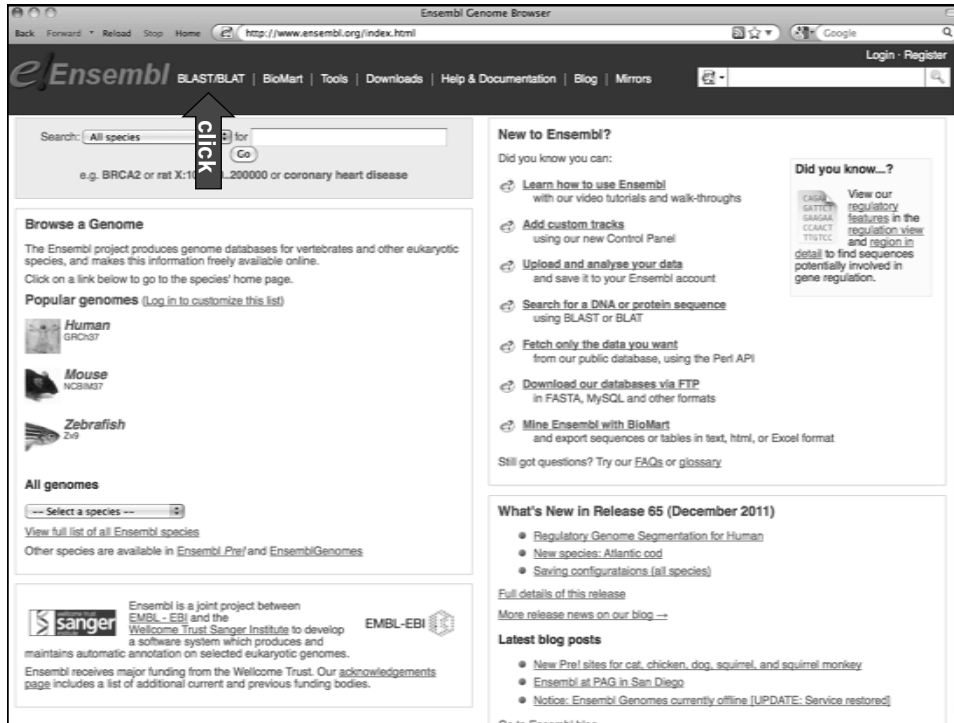
>hg19_refGene_NM_032291 range=chr1:66999625-66999824 5'pad=0 3'pad=0 strand=+ repeatMasking=none
ggagggtgtgtatttgggtacaagcggaggcgtgggggtgaagg
>hg19_refGene_NM_001145278 range=chr11:16766967-16767166 5'pad=0 3'pad=0 strand=+ repeatMasking=none
aaatggggacgggaataggtctgtgtctctccgggtattgtgtea
ggagatgcaaggctggctacctgtgacgcggtccaagctgaaggattg
ggcagagcagcagcagcgggtgagctcggcagcttgcctctctccc
>hg19_refGene_NM_001145277 range=chr11:16766967-16767166 5'pad=0 3'pad=0 strand=+ repeatMasking=none
agaccacgggttacagaggggtctgtccatggcgggcgagggcgtttc
tttgctcggaggggtgtctggaggaaggagaagcctctggaggaagga
gaagcctcggaggtgcgccgcacgtgtctgagccgggtttcagcag
aggggcacaaagaggggtgctgagcccgagctgcgcttagcc
>hg19_refGene_NM_001145277 range=chr11:16766967-16767166 5'pad=0 3'pad=0 strand=+ repeatMasking=none
agaccacgggttacagaggggtctgtccatggcgggcgagggcgtttc
tttgctcggaggggtgtctggaggaaggagaagcctctggaggaagga
                
```

## Ensembl

Identify genes that overlap with an oligo tag

<http://www.ensembl.org>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
 Division of Intramural Research





The screenshot shows the Ensembl BLAST search results page. The main content area is titled "Ensembl BLAST search" and contains three sections: "Alignment Locations vs. Karyotype", "Alignment Locations vs. Query", and "Alignment Summary".

The "Alignment Locations vs. Karyotype" section displays a karyotype with a red box highlighting a specific region. A callout box above it states: "100% identity over 100% of the query length".

The "Alignment Locations vs. Query" section shows a coverage plot and HSPs. A callout box above it displays: "Query Start 1 End 20 % ID 100.00".


The "Alignment Summary" section contains a table with columns for Subject, Chromosome, Supercontig, Clone, Contig, Lng, Stats, and Sort By. A callout box with the word "CLICK" points to the "Subject" column.

The table data is as follows:

Subject	Chromosome	Supercontig	Clone	Contig	Lng	Stats	Sort By
Name	Name	Name	Name	Name	Name	Score	>ID
Start	Start	Start	Start	Start	Start	E-val	>Score
1	20	Chr15	57210876	57210895	20	0.000	100.00
1	17	Chr17	79642589	79642575	17	2.2	100.00

At the bottom of the page, it says "Ensembl release 65 - Dec 2011 © WTSI / EBI" and "About Ensembl | Contact Us | Help".

**Ensembl**  
 Location tab  
<http://www.ensembl.org>


 NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
 Division of Intramural Research

**Ensembl Location tab: Region in detail**

click

translated exon  
untranslated exon

Genes

Gene Legend:
 

- CCDS set
- protein coding
- merged Ensembl/Havana
- processed transcript
- noncoding
- RNA gene

**Ensembl Location tab: Configure page**

Ensembl BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors

Human Configure Region Image Configure Overview Image Manage Configurations Custom Data

Location: Active tracks, Favourite tracks, Track order, Search results

Germline variation
 

- Enable/disable all dbSNP
- Sequence variants (dbSNP and all other sources)
- cbSNP variants

Key:
 

- Track style
- Forward strand
- Reverse strand
- Favourite track
- Track information

External tracks:
 

- DAS Distributed Annotation Source
- Custom track - uploaded data
- Custom track - UCSC-style web resource
- Custom data saved to your user account

Please note that the content of external tracks is not the responsibility of the Ensembl project. URL-based or DAS tracks may either slow down your ensembl browsing experience OR may be unavailable as these are served and stored from other servers elsewhere on the Internet.



Ensembl genome browser 65: Homo sapiens - Explore this variation - rs35615435

Ensembl Variation tab: Summary

Human (GRCh37) Location: 15:57,212,896-57,216,915 Variation: rs35615435

**rs35615435 SNP**

Source: dbSNP\_134 - Variants (including SNPs and indels) imported from dbSNP  
 Alleles: Reference/Alternative: A/G | Ancestral: A | Ambiguity code: R | MAF: 0.19 (G)  
 Location: Chromosome 15:57213283 (forward strand) | View in location tab  
 Validation status: This variation is validated by 1000 Genomes and also cluster, frequency  
 Synonyms: dbSNP rs59892738  
 HGVS names: This feature has 12 HGVS names - click the plus to show

Explore this variation help

Genomic context, Gene / Transcript, Population genetics, Individual genotypes, Linkage disequilibrium, Phenotype data, Phylogenetic context, Flanking sequence

Help with variations

YouTube videos: SNPs and other Variations - 1 of 2, SNPs and other Variations - 2 of 2, Clip: Genome Variation, BioMart: Variation IDs to HGNC Symbols

Reference materials: Ensembl variation data: background and terminology, Variation Quick Reference card

Additional resources: Accessing variation data with the Variation API, Genomes and SNPs in Malaria

Ensembl release 65 - Dec 2011 © WTSI / EBI

Ensembl Variation tab: Genomic context Population genetics

Context help

Ensembl/Havana... 5.00 Kb

Contigs

Sequence variant...

Population genetics help

1000 genomes alleles frequencies

CHB+JPT: A: 70%, G: 30%  
 YRI: A: 96%, G: 4%

1000 genomes (2)

Population	rsID	Submitter	Alleles A	Alleles G	Allele count	Genotype detail
1000GENOMES:pilot_1_CHB+JPT_low_coverage_panel	ss243177144	1000GENOMES	0.700	0.300	84 (A) / 36 (G)	Show
1000GENOMES:pilot_1_YRI_low_coverage_panel	ss226911693	1000GENOMES	0.958	0.042	113 (A) / 5 (G)	Show

Ensembl  
Gene tab  
<http://www.ensembl.org>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
Division of Intramural Research

Ensembl genome browser 65: Homo sapiens -- Region in detail -- Chromosome 15: 57,213,000 - 57,215,000

Location: 15:57212896-57216915 Go Gene: Go

Chromosome bands: Human RefSeq/CCDS set

Gene	Transcript	Protein	Location	Gene type
TCF12-202	ENST00000543236	ENSP00000442962	Chromosome 15: 57,213,322-57,279,190	Known protein coding
TCF12-025	ENST00000140262	ENSP00000442910		Known protein coding
TCF12-001				Known protein coding
TCF12-201				Known protein coding
TCF12-024				Known protein coding
TCF12-003				Known protein coding

click

Ensembl genome browser 65: Homo sapiens - Gene summary - Gene: TCF12 (ENSG00000140262)

Ensembl Gene tab: Gene summary

Human (GRCh37) Location: 15:57,212,896-57,216,915 Gene: TCF12 Transcript: TCF12-202

**Gene: TCF12 ENSG00000140262**

Description transcription factor 12 [Source:HGNC Symbol;Acc:11623]  
 Location Chromosome 15: 57,210,323-57,591,479 forward strand.  
 Transcripts This gene has 31 transcripts

Show All entries Show/hide columns Filter

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
TCF12-001	ENST00000267811	6061	ENSP00000267811	682	Protein coding	CCDS10158
TCF12-002	ENST00000333725	4719	ENSP00000331067	706	Protein coding	CCDS10160
TCF12-004	ENST00000543579	1809	ENSP00000440017	536	Protein coding	-
TCF12-006	ENST00000537840	1598	ENSP00000444696	446	Protein coding	-
TCF12-007	ENST00000559922	1598	ENSP00000453216	446	Protein coding	-
TCF12-008	ENST00000343827	3956	ENSP00000342459	512	Protein coding	CCDS42049
TCF12-009	ENST00000559710	1314	ENSP00000453264	316	Protein coding	-
TCF12-010	ENST00000559703	1544	ENSP00000454102	339	Protein coding	-
TCF12-025	ENST00000438423	4786	ENSP00000389940	706	Protein coding	CCDS10160
TCF12-026	ENST00000557843	4076	ENSP00000453737	682	Protein coding	CCDS10158
TCF12-027	ENST00000557947	575	ENSP00000454109	157	Protein coding	-
TCF12-028	ENST00000561152	675	ENSP00000453653	58	Protein coding	-
TCF12-201	ENST00000452095	4772	ENSP00000396881	702	Protein coding	-
TCF12-202	ENST00000543296	2474	ENSP00000442910	681	Protein coding	-
TCF12-203	ENST00000543417	1545	ENSP00000443452	294	Protein coding	-
TCF12-003	ENST00000559609	2252	ENSP00000453876	666	Nonsense mediated decay	-
TCF12-011	ENST00000561449	914	No protein product	-	Processed transcript	-
TCF12-013	ENST00000559216	475	No protein product	-	Processed transcript	-
TCF12-014	ENST00000560887	540	No protein product	-	Processed transcript	-
TCF12-016	ENST00000561346	852	No protein product	-	Processed transcript	-
TCF12-017	ENST00000561235	818	No protein product	-	Processed transcript	-
TCF12-018	ENST00000560784	1230	No protein product	-	Processed transcript	-
TCF12-020	ENST00000560506	575	No protein product	-	Processed transcript	-

click

Ensembl genome browser 65: Homo sapiens - Orthologues - Gene: TCF12 (ENSG00000140262)

Ensembl Gene tab: Orthologues

Selected orthologues

View sequence alignments of these homologues.

Show All entries Show/hide columns Filter

Species	Type	dN/dS	Ensembl identifier & gene name	Compare	Location	Target %id	Query %id
Alpaca ( <i>Vicugna pacos</i> )	1-to-1	n/a	ENSVPAG00000006545 TCF12 transcription factor 12 [Source:HGNC Symbol;Acc:11623]	Multi-species view Alignment (protein) Alignment (cDNA) Gene Tree (image)	GeneScaffold_1601.375959-706503.1	98	92
Anole Lizard ( <i>Anolis carolinensis</i> )	1-to-1	n/a	ENSACAG00000014277 TCF12 transcription factor 12 [Source:HGNC Symbol;Acc:11623]	Multi-species view Alignment (protein) Alignment (cDNA) Gene Tree (image)	GL343573.1-388289-434217:-1	78	79
Anole Lizard ( <i>Anolis carolinensis</i> )	Possible ortholog	n/a	ENSACAG000000027602	Multi-species view Alignment (protein) Alignment (cDNA) Gene Tree (image)	AAWZ02036688:11541-14447:-1	73	12
Armadillo ( <i>Dasypus novemcinctus</i> )	1-to-1	n/a	ENSNDQ000000013864 TCF12 transcription factor 12 [Source:HGNC Symbol;Acc:11623]	Multi-species view Alignment (protein) Alignment (cDNA) Gene Tree (image)	GeneScaffold_3602.38370-505693.1	60	58
Bushbaby ( <i>Otolemur</i> )	1-to-1	0.19081	ENSOGAG000000006485 TCF12	Multi-species view	GL873530.1-9653822-9743765.1	95	68

click



Ensembl genome browser 65: Homo sapiens - Supporting evidence - Transcript: TCF12-201

Ensembl BLAST/BLAT BioMart Tools Downloads More

Human (GRCh37) Location: 15:57,212,696-57,216,915 Gene: TCF12 Transcript: TCF12-201

**Transcript-based displays**

- Transcript summary
- Supporting evidence (39) **click**
- Sequence
  - Exons (21)
  - cDNA
  - Protein
- External References
  - General identifiers (37)
  - Ontology
    - Ontology graph (2)
    - Ontology table (2)
- Genetic Variation
  - Population comparison
  - Comparison image
- Protein Information
  - Protein summary
  - Domains & features (12)
  - Variations (57)
- External Data
  - Personal annotation
- ID History
  - Transcript history
  - Protein history

**Transcript: TCF12-201 ENST00000452095**

Description transcription factor 12 [Source:HGNC Symbol;Acc:11623]  
 Location Chromosome 15: 57,210,833-57,580,712 forward strand.  
 Gene This transcript is a product of gene ENSG00000140262 - This gene has 31 transcripts

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
TCF12-001	ENST00000267811	6061	ENSP00000267811	682	Protein coding	CCDS10158
TCF12-002	ENST00000333725	4719	ENSP00000331067	706	Protein coding	CCDS10160
TCF12-004	ENST00000543579	1809	ENSP00000440017	536	Protein coding	-
TCF12-006	ENST00000537840	1598	ENSP00000444696	446	Protein coding	-
TCF12-007	ENST00000559222	1598	ENSP00000453216	446	Protein coding	-
TCF12-008	ENST00000343827	3956	ENSP00000342459	512	Protein coding	CCDS42042
TCF12-009	ENST00000559710	1314	ENSP00000453264	316	Protein coding	-
TCF12-010	ENST00000559703	1544	ENSP00000454102	339	Protein coding	-
TCF12-025	ENST00000438423	4786	ENSP00000388940	706	Protein coding	CCDS10160
TCF12-026	ENST00000557843	4076	ENSP00000453737	682	Protein coding	CCDS10158
TCF12-027	ENST00000557947	575	ENSP00000454109	157	Protein coding	-
TCF12-028	ENST00000561152	675	ENSP00000453653	58	Protein coding	-
TCF12-201	ENST00000452095	4772	ENSP00000396881	702	Protein coding	-
TCF12-202	ENST00000543236	2474	ENSP00000442910	681	Protein coding	-
TCF12-203	ENST00000543417	1545	ENSP00000443452	294	Protein coding	-
TCF12-003	ENST00000559609	2252	ENSP00000453876	666	Nonsense mediated decay	-
TCF12-011	ENST00000561449	914	No protein product	-	Processed transcript	-
TCF12-013	ENST00000559216	475	No protein product	-	Processed transcript	-
TCF12-014	ENST00000560887	540	No protein product	-	Processed transcript	-
TCF12-016	ENST00000561346	852	No protein product	-	Processed transcript	-
TCF12-017	ENST00000561235	818	No protein product	-	Processed transcript	-
TCF12-018	ENST00000560784	1230	No protein product	-	Processed transcript	-
TCF12-020	ENST00000560506	575	No protein product	-	Processed transcript	-

Ensembl genome browser 65: Homo sapiens - Supporting evidence - Transcript: TCF12-201

Ensembl BLAST/BLAT BioMart Tools Downloads More

Human (GRCh37) Location: 15:57,212,696-57,216,915 Gene: TCF12 Transcript: TCF12-201

**Supporting evidence**

**Supporting evidence**

**click**

Ensembl/Envan...  
 Transcript supp...  
 NM\_207037.1  
 B4DGI9  
 Exon Support...  
 NM\_207037.1  
 NM\_207036.1  
 BC050556.1  
 NM\_009205.3  
 NM\_207038.1  
 BC051769.2  
 M83233.1  
 NM\_207040.1  
 B4DGI9  
 AK312710.1  
 AK302923.1  
 AK294991.1  
 AK308101.1  
 AK302749.1  
 AK304007.1  
 AK309028.1  
 Q99081-2  
 AK311709.1  
 CN420937.1  
 AL710129.1  
 Q9N0V9  
 D8225-960.1  
 Exon Support...  
 BK001049.1  
 Q86V42.1  
 AL831980.2  
 AK294617.1  
 Q86T1.1  
 Q99081.1  
 BX537967.1  
 M80627.1  
 BM4499590.1  
 B0710370.1  
 AK311359.1  
 BM470867.1  
 M85209.1  
 DK35423.1  
 A1436599.1  
 BQ010518.1  
 CN420934.1

Legend protein evidence EST evidence cDNA evidence



Ensembl genome browser 65: Homo sapiens - Protein sequence - Transcript: TCF12

Ensembl Transcript tab: Protein sequence

Transcript ID	ENST ID	Length	Protein product	Protein length	Processed transcript
TCF12-003	ENST00000558909	2252	ENSP00000453878	666	-
TCF12-011	ENST00000561449	914	No protein product	-	-
TCF12-013	ENST00000559216	475	No protein product	-	Processed transcript
TCF12-014	ENST00000560887	540	No protein product	-	Processed transcript
TCF12-016	ENST00000561346	852	No protein product	-	Processed transcript
TCF12-017	ENST00000561235	818	No protein product	-	Processed transcript
TCF12-018	ENST00000560784	1230	No protein product	-	Processed transcript
TCF12-020	ENST00000560506	575	No protein product	-	Processed transcript
TCF12-021	ENST00000561420	837	No protein product	-	Processed transcript
TCF12-022	ENST00000558210	639	No protein product	-	Processed transcript
TCF12-023	ENST00000558908	553	No protein product	-	Processed transcript
TCF12-024	ENST00000560190	1183	No protein product	-	Processed transcript
TCF12-025	ENST00000561454	564	No protein product	-	Processed transcript
TCF12-030	ENST00000560191	1205	No protein product	-	Processed transcript
TCF12-031	ENST00000560948	440	No protein product	-	Processed transcript
TCF12-032	ENST00000560836	212	No protein product	-	Processed transcript

Transcript and Gene level displays

Views in Ensembl are separated into gene based views and transcript based views according to which level the information is more appropriately associated. To flip between the two sets of views you can click on the Gene and Transcript tabs in the menu bar at the top of the page.

Protein sequence [help](#)

Key

Exons

```

MNPQQQMAALGOTKELSDLLDPSAMFSPVNSGKTRPTTLGSSQFSGSOTIREMKQLNS
KARQKQRIKGFDSHYSDHLNDRSLGASGLSPTTFMNSLNGKTSERGSFSLYSDPT
GLPGQSSLLRQDLGLGSPAQLSSGKPGTYYSFSAATSRRRFLSDSALDPLQAKKVR
KVPFGLPSSVYAPSPNSCDFNRESFSYSPFRPTSMFASTFNGDGTNNSDLWSSMGR
SQPFGILLQTSFSDGQSSYONLASSDGLSYFPPSSVSTPTMSTLPPMSFSGSSTSS
SPYVAASHTPPIGSDSLGTRGNAGSSQGDALGKALASIVSPTISFFSPNFPSTPV
GSPFLDTGTSQMPRFGGQAPSSPYSVNSLSLKNRVEQLREHLQDAMSLFKDVCQSRM
EDRLDLDDAIVLRNHAVGSPSTLPAGESDILHSLGFPHNAPIGLSNLYNGSSSLVASS
RSASMVGTHREDSVLSNGNSVLSSTVTTSSDLENKTKQENYRGLQSQGVTVTTEIKT
ENKEDENLHEPPSSDMMKSDDESQKDIKVSRRGKTSSTNEDDLNFPQKIEREKERM
ANARSELVARDIENAFKELGMCQLHFKSKKQVLLIHLQAVAVTLSEQQVREKRLN
PKAACLRREKESVAVSAEPTTLPGRDPTLPPGSRM
    
```

Ensembl release 65 - Dec 2011 © WTS

Permanent link - View in archive site

About Ensembl | Contact Us | Help

Ensembl genome browser 65: Homo sapiens - Protein sequence - Transcript: TCF12-201 (ENST00000049)

Ensembl archive

Transcript ID	ENST ID	Length	Protein product	Protein length	Processed transcript	Notes
TCF12-003	ENST00000558909	2252	ENSP00000453878	666	-	Nonsense mediated decay
TCF12-011	ENST00000561449	914	No protein product	-	-	Processed transcript
TCF12-013	ENST00000559216	475	No protein product	-	-	Processed transcript

The following archives are available for this page:

- Ensembl 64: Sep 2011 (GRCh37) - patched/updated gene set Sep 2011
- Ensembl 63: Jun 2011 (GRCh37) - gene set updated Apr 2011
- Ensembl 62: Apr 2011 (GRCh37) - gene set updated Apr 2011
- Ensembl 61: Feb 2011 (GRCh37) - patched/updated gene set Jan 2011
- Ensembl 60: Nov 2010 (GRCh37) - patched/updated gene set Oct 2010
- Ensembl 59: Aug 2010 (GRCh37)
- Ensembl 58: May 2010 (GRCh37) - patched/updated gene set May 2010
- Ensembl 57: Mar 2010 (GRCh37) - patched/updated gene set Jan 2010
- Ensembl 56: Sep 2009 (GRCh37) - patched/updated gene set Jul 2009
- Ensembl 55: Jul 2009 (GRCh37) - gene set updated May 2009
- Ensembl 54: May 2009 (NCBI36)
- Ensembl 53: Mar 2009 (NCBI36)
- Ensembl 52: Dec 2008 (NCBI36) - patched/updated gene set Oct 2008
- Ensembl 51: Nov 2008 (NCBI36) - patched/updated gene set Sep 2008
- Ensembl 50: Jul 2008 (NCBI36)
- Ensembl 48: Aug 2007 (NCBI36)

[More information about the Ensembl archives](#)

Ensembl release 65 - Dec 2011 © WTS / EBI


Permanent link - View in archive site

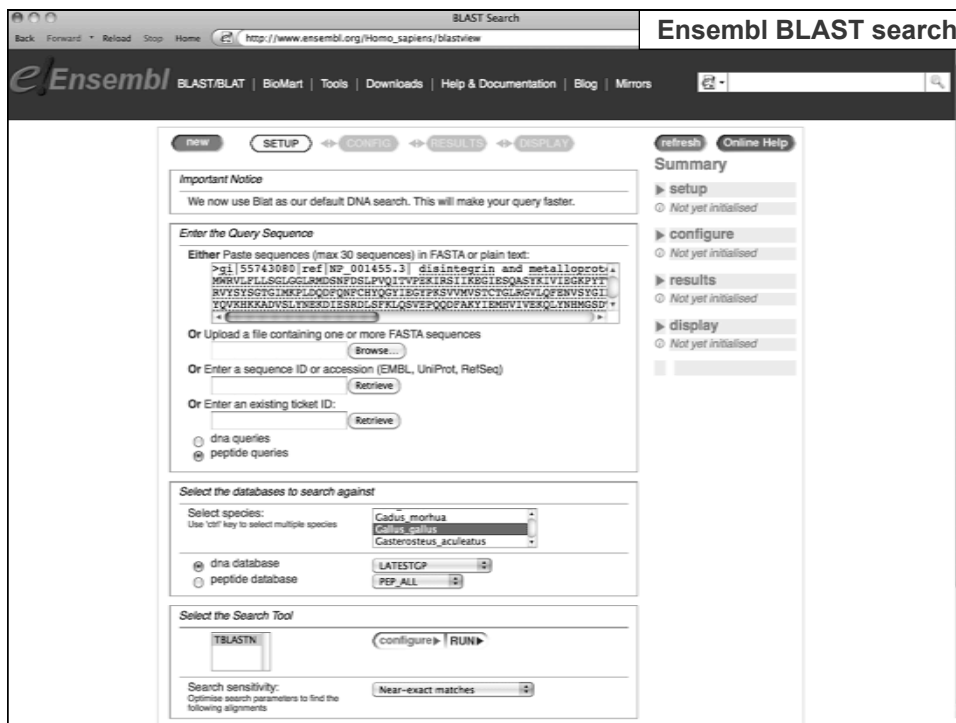
About Ensembl | Contact Us | Help

**Ensembl**

Find a chicken homolog of a human protein

<http://www.ensembl.org>

 NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
Division of Intramural Research



The screenshot shows the Ensembl BLAST search interface. At the top, there is a navigation bar with the Ensembl logo and links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. The main content area is titled "Ensembl BLAST search" and includes a "refresh" button and "Online Help" link. Below this, there is a "Summary" section with expandable options for "setup", "configure", "results", and "display", each with a "Not yet initialised" status. The main search area is divided into several sections: "Important Notice" (stating that BLAST is now the default DNA search), "Enter the Query Sequence" (with options to paste sequences, upload FASTA files, or enter sequence IDs), "Select the databases to search against" (with a species dropdown set to "Gallus gallus" and database options for "dna database" and "peptide database"), and "Select the Search Tool" (with "tblastn" selected). A "Search sensitivity" dropdown is set to "Near-exact matches".

▼ Alignment Summary (click arrow to hide)

Select rows to include in table, and type of sort (Use the 'ctrl' key to select multiples)

Query Subject Chromosome Supercontig Contig Stats Sort By

Start End Ori Start End Ori Score E-val %ID Length

**Ensembl BLAST search**

refresh

Query	Subject	Chromosome	Start	End	Ori	Score	E-val	%ID	Length
[A] [S] [G] [C]	4	669 +	Chr:8	29718636	29720642	+	1465	1.1e-129	33.05 708
[A] [S] [G] [C]	8	505 +	Chr:15	6293553	6296084	+	1194	2.7e-107	35.75 537
[A] [S] [G] [C]	278	668 +	Chr:15	6295085	6296212	+	1016	4.0e-86	37.99 408
[A] [S] [G] [C]	138	335 +	Chr:22	2488363	2488953	+	308	2.2e-70	31.19 218
[A] [S] [G] [C]	399	511 +	Chr:22	2501194	2501499	+	300	3.1e-78	39.13 115
[A] [S] [G] [C]	362	644 +	Chr:17	3075328	3076149	-	295	5.5e-21	29.28 321
[A] [S] [G] [C]	438	570 +	Chr:22	2490093	249044				
[A] [S] [G] [C]	425	659 +	Chr:1	200979512	200988				
[A] [S] [G] [C]	445	505 +	Chr:22	2479489	247961				
[A] [S] [G] [C]	445	535 +	Chr:6	3440871	344086				
[A] [S] [G] [C]	445	502 +	Chr:6	3298727	329881				
[A] [S] [G] [C]	212	270 +	Chr:22	2500007	250018				
[A] [S] [G] [C]	444	684 +	Chr:4	9312927	931291				
[A] [S] [G] [C]	339	404 +	Chr:22	2500735	250051				
[A] [S] [G] [C]	444	501 +	Chr:22	1062858	106301				
[A] [S] [G] [C]	329	456 +	Chr:6	10451971	104524				

Query location : ref|NP\_001455.3| 4 to 669 (+)

Database location : 5 29718636 to 29720642 (+)

Genomic location : 5 29718636 to 29720642 (+)

Alignment score : 1465

E-value : 1.1e-129

Alignment length : 708

Percentage identity: 33.05

Query: 4 VFLLEGLGLRMDSNFDELVPQIT----VPEKIRSIIEKIGIESQASVYKIVIEKPYTV 58

Sbjct: 29718636 VLVVLGLVCPPTDDESGPLHVGMVTVVPRQL-SPRADTNPLTVSYMLQVGRPQVL 29718812

Query: 59 NLM-QKNLPHNFRVYSYSGTGMKPLDQDF-QMPCYOGYIEGYPKSVVMVSTC-TGLR 115

Sbjct: 29718813 RLRPRKGLASPPFLVTVYDEDEGARRE-EQVTVQNCFCYOGVQSGPSGLVALTOGCRGLR 29718989

Query: 116 GVLFQFENVSYGIEPLESSVGFVHVIVQVKKKADVSLYNEK-DIESRDLSEFK-----LQ 168

Sbjct: 29718990 GVLNMGSTYIEPIPDPAFQHNLYRME---ADSDPMGPTCGLTPEELQYQKTVLPLWQ 29719160

Query: 169 S--VEPQ--QDF---AKYIEHVIIEKGLYHMGSD---TTVVAQKVFQILGILTAIFV 217

Sbjct: 29719161 APRTEDRYMLKDWHTRYVKLVVVDVNRFP--VRSORNEKVLRLQ-VLEVNNIGDGLYD 29719331

Query: 218 SFNITILSSLELWIDENKIATTOGANELLHTFLRWKTSYLVLR-PHDVAFLLVYRE--K 274

Sbjct: 29719332 QLSVQLFVGLIEWTNSNPINITKASKTLADFNRRKSDLYFRMHDHTAHLFAPQGGK 29719511

Query: 275 SNYVGTAFQGRMCDANYAGGVLEP-RTISLES LAVILQQLLSMGITDYDDINKQCQSG 333

Sbjct: 29719512 S--LGLAYLGSICDRQWSAAVDSYNNRRLS--SFIVTFVHELGNLGRHDE-RHCKCR 29719676

Query: 334 AVCCIM-NPEAIHFSQVIFNSCFEDFAHFISGQKSQLHNQPLRDPFF--KQAVCGNA 390

Sbjct: 29719677 KRCIMYSE----SDTDAFDCSYKDYFDLLGRGGS-CLYQAPALGYSYTLKRE-YCGNK 29719838

Query: 391 KLEAGBECDCGTQDCALIGETCCDIATCFRAGSNCBQPCENCLFMSKERCRPSFE 450

Sbjct: 29719839 IVEBGEQDCDGSKDCRR--DFCCB-PNCLTAGSVCAQKCKKCCQCLIPAGLICRARTG 29720009

# Ensembl

## Using BioMart to cross-reference data from different sources

<http://www.ensembl.org>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
 Division of Intramural Research

**Ensembl BioMart**  
 Get genomic coordinates, gene name, and RefSeq accessions for a list of ENSEMBL gene identifiers

**Step 1: Select Dataset**

Dataset: Danio rerio genes (Zv9)

Filters: [None selected]

Attributes: Ensembl Gene ID, Ensembl Transcript ID

Please restrict your query using criteria below

REGION:

GENE:

Limit to genes ... with Spaink Lab Leiden3 probe ID(s)

Only  
 Excluded

ID list limit

Ensembl Gene ID(s) [e.g. ENSG00000139618]

ENSDARG00000000906  
 ENSDARG00000002906  
 ENSDARG00000002507  
 ENSDARG00000004358  
 ENSDARG00000004561

Transcript count >=

Gene type

miRNA  
 misc\_RNA  
 Mt\_rRNA  
 Mt\_rRNA  
 protein\_coding

**Step 2: Select Filters (input)**

**Ensembl BioMart**  
 Get genomic coordinates, gene name, and RefSeq accessions for a list of ENSEMBL gene identifiers

Please select columns to be included in the output and hit

click

Attributes:

Ensembl Gene ID  
 Ensembl Transcript ID  
 Chromosome Name  
 Gene Start (bp)  
 Gene End (bp)  
 Associated Gene Name  
 RefSeq mRNA

Associated Gene Name  
 Associated Transcript Name  
 Associated Gene DB  
 Associated Transcript DB  
 Transcript count  
 % GC content  
 Gene Biotype

**Step 3: Select Attributes (output)**

External References (max 3)

PDB ID  
 Clone based Ensembl gene name  
 Clone based Ensembl transcript name  
 Clone based VEGA gene name  
 Clone based VEGA transcript name  
 EMBL (Genbank) ID  
 EntrezGene ID  
 VEGA transcript ID(s) (OTTT)  
 VEGA gene ID(s) (OTTG)  
 Ensembl transcript (where OTTT shares CDS with ENST)  
 HAVANA transcript (where ENST shares CDS with OTTT)  
 HAVANA transcript (where ENST identical to OTTT)  
 HGNC ID(s)  
 HGNC symbol  
 IPI ID  
 MEROPS ID  
 miRBase Accession(s)  
 miRBase ID(s)

RefSeq Protein ID  
 RefSeq Predicted Protein ID  
 Rfam ID  
 Rfam gene name  
 Rfam transcript name  
 Unigene ID  
 UniProt/TrEMBL Accession  
 UniProt/SwissProt ID  
 UniProt/SwissProt Accession  
 UniProt Gene Name  
 WikiGene name  
 WikiGene description  
 ZFIN ID  
 ZFIN symbol  
 ZFIN xpat  
 ZFIN transcript name  
 RefSeq mRNA  
 RefSeq mRNA predicted  
 RefSeq ncRNA  
 RefSeq ncRNA predicted

Ensembl BioMart  
 Get genomic coordinates, gene name, and RefSeq accessions for a list of ENSEMBL gene identifiers

Export all results to:    Unique results only

Email notification to:

View:  rows as   Unique results only

Ensembl Gene ID	Ensembl Transcript ID	Chromosome Name	Gene Start (bp)	Gene End (bp)	Associated Gene Name	RefSeq mRNA	RefSeq mRNA predicted
ENSDARG0000000906	ENSdart00000052660	16	23018783	23062136	skap2	NM_200628	
ENSDARG0000000906	ENSdart00000137344	16	23018783	23062136	skap2		
ENSDARG00000002006	ENSdart00000021596	16	20493224	20528393	nr1b	NM_131238	
ENSDARG00000002006	ENSdart00000147844	16	20493224	20528393	nr1b		
ENSDARG00000002006	ENSdart00000128914	16	20493224	20528393	nr1b		
ENSDARG00000002507	ENSdart00000138859	16	16045949	16118555	lga10		XM_0032001
ENSDARG00000002507	ENSdart0000011224	16	16045949	16118555	lga10		
ENSDARG00000004358	ENSdart0000012673	16	13722553	13799769	gr33a	NM_001002437	
ENSDARG00000004561	ENSdart00000142810	16	14722197	14861170	prfoc		XM_0019216
ENSDARG00000004561	ENSdart00000103886	16	14722197	14861170	prfoc		
ENSDARG00000004806	ENSdart00000121998	16	15611720	15622320	grwd1	NM_001003509	
ENSDARG00000005782	ENSdart00000138611	16	16979935	17345861	col14a1		
ENSDARG00000005782	ENSdart00000137912	16	16979935	17345861	col14a1		
ENSDARG00000005782	ENSdart00000027982	16	16979935	17345861	col14a1		XM_0019220
ENSDARG00000005782	ENSdart00000134087	16	16979935	17345861	col14a1		
ENSDARG00000007959	ENSdart00000137902	16	22955445	22973946	hba2b		
ENSDARG00000007959	ENSdart00000009429	16	22955445	22973946	hba2b	NM_201160	
ENSDARG00000007959	ENSdart00000134407	16	22955445	22973946	hba2b		
ENSDARG00000007959	ENSdart00000131452	16	22955445	22973946	hba2b		
ENSDARG00000009823	ENSdart00000148436	16	22143616	22239485	ankrd28		XM_684152
ENSDARG00000009823	ENSdart00000027020	16	22143616	22239485	ankrd28		
ENSDARG00000013371	ENSdart00000007842	16	14545332	14561307	isoc2	NM_001079953	
ENSDARG00000013371	ENSdart00000146997	16	14545332	14561307	isoc2		
ENSDARG00000018787	ENSdart00000159566	16	25621948	25537442	efna1b	NM_200783	
ENSDARG00000018787	ENSdart00000135279	16	25621948	25537442	efna1b		
ENSDARG00000019753	ENSdart00000131627	16	25838201	25958945	KCNJ3		XM_0019217
ENSDARG00000019753	ENSdart00000103211	16	25838201	25958945	KCNJ3		
ENSDARG00000023031	ENSdart00000009827	16	23011103	23013613	hoxa2b	NM_131106	

Ensembl BioMart:  
 Get predicted human orthologs for a list of ENSEMBL gene identifiers

Please select columns to be included in the output and

Features  Homologs  
 Structures  Variation  
 Transcript Event  Sequences

GENE:

ORTHOLOGS (Max select 3 orthologs):

**Atlantic Cod Orthologs**

Atlantic Cod Ensembl Gene ID  Atlantic Cod Chromosome End (bp)  
 Representative Protein or Transcript ID  Homology Type  
 Atlantic Cod Ensembl Protein ID  Ancestor  
 Chromosome Name  % Identity

**Human Orthologs**


Human Ensembl Gene ID  Homology Type  
 Representative Protein or Transcript ID  Ancestor  
 Human Ensembl Protein ID  dS  
 Human Chromosome  dN  
 Human Chromosome Start (bp)  % Identity  
 Human Chromosome End (bp)  Human % Identity

Ensembl Gene ID	Ensembl Transcript ID	Human Ensembl Gene ID	Human Ensembl Protein ID	% Identity
ENSDARG0000000906	ENSdart00000052660	ENSG00000005020	ENSP00000005587	58
ENSDARG0000000906	ENSdart00000137344	ENSG00000005020	ENSP00000005587	58
ENSDARG00000002006	ENSdart00000021596	ENSG000000204231	ENSP000000363812	70
ENSDARG00000002006	ENSdart00000147844	ENSG000000204231	ENSP000000363812	70
ENSDARG00000002006	ENSdart00000128914	ENSG000000204231	ENSP000000363812	70
ENSDARG00000002507	ENSdart00000138859			
ENSDARG00000002507	ENSdart0000011224			
ENSDARG00000004358	ENSdart0000012673			
ENSDARG00000004561	ENSdart00000142610	ENSG00000126583	ENSP00000263431	69
ENSDARG00000004561	ENSdart00000103886	ENSG00000126583	ENSP00000263431	69
ENSDARG00000004806	ENSdart00000121998	ENSG00000105447	ENSP00000253237	59
ENSDARG00000005782	ENSdart00000138611	ENSG00000187955	ENSP00000297848	59
ENSDARG00000005782	ENSdart00000137912	ENSG00000187955	ENSP00000297848	59
ENSDARG00000005782	ENSdart00000027982	ENSG00000187955	ENSP00000297848	59
ENSDARG00000005782	ENSdart00000134087	ENSG00000187955	ENSP00000297848	59

# NCBI

## View a genomic region between two SNPs

<http://www.ncbi.nlm.nih.gov/mapview>



NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
 Division of Intramural Research

Map Viewer query page

Back Forward Reload Stop Home <http://www.ncbi.nlm.nih.gov/mapview/>

NCBI Home GenBank BLAST

Map Viewer Home Help

The Map Viewer provides a wide variety of genome mapping and sequencing data. [More...](#)

**Search**

Search:

for:

**Tools Legend**

**News**

**Nasonia vitripennis build 2.1 released** Nov 1, 2011  
 Nasonia vitripennis build 2.1 has been released and is now a... [more](#)

**Oreochromis niloticus build Orenil1.0 released** Oct 31, 2011  
 Oreochromis niloticus build Orenil1.0 has been released and ... [more](#)

**Cavia porcellus build 1.1 released** Oct 31, 2011  
 Cavia porcellus build 1.1 has been released and is now avail... [more](#)

**Canis familiaris build 2.2 released** Oct 27, 2011  
 Canis familiaris build 2.2 has been released and is now avail... [more](#)

[Show all](#)

**Related Resources**

- NCBI Home
- NCBI Web Search
- NCBI Site map
- Genome Browser agreement
- Genome Biology
- Taxonomy
- Entrez (Global Query)
- BLAST
- Map Viewer FTP

Scientific name	Common name	Build	Tools
<b>Vertebrates (29)</b>			
<b>Mammals (22)</b>			
<b>Primates (6)</b>			
<i>Callithrix jacchus</i>	white-tufted-ear marmoset	<a href="#">Build 1.1</a>	<a href="#">Q</a> <a href="#">B</a> <a href="#">R</a>
<i>Homo sapiens</i>	human	<a href="#">Build 37.3</a>	<a href="#">Q</a> <a href="#">B</a> <a href="#">R</a> <a href="#">C</a> <a href="#">G</a>
		<a href="#">Build 36.3</a>	<a href="#">Q</a> <a href="#">B</a> <a href="#">R</a> <a href="#">C</a> <a href="#">G</a>
<i>Macaca mulatta</i>	rhesus macaque	<a href="#">Build 1.2</a>	<a href="#">Q</a> <a href="#">B</a> <a href="#">R</a> <a href="#">G</a>
<i>Nomascus leucogenys</i>	Northern white-cheeked gibbon	<a href="#">Build 1.1</a>	<a href="#">Q</a> <a href="#">B</a>
<i>Pan troglodytes</i>	chimpanzee	<a href="#">Build 3.1</a>	<a href="#">Q</a> <a href="#">B</a> <a href="#">R</a> <a href="#">G</a>
		<a href="#">Build 2.1</a>	<a href="#">Q</a> <a href="#">B</a> <a href="#">R</a>
<i>Pongo abelii</i>	Sumatran orangutan	<a href="#">Build 1.2</a>	<a href="#">Q</a> <a href="#">B</a> <a href="#">R</a>
<b>Rodents (4)</b>			
<i>Cavia porcellus</i>	Domestic guinea pig	<a href="#">Build 1.1</a>	<a href="#">Q</a> <a href="#">B</a>
<i>Cricetus griseus</i>	Chinese hamster	<a href="#">Build 1.1</a>	<a href="#">Q</a> <a href="#">B</a>
<i>Mus musculus</i>	laboratory mouse	<a href="#">Build 37.2</a>	<a href="#">Q</a> <a href="#">B</a> <a href="#">R</a> <a href="#">C</a> <a href="#">G</a>
		<a href="#">Build 36.1</a>	<a href="#">Q</a> <a href="#">B</a> <a href="#">R</a>
<i>Rattus norvegicus</i>	rat	<a href="#">RGSC v3.4</a>	<a href="#">Q</a> <a href="#">B</a> <a href="#">R</a> <a href="#">C</a> <a href="#">G</a>
<b>Monotremes (1)</b>			
<b>Marsupials (1)</b>			
<b>Other Mammals (19)</b>			
<b>Other Vertebrates (7)</b>			
<b>Invertebrates (17)</b>			
<b>Protozoa (19)</b>			
<b>Plants (118)</b>			
<b>Fungi (17)</b>			
<i>Aspergillus clavatus</i>		<a href="#">Build 1.1</a>	<a href="#">Q</a> <a href="#">B</a> <a href="#">R</a> <a href="#">G</a>
<i>Aspergillus fumigatus</i>		<a href="#">Build 2.1</a>	<a href="#">Q</a> <a href="#">B</a> <a href="#">R</a> <a href="#">G</a>
<i>Aspergillus niger</i>		<a href="#">Build 1.1</a>	<a href="#">Q</a> <a href="#">B</a> <a href="#">R</a> <a href="#">G</a>

Entrez Genome view  
 http://www.ncbi.nlm.nih.gov/projects/mapview/map\_search.cgi?taxid=9606&quer

### Map Viewer results page

NCBI NCBI Map Viewer

PubMed Nucleotide Protein Genome Gene Structure PopSet Taxonomy Help

Search for rs76552724 OR rs1326909 on chromosome(s) assembly All Find Advanced Search

*Homo sapiens (human)* genome view  
 Build 37.3 statistics Switch to previous build

Hits: 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

Search results for query "rs76552724 OR rs1326909": 4 hits

Chr	Assembly	Match	Map element	Type	Maps
8	reference	all matches	<a href="#">rs1326909</a>	SNP	Variation
		<a href="#">rs76552724</a>	<a href="#">rs76552724</a>	SNP	Variation
8	HuRef-Primary Assembly	all matches	<a href="#">rs1326909</a>	SNP	Variation
		<a href="#">rs76552724</a>	<a href="#">rs76552724</a>	SNP	Variation

Map Viewer  
 http://www.ncbi.nlm.nih.gov/projects/mapview/maps.cgi?taxid=9606&chr=8&MAP

### Map Viewer default view

NCBI NCBI Map Viewer

PubMed Entrez BLAST OMM Taxonomy Structure

Search Find Find in This View Advanced Search

*Homo sapiens (human)* Build 37.3 (Current) BLAST human sequences

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

Query: rs76552724 OR rs1326909 [clear]

Master Map: Variation Summary of Maps Maps & Options

Region Displayed: 37,566K-37,685K bp Download/View Sequence/Evidence

Rs	UniGene	Genes_seq	Variation	Map	Gene	Het	Validation	Genotypes Avail
rs76552724				▽	LTC	Inte	☆	☆
rs34410627				▽	LTC	Inte		
rs74435605				▽	LTC	Inte		
rs12545190				▽	LTC	Inte		→
rs78578713				▽	LTC	Inte		→
rs112360090				▽	LTC	Inte		→
rs74316734				▽	LTC	Inte		→
rs4739538				▽	LTC	Inte	☆	→
rs4739540				▽	LTC	Inte		→
rs4739541				▽	LTC	Inte	☆	→
rs762123				▽	LTC	Inte	☆	→
rs11781156				▽	LTC	Inte		→
rs112278207				▽	LTC	Inte		→
rs112352687				▽	LTC	Inte		→
rs6468438				▽	LTC	Inte		→
rs34967654				▽	LTC	Inte		→
rs13274161				▽	LTC	Inte		→
rs118116446				▽	LTC	Inte		→

**NCBI**

Change the maps displayed on the Map Viewer

<http://www.ncbi.nlm.nih.gov/mapview>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
Division of Intramural Research

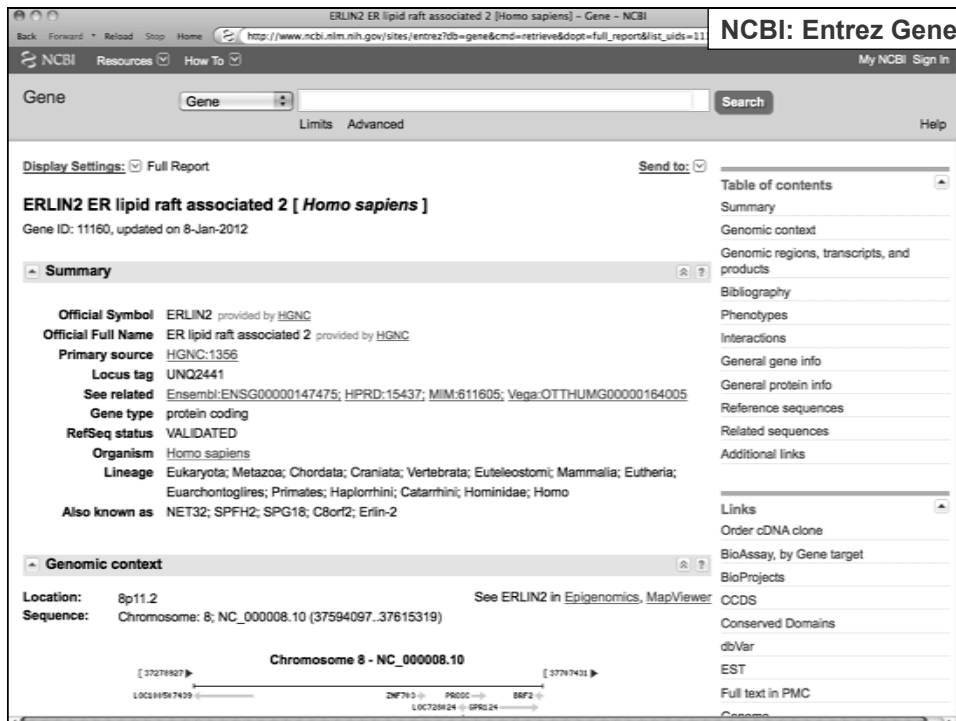
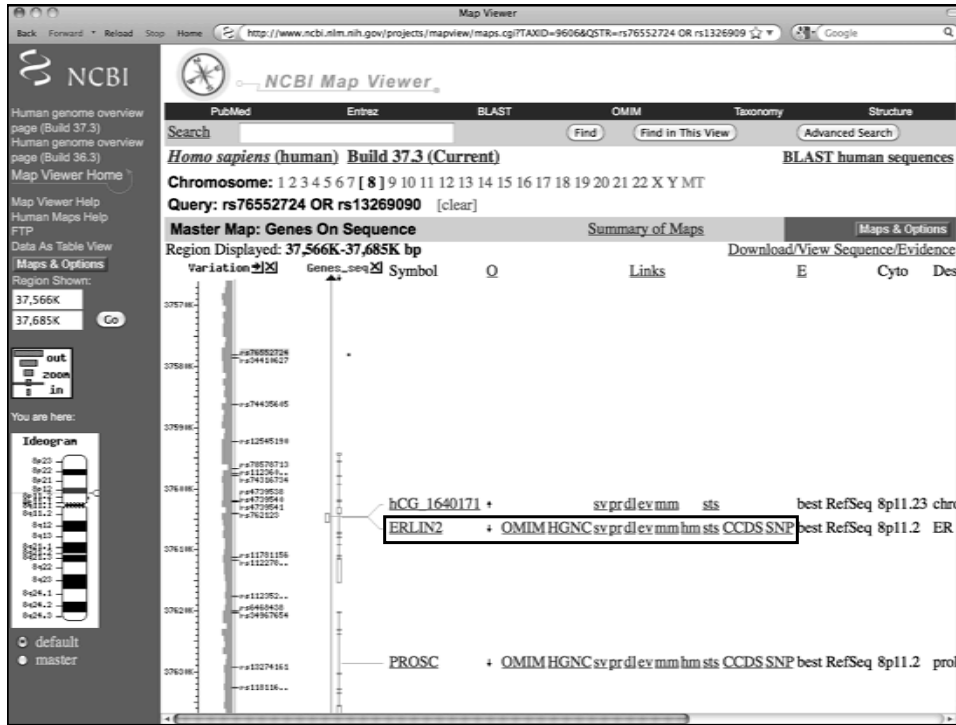
The screenshot shows the NCBI Map Viewer interface. The browser address bar displays <http://www.ncbi.nlm.nih.gov/projects/mapview/maps.cgi?taxid=9606&>. The page title is "Map Viewer Maps & Options". The "Maps & Options" dialog box is open, showing the following settings:

- Organism: **Homo sapiens**
- Chromosome: **8** Region Show: **37566100** **37685100**
- Available Tracks:
  - Sequence Maps
  - Assembly
  - Assembly regions
  - Clone
  - Component
  - Contig
  - CpG Island
  - Ensembl Genes
  - Ensembl Transcripts
  - FISH Clone (seq)
  - Cytogenetic Maps
  - Genetic Maps
  - RH Maps
- Tracks Displayed (left to right):
  - [GRCh37 p5-Primary Assembly] Variation
  - [GRCh37 p5-Primary Assembly] Gene

The dialog box has "Cancel" and "OK" buttons at the bottom right. The background shows a portion of the genome browser interface, including a zoom control and an ideogram.







ERLIN2 ER lipid raft associated 2 [Homo sapiens] - Gene - NCBI

http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=retrieve&opt=full\_report&list\_uids=11

**NCBI: Entrez Gene**

NCBI Reference Sequences (RefSeq)

RefSeqs maintained independently of Annotated Genomes

These reference sequences exist independently of genome builds. [Explain](#)

**mRNA and Protein(s)**

1. **NM\_001003790.2 → NP\_001003790.1 erlin-2 isoform 2**  
**Status: VALIDATED**

Description: Transcript Variant: This variant (2) uses different segments for its 5' UTR and for its 3' coding region and 3' UTR, compared to variant 1. The resulting protein (isoform 2) has a shorter and distinct C-terminus when it is compared to isoform 1. Variants 2 and 3 encode the same protein.

Source sequence(s): BC048308, BC067765, BI560439  
 Consensus CDS: CCDS34879.1  
 UniProtKB/Swiss-Prot: O94905

Related: ENSP00000335220, OTTHUMP00000225550, ENST00000335171, OTTHUMT00000376714

Conserved Domains (1) **summary**

<b>c02525</b>	Band_7; The band 7 domain of flotillin (reggie) like proteins. This group contains proteins similar to stomatin, prohibitin, flotillin, HfK/C and podocin. Many of these band 7 domain-containing proteins are lipid raft-associated. Individual proteins of this ...
Location: 24 - 141	
Blast Score: 633	

2. **NM\_001003791.2 → NP\_001003791.1 erlin-2 isoform 2**  
**Status: VALIDATED**

Description: Transcript Variant: This variant (3) uses a different segment for its 3' coding region and 3' UTR, compared to variant 1. The resulting protein (isoform 2) has a shorter and distinct C-terminus when it is compared to isoform 1. Variants 2 and 3 encode the same protein.

Source sequence(s): BC048308, BC067765, BP353279  
 Consensus CDS: CCDS34879.1  
 UniProtKB/Swiss-Prot: O94905

Related: ENSP00000380405, OTTHUMP00000225546, ENST00000397228

OMIM Entry - \*611605 - ENDOPLASMIC RETICULUM LIPID RAFT-ASSOCIATED PROTEIN 2; ERLIN2

http://omim.org/entry/611605

**NCBI: OMIM**

Home | About | Statistics | Downloads | Help | External Links | Copyright | Contact Us

Select Language

Search OMIM   Sort by:  Relevance  Date updated

Advanced Search: OMIM, Clinical Synopses, OMIM Gene Map  
 Search History: View, Clear

**\*611605**

**ENDOPLASMIC RETICULUM LIPID RAFT-ASSOCIATED PROTEIN 2; ERLIN2**

Alternative titles; symbols  
 SPFH DOMAIN-CONTAINING PROTEIN 2; SPFH2  
 CHROMOSOME 8 OPEN READING FRAME 2; CSORF2

HGNC Approved Gene Symbol: **ERLIN2**

Cytogenetic location: 8p11.23 Genomic coordinates (GRCh37): 8:37,594,096 - 37,615,318 (from NCBI)

**Gene Phenotype Relationships**

Location	Phenotype	Phenotype MIM number
8p11.23	Spastic paraplegia-38	611225

**TEXT**

**Cloning**

By genomic sequence analysis, followed by PCR and RACE of adult and fetal cDNA libraries, Itoigawa et al. (1999) cloned 2 splice variants of ERLIN2, which they designated CSORF2. The deduced 339- and 152-amino acid proteins share the first 141 N-terminal amino acids, then diverge. Both proteins have an N-glycosylation site and type-2 membrane topology and the longer protein has a lysine- and glutamic acid-rich region. Northern blot analysis detected ubiquitous expression of 1.6- and 2.5-kb transcripts; a minor 4.4-kb transcript was also observed.

Using monoclonal antibodies to human lipid raft proteins, Browman et al. (2006) identified ERLIN1 (611604) and ERLIN2 as components of lipid rafts. Immunohistochemical analysis of endogenous and fluorescence-tagged proteins revealed that ERLIN1 and ERLIN2 localized specifically to the endoplasmic reticulum (ER) and nuclear envelope. The 2 proteins share 83% identity, and both contain a conserved prohibitin (PHB; 176705) homology domain of about 160 amino acids.

Table of Contents - \*611605

- External Links:
- Genome
- DNA
- Protein
- Gene Info
- Variation
- Animal Models
- Cellular Pathways

NCBI: HomoloGene (hm)

Search HomoloGene for [ ] Go Clear

Display HomoloGene Show 20 Send to [ ]

All: 1 Fungi: 0 Mammals: 0

1: HomoloGene:5193. Gene conserved in Bilateria Download Links

Genes	Proteins
ERLIN2, <i>H.sapiens</i> ER lipid raft associated 2	NP_009106.1 339 aa
ERLIN2, <i>P.troglodytes</i> ER lipid raft associated 2	XP_001169738.1 339 aa
ERLIN2, <i>C.lupus</i> ER lipid raft associated 2	XP_848949.1 337 aa
ERLIN2, <i>B.taurus</i> ER lipid raft associated 2	NP_001040041.1 338 aa
Erlin2, <i>M.musculus</i> ER lipid raft associated 2	NP_705820.1 340 aa
Erlin2, <i>R.norvegicus</i> ER lipid raft associated 2	XP_214372.2 339 aa
ERLIN2, <i>G.gallus</i> ER lipid raft associated 2	XP_424380.1 342 aa
erlin2, <i>D.nerio</i> ER lipid raft associated 2	NP_001121887.1 331 aa
C42C1.15, <i>C.elegans</i> hypothetical protein	NP_502339.1 312 aa

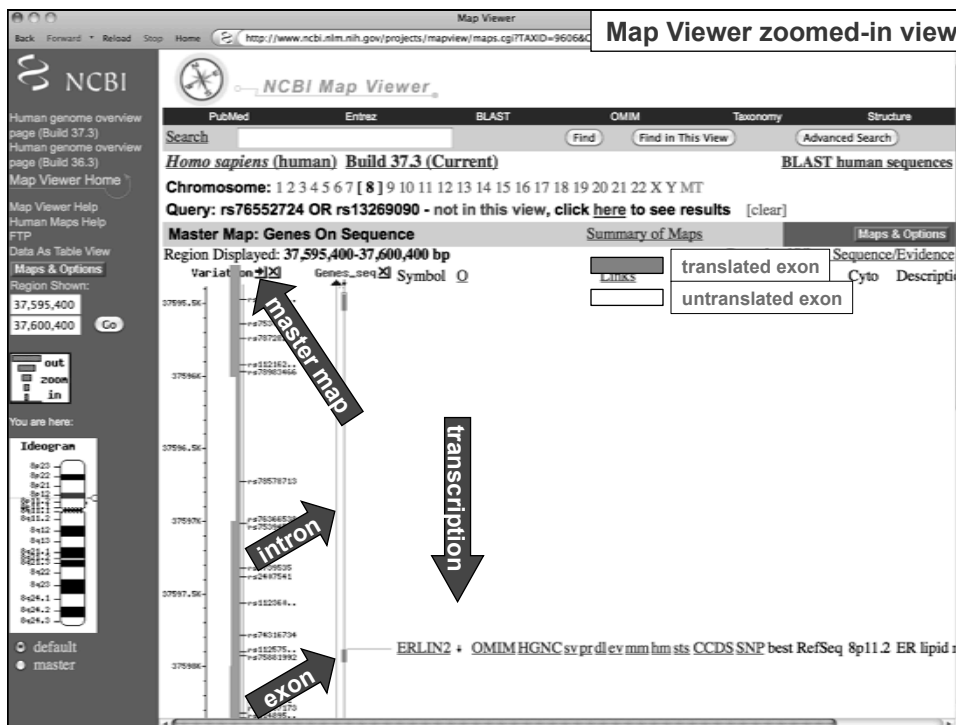
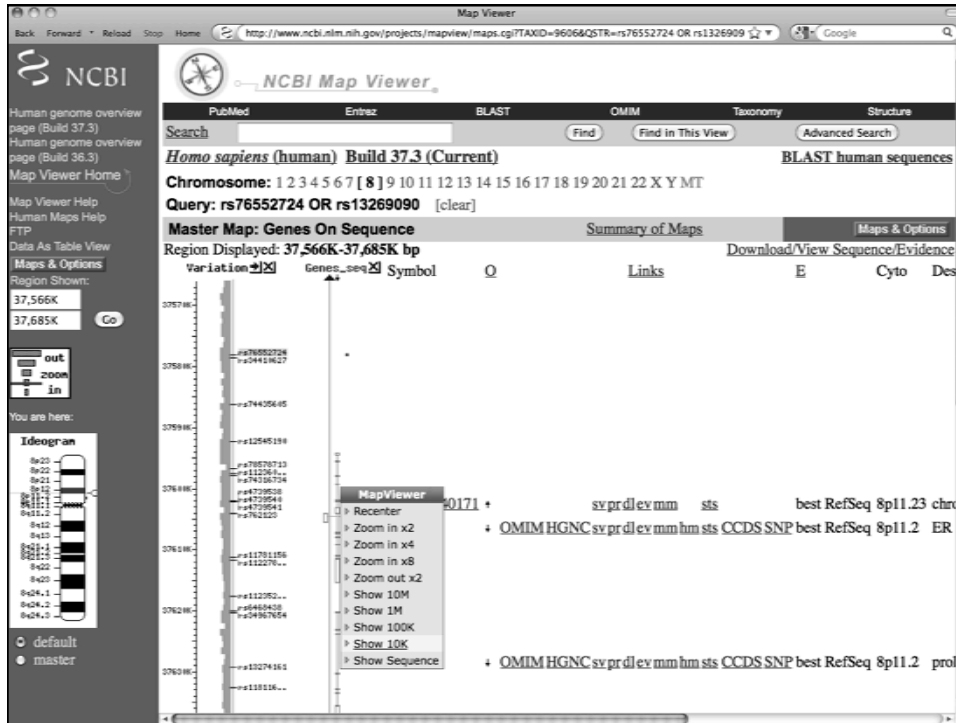
Protein Alignments Conserved Domains

**NCBI**

Zoom in to view greater detail

<http://www.ncbi.nlm.nih.gov/mapview>

NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
Division of Intramural Research



**Map Viewer SNP map**

**NCBI Map Viewer**

Human genome overview page (Build 37.3)  
 Human genome overview page (Build 36.3)  
 Map Viewer Home  
 Map Viewer Help  
 Human Maps Help  
 FTP  
 Data As Table View  
 Maps & Options  
 Region Show:  
 37,595,400  
 37,600,400  
 You are here:  
 Ideogram  
 default  
 master

**Homo sapiens (human) Build 37.3 (Current)** **BLAST human sequences**

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

Query: rs76552724 OR rs13269090 - not in this view, [click here to see results](#) [clear]

**Master Map: Variation** Summary of Maps Maps & Options

Region Displayed: 37,595,400-37,600,400 bp

Genes_seq	Variat	Map	Gene	Het	Validation	Genotypes Avail	Linkout	Avail
	rs117005490	▽	LTC	Inte				
	rs75303006	▽	LTC	Inte				
	rs78728254	▽	LTC	Inte				
	rs112162854	▽	LTC	Inte				
	rs78983466	▽	LTC	Inte				
	rs78578713	▽	LTC	Inte				
	rs76366538	▽	LTC	Inte				
	rs75394547	▽	LTC	Inte				
	rs4739535	▽	LTC	Inte				
	rs2407541	▽	LTC	Inte				
	rs112360090	▽	LTC	Inte				
	rs74316734	▽	LTC	Inte				
	rs112575270	▽	LTC	Inte				
	rs75881992	▽	LTC	Inte				
	rs2186291	▽	LTC	Inte				
	rs74657173	▽	LTC	Inte				
	rs114895254	▽	LTC	Inte				
	rs2154451	▽	LTC	Inte				

**L: Locus  
 T: Transcript  
 C: Coding region**

**NCBI: dbSNP**

**dbSNP Short Genetic Variations**

Protein Genome Structure PopSet Taxonomy OMIM Books SNP

Search for SNP on NCBI Reference Assembly

Reference SNP(refSNP) Cluster Report: rs75881992

RefSNP	Allele	HGVS Names	Links, Linkout
Organism: human (Homo sapiens)	Variation Class: SNV: single nucleotide variation	NC_000008.10:g.37597929A>G	
Molecule Type: Genomic	RefSNP Alleles: A/G	NM_001003790.2:c.154A>G	
Created/Updated in build: 131/135	Allele Origin:	NM_001003791.2:c.154A>G	
Map to Genome Build: 37.3	Ancestral Allele: A	NM_001003790.1:p.Met52Val	
Validation Status:	Clinical Source: unknown	NP_001003791.1:p.Met52Val	
	Clinical Significance: NA	NP_009106.1:p.Met52Val	
	MAF/MinorAlleleCount: G=0.001/3		
	MAF Source: 1000 Genomes		

SNP Details are organized in the following sections:  
 GeneView Map Submission Fasta Resource Diversity Validation

Function	mRNA				Protein		
	SNP to mRNA	Accession	Position	Allele change	Accession	Position	Residue change
missense	+	NM_001175.6	269	ATG => GTG	NP_009106.1	52	M [Met] => V [Val]

NC\_000008.10: 36M-36M (3.0Kbs+) Search & Go:

37,596 K 37,596,500 37,597 K 37,597,500 rs75881992 37,598

SNP  
 Genes  
 ERL12 NM\_001175.6 NP\_009106.1 NM\_001003791

## Additional resources

- **UCSC Human Genome Browser User Guide**  
<http://genome.ucsc.edu/goldenPath/help/>
- **Ensembl Tutorials and Worked Examples**  
<http://www.ensembl.org/info/website/tutorials/>
- **NCBI MapViewer Help**  
<http://www.ncbi.nlm.nih.gov/mapview/static/MapViewerHelp.html>



NATIONAL HUMAN GENOME RESEARCH INSTITUTE  
 Division of Intramural Research

## Current Protocols in Bioinformatics

### The UCSC Genome Browser

Donna Karolchik,<sup>1</sup> Angie S. Hinrichs,<sup>1</sup> and W. James Kent<sup>1</sup>

<sup>1</sup>Center for Biomolecular Science and Engineering, University of California Santa Cruz, Santa Cruz, California

UNIT 1.4

#### ABSTRACT

The University of California Santa Cruz (UCSC) Genome Browser based tool for quickly displaying a requested portion of a genome is accompanied by a series of aligned annotation "tracks." The annotation tracks are provided by the UCSC Genome Bioinformatics Group and external collaborators—genomic features, mRNA and expressed sequence tag alignments, simple nucleotide expression and regulatory data, phenotype and variation data, and pairwise comparative genomics data. All information relevant to a region of interest is displayed in one window, facilitating biological analysis and interpretation. Underlying the Genome Browser tracks can be viewed, downloaded, or interacted with using another Web-based application, the UCSC Table Browser. U

### Using the NCBI Map Viewer to Browse Genomic Sequence Data

Tyra G. Wolfsberg<sup>1</sup>

<sup>1</sup>Bethesda, Maryland

#### ABSTRACT

This unit includes a Basic Protocol with an introduction to the Map Viewer, describing how to perform a simple text-based search of genome annotations to view the genomic context of a gene, navigate along a chromosome, zoom in and out, and change the displayed maps to hide and show information. It also describes some of NCBI's sequence-

UNIT 1.5

### Using the Ensembl Genome Server to Browse Genomic Sequence Data

Xosé M. Fernández-Suárez<sup>1</sup> and Michael K. Schuster<sup>1</sup>

<sup>1</sup>EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom

UNIT 1.15

#### ABSTRACT

The Ensembl project provides a comprehensive source of automatic annotation of the human genome sequence, as well as other species of biomedical interest, with confirmed gene predictions that have been integrated with external data sources. This unit describes how to use the Ensembl genome browser (<http://www.ensembl.org/>), the public interface of the project. It describes how to find a gene or protein of interest, how to get additional information and external links, and how to use the comparative genomics tools. *Curr. Protoc. Bioinform.* 30:1.15.1-1.15.48. © 2010 by John Wiley & Sons, Inc.

Keywords: computer graphics • databases • genetic • genetic variation • genome • genome sequence • genome assembly • genomic sequence • gene map

links from the Map Viewer. The Alternate Protocols show how to perform a simple text-based search of genome annotations to view the genomic context of a gene, navigate along a chromosome, zoom in and out, and change the displayed maps to hide and show information. It also describes some of NCBI's sequence-

genome sequence, and also illustrate additional information. Alternate Protocol 1 shows how to perform and interpret the human genome. Alternate Protocol 2 demonstrates how to find a gene or protein of interest, how to get additional information and external links, and how to use the comparative genomics tools. *Curr. Protoc. Bioinform.* 30:1.15.1-1.15.48. © 2010 by John Wiley & Sons, Inc.

Access from NIH at

<http://onlinelibrary.wiley.com/book/10.1002/0471250953>